

5-1-2011

## Vol. 10, No. 1 (Full Issue)

JMASM Editors

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

---

### Recommended Citation

Editors, JMASM (2011) "Vol. 10, No. 1 (Full Issue)," *Journal of Modern Applied Statistical Methods*: Vol. 10: Iss. 1, Article 36.  
Available at: <http://digitalcommons.wayne.edu/jmasm/vol10/iss1/36>

This Full Issue is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

## Journal Of Modern Applied Statistical Methods

Shlomo S. Sawilowsky

*Editor*

College of Education  
Wayne State University

Harvey Keselman

*Associate Editor*

Department of Psychology  
University of Manitoba

Bruno D. Zumbo

*Associate Editor*

Measurement, Evaluation, & Research Methodology  
University of British Columbia

Vance W. Berger

*Assistant Editor*

Biometry Research Group  
National Cancer Institute

John L. Cuzzocrea

*Assistant Editor*

Educational Research  
University of Akron

Todd C. Headrick

*Assistant Editor*

Educational Psychology and Special Education  
Southern Illinois University-Carbondale

Alan Klockars

*Assistant Editor*

Educational Psychology  
University of Washington

## Journal Of Modern Applied Statistical Methods

### *Invited Articles*

- |        |                       |                                                                                                       |
|--------|-----------------------|-------------------------------------------------------------------------------------------------------|
| 2 – 7  | <b>Norman Cliff</b>   | A Test that Combines Frequency and Quantitative Information                                           |
| 8 – 18 | <b>Rand R. Wilcox</b> | Comparing the Strength of Association of Two Predictors via Smoothers or Robust Regression Estimators |

### *Regular Articles*

- |           |                                                                                                                                                  |                                                                                                             |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| 19 – 28   | <b>Daniel J. Mundfrom,<br/>Jay Schaffer,<br/>Myoung-Jin Kim,<br/>Dale Shaw,<br/>Ampai Thongteeraparp,<br/>Chana Preecha,<br/>Pornsin Supawan</b> | Number of Replications Required in Monte Carlo Simulation Studies: A Synthesis of Four Studies              |
| 29 – 50   | <b>Jack C. Hill,<br/>Shlomo S. Sawilowsky</b>                                                                                                    | Bias in Monte Carlo Simulations Due to Pseudo-Random Number Generator Initial Seed Selection                |
| 51 – 60   | <b>Housila P. Singh,<br/>Sunil Kumar,<br/>Sandeep Bhogal</b>                                                                                     | Estimation of Population Mean in Successive Sampling by Sub-Sampling Non-Respondents                        |
| 61 – 66   | <b>Rajesh Tailor,<br/>Balkishan Sharma</b>                                                                                                       | Improved Estimation of the Population Mean Using Known Parameters of an Auxiliary Variable                  |
| 67 – 76   | <b>Tze-San Lee</b>                                                                                                                               | Matched-Pair Studies with Misclassified Ordinal Data                                                        |
| 77 – 96   | <b>Guili Zhang,<br/>James Algina</b>                                                                                                             | A Robust Root Mean Square Standardized Effect Size in One Way Fixed-Effects ANOVA                           |
| 97 – 109  | <b>Gordon P. Brooks,<br/>George A. Johanson</b>                                                                                                  | Sample Size Considerations for Multiple Comparison Procedures in ANOVA                                      |
| 110 – 115 | <b>Sara Mann,<br/>James Chowhan</b>                                                                                                              | Weighting Large Datasets with Complex Sampling Designs: Choosing the Appropriate Variance Estimation Method |

116 – 122	<b>G. Srinivasa Rao</b>	Double Acceptance Sampling Plans Based on Truncated Life Tests for Marshall-Olkin Extended Lomax Distribution
123 – 131	<b>Ghassen El Montasser</b>	The Overall F-tests for Seasonal Unit Roots Under Nonstationary Alternatives: Some Theoretical Results and a Monte Carlo Investigation
132 – 138	<b>Tze-San Lee</b>	Fisher's Exact Test for Misclassified Data
139 – 175	<b>Ann A. O'Connell, Xing Liu</b>	Model Diagnostics for Proportional and Partial Proportional Odds Models
176 – 191	<b>Xuemei Pan, C. Mitchell Dayton</b>	Factors Influencing the Mixture Index of Model Fit in Contingency Tables Showing Independence
192 – 197	<b>Ayman Baklizi, Adil E. Yousif</b>	Inference in Simple Regression for the Intercept Utilizing Prior Information on the Slope
198 – 206	<b>Sally A. Lesik</b>	New Perspectives in Applying the Regression-Discontinuity Design for Program Evaluation: A Simulation Analysis
207 – 213	<b>S. N. Martysenko, E. A. Egorov</b>	Information Technology for Increasing Qualitative Information Processing Efficiency
214 – 225	<b>Madhusudan Bhandary, Koji Fujiwara</b>	An Exact Test for the Equality of Intraclass Correlation Coefficients Under Unequal Family Sizes
226 – 248	<b>Amery D. Wu, Bruno D. Zumbo, Linda S. Siegel</b>	General Piecewise Growth Mixture Model: Word Recognition Development for Different Learners in Different Phases
249 – 261	<b>Min Liu, Gregory R. Hancock, Jeffrey R. Harring</b>	Using Finite Mixture Modeling to Deal with Systematic Measurement Error: A Case Study
262 – 267	<b>Mahmoud M. Smadi, M. T. Alodat</b>	Bayesian Threshold Moving Average Models

268 – 276	<b>S. P. Ahmad, A. A. Khan, A. Ahmed</b>	Bayesian Regression Analysis with Examples in S-PLUS and R
277 – 286	<b>Miguel A. Padilla, Guili Zhang</b>	Estimating Internal Consistency Using Bayesian Methods
287 – 299	<b>Joseph L. Balloun, Hilton Barrett, Art Weinstein</b>	One is Not Enough: The Need for Multiple Respondents in Survey Research of Organizations
300 – 312	<b>John N. Dyer</b>	A Simulation Study of the Relative Efficiency of the Minimized Integrated Square Error Estimator (L2E) for Phase I Control Charting
313 – 321	<b>Androulla Michaeloudis</b>	Maximum Likelihood Solution for the Linear Structural Relationship with Three Parameters Known
322 – 328	<b>James F. Reed III</b>	Extension of Grizzle’s Classic Crossover Design
329 – 336	<b>Ramalingam Shanmugam</b>	Is Next Twelve Months Period Tumor Recurrence Free Under Restricted Rate Due to Medication? A Probabilistic Warning
337 – 348	<b>Narayan Chanra Sinha, M. Ataharul Islam, Kazi Saleh Ahamed</b>	Logistic Regression Models for Higher Order Transition Probabilities of Markov Chain for Analyzing the Occurrences of Daily Rainfall Data
349 – 361	<b>Ömer Eğecioğlu, Ayça Ebru Giritligil</b>	The Likelihood of Choosing the Borda- Winner with Partial Preference Rankings of the Electorate
<i>Emerging Scholars</i>		
362 – 372	<b>Albert K. Adusah*, Gordon P. Brooks</b>	Type I Error Inflation of the Separate- Variances Welch <i>t</i> test with Very Small Sample Sizes when Assumptions Are Met
373 – 402	<b>P. Borysov, Ping Sa</b>	A Robust One-Sided Variability Control Chart

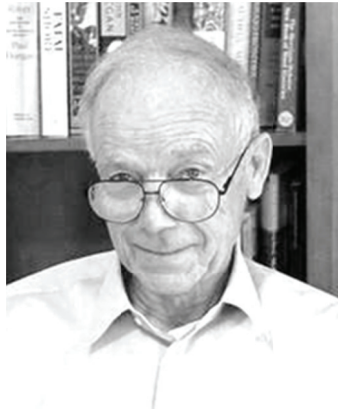
\*deceased

*JMASM* is an independent print and electronic journal (<http://www.jmasm.com/>), publishing (1) new statistical tests or procedures, or the comparison of existing statistical tests or procedures, using computer-intensive Monte Carlo, bootstrap, jackknife, or resampling methods, (2) the study of nonparametric, robust, permutation, exact, and approximate randomization methods, and (3) applications of computer programming, preferably in Fortran (all other programming environments are welcome), related to statistical algorithms, pseudo-random number generators, simulation techniques, and self-contained executable code to carry out new or interesting statistical methods.

Editorial Assistant: **Julie M. Smith**

### *Invited Articles*

## A Test That Combines Frequency and Quantitative Information



Norman Cliff  
University of Southern California,  
Los Angeles, CA USA

---

In many simple designs, observed frequencies in subclasses defined by a qualitative variable are compared to the frequencies expected on the basis of population proportions, design parameters or models. Often there is a quantitative variable which may be affected in the same way as the frequencies. Its differences among the groups may also be analyzed. A simple test is described that combines the effects on the frequencies and on the quantitative variable based on comparing the sums of the values for the quantitative value within each group to the random expectation. The sampling variance of the difference is derived and is shown to combine the qualitative and quantitative aspects in a logical way. A version of the test based on enumeration of the possible values of the sum is described, an example is analyzed, factors affecting the test's power are discussed and extensions are suggested.

Key words: Combined effects, frequencies, means.

---

#### Introduction

A method is proposed that combines two simple analytical paradigms dating from the early 20<sup>th</sup>

---

Norman Cliff is Professor Emeritus in the Department of Psychology. Norman Cliff has degrees in psychology from Wayne State and Princeton Universities. He was the longtime head of the Quantitative Psychology Program at the University of Southern California. He has over 100 research publications and three books. He is a Fellow of the American Association for the Advancement of Science. E-mail him at: nrcliff5@q.com.

century. In one, the frequencies with which observations fall into each of a set of categories are compared to the frequencies that are expected under the hypothesis of an a priori model. The discrepancies are analyzed according to the elementary Chi-square test. In the other paradigm, the groups are compared on a quantitative variable through a *t*-test or analysis of variance. The Chi-square and the *F*- or *t*-test are among the most elementary of inferential methods and hundreds of sources describe them. Alternatively, a more recent variant, such as a loglinear model (e.g., Agresti, 1983), can be used instead of the Chi-square, and location comparisons can be made by more

modern methods such as those described by Wilcox (1990, 2012) or an ordinal alternative (Cliff, 1993, 1996). The method described herein combines the frequency and quantitative information into a single test.

A common simple research context is one in which a population can be divided into  $g$  subpopulations, each with known relative frequency or theoretical probability  $\pi_i$ . Some members of the subpopulations then fall into a certain class, and the research question is whether members of the class come disproportionately from different subpopulations. A simple example is the hoary beads-of-different-colors-in-a-bag from which a random sample of beads is drawn and various questions about the contents of the bag can be investigated. Empirical examples might include investigating whether individuals with a certain disease come disproportionately from different city precincts, ethnicities, age groups, etc., each of whose population sizes are known; whether students taking high school advanced placement exams tend to come differentially from different schools, genders, etc.; whether psychotics tend to come from certain neighborhoods; whether the number of germinating plants tend to come from certain seed stocks, or netted fish tend to come from certain stocked batches or subspecies; whether defaulting mortgages tend to come from certain banks. Alternatively, there could be a model that determines the  $\pi$ .

Experimental contexts can also occur. Suppose three different inoculation regimens are employed, each on a large group. At a later time, incidences of the disease are recorded and the number that comes from each treatment is compared to its expected frequency based on the original group sizes. In all these cases it is assumed that there is an a priori probability that a random member of the observed class will come from subpopulation  $i$ . The observed number  $f_i$  that comes from  $i$  can be compared to  $n\pi_i$ , where  $n$  is the total number observed to fall in the class and elementary significance tests are applied to the results.

This article elaborates on such methods to cases where there is also an expected effect on an associated quantitative variable, specifically, for data such as: numbers written on the beads-in-a-bag; a measure of severity of disease; scores

of students on an exam; the measured sizes of plants or fish; amounts or dates of loan defaults. The groups could be compared quantitatively using some form of location comparison, such as, analysis of variance,  $t$ -test, modern more robust methods (Wilcox, 1990, 2012) or ordinal comparisons (Cliff, 1993, 1996).

Can the quantitative and qualitative information in testing a random model be combined? The traditional way this might be accomplished is to divide the quantitative variable into categories to form a cross-classification and then calculate expected cell frequencies or fit a loglinear model, etc. The qualitative variable could also be coded in some rational way and treated in parallel with the quantitative one via the general linear model. Here, combining quantitative and qualitative data more directly is suggested.

#### New Test Description

There are two beads-in-a-bag models to which the method can be applied. In the first, there is a large sack containing red and white beads. The supplier indicates that some beads, an equal number of red and white, have numbers written on them, and that the means of the red-bead numbers and white-bead numbers are the same. A sample of beads is taken, discarding those that do not have numbers, resulting in  $n$  numbered beads, some red and some white. The goal is to test the supplier's assurance of equal frequencies and equal means. A priori probabilities,  $\pi_r$  and  $\pi_w$  of 0.50, state that a numbered bead is white or red and the further hypothesis is that the means are the same for both red and white. In the general case, the a priori probabilities could be different, and/or there could be more than two colors of beads.

The second bead model uses two bags of beads, one red and one white. By hypothesis, equal proportions from red and white are numbered and the means of the numbers from red and white are equal. In this model, the plan is to sample  $s_w$  from the white bag and  $s_r$  from the red bag, once again discarding any unnumbered beads, and to determine how many of each are numbered and what the numbers are: if red and white beads are equally likely to be numbered, the probability that a numbered bead is white is  $s_w/(s_w + s_r)$ . The objective is to test



## A TEST THAT COMBINES FREQUENCY AND QUANTITATIVE INFORMATION

the combined hypothesis that the probabilities are as assumed and that the means are equal and the method generalizes to more than two colors.

The natural way to test either of the models is to calculate  $\sum_j x_{ij}$ , the sum of the scores  $x_{ij}$  by the  $j^{\text{th}}$  member of subpopulation  $i$  who are in the class, that is, are a numbered bead, and compare it to a random expectation. Here, the obvious candidate is  $n\pi_i m$ , where  $m$  is the overall mean of  $X$ ,  $n$  is the number falling in the class (numbered beads in the examples), and  $\pi_i$  is the a priori probability that the member came from subpopulation  $i$ . The difference,  $\sum_j x_{ij} - n\pi_i m$ , is a random variable that can be expected to be approximately normal under a wide variety of circumstances.

In order to assess whether the deviation could be consistent with a random model, it is essential to know the standard error of this difference. Its sampling variance,  $d_i^2$ , is  $E[(\sum_j x_{ij}) - n\pi_i m]^2$ . To determine its form, first consider the expectation at a fixed cell frequency  $f_i$  and make use of  $\sum_j x_{ij} = f_i m_i$  where  $m_i$  is the mean of the  $x_{ij}$  in  $i$ . The expected value of  $d_i^2$  at a given  $f_i$  is

$$E(d_i^2) = E[f_i^2 m_i^2 - 2f_i m_i n\pi_i + (n\pi_i m)^2],$$

and, because  $m_i$  is the mean of  $f_i$  cases,

$$E(f_i^2 m_i^2) = f_i^2 \mu^2 + f_i \sigma_x^2.$$

Next, take the expectation across the possible sample values of  $f_i$ ; where  $\mu^2$  and  $\sigma_x^2$  are constants, and  $f_i$  is a binomial in  $\pi_i$  and  $n$ . Thus, because the expected value of a squared random variable is again the sum of its squared mean and its variance,

$$E[f_i^2] = (n\pi_i)^2 + n(\pi_i - \pi_i^2).$$

Putting this back into  $d_i^2$  and collecting terms yields:

$$E(d_i^2) = n\mu^2(\pi_i - \pi_i^2) + n\pi_i \sigma_x^2.$$

This is exactly what one would expect: that the expected squared deviation under the null hypothesis is the sum of a term reflecting the expected deviation of the frequency from

expectation and one reflecting the expected deviation of the subgroup mean from the overall mean. Under the null hypothesis, the two deviations are independent; their terms are therefore additive.

Under broad conditions, that is, when  $n\pi_i$  is not too close to either  $n$  or zero and  $X$  is not far from normal with homogeneous variances across groups, the deviations  $\sum_j x_{ij} - n\pi_i \mu$  are approximately normal with the given variance, in this case the obvious test is to compute the ratio of the observed difference to  $d_i$ . In the application that this method was developed to solve,  $X$  was the first  $n$  integers so that  $\mu$  and  $\sigma^2$  were known parameters – in which case the ratio can be taken as a standard normal deviate.

However, in most applications  $m$  and  $s^2$  are estimates from the sample, the latter being a within-cells estimate. As was noted,  $d^2$  has two components, one identical to the denominator of the Chi-square test and one derived from the variance. When the latter is a sample estimate, the ratio is no longer a normal deviate, but tends to resemble a  $t$ -ratio to some degree. (Note that the unbiased estimate of  $\mu^2$  is  $m^2 - s^2/n$ .) Consequently, a slightly conservative approach is to interpret the ratio as a  $t$  with  $n - k$  df,  $k$  being the number of groups, although the expectation is that, in most contexts, the null sampling distribution may be very close to normal due to the influence of the first term in  $d^2$ . The method can be adapted to situations where  $n\pi_i$  is close to the extremes, offering some special advantages over simply comparing frequencies under those circumstances.

### Example

Table 1 contains artificial data that is used to illustrate the procedure. The data are analogous to what might be found if two groups of animals are given different cancer treatments. After a time the occurrence and size of lesions are determined, so  $x_{ij}$  is the size of the lesion in animal  $j$  from group  $i$ ; originally, there were  $s_1 = 15$  animals in treatment 1 and 10 in treatment 2, so the a priori probabilities that a given lesioned animal is in a given group are  $\pi_1 = 0.6$  and  $\pi_2 = 0.4$ , analogously to the second bead example. The expectation is that lesions will be more

common and larger in Group 1. Thirteen animals are found to have lesions in Group 1 and three in Group 2, so  $n = 16$ . The sizes of the lesions in each group are given in the upper part of the table along with the statistics for each group.

The lower part of the Table shows the components of  $d_i^2$  and the  $t$ 's for each group, which are found to be significant at the  $\alpha = 0.05$  level, one-tailed. A SAS macro was written to perform the analysis (by Professor Du Feng), but it is easily carried out in small samples with the aid of a pocket calculator. An analysis based on the rank-order version of the data gave highly similar results.

Power Considerations

It would seem natural to expect that including quantitative information would increase power over the simple frequency analysis, but one may wonder about the circumstances under which this might actually be true. Note that  $d_i^2$  has the appearance of combining expected frequency deviations and subgroup mean deviations, by adding these two components  $f_i m_i - n \pi_i m$  can be made into a form:

$$f_i m_i - n \pi_i = [f_i m_i - f_i m] + [f_i m - n \pi_i m].$$

After squaring the second bracketed term, call it  $a^2$ , and comparing it to the frequency part of  $d_i^2$ , their ratio would give exactly the same result as would be obtained in computing the  $t^{th}$

component of the Chi-square for testing observed frequencies; thus, the frequency component of this test is similar to the traditional test.

The mean difference component resembles a component of the F-test on mean differences, but is not identical. Dividing  $f_i^2(m_i - m)^2$  by  $f_i s^2$  would give a component of F, but the corresponding term in  $d_i^2$ ,  $n \pi_i$ , is the expected frequency, not  $f_i$ , of the observed group size itself, thus, these terms are similar, but are not the same.

However, the general circumstances under which using the combined test would be more powerful than simply using the frequencies can still be investigated. If  $b$  is defined as  $f_i[m_i - \mu]$  and  $e^2$  as the variance part of  $E(d_i^2)$ , then the ratio from the combined test is  $(a + b)^2/(c^2 + e^2)$ . The new ratio will be greater than the frequency ratio when

$$(a + b)^2/(c^2 + e^2) > a^2/c^2,$$

and, collecting some terms, this will be true when

$$(2ab + b^2)/e^2 > a^2/c^2.$$

This relation indicates that the new procedure is more likely to detect effects than simply testing the frequencies when both the mean and frequency effects are in the same direction as well as when the mean effect is relatively large.

Table 1: Artificial Animal Data to Illustrate the Combined Frequency and Quantitative Test

Group	Data	Statistics		
1	13.1, 2.5, 9.2, 6.2, 15.0, 12.1, 10.4, 17.4, 15.1, 6.0, 16.0, 6.1, 11.2	$\bar{x} = 11.55$		
2	3.1, 9.3, 8.6	$\bar{x} = 7.00$ $m = 10.69$ $s^2 = 16.86$		
Group	Analysis			
	$\sum x_{ij}$	$n \pi_i m$	$d_i^2$	$t$ -ratio
1	150.1	102.66	578.06	1.973
2	21.0	68.44	531.75	2.057

## A TEST THAT COMBINES FREQUENCY AND QUANTITATIVE INFORMATION

Another consequence of examining the ratio in this way is seeing that its two aspects are implicitly weighted by the relative magnitudes of variance and squared mean. The other factors,  $\pi_i - \pi_i^2$  and  $\pi_i$ , are similar in magnitude, their ratio being between 0.5 and 1.0. When the data consist of the first  $n$  positive integers, the ratio of squared mean to variance approaches 3.0 as  $n$  increases, indicating that frequency effects will always be emphasized relative to mean effects in such data.

The difference in influence can be even greater with some psychological variables whose mean and variance are set by convention. Many scholastic aptitude tests are scaled to have a mean about 500 and variance about 10,000, giving a ratio of about 25.0; the IQ scale is even more extreme, giving a ratio of squared mean to variance of more than 40.0. In such circumstances, the mean part of the proposed ratio has little effect because the proposed ratio approaches the traditional one for frequencies as  $\mu^2/\sigma^2$  increases.

In some research contexts,  $X$  has a well-established and empirically meaningful zero point. However, in others, such as the SAT and IQ scales, it merely represents a convenient reference. Where the origin of the scale is arbitrary, the user may feel that it is justifiable to give more nearly equal a priori weights to  $\mu^2$  and  $\sigma^2$ . However, it seems desirable that the lowest possible  $\sum x_{ij}$  value should be zero, occurring when  $f_i = 0$ . Thus, subtracting a constant to make the lowest observed score slightly positive seems to be the most that can be done to equate influences. However, if  $X$  is quasi-normal with lowest standardized value of around  $-3.5$  or  $-3.0$ , the ratio is still 9.0 to 12.0. Thus, making the analysis ordinal by converting the observed variable to the first  $n$  integers may be the most that can be done in equating influences of mean and variance.

### Exact Version

When  $n\pi_i$  is smaller than about five, the normality of the distribution of differences is likely to break down, making the assumed boundaries for an acceptance region unrealistic. In that circumstance, the researcher can construct cutoff values for the sum that

correspond nearly exactly to a given rejection probability. These probabilities are now defined under a randomization hypothesis rather than on the basis of parameter estimates.

A given set of  $n x_{ij}$  values, that is, from all groups in the sample, defines  $2^n$  possible values for  $\sum x_{ij}$ ; of these, a certain fraction, corresponding to the desired rejection level, give the smallest (largest) values for the sum. These can be enumerated; if the obtained sum falls within this set, the null hypothesis is rejected. This enumeration process may improve power in such cases by defining a finer-grained rejection region than the corresponding test that is based only on the frequencies or only on the means.

The method is suggested by the beads-in-a-bag models. Consider an obtained sum for Group  $i$  and ask: What is the probability of obtaining a sum this small (large) or smaller (larger) when drawing  $n$  times with probability  $\pi_i$ ? To illustrate with the example, the sum for Group 2 is 21.0,  $n$  is 16 and  $\pi_2$  is 0.40.

There are  $2^{16} = 65,336$  possible outcomes of randomly drawing a sum. Which are less than 21.0 and what are their respective probabilities? Of these outcomes, one has a sum of 0.0, that with  $f_2 = 0$ . This will happen with binomial probability  $2.82 \times 10^{-4}$ . There are 16 draws with  $f_2 = 1$ , each with probability  $1.88 \times 10^{-4}$ , and all have sums less than 21.0. There are 120 with  $f_2 = 2$ , all with probability  $1.25 \times 10^{-4}$ , but only 55 of them have sums less than 21.0. When  $f_2 = 3$ , there are only 23 that are less than 21.0, each having probability  $8.35 \times 10^{-5}$ . No combination of four has a sum below that limit.

Summing the probabilities of the instances that have sums less than 21.0 it is found that, under randomization,  $0.000282 + 16 \times 0.000188 + 55 \times 0.000125 + 23 \times 0.0000835 = 0.012192$  is the probability of obtaining a sum of 21.0 or less for Group 2, which is just short of the 0.01 significance level. By contrast, if only the frequencies are considered, the corresponding binomial probability of  $f_2 = 3$  or fewer is 0.0652. Also, the  $t$ -test in Table 1 yielded a significance level of about 0.04, less extreme than the probability obtained by enumeration.

## Applications

Applied contexts having the characteristics that are appropriate to the method seem likely to be fairly common. Consider a state infectious disease-monitoring agency that observes an outbreak of a disease such as meningitis, and tabulates the locations, by district, of the disease. It might hope to identify the origin of the outbreak by tabulating frequency by district and comparing them to expectations based on district sizes. Here,  $n$  is the total number of meningitis cases and the  $\pi_i$  are defined by the relative sizes of the populations of the different districts. If the agency records the days since diagnosis of each case and uses it as the quantitative variable in the present method, an easier identification of the outbreak's focus may be possible.

Consider also a bank-regulating agency such as the Federal Deposit Insurance Corporation that is observing a group of banks to assess the riskiness of their policies. It knows the number of mortgages issued by the banks and records the defaults that occur for each,  $n$  being the total number of mortgages that are in default and the  $p_i$  are defined by the number of mortgages issued by each bank. Using either the days since default or the amount of the default as well as the frequency of default might well give a more sensitive measure of the banks' statuses than frequency alone.

In psychology, suppose individuals are given training in problem-solving. After training, they and a control group are given a problem to solve under a time-limit. Some individuals are successful and some not,  $n$  being successful, and the time taken to success is recorded. If there are  $s_t$  individuals in the trained group and  $s_c$  in the control,  $\pi_t = s_t/(s_t + s_c)$ , and similarly for  $\pi_c$ , represent the a priori probabilities that a success comes from the respective groups. Here, in order for the time variable to operate in the appropriate direction, it is best recorded as time remaining before the cut-off signal in order that small means and small frequencies are expected to go together.

In a study of differences in criminal recidivism, released convicts who have been under different prison regimens or treatments or who belong to different natural groups can be followed for a period. The frequency of re-

incarceration can be combined with the length of sentence and analyzed in the proposed way. The method could also be applied to studies of the effects of educational treatments.

Many other potential applications exist; the key to the relevance of the method is the expectation that frequency and some quantitative variable will act in the same direction. It has been noted that treating the quantitative variable as a rank order may have some advantages.

It has been assumed that the qualitative variable consists of a single dimension of classification, but it seems in principle that this limitation is not necessary. The classification could have two or more ways as in a factorial or nested design and the relevant quantities could be computed for various effects. Another possible complication is dealing with more than one quantitative variable. Could the variables be combined by forming an optimally weighted composite of the observed variables? That optimization might be complicated by the necessity of keeping the composite positive. Investigation of such a possibility is beyond the scope of the present article.

## Acknowledgements

This article benefited from discussions with and suggestions by Texas Tech University Professors Clyde Martin, Department of Mathematics and Statistics, e-mail: clyde.f.martin@ttu.edu; Du Feng, Human Development and Family Studies, e-mail: du.feng@ttu.edu; and Rand Wilcox, Department of Psychology, University of Southern California, e-mail: rwilcox@wilcox.usc.edu.

## References

- Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.
- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114, 494.
- Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Erlbaum.
- Wilcox, R.W. (1990). Comparing the means of independent groups. *Biometrical Journal*, 32, 771-780.
- Wilcox, R. W. (2012). *Statistics for the social and behavioral sciences*. New York: CRC Press.

## Comparing the Strength of Association of Two Predictors via Smoothers or Robust Regression Estimators



Rand R. Wilcox  
University of Southern California,  
Los Angeles, CA USA

---

Consider three random variables,  $Y$ ,  $X_1$  and  $X_2$ , having some unknown trivariate distribution and let  $\eta_j^2$  ( $j = 1, 2$ ) be some measure of the strength of association between  $Y$  and  $X_j$ . When  $\eta_j^2$  is taken to be Pearson's correlation numerous methods for testing  $H_0 : \eta_1^2 = \eta_2^2$  have been proposed. However, Pearson's correlation is not robust and the methods for testing  $H_0$  are not level robust in general. This article examines methods for testing  $H_0$  based on a robust fit. The first approach assumes a linear model and the second approach uses a nonparametric regression estimator that provides a flexible way of dealing with curvature. The focus is on the Theil-Sen estimator and Cleveland's LOESS smoother. It is found that a basic percentile bootstrap method avoids Type I errors that exceed the nominal level. However, situations are identified where this approach results in Type I error probabilities well below the nominal level. Adjustments are suggested for dealing with this problem.

Key words: Explanatory power, Theil-Sen estimator, nonparametric regression, bootstrap methods, kernel density estimators.

---

### Introduction

Consider three random variables,  $Y$ ,  $X_1$  and  $X_2$  having some unknown trivariate distribution

and let  $\eta_j^2$  be some measure of association between  $Y$  and  $X_j$  ( $j=1, 2$ ). This article considers the problem of testing

$$H_0 : \eta_1^2 = \eta_2^2 \quad (1.1)$$

---

Rand R. Wilcox is a Professor of Psychology. He is the author of eight textbooks on statistics, including *Modern Statistics for the Social and Behavioral Sciences* (2012, New York, CRC Press). Email him at: [rwilcox@usc.edu](mailto:rwilcox@usc.edu).

when  $\eta_j^2$  is a robust version of explanatory power, which is estimated via the Theil (1950) and Sen (1968) regression estimator or the robust version of Cleveland's (1979) smoother

(LOESS). For the special case where  $\eta_j^2$  is Pearson's correlation,  $\rho$ , numerous methods for testing  $H_0 : \rho_1 = \rho_2$ , as well as  $H_0 : \rho_1^2 = \rho_2^2$ , have been proposed by many authors (Hittner, May & Silver, 2003; Hotelling, 1940; Olkin, 1967; Dunn & Clark, 1971; Meng, Rosenthal & Rubin, 1992; Steiger, 1980; Wilcox & Tian, 2008; Wilcox, 2009; Williams, 1959; Zou, 2007). A general concern, however, is that  $r$  – the usual estimate of  $\rho$  – is not robust, roughly meaning that even a single outlier can result in a large value for  $r$  when there is little or no association among the bulk of the points. Similarly, a strong association among the bulk of the points can be masked by one or more outliers (Wilcox, 2005). Thus,  $\rho$  is not robust in the general sense as summarized by Huber (1981) and as illustrated by Wilcox (2005, p. 385).

Another concern is curvature. Experience with smoothers indicates that approximating the regression line with the usual parametric models can be unsatisfactory, which in turn raises concerns about how to measure the overall strength of association. A relatively simple strategy is to approximate the regression line with some type of nonparametric regression estimator or smoother (e.g., Efromovich, 1999; Eubank, 1999; Fan & Gijbels, 1996; Fox, 2001; Green & Silverman, 1993; Györfi, et al., 2002; Härdle, 1990; Hastie & Tibshirani, 1990) that can be used to estimate a robust measure of the strength of the association; this is the approach employed herein.

It is noted that there is a vast literature on identifying and ordering the importance of predictor variables; see for example Lafferty and Wasserman (2008) and the references they cite. It seems that none of these methods are based on a robust measure of association. Moreover, the precision of the resulting ordering is typically unclear. Thus, an additional goal of this research is to consider a formal hypothesis testing approach for determining which of two predictors has a stronger association with the outcome variable of interest, in contrast to merely estimating which has the stronger association.

## Background

Basic results and methods used to measure and estimate the strength of an association are first reviewed. Consider the situation where the conditional mean of  $Y$ , given  $X$ , is assumed to be  $Y = \beta_0 + \beta_1 X$  and ordinary least squares is used to estimate the unknown slope,  $\beta_1$ , and intercept,  $\beta_0$ . Let  $\hat{Y} = b_0 + b_1 X$ , where  $b_0$  and  $b_1$  are the least squares estimates of  $\beta_0$  and  $\beta_1$ , respectively, based on the random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ . It is well known (and readily verified) that

$$r^2 = \frac{\hat{\sigma}^2(\hat{Y})}{\hat{\sigma}^2(Y)}, \quad (2.1)$$

where  $\hat{\sigma}^2(\hat{Y})$  is the usual sample variance based on  $\hat{Y}_i = b_0 + b_1 X_i$ ,  $i = 1, \dots, n$ . Slightly extending an approach to measuring the strength of an association used by Doksum and Samarov (1995), there is a simple and seemingly natural robust generalization of  $r^2$ . First, replace  $\hat{Y}$  with  $\tilde{Y}$ , where  $\tilde{Y}$  is any fit to the data, which might be obtained via a robust regression estimator (using a linear model) or some smoother that deals with curvature in a reasonably flexible manner. Next, let  $\tau^2(Y)$  be some robust measure of variation associated with the marginal distribution of  $Y$ . It is assumed that  $\tau^2(Y)$  has been chosen so that if there is no variation,  $\tau^2(Y) = 0$ . A general approach to measuring the strength of the association between  $Y$  and  $X$  is then

$$\eta^2 = \frac{\tau^2(\tilde{Y})}{\tau^2(Y)} \quad (2.2)$$

which Doksum and Samarov (1995) call explanatory power. To make  $\eta^2$  practical, there are the issues of choosing  $\tau$  and some method for computing  $\tilde{Y}$ . First consider  $\tau$ . There are many robust alternatives to the usual variance (Wilcox, 2005). Lax (1985) compared the

## COMPARING THE STRENGTH OF ASSOCIATION OF TWO PREDICTORS

efficiency of many scale estimators and concluded that two so-called A-estimators are best, one of which corresponds to what Shoemaker and Hettmansperger (1982) term the percentage bend midvariance. The other A-estimator found to perform well by Lax corresponds to what Shoemaker and Hettmansperger call the biweight midvariance.

Bickel and Lehmann (1976) argue that if both  $X$  and  $Y$  have symmetric distributions about zero, and if  $|X|$  is stochastically larger than  $|Y|$ , then it should be the case that a measure of scale should be larger for  $X$  than it is for  $Y$ . Bickel and Lehmann define a measure of scale that satisfies this property to be a measure of dispersion. Shoemaker and Hettmansperger show that the percentage bend midvariance is a measure of dispersion but the biweight midvariance is not. A possible reason for preferring the biweight midvariance is that it has the highest possible breakdown point, namely .5. Here the focus is on the percentage bend midvariance, but this is not to suggest that all other measures of scale be eliminated from consideration.

Recently, Randal (2008) expanded on Lax's study and concluded that the two A-estimators recommended by Lax perform relatively well. However, Randal's study did not include Rocke's (1996) TBS (translated biweight S) estimator, and the tau measure of scale introduced by Yohai and Zamar (1988). As a partial check on the relative merits of these estimators, simulations based on 5,000 replications were used to estimate the standard error of the logarithm of these estimators when  $n = 20$  for the same distributions used by Lax and Randal. (For this study the tau estimator was computed as described by Marrona & Zamar, 2002.) For a standard normal distribution, the results were 0.402, 0.388 and 0.530 for the percentage bend midvariance, tau and TBS, respectively.

For a 1-wild distribution (generate data from a normal distribution and multiply one value by 10), the results were 0.398, 0.420 and 0.516. For a slash distribution ( $Z / U$ , where  $Z$  has a standard normal distribution and  $U$  a uniform distribution), the results were 0.744, 0.631 and 0.670. No single estimator dominates.

Although the focus here is on the percentage bend midvariance, it seems that the tau measure of scale deserves serious consideration based on these limited results.

For a random sample  $Y_1, \dots, Y_n$ , the percentage bend midvariance is computed as follows. Let  $f$  be the value of  $(1 - \beta)n + .5$  rounded down to the nearest integer. The parameter  $\beta$  determines the finite breakdown point of the percentage bend midvariance, meaning the proportion of points that must be altered to make the estimate arbitrarily large. Roughly,  $\beta$  reflects the proportion of outliers that can be tolerated. Here  $\beta = 0.2$  is used, which is motivated in part by the desire to obtain good efficiency under normality. Let  $W_i = |Y_i - M|$ ,  $i = 1, \dots, n$ , and let  $W_{(1)} \leq \dots \leq W_{(n)}$  be the  $W_i$  values written in ascending order. Let

$$\hat{\omega}_\beta = W_{(f)},$$

be the  $f^{\text{th}}$  largest of the  $W_i$  values, let  $M$  be the usual sample median based on  $Y_1, \dots, Y_n$  and let

$$U_i = \frac{Y_i - M}{\hat{\omega}_\beta}$$

where  $a_i = 1$  if  $|U_i| < 1$ ; otherwise  $a_i = 0$ . The estimated percentage bend midvariance is

$$\hat{\zeta}^2 = \frac{n\hat{\omega}_\beta^2 \sum \Psi^2(U_i)}{(\sum a_i)^2}, \quad (2.3)$$

where  $\Psi(x) = \max[-1, \min(1, x)]$ .

Henceforth, it is assumed that  $\tau^2$  is the percentage bend midvariance and that  $\eta^2$  is estimated with

$$\hat{\eta}^2 = \frac{\hat{\tau}^2(\tilde{Y})}{\hat{\tau}^2(Y)}. \quad (2.4)$$

There remains the problem of choosing a method for computing  $\tilde{Y}$ . First consider the situation where

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (2.5)$$

where  $X$  and  $\varepsilon$  are independent random variables. If  $b_0$  and  $b_1$  are estimates of  $\beta_0$  and  $\beta_1$ , respectively,  $\tilde{Y}_i = b_0 + b_1 X + \varepsilon$  ( $i = 1, \dots, n$ ) can be used to compute  $\hat{\tau}^2(\tilde{Y})$ , which in turn can be used to compute  $\hat{\eta}^2$ . Wilcox (in press b) considered several robust regression estimates of  $\beta_0$  and  $\beta_1$  with the goal of estimating  $\eta^2$  with  $\hat{\eta}^2$ . Based on mean squared error and bias, it was found that the Theil-Sen estimate of  $\beta_0$  and  $\beta_1$  performs relatively well, thus it is used here.

Consider all pairs of points for which the two predictor values differ. The Theil-Sen estimator computes the slope for all such pairs of points and the estimate of  $\beta_1$ , for example  $\beta_{1ts}$ , is taken to be the median of all these slopes. The intercept is taken to be  $b_{0ts} = M_y - b_{1ts} M_x$ , where  $M_y$  is the usual median of  $Y_1, \dots, Y_n$ . The breakdown point of this estimator is approximately 0.29, where roughly, the breakdown point of an estimator is the proportion of points that must be altered to make it arbitrarily large or small. Moreover, the Theil-Sen estimator has excellent efficiency compared to many other robust estimators that have been proposed.

Next consider the more general case

$$Y = m(X) + \varepsilon \quad (2.6)$$

where  $m(X)$  is some unknown function of  $X$  and  $\varepsilon$  is some random variable that is independent of  $X$ . Wilcox (in press b) considered various nonparametric regression estimators with the goal of estimating  $\eta^2$ . In terms of mean squared error and bias, a so-called running interval smoother (Wilcox, 2005), as well as a method based on a cubic B-spline

(Hastie & Tibshirani, 1990) were found to be relatively unsatisfactory. Bootstrap bagging combined with these estimators was considered, but was found to perform poorly. No method dominated, but a rough guideline is that, when a linear model holds, the Theil-Sen estimator is a good choice, otherwise use Cleveland's (1979) LOESS. A nonparametric estimator derived by Wood (2004) was found to perform relatively well when a linear model holds, but the Theil-Sen estimator seems preferable. Finally, when there is curvature LOESS was generally more satisfactory.

To briefly outline Cleveland's method, consider the random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ . For any  $x$ , let

$$\delta_i = |X_i - x|.$$

Put the  $\delta_i$  values in ascending order and retain the  $\kappa n$  pairs of points that have the smallest  $\delta_i$  values, where  $\kappa$  is some number between 0 and 1 and is called the span. Let

$$Q_i = \frac{|x - X_i|}{\delta_m},$$

where  $\delta_m$  is the maximum of the retained  $\delta_i$  values. If  $0 \leq Q_i < 1$ , set

$$w_i = (1 - Q_i^3)^3,$$

otherwise  $w_i = 0$ . Let  $m(X)$  be the estimate of  $Y$  given  $X = x$  and use weighted least squares to estimate  $m(X)$  using  $w_i$  as weights. Both R and S-PLUS provide access to a function, called `lowess`, that performs a robust version proposed by Cleveland, and the R version was used in the simulations reported here using the default value  $\kappa = .75$ . Cleveland's robust method in effect gives little or no weight to extreme  $Y$  values. (An outline of these additional computations also can be found in Härdle, 1990, p. 192.)



## COMPARING THE STRENGTH OF ASSOCIATION OF TWO PREDICTORS

### Methodology

Testing (1.1) when (2.5) Is True

Consider the case where (2.5) is true and  $\eta^2$  is estimated via the Theil-Sen estimator. The initial strategy considered for testing (1.1) was a basic percentile bootstrap method (Efron & Tibshirani, 1993). Let  $(X_{i1}, X_{i2}, Y_i)$ ,  $i = 1, \dots, n$ , be a random sample. A bootstrap sample is obtained by resampling with replacement  $n$  vectors from this random sample yielding, for example,  $(X_{i1}^*, X_{i2}^*, Y_i^*)$ . Let  $\hat{\eta}_j^2$  be the estimate of  $\eta_j^2$  based on this bootstrap sample. Repeat this process  $B$  times yielding  $\hat{\eta}_{jb}^2$ ,  $b = 1, \dots, B$ . Let

$$P = \frac{1}{B} \sum I_b$$

where the indicator function  $I_b = 1$  if  $\hat{\eta}_{1b}^2 > \hat{\eta}_{2b}^2$ , otherwise  $I_b = 0$ . A (generalized) p-value is (Liu & Sing, 1997) is then:

$$p = 2 \min(P, 1 - P).$$

Let  $\rho_{12}$  be Pearson's correlation between  $X_1$  and  $X_2$ . Initial simulations revealed that when testing at the  $\alpha = .05$  level, the basic percentile bootstrap method described performs reasonably well when  $\rho_{12} = 0$ . However, as  $\rho_{12}$  increases, the estimate of the actual Type I error probability decreased. For  $\rho_{12} = 0.7$  the estimates were less than 0.01.

The first attempt at reducing this problem was to use a bootstrap estimate of the squared standard error of  $\eta_1^2 - \eta_2^2$ , say,  $\hat{Y}^2$ , and then use the test statistic  $(\eta_1^2 - \eta_2^2) / \hat{Y}$ . However, in simulations with  $B = 100$ , Type I errors were found to be relatively sensitive to the distributions generating the data; increasing  $B$  to 400 reduced this problem somewhat but did not eliminate it, thus this approach was abandoned.

Consider the case  $\rho_{12} = 0$  and let  $D = \hat{\eta}_1^2 - \hat{\eta}_2^2$ . It was found that Type I error control is improved if, rather than a single

bootstrap sample, two independent bootstrap samples are used. That is, take a bootstrap sample from  $(X_{i1}, Y_i)$  and compute a bootstrap estimate of  $\eta_1^2$ , for example,  $\hat{\eta}_1^2$ , take a new, independent bootstrap sample from  $(X_{i2}, Y_i)$  yielding  $\hat{\eta}_2^2$  and let  $D = \hat{\eta}_1^2 - \hat{\eta}_2^2$ . Repeating this process  $B$  times yields  $D_1, \dots, D_B$ , which can be used to estimate  $P = P(D < 0)$  in the manner already described. This in turn yields the generalized p value. Once again control over the probability of a Type I error was found to be unsatisfactory. However, it was found that control over the Type I error probability was improved if, instead of estimating  $P = P(D < 0)$  with the bootstrap samples in the usual way, a kernel density estimate is used; this strategy was based on results from Racine and MacKinnon (2007).

Generally, kernel density estimates of the distribution of  $D$  take the form

$$\hat{f}(d) = \frac{1}{nh} \sum K\left(\frac{d - D_i}{\lambda}\right),$$

where  $K$  is some probability density function and  $\lambda$  is a constant to be determined called the span or smoothing parameter. Given  $h$  and a choice for  $K$ , which is assumed to be defined over some known interval  $(\ell, u)$ , an estimate of  $P(D < 0)$  is

$$P(D < 0) = \frac{1}{nh} \sum_{i=1}^n \int_{\ell}^0 K\left(\frac{t - D_i}{\lambda}\right) dt.$$

The focus here is on the Epanechnikov kernel where, for  $|t| < \sqrt{5}$ ,

$$K(t) = \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5} t^2\right),$$

otherwise  $K(t) = 0$ .

Following Silverman (1986, pp. 47-48), the span is taken to be

$$\lambda = 1.06 \frac{A}{n^{1/5}},$$

where

$$A = \min(s, IQR / 1.34)$$

and  $s$  is the standard deviation, and IQR is the interquartile range.

From Silverman (1986), one possible way of improving on the basic kernel density estimator, is to use an adaptive method. Let  $\tilde{f}(X_i)$  be an initial estimate of  $f(X_i)$ . Here,  $\tilde{f}(X_i)$  is based on the so-called expected frequency curve (Wilcox, 2005, pp. 48-49). Let

$$\log(g) = \frac{1}{n} \sum \log(f(X_i))$$

and

$$\omega_i = \tilde{f}(X_i / g)^{-a}$$

where  $a$  is a sensitivity parameter satisfying  $0 \leq a \leq 1$ . Based on comments by Silverman (1986), if  $\alpha = 0.5$  is used, then the adaptive kernel estimate of the probability density function  $f$  is taken to be

$$\hat{f}(t) = K\{\lambda^{-1}\omega_i^{-1}(t - X_i)\}.$$

Henceforth, it is assumed that the adaptive method described is used to estimate  $P(D < 0)$  based on  $D_1^*, \dots, D_B^*$ , and the corresponding p-value is denoted by  $p$ .

There remains the problem of dealing with the general case  $\rho_{12} \neq 0$ . If it is assumed that there is normality and  $\rho_{12}$  is known, then simulations can be used to determine  $p_{adj}$  so that for some choice for  $\alpha$ ,  $P(p \leq p_{adj}) = \alpha$ . In particular, imagine that simulations with  $N$  replications are performed resulting in the p-values,  $p_1, \dots, p_N$ . Arranging these  $N$  values in ascending order yielding  $p_{(1)} \leq \dots \leq p_{(N)}$  and letting  $C = \alpha N$  round to the nearest integer results in the adjusted p-value  $p_{adj} = p_{(C)}$ .

A simple approach when dealing with  $\rho_{12}$  unknown is to replace  $\rho_{12}$  with  $r_{12}$  in such a simulation. Execution time was found to be reasonably low, but to reduce it further, the following approach was considered when  $\alpha = .05$ . The value of  $p_{adj}$  was determined with  $n = 20$  for  $\rho_{12} = 0, 0.2, 0.5$  and  $0.8$ . When  $\rho_{12}$  is known, it was found that  $p_{adj}$  is given approximately by  $0.352|p_{adj}| + 0.049$ . But when  $\alpha = .05$  the actual level can exceed  $0.075$  due to situations where  $|r_{12}|$  exceeds  $|\rho_{12}|$  resulting in over adjusting the critical p-value. In this situation, the additional concern is that  $r_{12}$  is not robust, and there is the issue of how to adjust the critical p-value when  $n > 20$ .

To deal with the lack of robustness associated with Pearson's correlation,  $r_{12}$  was replaced by Kendall's tau, resulting in  $r_{k12}$ . The population analog of  $r_{k12}$  is denoted by  $\rho_{k12}$ . Next, a 0.95 confidence interval for  $\rho_{k12}$  was computed using a basic percentile bootstrap method (Wilcox, 2005, p. 403), which has low execution time, even when the sample size is large. If this interval contains zero, let  $\tilde{p} = 0.05$ , Otherwise, let  $\tilde{p} = .352 |r_{k12}| + 0.049$ . Rejecting (1) when the p-value is less than or equal to  $\tilde{p}$  will be called method BTS.

This approximation depends on the sample size,  $n$ , but a convenient feature is that it was found to change slowly as  $n$  gets large. In particular, it continues to perform well when  $n = 100$ . For  $n = 200$  this is no longer the case, but with  $n \geq 100$  the adjustment makes little difference. So the suggestion is to use method BTS when  $n \leq 100$ , otherwise reject if the p-value is less than or equal to  $\alpha$ .

#### Testing (1.1) when (2.6) Is True

Consider now the more general case where the regression line is given by (2.6). Method BTS can be extended in an obvious way. In particular, again the strategy is to use independent bootstrap samples to estimate  $\eta_1^2$  and  $\eta_2^2$  and the adaptive kernel density

## COMPARING THE STRENGTH OF ASSOCIATION OF TWO PREDICTORS

estimation method for computing a p-value is used. However, now the actual level of the test is more sensitive to  $\rho_{12}$  and for the case  $\alpha = .05$ , a modification of  $\tilde{p}$  is required. As was the case when (2.5) is assumed, simulations indicate that if (1.1) is rejected when the p-value is less than or equal to  $\alpha$ , the actual level will be less than or equal to  $\alpha$ ; avoiding actual Type I error probabilities substantially less than the nominal level is more difficult in this case. Based on preliminary simulations, under normality, when testing at the  $\alpha = .05$  level, the following approach performed best among the methods considered. Let

$$\tilde{p} = .25 | r_{k12} | + .05 + (100 - n) / 10000,$$

$\hat{p} = \max(.05, \tilde{p})$ , and reject (1.1) if  $p \leq \hat{p}$ . For  $n > 200$ ,  $\hat{p}$  is taken to be 0.05 and this will be called method SM. Note that in contrast to method BTS, a confidence interval for  $\rho_{k12}$  is not used.

### Results

Simulations were used as a partial check on the actual level of methods SM and BTS when testing at the 0.05 level. Values for  $X_1$  and  $X_2$  were generated from a bivariate distribution for which the marginal distributions belong to the family of g-and-h distributions, which contains the standard normal as a special case. The R function `rmul` was used, in conjunction with the function `ghdist`, which are part of the library of R functions described in Wilcox (2005).

The R function `rmul` generates data from an m-variate distribution having a population correlation matrix  $\mathbf{R}$  by first forming the Cholesky decomposition  $\mathbf{U}\mathbf{U}' = \mathbf{R}$ , where  $\mathbf{U}$  is the matrix of factor loadings of the principal components of the square-root method of factoring a correlation matrix, and  $\mathbf{U}'$  is the transpose of  $\mathbf{U}$ . Next, an  $n \times m$  matrix of data,  $\mathbf{X}$ , for which the marginal distributions are independent, is generated, then  $\mathbf{X}\mathbf{U}$  produces an  $n \times m$  matrix of data that has population correlation matrix  $\mathbf{R}$ .

To elaborate, let  $Z$  be a standard normal distribution. For  $g > 0$ , let

$$X = \frac{\exp(gZ) - 1}{h} \exp(hZ^2 / 2),$$

and for  $g = 0$ , let

$$X = \exp(hZ^2 / 2),$$

in which case  $\mathbf{X}$  has a g-and-h distribution where g and h are parameters that determine the first four moments. When  $g = h = 0$ ,  $\mathbf{X}$  has a standard normal distribution. With  $g = 0$  this distribution is symmetric and it becomes increasingly skewed as g gets large. As h gets large, the g-and-h distribution becomes more heavy-tailed. Table 1 shows the skewness ( $\kappa_1$ ) and kurtosis ( $\kappa_2$ ) for each distribution considered in the simulations used herein. They correspond to a standard normal ( $g = h = 0$ ), a symmetric heavy-tailed distribution ( $h = 0.2, g = 0.0$ ), an asymmetric distribution with relatively light tails ( $g = 0.2, h = 0$ ) and an asymmetric distribution with relatively heavy tails ( $g = h = 0.2$ ).

Table 1: Some Properties of the g-and-h Distribution

g	h	$\kappa_1$	$\kappa_2$
0.0	0.0	0.00	3.0
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

Simulations were run with  $\rho_{12} = 0$  and 0.7, where  $\varepsilon$  has the same distribution as  $X_1$  and  $X_2$ . Additional simulations were run where  $X_1$  is normal but  $X_2$  has one of the non-normal g-and-h distributions previously described.

Table 2 shows the estimated probability of a Type I error based on 1,000 replications when using method BTS,  $n = 50$  and  $Y = X_1 + X_2 + \varepsilon$ . The columns headed by  $X_1 \sim X_2$  indicate that  $X_1$  and  $X_2$  have

RAND R. WILCOX

identical distributions, while  $X_1 \sim N(0,1)$  means that  $X_1$  has a standard normal distribution and  $X_2$  has the g-and-h distribution indicated. Table 3 shows the results when using method SM when  $Y = X_1 + X_2 + \varepsilon$  and  $Y = X_1^2 + X_2^2 + \varepsilon$  with  $n = 60$ . As is evident, method BTS performs reasonably well in terms of avoiding a Type I error well above the nominal level, at least for the situations considered. A deficiency of the method is that the estimates drop below 0.025 in some situations. Method SM also performs reasonably well, but the actual level drops well below the nominal level in some situations.

Table 2: Estimated Type I Error Rates, n=50, Method BTS,  $\alpha=0.05$

			$X_1 \sim X_2$	$X_1 \sim N(0,1)$
g	h	$\rho_{12}$	$\hat{\alpha}$	$\hat{\alpha}$
0.0	0.0	0.0	0.039	-
-	-	0.7	0.048	-
0.0	0.2	0.0	0.026	0.054
-	-	0.7	0.023	0.059
0.2	0.0	0.0	0.041	0.027
-	-	0.7	0.063	0.048
0.2	0.2	0.0	0.064	0.056
-	-	0.7	0.044	0.071

Table 3: Estimated Type I Error Rates, n=60,  $\alpha=0.05$ , Method SM

			$X_1 \sim X_2$	$X_1 \sim X_2$	$X_1 \sim N(0,1)$	$X_1 \sim N(0,1)$
g	h	$\rho_{12}$	$Y = X_1 + X_2 + \varepsilon$	$Y = X_1^2 + X_2^2 + \varepsilon$	$Y = X_1 + X_2 + \varepsilon$	$Y = X_1^2 + X_2^2 + \varepsilon$
0.0	0.0	0.0	0.036	0.026	-	-
		0.5	0.020	0.034	-	-
		0.7	0.008	0.014	-	-
0.0	0.2	0.0	0.036	0.012	0.048	0.049
		0.5	0.014	0.020	0.022	0.054
		0.7	0.008	0.012	0.014	0.021
0.2	0.0	0.0	0.032	0.024	0.036	0.022
		0.5	0.014	0.024	0.023	0.036
		0.7	0.014	0.016	0.008	0.022
0.2	0.2	0.0	0.026	0.014	0.040	0.024
		0.5	0.004	0.016	0.026	0.042
		0.7	0.008	0.008	0.010	0.023

## COMPARING THE STRENGTH OF ASSOCIATION OF TWO PREDICTORS

### Power

There is the issue of how much power is sacrificed if method SM is used rather than BTS when the regression line is straight. Table 4 shows the probability of rejecting when  $Y = X_1 + \varepsilon$ . As is evident, both methods have fairly high power for this special case and BTS can offer a substantial gain in power when the regression line is straight.

Table 4: Power Comparisons,  
n=60,  $\alpha=0.05$ ,  $Y = X_1 + \varepsilon$

g	h	$\rho_{12}$	BTS	SM
0.0	0.0	0.0	0.960	0.833
-	-	0.5	0.861	0.659
0.0	0.2	0.0	0.930	0.668
-	-	0.5	0.777	0.467
0.2	0.0	0.0	0.968	0.807
-	-	0.5	0.836	0.710
0.2	0.2	0.0	0.942	0.672
-	-	0.5	0.772	0.460

### An Illustration

In an unpublished study by Doi, a general goal was to identify good predictors of reading ability in children. Two of the predictors were a measure of letter naming speed and the speed at which lowercase letters could be identified. The outcome of interest was a measure of reading comprehension. A scatterplot of the data and the LOESS estimate of the regression strongly suggests that there is curvature, and a test of the hypothesis that the regression line is straight (using the method in Wilcox, 2005, section 11.5.1) is rejected at the 0.05 level; thus method SM is used and it rejects at the 0.05 level. The estimated explanatory power for the plot in the left panel is 0.444, and in the right panel it is 0.171. These results suggest that naming speed has a stronger association with comprehension.

If the apparent curvature is ignored, BTS also rejects at the 0.05 level, but now the estimated explanatory is 0.351 for the left panel and 0.142 for the right. That is, the estimated

difference in explanatory power is substantially smaller compared to using a smoother. If instead Pearson correlations are compared using the method in Zou (2007), the 0.95 confidence interval for the difference is  $(-0.490, 0.024)$ . Therefore, fail to reject at the 0.05 level.

### Conclusion

In summary, numerous methods for comparing two predictors were considered based on a robust measure of the strength of the association. Two methods were found that perform reasonably well in simulations, one of which is based on a smoother and so provides a flexible approach to curvature. All indications are that Type I errors that exceed the nominal level are avoided using a basic percentile bootstrap method; however, there is a practical problem that the actual level can drop well below the nominal level, particularly when the sample size is small. Adjustments were suggested that substantially reduce this problem among the situations considered. The adjustment used by method BTS performed reasonably well in simulations, but when using method SM, situations occurred where the actual level drops well below the nominal level even with  $n = 60$ . In principle, if there are  $p$  predictors and the goal is to compare subsets of  $k$  predictors, a strategy similar to those used here could be used, but it remains to be determined whether reasonable control over the probability of a Type I error can be achieved.

Regarding the use of a bootstrap method, Hall and Wilson (1991) argue in favor of using a pivotal test statistic, which is not done here. When working with means, more recent results, summarized in Wilcox (2005), also support the conclusion that a pivotal test statistic be used. When working with robust estimators, however, there are general situations where a percentile bootstrap method has a substantial advantage. In addition, when using a percentile bootstrap method, there is no need to approximate the null distribution of some test statistic (Liu & Singh, 1997). Roughly, the percentile bootstrap method is based on determining how deeply the null value is nested within the sampling distribution of some estimator. Finally, R functions for applying the methods considered are available from the

author; download the file Rallfun-v17 from www-rcf.usc.edu~rwilcox. The function sm2strv7 performs method SM, and the function ts2str performs methods BTS.

References

- Bickel, P. J., & Lehmann, E. L. (1976). Descriptive statistics for nonparametric models III dispersion. *Annals of Statistics*, 4, 1139-1158.
- Doksum, K. A., & Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Annals of Statistics*, 23, 1443-1473.
- Dunn, O. J., & Clark, V. A. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association*, 66, 904-908.
- Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory and applications*. New York: Springer-Verlag.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: Marcel Dekker.
- Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. Boca Raton, FL: CRC Press.
- Fox, J. (2001). *Multiple and generalized nonparametric regression*. Thousand Oaks, CA: Sage.
- Godfrey, L. G. (2006). Tests for regression models with heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, 50, 2715-2733.
- Green, P. J., & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: A roughness penalty approach*. Boca Raton, FL: CRC Press.
- Györfi, L., Kohler, M., Krzyzk, A., & Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer-Verlag.
- Hall, P., & Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47, 757-762.
- Härdle, W. (1990). *Applied nonparametric regression*. Econometric Society Monographs, No. 19, Cambridge, UK: Cambridge University Press.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall.
- Hendrickson, G. F., Stanley, J. C., & Hills, J. R. (1970). Olkin's new formula for significance levels of  $r_{13}$  vs.  $r_{23}$  compared with Hotelling's method. *American Educational Research Journal*, 7, 189-195.
- Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology*, 130, 149-168.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In *Exploring data tables, trends, and shapes*, D. Hoaglin, F. Mosteller, & J. Tukey (Eds.), 461-515. New York: Wiley.
- Hotelling, H. (1940). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Statistics*, 11, 271-283.
- Huber, P. J. (1981). *Robust statistics*. New York: Wiley.
- Lafferty, J., & Wasserman, L. (2008). RODEO: Sparse, greedy nonparametric regression. *Annals of Statistics*, 3, 28-63.
- Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association*, 80, 736-741.
- Liu, R. G., & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, 92, 266-277.
- Maronna, R. A., & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44, 307-317.
- Meng, X.-L., Rosenthal, R. & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.
- Olkin, I. (1967). Correlations revisited. In *Improving Experimental Design and Statistical Analysis*, J. C. Stanley (Ed.), 102-128. Chicago: Rand McNally

## COMPARING THE STRENGTH OF ASSOCIATION OF TWO PREDICTORS

Racine, J. S., & MacKinnon, J. G. (2007). Inference via kernel smoothing of bootstrap p values. *Computational Statistics & Data Analysis*, 51, 5940-5957.

Randal, J. A. (2008). A reinvestigation of robust scale estimation in finite samples. *Computational Statistics & Data Analysis*, 52, 5014-5021.

Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimensions. *Annals of Statistics*, 24, 1327-1345.

Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63, 1379-1389.

Shoemaker, L. H., & Hettmansperger, T. P. (1982) Robust estimates and tests for the one- and two-sample scale models. *Biometrika*, 69, 47-54.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245-251.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae*, 12, 85-91.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. New York: Academic Press.

Wilcox, R. R. (2009). Comparing Pearson correlations: Dealing with heteroscedasticity and nonnormality. *Communications in Statistics-Simulations and Computations*, 38, 2220-2234.

Wilcox, R. R. (in press). Measuring and detection associations: Methods based on robust regression estimators or smoothers that allow curvature. *British Journal Mathematical and Statistical Psychology*.

Wilcox, R. R., & Tian, T. (2008). Comparing dependent correlations. *Journal of General Psychology*, 135, 105-112.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B*, 21, 396-399.

Yohai, V. J., & Zamar, R. H. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 83, 406-414.

Yu, M. C., & Dunn, O. J. (1982). Robust tests for the equality of two correlation coefficients: A Monte Carlo study. *Educational and Psychological Measurement*, 42, 987-1004.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, 12, 399-413.

*Regular Articles*  
**Number of Replications Required in Monte Carlo Simulation Studies:  
A Synthesis of Four Studies**

Daniel J. Mundfrom  
New Mexico State University,  
Las Cruces, NM USA

Jay Schaffer  
University of Northern Colorado,  
Greeley, CO USA

Myoung-Jin Kim  
Mennonite College of Nursing  
at Illinois State University,  
Normal, IL USA

Dale Shaw  
University of Northern Colorado,  
Greeley, CO USA

Ampai Thongteeraparp  
Kasetsart University,  
Bangkean, Bangkok, Thailand

Chana Preecha  
Thepsatri Rajabhat University,  
Muang, Lopburi, Thailand

Pornsini Supawan  
Rajabhat Rajanagarindra University,  
Chacoengsao, Chaiyaphum, Thailand

---

Monte Carlo simulations are used extensively to study the performance of statistical tests and control charts. Researchers have used various numbers of replications, but rarely provide justification for their choice. Currently, no empirically-based recommendations regarding the required number of replications exist. Twenty-two studies were re-analyzed to determine empirically-based recommendations.

Key words: Simulation studies, number of replications, ANOVA, multiple comparisons, regression, control charts.

---

### Introduction

Monte Carlo simulation has become an

---

Daniel J. Mundfrom is an Associate Professor of Applied Statistics, Department of Economics and Professor Emeritus of Applied Statistics and Research Methods at the University of Northern Colorado. E-mail: daniel.mundfrom@eku.edu. Jay Schaffer is an Associate Professor of Applied Statistics and Research Methods. E-mail: jay.schaffer@unco.edu. Myoung-Jin Kim is an Assistant Professor. E-mail: mkim2@ilstu.edu. Dale Shaw is Professor Emeritus of Applied Statistics and Research Methods. E-mail: dale.shaw@unco.edu. Ampai Thongteeraparp is an Assistant Professor of Statistics in the Department of Statistics, Faculty of Science. E-mail: fsciamu@ku.ac.th. Chana Preecha is an Assistant Professor of Statistics on the Faculty of Science and Technology. E-mail: chanaua@yahoo.com. Pornsini Supawan is an Assistant Professor of Statistics on the Faculty of Science. E-mail: pornsin.s@gmail.com.

important and popular research tool used by quantitative researchers in a variety of disciplines (Fan, Felsővályi, Sivo & Keenan, 2002). The Monte Carlo method provides approximate solutions to a variety of mathematical problems by performing statistical sampling experiments via computer. Monte Carlo simulation offers researchers an alternative to the theoretical approach; this is important because many situations exist in which implementing a theoretical approach is difficult – and finding an exact solution is even more difficult. In addition, computing power has become increasingly less expensive and computers are more widely available than ever before.

An important question to address when conducting a Monte Carlo simulation study is how many replications are needed to obtain accurate results. With advanced computers, researchers are able to run in excess of 10,000 replications in their studies (see, for example, Kaplan, 1983; Klockars & Hancock, 1992; Gamage & Weerahandi, 1998; Alyounes, 1999).



## NUMBER OF REPLICATIONS REQUIRED IN MONTE CARLO SIMULATION STUDIES

According to Brooks (2002), simulations may produce inaccurate estimates if an insufficient number of replications are used. Hutchinson and Bandalos (1997) also criticized:

With too few replications, idiosyncratic results based on a particular sample are more likely to arise. Unfortunately for simulation researchers there are no definitive guidelines for selecting the appropriate number of replications. The specific number will depend on the type of phenomenon being studied, the extent to which the steps of the simulation can be automated, as well as available computer resources. (p. 238)

The choice of the number of replications used in simulation studies appear to be made solely by the judgment of the researchers; this is surmised due to the many simulation studies that have been conducted without any justification provided for the number of replications used (see, for example, Fellner, 1990; Neubauer, 1997; Khoo & Quah, 2002; Khoo & Quah, 2003; Khoo, 2003; Khoo, 2004). Currently, however, no empirically-based recommendations for general guidelines regarding the required number of replications a researcher should use in order to achieve accurate results exist. The obtained results from a Monte Carlo study might be invalid if too few replications were used, whereas time and resources may have been wasted if more replications were used than were necessary. In addition, with the same amount of time and resources but fewer replications, more conditions could be investigated.

The purpose of this synthesis was to: (1) provide information regarding the minimum number of replications required to reproduce a reported statistic, within a specified degree of accuracy, in 22 published Monte Carlo studies from a variety of areas, and (2) provide general recommendations regarding the minimum number of replications needed for future simulation studies.

### Methodology

An extensive review of the literature was conducted in various fields of study, identifying

research that used Monte Carlo simulations to estimate characteristics of interest (e.g., Type I error rates, power and average run length). Through four dissertations, 22 studies were selected such that each provided sufficient information regarding methodology to replicate.

Each study was re-analyzed using the same number of replications as in the original study to produce results that were considered the standard to be met by the re-analyses using a different number of replications. Using a decreasing (or increasing) number of replications, the simulations were repeated until the minimum number of replications was found that produced stable results.

For example, if the original study used 10,000 replications, the process started with 10,000 replications to reproduce the original results and identify the standard to be met, and then the study was re-done with the number of replications cut in half to 5,000. If the results were reproduced, the replications were cut to 2,500; conversely, if the results were not reproduced the replications were increased to 7,500. This iterative process, either reducing the number of replications by cutting in half the number of replications used in the previous step, or increasing the number of replications used by splitting the difference between the last two numbers of replications used (e.g., 5,000 and 10,000), continued until stable results were obtained. After the simulations were completed, recommendations were put forth for the minimum number of replications necessary to estimate a particular parameter within a defined degree of accuracy.

In order to define a specified degree of accuracy, an error band was created by adding/subtracting some percentage to/from each statistic of interest. Bradley (1978) presented two intervals to examine the robustness of hypothesis testing by examining Type I error rate,  $\alpha$ . These two intervals were described as a fairly stringent error band,  $\alpha \pm 0.1 \alpha$ , and a fairly liberal error band,  $\alpha \pm 0.4 \alpha$ . If  $\alpha = 0.05$ , these error bands become  $\pm 0.005$  and  $\pm 0.02$  respectively. Bradley's criteria were used in these dissertations.

Dissertation I: ANOVA Simulation Studies (Preecha, 2004)

This study replicated 5 simulation studies related to ANOVA. The studies included:

(1) Brown and Forsythe (1974) examined the small sample behavior of various statistics testing the equality of several means. They used 10,000 replications and examined both Type I error rate and power. No justification was provided for the number of replications used or for how the accuracy of results was determined. The four statistics compared were the:

- (a) ANOVA F-statistic;
- (b) Modified F-statistic;
- (c) Welch and James statistic (Welch, 1947); and the
- (d) Welch and James statistic (Welch, 1951).

(2) Alyounes (1999) compared the Type I error rate and power for the Kruskal-Wallis test and the Welch test to the F-test, followed by four post hoc procedures. They used 21,000 replications but provided no justification for that number. Bradley's stringent criterion and Robey and Barcikowski's intermediate criterion were used to examine the robustness of the tests compared. The parametric and nonparametric omnibus tests and the post hoc comparisons used were the:

- (a) ANOVA F-test;
- (b) Welch test;
- (c) Kruskal Wallis test;
- (d) Tukey-Kramer test;
- (e) Games-Howell test;
- (f) Joint ranking (Improved Dunn) test; and
- (g) Separate ranking test.

(3) Gamage and Weerahandi (1998) examined the size performance of four tests in a one-way ANOVA. They compared the Type I error rate and power of the Generalized F-test to the classical F-test, the F-test using weighted least squares to adjust for heteroscedasticity, the Brown-Forsythe test, and the Welch test using 20,000 replications.

No justification was provided for the number of replications used or for how the accuracy of results was determined. The statistics compared were the:

- (a) Generalized F-test;
- (b) ANOVA F-test;
- (c) F-test using weighted-least squares;
- (d) Brown-Forsythe test; and the
- (e) Welch test.

(4) Kim (1997) examined three robust tests for ANOVA using weighted likelihood estimation, comparing Type I error rate and power with 5,000 replications. No justification was provided for the number of replications used or for how the accuracy of results was determined. The statistics compared were the:

- (a) Basu-Sarkar-Basu test;
- (b) Modified Welch Test with weighted likelihood estimators; and the
- (c) Modified Brown-Forsythe test using weighted likelihood estimators.

(5) Kaplan (1984) examined the comparative effects of violations of homogeneity of variance on two tests when the underlying populations were normal, but sample sizes were unequal. She compared Type I error rate and power using 20,000 replications. She provided no justification for the number of replications used, but used the estimated standard error when examining a single proportion and the estimated standard error of the difference between two proportions when comparing two independent proportions. The tests compared were the:

- (a)  $\chi^2$ -approximation of the Kruskal-Wallis statistic; and the
- (b) Incomplete Beta approximation of the Kruskal-Wallis statistic.

Dissertation I: Results and Discussion

Each of the five studies investigated Type I error rates and power. Using  $\pm 0.005$  for Type I error (Bradley's fairly stringent criterion) and  $\pm 0.02$  for power (Bradley's fairly liberal criterion), the minimum number of replications

## NUMBER OF REPLICATIONS REQUIRED IN MONTE CARLO SIMULATION STUDIES

were found that produced stable results. Table 1 displays the number of replications used in the original study along with the recommended minimum number of replications needed to produce similar results. In each situation, it appears that fewer replications could have been used to predict power and in all but one situation, fewer replications could have been used to estimate Type I error. In that one situation a larger number of replications was required to get a stable estimate of the Type I error rate.

Table 1: Number of Replications Used By the Original Study Along With the Recommended Minimum Number of Replications Required To Produce Stable Results

Study	Original	Replications Recommended	
	Replications Used	Type I Error	Power
1	10,000	5,000 – 10,000	5,000
2	21,000	10,500	5,250
3	20,000	7,500	5,000
4	5,000	7,500	2,500
5	20,000	5,000	5,000

### Dissertation II: Multiple Comparison Simulation Studies (Ussawarujikulchai, 2004)

The second dissertation replicated 5 simulation studies related to multiple comparison tests after a significant ANOVA was found. The studies included:

- (1) Seaman, Levin and Serlin (1991) examined the Type I error rate of several multiple comparison procedures using 5,000 replications to compare 5 treatment groups with sample sizes of  $n = 10$ . Three groups had means set equal to 0 and the other groups had means set to 0.8560. The procedures compared were:
  - (a) Standard Bonferroni;
  - (b) Tukey test;
  - (c) Holm test;
  - (d) Fisher LSD test;
  - (e) Hayter-Fisher Modified LSD test;
  - (f) REGWQ test;
  - (g) Newman-Kuels test;
  - (h) Duncan test;
  - (i) Shaffer test;
  - (j) Protected Shaffer test; and
  - (k) Ramsey's Model-Testing approach.
  
- (2) Klockars and Hancock (1992) examined the power of five multiple comparison procedures against the standard Bonferroni procedure when applied to complete sets of orthogonal contrasts. They used 20,000 replications with both  $k = 4$  and  $k = 5$  treatment groups partitioned into  $k-1$  orthogonal contrasts. The procedures they compared were:
  - (a) Holm test;
  - (b) Hochberg test;
  - (c) Hommel test;
  - (d) Protected Shaffer test;
  - (e) Modified Stageswise Protected test; and
  - (f) Standard Bonferroni procedure.
  
- (3) Hsiung and Olejnik (1994) examined the Type I error rate of several multiple comparison procedures for all pairwise contrasts when population variances differed in both balanced and unbalanced one-factor designs. They used 10,000 replications for each of  $k = 4$  and  $k = 6$  treatment groups. The multiple comparison procedures they compared were:
  - (a) Games-Howell test;
  - (b) Dunnett T3 test;
  - (c) Dunnett C test;
  - (d) Holland-Copenhaver test;
  - (e) Shaffer test; and
  - (f) Protected Shaffer test.
  
- (4) Morikawa, Terao and Iwasaki (1996) examined the Type I error rate and power of several multiple comparison procedures for pairwise comparisons. They used 1,000 replications with each of  $k = 3$  and  $k = 4$

treatment groups and sample sizes of 10, 20 and 50 to examine both any-pair power and all-pairs power. The procedures they compared were:

- (a) Tukey test;
  - (b) Standard Bonferroni test;
  - (c) Holm test;
  - (d) Shaffer test;
  - (e) Hommel test;
  - (f) Hochberg test; and the
  - (g) Rom test.
- (5) Ramsey (2002) examined the power of five pairwise multiple comparison procedures using 10,000 replications with 4 treatment groups and a sample size of 16. Both any-pair power and all-pairs power were examined for three different mean configurations-maximum range, equally spaced, and minimum range. The procedures compared were:

- (a) Tukey test;
- (b) Hayter-Fisher Modified LSD test;
- (c) Shaffer-Welsch test;
- (d) Shaffer test; and the
- (e) Holland-Copenhaver test.

Dissertation II: Results and Discussion

Each of these five studies investigated either Type I error rate, power, or both. Using  $\pm 0.005$  for Type I error (Bradley’s fairly stringent criterion) and  $\pm 0.02$  for power (Bradley’s fairly liberal criterion), the minimum number of replications were found that produced stable results. Table 2 displays the number of replications used by the original study along with the recommended minimum number of replications needed to produce stable results. It appears that fewer replications could have been used to predict power in studies 2 and 5, while too few replications were used in study 4. To predict Type I error, it appears that study 3 could have used fewer replications, whereas study 4 again could have used more replications.

Table 2: Number of Replications Used By the Original Study Along With the Recommended Minimum Number of Replications Required To Produce Stable Results

Study	Original	Replications Recommended	
	Replications Used	Type I Error	Power
1	5,000	5,000	---
2	20,000	---	3,750
3	10,000	5,000	---
4	1,000	8,000	4,000
5	10,000	---	3,750

Dissertation III: Regression Simulation Studies (Supawan, 2004)

The third dissertation replicated 6 simulation studies related to multiple linear regression. The studies included:

- (1) Griffiths and Surekha (1986) examined the Type I error rate and power of three tests for heteroscedasticity. They used 5,000 replications, but provided no justification for that choice. The tests they compared were:

- (a) Szroeter Test;
- (b) Breusch-Pagan Test; and
- (c) Goldfeld-Quandt Test.

- (2) Pfaffenberger and Dielman (1991) examined the Type I error rate and power of the Filliben test for normality of regression residuals using 6 different statistics. They used 5,000 replications, justifying this choice by their desire to control the maximum standard deviation of the rejection percentage to be  $< 1.0\%$ . The six statistics they examined were:

- (a) Means and the z-transformed residuals;
- (b) Medians and the z-transformed residuals;
- (c) Means and standardized residuals;
- (d) Medians and standardized residuals;

## NUMBER OF REPLICATIONS REQUIRED IN MONTE CARLO SIMULATION STUDIES

- (e) Means and studentized deleted residuals; and
  - (f) Medians and studentized deleted residuals.
- (3) Godfrey (1978) examined the power of the  $\chi^2(1)$  heteroscedasticity test for two multiplicative models, Uniform (1,31) and Lognormal (3, 1) using 1,000 replications, but providing no justification for this choice.
- (4) Flack and Chang (1987) examined the effects of sample size and the number of noise variables on the frequency of selecting noise variables by using  $R^2$  selection. They used 50 replications, justifying the choice by their belief that it was sufficient to give reliable results.
- (5) Hurvich and Tsai (1990) examined the effect of Akaike's Information Criterion (AIC) for model selection on the coverage rates of confidence regions of linear regression. They used 500 replications with no justification provided for their choice.
- (6) Olejnik, Mills and Keselman (2000) examined the accuracy of using stepwise regression compared with Wherry's  $R^2_{\text{adjusted}}$  and Mallow's  $C_p$  to select the model in all possible regressions by considering the effect of sample size, the number of noise variables and the correlation between authentic variables. They used 1,000 replications, but provided no justification for their choice.

### Dissertation III: Results and Discussion

Studies 1-3 investigated either Type I error rate, power, or both. Using  $\pm 0.005$  for Type I error (Bradley's fairly stringent criterion) and  $\pm 0.02$  for power (Bradley's fairly liberal criterion), the minimum number of replications were found that produced stable results. Table 3 displays the number of replications used by the original study along with the recommended minimum number of replications needed. In all but two situations, it appears that fewer replications could have been used to predict Type I error and power, with only Study #1

needing substantially more replications than were used to get a stable prediction for power.

Table 3: Number of Replications Used By the Original Study Along With the Recommended Minimum Number of Replications Required To Produce Stable Results

Study	Original	Replications Recommended	
	Replications Used	Type I Error	Power
1	5,000	4,600	7,000
2	5,000	4,200	1,300
3	1,000	---	1,250

Studies 4-6 investigated the proportion of variables selected to be included in the multiple linear regression model. Using  $\pm 0.005$  for the proportion of variables selected (Bradley's fairly stringent criterion), the minimum number of replications were found that produced stable results. Table 4 displays the number of replications used by the original study along with the recommended minimum number of replications needed. In each instance, it appears that more replications than were used in the original studies were required to obtain stable results.

Table 4: Number of Replications Used By the Original Study Along With the Recommended Minimum Number of Replications Required To Produce Stable Results

Study	Original	Replications Recommended
	Replications Used	Proportion of Variables Selected
4	50	1,900
5	500	2,000
6	1,000	1,900

Dissertation IV: Quality Control Simulation Studies (Kim, 2005)

The fourth dissertation replicated 6 simulation studies examining the average run length, ARL, of various statistical process control charts. The studies included:

- (1) Khoo (2004) examined the ARL property of the Shewhart chart using individual observations for 18 different shifts of size  $\delta$ . They used 10,000 replications with no justification provided.
- (2) Fellner (1990) examined the ARL property of the cumulative sum or CUSUM chart using individual observations for 6 different shifts of size  $\delta$ . A two-sided CUSUM control chart using decision values  $H = 2, 3, 4, 5, 6$  and reference value  $K = 0.5$  was studied. A total of 30 different scenarios were simulated using 10,000 replications with no justification provided.
- (3) Neubauer (1997) examined the ARL property of the exponentially weighted moving average (EWMA) chart using individual observations for 31 different shifts of size  $\delta$ . The EWMA control chart studied used a weighting constant  $\lambda = 0.2$  and width of the control limits  $L = 2.86$ ; 10,000 replications were used with no justification provided.
- (4) Khoo and Quah (2003) examined the ARL property of the Hotelling  $\chi^2$  chart using individual observation vectors for 18 different shifts of size  $\delta$ . Only the bivariate case was considered for shifts of size  $\delta$ . They used 10,000 replications, but provided no justification.
- (5) Khoo and Quah (2002) examined the ARL property of two multivariate CUSUM or MCUSUM charts using individual observation vectors for 11 different shifts of size  $\delta$ . The MC1 control chart studied used  $p = 2, 3$ , and 10 variables with reference value  $k = 0.5$  and the MC2 control chart studied used  $p = 2, 3$ , and 10 variables with reference values  $k = 2.5, 3.5$ , and 10.5. A total of 33 different scenarios were

simulated for each MCUSUM chart. They used 10,000 replications, but provided no justification.

- (6) Khoo (2003) examined the ARL property of the multivariate EWMA or MEWMA chart using individual observation vectors for 6 different shifts of size  $\delta$ . The MEWMA control chart studied used  $p = 2, 4$ , and 10 variables and weighting constants  $\lambda = 0.05, 0.10$ , and 0.20. A total of 54 different scenarios were simulated. They used 10,000 replications, but provided no justification.

#### Dissertation IV: Results and Discussion

Statistical control charts are based on the same principles as hypothesis testing. A process is said to be out-of-control if the test of hypotheses is rejected and in-control when it is not rejected, thus, control charts have Type I error rates and power. However, they are typically measured through a different metric, the average run length (ARL). When the process has not changed or shifted, type I error rates can be determined through an in-control ARL. However, when the process has shifted, power can be measured through an out-of-control ARL.

A modified error band, incorporating ARL (e.g.  $ARL \pm 0.1ARL$ ), was used by Chang & Gan (2004) to examine the robustness of the Shewhart control chart with respect to both ARL and SDRL (standard deviation of run length). Chakraborti & van de Wiel (2005) stated this 10% error band might be too wide to detect practical departures of the simulated results from the target value. They used a 2% error band,  $ARL \pm 0.02ARL$ , to examine the robustness of a non-parametric control chart with respect to its ARL. The 2% error band was used in Dissertation IV.

Table 5 displays the number of replications used by the original study along with the recommended ranges for the minimum number of replications needed to produce stable results for various size shifts within the process. Each process shift is recorded in standard deviations. It appears that fewer replications could have been used to predict ARL in each study, particularly when the shift in the process is large.

## NUMBER OF REPLICATIONS REQUIRED IN MONTE CARLO SIMULATION STUDIES

Table 5: Number of Replications Used By the Original Study Along With the Recommended Ranges for the Minimum Number of Replications Required To Produce Stable Results for Various Shifts

Study	Original Replications Used	Replications Recommended						
		Shift						
		0.0	0.1-1.0	1.1-2.0	2.1-3.0	3.1-4.0	4.1-5.0	5.1-6.0
1	10,000	6,329	2,500-3,985	2,031-4,375	1,093-2,265	546	78	---
2	10,000	2,187-7,891	1,093-3,203	312-1,953	78-390	78-703	78-312	---
3	10,000	5,391	1,797-6,407	312-1,641	78-703	---	---	---
4	10,000	5,391	3,125-5,703	3,593-5,703	1,171	937-1,015	312	234
5a	10,000	2,891-7,657	1,015-4,843	156-1,407	156-546	156-703	78-156	---
5b	10,000	4,845-5,157	2,187-5,547	937-2,109	546-2,109	156-625	156-390	---
6	10,000	2,036-9,921	625-5,235	625-1,797	312-703	78-546	78-390	---

### Conclusion

Monte Carlo simulations have been used extensively in studying the performance of various statistical tests and control charts. Researchers have used a wide range (50-21,000 in the 22 studies replicated herein) of replications in their studies, but seldom provided justifications for the number of replications they used. Currently, there are no empirically based recommendations regarding the required number of replications to ensure accurate results.

Through 4 dissertations, 22 studies from various fields were re-analyzed to provide empirically based recommendations for future simulation studies. In many cases, fewer replications than were used in the original studies were needed to produce stable estimates of the results. In all but two of the situations in which more replications than what was used originally were needed, the original studies began with 1,000 or fewer replications. In general, for most of the studies replicated and most of the statistics calculated, the minimum recommended number of replications was

always less than 10,000 and in many cases was less than 5,000. In several situations investigated in these dissertations, 5,000 replications were not sufficient, but seldom were more than 7,500 replications needed. It appears to be the case, generally, that 7,500 to 8,000 replications are sufficient to produce stable results, and in a number of situations, depending upon what characteristic is being estimated, 5,000 replications may be enough.

### References

- Alyounes, Y. (1999). *A realistic look at one-way ANOVA: A comparison of parametric and nonparametric omnibus tests and their post hoc comparison procedures*. Unpublished Doctoral Dissertation, Ohio University, Athens, OH.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

- Brooks, C. (2002). *Solutions to the review questions at the end of chapter 10 in introductory economics for finance*. Retrieved from <http://www.cambridge.org/resources/0521790182/1565-69060.doc>.
- Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, *16*(1), 129-132.
- Chang, T. C., & Gan, F. F. (2004). Shewhart charts for monitoring the variance components. *Journal of Quality Technology*, *36*, 293-308.
- Chakraborti, S., & van de Wiel, M. A. (2005). *A nonparametric control chart based on the Mann-Whitney Statistic*. Unpublished manuscript. Retrieved from <http://projecteuclid.org/euclid.imsc/1207058271>, Digital Object Identifier: doi:10.1214/193940307000000112.
- Fan, X., Felsövályi, Á., Sivo, S. A., & Keenan, S. C. (2002). *SAS for Monte Carlo studies: A guide for quantitative researchers (1<sup>st</sup> Ed.)*. North Carolina: SAS Institute Inc.
- Flack, V. F., & Chang, P. C. (1987). Frequency of selecting noise variables in subset regression analysis: A simulation study. *The American Statistician*, *41*, 84-86.
- Fellner, W. H. (1990). Algorithm AS 258: Average run lengths for cumulative sum schemes. *Applied Statistics*, *39*, 402-412.
- Gamage, J., & Weerahandi, S. (1998). Size performance of some tests in one-way ANOVA. *Communications in Statistics – Simulation and Computation*, *27*(3), 625-640.
- Godfrey, L. G. (1978). Testing for multiplicative heteroscedasticity. *Journal of Econometrics*, *8*, 227-236.
- Griffiths, W. E., & Surekha, K. (1986). A Monte Carlo evaluation of the power of some tests for heteroscedasticity. *Journal of Econometrics*, *31*, 219-231.
- Hsiung, T. H., & Olejnik, S. (1994). Power of pairwise multiple comparison in the unequal variance case. *Communications in Statistics – Simulation and Computation*, *23*(3), 691-710.
- Hurvich, C. M., & Tsai, C. L. (1990). The impact of model selection on inference in linear regression. *The American Statistician*, *44*, 214-217.
- Hutchinson, S. R., & Bandalos, D. L. (1997). A guide to Monte Carlo simulation research for applied researchers. *Journal of Vocational Education Research*, *22*, 233-245.
- Kaplan, A. M. (1984). *A comparison of the Kruskal-Wallis test with ANOVA under violations of homogeneity of variance*. Unpublished Doctoral Dissertation, Temple University, Philadelphia, PA.
- Khoo, M. B. C. (2003). Increasing the sensitivity of MEWMA control chart. *Quality Engineering*, *16*, 75-85.
- Khoo, M. B. C. (2004). Performance measures for the Shewhart  $\bar{x}$  control chart. *Quality Engineering*, *16*, 585-590.
- Khoo, M. B. C., & Quah, S. H. (2002). Computing the Percentage Points of the Run-Length Distributions of Multivariate CUSUM Control Charts. *Quality Engineering*, *15*, 299-310.
- Khoo, M. B. C., & Quah, S. H. (2003). Incorporating run rules into Hotelling's  $\chi^2$  control charts. *Quality Engineering*, *15*, 671-675.
- Kim, C. (1997). *Robust tests using weighted likelihood estimation*. Unpublished Doctoral Dissertation, Oklahoma State University, Stillwater, OK.
- Kim, M. (2005). *Number of replications required in control chart Monte Carlo simulation studies*. Unpublished Doctoral Dissertation, University of Northern Colorado, Greeley, CO.
- Klockers, A. J., & Hancock, G. R. (1992). Power of recent multiple comparison procedures as applied to a complete set of planned orthogonal contrasts. *Psychological Bulletin*, *111*(3), 505-510.
- Morikawa, T., Terao, A., & Iwasaki, M. (1996). Power evaluation of various modified Bonferroni procedures by a Monte Carlo study. *Journal of Biopharmaceutical Statistics*, *6*(3), 343-359.
- Olejnik, S., Mills, J., & Keselman, H. (2000). Using Wherry's adjusted  $R^2$  and Mallows'  $C_p$  for model selection from all possible regressions. *The Journal of Experimental Education*, *68*(4), 365-380.



## NUMBER OF REPLICATIONS REQUIRED IN MONTE CARLO SIMULATION STUDIES

Neubauer, A. S. (1997). The EWMA control chart; properties and comparison with other quality-control procedures by computer simulation. *Clinical Chemistry*, 43, 594-601.

Pfaffenberger, R. C., & Dielman, T. E. (1991). Testing normality of regression disturbances: A Monte Carlo study of the Filliben test. *Computational Statistics & Data Analysis*, 11, 265-273.

Preecha, C. (2004). *Numbers of replications required in ANOVA simulation studies*. Unpublished Doctoral Dissertation, University of Northern Colorado, Greeley, CO.

Ramsey, P. H. (2002). Comparison of closed testing procedures for pairwise testing of means. *Psychological Methods*, 7(4), 504-523.

Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. *Psychological Bulletin*, 110(3), 557-586.

Supawan, P. (2004). *An examination of the number of replications required in regression simulation studies*. Unpublished Doctoral Dissertation, University of Northern Colorado, Greeley, CO.

Ussawarujikulchai, A. (2004). *Number of replications required in multiple comparison procedures Monte Carlo simulation studies*. Unpublished Doctoral Dissertation, University of Northern Colorado, Greeley, CO.

Welch, B. L. (1947). The generalization of Students' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28-35.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4), 330-336.

## Bias in Monte Carlo Simulations Due To Pseudo-Random Number Generator Initial Seed Selection

Jack C. Hill  
Beaumont Health System

Shlomo S. Sawilowsky  
Wayne State University

---

Pseudo-random number generators can bias Monte Carlo simulations of the standard normal probability distribution function with initial seeds selection. Five generator designs were initial-seeded with values from  $10000_{\text{HEX}}$  to  $1\text{FFFF}_{\text{HEX}}$ , estimates of the mean were calculated for each seed, the distribution of mean estimates was determined for each generator and simulation histories were graphed for selected seeds.

Key words: Pseudo-random generator, number sequences, initial seed, bias, Monte Carlo.

---

### Introduction

It is possible, using a vetted pseudo-random number generator and a specific initial seed, to produce number sequences that may have very nonrandom-like characteristics; however, restarting the algorithm with a new seed will produce an excellent number sequence. Park and Miller (1988) asserted that coding errors in the algorithm might be responsible, although, in a vetted generator, coding errors are unlikely. (See Fishman (1995) for descriptions of other possible reasons for this phenomenon.)

### Pseudo-Random Number Generator Design

Lehmer (1951) suggested a simple computer algorithm to generate a random number sequence whose period is limited only by the arithmetic registers of the digital computer and the parameters selected. This linear recurrence (congruential) generator is the basic and most widely used class of random number generators. Linear recurrence (congruential) generators are of the form

$$Z_i = \sum (A_s Z_{i-1} + B) \text{ mod } M$$
$$\text{for } 0 \leq i < 2^{31} - 1 \text{ and } 1 \leq s \leq X$$
(1.1)

where  $0 < A_1, A_2, \dots, A_x < M$ ;  $A, B$  and  $M$  are integers, and  $X$  is the order of the generator (the number of non-zero  $A_s$ ). The most widely used first-order,  $X = 1$ , form of the linear congruential generator is the multiplicative (linear) congruential generator:

$$Z_i = (AZ_{i-1} + B) \text{ mod } M$$
$$\text{for } 0 \leq i < 2^{31} - 1 \text{ and } Z_0 \text{ is the initial seed}$$
(1.2)

where  $0 < A < M$ ;  $A$  and  $M$  are integers and  $B = 0$ .

After the terminology and symbols of L'Ecuyer (1993), the following short notation defines random number generators:

$$\text{XXX} (M, [A_{1,2,3\dots n}], B, Z_0),$$

where XXX is the abbreviation of the type of generator (e.g., MCG = Multiplicative Congruential Generator; MRG = Multiple Recursive Generator; and MWC = Multiply-With-Carry multiplicative congruential generator),  $M$  is the modulus,  $A$  is the array of multipliers or coefficients,  $B$  is the increment, and  $Z_0$  is the initial seed.

---

Jack Hill is a BI Database Architect and adjunct faculty in statistics and quantitative analysis. Email him at: hill3jc@cmich.edu. Shlomo Sawilowsky is a professor of educational statistics, and editor of *JMASM*. Email him at: shlomo@wayne.edu.

## PSEUDO-RANDOM NUMBER INITIAL SEED BIAS IN MONTE CARLO SIMULATIONS

MCG ( $2^{31}$ , 65539, 0, 1), also known as RANDU, is an example of a generator whose number sequences are recognized as poorly randomized and whose period is relatively short at  $2^{29}$  (Park & Miller, 1988):

$$Z_i = 65539 * Z_{i-1} \text{ mod } 2^{31}$$

for  $1 \leq i \leq 2^{31} - 1$  and  $Z_0$  the initial seed

(1.3)

The set, S, of all numbers, period,  $p = 2^{29}$ , can be generated using an initial seed,  $Z_0 = 1$  (or any other value,  $0 < Z_0 \leq 2^{31}-1$ ). An initial seed ( $Z_0$ ) determines the first number in the set S where values will be drawn, and each subsequent value is determined by recursively evaluating equation 1.3.

MCG ( $2^{31}-1$ , 16807, 0, 1), also known as Minimum Standard (Park & Miller, 1988), used in RANGEN (Sawilowsky & Blair, 1987; Fahome & Sawilowsky, 2001) is included as an example of a generator whose number sequences are generally recognized as good for a  $2^{31}-1$  period generator. Park and Miller considered this an example of a good minimum standard generator and it has endured decades of study and testing:

$$Z_i = 16807 * Z_{i-1} \text{ mod } (2^{31} - 1)$$

for  $1 \leq i \leq 2^{31} - 1$  and  $Z_0$  the initial seed

(1.4)

where  $Z_i$  is the current value,  $Z_{i-1}$  is the prior value. When  $i = 0$ ,  $Z_0$  is the initial seed, the generator has been generally accepted as a good random number generator and has been used in numerous Monte Carlo studies. The set, S, of all numbers, period,  $p = 2^{31}-1$ , can be generated using an initial seed,  $Z_0 = 1$  (or any other value,  $0 < Z_0 \leq 2^{31}-1$ ). An initial seed,  $Z_0$ , determines the first number in the set, S, where values will be drawn, and each subsequent value is determined by recursively evaluating equation 1.4.

MCG (231-1, 630360016, 0, 1) is also reported to be a good MCG (Entacher, 1998):

$$Z_i = 630360016 * Z_{i-1} \text{ mod } (2^{31} - 1)$$

for  $1 \leq i \leq 2^{31} - 1$  and  $Z_0$  the initial seed

(1.5)

The set, S, of all numbers, period,  $p = 2^{31}-1$ , can be generated using an initial seed,  $Z_0 = 1$  (or any other value,  $0 < Z_0 \leq 2^{31}-1$ ). An initial seed,  $Z_0$ , determines the first number in the set, S, where values will be drawn, and each subsequent value is determined by recursively evaluating equation 1.5.

L'Ecuyer (1993) reported testing several higher-order linear congruential generators or multiple recursive generators having the form

$$Z_i = (A_1 Z_{i-1} + A_2 Z_{i-2} + \dots + A_x Z_{i-x}) \text{ mod } M$$

for  $0 \leq i < 2^{31} - 1$  and  $Z_{0,1,2,\dots,x-1}$  are the initial seeds

(1.6)

where  $0 < A_1, A_2, \dots, A_x < M$ ; A and M are integers and X is the generator's order.

One of L'Ecuyer's best performing multiple recursive generators is MRG (231-1, [2001982722, 1412284257, 1155380217, 1668339922], 0, 1), also known as LECUYER, a 4<sup>th</sup>-order multiple recursive generator whose output sequence is reported to be good:

$$Z_i = (2001982722 * Z_{i-1} + 1412284257 * Z_{i-2} + 1155380217 * Z_{i-3} + 1668339922 * Z_{i-4}) \text{ mod } (2^{31} - 1)$$

for  $1 \leq i \leq 2^{31}$  and  $Z_{0,1,2,\dots,x-1}$  are the initial seeds

(1.7)

The set, S, of all numbers, period,  $p \equiv 2^{158}$ , can be generated using an initial seed,  $Z_0 = 1$  (or any other value,  $0 < Z_0 \leq 2^{31}-1$ ). An initial seed,  $Z_0$ , determines the first number in the set, S, where values will be drawn, and each subsequent value is determined by recursively evaluating Equation 1.7.

Marsaglia (1994b) proposed a variation of the multiplicative congruential generator called multiply-with-carry of the form

$$Z_i = (A_1 Z_{i-4} + A_2 Z_{i-3} + A_3 Z_{i-2} + A_4 Z_{i-1} + C) \text{ mod } M$$

for  $0 \leq i < 2^{31} - 1$  and  $Z_{0,1,2,3}$  are initial seeds

(1.8)

where  $0 < A_1, A_2, A_3, A_4 < M$ ; A and M are integers (Gentle, 2003). Marsaglia's design generates a 64-bit sum of four products and a carry; these 64 bits are separated into two 32-bit words. The upper 32-bits become the new carry, C, the lower 32-bits (modulo M) is the new  $Z_i$  and each subsequent value is determined by recursively evaluating equation 1.8.

MWC (232, [211111111111, 1492, 1776, 5115], C, 1) (Marsaglia, 1994a), is implemented in a subroutine library known as MOTHER. (The version considered here is due Miller, 1995, adapted and updated to Fortran 90 by Blair, 1999, in a subroutine library called BFRA. It is an implementation of Luescher's, 1994, algorithm and James', 1994, Fortran 77 coding of the "Luxury" generator.) It is included as an example of a popular form of the multiply-with-carry, random number generators:

$$Z_i = (211111111111 * Z_{i-1} + 1492 * Z_{i-2} + 1776 * Z_{i-3} + 5115 * Z_{i-4} + C) \text{ mod } 2^{32}$$

for  $0 \leq i < 2^{31} - 1$  and  $Z_{0,1,2,3}$  are initial seeds

(1.9)

The set, S, of all numbers, period,  $p = 2^{158}$ , can be generated using an initial seed,  $Z_0 = 1$  (or any other value,  $0 < Z_0 \leq 2^{31} - 1$ ). An initial seed,  $Z_0$ , determines the first number in the set, S, where values will be drawn, and each subsequent value is determined by recursively evaluating equation 1.9.

#### Pseudo-Random Number Sequences

L'Ecuyer and Hellekalek (1997) suggested a mental-model of a huge roulette wheel to visualize the sequence of unscaled numbers from a generator such as RANGEN (equation 1.4) where the sequenced numbers appear once, in the range of 1 to  $2^{31} - 1$ . The order of the number sequence is determined by the RANGEN algorithm based on the previous number (seed). Spinning the roulette wheel would afford the researcher a random starting

point (initial seed) in the number sequence; however, because the number sequence is calculated based on the previous number, the sequence is truly deterministic in nature; it only has the appearance of randomness.

These number sequences can be easily scaled to the unit interval,  $U[0,1]$ , by dividing each number by P, the Period of the generator. RANGEN's period =  $2^{31} - 1$ . Resetting the algorithm to the same initial seed reproduces the identical random number-sequence. Individual numbers from these sequences are used as indices that determine the locations on a probability density function or in the data table where samples should be drawn.

It should be noted that, due to their deterministic nature, different initial seeds could generate overlapping number sequences. For example, a hypothetical random number generator using an initial seed  $S_0$  calculated the following number sequence where  $S_0 = N_0$ :

$$N_0, N_1, N_2, N_3, N_4, N_5, N_6, N_7, N_8, N_9, N_{10} \dots N_n$$

If a different initial seed,  $S_0 = N_3$ , is used, the following number sequence would be calculated:

$$N_3, N_4, N_5, N_6, N_7, N_8, N_9, N_{10} \dots N_n$$

These number sequences clearly overlap beginning at  $N_3$ ; however, the Monte Carlo simulations reported herein were not adjusted for possible overlapping number sequences.

#### Methodology

The purpose of this study was to determine the bias introduced into Monte Carlo simulation studies resulting from initial seed selection. The random number sequences of five random number generators were analyzed. Each random number generator was seeded with bit-patterns sequenced from an initial value of 10000<sub>Hex</sub> (65,536) to a final value of 1FFFF<sub>Hex</sub> (131,071).

The process was as follows: Use each random number generator's scaled number sequence from the unit interval ( $U[0,1]$ ,  $\text{Seed}_0 = S_0$ ). Select a sample, S, of sample size,  $N = 10$ , from the standard normal probability density function. Calculate the sample mean,  $X_S = 1/N \sum S$ , of these values. Continue sampling for T trials and calculate the overall mean,  $X_T = 1/T \sum X_S$ , of

## PSEUDO-RANDOM NUMBER INITIAL SEED BIAS IN MONTE CARLO SIMULATIONS

T trials. For the standard normal probability density function, with a large number of trials, T, it is expected that the sample statistic,  $X_T$ , will converge on the population parameter mean,  $\mu=0$ , and the population standard deviation,  $\sigma=1$ .

### Results

The population mean and standard deviation were estimated for each  $1000_{\text{HEX}}$  (65,536)  $\leq$  Initial Seed  $\leq 1\text{FFFF}_{\text{HEX}}$  (131,071) using five random number generators at Trials = 1,000, 10,000, 100,000 and 1,000,000 (Table 1.1). With the exception of MOTHER, the average mean estimate Z score = 0.000 and average standard deviation estimate = 1.000. The distribution of

mean estimate Z scores for each random number generator at each trial is normal-like, mean = 0, and standard deviation = 1.00 with minimum Z scores of -5.07 to -3.34 and maximum Z scores of 3.09 to 4.59 (Table 1.2).

### RANGEN Pseudo-Random Number Generator

RANGEN's mean estimate distributions at Trials = 1,000, 10,000, 100,000 and 1,000,000 are typical of the five random number generators studied (Figures 1.1, 1.2, 1.3, 1.4). Each graph shows the normalized SPSS histogram of all mean estimates from Monte Carlo simulations using each initial seed with the standard normal distribution superimposed.

Table 1.1: Mean and Standard Deviation Estimates for Each Random Number Generator of Monte Carlo Simulations of the Standard Normal Probability Density Function Using All Initial Seeds from 10000Hex (65,536) through 1FFFFHex (131,072); Sample Size = 10, Trials = 1,000, 10,000, 100,000 and 1,000,000

Trials	1,000		10,000		100,000		1,000,000	
RNG	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
RANGEN	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
ENTACHER	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
LECUYER	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
MOTHER	-0.001	1.000	-0.005	1.000	-0.008	1.000	-0.005	1.000
RANDU	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000

Table 1.2: The Maximum and Maximum Mean Estimates For Each Random Number Generator of Monte Carlo Simulations of the Standard Normal Probability Density Function Using All Initial Seeds from 10000Hex (65,536) through 1FFFFHex (131,072); Sample Size = 10, Trials = 1,000, 10,000, 100,000, and 1,000,000

Trials	1,000		10,000		100,000		1,000,000	
RNG	Min	Max	Min	Max	Min	Max	Min	Max
RANGEN	-3.83	4.04	-4.46	4.18	-4.31	4.12	-3.34	3.09
ENTACHER	-4.35	4.05	-4.08	4.41	-4.14	4.15	-4.34	4.35
LECUYER	-4.01	4.01	-4.51	4.23	-4.13	4.32	-4.00	4.07
MOTHER	-4.02	4.10	-4.51	4.42	-5.07	4.58	-4.32	4.59
RANDU	-3.86	4.18	-4.21	4.04	-3.76	3.79	-3.86	3.84

Figure 1.1: RANGEN Random Number Generator  
 A Normalized SPSS Histogram Graph of Monte Carlo Simulations of Mean Estimates of the Standard Normal Probability Density Function Using All Initial Seeds from 10000<sub>Hex</sub> (65,536) through 1FFFF<sub>Hex</sub> (131,072); Sample Size = 10, Trials = 1,000

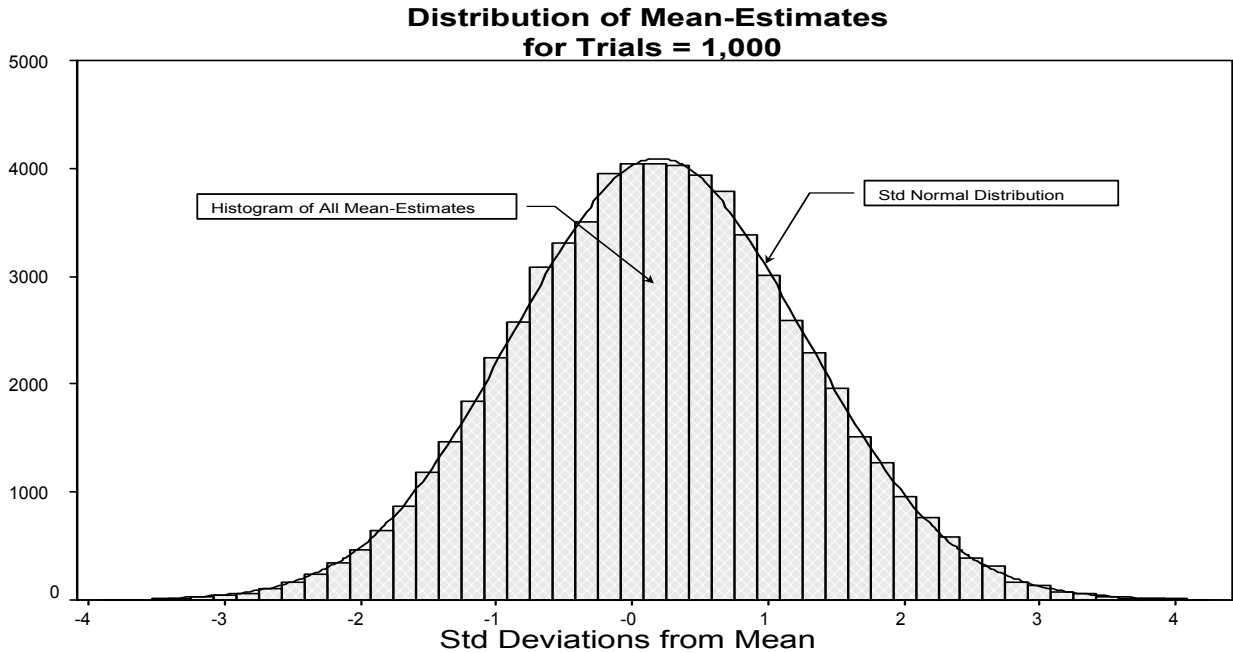
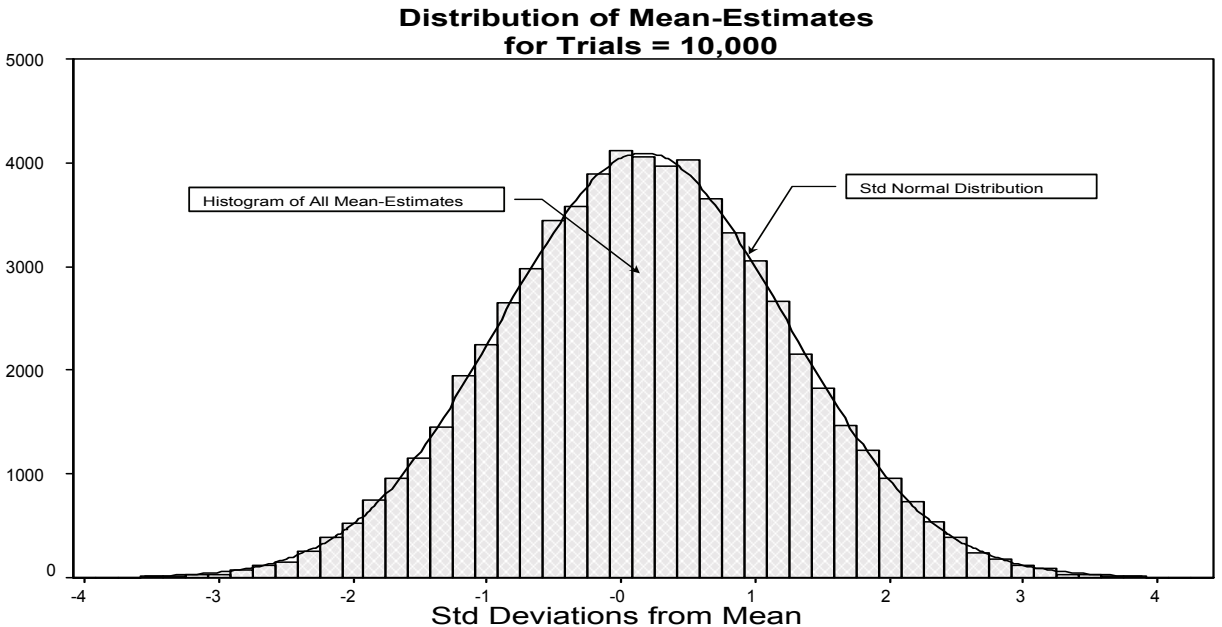


Figure 1.2: RANGEN Random Number Generator  
 A Normalized SPSS Histogram Graph of Monte Carlo Simulations of Mean Estimates of the Standard Normal Probability Density Function Using All Initial Seeds from 10000<sub>Hex</sub> (65,536) through 1FFFF<sub>Hex</sub> (131,072); Sample Size = 10, Trials = 10,000



# PSEUDO-RANDOM NUMBER INITIAL SEED BIAS IN MONTE CARLO SIMULATIONS

Figure 1.3: RANGEN Random Number Generator

A Normalized SPSS Histogram Graph of Monte Carlo Simulations of Mean Estimates of the Standard Normal Probability Density Function Using All Initial Seeds from 10000<sub>Hex</sub> (65,536) through 1FFFF<sub>Hex</sub> (131,072); Sample Size = 10, Trials = 100,000

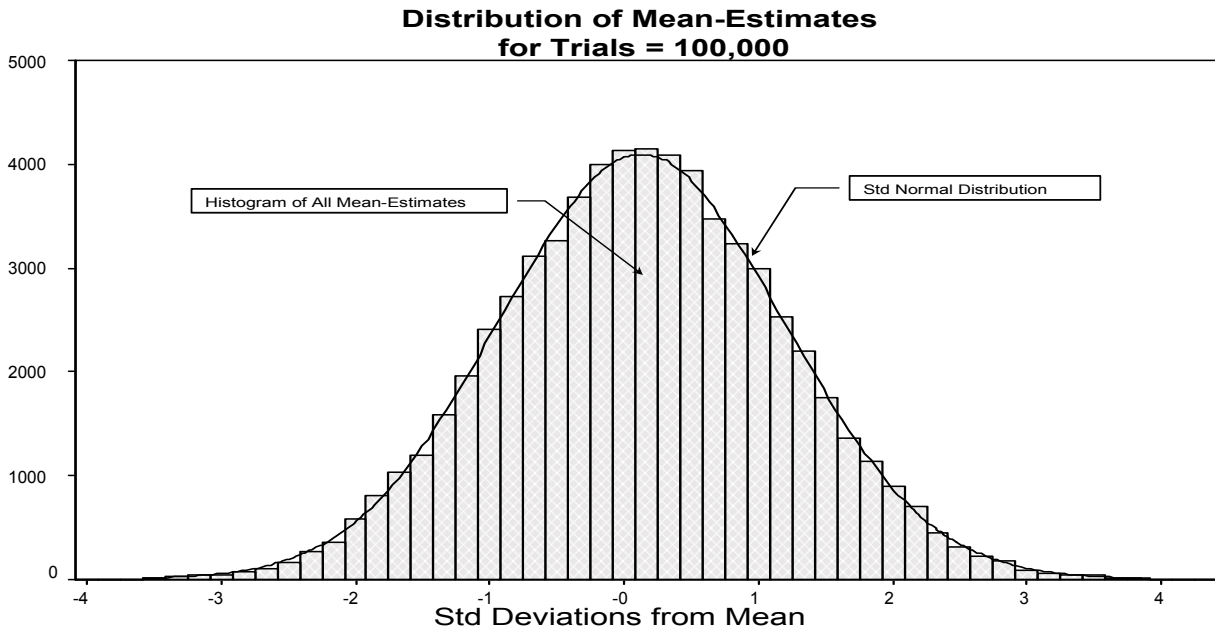


Figure 1.4: RANGEN Random Number Generator

A Normalized SPSS Histogram Graph of Monte Carlo Simulations of Mean Estimates of the Standard Normal Probability Density Function Using All Initial Seeds from 10000<sub>Hex</sub> (65,536) through 1FFFF<sub>Hex</sub> (131,072); Sample Size = 10, Trials = 1,000,000

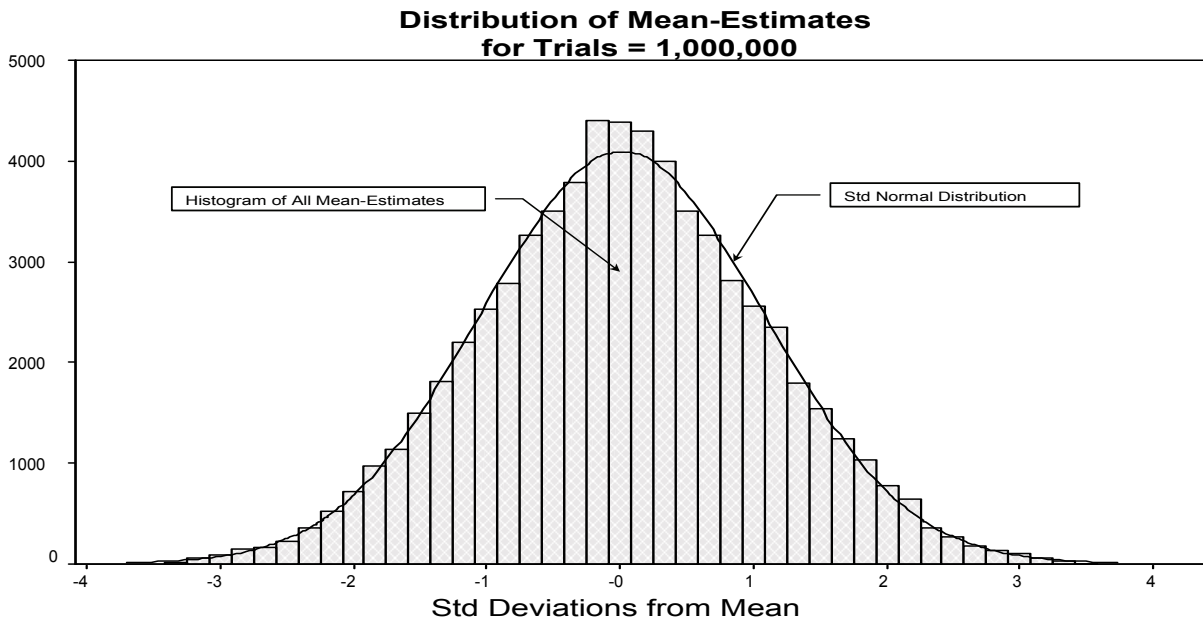


Table 1.3: RANGEN Random Number Generator

SPSS Extreme Values Table of Normalized Mean Estimates from Monte Carlo Simulations of the Standard Normal Probability Density Function Using All Initial Seeds,  $S_i = 10000\text{Hex}$  through  $1\text{FFFFHex}$  at Trials = 1,000, 10,000, 100,000, and 1,000,000

Trials	1,000		10,000		100,000		1,000,000	
Deviation	Seed	z-Score	Seed	z-Score	Seed	z-Score	Seed	z-Score
Above	66184	4.04	93552	4.18	92157	4.12	72114	3.09
	114634	3.80	68018	3.95	91955	4.11	70120	3.08
	87579	3.80	90064	3.84	91223	3.94	115292	3.04
	98710	3.73	85895	3.73	117600	3.94	118753	3.04
	118038	3.73	73471	3.73	116943	3.88	95913	3.03
Below	108848	-3.65	96630	-4.04	82111	-3.92	76920	-3.31
	79125	-3.69	114546	-4.14	121739	-4.16	68326	-3.31
	73787	-3.73	101156	-4.18	113374	-4.16	79462	-3.32
	87518	-3.74	105875	-4.34	65647	-4.17	77341	-3.32
	102745	-3.83	125434	-4.46	105966	-4.31	106675	-3.34

RANGEN Monte Carlo Simulation (MCS) History Graphs

For RANGEN’s distribution of mean estimates at Trial = 1,000,000, fifteen initial seeds were chosen, five from each extreme (the worst) and five from the center (the best). The mean estimates’ initial seeds were used to capture the cumulative (running) average at 10,000-trial intervals of a Monte Carlo simulation of the standard normal probability distribution function. These data were normalized and graphed, (Figure 1.5), as a graphical history of the Monte Carlo simulation’s mean estimate at 10,000-trial intervals. RANGEN’s cumulative mean-estimates undulated wildly  $\pm 15$  standard deviations for the first 100,000 trials and then converged on their respective region of the mean estimate distribution.

Individual RANGEN MCS history graphs for the two worst initial seeds (Table 1.3), 106675 and 72114, and closest to the true mean, 117656, were isolated. The RANGEN MCS history graph (Figure 1.6) for initial seed,  $S_i = 106675$ , demonstrated an immediate

negative departure from zero of  $-6$  standard deviations at trials = 20,000; a recovery to  $-2$  standard deviations at trials = 100,000; a sharp drop to  $-5$  standard deviations at trials = 140,000; another recovery attempt at trials = 190,000; and a final convergence on  $-3$  standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

The RANGEN MCS history graph (Figure 1.7) for initial seed,  $S_i = 117656$ , demonstrated an initial negative departure from zero,  $-9$  standard deviations, at trials = 10,000; a recovery,  $-0.5$  standard deviations, at trials = 80,000; and a final convergence on  $0.0$  standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated some wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

The RANGEN MCS history graph (Figure 1.8) for initial seed,  $S_i = 72114$ , demonstrated an initial negative departure from



# PSEUDO-RANDOM NUMBER INITIAL SEED BIAS IN MONTE CARLO SIMULATIONS

zero, -13 standard deviations at trials = 10,000; a recovery, +9 standard deviations, at trials = 50,000; a sharp drop, +2 standard deviations, at trials = 100,000; and a final convergence on +3 standard deviations. Individual 10,000-trial intervals, designated as ♦ on the graph, demonstrated wildly fluctuating mean estimates of ±15 standard deviations throughout the simulation run.

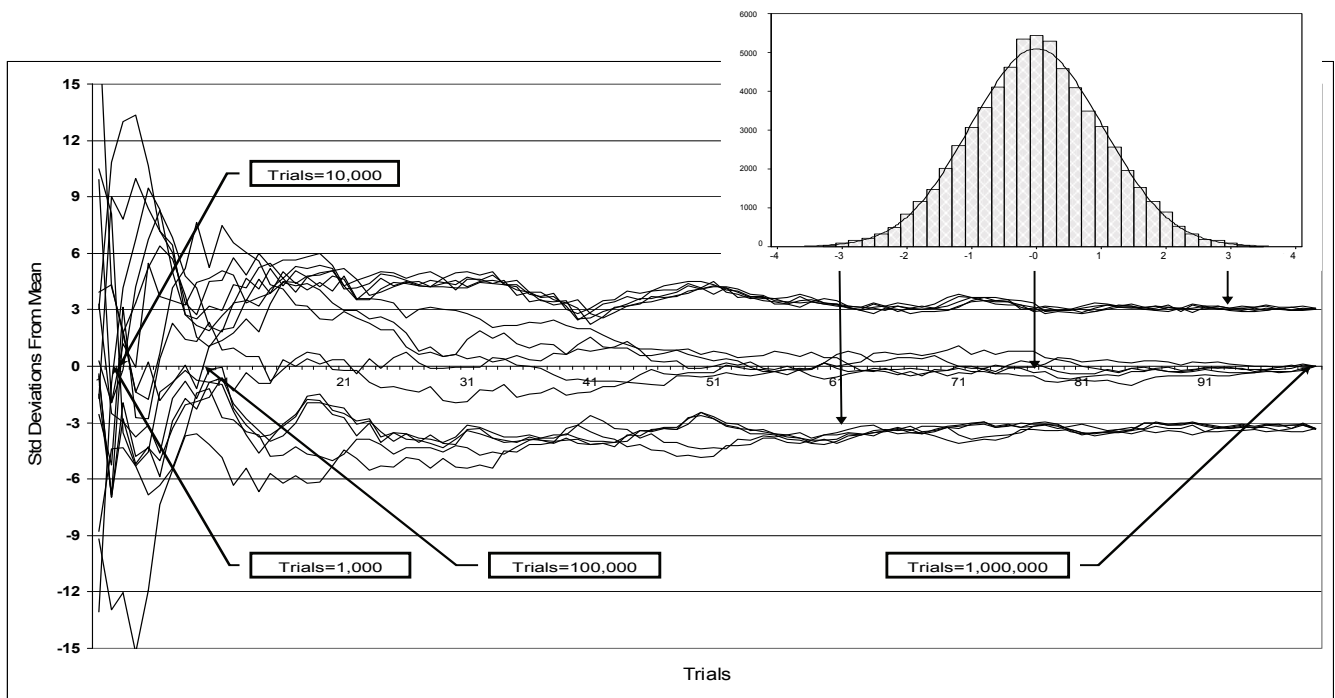
## MOTHER Pseudo-Random Number Generator

For MOTHER's distribution of mean estimates at Trial = 1,000,000, fifteen initial seeds were chosen, five from each extreme, the worst, and five from the center, the best. The mean estimates' initial seeds were used to capture the cumulative (running) average at

10,000-trial intervals of a Monte Carlo simulation of the Standard Normal probability distribution function. These data were normalized and graphed, (Figure 2.1), as graphical history of the Monte Carlo simulation's mean-estimate at 10,000-trial intervals. MOTHER's cumulative mean estimates undulated wildly ±15 standard deviations for the first 100,000 trials and then converged on their respective region of the mean estimate distribution.

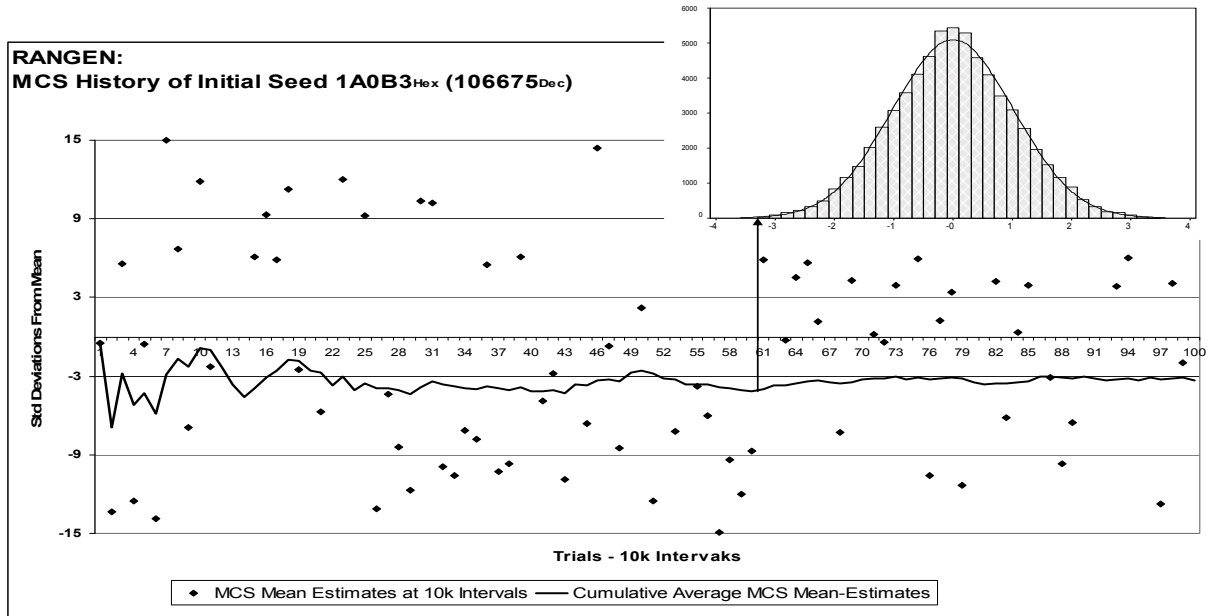
Individual MOTHER MCS history graphs for the two worst initial seeds, 83815 and 112145, and closest to the true mean, 126281, were isolated. The MOTHER MCS history graph (Figure 2.2) for initial seed,  $S_i = 83815$ , demonstrated an immediate negative

Figure 1.5: RANGEN Random Number Generator



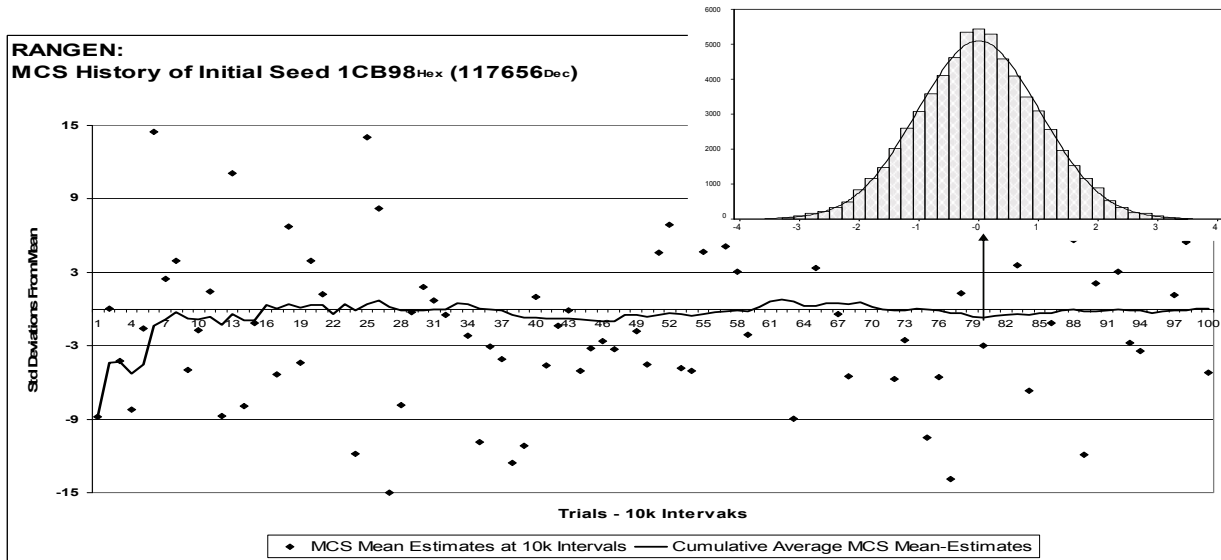
Each line in the graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Fifteen initial seeds (five each: high, middle and low mean estimates) are shown at 10k intervals, total trials = 1,000,000. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Down arrows correlate the  $-3\sigma$ ,  $0$ , and  $+3\sigma$  points of the standard normal probability density function to the horizontal cumulative averages and their deviations from the mean.

Figure 1.6: RANGEN Random Number Generator



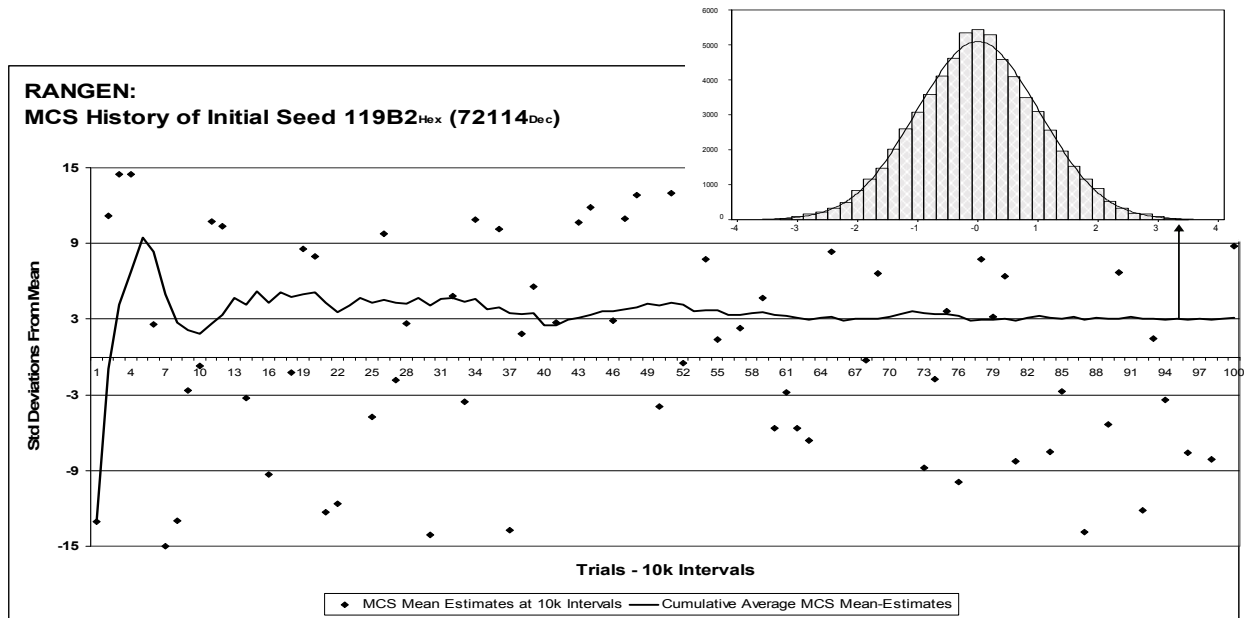
This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 106675, is the worst mean estimate below the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 106675, on the standard normal probability density function.

Figure 1.7 RANGEN Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 117656, is the best mean estimate. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 117656, on the standard normal probability density function.

Figure 1.8: RANGEN Random Number Generator



This Graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 72114, is the worst mean estimate above the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 72114, on the standard normal probability density function.

departure from zero, -6 standard deviations, at trials = 10,000; a recovery to -3 standard deviations at trials = 50,000; and a final gradual convergence on -4 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

The MOTHER MCS history graph (Figure 2.3) for initial seed,  $S_i = 126281$ , demonstrated an initial negative departure from zero, -15 standard deviations, at trials = 10,000 and a final convergence on 0.0 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated some wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

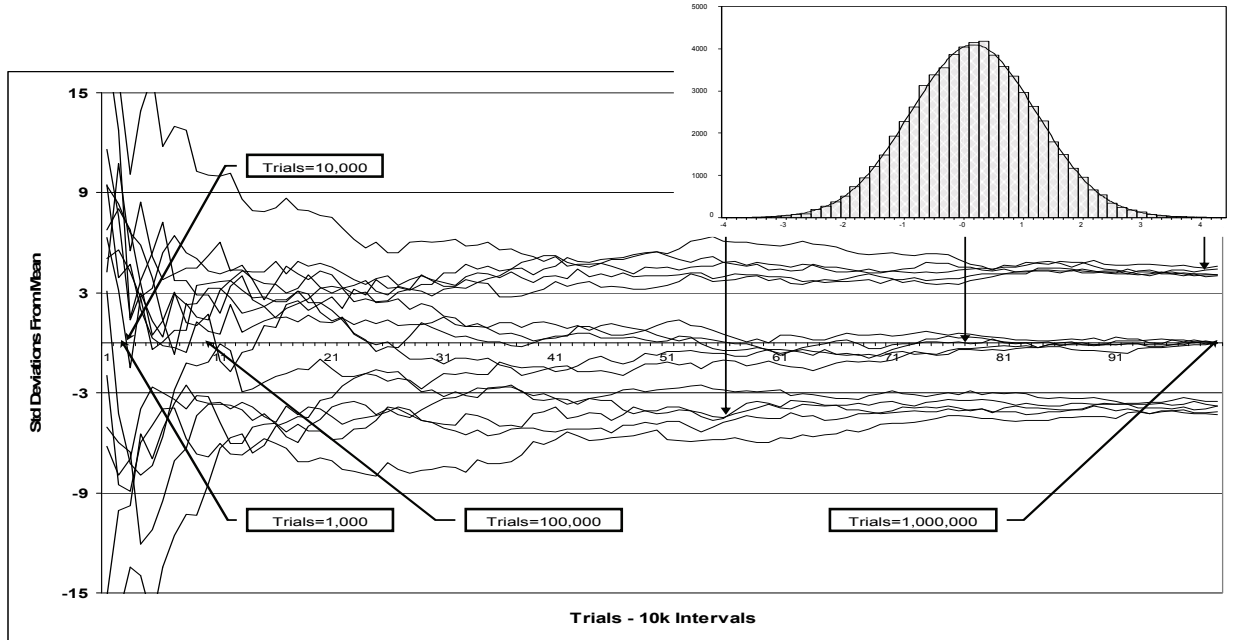
The MOTHER MCS history graph (Figure 2.4) for initial seed,  $S_i = 112145$ , demonstrated an initial departure from zero, 15 standard deviations, at trials = 10,000 and a final convergence on +4 standard deviations.

Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

#### ENTACHER Pseudo-Random Number Generator

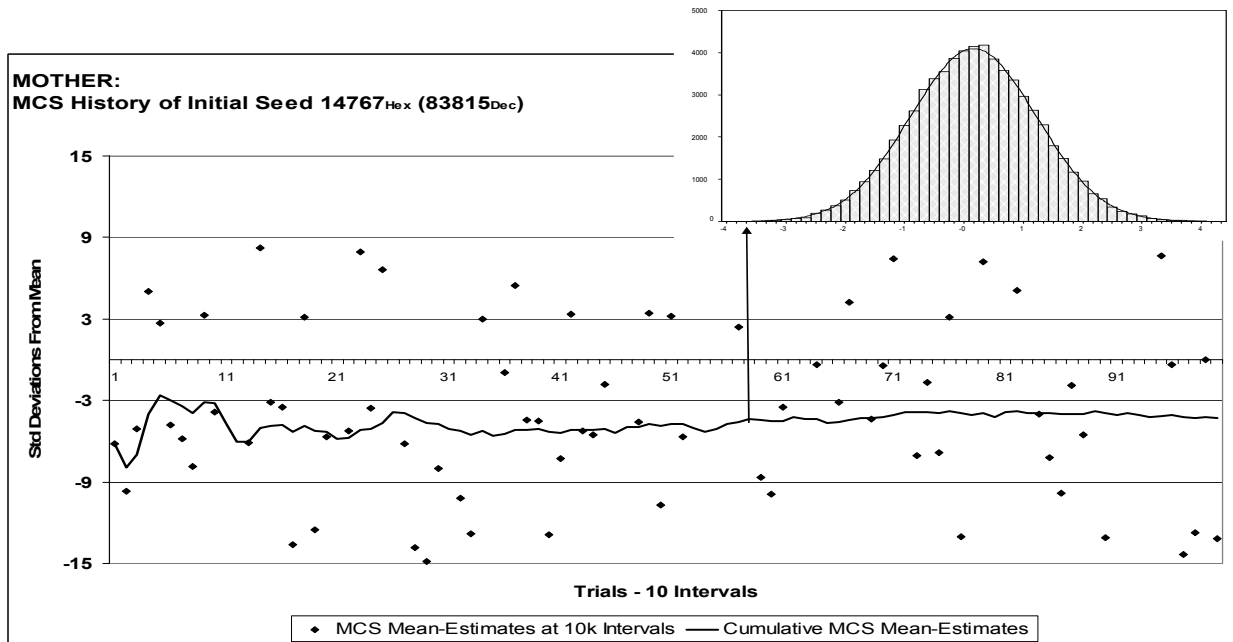
For ENTACHER's distribution of mean estimates at Trial = 1,000,000, fifteen initial seeds were chosen, five from each extreme (the worst) and five from the center (the best). The mean estimates' initial seeds were used to capture the cumulative (running) average at 10,000-trial intervals of a Monte Carlo simulation of the Standard Normal probability distribution function. These data were normalized and graphed, (Figure 3.1), as graphical history of the Monte Carlo simulation's mean-estimate at 10,000-trial intervals. ENTACHER's cumulative mean estimates undulated wildly ( $\pm 15$  standard deviations) for the first 100,000 trials and then converged on their respective region of the mean estimate distribution.

Figure 2.1: MOTHER Random Number Generator



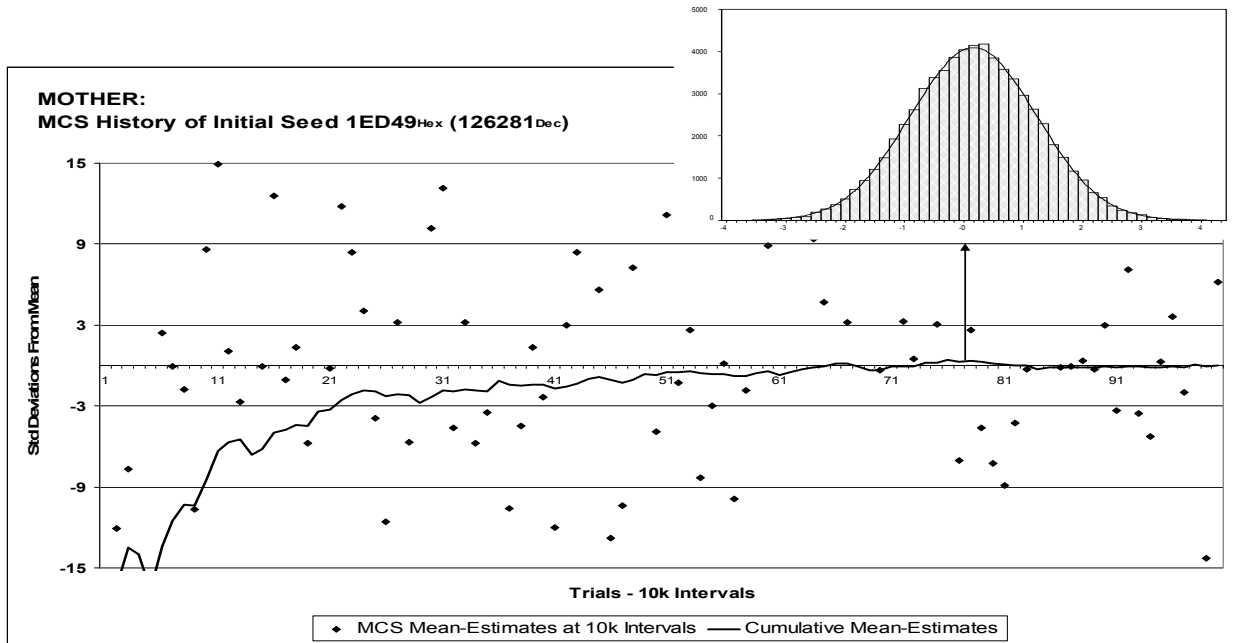
Each line in the graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Fifteen initial seeds (five each: high, middle and low mean estimates) are shown at 10k intervals, total trials = 1,000,000. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Down arrows correlate the  $3\sigma$ , 0, and  $+3\sigma$  points of the standard normal probability density function to the horizontal cumulative averages and their deviations from the mean.

Figure 2.2: MOTHER Random Number Generator



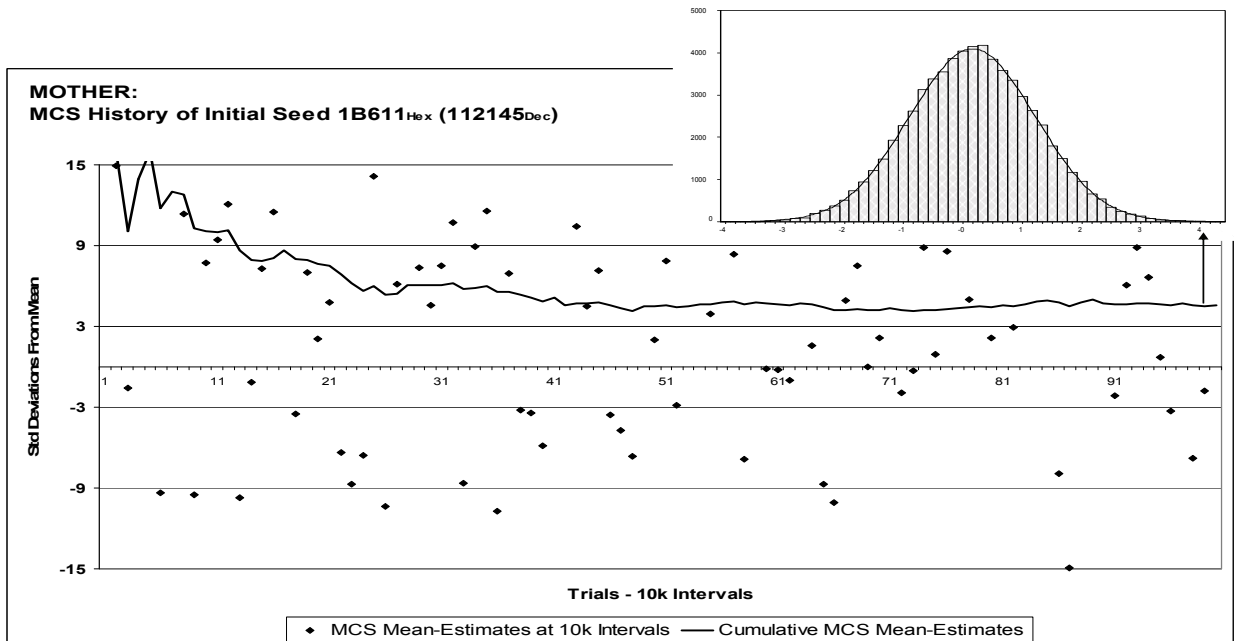
This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 83815, is the worst mean estimate below the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 83815, on the standard normal probability density function.

Figure 2.3: MOTHER Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 126281, is the best mean estimate. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 126281, on the standard normal probability density function.

Figure 2.4 MOTHER Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 112145, is the worst mean estimate above the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 112145, on the standard normal probability density function.

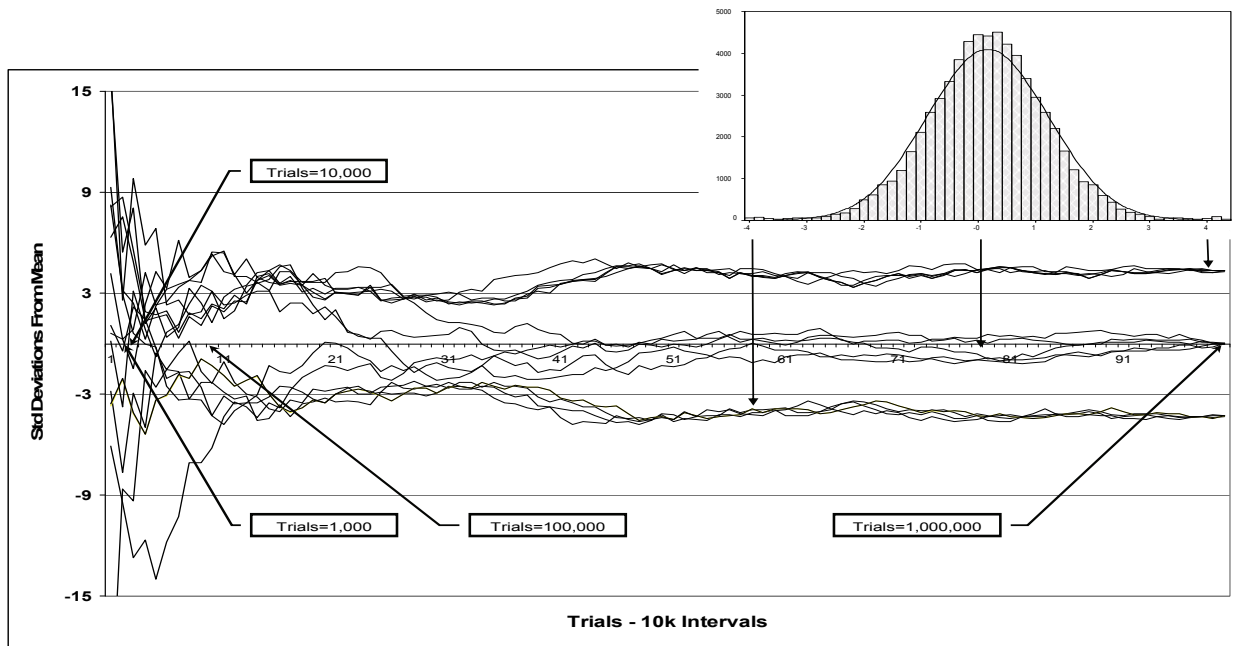
Individual ENTACHER MCS history graphs for the two worst initial seeds, 82315 and 73489, and closest to the true mean, 120803, were isolated. The ENTACHER MCS history graph (Figure 3.2) for initial seed,  $S_i = 82315$ , demonstrated an immediate departure from zero, -15 standard deviations, at trials = 10,000; a recovery to -2 standard deviations at trials = 30,000; and a final convergence on -4 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

The ENTACHER MCS history graph (Figure 3.3) for initial seed,  $S_i = 120803$ , demonstrated an immediate departure from zero, +8 standard deviations, at trials = 10,000; a recovery to +0.5 standard deviations at trials =

30,000; a sharp jump to +5 standard deviations at trials = 80,000; and a final convergence on 0.0 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

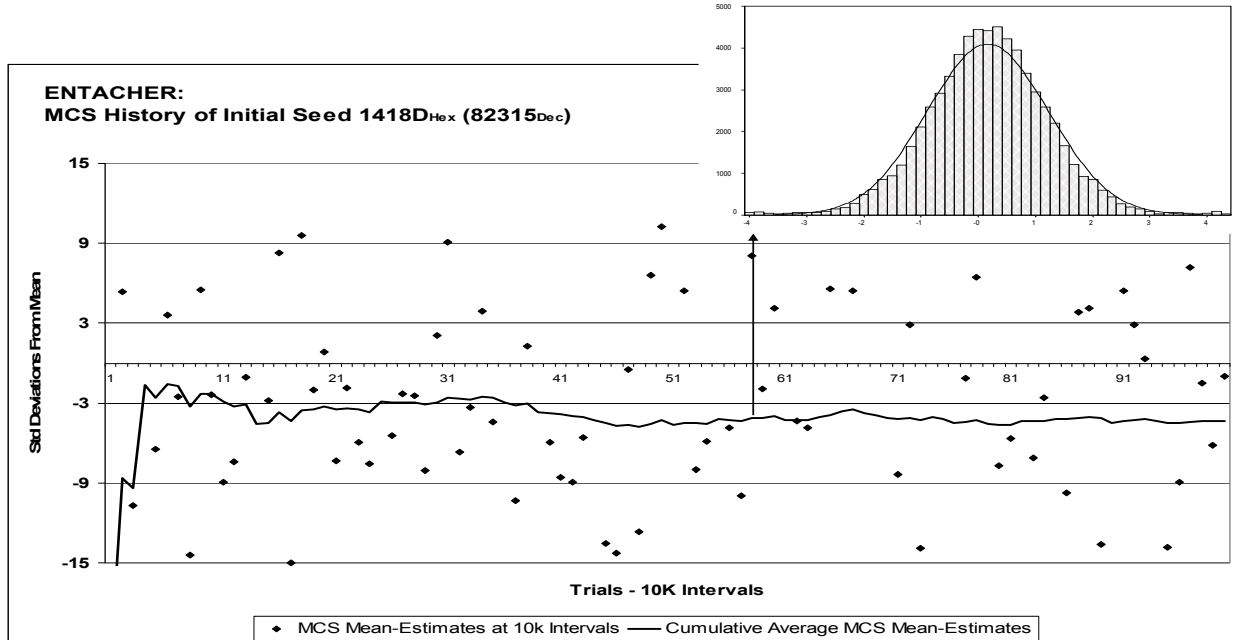
The ENTACHER MCS history graph (Figure 3.4) for initial seed,  $S_i = 73489$ , demonstrated an initial departure from zero, +15 standard deviations, at trials = 10,000; a recovery, +4 standard deviations, at trials = 200,000; and a final convergence on +4 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

Figure 3.1: ENTACHER Random Number Generator



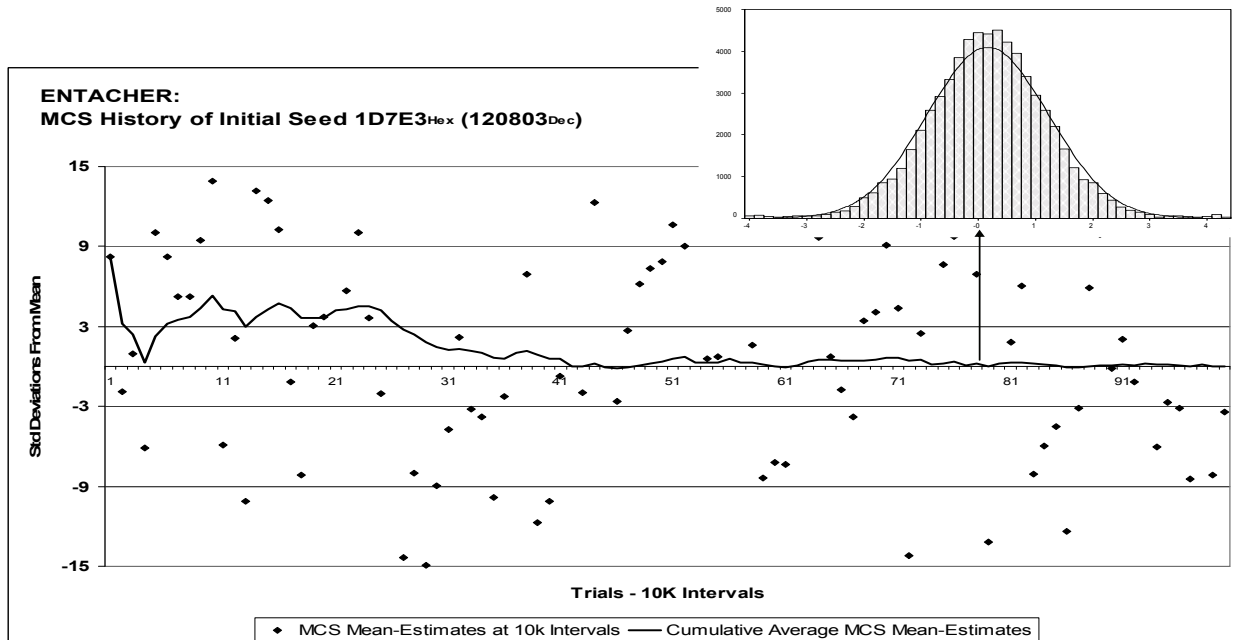
Each line in the graph follows the cumulative mean estimate of a Monte Carlo Simulation of the standard normal probability density function. Fifteen initial seeds (five each: high, middle and low mean estimates) are shown at 10k intervals, total trials = 1,000,000. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Down arrows correlate the  $3\sigma$ , 0, and  $+3\sigma$  points of the standard normal probability density function to the horizontal cumulative averages and their deviations from the mean.

Figure 3.2: ENTACHER Random Number Generator



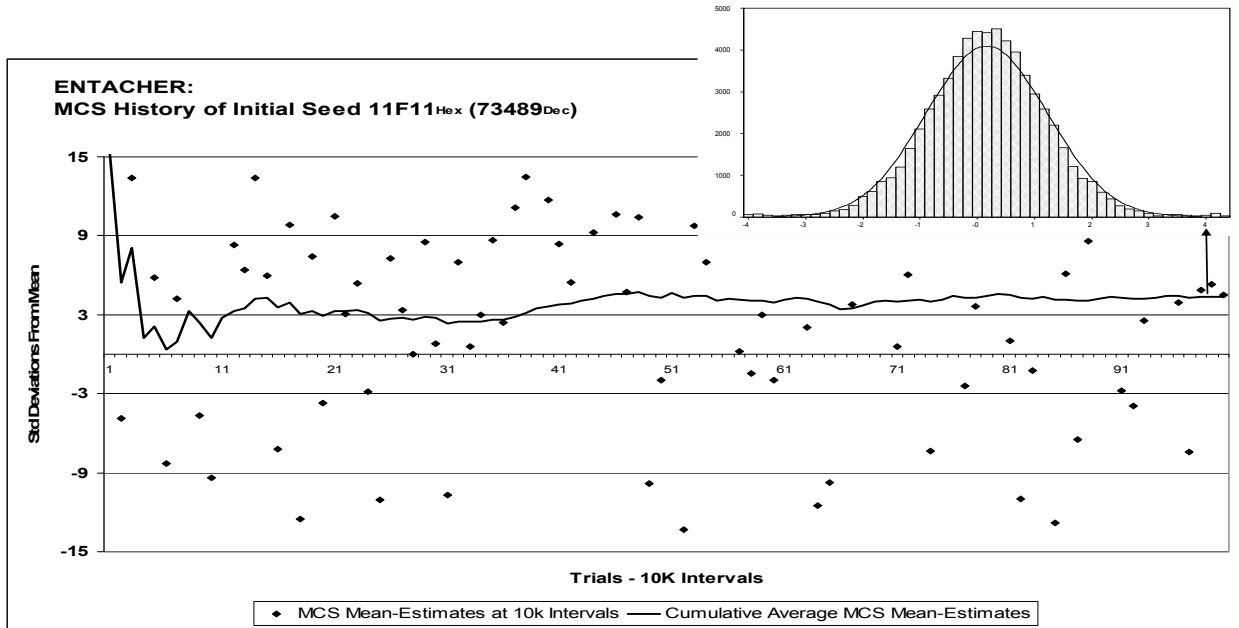
This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 82315, is the worst mean estimate below the mean. Each ♦ is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 82315, on the standard normal probability density function.

Figure 3.3: ENTACHER Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 120803, is the best mean estimate. Each ♦ is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 120803, on the standard normal probability density function.

Figure 3.4: ENTACHER Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 73489, is the worst mean estimate above the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 73489, on the standard normal probability density function.

LECUYER Pseudo-Random Number Generator

For LECUYER's distribution of mean estimates at Trial = 1,000,000, fifteen initial seeds were chosen, five from each extreme (the worst) and five from the center (the best). The mean estimates' initial seeds were used to capture the cumulative (running) average at 10,000-trial intervals of a Monte Carlo simulation of the Standard Normal probability distribution function. These data were normalized and graphed, (Figure 4.1), as graphical history of the Monte Carlo simulation's mean estimate at 10,000-trial intervals. LECUYER's cumulative mean-estimates undulated wildly ( $\pm 15$  standard deviations) for the first 100,000 trials and then converged on their respective region of the mean estimate distribution.

Individual LECUYER MCS history graphs for the two worst initial seeds, 92765 and 102702, and closest to the True Mean, 105488, were isolated. The LECUYER MCS history graph (Figure 4.2) for initial seed,  $S_i = 92765$ ,

demonstrated an immediate departure from zero, -15 standard deviations, at trials = 10,000; a recovery to -3 standard deviations at trials = 170,000; and a final convergence on -4 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

The LECUYER MCS history graph (Figure 4.3) for initial seed,  $S_i = 105488$ , demonstrated an initial departure from zero, -1 standard deviations, at trials = 10,000; a spike to +6 standard deviations, at trials = 30,000; and a final convergence on 0.0 standard deviations. Individual 10,000-trial intervals, designated as  $\blacklozenge$  on the graph, demonstrated some wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

The LECUYER MCS history graph (Figure 4.4) for initial seed,  $S_i = 102702$ , demonstrated an initial departure from zero, +2 standard deviations, at trials = 10,000; a spike to



## PSEUDO-RANDOM NUMBER INITIAL SEED BIAS IN MONTE CARLO SIMULATIONS

+11 standard deviations at trials = 20,000; a sharp drop to +2 standard deviations, at trials = 30,000; and a final convergence on +4 standard deviations. Individual 10,000-trial intervals, designated as ♦ on the graph, demonstrated wildly fluctuating mean estimates of ±15 standard deviations throughout the simulation run.

### RANDU Pseudo-Random Number Generator

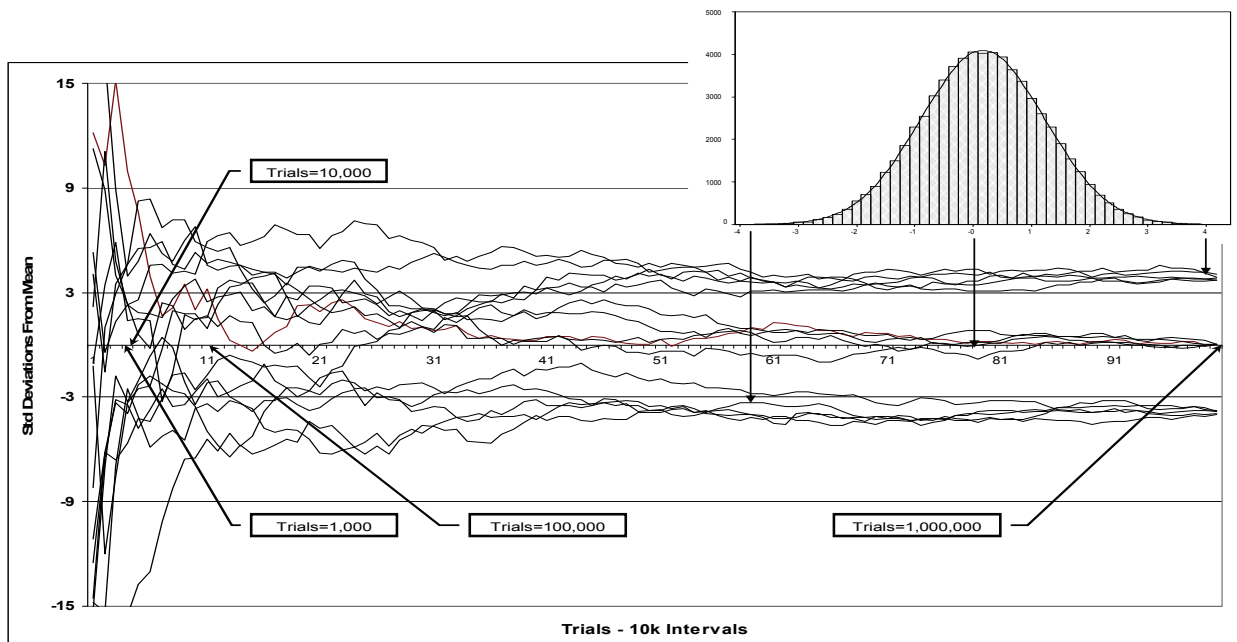
For RANDU's distribution of mean estimates at Trial = 1,000,000, fifteen initial seeds were chosen, five from each extreme, the worst, and five from the center, the best. The mean estimates' initial seeds were used to capture the cumulative (running) average at 10,000-trial intervals of a Monte Carlo simulation of the Standard Normal probability distribution function. These data were normalized and graphed, (Figure 5.1), as graphical history of the Monte Carlo simulation's mean estimate at 10,000-trial

intervals. RANDU's cumulative mean estimates undulated wildly (±15 standard deviations) for the first 100,000 trials and then converged on their respective region of the mean estimate distribution.

Individual RANDU MCS history graphs for the two worst initial seeds, 128395 and 93665, and closest to the True Mean, 81514, were isolated. The RANDU MCS history graph (Figure 5.2) for initial seed,  $S_i = 128395$ , demonstrated an immediate departure from zero, -15 standard deviations and a gradual convergence on -4 standard deviations. Individual 10,000-trial intervals, designated as ♦ on the graph, demonstrated wildly fluctuating mean estimates of ±15 standard deviations throughout the simulation run.

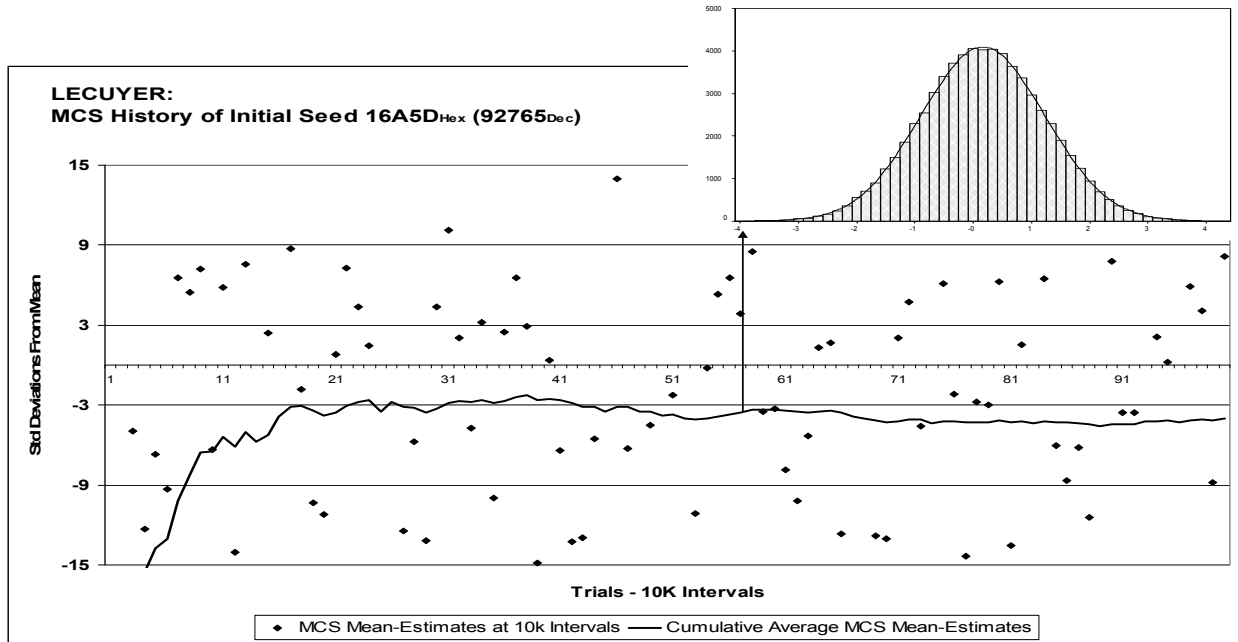
The RANDU MCS history graph (Figure 5.3) for initial seed,  $S_i = 81514$ , demonstrated an initial departure from zero, +12 standard deviations, at trials = 10,000; a recovery, 0.5 standard deviations, at trials =

Figure 4.1 LECUYER Random Number Generator



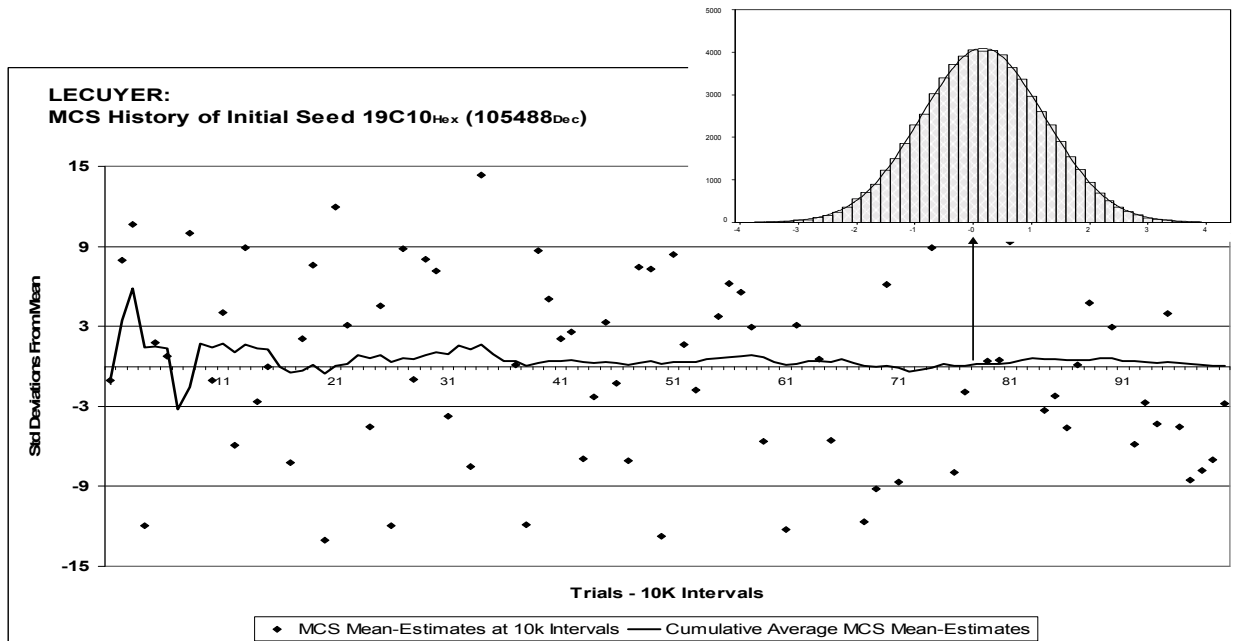
Each line in the graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Fifteen initial seeds (five each: high, middle and low mean estimates) are shown at 10k intervals, total trials = 1,000,000. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Down arrows correlate the  $3\sigma$ , 0, and  $+3\sigma$  points of the standard normal probability density function to the horizontal cumulative averages and their deviations from the mean.

Figure 4.2: LECUYER Random Number Generator



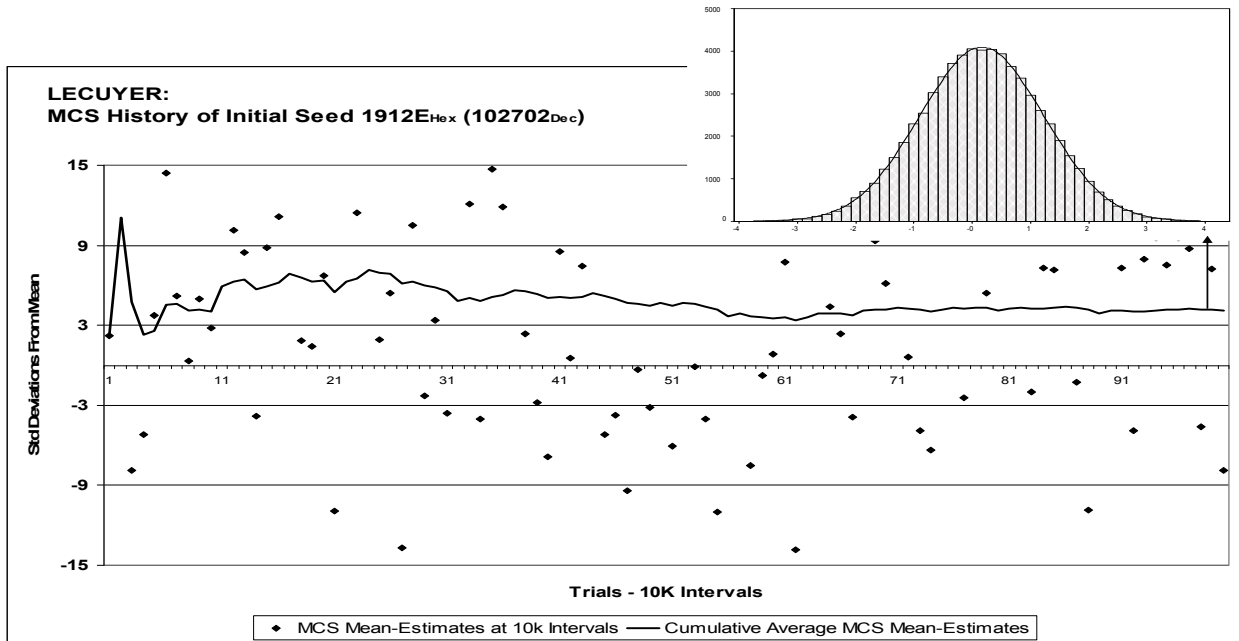
This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 92765, is the worst mean estimate below the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 92765, on the standard normal probability density function.

Figure 4.3: LECUYER Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 105488, is the best mean estimate. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 105488, on the standard normal probability density function.

Figure 4.4: LECUYER Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 102702, is the worst mean estimate above the mean. Each  $\diamond$  is the average mean estimate of its 10k trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 102702, on the standard normal probability density function.

50,000; and a final convergence on 0.0 standard deviations. Individual 10,000-trial intervals, designated as  $\diamond$  on the graph, demonstrated some wildly fluctuating mean estimates of  $\pm 15$  standard deviations but many were  $\pm 15$  standard deviations throughout the simulation run.

The RANDU MCS history graph (Figure 5.4) for initial seed,  $S_i = 93665$ , demonstrated an initial negative departure from zero, -13 standard deviations, at trials = 10,000; a recovery, +9 standard deviations, at trials = 50,000; a sharp drop, +2 standard deviations, at trials = 100,000; and a final convergence on +3 standard deviations. Individual 10,000-trial intervals, designated as  $\diamond$  on the graph, demonstrated wildly fluctuating mean estimates of  $\pm 15$  standard deviations throughout the simulation run.

#### Conclusion

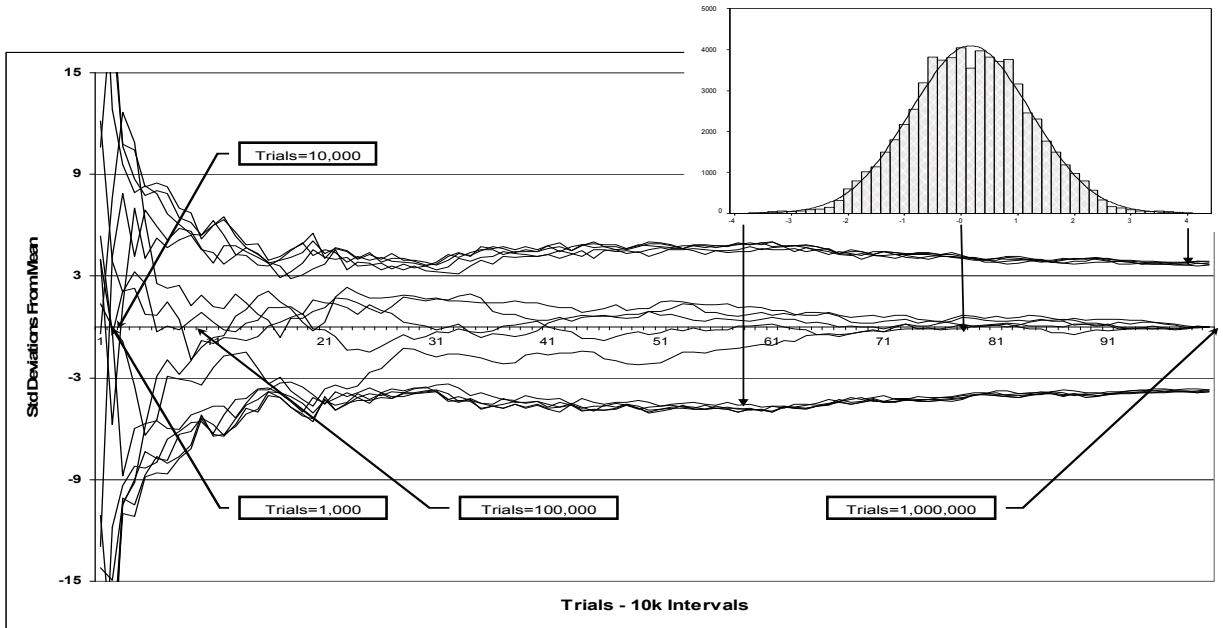
Monte Carlo simulations of the standard normal probability density function using all initial seed between  $10000_{Hex}$  (65,536) through  $1FFFF_{Hex}$

(131,071) did not always yield accurate estimates of the True Mean,  $\mu = 0$ . It was observed that the distribution of all mean estimates by Monte Carlo simulation is normal-like where about 67% of the mean estimates are within  $\pm 1\sigma$  and about 95% of the mean estimates are within  $\pm 2\sigma$ . Conversely, about 5% of the mean estimates exceed  $2\sigma$ , and the maximum mean estimates were  $\pm 5\sigma$ .

Most researchers have recognized that Monte Carlo simulations should be run using several initial seeds and this study supported that practice. Most researchers have also recognized that Monte Carlo simulations should be run for Trials = 1,000,000 or more. This study demonstrated that most mean estimates converged on their final value by trial = 500,000 and did not change very much with additional trials. Each random number generator exhibited similarly distributed mean estimates.

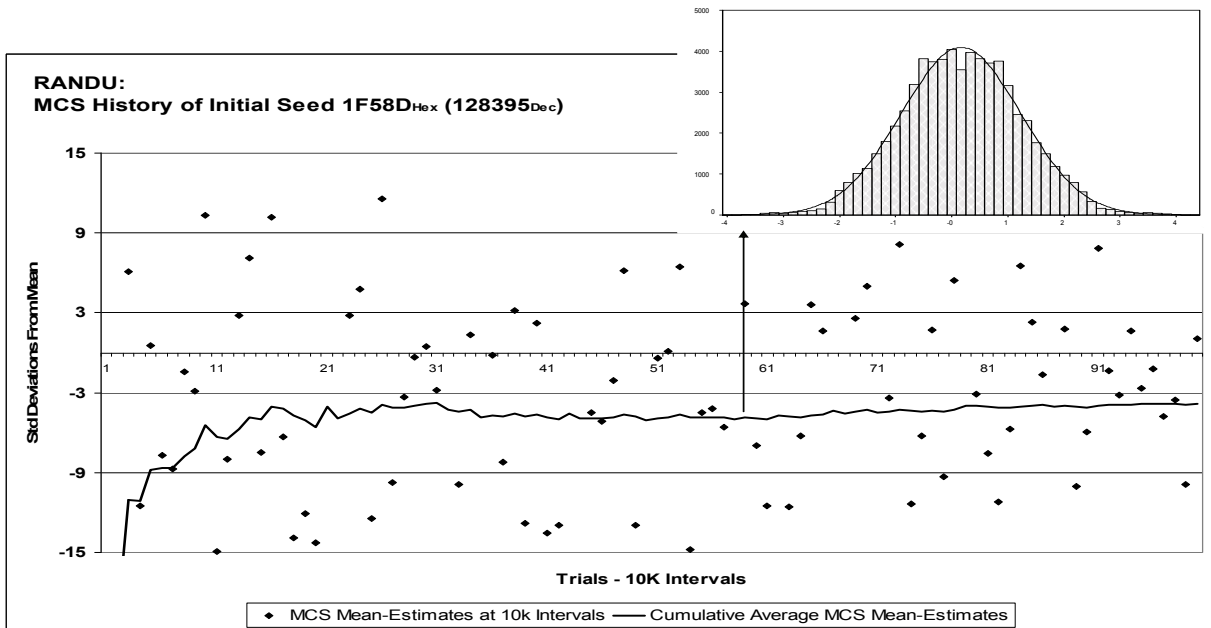
Good and bad initial seed numbers were not identified, but it was shown that errors of  $2\sigma$  or more could occur in Monte Carlo simulations

Figure 5.1: RANDU Random Number Generator



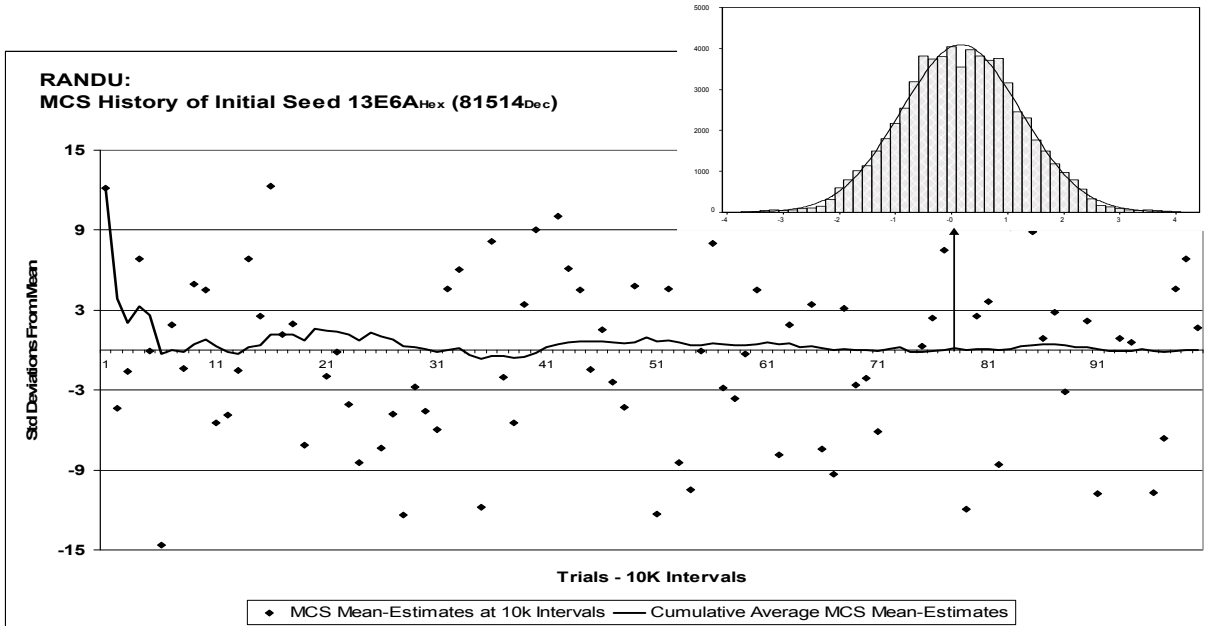
Each line in the graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Fifteen initial seeds (five each: high, middle and low mean estimates) are shown at 10k intervals, total trials = 1,000,000. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Down arrows correlate the  $3\sigma$ , 0, and  $+3\sigma$  points of the standard normal probability density function to the horizontal cumulative averages and their deviations from the mean.

Figure 5.2: RANDU Random Number Generator



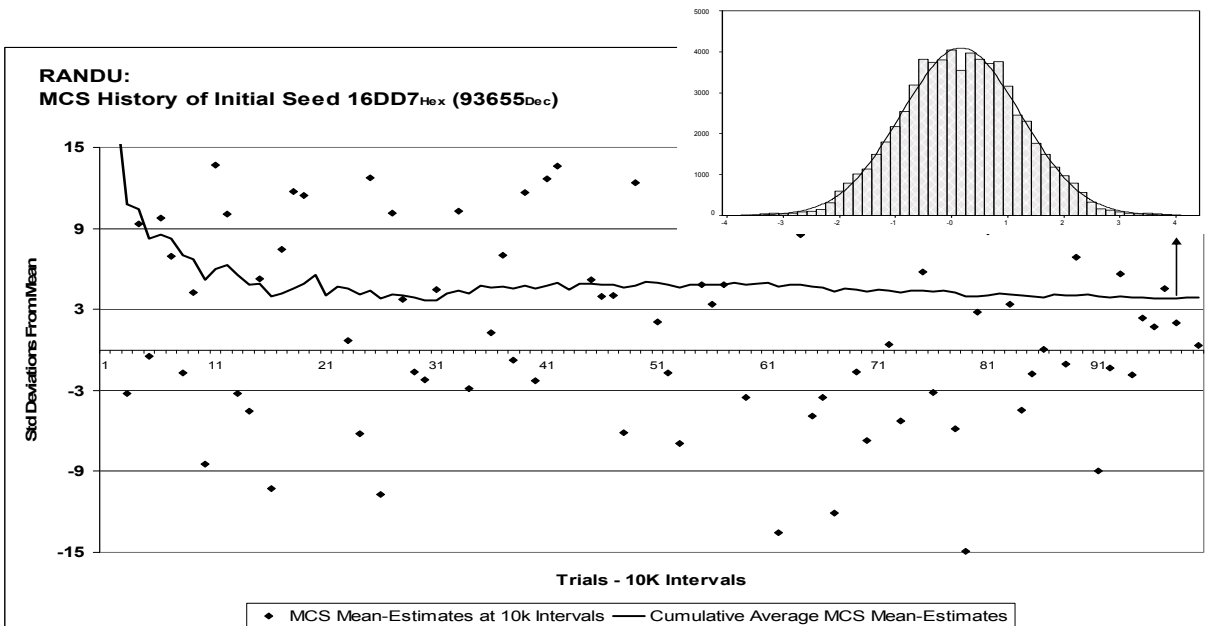
This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 128395, is the worst mean estimate below the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 128395, on the standard normal probability density function.

Figure 5.3 RANDU Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 81514, is the best mean estimate. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 81514, on the standard normal probability density function.

Figure 5.4: RANDU Random Number Generator



This graph follows the cumulative mean estimate of a Monte Carlo simulation of the standard normal probability density function. Initial seed, 93655, is the worst mean estimate above the mean. Each  $\blacklozenge$  is the average mean estimate of its 10k-trial interval. Inset graph is the distribution of all 65,536 initial seeds' mean estimates at 1,000,000 trials with the standard normal probability density function superimposed. Up arrow correlates the location of initial seed, 93655, on the standard normal probability density function.

based on initial seed selection. This study did not intend to identify a list of bad initial seeds that introduced significant error in Monte Carlo simulations of the standard normal probability density function, and then, assume they would also introduce significant error in other distributions or data sets; additional study is required to make this assumption.

Because the mean estimates randomly fluctuated throughout the simulation, any strategy that runs the simulation for  $T$  trials, discards these samples, and starts sampling at  $Trials = T+1$ , will repeat the  $\pm 15$  standard deviation fluctuations and gain nothing. In this study, every Monte Carlo simulation of the standard normal probability density function exhibited this phenomenon: randomly fluctuating mean estimates until the average of all the simulations stabilized. It is well known that averaging multiple samples will decrease the effects of random variations, thus, as samples accumulate, the average mean estimate stabilizes. However, discarding samples or resetting the average = 0 will also discard the accumulated diminishing effect of random variations and the mean estimates will wildly fluctuate, again.

The overall mean estimate of Monte Carlo simulations of the standard normal probability density function using all initial seed between  $10000_{Hex}$  (65,536) through  $1FFFF_{Hex}$  (131,071) was 0.0 at  $Trial = 1,000$ , 0.0 at  $Trial = 10,000$ , 0.0 at  $Trial = 100,000$ , and 0.0 at  $Trial = 1,000,000$  for RANGEN, ENTACHER, LECUYER and RANDU pseudo-random number generators. The standard deviation estimates were 1.00 at all trials for each of these pseudo-random number generators.

#### Questions For Additional Research

Did these findings suggest relatively short simulation runs, trials = 1,000 or less, using tens of thousands initial seeds would yield more accurate mean estimates than typical Monte Carlo long simulation runs, trials = 500,000 or more, using a few initial seeds?

The total number of samples drawn from the population need not be increased. For example, typical Monte Carlo simulations requiring 100,000,000 samples (10 seeds, sample size 10, and 1,000,000 trials) cannot

guarantee an accurate mean-estimate. However, Monte Carlo simulations of 10,000 seeds, sample size 10, and 1,000 trials also require 100,000,000 samples and may yield estimates the mean with greater accuracy.

This study has shown that good pseudo-random number generators like RANGEN, ENTACHER, LECUYER and MOTHER yielded accurate mean estimates by Monte Carlo simulation of the standard normal probability density function when tens of thousands initial seeds were used with short simulation runs of trials = 1,000. It was also shown that a bad pseudo-random number generator, such as RANDU, could yield accurate mean estimates, which is contrary to its reported poor performance in Monte Carlo simulations.

This study raised many unanswered questions: Are these findings limited to Monte Carlo simulation of the standard normal probability density function, or will other data distributions exhibit similar findings? Will other pseudo-random number generators exhibit similar findings? Will Monte Carlo simulation runs of  $10 \leq Trials \leq 1,000$  using tens of thousands initial seeds continue to yield accurate mean estimates of the standard normal probability density function?

#### References

- Blair, R. C. (1999). *BFRA ver. 2.1 users' manual*. College of Public Health, University of South Florida.
- Entacher, K. (1998). Bad subsequences of well-known linear congruential pseudorandom number generators. *ACM Transactions on Modeling and Computer Simulations*, 8(1), 61-70.
- Fahome, G., & Sawilowsky, S. S. (2001). *Rangen 2.0 for Fortran 90*. College of Education, Wayne State University.
- Fishman, G. S. (1995). *Monte Carlo: concepts, algorithms, and applications*. New York: Springer-Verlag.
- James, F. (1994). RANLUX: A Fortran implementation of the high-quality pseudo-random number generator of Lüscher. *Computer Physics Communications*, 79, 111-114.

L'Ecuyer, P. (1993). Two approaches for estimating the gradient in functional form. *Proceedings of the 1993 Winter Simulation Conference*.338-346.

L'Ecuyer, P., &Hellegalek, P. (1997). Random number generators: selection criteria and testing. *Random and Quasi-Random Point Sets, Lectures Notes In Statistics (138)*, Springer, 223-266.

Lehmer, D. H. (1951).Mathematical methods in large-scale computing units.2<sup>nd</sup> *Symposium on Large-Scale Digital Calculating Machines, Harvard University Press, 26(141)*, 141-146.

Lüscher, M. (1994). A portable high-quality random number generator for lattice field theory calculations. *Computer Physics Communications, 79*, 100-110.

Marsaglia, G. (1994a). *The mother of all random generators*.Posted by Bob Wheeler to sci.stat.consult and sci.math.num-analysis on behalf of George Marsaglia on October 28, 1994.

Marsaglia, G. (1994b). Multiply-with-carry generators. Diehard CD-ROM, available online at ftp.csis.hku.hk.

Miller, A. (1995). MODULE luxury. <http://jblevins.org/mirror/amiller/luxury.f90>

Park, S. K., & Miller, K. W. (1988). Random number generators: Goods ones are hard to find. *Communications of the ACM, 31(10)*, 1192-1201.

Sawilowsky, S. S., & Blair, R. C. (1987). *RANGENPC 1.0 for Fortran 77*. College of Education, Wayne State University.

## Estimation of Population Mean In Successive Sampling by Sub-Sampling Non-Respondents

Housila P. Singh  
Vikram University,  
Ujjain, India

Sunil Kumar  
University of Jammu,  
(J & K), India

Sandeep Bhogal  
Shri Mata Vaishno Devi University,  
Kakryal, Jammu (J & K), India

---

The estimation of the population mean in mail surveys is investigated in the context of sampling on two occasions where the population mean of the auxiliary variable is available in the presence of non-response only for the current occasion in two occasion successive sampling. The behavior of the proposed estimator is compared with the estimator for the same situation but in the absence of non-response. An empirical illustration demonstrates the performance of the proposed estimator.

Key words: Variance, study variable, auxiliary variable, non-response, successive sampling.

---

### Introduction

A very important problem for many countries is the management and conservation of food resources. However, it commonly occurs that the classical theory of sampling cannot be directly applied in situations calling for quantification of environmental resources. If a population is subject to change, a survey carried out on a single occasion cannot of itself give any information of the nature or rate of such change (Miranda, 2007, p. 385).

The problem of sampling on two successive occasions was first considered by Jessen (1942) and has also been discussed by Patterson (1950), Narain (1953), Eckler (1955), Adhvaryu (1978), Sen (1979), Gorden (1983) and Arnab and Okafor (1992). In addition to the information from previous research, Singh, et al. (1991), Artes and Garcia (2001), Singh and Singh (2001), Garcia and Artes (2002), Singh (2003) and Singh and Vishwakarma (2007),

used auxiliary information on current occasion for estimating the current population mean in two-occasion successive sampling.

It is common experience in sample surveys that a proportion of people among those invited to participate in a non-compulsory interview survey, or other study, choose not to take part or are unobtainable for other reasons. Non-response covers all causes of non-participation including, direct refusals, people who are away temporarily on holiday and non-contacts for other reasons. Those who are found to be outside the scope of the survey are classified as ineligible and excluded altogether. Ineligibles include people who had died or moved to an area outside the survey area, businesses that had closed down and changed addresses.

Hansen and Hurwitz (1946) were the first to suggest a technique of handling non-response in mail surveys. Cochran (1977), Okafor and Lee (2000) extended the Hansen and Hurwitz technique to the case when along with the information on character under study, information is also available on an auxiliary character. More recently Choudhary, et al. (2004), Okafor (2005) and Singh and Priyanka (2007) used the Hansen and Hurwitz technique for estimating the population mean on current occasion in the context of sampling on two occasions. This article investigates successive sampling theory in the presence of non-response and examines the efficiency over the estimate

---

Housila P. Singh is a Professor in the School of Studies in Statistics. Email him at: hpsujn@rediffmail.com. Sunil Kumar is in the Department of Statistics, University of Jammu. Email him at: sunilbhogal06@gmail.com. Sandeep Bhogal is in the School of Mathematics, SMVDU. Email him at: sandeep.bhogal@smvdu.ac.in.



defined for the same situation with complete response.

Building an Estimator

Suppose that the two samples are of size  $n$  on both occasions and simple random sampling and the size of the population  $N$  is used which is sufficiently large for the correlation factor to be ignored.

Let  $U = (U_1, U_2, \dots, U_N)$  represent the total population of  $N$  identifiable units that have been sampled over two occasions. Let  $x(y)$  be the character under study on the first (second) occasions respectively. It is deduced that information on an auxiliary variable  $x$  is available on both the occasions with known population mean. A simple random sample without replacement of  $n$  units is taken on the first occasion.

On the second occasion, a simple random sample without replacement of  $m = n\lambda$  units is retained while an independent sample of  $u = n\mu = n - m$  units is selected so that the sample size on both the occasions is the same,  $n$  units. It is assumed that there is non-response at the second (current) occasion, so that the population can be divided into two classes, those who will respond at the first attempt and those who will not: let the sizes of these two classes be  $N_1$  and  $N_2$  respectively. Assume that in the matched (unmatched) portion of the sample on two occasions  $m_1(u_1)$  units respond and  $m_2(u_2)$  units do not. Let  $m_{h_2}(u_{h_2})$  units denote the size of the sub-sample drawn from the non-response class from the matched (unmatched) portion of the sample on the two occasions for collecting information through personal interview.

This study considers the same situation as outlined in Singh and Kumar (2010), where the information on the auxiliary variable is completely available for all the second phase sample of size  $n$  units while, out of  $n$  sample units on the current occasion, some units refused to respond on the study variable  $y$ . Hansen and Hurwitz (1946) technique to sub sampling from  $m_2(u_2)$  non-respondents of size  $m_{h_2}(u_{h_2})$  units

selected at random and is enumerated by direct interview, such that, by  $(m_{h_2} = m_2/k)$   $(u_{h_2} = u_2/k)$ ,  $k > 1$ , one will obtain the estimate

$$\left( \bar{y}_{2m_{h_2}} = \sum_{j=1}^{m_{h_2}} y_j / m_{h_2} \right) \left( \bar{y}_{2u_{h_2}} = \sum_{j=1}^{u_{h_2}} y_j / u_{h_2} \right).$$

Using  $\bar{y}_{2m_2}(\bar{y}_{2u_2})$ , an unbiased estimator  $\bar{y}^*$  of the population mean  $\bar{Y}$  of the study variable  $y$  on the current occasion will be constructed. For these  $m_{h_2}(u_{h_2})$  units selected from  $m_2(u_2)$  non-respondent units one can also obtain the estimate

$$\left( \bar{x}_{2m_{h_2}} = \sum_{j=1}^{m_{h_2}} x_j / m_{h_2} \right) \left( \bar{x}_{2u_{h_2}} = \sum_{j=1}^{u_{h_2}} x_j / u_{h_2} \right)$$

and using this estimate results in the unbiased estimate  $\bar{x}^*$  on the current occasion.

Further, an estimator is constructed when there is non-response only on the second occasion as:

$$t_m = \bar{y}_m^* + \lambda_1 (\bar{x}_n - \bar{x}_m^*) + \lambda_2 (\bar{x}_m - \bar{x}_n), \tag{2.1}$$

where  $\lambda_1$  and  $\lambda_2$  are suitably chosen constants,

$$\bar{y}_m^* = \frac{m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_{h_2}}}{m}$$

is the Hansen and Hurwitz (1946) estimator for the population mean  $\bar{Y}$  for matched portion of the sample on second occasion;

$$\bar{x}_m^* = \frac{m_1 \bar{x}_{m_1} + m_2 \bar{x}_{m_{h_2}}}{m}$$

is the Hansen and Hurwitz (1946) estimator for the population mean  $\bar{X}$  for matched portion of the sample on second occasion;

$$\bar{x}_n = \sum_{i=1}^n x_i / n$$

is the estimate of the population mean  $\bar{X}$  of the sample;

$$\bar{x}_m = \sum_{i=1}^m x_i / m$$

is the estimate of the population mean  $\bar{X}$  on second occasion for the matched portion of the sample;

$$\bar{x}_{m_1} = \sum_{i=1}^{m_1} x_i / m_1$$

is the estimate of the population mean  $\bar{X}_1$  on second occasion for the matched portion of the sample;

$$\bar{x}_{m_{h_2}} = \sum_{i=1}^{m_{h_2}} x_i / m_{h_2}$$

is the sub-sample mean of variable  $x$  based on  $m_{h_2}$  units on the second occasion;

$$\bar{y}_{m_1} = \sum_{i=1}^{m_1} y_i / m_1$$

is the estimate of the population mean  $\bar{Y}_1$  on second occasion for the matched portion of the sample; and

$$\bar{y}_{m_{h_2}} = \sum_{i=1}^{m_{h_2}} y_i / m_{h_2}$$

is the sub-sample of variable  $y$  based on  $m_{h_2}$  units on the second occasion.

The variance of  $t_m$  (if fpc is ignored) to the first degree of approximation is given by

$$\begin{aligned} Var(t_m) = & \left[ \left( \frac{1}{m} - \frac{1}{n} \right) \{ S_y^2 + (\lambda_1 - \lambda_2)^2 S_x^2 - 2(\lambda_1 - \lambda_2) \beta S_x^2 \} \right. \\ & \left. + \frac{W_2(k-1)}{m} \{ S_{y(2)}^2 + \lambda_1(\lambda_1 - 2\beta_{(2)}) S_{x(2)}^2 \} + \frac{1}{n} S_y^2 \right] \end{aligned} \quad (2.2)$$

where

$$W_2 = N_2 / N;$$

$$\beta = \rho(S_y / S_x);$$

$$\beta_{(2)} = \rho_{(2)}(S_{y(2)} / S_{x(2)});$$

$$k = (u_2 / u_{h_2}) = (m_2 / m_{h_2});$$

and  $\rho$  and  $\rho_{(2)}$  are the correlation coefficient between the variables ( $y$  and  $x$ ) and ( $y_{(2)}$  and  $x_{(2)}$ );

$$S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 / (N - 1)$$

denotes the population mean square of the variable  $y$ ;

$$S_x^2 = \sum_{i=1}^N (x_i - \bar{X})^2 / (N - 1)$$

denotes the population mean square of the variable  $x$ ;

$$S_{y(2)}^2 = \sum_{i=1}^{N_2} (y_i - \bar{Y}_{(2)})^2 / (N_2 - 1)$$

denotes the population mean square pertaining to the non-response class of the variable  $y$ ;

$$S_{x(2)}^2 = \sum_{i=1}^{N_2} (x_i - \bar{X}_{(2)})^2 / (N_2 - 1)$$

denotes the population mean square pertaining to the non-response class of the variable  $x$ .

Differentiating the variance of  $t_m$ , that is,  $Var(t_m)$  at (2.2) with respect to  $\lambda_1$  and  $\lambda_2$ , and equating to zero, results in the optimum values of  $\lambda_1$  and  $\lambda_2$  as

$$\lambda_1 = \rho_{(2)}(S_{y(2)}/S_{x(2)}) = \beta_{(2)}$$

and

$$\begin{aligned} \lambda_2 &= \left\{ \rho_{(2)}(S_{y(2)}/S_{x(2)}) \right\} - \left\{ \rho(S_y/S_x) \right\} \\ &= (\beta_{(2)} - \beta) \end{aligned}$$

Substituting the optimum values of  $\lambda_1$  and  $\lambda_2$  in (2.1), results in the optimum estimate of the estimator  $t_m$  as

$$t_m^{(0)} = \bar{y}_m^* - \beta_{(2)}(\bar{x}_m^* - \bar{x}_m) - \beta(\bar{x}_m - \bar{x}_n), \quad (2.3)$$

with variance (ignoring fpc), the result is

$$Var(t_m^{(0)}) = \left[ \left( \frac{1}{m} - \frac{1}{n} \right) (1 - \rho^2) S_y^2 + \frac{W_2(k-1)}{m} (1 - \rho_{(2)}^2) S_{y(2)}^2 + \frac{1}{n} S_y^2 \right] \quad (2.4)$$

In practice  $\beta_{(2)}$  and  $\beta$  are usually unknown, it lacks the practical utility of the optimum estimator  $t_m^{(0)}$ , thus it is advisable to replace  $\beta_{(2)}$  and  $\beta$  by their consistent estimates  $\hat{\beta}_{(2)}^*$  and  $\hat{\beta}^*$  respectively in (2.3) to calculate an estimate of the population mean  $\bar{Y}$  based on matched portion on second occasion as

$$\hat{t}_m^{(0)} = \bar{y}_m^* - \hat{\beta}_{(2)}^*(\bar{x}_m^* - \bar{x}_m) - \hat{\beta}^*(\bar{x}_m - \bar{x}_n), \quad (2.5)$$

where

$$\hat{\beta}_{(2)}^* = s_{xy(2)}^* / s_{x(2)}^{*2},$$

$$\hat{\beta}^* = s_{xy}^* / s_x^{*2},$$

$$s_{xy}^* = \frac{1}{(m-1)} \left\{ \sum_{j=1}^{m_1} x_j y_j + k \sum_{j=1}^{m_{h_2}} x_j y_j - m \bar{x}_m \bar{y}_m^* \right\},$$

$$s_x^{*2} = \frac{1}{(m-1)} \left\{ \sum_{j=1}^{m_1} x_j^2 + k \sum_{j=1}^{m_{h_2}} x_j^2 - m \bar{x}_m \bar{x}_m^* \right\},$$

$$s_{xy(2)}^* = \frac{1}{(m_{h_2}-1)} \left[ \sum_{j=1}^{m_{h_2}} (x_j - \bar{x}_{m_{h_2}}) (y_j - \bar{y}_{m_{h_2}}) \right],$$

$$s_{x(2)}^{*2} = \frac{1}{(m_{h_2}-1)} \sum_{j=1}^{m_{h_2}} (x_j - \bar{x}_{m_{h_2}})^2,$$

$$\bar{x}_{m_{h_2}} = \frac{1}{m_{h_2}} \sum_{j=1}^{m_{h_2}} x_j,$$

$$\bar{y}_{m_{h_2}} = \frac{1}{m_{h_2}} \sum_{j=1}^{m_{h_2}} y_j,$$

and

$$\bar{x}_m = \frac{1}{m} \sum_{j=1}^m x_j.$$

It can be shown to the first degree of approximation that

$$\begin{aligned} Var(\hat{t}_m^{(0)}) &= \\ &= Var(t_m^{(0)}) \\ &= \left[ \left( \frac{1}{m} - \frac{1}{n} \right) (1 - \rho^2) S_y^2 + \right. \\ &\quad \left. \left[ \frac{W_2(k-1)}{m} (1 - \rho_{(2)}^2) S_{y(2)}^2 + \frac{1}{n} S_y^2 \right] \right] \end{aligned} \quad (2.6)$$

where  $Var(t_m^{(0)})$  is given by (2.4) (Singh & Kumar, 2008).

Hence, an estimate of the population mean  $\bar{Y}$  of the study variable  $y$  is constructed in the presence of non-response on the current occasion by combining the two independent

estimates  $\bar{y}_u^*$  and  $\hat{t}_m^{(0)}$  with  $\alpha$  an unknown constant as

$$T_{21} = \alpha \bar{y}_u^* + (1 - \alpha) \hat{t}_m^{(0)}, \quad (2.7)$$

where

$$\bar{y}_u^* = \frac{u_1 \bar{y}_{u_1} + u_2 \bar{y}_{u_{h_2}}}{u},$$

$$\bar{y}_{u_1} = \sum_{i=1}^{u_1} y_i / u_1,$$

and

$$\bar{y}_{u_{h_2}} = \sum_{i=1}^{u_{h_2}} y_i / u_{h_2}.$$

The variance of  $\bar{y}_u^*$ , the Hansen and Hurwitz (1946) estimator is

$$Var(\bar{y}_u^*) = \left( \frac{1}{u} - \frac{1}{N} \right) S_y^2 + \frac{W_2(k-1)}{u} S_{y(2)}^2. \quad (2.8)$$

The variance of  $T_{21}$  at (2.7) to the first degree of approximation is given by

$$Var(T_{21}) = \alpha^2 Var(\bar{y}_u^*) + (1 - \alpha)^2 Var(\hat{t}_m^{(0)}). \quad (2.9)$$

Because, the variance of  $T_{21}$  in equation (2.9) is a function of unknown constant  $\alpha$ , it is minimized with respect to  $\alpha$  and subsequently the optimum value of  $\alpha$  is obtained as

$$\alpha_{opt} = \frac{Var(\hat{t}_m^{(0)})}{Var(\bar{y}_u^*) + Var(\hat{t}_m^{(0)})}. \quad (2.10)$$

Using the optimum value of  $\alpha$  from (2.10) in (2.9), results in the optimum variance of  $T_{21}$  as

$$Var(T_{21})_{opt} = \frac{Var(\bar{y}_u^*) Var(\hat{t}_m^{(0)})}{Var(\bar{y}_u^*) + Var(\hat{t}_m^{(0)})}. \quad (2.11)$$

Further, substituting the values from (2.4) and (2.8) in (2.11), the optimum variance of  $T_{21}$  is simplified as

$$Var(T_{21})_{opt} = \left( \frac{S_y^2 + W_2(k-1) S_{y(2)}^2}{n} \right) \left[ \frac{\{(1 - q\rho^2) S_y^2 + A\}}{\left\{ (1 - q^2 \rho^2) S_y^2 + \left[ W_2(k-1)(1 - q\rho_{(2)}^2) S_{y(2)}^2 \right] \right\}} \right], \quad (2.12)$$

where

$$A = W_2(k-1)(1 - \rho_{(2)}^2) S_{y(2)}^2.$$

To determine the optimum value of  $q$  so that population mean  $\bar{Y}$  of study variable  $y$  may be estimated with maximum precision, minimize  $Var(T_{21})_{opt}$  in (2.12) with respect to  $q$  and the optimum value of  $q$  is obtained as

$$\begin{aligned} q &= \frac{(A + S_y^2) \pm \sqrt{(1 - \rho^2) S_y^4 + W_2(k-1) \left\{ A + (2 - \rho^2 - \rho_{(2)}^2) S_y^2 \right\} S_{y(2)}^2}}{\rho^2 S_y^2} \\ &= q_0 \end{aligned} \quad (2.13)$$

The real value of  $q_0$  exists if

$$\left[ (1 - \rho^2) S_y^4 + W_2(k-1) \left\{ A + (2 - \rho^2 - \rho_{(2)}^2) S_y^2 \right\} S_{y(2)}^2 \right] \geq 0.$$

For certain situations, there might be two values of  $q_0$  satisfying the above condition, hence when selecting a value of  $q_0$ , it should be remembered that the existence of  $q_0$  depends on the limit  $0 \leq q_0 \leq 1$ ; all other values of  $q_0$  are inadmissible. In the case where both the values

of  $q_0$  are admissible, choose the minimum as  $q_0$ .

Further, substituting the value of  $q_0$  from (2.13) in (2.12),

$$Var(T_{21})_{opt^*} = \left( \frac{S_y^2 + W_2(k-1)S_{y(2)}^2}{n} \right) \left[ \frac{\left\{ (1 - q_0 \rho^2) S_y^2 + A \right\}}{\left\{ (1 - q_0^2 \rho^2) S_y^2 + W_2(k-1)(1 - q_0 \rho_{(2)}^2) S_{y(2)}^2 \right\}} \right], \quad (2.14)$$

where  $Var(T_{21})_{opt^*}$  is the optimum variance of  $T_{21}$  with respect to both  $\alpha$  and  $q$ .

#### Efficiency Comparison

To determine the effect of non-response in successive sampling, calculate the percent relative loss in efficiency of  $T_{21}$  with respect to the estimator under the same circumstances but in absence of non-response. The estimator is defined as

$$T_{21}^* = \phi \bar{y}_u + (1 - \phi) t_{ld};$$

where

$$\bar{y}_u = \sum_{i=1}^u y_i / u; \quad t_{ld} = \bar{y}_m + \hat{\beta}(\bar{x}_n - \bar{x}_m),$$

$$\hat{\beta} = \left\{ \sum_{i=1}^m (x_i - \bar{x}_m)(y_i - \bar{y}_m) \right\} / \left\{ \sum_{i=1}^m (x_i - \bar{x}_m)^2 \right\}$$

and  $\phi$  is an unknown constant to be determined under certain criterion. Because  $T_{21}^*$  is an unbiased estimator of  $\bar{Y}$  and is based on two independent samples the covariance terms vanishes, therefore following the procedure of Sukhatme, et al. (1984), the optimum variance of  $T_{21}^*$  can be obtained as

$$Var(T_{21}^*)_{opt^*} = \frac{(1 - q_1 \rho^2) S_y^2}{n(1 - q_1^2 \rho^2)}, \quad (3.1)$$

where

$$q_1 = \frac{1 \pm \sqrt{(1 - \rho^2)}}{\rho^2}.$$

To select the optimum value of  $q_1$ , it is important to remember that  $0 \leq q_1 \leq 1$ , however, if both values of  $q_1$  are admissible, then the least of two values of  $q_1$  should be chosen. Thus, the percentage loss in precision of  $T_{21}^*$  with respect to  $T_{21}$  both at optimality condition is given by

$$L = \frac{Var(T_{21})_{opt^*} - Var(T_{21}^*)_{opt^*}}{Var(T_{21})_{opt^*}} \times 100. \quad (3.2)$$

#### Results

Table 1 shows the percentage loss in precision observed wherever the optimum value of  $q$  exists when non-response is taken into account at current occasion. For fixed values of  $\rho$ ,  $\rho_{(2)}$ ,  $(k-1)$  and  $W_2$ , for  $S_y < S_{y(2)}$ , the loss in precision decreases with the increase in the value of  $S_y$ ; for  $S_y > S_{y(2)}$ , the loss in precision shows negative values with the decrease in the value of  $S_{y(2)}$ ; and for  $S_y = S_{y(2)}$ , the loss in precision remains constant. For fixed values of  $S_y$ ,  $S_{y(2)}$ ,  $(k-1)$  and  $W_2$ , the loss in precision shows negative values for  $\rho < \rho_{(2)}$  with the decrease in the value of  $\rho$  and for  $\rho > \rho_{(2)}$ , the loss in precision decreases with increase in the value of  $\rho_{(2)}$  while it remains constant for  $\rho = \rho_{(2)}$ . Table 2 shows that, for the increased values of  $W_2$ , the percentage loss in precision increases and it decreases with the decreases in the value of  $(k-1)$ .

Table 1: Percentage Loss in Precision of  $T_{21}^*$  over  $T_{21}$   
for Different Values of  $\rho$ ,  $\rho_{(2)}$ ,  $S_y$  and  $S_{y(2)}$

$$\rho = 0.8, \rho_{(2)} = 0.2, (k-1) = 1, W_2 = 0.8$$

$S_y < S_{y(2)}$		$L$	$S_y > S_{y(2)}$		$L$	$S_y = S_{y(2)}$		$L$
$S_y$	$S_{y(2)}$		$S_y$	$S_{y(2)}$		$S_y$	$S_{y(2)}$	
0.2	0.8	90.59	0.8	0.7	*	0.8	0.8	22.07
0.3	0.8	80.43	0.8	0.6	*	0.8	0.8	22.07
0.4	0.8	68.44	0.8	0.5	*	0.8	0.8	22.07

$$(k-1) = 0.5, W_2 = 0.7, S_y = 0.4, S_{y(2)} = 0.8$$

$\rho < \rho_{(2)}$		$L$	$\rho > \rho_{(2)}$		$L$	$\rho = \rho_{(2)}$		$L$
$\rho$	$\rho_{(2)}$		$\rho$	$\rho_{(2)}$		$\rho$	$\rho_{(2)}$	
0.7	0.8	*	0.7	0.3	39.01	0.8	0.8	16.67
0.6	0.8	*	0.7	0.4	37.04	0.8	0.8	16.67
0.5	0.8	*	0.7	0.5	33.94	0.8	0.8	16.67

Table 2: Percentage Loss in Precision of  $T_{21}^*$  over  $T_{21}$   
for Different Values of  $W_2$  and  $(k-1)$

$\rho = 0.8, \rho_{(2)} = 0.2,$ $(k-1) = 0.5, S_y = 0.2, S_{y(2)} = 0.7$		$\rho = 0.8, \rho_{(2)} = 0.4, W_2 = 0.3,$ $S_y = 0.3, S_{y(2)} = 0.8$	
$W_2$	$L$	$(k-1)$	$L$
0.3	19.32	1.5	65.95
0.4	37.83	1.0	54.38
0.5	49.40	0.5	30.03

A tangible idea regarding obtaining cost saving through mail surveys in the context of successive sampling on two occasions for different assumed values of  $\rho$ ,  $\rho_{(2)}$ ,  $S_y$ ,  $S_{y(2)}$ ,  $W_2$  and  $k$  is shown in Tables 3 and 4. Also, let

$N = 500$  and  $n = 50$  and  $c_1/c_0 = 4$ ,  $c_2 = 45$ , where  $c_0$  is the cost per unit for mailing a questionnaire (Rs. 1.00),  $c_1$  is the cost per unit of processing the results from the first attempt respondents (Rs. 4.00),  $c_2$  is the cost per unit

for collecting data through personal interview (Rs. 45.00). Denote  $C$  = total cost incurred in collecting the data by personal interview from the whole sample, that is, when there is no non-response. Assuming that the cost incurred on data collection for the matched and unmatched portion of the sample are same and also cost incurred on data collection on both the occasions is same, the cost function in this case is given by:

$$C = 2nc_2. \quad (3.3)$$

Setting the values of  $n$  and  $c_2$  in (3.3), the total cost work out to be Rs. 4500.00.

Further, let  $n_1$  denote number of units which respond at the first attempt and  $n_2$  denote number of units which do not respond. The cost function for the case when there is non-response on both occasions is given by

$$C_1 = 2\{c_0n + c_1n_1 + (c_2n_2/k)\}.$$

The expected cost is given by

$$E(C_1) = 2n_0\{c_0 + c_1W_1 + (c_2W_2/k)\} = C_1^*,$$

where

$$W_1 = N_1/N$$

and

$$W_2 = N_2/N,$$

such that

$$W_1 + W_2 = 1,$$

$$n_0 = \frac{n(1 - q_1^2\rho^2)(1 - B)(S_y^2 + W_2(k - 1)S_{y(2)}^2)}{(1 - q_1\rho^2)(1 - q_0B)S_y^2},$$

and

$$B = \frac{\{q_0\rho^2S_y^2 + W_2(k - 1)\rho_{(2)}^2S_{y(2)}^2\}}{\{S_y^2 + W_2(k - 1)S_{y(2)}^2\}}.$$

From Table 3 it is noted that for fixed values of  $\rho$ ,  $\rho_{(2)}$ ,  $(k - 1)$  and  $W_2$ , for the case

$S_y < S_{y(2)}$ , the cost savings increases with decreases in the value of  $S_y$ . For the case  $S_y > S_{y(2)}$ , the cost savings decreases with the decreases in the value of  $S_{y(2)}$ , and for the case  $S_y = S_{y(2)}$ , it remains constant. Further, for the fixed values of  $(k - 1)$ ,  $W_2$ ,  $S_y$  and  $S_{y(2)}$ , for the case  $\rho < \rho_{(2)}$ , the cost savings decreases with the decreases in the value of  $\rho$  and for the case  $\rho > \rho_{(2)}$ , it also decreases with the increase in the value of  $\rho_{(2)}$  but it remains constant for the case  $\rho = \rho_{(2)}$ . It is to be observed from Table 4 that increases in the values of  $W_2$  and decreases in the value of  $(k - 1)$ , the cost savings increase respectively.

#### References

- Adhvaryu, D. (1978). Successive sampling using multi auxiliary information. *Sankhya, C, 2*, 167-173.
- Arnab, R. and Okafor, F. C. (1992). A note on double sampling over two occasions. *Pakistan Journal of Statistics, B, 8*(3), 9-18.
- Artes, E. and Garcia, A. (2001). Estimation of current population ratio in successive sampling. *Journal of the Indian Society of Agricultural Statistics, 54*(3), 342-354.
- Choudhary, R. K., Bathla, H. V. L., & Sud, U. C. (2004). On non-response in sampling on two occasions. *Journal of the Indian Society of Agricultural Statistics, 58*, 331-343.
- Cochran, W. G. (1963). *Sampling Techniques. 2<sup>nd</sup> Ed.* New York: Wiley.
- Eckler, A. R. (1955). Rotation sampling. *Annals of Mathematical Statistics, 26*, 664-685.
- Garcia, A., & Artes, E. (2002). Improvement on estimating current population ratio in successive sampling. *Brazilian Journal of Probability Statistics, 16*, 107-122.
- Gorden, L. (1983). Successive sampling in finite populations. *The Annals of Statistics, 11*, 702-706.

Table 3: Sample Sizes and Corresponding Expected Cost of Survey, which Give Equal Precision of Proposed Estimate  $T_{21}^*$  over  $T_{21}$  for Different Values of  $\rho$ ,  $\rho_{(2)}$ ,  $S_y$  and  $S_{y(2)}$

$$\rho = 0.4, \rho_{(2)} = 0.8, (k-1) = 0.5, W_2 = 0.2,$$

$S_y < S_{y(2)}$		$C_1^*$	$S_y > S_{y(2)}$		$C_1^*$	$S_y = S_{y(2)}$		$C_1^*$
$S_y$	$S_{y(2)}$		$S_y$	$S_{y(2)}$		$S_y$	$S_{y(2)}$	
0.7	0.8	2354.88	0.8	0.7	2288.02	0.6	0.6	2317.93
0.6	0.8	2407.35	0.8	0.6	2260.93	0.6	0.6	2317.93
0.5	0.8	2483.36	0.8	0.5	2237.14	0.6	0.6	2317.93

$$(k-1) = 0.5, W_2 = 0.5, S_y = 0.7, S_{y(2)} = 0.8$$

$\rho < \rho_{(2)}$		$C_1^*$	$\rho > \rho_{(2)}$		$C_1^*$	$\rho = \rho_{(2)}$		$C_1^*$
$\rho$	$\rho_{(2)}$		$\rho$	$\rho_{(2)}$		$\rho$	$\rho_{(2)}$	
0.5	0.6	3496.39	0.7	0.3	3762.69	0.6	0.6	3612.65
0.4	0.6	3476.17	0.7	0.4	3732.99	0.6	0.6	3612.65
0.3	0.6	3450.64	0.7	0.5	3693.69	0.6	0.6	3612.65

Table 4: Sample Sizes and Corresponding Expected Cost of Survey, which Give Equal Precision of Proposed Estimate  $T_{21}^*$  over  $T_{21}$  for Different Values of  $W_2$  and  $(k-1)$

$\rho = 0.2, \rho_{(2)} = 0.7,$ $(k-1) = 0.5, S_y = 0.8, S_{y(2)} = 0.4$		$\rho = 0.4, \rho_{(2)} = 0.5,$ $W_2 = 0.3, S_y = 0.7, S_{y(2)} = 0.8$	
$W_2$	$C_1^*$	$(k-1)$	$C_1^*$
0.2	2260.11	1.5	1802.75
0.3	3161.32	1.0	2213.82
0.4	4075.08	0.5	3523.25

Hansen, M. H., & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.

Jessen, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agricultural Experiment Statistical Research Bulletin*, 304.



## POPULATION MEAN ESTIMATION IN SUCCESSIVE NON-RESPONDENT SAMPLING

Miranda, M. D. M., Garcia, M. R., Cebrian, A. A., Montoya, Y. R., & Aguilera, S. G. (2007). Quantile estimation under successive sampling. *Computational Statistics*, 20(3), 385-399.

Narain, R. D. (1953). On the recurrence formula in sampling on successive occasions. *Journal of the Indian Society of Agricultural Statistics*, 5, 96-99.

Okafor, F. C. (2005). Sub-sampling the non-respondents in two-stage sampling over two successive occasions. *Journal of the Indian Statistical Association*, 43(1), 33-49.

Okafor, F. C., & Lee, H. (2000). Double sampling for ratio and regression estimation with sub sampling the non-respondent. *Survey Methodology*, 26(2), 183-188.

Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society, Sec. B*, 12, 241-255.

Sen, A. R. (1979). Sampling theory on repeated occasions with ecological applications. *Sampling Biological Populations*, 315-328.

Singh, G. N. (2003). Estimation of population mean using auxiliary information on recent occasion in h occasions successive sampling. *Statistics in Transition*, 6, 523-532.

Singh, G. N., & Priyanka, K. (2007). Estimation of population mean at current occasion in successive sampling under a super-population model. *Model Assisted Statistics and Applications*, 2, 189-200.

Singh, G. N., & Singh, V. K. (2001). On the use of auxiliary information in successive sampling. *Journal of the Indian Society for Agricultural Statistics*, 54(1), 1-12.

Singh, H. P., & Vishwakarma, G., K. (2007). A general class of estimators in successive sampling. *METRON*, LXV, 2, 201-227.

Singh, H. P., & Kumar, S. (2010). Estimation of mean in presence of non-response using two phase sampling scheme. *Statistical Papers*, 50, 559-582.

Singh, V. K., Singh, G. N., & Shukla, D. (1991). An efficient family of ratio-cum difference type estimators in successive sampling over two occasions. *Journal of Scientific Research*, 41C, 149-159.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., & Asok, C. (1984). *Sampling theory of surveys with applications*. Ames, IA, USA: Iowa State University Press and New Delhi, India: Indian Society of Agricultural Statistics.

## Improved Estimation of the Population Mean Using Known Parameters of an Auxiliary Variable

Rajesh Tailor    Balkishan Sharma  
Vikram University  
Ujjain, M.P., India

An improved ratio-cum-product type estimator of the finite population mean is proposed using known information on the coefficient of variation of an auxiliary variate and correlation coefficient between a study variate and an auxiliary variate. Realistic conditions are obtained under which the proposed estimator is more efficient than the simple mean estimator, usual ratio and product estimators and estimators proposed by Singh and Diwivedi (1981), Pandey and Dubey (1988), Upadhaya and Singh (1999), and Singh, et al., (2004). An empirical study supports theoretical findings.

Key words: Study variate, auxiliary variate, population mean, correlation coefficient, coefficient of variation.

### Introduction

Auxiliary information is frequently used at the estimation stage in order to improve the efficiency of the estimator(s) of the parameter(s) of a variate under study; ratio, product and regression methods of estimation are examples. When the correlation between study variate and the auxiliary variate is positive (high), the ratio method of estimation is used for estimating the population mean. Conversely, if the correlation is negative, the product method of estimation is preferred.

Consider a finite population  $U = (U_1, U_2, \dots, U_N)$  of  $N$  units. Let  $y_i$  and  $x_i$  be the values of the study variate  $y$  and auxiliary variate  $x$  respectively on the  $i^{\text{th}}$  unit  $U_i$  ( $i=1,2,3,\dots,N$ ). For estimating the population mean,  $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$ , of the study variate  $y$ , a

simple random sample of size  $n$  is drawn using the simple random sampling without replacement (SRSWOR) technique from  $U$ .

When the population mean  $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$ , of the auxiliary variate  $x$  is known, the classical ratio and product estimators of  $\bar{Y}$  are respectively defined by the ratio estimator

$$\hat{Y}_1 = \bar{y} \left( \frac{\bar{X}}{\bar{x}} \right) \quad (1.1)$$

and the product estimator

$$\hat{Y}_2 = \bar{y} \left( \frac{\bar{x}}{\bar{X}} \right) \quad (1.2)$$

where  $\bar{y} = \sum_{i=1}^n y_i / n$  and  $\bar{x} = \sum_{i=1}^n x_i / n$  are the sample means of  $y$  and  $x$  respectively based on  $n$  observations.

When the population mean  $\bar{X}$  and coefficient of variation ( $C_x$ ) of auxiliary variate  $x$  are known, Sisodia and Dwivedi (1981) suggested using a ratio type estimator for  $\bar{Y}$  as

Rajesh Tailor is a Reader in the School of Studies in Statistics. His research interests are in the field of sampling techniques. Email: tailorraj@gmail.com. Balkishan Sharma is a research scholar working on his Ph.D. under the supervision of Dr. Rajesh Tailor.

## IMPROVED ESTIMATION OF POPULATION MEAN

$$\hat{Y}_3 = \bar{y} \left( \frac{\bar{X} + C_x}{\bar{x} + C_x} \right). \quad (1.3)$$

Using the same of information, Pandey and Dubey (1988) suggested a product type estimator for  $\bar{Y}$  as

$$\hat{Y}_4 = \bar{y} \left( \frac{\bar{x} + C_x}{\bar{X} + C_x} \right) \quad (1.4)$$

Further, when the population mean  $\bar{X}$  of  $x$  and the correlation coefficient ( $\rho$ ) between  $y$  and  $x$  are known, Singh and Tailor (2003) suggested ratio and product type estimators for  $\bar{Y}$  respectively as

$$\hat{Y}_5 = \bar{y} \left( \frac{\bar{X} + \rho}{\bar{x} + \rho} \right) \quad (1.5)$$

and

$$\hat{Y}_6 = \bar{y} \left( \frac{\bar{x} + \rho}{\bar{X} + \rho} \right) \quad (1.6)$$

Kadilar and Singi (2006) suggested a ratio-type and a product type estimator for  $\bar{Y}$ , using coefficient of variation  $C_x$  and correlation coefficient ( $\rho$ ), as

$$\hat{Y}_7 = \bar{y} \left( \frac{\bar{X}C_x + \rho}{\bar{x}C_x + \rho} \right) \quad (1.7)$$

This study proposes a ratio-cum-product estimator utilizing the knowledge on  $\bar{X}$ ,  $C_x$  and  $\rho$  and its properties are examined.

### Proposed Ratio-Cum-Product Estimator

Motivated by Singh and Tailor (2005), the proposed ratio-cum-product estimator for  $\bar{Y}$  is

$$\hat{Y}_B = \bar{y} \left[ \alpha \left( \frac{\bar{X}C_x + \rho}{\bar{x}C_x + \rho} \right) + (1 - \alpha) \left( \frac{\bar{x}C_x + \rho}{\bar{X}C_x + \rho} \right) \right] \quad (2.1)$$

where  $\alpha$  is a suitably chosen scalar. It should be noted that  $\alpha = 1$ ,  $\hat{Y}_B$  reduces to the estimator  $\hat{Y}_7$  suggested by Kadilar and Cingi (2006) and for the  $\alpha = 0$  product version of the  $\hat{Y}_7$ . Thus, these two estimators are particular cases of the proposed estimator  $\hat{Y}_B$ . To obtain the bias and MSE of  $\hat{Y}_B$ ,  $\bar{y} = \bar{Y}(1 + e_0)$  and  $\bar{x} = \bar{X}(1 + e_1)$  such that  $E(e_0) = E(e_1) = 0$  and

$$E(e_0^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_y^2,$$

$$E(e_1^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_x^2,$$

$$E(e_0 e_1) = \left( \frac{1}{n} - \frac{1}{N} \right) \rho C_y C_x$$

where

$$C_y = S_y / \bar{Y},$$

$$C_x = S_x / \bar{X},$$

$$\rho = S_{yx} / S_x S_y,$$

$$K = \rho C_y / C_x,$$

and

$$S_x^2 = \sum_{i=1}^N (x_i - \bar{X})^2 / (N - 1),$$

$$S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 / (N - 1),$$

$$S_{xy} = \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X}) / (N - 1).$$

Expressing (2.1) in terms of  $e_i$ 's results in

$$\hat{Y}_B = \bar{Y}(1 + e_0) \left[ \alpha(1 + \lambda_3 e_1)^{-1} + (1 - \alpha)(1 + \lambda_3 e_1) \right], \quad (2.2)$$

where

$$\lambda_3 = \bar{X}C_x / (\bar{X}C_x + \rho).$$

To the first degree of approximation, the bias and mean squared error of  $\hat{Y}_B$  respectively are

$$B(\hat{Y}_B) = \frac{(1-f)}{n} \bar{Y} \lambda_3 C_x^2 [K + \alpha(\lambda_3 - 2K)], \quad (2.3)$$

and

$$MSE(\hat{Y}_B) = \frac{(1-f)}{n} \bar{Y}^2 [C_y^2 + (1-2\alpha)\lambda_3 C_x^2 \{(1-2\alpha)\lambda_3 + 2K\}]. \quad (2.4)$$

Thus, with  $\alpha = K/(2K - \lambda_3)$ , the estimator  $\hat{Y}_B$  is almost unbiased. It is also observed from (2.3) that the bias of  $\hat{Y}_B$  is negligible for large sample.

The mean squared error of  $\hat{Y}_B$  in (2.4) is minimized for

$$\alpha = \frac{(\lambda_3 + K)}{2\lambda_3} = \alpha_0. \quad (2.5)$$

Substitution of (2.5) in (2.1) yields the asymptotically optimum estimator (AOE) for  $\bar{Y}$  as

$$\hat{Y}_B^{(opt)} = \frac{\bar{y}}{2\lambda_3} \left[ (\lambda_3 + K) \left( \frac{\bar{X}C_x + \rho}{\bar{x}C_x + \rho} \right) + (\lambda_3 - K) \left( \frac{\bar{x}C_x + \rho}{\bar{X}C_x + \rho} \right) \right] \quad (2.6)$$

and placing (2.5) in (2.3) and (2.4), results in the bias and variance of  $\hat{Y}_B^{(opt)}$  respectively as

$$B(\hat{Y}_B^{(opt)}) = \frac{(1-f)}{2n} \bar{Y} C_x^2 (\lambda_3 - K)(\lambda_3 + 2K), \quad (2.7)$$

and

$$MSE(\hat{Y}_B^{(opt)}) = \frac{(1-f)}{n} S_y^2 (1 - \rho^2). \quad (2.8)$$

It is clear that mean squared error of  $\hat{Y}_B^{(opt)}$  is the same as that of the approximate variance of the usual linear regression estimator  $\bar{y}_{lr} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x})$ , where  $\hat{\beta}$  is the sample regression coefficient of y on x.

#### Efficiency Comparisons

Under simple random sampling without replacement (SRSWOR), the variance of sample mean  $\bar{y}$  is

$$V(\bar{y}) = \frac{(1-f)}{n} \bar{Y}^2 C_y^2 \quad (3.1)$$

and the mean squared error of  $\hat{Y}_i$  (i=1 to 8) to the first degree of approximation are respectively given by:

$$MSE(\hat{Y}_1) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 - 2K)] \quad (3.2)$$

$$MSE(\hat{Y}_2) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 + 2K)] \quad (3.3)$$

$$MSE(\hat{Y}_3) = \theta \bar{Y}^2 [C_y^2 + \lambda_1 C_x^2 (\lambda_1 - 2K)] \quad (3.4)$$

$$MSE(\hat{Y}_4) = \theta \bar{Y}^2 [C_y^2 + \lambda_1 C_x^2 (\lambda_1 + 2K)] \quad (3.5)$$

$$MSE(\hat{Y}_5) = \theta \bar{Y}^2 [C_y^2 + \lambda_2 C_x^2 (\lambda_2 - 2K)] \quad (3.6)$$

$$MSE(\hat{Y}_6) = \theta \bar{Y}^2 [C_y^2 + \lambda_2 C_x^2 (\lambda_2 + 2K)] \quad (3.7)$$

$$MSE(\hat{Y}_7) = \theta \bar{Y}^2 [C_y^2 + \lambda_3 C_x^2 (\lambda_3 - 2K)] \quad (3.8)$$

where

$$\lambda_1 = \frac{\bar{X}}{\bar{X} + C_x},$$

$$\lambda_2 = \frac{\bar{X}}{\bar{X} + \rho},$$

IMPROVED ESTIMATION OF POPULATION MEAN

$$\lambda_3 = \frac{\overline{XC}_x}{\overline{XC}_x + \rho},$$

$$C_y = S_y / \bar{Y},$$

$$\theta = \left( \frac{1}{n} - \frac{1}{N} \right),$$

$$K = \frac{\rho C_y}{C_x}$$

and

$$S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 / (N-1).$$

From (2.4) and (3.1), it is observed that  $\hat{Y}_B$  is more efficient than the usual unbiased estimator  $\bar{y}$  if:

$$\left. \begin{array}{l} \text{either } \frac{1}{2} < \alpha < \left( \frac{1+K}{2+\lambda_3} \right) \\ \text{or } \left( \frac{1+K}{2+\lambda_3} \right) < \alpha < \frac{1}{2} \end{array} \right\} \quad (3.9)$$

A comparison of (2.4) and (3.2) shows that  $\hat{Y}_B$  is more efficient than the usual ratio estimator  $\hat{Y}_1$  if:

$$\left. \begin{array}{l} \text{either } \frac{(1+\lambda_3)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3+2K-1}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3+2K-1}{2\lambda_3} \right) < \alpha < \frac{(1+\lambda_3)}{2\lambda_3} \end{array} \right\} \quad (3.10)$$

From (2.4) and (3.3) it is clear that  $\hat{Y}_B$  would be more efficient than  $\hat{Y}_2$  if:

$$\left. \begin{array}{l} \text{either } \left( \frac{\lambda_3+2K+1}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3-1)}{2\lambda_3} \\ \text{or } \frac{(\lambda_3-1)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3+2K+1}{2\lambda_3} \right) \end{array} \right\} \quad (3.11)$$

Comparing (2.4) and (3.4), it is observed that  $\hat{Y}_B$  is more efficient than the Sisodia and Dwivedi (1981) estimator  $\hat{Y}_3$  if

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3+\lambda_1)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3-\lambda_1+2K}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3-\lambda_1+2K}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3+\lambda_1)}{2\lambda_3} \end{array} \right\} \quad (3.12)$$

Comparing (2.4) and (3.5), it is observed that  $\hat{Y}_B$  is more efficient than the Pandey and Dubey (1988) estimator  $\hat{Y}_4$  if

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3-\lambda_1)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3+\lambda_1+2K}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3+\lambda_1+2K}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3-\lambda_1)}{2\lambda_3} \end{array} \right\} \quad (3.13)$$

Comparing (2.4) and (3.6), conditions under which suggested estimator  $\hat{Y}_B$  is more efficient than the Singh and Tailor (2003) ratio type estimator  $\hat{Y}_5$  when

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3+\lambda_2)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3-\lambda_2+2K}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3-\lambda_2+2K}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3+\lambda_2)}{2\lambda_3} \end{array} \right\} \quad (3.14)$$

Similarly conditions under which suggested estimator  $\hat{Y}_B$  is more efficient than the Singh and Tailor (2005) product type estimator  $\hat{Y}_6$  when

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3 - \lambda_2)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3 + 2K + \lambda_2}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3 + 2K + \lambda_2}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3 - \lambda_2)}{2\lambda_3} \end{array} \right\} \quad (3.15)$$

Comparing (2.4) and (3.8), it is observed that  $\hat{Y}_B$  is more efficient than the Kadilar and Cingi (2006) ratio type estimator  $\hat{Y}_7$ , if

$$\left. \begin{array}{l} \text{either } \frac{1}{2} < \alpha < \frac{K}{\lambda_3} \\ \text{or } \frac{K}{\lambda_3} < \alpha < \frac{1}{2} \end{array} \right\} \quad (3.16)$$

#### Empirical Study

To analyze the performance of the proposed estimator compared to other estimators, three natural population data sets were considered. The description of the populations is given by Annexure.

Population I (Sukhatme & Sukhatme, 1970, p. 256):

y: Number of villages in the circles  
 x: A circle consisting more than five villages  
 $\bar{Y} = 3.360$   
 $\bar{X} = 0.1236$   
 $\rho = 0.766$   
 $C_y = 0.60400$   
 $C_x = 2.19012$

Population II (Cochran, 1977):

y: The number of persons per block  
 x: The number of rooms per block  
 $\bar{Y} = 101.1$

$\bar{X} = 58.80$   
 $\rho = 0.6500$   
 $C_y = 0.14450$   
 $C_x = 0.1281$

Population III (Kadilar & Singi, 2003):

y: Level of apple production  
 x: number of apple trees  
 $\bar{Y} = 625.37$   
 $\bar{X} = 13.93$   
 $\rho = 0.865$   
 $C_y = 1.866$   
 $C_x = 1.653,$

#### Results

Table 4.1 shows a significant gain in efficiency by using proposed estimator  $\hat{Y}_B$  ( $\hat{Y}_B^{(opt)}$ ) over the unbiased estimator  $\hat{Y}$ , the usual ratio estimator  $\hat{Y}_1$ , the product estimator  $\hat{Y}_2$ , the Sisodiya and Dwivedi (1981) estimator  $\hat{Y}_3$ , the Pandey and Dubey (1988) estimator  $\hat{Y}_4$ , the Singh and Tailor. (2003) estimators  $\hat{Y}_5$  and  $\hat{Y}_6$ , and the Kadilar and Singi (2006) estimator  $\hat{Y}_7$ .

Table 4.2 illustrates the wide range of  $\alpha$  in which a suggested estimator  $\hat{Y}_B$  or  $\hat{Y}_B^{(opt)}$  is more efficient than all estimators considered in this study; it shows that even if the scalar  $\alpha$  deviates from its optimum value ( $\alpha_{opt}$ ), the suggested estimator  $\hat{Y}_B^{(opt)}$  will yield better estimates than  $\hat{Y}, \hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \hat{Y}_4, \hat{Y}_5, \hat{Y}_6$  and  $\hat{Y}_7$ . Therefore, the suggested estimator  $\hat{Y}_B^{(opt)}$  is recommended for use in practice.

## IMPROVED ESTIMATION OF POPULATION MEAN

Table 4.1: Percent Relative Efficiencies of  $\hat{Y}$ ,  $\hat{Y}_1$ ,  $\hat{Y}_2$ ,  $\hat{Y}_3$ ,  $\hat{Y}_4$ ,  $\hat{Y}_5$ ,  $\hat{Y}_6$ ,  $\hat{Y}_7$  and  $\hat{Y}_B$  or  $\hat{Y}_B^{(opt)}$   
With Respect To  $\hat{Y}$

Estimators	$\hat{Y}$	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	$\hat{Y}_4$	$\hat{Y}_5$	$\hat{Y}_6$	$\hat{Y}_7$	$\hat{Y}_B^{(opt)}$
Population I	100.00	11.64	5.08	134.99	74.95	207.47	49.37	224.25	241.99
Population II	100.00	157.87	34.03	158.09	34.10	158.99	34.38	165.29	173.16
Population III	100.00	396.49	30.15	388.92	33.37	395.67	31.86	396.97	397.18

Table 4.2: Range of  $\alpha$  in Which  $\hat{Y}_B$  is Better than  $\hat{Y}$ ,  $\hat{Y}_1$ ,  $\hat{Y}_2$ ,  $\hat{Y}_3$ ,  $\hat{Y}_4$ ,  $\hat{Y}_5$ ,  $\hat{Y}_6$ ,  $\hat{Y}_7$

		Population		
		I	II	III
Range of $\alpha$ in which $\hat{Y}_B$ or $\hat{Y}_B^{(opt)}$ is Better Than	$\hat{Y}$	(0.5, 1.31)	(0.5, 1.30)	(0.5, 1.51)
	$\hat{Y}_1$	(-0.413, 0.165)	(0.638, 0.883)	(0.924, 0.946)
	$\hat{Y}_2$	(-0.096, 0.22)	(-0.037, 1.56)	(-0.017, 1.89)
	$\hat{Y}_3$	(0.041, 0.082)	(0.639, 0.883)	(0.895, 0.975)
	$\hat{Y}_4$	(0.027, 0.096)	(-0.036, 1.56)	(0.034, 1.84)
	$\hat{Y}_5$	(0.052, 0.057)	(0.512, 0.879)	(0.918, 0.964)
	$\hat{Y}_6$	(0.016, 0.087)	(-0.032, 1.24)	(0.010, 1.88)
	$\hat{Y}_7$	(0.5, 0.809)	(0.5, 0.796)	(0.5, 0.013)
Optimum Value of $\alpha$ ( $\alpha_0$ )		(0.0617)	(0.7612)	(0.9350)

### References

Sukhatme, P. V., & Sukhatme, B. V. (1970). *Sampling theory of surveys with applications*. Ames, IA: Iowa State University Press.

Das, A. K. (1988). *Contribution to the theory of sampling strategies based on auxiliary information*. Ph.D. Thesis, BCKV, West Bengal, India.

Pandey, B. N., & Dubey, V. (1988). Modified product estimator using coefficient of variation of auxiliary variable. *Assam Statistical Review*, 2, 64-66.

Singh, H. P., & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6, 555-560.

Sisodia, B. V. S., & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of Indian Society of Agricultural Statistics*, 33(1), 13-18.

Kadilar, C., & Cingi, H. (2006a). A new ratio estimator using correlation coefficient. *Inter Stat*, 1-11

Kadilar, C., & Cingi, H. (2006b). Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters*, 19, 75.

## Matched-Pair Studies with Misclassified Ordinal Data

Tze-San Lee  
Western Illinois University  
Macomb, IL USA

The problem of matched-pair studies with misclassified ordinal data is considered. Misclassification is assumed to occur only between the adjacent columns/rows. Bias-adjusted generalized odds ratio and a test for marginal homogeneity are presented to account for misclassification bias. Data from lambing records of 227 Merino ewes are used to illustrate how to calculate these bias-adjusted estimators and – because validation data are not available – a sensitivity analysis is conducted.

Key words: Matched-pair, misclassification, ordinal scale.

### Introduction

Matched-pair studies with ordered categorical variables have received much attention in the literature (see Agresti, 1983, 1984; Clayton, 1974; McCullagh, 1977; Stuart, 1953, 1955). A few published studies investigated matched-pair with misclassified data (Greenland, 1982, 1989; Greenland & Kleinbaum, 1983; Lee, 2010); however, these studies consider only  $2 \times 2$  contingency tables with misclassified data. To date, matched-pair studies with misclassified data have not been investigated when the number of exposure categories is greater than two. A matched-pair misclassification problem is considered here with an exposure variable that has  $K (\geq 3)$  ordered levels. The generalized odds ratio is used for measuring the association in contingency tables with misclassified ordinal data and a test for marginal homogeneity proposed by Stuart (1955) is modified to manage the misclassified data.

### Methodology

Consider a 1:1 matched-pair study where X represents the case and Y represents the control population and the same exposure variable with

$K (\geq 3)$  ordered levels is used. Assume that a  $K \times K$  contingency table is realized with the following frequency counts:

$$A = [n_{ij}]_{i,j=1,\dots,K} \quad (1)$$

where  $\{n_{ij}\}$  are assumed to follow a multinomial distribution with parameters  $n (= \sum_{i,j} n_{ij})$  and the cell probability  $\{p_{ij} > 0\}$ . A naïve estimator ( $\hat{p}_{ij}$ ) for  $p_{ij}$  is given by

$$\hat{p}_{ij} = n_{ij} / n. \quad (2)$$

### Generalized Odds Ratio

As a measure for the association between X and Y, a generalized odds ratio (GOR),  $\zeta$ , is defined by Agresti (1980) as:

$$\zeta = \frac{p_C}{p_D} \equiv \frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^K p_{ij}}{\sum_{i=2}^K \sum_{j=1}^{i-1} p_{ij}}, \quad (3)$$

where  $p_C$  (or  $p_D$ ) denotes the probability of a randomly selected matched-pair in which a case has a higher (or lower) level of exposure than his/her matched control. A naïve estimator, denoted by  $\hat{\zeta}$ , for (3) is obtained by replacing

Tze-San Lee is presently working at the Centers for Disease Control and Prevention, Mail Stop F-58, Chamblee, GA 34301, USA. Email: tj13@cdc.gov.



the unknown parameters  $p_{ij}$ ,  $i, j = 1, \dots, K$  by the sample estimator  $\hat{p}_{ij}$  shown in (2). Note that this naïve estimator of equation 3 could have substantial bias if the observed data in (1) are misclassified. Due to the faster convergence of  $\ln(\hat{\zeta})$ , a natural logarithm of  $\hat{\zeta}$ , to normality, is preferred to find a large sample Wald's  $100(1 - p)\%$  confidence interval:

$$\begin{aligned} & [\hat{\zeta} \cdot \exp(-z_{\alpha/2} \sqrt{\hat{\sigma}^2(\ln(\hat{\zeta}))}), \\ & \hat{\zeta} \cdot \exp(z_{\alpha/2} \sqrt{\hat{\sigma}^2(\ln(\hat{\zeta}))})] \end{aligned} \quad (4a)$$

for  $\zeta$ , where  $z_{\alpha/2}$  is the  $(\alpha/2)^{\text{th}}$  upper-tail percentile of the standard normal distribution. The asymptotic variance of  $\ln(\hat{\zeta})$  is given by Agresti (1980) as

$$\hat{\sigma}^2(\ln(\hat{\zeta})) = \sqrt{n^{-1}(\hat{p}_C^{-1} + \hat{p}_D^{-1})}, \quad (4b)$$

where  $\hat{p}_C$  (or  $\hat{p}_D$ ) is obtained by substituting equation 2 for  $p_{ij}$  in  $p_C$  (or  $p_D$ ).

#### A Test for Marginal Homogeneity

A global test for marginal homogeneity was proposed by Stuart (1955). A drawback of this global test is its failure to account for an ordinal nature in the categorical level of the exposure variable. Assume that the ordinal nature of the exposure variable is quantified by a variable  $U$  taking the score values  $u_k$  ( $u_1 < u_2 < \dots < u_K$ ) at the  $k^{\text{th}}$  level. Thus, the score test for the significance of the  $\beta$  coefficient in a linear trend model  $H_0: \beta_0 = \beta_1$ , where  $p_{i+} = \alpha_0 + \beta_0 u_i$  and  $p_{+j} = \alpha_1 + \beta_1 u_j$  is defined by

$$\hat{S} = \frac{n \cdot \left[ \sum_{j=1}^{K-1} \sum_{i=j+1}^K (\hat{p}_{ji} - \hat{p}_{ij})(u_j - u_i) \right]^2}{\sum_{j=1}^{K-1} \sum_{i=j+1}^K \hat{p}_{ij} (u_j - u_i)^2}, \quad (5)$$

where  $\hat{p}_{ij}$  is defined by equation 2,  $\hat{P}_{ij} = \hat{p}_{ij} + \hat{p}_{ji}$ , and  $u_i = i - 1$ ,  $i = 1, \dots, K$ . By a large sample theory (5) is distributed as a Chi-square distribution with 1 degree of freedom (Breslow, 1982).

#### Misclassification Probability

Suppose that the observed  $K \times K$  contingency table shown in (1) were misclassified with respect to both  $X$  and  $Y$ . Let  $X^*$  and  $Y^*$  be the classified surrogate variables for  $X$  and  $Y$ , respectively. Furthermore, assume that only adjacent rows in  $X^*$  or adjacent columns in  $Y^*$  are misclassified. For  $Z = X, Y$ , the misclassification probabilities (MPs) for the row or column variable are defined as follows:

$$\begin{aligned} \phi_{Z[k;j]} &= \Pr(Z^* = k + 1 | Z = k; \bar{Z} = j), \\ \overline{\phi_{Z[k;j]}} &= 1 - \phi_{Z[k;j]}; \end{aligned} \quad (6a)$$

and

$$\begin{aligned} \psi_{Z[j;k]} &= \Pr(Z^* = j - 1 | Z = j; \bar{Z} = k), \\ \overline{\psi_{Z[j;k]}} &= 1 - \psi_{Z[j;k]}; \end{aligned} \quad (6b)$$

where  $\bar{Z} = Y$  if  $Z = X$ ; and vice versa. Note that, due to symmetry,  $\phi_{Z[k;j+1]} = \psi_{Z[k+1;j]}$ . If

$$p = [p_{11}, \dots, p_{1K}, p_{21}, \dots, p_{2K}, \dots, p_{K1}, \dots, p_{KK}]^T, \quad (7a)$$

and

$$\hat{p} = [\hat{p}_{11}, \dots, \hat{p}_{1K}, \hat{p}_{21}, \dots, \hat{p}_{2K}, \dots, \hat{p}_{K1}, \dots, \hat{p}_{KK}]^T, \quad (7b)$$

then the expected value of a naïve estimator for equation 2 is given by

$$E(\hat{p}) = Wp \quad (8)$$

where the misclassification matrix  $W$  is a  $K^2 \times K^2$  matrix defined, respectively, by

$$W = \begin{bmatrix} W_{11} & W_{12} & 0 & 0 & \dots & \dots \\ W_{21} & W_{22} & W_{23} & 0 & 0 & \dots \\ 0 & W_{32} & W_{33} & W_{34} & \dots & \dots \\ 0 & 0 & W_{43} & W_{44} & \dots & \dots \\ \vdots & \vdots & 0 & W_{54} & \dots & \dots \\ \vdots & \vdots & \vdots & 0 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & \vdots \end{bmatrix} \quad (9)$$

where

$$\begin{aligned} W_{kl} &= [w_{ij}^{[kl]}], i, j, k, l = 1, 2, \dots, K; w_{11}^{[11]} = \overline{\varphi_{X[1;1]}} - \varphi_{Y[1;1]}, \\ w_{KK}^{[11]} &= \overline{\psi_{Y[K;1]}} - \varphi_{X[1;K]}, w_{22}^{[11]} = \overline{\psi_{Y[2;1]}} - \varphi_{X[1;1]} - \varphi_{Y[2;1]}, \\ w_{kk}^{[11]} &= \overline{\psi_{Y[k-1;1]}} - \varphi_{X[1;k-1]} - \varphi_{X[2;k-1]} - \dots - \psi_{Y[k-1;k-1]}, \\ k &= 3, \dots, K-1; w_{k,k+1}^{[11]} = \psi_{Y[k;k-1]}, k = 1, \dots, K-1; \\ w_{k+1,k}^{[11]} &= \varphi_{Y[k-1;k]}, k = 2, \dots, K, 0 \text{ elsewhere}; \\ W_{12} &= \text{diag}[\psi_{Y[2;1]}, \psi_{Y[3;1]}, \dots, \psi_{Y[K;1]}]; \\ W_{21} &= \text{diag}[\varphi_{Y[1;1]}, \varphi_{Y[2;1]}, \dots, \varphi_{Y[K-1;1]}]; w_{K+i,K+i}^{[22]} = \varphi_{X[1;1]}, \\ i &= 1, \dots, K-1; w_{2K,2K}^{[22]} = \overline{\psi_{Y[2K;2K-1]}} - \psi_{X[2K-1;2K]} - \varphi_{X[2K-1;2K]}, \\ w_{kk}^{[22]} &= \overline{\psi_{X[k-K;k-K]}} - \varphi_{X[k-K;k-K]} - \psi_{Y[k-K;k]} - \varphi_{Y[k-K;k-K]}, \\ k &= K+2, \dots, 2K-1, \dots, w_{k,k}^{[kk]} = \varphi_{X[k-K;k]}, \\ k &= (K-1)K+1, \dots, K^2-1, \\ w_{k,k}^{[kk]} &= \overline{\psi_{X[K-1;K(K-1)+1]}} - \varphi_{Y[1;K(K-1)+1]} - \psi_{X[K-1;k]}, \\ k &= K(K-1)+1, \dots, K^2-1, w_{K^2,K^2}^{[KK]} = \overline{\psi_{X[K-1;K^2]}} - \psi_{Y[K-1;K^2]}. \end{aligned}$$

Note that W is a block tri-diagonal matrix.

### Bias-Adjusted Cell Proportion

Using equation 8, a bias-adjusted cell proportion (BACP) estimator for p is given by

$$\tilde{p} = [\tilde{p}_{11}, \dots, \tilde{p}_{1K}, \dots, \tilde{p}_{K1}, \dots, \tilde{p}_{KK}]^T = W^{-1} \hat{p} = V \hat{p}, \quad (10a)$$

where  $\hat{p}$  is defined by (2) and V is defined by

$$W^{-1} \equiv V = [v_{ij}]_{i,j=1,\dots,K^2}. \quad (10b)$$

The appendix shows how to calculate its inverse V of the misclassification matrix W for K = 3, which was used to analyze the data for Table 1. When K = 3, then for i, j = 1, 2, 3

$$\tilde{p}_{ij} = \sum_{k=1}^3 (v_{ik} \hat{p}_{ij} + v_{i,k+3} \hat{p}_{i+1,j} + v_{i,k+6} \hat{p}_{i+2,j}), \quad (11)$$

where  $\{v_{ij}\}$ , i, j = 1, 2, ..., 9 are given respectively by equation A5 in the appendix with

$$\begin{aligned} \det(W_{11}) &= 1 - 2\varphi_{Y[1;1]} - 3\varphi_{Y[2;1]} - 2\psi_{Y[3;1]} \\ &\quad + 5\varphi_{Y[2;1]} (\varphi_{Y[1;1]} + \psi_{Y[3;1]}) \\ &\quad + 4\varphi_{Y[1;1]} \psi_{Y[3;1]} (1 - 2\overline{\varphi_{Y[2;1]}}) \end{aligned} \quad (12)$$

$$\begin{aligned} \det(\Delta_i) &= \delta_{11}^{[i]} (\delta_{22}^{[i]} \delta_{33}^{[i]} - \delta_{23}^{[i]} \delta_{32}^{[i]}) \\ &\quad - \delta_{12}^{[i]} (\delta_{21}^{[i]} \delta_{33}^{[i]} - \delta_{31}^{[i]} \delta_{23}^{[i]}) \\ &\quad + \delta_{13}^{[i]} (\delta_{21}^{[i]} \delta_{32}^{[i]} - \delta_{31}^{[i]} \delta_{22}^{[i]}), i = 1, 2 \end{aligned} \quad (13 \& 14)$$

Where  $\{\delta_{jk}^{[i]}\}$ , i = 1, 2 are given, respectively, by equations A5 and A6 in the appendix.

A set of MPs is said to be feasible if the values of all three determinants,  $\det(W_{11})$ ,  $\det(\Delta_1)$  and  $\det(\Delta_2)$ , from (12), (13) and (14), are nonzero for the given set of equation 6. Furthermore, a set of MPs is said to be admissible if - for all feasible  $\varphi_{Z[i;j]}$  and  $\psi_{Z[j;i]}$  - where Z = X, Y, the constraint of the sum of total probability  $\{\tilde{p}_{ij}\}$ , i, j = 1, 2, 3, that is,

$$\sum_{i=1}^3 \sum_{j=1}^3 \tilde{p}_{ij} = 1, \text{ is satisfied where } 0 < \tilde{p}_{ij} < 1.$$

### Bias-Adjusted Generalized Odds Ratio

By substituting (11) into (3), a bias-adjusted generalized odds ratio (BAGOR) is thus defined by

## MATCHED-PAIR STUDIES WITH MISCLASSIFIED ORDINAL DATA

$$\zeta = \frac{\bar{p}_C}{\bar{p}_D} \equiv \frac{\bar{p}_{12} + \bar{p}_{13} + \bar{p}_{23}}{\bar{p}_{21} + \bar{p}_{31} + \bar{p}_{32}} = \frac{\sum_{i,j=1}^3 v_{ij}^* \hat{p}_{ij}}{\sum_{i,j=1}^3 v_{ij}^{**} \hat{p}_{ij}}, \quad (15a)$$

where  $\{\hat{p}_{ij}\}$  are given by (2),  $\{v_{ij}^*\}$  and  $\{v_{ij}^{**}\}$  are given respectively for  $j = 1, 2, 3$ , by

$$\begin{aligned} v_{1j}^* &= v_{2j} + v_{3j} + v_{6j}, \\ v_{2j}^* &= v_{2,j+3} + v_{3,j+3} + v_{6,j+3}, \\ v_{3j}^* &= v_{2,j+6} + v_{3,j+6} + v_{6,j+6}, \end{aligned} \quad (15b)$$

and

$$\begin{aligned} v_{1j}^{**} &= v_{4j} + v_{7j} + v_{8j}, \\ v_{2j}^{**} &= v_{4,j+3} + v_{7,j+3} + v_{8,j+3}, \\ v_{3j}^{**} &= v_{4,j+6} + v_{7,j+6} + v_{8,j+6}, \end{aligned} \quad (15c)$$

and  $\{v_{ij}\}$  are given by equation A7 in the appendix. Using the delta method (Goodman & Kruskal, 1972), the asymptotic variance of  $\ln(\zeta)$  is given by

$$\sigma^2(\ln(\zeta)) = \frac{\sum_{i,j=1}^3 v_{ij}^{**} \sum_{i,j=1}^3 v_{ij}^* p_{ij} - \sum_{i,j=1}^3 v_{ij}^* \sum_{i,j=1}^3 v_{ij}^{**} p_{ij}}{n(\zeta \sum_{i,j=1}^3 v_{ij}^{**} p_{ij})^2}, \quad (15d)$$

where  $\{v_{ij}^*\}$  and  $\{v_{ij}^{**}\}$  are given, respectively, by equations 15(b-c). A large sample Wald's  $100(1-\alpha)\%$  confidence interval is given by

$$\left[ \zeta \exp\left(-z_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2(\ln(\zeta))}\right), \zeta \exp\left(-z_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2(\ln(\zeta))}\right) \right], \quad (15e)$$

where

$$\hat{\sigma}^2(\ln(\zeta)) = \sigma^2(\ln(\zeta)) \Big|_{p_{ij} = \bar{p}_{ij}}.$$

Bias-Adjusted Test for Marginal Homogeneity  
Substituting equation (11) into (5) for  $\hat{p}_{ij}$ , a bias-adjusted test for marginal homogeneity (BATMH) is given by

$$\tilde{S} = \frac{n \cdot \left[ \sum_{j=1}^{K-1} \sum_{i=j+1}^K (\bar{p}_{ji} - \bar{p}_{ij})(z_j - z_i) \right]^2}{\sum_{j=1}^{K-1} \sum_{i=j+1}^K \bar{P}_{ij} (z_j - z_i)^2}, \quad (16)$$

where  $\{\bar{p}_{ij}\}$  are given by equation (11), and  $\bar{P}_i = \bar{p}_{i+} + \bar{p}_{+i}$ .

### Results

Table 1 shows the first and second lambing records of a flock of 227 Merino ewes from 1952-1953 (Tallis, 1962). If the data in Table 1 are not misclassified, then the naïve GOR can be calculated as 1.22 (95% CI: 1.12–1.32) using equations 3 and 4. This implies that a significant association exists between the number of lambing records in 1952 and 1953. Also, the test value of equation 5 is obtained as  $\hat{S} = 70.0$  with  $p < 0.0001$  which indicates that the marginal distribution of the lambing records in 1952 is significantly different from that of 1953.

Table 1: Cross-Classification of Ewes According to Number of Lambs Born in Consecutive Years

Number of Lambs (1953)	Number of Lambs (1952)			Total
	0	1	2	
0	58	52	1	111
1	26	58	3	87
2	8	12	9	29
Total	92	122	13	227

Suppose that errors are present in the classification of the lambing records in Table 1; in that case, the bias-adjusting method would be applied. In order to use the formula of equation 11, the true MPs must be calculated. In order to accomplish this task, it is necessary to know the true cell counts; Through the use of theory of counterfactuals it is intuitively clear that the issue of getting a true table is simply a counterfactual of the observed [misclassified] table which is thought the factual one (Lewis, 1973). Hence, the above idea may be applied to obtain the hypothetically true cell counts by reshuffling the number of misclassified subjects in the observed table.

Because the row/column marginal totals in case-control studies have to be kept as being fixed, four out of nine cells can be chosen as free parameters to construct the true (counterfactual) table. By noting that there are two cells (1,3) and (2,3) with small observed counts, these two cells and two other cells (2,1) and (3,2) are selected as free parameters to construct ten true tables (column 2, Table 2). With 1 in the (1,3) cell to be kept unchanged, the assumed number of under- or over-misclassified subjects starts with the (2,3) cell and then increases one by one up to seven in that cell, while the assumed number of under- or over-misclassified subjects the other two cells (2,1) and (3,2) are chosen

discreetly we ended up with eight true cell count tables (#1 to #8, column 2 of Table 2); True cell counts in #9 and #10 of Table 2 are similarly constructed.

With the true cell counts as given, it is a matter of straightforward calculation to obtain true MPs; the MPs are calculated as the ratio of difference between true and observed marginal totals divided by their sum. These corresponding MPs were calculated (column 3, Table 2): the details are similar to that of Lee (2009a, 2010) and are hence omitted here. In order to check the feasibility of the MPs, three determinants (equations 12-14),  $\det(W_{11})$ ,  $\det(\Delta_1)$  and  $\det(\Delta_2)$ , were calculated. After examining their values, they are all feasible because all the determinant values are positive (columns 2-4, Table 3).

Although all MPs are feasible, it is interesting to note that only five out of ten (#1 to #5) are admissible because (1) they are positive real numbers between 0 and 1, and (2) they satisfy the constraint on the total probability sum:

$$\sum_{i=1}^3 \sum_{j=1}^3 \tilde{p}_{ij} = 1 \text{ (column 5, Table 3). As a result,}$$

BAGORs and BATMHs were calculated for models #1 to #5 (columns 2 and 3, Table 4). number of under- or over-misclassified subjects

Table 2: Ten Assumed True Cell Counts and their Corresponding MPs for Table 1

#	$(n_{11}, n_{12}, n_{13}; n_{21}, n_{22}, n_{23}; n_{31}, n_{32}, n_{33})$	$(a_{x[1]}, b_{x[1]}, d_{x[1]}; a_{x[2]}, b_{x[2]}, d_{x[2]}; a_{x[3]}, b_{x[3]}, d_{x[3]})^*$
1	(57,53,1;27,57,3;8,12,9)	(0.8,3,0;6,2,0;0,0,0)
2	(57,53,1;27,56,4;8,13,8)	(0.8,3,0;6,4,50;0,1,30)
3	(56,54,1;27,55,5;9,13,7)	(9,6,0;6,7,80;30,10,60)
4	(55,55,1;28,53,6;9,14,6)	(10,9,0;10,10,110;30,30,10)
5	(54,56,1;29,51,7;9,15,5)	(20,10,0;20,20,130;30,40,140)
6	(48,62,1;30,49,8;14,11,4)	(50,30,0;20,20,150;140,10,190)
7	(45,65,1;31,47,9;16,10,3)	(60,40,0;30,30,170;170,30,250)
8	(49,61,1;25,52,10;18,9,2)	(40,30,0;7,20,180;190,50,320)
9	(55,54,2;24,60,3;13,8,8)	(10,6,170;10,4,0;120,70,30)
10	(55,54,2;22,61,4;15,7,7)	(10,6,170;30,6,50;150,90,60)

\*All entries inside the parenthesis defined by equations A1 and A2 in the appendix need to multiply by  $10^{-3}$ .

## MATCHED-PAIR STUDIES WITH MISCLASSIFIED ORDINAL DATA

The value of BAGOR/BATMH was not computed if the corresponding BACPs were inadmissible.

Table 4 shows that admissible BACPs the BAGOR ( $\check{\zeta} = 2.08$ ) is biased further away

from the value of the null hypothesis than the classical estimator ( $\hat{\zeta} = 1.22$ ), but the BATMHs ( $\check{S} = 12.0$  is biased toward the null value than the classical estimator ( $\hat{S} = 70.0$ ).

Table 3: Feasibility and Admissibility of MPs and/or BACP in Table 2

#	$\det(W_{11})$	$\det(\Delta_1)$	$\det(\Delta_2)$	$(\check{P}_{11}, \check{P}_{12}, \check{P}_{13}, \check{P}_{21}, \check{P}_{22}, \check{P}_{23}, \check{P}_{31}, \check{P}_{32}, \check{P}_{33})$
1	0.99	0.96	0.96	(0.260, 0.0008, 0.004, 0.11, 0.26, 0.22, 0.03, 0.05, 0.04)
2	0.99	0.82	0.74	(0.26, 0.007, 0.005, 0.11, 0.26, 0.22, 0.03, 0.05, 0.03)
3	0.96	0.69	0.54	(0.27, 0.02, 0.008, 0.11, 0.26, 0.21, 0.03, 0.04, 0.03)
4	0.95	0.58	0.39	(0.27, 0.04, 0.02, 0.11, 0.26, 0.22, 0.02, 0.04, 0.03)
5	0.93	0.49	0.28	(0.27, 0.07, 0.04, 0.11, 0.26, 0.22, 0.02, 0.03, 0.02)
6	0.82	0.38	0.15	(0.32, 0.13, 0.06, 0.10, 0.25, 0.21, 0.02, -0.01, 0.02)
7	0.77	0.32	0.08	(0.34, 0.22, 0.14, 0.09, 0.22, 0.21, 0.01, -0.04, 0.006)
8	0.84	0.36	0.05	(0.37, 0.59, 0.54, 0.10, 0.11, 0.20, 0.008, -0.06, -0.002)
9	0.64	0.60	0.34	(0.32, -0.09, 0.02, 0.10, 0.28, 0.19, 0.01, -0.008, 0.04)
10	0.64	0.48	0.21	(0.34, -0.11, 0.006, 0.10, 0.28, 0.19, -0.003, -0.04, 0.03)

Table 4: Estimated BAGORs with 95% Confidence Interval (CI) and BATMHs with p-value for Table 3

#	$\check{\zeta}$ (95% CI)	$\check{S}$ (p - value)
1	0.74 (0.26 – 2.12)	0.49 (0.31)
2	1.17 (0.66 – 2.08)	2.05 (0.02)
3	1.34 (0.72 – 2.48)	4.05 (< 0.0001)
4	1.60 (0.86 – 3.00)	6.80 (< 0.0001)
5	2.08 (1.06 – 4.07)	12.0 (< 0.0001)
6-10	*	*

\*MPs are not admissible; thus values of  $\check{\zeta}/\check{S}$  are not calculated.

### Conclusion

A new method is presented here to study the misclassification problem associated with matched-pair case-control studies for the polytomous exposure variable. Based on results from this study, the following conclusions are put forth:

1. Determining whether there are classification errors in the collected data is a difficult issue. Strictly, this requires the principal investigator using personal expertise to exercise subjective judgment on the collected data. However, from the sensitivity analysis of this data set of lambing records, the method presented herein can vindicate itself empirically. Note that this example indicates that, at most, one record in the (1,3) cell can be under- or over-misclassified. It is impossible to have more than one record misclassified in that cell due to the occurrence of inadmissible MPs.
2. This method does not require non-differential misclassification as an assumption. In fact, differential misclassification is inclined to be the norm rather than exception in practical applications. Indeed, the example provided shows that, even if both the column and the row marginal totals misclassify, just the same number of records to their corresponding MPs are not the same because they have different marginal totals for the column and row respectively.
3. The direction of the bias is not the same for two measures of association - it depends on which measure is used.
4. The close-form formula for this method are derived only for  $K = 3$ . For  $K = 4, 5, 6$  it is workable to obtain the closed-form formula by hand. For much bigger values of  $K$ , it is a formidable task to work out all the details by hand. Fortunately, there is an alternative way to bypass the necessity of getting closed-form formula. Taking a closer look at two criteria for MA: feasibility and admissibility, it is found that feasibility is not essential, but admissibility is critical,

meaning that it is not necessary to pay much attention to feasibility, the main focus is only on admissibility. Hence, instead of getting closed-form formula, equation 10 can be solved numerically for BACP and the admissibility of MP checked by examining whether all components of BACP are positive real numbers between 0 and 1.

5. The confidence interval given by equations 4 or 15(e) is large sample asymptotic. If the sample sizes are small, an exact confidence interval should instead be used (Lui, 2002).

Although the traditional naïve estimator can be viewed as a special case of bias-adjusted estimator when all misclassification probabilities are zero, a huge difference exists between these two estimators. Note that a bias-adjusted generalized odds ratio as shown in equation (11) uses both the concordant and discordant data in the observed table, while the naïve estimator shown in (3) uses only the discordant data. As a result, a bias-adjusted generalized odds ratio will be more efficient than the naïve one.

Finally, a limitation of this study is that the results presented do not apply to a situation in which more than two adjacent columns/rows are misclassified in the contingency table. Clearly, the question remains open regarding how to adjust the naïve estimator for the misclassification bias if the assumption of only two adjacent columns/rows being misclassified is not satisfied.

### Acknowledgements

All numerical calculations obtained in Tables 2-4 are facilitated by using the Microsoft EXCEL spreadsheet. This article is an edited version of Lee (2009b), except that the data used for illustration has been changed from the unaided eyesight data of employees at Royal Ordnance factories in 1943-1946 to the current lambing records of Merino ewes in 1952-1953. Also, I was grateful for the editorial help which greatly improved the presentation of this paper.

### References

Agresti, A. (1980). Generalized odds ratios for ordinal data. *Biometrics*, 36, 59-67.

## MATCHED-PAIR STUDIES WITH MISCLASSIFIED ORDINAL DATA

Agresti, A. (1983). Testing marginal homogeneity for ordinal categorical variables. *Biometrics*, 39, 505-510.

Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

Breslow, N. (1982). Covariance adjustment of relative-risk estimates in matched studies. *Biometrics*, 38, 661-672.

Clayton, D. G. (1974). Some odds ratio statistics for the analysis of ordered categorical data. *Biometrika*, 61, 525-531.

Goodman, L. A., & Kruskal, W. H. (1972). Measures of association for cross classifications IV: Simplification of asymptotic variances. *Journal of the American Statistical Association*, 67, 415-421.

Greenland, S. (1982). The effect of misclassification in matched-pair case-control studies. *American Journal of Epidemiology*, 116, 402-406.

Greenland, S. (1989). On correcting misclassification in twin studies and other matched pair studies. *Statistics in Medicine*, 8, 825-829.

Greenland, S. and Kleinbaum, D. G. (1983). Correcting for misclassification in two-way tables and matched-pair studies. *International Journal of Epidemiology*, 12, 93-97.

Lee, T-S. (2009a). Bias-adjusted exposure odds ratio for misclassified data. *The Internet Journal of Epidemiology*, 6(2). Accessed from <http://www.ispub.com/journal/the-internet-journal-of-epidemiology/volume-6-number-2/bias-adjusted-exposure-odds-ratio-for-misclassification-data-1.html>.

Lee, T-S. (2009b). Misclassified matched studies with ordinal data. Proceedings of the 6<sup>th</sup> Sino-International Symposium on Probability, Statistics, and Quantitative Management, pp. 48-67, Taipei, Taiwan.

Lee, T-S. (2010). Bias-adjusted McNemar's test for misclassified data. *Journal of Probability & Statistical Science*, 8, 81-95.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Lui, K.-J. (2002). Notes on estimation of the general odds ratio and the general risk difference for paired-sample data. *Biometrical Journal*, 44, 957-968.

McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika*, 64, 449-453.

Stuart, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, 40, 105-110.

Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42, 412-416.

Tallis, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, 18, 342-353.

### Appendix

For  $j = 1, 2, 3$ , let

$$\begin{aligned} a_{Z[j]} &= \varphi_{Z[1;j]}, b_{Z[j]} = \varphi_{Z[2;j]}, c_{\bar{Z}[j]} = \\ \psi_{Z[2;j]}, d_{Z[j]} &= \psi_{Z[3;j]}, \end{aligned} \quad (A1)$$

Where  $\bar{Z} = Y$  if  $Z = X$ , and vice versa.

Because of the symmetry in matched-pair studies, it is reasonable to assume that

$$\begin{aligned} a_{Z[1]} &= a_{\bar{Z}[1]}, a_{Z[2]} = b_{Z[1]} = c_{Z[1]}, b_{Z[2]} = \\ b_{\bar{Z}[2]} &= c_{Z[2]} = c_{\bar{Z}[2]}, b_{Z[3]} = c_{Z[3]} = \\ d_{\bar{Z}[2]}, a_{Z[3]} &= d_{\bar{Z}[1]}, d_{Z[3]} = d_{\bar{Z}[3]}. \end{aligned} \quad (A2)$$

For  $K = 3$ , the matrix  $W$  in equation 9 was given by

$$W_{9 \times 9} = \begin{bmatrix} W_{11} & W_{12} & 0 \\ W_{21} & W_{22} & W_{23} \\ 0 & W_{32} & W_{33} \end{bmatrix}, \quad (A3)$$

where

$$W_{11} = \begin{bmatrix} 1 - 2a_{X[1]} & b_{X[1]} & 0 \\ a_{X[1]} & 1 - 3b_{X[1]} & d_{X[1]} \\ 0 & b_{X[1]} & 1 - 2d_{X[1]} \end{bmatrix},$$

$$W_{12} = \text{diag}[c_{Y[1]}, c_{Y[2]}, c_{Y[3]}], W_{13} = W_{31} = [0], W_{21} = \text{diag}[a_{X[1]}, b_{X[1]}, d_{X[1]}];$$

$$W_{22} = \begin{bmatrix} 1 - 3a_{X[2]} & b_{X[2]} & 0 \\ a_{X[2]} & 1 - 4b_{X[2]} & d_{X[2]} \\ 0 & b_{X[2]} & 1 - 3d_{X[2]} \end{bmatrix},$$

$$\begin{aligned}
 W_{23} &= \text{diag}[d_{Y[1]}, d_{Y[2]}, d_{Y[3]}] = \\
 &\text{diag}[a_{X[3]}, b_{X[3]}, d_{X[3]}], W_{32} = \\
 &\text{diag}[b_{Y[1]}, b_{Y[2]}, b_{Y[3]}] = \\
 &\text{diag}[a_{X[2]}, b_{X[2]}, d_{X[2]}]; \\
 W_{33} &= \begin{bmatrix} 1 - 2a_{X[3]} & b_{X[3]} & 0 \\ a_{X[3]} & 1 - 3b_{X[3]} & d_{X[3]} \\ 0 & b_{X[3]} & 1 - 2d_{X[3]} \end{bmatrix},
 \end{aligned}$$

where  $\text{diag}[d_{11}, d_{22}, d_{33}]$  denotes a  $3 \times 3$  diagonal matrix.

Solving the matrix equation of  $W_{9 \times 9} \cdot V_{9 \times 9} = I_{9 \times 9}$ , where  $V_{9 \times 9}$  was given by equation 10(b) and  $I_{9 \times 9}$  was a  $9 \times 9$  identity matrix, results in

$$\begin{aligned}
 V_{31} &= \Delta_2^{-1} W_{32} W_{21}, \\
 V_{21} &= -\Delta_1^{-1} (W_{11} W_{23} V_{31} + W_{21}), \\
 V_{11} &= W_{11}^{-1} (I_{3 \times 3} - W_{12} V_{21}), \\
 V_{32} &= -\Delta_2^{-1} W_{32} W_{11}, \\
 V_{22} &= \Delta_1^{-1} W_{11} (I_{3 \times 3} - W_{23} V_{32}), \quad (\text{A4}) \\
 V_{12} &= -W_{11}^{-1} W_{12} V_{22}, \\
 V_{33} &= \Delta_2^{-1} \Delta_1, \\
 V_{23} &= -\Delta_1^{-1} W_{11} W_{23} V_{33}, \\
 V_{13} &= -W_{11}^{-1} W_{12} V_{23}.
 \end{aligned}$$

where  $\Delta_1$  and  $\Delta_2$  were defined, respectively, by

$$\Delta_1 = W_{11} W_{22} - W_{21} W_{12},$$

and

$$\Delta_2 = \Delta_1 W_{33} - W_{32} W_{11} W_{23}.$$

If  $\{\delta_{ij}^{[1]}\}$  and  $\{\delta_{ij}^{[2]}\}$  denote the  $(i, j)^{\text{th}}$  entry of  $\Delta_1$  and  $\Delta_2$ , then

$$\delta_{11}^{[1]} = 1 - 2a_{X[1]} - 3a_{X[2]} + 5a_{X[1]}a_{X[2]} + a_{X[2]}b_{X[1]},$$

$$\delta_{12}^{[1]} = b_{X[2]}(1 - 2a_{X[1]}) + b_{X[1]}(1 - 4b_{X[2]}),$$

$$\delta_{13}^{[1]} = b_{X[1]}d_{X[2]},$$

$$\delta_{21}^{[1]} = a_{X[1]}(1 - 3a_{X[2]}) + a_{X[2]}(1 - 3b_{X[1]}),$$

$$\delta_{22}^{[1]} = a_{X[1]}b_{X[2]} + 1 - 3b_{X[1]} - 4b_{X[2]} + 11b_{X[1]}b_{X[2]} + d_{X[1]}b_{X[2]}, \quad (\text{A5})$$

$$\delta_{23}^{[1]} = b_{X[1]}(1 - 4b_{X[2]}) + b_{X[2]}(1 - 2d_{X[1]});$$

$$\delta_{31}^{[1]} = b_{X[1]}a_{X[2]},$$

$$\delta_{32}^{[1]} = b_{X[1]}(1 - 4b_{X[2]}) + b_{X[2]}(1 - 2d_{X[1]}),$$

$$\delta_{33}^{[1]} = b_{X[1]}d_{X[2]} + 1 - 2d_{X[1]} - 3d_{X[2]} + 5d_{X[1]}d_{X[2]};$$

$$\delta_{11}^{[2]} = (1 - 2a_{X[3]})\delta_{11}^{[1]} + a_{X[3]}\delta_{12}^{[1]} - a_{X[2]}a_{X[3]}(1 - 2a_{X[1]}),$$

$$\delta_{12}^{[2]} = b_{X[3]}\delta_{11}^{[1]} + (1 - 3b_{X[3]})\delta_{12}^{[1]} + b_{X[3]}\delta_{13}^{[1]} - b_{X[1]},$$

$$\delta_{13}^{[2]} = d_{X[3]}\delta_{12}^{[1]} + (1 - 2d_{X[3]})\delta_{13}^{[1]},$$

$$\delta_{21}^{[2]} = (1 - 2a_{X[3]})\delta_{21}^{[1]} + a_{X[3]}\delta_{22}^{[1]} - a_{X[1]},$$

$$\delta_{22}^{[2]} = b_{X[3]}\delta_{21}^{[1]} + (1 - 3b_{X[3]})\delta_{22}^{[1]} + b_{X[3]}\delta_{23}^{[1]} - b_{X[2]}b_{X[3]}(1 - 3b_{X[1]}), \quad (\text{A.6})$$

$$\delta_{23}^{[2]} = d_{X[3]}\delta_{22}^{[1]} + (1 - 2d_{X[3]})\delta_{23}^{[1]} - d_{X[1]};$$

$$\delta_{31}^{[2]} = (1 - 2a_{X[3]})\delta_{31}^{[1]} + a_{X[3]}\delta_{32}^{[1]},$$

$$\delta_{32}^{[2]} = b_{X[3]}\delta_{31}^{[1]} + (1 - 3b_{X[3]})\delta_{32}^{[1]} + b_{X[3]}\delta_{33}^{[1]} - b_{X[1]},$$

$$\delta_{33}^{[2]} = d_{X[3]}\delta_{32}^{[1]} + (1 - 2d_{X[3]})\delta_{33}^{[1]} - d_{X[2]}d_{X[3]}(1 - 2d_{X[1]}).$$



## MATCHED-PAIR STUDIES WITH MISCLASSIFIED ORDINAL DATA

The nine equations in A4 were solved by grouping them into three sets. First, the first three equations were solved together in A4 letting the entries for  $V_{31}$ ,  $V_{21}$ , and  $V_{11}$  be denoted, respectively, by  $\{a_{ij}\}$ ,  $\{s_{ij}\}$ , and  $\{x_{ij}\}$ , namely,  $V_{31} = [a_{ij}]$ ,  $V_{21} = [s_{ij}]$ , and  $V_{11} = [x_{ij}]$ . Next the second set of three equations in A4 were solved for  $V_{32}$ ,  $V_{22}$ , and  $V_{12}$ ; the entries of these matrices are given, respectively, by  $V_{32} = [b_{ij}]$ ,  $V_{22} = [t_{ij}]$  and  $V_{12} = [y_{ij}]$ . Finally, after solving equations A4 for  $V_{33}$ ,  $V_{23}$ , and  $V_{13}$ , the entries of these matrices are given, respectively, by  $V_{33} = [c_{ij}]$ ,  $V_{23} = [u_{ij}]$  and  $V_{13} = [z_{ij}]$ .

Putting together the above result, the inverse of the misclassification matrix  $W$  was thus obtained as follows:

$$V_{9 \times 9} = [v_{ij}]$$

$$= \begin{bmatrix} x_{11} & x_{12} & x_{13} & y_{11} & y_{12} & y_{13} & z_{11} & z_{12} & z_{13} \\ x_{21} & x_{22} & x_{23} & y_{21} & y_{22} & y_{23} & z_{21} & z_{22} & z_{23} \\ x_{31} & x_{32} & x_{33} & y_{31} & y_{32} & y_{33} & z_{31} & z_{32} & z_{33} \\ s_{11} & s_{12} & s_{13} & t_{11} & t_{12} & t_{13} & u_{11} & u_{12} & u_{13} \\ s_{21} & s_{22} & s_{23} & t_{21} & t_{22} & t_{23} & u_{21} & u_{22} & u_{23} \\ s_{31} & s_{32} & s_{33} & t_{31} & t_{32} & t_{33} & u_{31} & u_{32} & u_{33} \\ a_{11} & a_{12} & a_{13} & b_{11} & b_{12} & b_{13} & c_{11} & c_{12} & c_{13} \\ a_{21} & a_{22} & a_{23} & b_{21} & b_{22} & b_{23} & c_{21} & c_{22} & c_{23} \\ a_{31} & a_{32} & a_{33} & b_{31} & b_{32} & b_{33} & c_{31} & c_{32} & c_{33} \end{bmatrix}.$$

where the closed-form solutions for all the entries  $\{v_{ij}\}$  can be found in the appendix of Lee (2009b).

## A Robust Root Mean Square Standardized Effect Size in One-Way Fixed-Effects ANOVA

Guili Zhang  
East Carolina University

James Algina  
University of Florida

---

A robust Root Mean Square Standardized Effect Size (RMSSE<sub>R</sub>) was developed to address the unsatisfactory performance of the Root Mean Square Standardized Effect Size. The coverage performances of the confidence intervals (CI) for RMSSE<sub>R</sub> were investigated. The coverage probabilities of the non-central  $F$  distribution-based CI for RMSSE<sub>R</sub> were adequate.

Key words: Confidence interval, effect size, root mean square standardized effect size, non-central  $F$  distribution-based confidence interval, percentile bootstrap, coverage probability, robust root mean square standardized effect size.

---

### Introduction

Using an effect size (ES) in addition to or in place of a hypothesis test has been enthusiastically advocated by many statistical methodologists because ESs are regarded as more appropriate and more informative (Cohen, 1965, 1994; Cumming & Finch, 2005; Finch, et al., 2002; Hays, 1963; Meehl, 1967; Nickerson, 2000; Steiger, 2004; Steiger & Fouladi, 1997; Zhang, 2009; Zhang & Algina, 2008). Reporting an ES has become mandatory or strongly recommended in some editorial policies in the last two decades (Murphy, 1997; Thompson, 1994). *The Publication Manual of the American Psychological Association* (2001) stated that it is almost always necessary to include some index of ES or strength of relationship in the results section of a research report.

The APA Task Force on Statistical Inference (Wilkinson and the Task Force on Statistical Inference, 1999) not only supports the use of ESs but also requires researchers to provide confidence intervals (CI) for all principal outcomes. A CI for an ES is recommended as a superior replacement for significance testing because it is argued that CI contains all the information found in the significance tests and vital information not provided by the significance tests about the magnitude of effects and precision of estimates (Cohen, 1994; Steiger & Fouladi, 1997; Wilkinson, et al., 1999; Cumming & Finch, 2001, 2005; Zhang, 2009).

The increased interests in ES and CI have motivated explorations of their usefulness and effectiveness within recent years (Algina & Keselman, 2003a, 2003b; Bird, 2002; Cumming & Fitch, 2001; Zhang & Algina, 2008). In the two group case, it has been reported that - in both the independent and dependent samples cases - CIs for Cohen's  $\delta$ , arguably the most widely accepted ES index for a pairwise contrast on means, may be misleading due to poor coverage probability when data are nonnormal and can grossly misrepresent the degree to which two distributions differ (Algina & Keselman, 2003b; Algina, et al., 2006; Algina, et al., 2005a; Kelly, 2005; Wilcox & Keselman, 2003). However, research has shown that the CIs for  $\delta_r$ , a robust version of  $\delta$  based on trimmed

---

Guili Zhang is an Assistant Professor of Research and Evaluation Methodology in the Department of Curriculum and Instruction, East Carolina University. Her research interests are in applied statistics. Email her at: zhangg@ecu.edu. James Algina is a Professor of Research and Evaluation Methodology at the University of Florida. His research interests are in psychometric theory and applied statistics. Email him at: algina@ufl.edu.

means and Winsorized variances, have better coverage probability than do CIs for Cohen’s  $\delta$  under data nonnormality (Algina & Keselman, 2003b).

In the more than two group case, Zhang and Algina (2008) investigated the coverage performance of the noncentral  $F$  distribution-based CI and the percentile bootstrap CI for one of the most commonly used generalized effect size indices, the Root Mean Square Standardized Effect Size (RMSSE), proposed by Steiger and Fouladi (1997), denoted by

$$f^* = \sqrt{\frac{\sum_{j=1}^J (\mu_j - \mu)^2}{(J-1)\sigma^2}}$$

in a one-way, fixed-effects, between-subjects ANOVA. Both CIs were implemented for all combinations of the following five factors: (1) five population distributions including the normal distribution and four additional cases from the family of the  $g$  and  $h$  distributions that are nonnormal (Hoaglin, 1983, Martinez & Iglewicz, 1984); (2) two numbers of levels for treatment groups:  $J = 3$  and  $J = 6$ ; (3) three cell sample sizes in each treatment; (4) six values of population RMSSEs; and (5) two mean configurations, the equally spaced mean configuration and the one extreme mean configuration. Each condition was replicated 2,500 times and the number of bootstrap replications in the bootstrap procedure was 1,000. Zhang and Algina found that both the noncentral  $F$  distribution-based CI and the percentile bootstrap CI for RMSSE yielded inadequate coverage probabilities under data nonnormality.

According to arguments in Wilcox and Keselman (2003) about the robustness of  $\delta$  in the two-group case, it is not surprising that  $f^*$  is not an entirely adequate measure of group separation because  $f^*$  is formulated with least-square parameters which are affected by skewed data, long tails and/or outlying values. It is therefore imperative to develop a robust version of the RMSSE to ensure the appropriate and effective use of the ES in ANOVA.

### Methodology

The unsatisfactory coverage performance of the CIs for  $f^*$  reported by Zhang and Algina (2008) is understandable: This is because the problems that trouble Cohen’s  $\delta$  and its CI are very likely to also haunt  $f^*$  and its CI, as  $f^*$  is a generalized  $\delta$  and is formulated with the nonrobust least-square means and variances. It is well known that when the distribution of the data is not normal, the least-square means and standard deviations can work poorly because they are affected by the skewness of the data and by the outliers in the data; consequently  $f^*$  may be misleading as a measure of population separation. Therefore, a robust version of  $f^*$  that is parallel to  $\delta_R$ , the robust effect size in the two-group case, is strongly desired. The purposes of the study are:

- a. To develop a robust RMSSE,  $f_R^*$ ,
- b. To develop a noncentral  $F$  distribution-based CI for  $f_R^*$ , and
- c. To investigate the performance of the noncentral  $F$  distribution-based and percentile bootstrap CI for  $f_R^*$ .

Note that  $f^*$  and  $f_R^*$  are two different parameters based on different measures of location and variability and, unless the data are normally distributed,  $f^*$  and  $f_R^*$  will not be equal. The parameter  $f^*$  is used to characterize the amount of difference among the population means, while  $f_R^*$  represents the amount of difference among the population trimmed means.

### Robust Root Mean Square Standardized Effect Size and Its Confidence Interval

To overcome the weaknesses in  $f^*$ , a robust version of the generalized effect size was developed, the Robust Root Mean Square Standardized Effect Size (RMSSE<sub>R</sub>), denoted by  $f_R^*$  in this study. The value of  $f_R^*$  is defined by using robust parameters (20% trimmed means

and Winsorized variances) as opposed to the least-square parameters (means and variances). Trimmed means are used because it has been shown that the impact of outliers on trimmed means can be much less disturbing than on the usual means (Wilcox, 2005). The Winsorized variance is used because the sample Winsorized variance is used in hypothesis testing based on trimmed means. Both the trimmed mean and the Winsorized variance are robust parameters as judged by the criteria of qualitative robustness, quantitative robustness and infinitesimal robustness (Wilcox, 2005, Section 2.1 describes these criteria).

In a balanced one-way between-subjects ANOVA design,  $f_R^*$  is defined as

$$f_R^* = .642 \sqrt{\frac{\sum_{j=1}^J (\mu_{Tj} - \mu_T)^2}{(J-1)\sigma_w^2}}, \quad (1)$$

where  $\mu_{Tj}$  is the trimmed mean for the  $j^{\text{th}}$  level,  $\mu_T$  is the grand mean based on the trimmed means, and  $\sigma_w^2$  is the within-level Winsorized variance, which is assumed to be constant across levels. The quantity 0.642 is the square root of the population Winsorized variance for a standard normal distribution, therefore, including 0.642 in the definition of the robust effect ensures that  $f_R^* = f^*$  when the data are drawn from normal distributions with equal variances.

An estimate of  $f_R^*$  can be attained from sample statistics by applying the following formula:

$$\hat{f}_R^* = .642 \times \sqrt{\frac{\sum_{j=1}^J (\bar{Y}_{Tj} - \bar{Y}_T)^2}{(J-1)S_{Wp}^2}}, \quad (2)$$

where  $\bar{Y}_{Tj}$  is the trimmed sample mean for the  $j^{\text{th}}$  level,  $\bar{Y}_T$  is the sample grand trimmed mean,

and  $S_{Wp}^2$  is the sample pooled within-level Winsorized variance.

The quantity  $S_{Wp}^2$  is obtained by using

$$S_{Wp}^2 = \frac{\sum_{j=1}^J (n_j - 1)S_{Wj}^2}{\sum_{j=1}^J n_j - J}. \quad (3)$$

A CI for  $f_R^*$  can be constructed based on the noncentral  $F$  distribution. Consider a one-way, between-subjects, fixed-effects ANOVA with  $n_j$  observations in the  $j^{\text{th}}$  group and  $J$  groups. The robust  $F$  statistic is calculated by using (Yuen, 1974)

$$F_R = \frac{MS_{RB}}{MS_{RW}}, \quad (4)$$

where  $MS_{RB}$  and  $MS_{RW}$  are the robust mean square between and robust mean square within respectively, and are calculated by using:

$$MS_{RB} = \frac{\sum_{j=1}^J h_j (\bar{Y}_{Tj} - \bar{Y}_T)^2}{J-1} \quad (5)$$

and

$$MS_{RW} = \frac{\sum_{j=1}^J \sum_{i=1}^{n_j} (Y_{Wij} - \bar{Y}_{Wj})^2}{\sum_{j=1}^J h_j - J}, \quad (6)$$

where  $Y_{Wij}$  is the  $i^{\text{th}}$  Winsorized score in group  $j$ , and  $\bar{Y}_{Wj}$  is the Winsorized mean for group  $j$ . The robust  $F$  statistic has robust noncentrality parameter

$$\lambda_R = \frac{\sum_{j=1}^J h_j (\mu_{Tj} - \mu_T)^2}{\tilde{\sigma}_w^2}, \quad (7)$$

where  $\tilde{\sigma}_W^2$  is an adjusted version of the population Winsorized variance:

$$\tilde{\sigma}_W^2 = \frac{N - J}{\sum_{j=1}^J h_j - J} \sigma_W^2. \quad (8)$$

The lower limit of the 95% CI for  $\lambda_R$  is the robust noncentrality parameter for the noncentral  $F$  distribution in which the calculated robust  $F$  statistic is the 0.975 quantile. The upper limit of the 95% confidence interval for  $\lambda_R$  is the robust noncentrality parameter for the noncentral  $F$  distribution in which the calculated robust  $F$  statistic is the 0.025 quantile of the distribution.

In a balanced one-factor between-subject design with equal  $n_s$ ,  $f_R^*$  can be written as a function of  $\lambda_R$ :

$$f_R^* = \sqrt{\frac{.4129 \times \left( \sum_{j=1}^J n_j - J \right)}{\left( \sum_{j=1}^J h_j - J \right) \times (J - 1) h}} \lambda_R. \quad (9)$$

To find a  $(1 - \alpha)\%$  (95% in this study) CI for  $f_R^*$ , the noncentral  $F$  distribution was first used to find a 95% CI for  $\lambda_R$ . After the CI on  $\lambda_R$  is found, equation 9 is applied to transform the endpoints of the CI for  $\lambda_R$  to obtain the endpoints for the CI for  $f_R^*$ .

Although the noncentral  $F$  distribution can be used to obtain a CI for  $f_R^*$ , because this CI construction method is based on the assumption that the data are drawn from a normal distribution, when the data are nonnormal the coverage probability for this interval may be poor and the percentile bootstrap CI may have better coverage probability (Algina & Keselman, 2003b; Efron & Tibshirani, 1993). Therefore, the performances of the percentile bootstrap method for the construction of CIs for  $f_R^*$  were examined and compared to the

noncentral  $F$  distribution-based method in terms of the probability coverage and interval width.

#### Coverage Performance of the Confidence Interval for Robust Root Mean Square Standardized Effect Size

To investigate the coverage performance of the CIs for  $f_R^*$ , the noncentral  $F$  distribution-based and the percentile bootstrap CIs were implemented for all combinations of the following five factors: (1) five population distributions including the normal distribution and four additional cases from the family of the  $g$  and  $h$  distributions that are nonnormal (Hoaglin, 1983, Martinez & Iglewicz, 1984); (2) two numbers of levels for treatment groups:  $J = 3$  and  $J = 6$ ; (3) three cell sample sizes in each treatment; (4) six values of population RMSSE<sub>R</sub>; (5) two mean configurations, the equally spaced mean configuration and the one extreme mean configuration. The nominal confidence level for all intervals investigated was 0.95 and each condition was replicated 2,500 times. The number of bootstrap replications in the bootstrap procedure was 1,000.

#### Conditions

Data for all five distributions were generated from the  $g$  and  $h$  distributions: (1)  $g = h = 0$ , the standard normal distribution ( $\gamma_1 = \gamma_2 = 0$ ), where  $\gamma_1 = \sqrt{\beta_1}$  and is the skewness, and  $\gamma_2 = \beta_2$  and is the kurtosis, (2)  $g = .76$  and  $h = -.098$ , a distribution with the skewness and kurtosis of an exponential distribution ( $\gamma_1 = 2$ ,  $\gamma_2 = 6$ ), (3)  $g = 0$  and  $h = .225$  ( $\gamma_1 = 0$  and  $\gamma_2 = 154.84$ ), (4)  $g = h = .225$  ( $\gamma_1 = 4.90$  and  $\gamma_2 = 4673.80$ ), and (5)  $g = 0$  and  $h = .109$  ( $\gamma_1 = 0$  and  $\gamma_2 = 6$ ), a distribution with the skewness and kurtosis of a double exponential distribution.

The four nonnormal distributions cover a wide range of nonnormality including distributions that are strongly nonnormal. Such a selection of distributions allows the researcher to investigate the performances of the CIs under a wide range of the data conditions. The goal is to find which procedure or procedures are likely to

work well over a wide range of distributions because it is impossible for any one of the simulations to include every possible distribution that might be encountered in real data or to anticipate what types of distributions are realistic in all of social and behavioral science fields. The inclusion of the normal distribution provides a reference for judgments on the CIs' performance under data that deviate from normality.

The numbers of treatment groups investigated were 3 and 6 ( $J = 3$  and  $J = 6$ ), and sample sizes in each treatment included were  $n_j = 20$  to 50 in steps of 15. In other words, the treatment groups have equal sample size and the sample sizes investigated were 20, 35 and 50. The number of treatment groups equal to 3 and 6 was selected because this covers the likely range encountered in most research in the social and behavioral sciences. Sample sizes ranging from 20 to 50 are fairly typical of sample sizes used in social science research, although clearly do not cover sample sizes found in very small or very large studies.

The treatment group means followed two mean configurations: the equally spaced mean configuration and the one extreme mean configuration. A mean configuration is a specification of the arrangement of the treatment groups means. Denoting the smallest and the largest means by  $\mu_{min}$  and  $\mu_{max}$ , if the means other than  $\mu_{min}$  and  $\mu_{max}$  are equally spaced between these two extremes, the configuration is referred to as an equally spaced configuration (Cohen, 1969). If one of the means is equal to  $\mu_{min}$  and the rest of the means are all equal to  $\mu_{max}$ , or, if one of the means is equal to  $\mu_{max}$  and the rest of the means are equal to  $\mu_{min}$ , then the configuration is called a one extreme mean configuration. Mean configurations are an artifice adopted because the actual configuration of means in social science research is quite variable. Nevertheless, the selected configurations cover a range of possibilities and will allow determination of whether results tend to generalize over configurations.

Six values of  $f_R^*$  were investigated: 0, 0.1, 0.25, 0.40, 0.55 and 0.70. Defining

$$\delta_{max} = \frac{\mu_{max} - \mu_{min}}{\sigma} \quad (10)$$

as Cohen's effect size for the largest and smallest means, under the equally spaced mean configurations, these population  $f_R^*$  values approximately correspond to  $\delta_{max}$  of 0, 0.2, 0.5, 0.8, 1.10 and 1.40, respectively. Under the one extreme mean configuration, these population  $f_R^*$  values roughly correspond to  $\delta_{max}$  of 0, 0.173, 0.433, 0.693, 0.952, and 1.212. Therefore, a  $f_R^*$  of 0 indicates no effect, .1 a small effect, 0.25 a medium effect, 0.40 a large effect, and 0.55 and 0.70 very large effects.

The nominal confidence level for all intervals investigated was .95 and each condition was replicated 2,500 times, assuring sufficient precision for an adequate initial investigation into the sampling behaviors of the CIs. The number of bootstrap replications in the bootstrap procedure was 1,000.

#### Analyses Conducted

The study was designed to investigate the robustness of the noncentral  $F$  distribution-based CIs and the percentile bootstrap CIs for  $f_R^*$  to sampling from nonnormal distributions. Coverage probabilities for the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  were estimated. Additionally, the average width of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  were compared.

Variables conforming to a  $g$  and  $h$  distributions are transformations of a standard normal distribution. When  $g$  and  $h$  are both nonzero,

$$Y = \frac{\exp(gZ) - 1}{g} \exp\left(\frac{hZ^2}{2}\right) \quad (11)$$

where  $Z$  is a standard normal variable, and  $Y$  is the  $g$  and  $h$  distributed variable. When  $g$  is zero,

$$Y = Z \exp\left(\frac{hZ^2}{2}\right). \quad (12)$$

Standard normal variables ( $Z_{ij}$ ) were generated by using RANNOR function in SAS (SAS, 1999). Then the  $Z_{ij}$  were converted to the desired  $g$  and  $h$  distributed random variable by using Equations 11 and 12. To create scores corresponding to the selected values of  $f_R^*$ , it is necessary to linearly transform the  $g$  and  $h$  distributed variables. Data were generated for three samples and six samples in each replication of each condition by the following steps: First, for the first sample  $n_1$  scores were generated from the appropriate distribution. Secondly,  $n_2$  scores from the same distribution were generated and a constant was added to each score. Thirdly,  $n_3$  scores from the same distribution were generated and a constant was added to each score and so forth until  $n_j$  scores from the same distribution were generated and a constant was added to each score. The constants were chosen such that the population  $\text{RMSSE}_R$ ,  $f_R^*$  would equal the following values: 0, 0.1, 0.25, 0.40, 0.55, and 0.70.

For the equally spaced mean configuration, the  $Y$  variables were obtained by using

$$Y_{ij} = X_{ij} + (j-1) \sqrt{\frac{12}{J(J+1)}} f_R^* \frac{\sigma_W}{.642}, \quad j = 1, \dots, J. \quad (13)$$

For the configuration with one extreme mean,  $Y_{ij} = X_{ij}$  for groups  $j = 1, \dots, J-1$ . For group  $J$  the transformation was

$$Y_{ij} = X_{ij} + \sqrt{J} f_R^* \frac{\sigma_W}{.642}. \quad (14)$$

To find a  $(1-\alpha)\%$  (95% in the current study) confidence interval for  $f_R^*$ , the noncentral  $F$  distribution is first used to obtain a 95% confidence interval on  $\lambda_R$ , the robust noncentrality parameter of the  $F$  distribution. Once the CI for  $\lambda_R$  is found, the endpoints of the CI for  $\lambda_R$  are transformed to endpoints for

$f_R^*$  by applying Equation 9. Notice the CI for  $f_R^*$  constructed by the noncentral  $F$  distribution-based method will result in coverage probability of 0.975 when  $f_R^* = 0$  because the probability noncoverage from the lower side of the distribution will be 0 instead of 0.025.

To apply the percentile bootstrap method, the following steps are completed 1,000 times within each replication of a condition.

1. A sample of size  $n_j$  is randomly selected with replacement from the scores for the group  $j$ ,  $j = 1, \dots, J$ . These  $J$  samples are combined to form a bootstrap sample.
2. The parameter  $f_R^{*2}$  is estimated by using

$$\hat{f}_R^{*2} = \frac{.642^2}{n} \frac{\sum_{j=1}^J n_j - J}{\sum_{j=1}^J h_j - J} (F_R - 1). \quad (15)$$

3. The 1,000  $f_R^{*2}$  estimates are ranked from low to high. The lower limit of the CI for  $f_R^{*2}$  is determined by finding the 26<sup>th</sup> estimate in the rank order [i.e., the  $(0.025 \times 1,000 + 1)$ <sup>th</sup> estimate]; and the 975<sup>th</sup> estimate is the upper limit of the CI for  $f_R^{*2}$  (i.e. the  $(0.975 \times 1,000)$ <sup>th</sup> estimate].
4. The lower limit of the CI for  $f_R^*$  is equal to the square root of the lower limit of the CI for  $f_R^{*2}$  if the latter lower limit is larger than zero and is zero otherwise. The upper limit of the CI for  $f_R^*$  is equal to the square root of the upper limit of the CI for  $f_R^{*2}$ .

### Results

The estimated coverage probabilities of the noncentral  $F$  distribution-based CIs for  $f_R^*$  are reported in Tables 1-4. The average widths of the noncentral  $F$  distribution-based CIs for  $f_R^*$  are shown in Tables 5-8.

Estimated Coverage Probabilities of Confidence Intervals for  $f_R^*$

In Tables 1 to 4, the estimated coverage probabilities of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  are presented with estimates outside the [.94, .96] interval bolded, and estimates outside of the interval [.925, .975] bolded and underlined.

The pattern of results for the noncentral  $F$  distribution-based CI for  $f_R^*$  looks strikingly similar across Tables 1 to 4. First, when sampling from a normal distribution, the coverage probability of the noncentral  $F$  distribution-based CIs should be 0.975 when  $f_R^* = 0$ , and the results in Tables 1 to 4 are consistent with the theory. When  $f_R^* > 0$ , the coverage probability of the noncentral  $F$  distribution-based CI is expected to be 0.95 under normality and the results presented in Tables 1-4 are consistent with this expectation.

Second, considering the results in all four tables, coverage probability for the noncentral  $F$  distribution-based CI for  $f_R^*$  tends to be appreciably better than for the bootstrap CI both when sampling from normal and nonnormal distributions. When sampling from the normal distribution, when  $J = 3$  the coverage probability for the noncentral  $F$  distribution-based CI is outside the [.925, .975] interval in only 1 case out of a total of 36, while the bootstrap CI has a total of 20 cases outside this interval. Under normality, when  $J = 6$ , the noncentral  $F$  distribution-based CI coverage probabilities are outside [.925, .975] in 2 out of 36 cases, while the bootstrap CI coverage probabilities are outside this interval in 6 out of 36 cases.

For the nonnormal distributions, the noncentral  $F$  distribution-based CI for  $f_R^*$  has noticeably fewer coverage probabilities that are outside the criterion intervals than does the bootstrap CI under each of the four distribution conditions. The number of cases that are outside the [.925, .975] criterion interval, out of a total of 72 cases under each nonnormal distribution for the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$ , are: 7 versus 31 for the  $g$

$= 0$  and  $h = 0.109$  distribution; 7 versus 40 for the  $g = 0$  and  $h = 0.225$  distribution; 20 versus 38 for the  $g = 0.760$  and  $h = -0.098$  distribution; and 6 versus 41 for the  $g = 0.225$  and  $h = 0.225$  distribution.

Third, the performance of the noncentral  $F$  distribution-based CI under the four nonnormal distributions reveals some common characteristics across Table 1 to Table 4. When  $f_R^* = 0$ , roughly 50% of the coverage probabilities tend to be outside [.925, .975]. Of the coverage probabilities that are inside the interval, most are for  $J = 3$  when the data are sampled from either the  $g = 0$  and  $h = 0.225$  distribution or the  $g = 0.225$  and  $h = 0.225$  distribution.

The coverage probabilities of the noncentral  $F$  distribution-based CI for  $f_R^*$  are all inside either [.925, .975] or both intervals when  $f_R^*$  is 0.10, 0.25 or 0.40. The coverage probabilities of noncentral  $F$  distribution-based CI for  $f_R^*$  are also all inside either the [.925, .975] interval or both intervals when  $f_R^*$  is 0.55 except when  $n = 35$  and the data are sampled from the  $g = 0.760$  and  $h = -0.098$  distribution with the means following the equally spaced mean configuration. Even when  $f_R^* = 0.70$ , the coverage probabilities still tend to be inside the [.925, .975] interval. The exceptions occur mostly for the  $g = 0.760$  and  $h = -0.098$  distribution in combination with the equally spaced mean configuration. Other exceptions involve the  $g = 0$  and  $h = 0.225$  distribution when  $n = 35$ ,  $J = 6$ , and the  $g = 0.225$  and  $h = 0.225$  distribution when  $n = 35$  with the group means following the equally spaced mean configuration.

Overall, under all data distributions, the coverage probabilities of the noncentral  $F$  distribution-based CI for  $f_R^*$  are adequate by the [.925, .975] criterion except for some cases of  $f_R^* = 0$  and a few cases when  $f_R^* = 0.70$ . When  $f_R^* = 0$  the probability coverage of the noncentral  $F$  distribution-based CI for  $f_R^*$  tends to exceed 0.975, and when  $f_R^* = 0.70$  the



probability coverage of the noncentral  $F$  distribution-based CI for  $f_R^*$  tends to go below 0.925. It is observed that, excluding  $f_R^* = 0$ , the coverage performance of the noncentral  $F$  distribution-based CIs for  $f_R^*$  becomes less satisfactory when  $f_R^*$  gets larger.

The results of the bootstrap CIs for  $f_R^*$  are also presented in Tables 1 to 4. When sampling from normal distributions, when  $f_R^* = 0$  and  $J = 3$ , the coverage probabilities of the bootstrap CI for  $f_R^*$  are all above 0.975, but when  $f_R^* = 0$  and  $J = 6$  they are outside the [.94, .96] interval only when  $n = 20$ . Under normality, when  $f_R^* = 0.10$  or 0.25, the coverage probabilities of the bootstrap CI for  $f_R^*$  are all outside the [.925, .975] criterion interval when  $J = 3$ , but all inside the [.94, .96] interval when  $J = 6$  except when  $f_R^* = 0.10$  and  $n = 20$ . When  $f_R^* \geq 0.40$ , coverage probabilities tend to be inside [.925, .975] for both levels of  $J$ , except when  $f_R^* = 0.40$  and  $n = 20$  for  $J = 3$ , when  $f_R^* = 0.70$ ,  $n = 20$  for  $J = 6$  and the equally spaced mean configuration, and when  $f_R^* = 0.55$ ,  $n = 20$  for  $J = 6$  and the one extreme configuration.

Under the four nonnormal distributions, when  $f_R^* = 0$ , the coverage probability of the bootstrap CI for  $f_R^*$  tends to be outside the [.925, .975] criterion interval when  $J = 3$ . Roughly 50% are inside [.925, .975] when  $J = 6$ , mostly associated with larger sample sizes. When  $f_R^* = 0.10$  or 0.25, the coverage probability of bootstrap CI for  $f_R^*$  tends to be outside the [.925, .975] criterion interval when  $J = 3$ , and inside the [.925, .975] criterion interval when  $J = 6$  except when sample size is small for some data distributions. For example, for the  $g = 0$  and  $h = 0.109$  distribution, when  $J = 6$  and  $f_R^* = 0.10$  or 0.25, the coverage probabilities of the bootstrap CI are all within [.925, .975] except when  $n = 35$  and the mean configuration is the one extreme mean configuration. For the other three nonnormal distributions, the

coverage probabilities are outside the [.925, .975] interval mostly when  $n = 20$  and  $J = 3$ .

The coverage probability tends to be inside either the [.925, .975] interval or both intervals in most conditions when  $f_R^* \geq 0.40$ , except when  $n = 20$  and a few cases when  $n = 35$ . The inadequate coverage probabilities under  $n = 35$  mostly occur in the conditions with  $J = 6$ . Overall, the performance of the coverage probability of the bootstrap CI for  $f_R^*$  is much less adequate than is the performance of the noncentral  $F$  distribution-based CIs for  $f_R^*$ . Typically the coverage probability of the bootstrap CI is too high.

#### Average Widths of Confidence Intervals for $f_R^*$

The average widths of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  under  $J = 3$  and the equally spaced mean configuration are presented in Table 5. It is observed that, generally, the average widths of the noncentral  $F$  distribution-based CIs for  $f_R^*$  are shorter than those of the bootstrap CIs for  $f_R^*$ . The difference between the widths of the two kinds of CIs has a tendency to become smaller when sample size gets larger. For both the noncentral  $F$  distribution-based and the bootstrap CIs for  $f_R^*$ , the average width of the CIs gets narrower as the sample size increases and as the population effect size  $f_R^*$  decreases.

Across distributions, there is only a very trivial difference in the width of the noncentral  $F$  distribution-based CIs for  $f_R^*$ . Similar to the pattern in the widths of the bootstrap CIs for  $f_R^*$  observed and reported by Zhang and Algina (2008), the widths of the bootstrap CIs for  $f_R^*$  fluctuate very little across data distribution conditions.

Presented in Table 6, the average widths of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  under  $J = 3$  and the one extreme mean configuration shows little difference from those from the widths for the equally spaced mean configuration in Table 5.

This suggests that the type of mean configuration does not affect the precision of estimation for  $f_R^*$ .

Table 7 shows the average widths of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  under  $J = 6$  and the equally spaced mean configuration. It is fairly apparent that, when  $J$  increases from 3 to 6, the intervals become narrower. This is observed for all combinations of conditions. It is also observed that, generally, the average widths of the noncentral  $F$  distribution-based CIs for  $f_R^*$  are shorter than those of the bootstrap CIs for  $f_R^*$ . This difference is consistent across all combinations of conditions. Furthermore, for both the noncentral  $F$  distribution-based and the bootstrap CIs for  $f_R^*$ , the average width of the CIs gets narrower as the sample size increases and the population effect size  $f_R^*$  decreases. Across distributions, there is very little difference in the widths of the noncentral  $F$  distribution-based CIs, and the widths of the bootstrap CIs for  $f_R^*$  also remain quite constant across data distribution conditions.

The average widths of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  under  $J = 6$  and the one extreme mean configuration are presented in Table 8. Again there is little difference between these widths and the widths in the equally spaced mean configuration in Table 7, in terms of values as well as patterns observed. This suggests that the type of mean configuration does not strongly affect the estimation accuracy for  $f_R^*$ .

### Conclusion

Confidence intervals for effect size have been strongly advocated by statistical methodologists to be used as a useful supplement to and maybe even a superior replacement for the traditional hypothesis testing. Despite the increasing need for using CIs, much remains to be known about the robustness of the CIs in order to ensure their proper usage. Investigation and evaluation of the performance of the CIs and their robustness under various conditions are urgently needed.

In the two-group case, it has been reported that in both the independent samples and dependent samples case CIs for Cohen's  $\delta$  may be misleading because of poor coverage probability when data are nonnormal (Algina & Keselman, 2003b; Algina, et al., 2005a, Algina, et al., 2006; Kelly, 2005). A second problem with using Cohen's  $\delta$  is that, although it is intended as a measure of group separation, it is not always an adequate measure of group separation due to the fact  $\delta$  can be dramatically affected by outliers and long-tailed distributions (Keselman & Wilcox, 2003). Algina, et al. (2005b) recommended a robust version of Cohen's  $\delta$  defined by

$$\delta_R = .642 \left( \frac{\mu_{t2} - \mu_{t1}}{\sigma_w} \right).$$

Algina and Keselman (2003b) and Algina, et al. (2005b) reported that CIs for  $\delta_R$  have better coverage probability than do CIs for Cohen's  $\delta$ , and that the actual coverage probability is closer to the nominal coverage probability for CIs constructed by using the percentile bootstrap than for the CIs constructed by using the noncentral  $t$  distribution-based method.

In the more than two group cases, Zhang and Algina (2008) examined the coverage performance of the CIs for the Root Mean Square Standardized Effect (RMSSE,  $f^*$ ) proposed by Steiger and Fouladi (1997), which is one of the generalized ES measures in ANOVA. The findings of their study indicated that the coverage probabilities of the CIs for  $f^*$  were not adequate under data nonnormality. This is not surprising because  $f^*$  is formulated with least-square parameters which are affected by skewed data, long tails and/or outlying values.

This study proposed a robust version of  $f^*$ ,  $f_R^*$ , by substituting robust estimators, i.e., trimmed means and Winsorized variances, for the least-square values. The coverage performances of the noncentral  $F$  distribution-based and the percentile bootstrap CIs for  $f_R^*$  were examined in this investigation.

ROBUST RMSSE

Table 1: Estimated Coverage Probabilities for Nominal 95% Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ : J = 3, Equally Spaced Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	<b>.973</b>	<b><u>.994</u></b>	<b>.968</b>	<b><u>.993</u></b>	<b><u>.977</u></b>	<b><u>.994</u></b>	<b>.975</b>	<b><u>.992</u></b>	<b>.975</b>	<b><u>.994</u></b>
	35	<b>.968</b>	<b><u>.990</u></b>	<b><u>.977</u></b>	<b><u>.993</u></b>	<b>.972</b>	<b><u>.990</u></b>	<b><u>.979</u></b>	<b><u>.993</u></b>	<b>.974</b>	<b><u>.996</u></b>
	50	<b>.972</b>	<b><u>.993</u></b>	<b><u>.976</u></b>	<b><u>.994</u></b>	<b>.974</b>	<b><u>.992</u></b>	<b><u>.977</u></b>	<b><u>.992</u></b>	<b>.974</b>	<b><u>.990</u></b>
.10	20	.943	<b><u>.991</u></b>	.951	<b><u>.994</u></b>	.948	<b><u>.993</u></b>	.953	<b><u>.990</u></b>	.955	<b><u>.993</u></b>
	35	.957	<b><u>.987</u></b>	.951	<b><u>.989</u></b>	.954	<b><u>.988</u></b>	.949	<b><u>.988</u></b>	.954	<b><u>.990</u></b>
	50	.943	<b><u>.983</u></b>	.956	<b><u>.988</u></b>	.954	<b><u>.988</u></b>	.958	<b><u>.988</u></b>	.952	<b><u>.989</u></b>
.25	20	.942	<b><u>.981</u></b>	.952	<b><u>.993</u></b>	.947	<b><u>.989</u></b>	<b>.961</b>	<b><u>.988</u></b>	.943	<b><u>.988</u></b>
	35	.945	<b><u>.981</u></b>	.950	<b><u>.982</u></b>	.946	<b><u>.981</u></b>	.954	<b><u>.984</u></b>	.952	<b><u>.986</u></b>
	50	.953	<b><u>.978</u></b>	.940	<b><u>.981</u></b>	.951	<b><u>.987</u></b>	.950	<b><u>.978</u></b>	.945	<b><u>.982</u></b>
.40	20	.942	<b><u>.976</u></b>	.951	<b><u>.988</u></b>	.945	<b><u>.990</u></b>	.940	<b><u>.980</u></b>	<b>.939</b>	<b><u>.991</u></b>
	35	.954	<b>.970</b>	<b>.934</b>	<b>.964</b>	.944	<b>.969</b>	.949	<b>.974</b>	.954	<b><u>.978</u></b>
	50	.950	<b>.961</b>	.943	<b>.962</b>	<b>.935</b>	.960	<b>.939</b>	<b>.961</b>	.949	<b>.968</b>
.55	20	.943	<b>.973</b>	<b>.939</b>	<b><u>.978</u></b>	<b>.937</b>	<b><u>.979</u></b>	<b>.932</b>	<b>.972</b>	<b>.938</b>	<b><u>.977</u></b>
	35	.944	<b>.968</b>	.946	<b>.963</b>	.940	<b>.969</b>	<b><u>.924</u></b>	.960	.940	<b>.966</b>
	50	.947	.960	<b>.934</b>	.952	.940	.959	<b>.929</b>	<b>.963</b>	<b>.929</b>	.958
.70	20	.945	<b>.968</b>	.940	<b>.972</b>	<b>.938</b>	<b><u>.977</u></b>	<b><u>.916</u></b>	<b>.973</b>	<b>.936</b>	<b><u>.980</u></b>
	35	<b>.935</b>	.958	.944	<b>.964</b>	<b>.935</b>	<b>.965</b>	<b><u>.923</u></b>	<b>.969</b>	<b><u>.924</u></b>	<b>.964</b>
	50	.942	<b>.962</b>	.944	<b>.967</b>	<b>.928</b>	<b>.967</b>	<b><u>.923</u></b>	<b>.968</b>	<b>.935</b>	<b>.963</b>

Note: Bold values are estimates outside the interval  $[.94, .96]$  and bold underlined values are outside the interval  $[.925, .975]$ .

Table 2: Estimated Coverage Probabilities for Nominal 95% Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ :  $J = 3$ , One Extreme Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	.973	<b><u>.994</u></b>	.972	<b><u>.992</u></b>	.974	<b><u>.996</u></b>	<b><u>.983</u></b>	<b><u>.996</u></b>	.972	<b><u>.992</u></b>
	35	<b><u>.981</u></b>	<b><u>.993</u></b>	<b><u>.979</u></b>	<b><u>.993</u></b>	.975	<b><u>.991</u></b>	<b><u>.980</u></b>	<b><u>.991</u></b>	.969	<b><u>.991</u></b>
	50	.974	<b><u>.991</u></b>	<b><u>.976</u></b>	<b><u>.994</u></b>	.973	<b><u>.991</u></b>	<b><u>.980</u></b>	<b><u>.995</u></b>	.970	<b><u>.993</u></b>
.10	20	.952	<b><u>.990</u></b>	.944	<b><u>.991</u></b>	.946	<b><u>.991</u></b>	.951	<b><u>.990</u></b>	.953	<b><u>.994</u></b>
	35	.949	<b><u>.986</u></b>	.948	<b><u>.990</u></b>	.944	<b><u>.990</u></b>	<b><u>.938</u></b>	<b><u>.986</u></b>	.944	<b><u>.987</u></b>
	50	.945	<b><u>.987</u></b>	.949	<b><u>.986</u></b>	.948	<b><u>.990</u></b>	.958	<b><u>.986</u></b>	.956	<b><u>.987</u></b>
.25	20	.946	<b><u>.982</u></b>	.945	<b><u>.985</u></b>	.952	<b><u>.987</u></b>	.954	<b><u>.991</u></b>	.954	<b><u>.990</u></b>
	35	.952	<b><u>.984</u></b>	.942	<b><u>.982</u></b>	.950	<b><u>.986</u></b>	.948	<b><u>.981</u></b>	.947	<b><u>.986</u></b>
	50	<b><u>.938</u></b>	<b><u>.976</u></b>	.953	<b><u>.981</u></b>	.951	<b><u>.982</u></b>	.950	<b><u>.984</u></b>	.949	<b><u>.983</u></b>
.40	20	.949	<b><u>.980</u></b>	.943	<b><u>.984</u></b>	.942	<b><u>.988</u></b>	<b><u>.938</u></b>	<b><u>.984</u></b>	.942	<b><u>.992</u></b>
	35	.943	<b><u>.966</u></b>	.949	<b><u>.972</u></b>	.946	<b><u>.976</u></b>	.948	<b><u>.972</u></b>	.948	<b><u>.973</u></b>
	50	.952	<b><u>.962</u></b>	.950	<b><u>.964</u></b>	.952	<b><u>.970</u></b>	.946	<b><u>.965</u></b>	.945	<b><u>.966</u></b>
.55	20	.952	<b><u>.975</u></b>	.943	<b><u>.975</u></b>	.941	<b><u>.980</u></b>	.940	<b><u>.981</u></b>	.943	<b><u>.984</u></b>
	35	.943	.958	.947	<b><u>.966</u></b>	<b><u>.938</u></b>	<b><u>.963</u></b>	<b><u>.936</u></b>	<b><u>.968</u></b>	.943	<b><u>.970</u></b>
	50	.942	<b><u>.961</u></b>	<b><u>.931</u></b>	.953	.943	<b><u>.962</u></b>	<b><u>.935</u></b>	<b><u>.964</u></b>	<b><u>.934</u></b>	.958
.70	20	.944	<b><u>.970</u></b>	<b><u>.937</u></b>	<b><u>.976</u></b>	<b><u>.931</u></b>	<b><u>.972</u></b>	<b><u>.930</u></b>	<b><u>.982</u></b>	<b><u>.936</u></b>	<b><u>.982</u></b>
	35	.941	.960	<b><u>.938</u></b>	<b><u>.964</u></b>	<b><u>.932</u></b>	<b><u>.965</u></b>	<b><u>.924</u></b>	<b><u>.966</u></b>	<b><u>.934</u></b>	<b><u>.966</u></b>
	50	<b><u>.939</u></b>	.957	.940	<b><u>.962</u></b>	<b><u>.938</u></b>	<b><u>.966</u></b>	<b><u>.932</u></b>	<b><u>.965</u></b>	<b><u>.935</u></b>	<b><u>.967</u></b>

Note: Bold values are estimates outside the interval  $[.94, .96]$  and bold underlined values are outside the interval  $[.925, .975]$ .

ROBUST RMSSE

Table 3: Estimated Coverage Probabilities for Nominal 95% Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ : J = 6, Equally Spaced Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	<b><u>.976</u></b>	<b><u>.978</u></b>	<b><u>.977</u></b>	<b><u>.982</u></b>	<b><u>.976</u></b>	<b><u>.986</u></b>	<b><u>.980</u></b>	<b><u>.983</u></b>	<b><u>.978</u></b>	<b><u>.986</u></b>
	35	<b><u>.972</u></b>	<b><u>.969</u></b>	<b><u>.969</u></b>	<b><u>.972</u></b>	<b><u>.969</u></b>	<b><u>.974</u></b>	<b><u>.979</u></b>	<b><u>.977</u></b>	<b><u>.977</u></b>	<b><u>.974</u></b>
	50	<b><u>.974</u></b>	<b><u>.966</u></b>	<b><u>.972</u></b>	<b><u>.965</u></b>	<b><u>.975</u></b>	<b><u>.974</u></b>	<b><u>.979</u></b>	<b><u>.970</u></b>	<b><u>.975</u></b>	<b><u>.972</u></b>
.10	20	.950	<b><u>.977</u></b>	.950	<b><u>.974</u></b>	.948	<b><u>.979</u></b>	.947	<b><u>.980</u></b>	.944	<b><u>.982</u></b>
	35	.948	<b><u>.963</u></b>	.952	<b><u>.969</u></b>	.950	<b><u>.970</u></b>	.951	<b><u>.974</u></b>	.951	<b><u>.972</u></b>
	50	.948	<b><u>.962</u></b>	.952	<b><u>.965</u></b>	.954	<b><u>.971</u></b>	.944	<b><u>.967</u></b>	.947	<b><u>.966</u></b>
.25	20	.945	<b><u>.966</u></b>	.947	<b><u>.975</u></b>	<b><u>.938</u></b>	<b><u>.983</u></b>	.952	<b><u>.982</u></b>	.943	<b><u>.983</u></b>
	35	.943	.960	.947	<b><u>.970</u></b>	.948	<b><u>.981</u></b>	.942	<b><u>.969</u></b>	.950	<b><u>.978</u></b>
	50	.950	<b><u>.962</u></b>	.940	<b><u>.961</u></b>	.944	<b><u>.970</u></b>	.944	<b><u>.970</u></b>	.944	<b><u>.969</u></b>
.40	20	.947	<b><u>.972</u></b>	.943	<b><u>.980</u></b>	.948	<b><u>.990</u></b>	<b><u>.938</u></b>	<b><u>.981</u></b>	<b><u>.935</u></b>	<b><u>.989</u></b>
	35	.946	<b><u>.968</u></b>	.946	<b><u>.971</u></b>	.945	<b><u>.978</u></b>	<b><u>.938</u></b>	<b><u>.972</u></b>	.942	<b><u>.976</u></b>
	50	.954	<b><u>.968</u></b>	.944	<b><u>.966</u></b>	.942	<b><u>.971</u></b>	<b><u>.936</u></b>	<b><u>.969</u></b>	.944	<b><u>.970</u></b>
.55	20	.949	<b><u>.974</u></b>	<b><u>.939</u></b>	<b><u>.980</u></b>	<b><u>.936</u></b>	<b><u>.986</u></b>	<b><u>.926</u></b>	<b><u>.985</u></b>	<b><u>.936</u></b>	<b><u>.989</u></b>
	35	.955	<b><u>.971</u></b>	.947	<b><u>.973</u></b>	.942	<b><u>.973</u></b>	<b><u>.915</u></b>	<b><u>.969</u></b>	<b><u>.934</u></b>	<b><u>.982</u></b>
	50	.946	<b><u>.962</u></b>	.943	<b><u>.967</u></b>	.948	<b><u>.973</u></b>	<b><u>.927</u></b>	<b><u>.969</u></b>	<b><u>.928</u></b>	<b><u>.967</u></b>
.70	20	.949	<b><u>.980</u></b>	<b><u>.938</u></b>	<b><u>.984</u></b>	<b><u>.930</u></b>	<b><u>.983</u></b>	<b><u>.902</u></b>	<b><u>.984</u></b>	<b><u>.939</u></b>	<b><u>.988</u></b>
	35	.943	<b><u>.967</u></b>	<b><u>.934</u></b>	<b><u>.970</u></b>	<b><u>.921</u></b>	<b><u>.968</u></b>	<b><u>.907</u></b>	<b><u>.972</u></b>	<b><u>.921</u></b>	<b><u>.971</u></b>
	50	.941	<b><u>.962</u></b>	.944	<b><u>.972</u></b>	<b><u>.933</u></b>	<b><u>.968</u></b>	<b><u>.914</u></b>	<b><u>.970</u></b>	<b><u>.928</u></b>	<b><u>.966</u></b>

Note: Bold values are estimates outside the interval  $[.94, .96]$  and bold underlined values are outside the interval  $[.925, .975]$ .

Table 4: Estimated Coverage Probabilities for Nominal 95% Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ :  $J = 6$ , One Extreme Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	<b>.972</b>	<b><u>.976</u></b>	<b><u>.979</u></b>	<b><u>.981</u></b>	<b><u>.977</u></b>	<b><u>.983</u></b>	.974	<b><u>.984</u></b>	<b><u>.977</u></b>	<b><u>.989</u></b>
	35	<b>.974</b>	<b>.969</b>	<b>.971</b>	<b>.968</b>	<b><u>.976</u></b>	<b><u>.976</u></b>	<b>.975</b>	<b>.972</b>	<b><u>.978</u></b>	<b><u>.976</u></b>
	50	<b><u>.978</u></b>	<b>.973</b>	<b><u>.976</u></b>	<b>.970</b>	<b><u>.977</u></b>	<b>.974</b>	<b><u>.982</u></b>	<b><u>.976</u></b>	<b>.971</b>	<b>.965</b>
.10	20	<b>.962</b>	<b><u>.976</u></b>	.945	<b>.975</b>	.953	<b><u>.982</u></b>	.943	<b><u>.978</u></b>	.954	<b><u>.981</u></b>
	35	.954	<b>.971</b>	.955	<b>.970</b>	.955	<b>.969</b>	.945	<b>.962</b>	.948	<b>.975</b>
	50	.948	.960	.948	<b>.965</b>	.951	<b>.964</b>	.954	<b>.971</b>	.956	<b>.972</b>
.25	20	.951	<b>.974</b>	<b>.937</b>	<b>.974</b>	.954	<b><u>.987</u></b>	.949	<b><u>.978</u></b>	.951	<b><u>.988</u></b>
	35	.952	<b>.967</b>	.950	<b><u>.976</u></b>	.945	<b>.974</b>	.955	<b>.973</b>	.951	<b>.974</b>
	50	.953	<b>.965</b>	.946	<b>.961</b>	.943	<b>.970</b>	.948	<b>.970</b>	.951	<b>.973</b>
.40	20	.945	<b>.972</b>	.950	<b><u>.984</u></b>	.945	<b><u>.983</u></b>	.952	<b><u>.983</u></b>	<b>.938</b>	<b><u>.988</u></b>
	35	<b>.939</b>	.958	.944	<b>.969</b>	.942	<b><u>.977</u></b>	.952	<b><u>.979</u></b>	<b>.938</b>	<b>.972</b>
	50	.941	.956	.945	<b>.968</b>	<b>.936</b>	<b>.971</b>	.951	<b>.975</b>	.945	<b>.975</b>
.55	20	.944	<b><u>.976</u></b>	.943	<b><u>.982</u></b>	<b>.936</b>	<b><u>.987</u></b>	.942	<b><u>.989</u></b>	<b>.938</b>	<b><u>.988</u></b>
	35	.949	<b>.970</b>	.940	<b>.973</b>	<b>.934</b>	<b>.970</b>	<b>.937</b>	<b><u>.981</u></b>	<b>.935</b>	<b>.970</b>
	50	.950	<b>.963</b>	<b>.939</b>	<b>.961</b>	<b>.932</b>	<b>.967</b>	<b>.927</b>	<b>.968</b>	<b>.931</b>	<b>.966</b>
.70	20	<b>.935</b>	<b>.972</b>	<b>.938</b>	<b><u>.982</u></b>	<b>.929</b>	<b><u>.985</u></b>	<b>.925</b>	<b><u>.992</u></b>	<b>.928</b>	<b><u>.987</u></b>
	35	.946	<b>.970</b>	<b>.936</b>	<b>.972</b>	<b><u>.917</u></b>	<b>.964</b>	<b><u>.920</u></b>	<b>.973</b>	<b>.926</b>	<b>.969</b>
	50	.946	<b>.966</b>	<b>.932</b>	<b>.961</b>	<b>.930</b>	<b>.972</b>	<b><u>.908</u></b>	<b>.961</b>	<b>.930</b>	<b>.968</b>

Note: Bold values are estimates outside the interval  $[.94, .96]$  and bold underlined values are outside the interval  $[.925, .975]$ .

ROBUST RMSSE

Table 5: Average Widths of Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ : J=3, Equally Spaced Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	.491	.618	.494	.611	.488	.596	.490	.601	.487	.595
	35	.366	.437	.361	.430	.365	.428	.364	.425	.368	.428
	50	.301	.354	.304	.354	.306	.353	.305	.351	.304	.352
.10	20	.512	.637	.509	.625	.515	.618	.514	.625	.515	.617
	35	.397	.461	.390	.453	.395	.450	.396	.453	.386	.444
	50	.332	.375	.334	.378	.332	.374	.335	.376	.332	.374
.25	20	.609	.725	.607	.714	.601	.701	.615	.725	.604	.703
	35	.487	.550	.485	.540	.480	.533	.485	.540	.481	.533
	50	.423	.466	.421	.464	.421	.462	.421	.463	.421	.461
.40	20	.702	.828	.702	.823	.692	.813	.698	.840	.699	.818
	35	.545	.613	.543	.611	.543	.613	.544	.623	.542	.613
	50	.454	.499	.454	.501	.454	.506	.454	.511	.453	.505
.55	20	.760	.899	.758	.905	.755	.906	.755	.937	.756	.914
	35	.562	.625	.562	.634	.561	.641	.562	.662	.561	.647
	50	.463	.501	.463	.508	.463	.513	.463	.532	.463	.518
.70	20	.794	.947	.791	.961	.789	.980	.793	1.026	.789	.987
	35	.579	.644	.578	.655	.577	.668	.579	.698	.578	.679
	50	.479	.517	.478	.526	.477	.536	.478	.565	.478	.545

Note: Results are based on 2,500 replications.

Table 6: Average Widths Of Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ : J=3, One Extreme Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	.487	.616	.497	.613	.489	.598	.484	.596	.494	.598
	35	.366	.436	.364	.432	.367	.430	.363	.425	.370	.430
	50	.304	.357	.309	.357	.305	.352	.305	.350	.303	.351
.10	20	.518	.638	.517	.631	.515	.620	.510	.619	.510	.613
	35	.391	.457	.390	.452	.390	.450	.391	.450	.392	.450
	50	.334	.381	.333	.378	.334	.376	.333	.373	.337	.378
.25	20	.611	.726	.607	.715	.608	.709	.606	.715	.598	.696
	35	.485	.546	.486	.543	.484	.539	.485	.542	.485	.538
	50	.420	.464	.422	.464	.421	.461	.422	.462	.423	.461
.40	20	.705	.834	.696	.815	.697	.813	.700	.833	.695	.813
	35	.544	.612	.544	.611	.543	.614	.544	.614	.542	.610
	50	.454	.501	.454	.501	.454	.503	.454	.503	.454	.503
.55	20	.760	.899	.756	.901	.756	.912	.759	.926	.755	.910
	35	.562	.627	.561	.635	.561	.642	.563	.652	.561	.642
	50	.463	.501	.463	.508	.463	.513	.463	.520	.462	.514
.70	20	.793	.942	.791	.956	.790	.979	.792	1.006	.789	.982
	35	.580	.646	.578	.652	.579	.671	.580	.692	.578	.670
	50	.478	.518	.478	.526	.478	.539	.478	.557	.478	.540

Note: Results are based on 2,500 replications.



ROBUST RMSSE

Table 7: Average Widths of Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ : J = 6, Equally Spaced Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	.352	.487	.356	.484	.345	.474	.345	.480	.348	.475
	35	.259	.351	.261	.350	.262	.348	.256	.346	.260	.346
	50	.218	.289	.219	.289	.216	.286	.215	.286	.219	.287
.10	20	.372	.499	.379	.501	.375	.493	.374	.499	.373	.491
	35	.294	.370	.290	.367	.294	.368	.290	.367	.290	.364
	50	.253	.310	.252	.308	.251	.307	.251	.306	.249	.305
.25	20	.464	.559	.460	.554	.459	.551	.463	.561	.457	.550
	35	.363	.406	.362	.404	.361	.405	.360	.406	.360	.405
	50	.302	.326	.301	.326	.301	.326	.301	.328	.301	.326
.40	20	.506	.577	.505	.582	.503	.590	.503	.598	.502	.591
	35	.361	.389	.362	.392	.361	.396	.361	.405	.361	.398
	50	.293	.309	.293	.311	.293	.314	.293	.321	.293	.315
.55	20	.506	.572	.506	.583	.506	.600	.506	.619	.507	.610
	35	.363	.392	.363	.399	.363	.405	.364	.422	.363	.411
	50	.299	.318	.299	.323	.299	.328	.299	.340	.298	.331
.70	20	.517	.590	.516	.605	.516	.629	.517	.666	.516	.641
	35	.377	.412	.376	.421	.376	.433	.377	.461	.376	.442
	50	.311	.336	.311	.342	.310	.350	.311	.372	.310	.356

Note: Results are based on 2,500 replications.

Table 8: Average Widths of Noncentral F Distribution-Based (NCF) and Percentile Bootstrap (Boot) CIs for  $f_R^*$ :  $J = 6$ , One Extreme Mean Configuration

$f_R^*$	$n$	Normal		$g = .000$ $h = .109$		$g = .000$ $h = .225$		$g = .760$ $h = -.098$		$g = .225$ $h = .225$	
		NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot	NCF	Boot
.00	20	.353	.488	.348	.482	.350	.477	.349	.484	.346	.476
	35	.259	.350	.260	.349	.260	.347	.258	.348	.259	.346
	50	.219	.290	.215	.287	.218	.288	.215	.286	.217	.287
.10	20	.381	.504	.377	.499	.375	.492	.377	.502	.379	.494
	35	.295	.372	.294	.369	.292	.365	.292	.367	.293	.366
	50	.251	.308	.252	.308	.251	.306	.252	.307	.251	.306
.25	20	.463	.558	.459	.553	.460	.552	.462	.558	.461	.553
	35	.362	.406	.361	.404	.361	.405	.363	.401	.361	.403
	50	.302	.325	.301	.325	.301	.326	.303	.321	.302	.325
.40	20	.505	.576	.505	.582	.505	.589	.506	.583	.504	.589
	35	.361	.388	.362	.393	.362	.398	.361	.389	.361	.395
	50	.293	.308	.293	.311	.293	.315	.293	.312	.293	.312
.55	20	.506	.570	.505	.584	.506	.599	.506	.600	.506	.603
	35	.363	.393	.363	.399	.363	.405	.363	.411	.363	.407
	50	.299	.317	.299	.321	.299	.327	.299	.335	.298	.328
.70	20	.518	.591	.517	.607	.516	.634	.517	.643	.516	.640
	35	.376	.410	.376	.423	.376	.433	.376	.452	.376	.440
	50	.311	.334	.310	.341	.310	.349	.311	.368	.310	.353

Note: Results are based on 2,500 replications.

Comparisons were made to the CIs for  $f_R^*$  constructed by using the noncentral  $F$  distribution-based and the bootstrap methods in terms of the probability coverage and interval width.

The robustness of the CIs for  $f_R^*$  was investigated in a one-way, fixed-effects, between-subjects ANOVA. The study conditions incorporated five population distributions including the normal distribution and four additional cases from the family of the  $g$  and  $h$  distributions that are nonnormal; two number of levels for the number of treatment groups:  $J = 3$  and  $J = 6$ ; three cell sample sizes in each treatment ( $n = 20, 35$  and  $50$ ); six values of population  $\text{RMSSE}_R$  ( $0.00, 0.10, 0.25, 0.40, 0.55$  and  $0.70$ ); and two mean configurations: the equally spaced mean configuration and the one extreme mean configuration. The nominal confidence level for all intervals investigated was  $0.95$  and each condition was replicated  $2,500$  times. The number of bootstrap replications in the bootstrap procedure was  $1,000$ .

The results indicated that the coverage probabilities of the noncentral  $F$  distribution-based CIs for  $f_R^*$  introduced in this study, which was formulated with robust trimmed means and Winsorized variances, were generally adequate, that is, generally either within our lenient criterion of robustness  $[.925, .975]$ , or both the lenient criterion of robustness and the strict criterion interval  $[.94, .96]$ . There were only a few cases in which the noncentral  $F$  distribution-based CIs for  $f_R^*$  broke down. These include some cases of  $f_R^* = 0$ , and when  $f_R^* = .70$  for small sample sizes under nonnormal data distributions, especially under the  $g = .760, h = -.098$  distribution.

For the bootstrap CIs for  $f_R^*$ , the probability coverage were not adequate when  $J = 3$  and  $f_R^* \leq .25$  or when  $J = 6$  and sample size was small, especially when sample size was  $20$ . In particular, when  $J = 3$ , over half of the estimated coverage probabilities were outside of the  $[.925, .975]$  interval. These probability

coverages mostly occurred when  $f_R^* \leq .25$ . When  $J = 6$ , the bootstrap CIs were mostly inside the  $[.925, .975]$  criterion interval under normality. However, under all other data distribution conditions, they were outside of the interval when sample size was small: most cases for  $n = 20$  as well as some cases for  $n = 35$ .

For both the noncentral  $F$  distribution-based and the bootstrap CIs for  $f_R^*$ , the mean configuration did not appear to alter the pattern of the probability coverage performance. However, sample sizes seem to be slightly positively related to probability coverage. The widths of the noncentral  $F$  distribution-based CIs for  $f_R^*$  were shorter than those of the bootstrap CIs under the same condition. Therefore, not only does the noncentral  $F$  distribution-based CI for  $f_R^*$  have better coverage probability than the bootstrap CIs for  $f_R^*$ , they are also narrower than those of the bootstrap CI. Both the widths of the noncentral  $F$  distribution-based and bootstrap CIs for  $f_R^*$  remained relatively unchanged across data distributions. In other words, the widths of the bootstrap CIs for  $f_R^*$  fluctuated very little across data distribution conditions.

For both the noncentral  $F$  distribution-based and the bootstrap CIs for  $f_R^*$ , as the number of levels of  $J$  increases, the width of the estimated CIs becomes narrower. For both the noncentral  $F$  distribution-based and the bootstrap CIs for  $f_R^*$ , under the same condition, the average width of the CIs becomes narrower as the sample size increases and the population effect size  $f_R^*$  decreases.

In summary, both the noncentral  $F$  distribution-based and the bootstrap CIs for  $f^*$ , which are based on the usual least-square estimators, yielded inadequate coverage probabilities. Thus, an important task to help researchers who want to set a CI around  $f^*$  is developing a better interval than the noncentral  $F$  distribution-based or percentile bootstrap CI. The noncentral  $F$  distribution-based CIs for  $f_R^*$ , which was proposed in the current study and

was formulated with the robust parameters including the trimmed means and Winsorized variances, yielded fairly adequate coverage probabilities and better coverage probability than the percentile bootstrap CI. Accordingly, researchers who want to set a CI for  $f_R^*$  can use the CI constructed by using the noncentral  $F$  distribution and will enjoy the additional benefit of using a robust measure of effect size, that is, a measure that is not likely to be strongly affected by outlying data points.

#### References

- Algina, J., & Keselman, H. J. (2003a). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement, 63*, 537-553.
- Algina, J., & Keselman, H. J. (2003b, May). Confidence intervals for effect sizes. Paper presented at a conference in honor of H. Swaminathan, Amherst, MA.
- Algina, J., Keselman, H. J., & Penfield, R. D. (2006). Confidence interval coverage for Cohen's effect size Statistic. *Educational and Psychological Measurement, 66*, 945-960.
- Algina, J., Keselman, H. J., & Penfield, R. D. (2005a). Effect sizes and their intervals: the two-level repeated measures case. *Educational and Psychological Measurement, 65*, 241-258.
- Algina, J., Keselman, H. J., & Penfield, R. D. (2005b). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods, 10*, 317-328.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association (5<sup>th</sup> Ed.)*. APA: Washington, DC.
- Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology*, 95-121. New York: Academic Press.
- Cohen, J. (1969). *Statistical power analyses for the behavioral sciences*. New York: Academic Press, Inc.
- Cohen, J. (1994). The earth is round ( $p < 0.05$ ). *American Psychologist, 49*, 997-1003.
- Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532-574.
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 178-180.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Hays, W. L. (1963). *Statistics*. New York: Holt, Rinehart & Winston.
- Hoaglin, D. C. (1983). Summarizing shape numerically: The g-and h distributions. In D. D. Hoaglin, F. Mosteller, & J. W. Tukey, (Eds.), *Data analysis for tables, trends, and shapes: Robust and exploratory techniques*. New York: Wiley.
- Martinez, J., & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Communications in Statistics, Theory and Methods, 13*, 353-369.
- Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103-115.
- Murphy, K. R. (1997). Editorial. *Journal of Applied Psychology, 82*, 3-5.
- Nickerson, R. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.
- SAS Institute Inc. (1999). *SAS/IML user's guide, version 8*. Cary, NC: Author.
- Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164-182.
- Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. Harlow, S. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* Hillsdale, NJ: Erlbaum.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

## ROBUST RMSSE

Wilcoxon, R. R. (2003). *Applying contemporary statistical techniques*. San Diego: Academic Press.

Wilcoxon, R. R. (2005). *Introduction to robust estimation and hypothesis testing (2<sup>nd</sup> Ed.)*. San Diego: Elsevier Academic Press.

Wilcoxon, R. R., & Keselman. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8, 254-274.

Wilkinson, L., & Task force on Statistical Inference (1999). Statistical methods in psychology journals. *American Psychologist*, 54, 594-604.

Zhang, G. (2009). t Test: The good, the bad, the ugly, & the remedy. *Middle Grades Research Journal*, 4(2), 25-34.

Zhang, G., & Algina, J. (2008). Coverage performance of the non-central F-based and percentile bootstrap confidence intervals for Root Mean Square Standardized Effect Size in one-way fixed-effects ANOVA. *Journal of Modern Applied Statistical Methods*, 7(1), 56-76.

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, 61, 165-170.

## Sample Size Considerations for Multiple Comparison Procedures in ANOVA

Gordon P. Brooks    George A. Johanson  
Ohio University,  
Athens, Ohio USA

---

Adequate sample sizes for omnibus ANOVA tests do not necessarily provide sufficient statistical power for post hoc multiple comparisons typically performed following a significant omnibus  $F$  test. Results reported support a comparison-of-most-interest approach for sample size determination in ANOVA based on effect sizes for multiple comparisons.

Key words: Sample size, multiple comparison procedures, Tukey, ANOVA.

---

### Introduction

The determination of an appropriate sample size is an often difficult, but critically important, element in the research design process. One of the chief functions of experimental design is to ensure that a study has adequate statistical power to detect meaningful differences, if indeed they exist (e.g., Hopkins & Hopkins, 1979). There is a very good reason why researchers should worry about statistical power a priori: If researchers are going to invest time and money in carrying out a study, then they would want to have a reasonable chance, perhaps 70% or 80%, to find a statistically significant difference between groups if it does exist in the population. Thus, a priori power, the probability of rejecting a null hypothesis that is indeed false, will inform researchers about how many subjects per group will be needed for adequate power (Light, Singer & Willett, 1990).

Among the most important matters impacting the choice of sample size is the

particular statistical analysis that will be used to analyze data. For example, when a  $t$  test is used, the researcher commonly estimates an expected, standardized group mean difference effect size (such as Cohen's  $d$ ) in order to determine an appropriate sample size. Sample sizes in analysis of variance (ANOVA) are often based on an effect size that represents an overall standardized difference in the means (such as Cohen's  $f$ ), but these recommended sample sizes provide statistical power only for the omnibus null hypothesis (overall ANOVA) that no group means differ. Adequate sample size for the omnibus test does not necessarily provide sufficient statistical power for the post hoc multiple comparisons typically performed following a statistically significant (exploratory) omnibus test and in many cases the multiple comparisons are of most interest to a researcher.

The purpose of this study was to determine whether the knowledge that multiple comparison procedures will be used following a statistically significant omnibus ANOVA can be helpful in choosing a sample size for a given study. In particular, results using the Tukey HSD post hoc multiple comparison procedure (MCP) were examined to determine whether specific recommendations can be made about sample sizes when the Tukey MCP is used and three groups are compared. This evidence was used to reach conclusions about whether such an approach to sample size selection has merit. Note that this is a presentation of a new approach to sample size selection – specifically, a new way to think about effect sizes – for

---

Gordon P. Brooks is an Associate Professor of Educational Research and Evaluation. His research interests include statistics education, power and sample size analysis and Monte Carlo programming. Email him at: brooksg@ohio.edu. George A. Johanson is a Professor Emeritus of Educational Research and Evaluation. His research interests include survey research methods and differential item and person functioning. Email him at: johanson@ohio.edu.

exploratory ANOVA where post hoc comparisons are relevant. Other approaches are both more appropriate and more powerful when planned comparisons are made in a confirmatory analysis.

#### Theoretical Framework

Several factors play a role in sample size determination, including that after the statistical method and the directionality of the statistical alternative hypotheses have been decided, sample size, level of significance, effect size and statistical power are all functionally related. Other issues also impact statistical power, such as the reliability of measurements, unequal group sizes and unequal group variances. However, little consideration has been given to the role of post hoc multiple comparison tests in choosing adequate sample sizes.

In order to maintain reasonable experiment-wise Type I error rates when group means are compared, researchers often use ANOVA followed by an appropriate MCP. The overall ANOVA is tested using an omnibus test at a predetermined level of significance (e.g., 0.05). The post hoc tests that follow a statistically significant omnibus test are then often performed at an adjusted level of significance, based on the number of comparisons to be made.

For example, when comparing four groups, six pairwise group mean comparisons possible. If the researcher wishes to perform all six pairwise comparisons, the per comparison (i.e., per test) level of significance would be adjusted so that the entire set of follow-up tests does not exceed the experiment-wise alpha (e.g., if experiment-wise alpha is 0.05, the adjusted per comparison alpha might be  $0.05/6 = 0.0083$ , using a Bonferroni approach). Each MCP performs this adjustment differently, resulting in different performance for each in terms of Type I error and statistical power (e.g., Carmer & Swanson, 1973; Einot & Gabriel, 1975; Toothaker, 1991).

Several methods exist for determining sample size for ANOVA. Most common are statistical power approaches based on Cohen's (1988)  $f$  effect size, which represents the standardized variability of the group means about the grand mean (Stevens, 2007). This

method (and other similar methods) concentrates on the statistical power of the omnibus test in ANOVA. Others, Hinkle, Wiersma and Jurs (2003) and Levin (1975), for example, have recommended approaches based on how large the sample must be to detect a predetermined mean difference effect size between any two groups, or two extreme groups. Although Levin's approach is designed for use with the Scheffé multiple comparison procedure, Hinkle, et al. base their method on Cohen's  $d$  effect size for comparison between the two groups with the largest (most extreme) mean differences, and therefore do not consider the adjustments to alpha for multiple comparison procedures. Pan and Dayton (2005) provided sample size requirements for patterns of ordered means, but focused on an information criteria approach to pair-wise comparison procedures.

#### Comparison-of-Most-Interest

When determining sample sizes for a factorial ANOVA, researchers may choose the sample size that provides sufficient statistical power for all sources of variation (e.g., main effects and interactions). Alternatively, researchers may determine which effect is most important to them and select a sample size based on the expected effect size for that particular source of variation. For example, researchers may have most interest in the interaction effect or a particular main effect. Depending on the structure of the cell means, these effect sizes can vary and therefore result in different required sample sizes for the various main effects and interaction effects.

The approach presented in this study is based loosely on this effect-of-most-interest approach from factorial ANOVA as applied to one-way ANOVA: That is, beyond determining the sample size required for an omnibus test in one-way ANOVA, the new approach also determines the sample sizes required for the follow-up tests from a given set of population means.

For example, in a 3-group study the researcher may be able to estimate that a large effect exists between a control group and two types of treatment, but may expect a much smaller difference between two types of treatment. The comparison-of-most-interest may

be the difference between the treatments and the control; however, the much smaller difference between the two treatments may be the most interesting. The researcher would use this information to determine an appropriate sample size for the study by selecting a sample size large enough for the smaller effect size between the types of treatment. This differs from an a priori set of planned comparisons in that the researcher may have a special interest in particular comparisons, but not have specific alternative research hypotheses to predict the direction of the mean differences. The procedure studied here is an adaptation of the Hinkle, et al. (2003) approach that looks at meaningful effect sizes between any groups rather than the Hinkle, et al. difference between only the two most extreme groups.

Even in an exploratory ANOVA, it is rarely satisfactory knowing only that a difference exists in the means (as given by the omnibus test); researchers typically also want to know between which groups the differences exist. Without consideration of the multiple comparison procedures during the sample size analysis, it is possible to find a statistically significant omnibus test with no pairwise group differences determined to be statistically significant in post hoc tests. Although other potential reasons for such a result exist, it may sometimes be an issue of statistical power.

#### An Example of the Problem

Suppose a researcher is analyzing the mean differences for three groups, where the means for groups 1 and 2 are both 0.0, but the third group mean is 0.8. This represents a relatively large pairwise difference between group 3 and both groups 1 and 2. Using the Cohen (1988) effect size,  $f$ , for ANOVA, this might be characterized as a relatively large effect: Cohen's large effect size is  $f = 0.40$  and in this example  $f = 0.38$ . Cohen's sample size analysis, as implemented by the SPSS SamplePower program, indicates that 24 cases per group are required to achieve statistical power of 0.80 for the omnibus test in such a situation.

When performing a Monte Carlo analysis for this condition using the MC4G program (Brooks, 2008), approximately 80.8%

of 100,000 samples resulted in statistically significant omnibus  $F$  statistics for the ANOVA among the three groups. However, the number of correct statistically significant Tukey HSD comparisons between groups 1 and 3 and between groups 2 and 3 (with a sample size of 24 in each group), was approximately 64.7%. At the adjusted alpha used by the Tukey HSD procedure, approximately 1.9% of the comparisons between groups 1 and 2 were statistically significant (and therefore Type I errors because both group 1 and 2 had the same mean).

These illustrative power analysis results imply that a number of samples from among the 100,000 had statistically significant omnibus  $F$  statistics while, at most, one of the non-null Tukey post hoc comparisons was statistically significant. The MC4G program reported that approximately 78.9% of samples had at least one significant Tukey comparison following a significant omnibus test. However, because only 64.7% of each non-null comparison were statistically significant, and because the group 1 versus group 2 comparison was significant as a Type I error in about 1.9% of the samples, this implies that - in many of those samples - only one of the two large, non-null comparisons was statistically significant.

From another perspective, in order to reach statistical power of 0.80 for the two non-null Tukey comparisons (i.e., group 1 vs. group 3 and group 2 vs. group 3), 32 cases are needed per group, for a total sample size of 96 (compared to 24 per group based solely on the omnibus test). With a total sample size of 96 the omnibus  $F$  test, however, had a power rate of approximately 0.91.

#### Methodology

An existing Monte Carlo program was modified so that it can ascertain appropriate sample sizes for pairwise comparisons calculated using the Tukey multiple comparison procedure. The MC4G: Monte Carlo Analyses for up to 4 Groups program was originally developed by one of the authors to perform Monte Carlo analyses for  $t$  tests and ANOVA in a Windows environment (Brooks, 2008). The current version of the program (MC4G version v2008)



## MCP SAMPLE SIZES

was upgraded to include the sample size analyses required for this study.

The MC4G program was compiled in Delphi 2007. The program uses the L'Ecuyer (1988) uniform pseudorandom number generator. Specifically, the FORTRAN code of Press, et al. (1992), was translated into Delphi Pascal. The L'Ecuyer generator was chosen due to its large period and because combined generators are recommended for use with the Box-Muller method for generating random normal deviates (Park & Miller, 1988), as is the case in MC4G. The computer algorithm for the Box-Muller method used in MC4G was adapted for Delphi Pascal from the standard Pascal code provided by Press, et al. (1989). Simulated samples were chosen randomly to test program function by comparison with results provided by SPSS.

### Monte Carlo Design

In all simulations, normally distributed standardized data were generated to fit the given conditions for each simulation; that is, all variances were set to 1.0, while group means varied between 0.0 and 0.8, depending on the given effect size. A minimum of 10,000 replications were performed for the final sample size analysis in each condition. Specifically, a default value of 20,000 was used with the MC4G sample size analysis, which guaranteed that the final results would be based on at least 10,000 iterations (i.e., simulated samples). Samples sizes for all three groups were restricted to be equal. Some of the Monte Carlo simulations were run multiple times with different seeds to verify that the results were not an artifact of a poor seed choice.

Conditions included varying standardized mean differences among groups for a three-group ANOVA. In particular, groups varied such that all possible non-redundant patterns of pairwise mean differences were varied across groups from 0.0 to 0.8. The minimum non-null standardized mean difference between groups of 0.2 was chosen because of the very large sample sizes required for smaller effects; the maximum of 0.8 was chosen because of the very small sample sizes required when the mean differences are larger.

For example, whether the three group means were set at 0.2, 0.4 and 0.6 or at 0.3, 0.5 and 0.7, the pattern for both resulting standardized mean difference effect sizes (all standard deviations were 1.0) would be 0.2, 0.2 and 0.4, respectively. The mean differences - as effect sizes - are the key to the sample size analyses, not the absolute sizes of the means. Therefore, each pattern of mean differences was only included once. The result was 16 non-redundant comparison patterns that fit the mean difference conditions described (see Table 1).

### Results

Three primary findings of interest were observed from this study. First, when the pattern of means resulted in a pattern where two of the three means are equal - and different from the third - there was a consistent pattern of sample sizes required for the comparison relative to the sample size required for the omnibus test. Second, when the pattern of means resulted in two of the three mean differences being equal - and different from the third - there was a consistent pattern of sample sizes required for the comparison relative to the sample size required for the omnibus test. Third, no matter what the pattern of means, a given absolute standardized mean difference effect size consistently required the same sample size to achieve the power desired.

### Two Equal Means

In situations where two groups had the same mean and a third group mean differed, the non-null multiple comparisons required larger sample sizes than the omnibus ANOVA. For example, the condition where the pattern of standardized means was 0.0, 0.0 and 0.5 (therefore a pattern of mean differences of 0.0, 0.5 and 0.5) resulted in per group sample sizes of roughly 81 cases to achieve power of 0.80 for the two multiple comparisons with a standardized mean difference of 0.5 (see Table 2). This was compared to the 60 cases per group needed to achieve statistical power of 0.80 for the omnibus test.

All patterns with two similar means, regardless of the magnitude of the mean differences, resulted in a relative efficiency of sample sizes (omnibus per group sample size

BROOKS & JOHANSON

Table 1: Patterns of Means Studied

Analysis	Group 1 Mean	Group 2 Mean	Group 3 Mean	Comparison Pattern <sup>a</sup>	Cohen <i>f</i> Effect Size	Cohen Total N	Cohen N Per Group
1	0.0	0.0	0.2	0.0, 0.2, 0.2	0.0943	1089	363
2	0.0	0.0	0.3	0.0, 0.3, 0.3	0.1414	486	162
3	0.0	0.0	0.4	0.0, 0.4, 0.4	0.1886	276	92
4	0.0	0.0	0.5	0.0, 0.5, 0.5	0.2357	177	59
5	0.0	0.0	0.6	0.0, 0.6, 0.6	0.2828	126	42
6	0.0	0.0	0.7	0.0, 0.7, 0.7	0.3300	93	31
7	0.0	0.0	0.8	0.0, 0.8, 0.8	0.3771	72	24
8	0.0	0.2	0.4	0.2, 0.2, 0.4	0.1633	366	122
9	0.0	0.2	0.5	0.2, 0.3, 0.5	0.2055	234	78
10	0.0	0.2	0.6	0.2, 0.4, 0.6	0.2494	159	53
11	0.0	0.2	0.7	0.2, 0.5, 0.7	0.2944	117	39
12	0.0	0.2	0.8	0.2, 0.6, 0.8	0.3399	87	29
13	0.0	0.3	0.6	0.3, 0.3, 0.6	0.2449	165	55
14	0.0	0.3	0.7	0.3, 0.4, 0.7	0.2867	123	41
15	0.0	0.3	0.8	0.3, 0.5, 0.8	0.3300	93	31
16	0.0	0.4	0.8	0.4, 0.4, 0.8	0.3266	96	32

<sup>a</sup>Comparison pattern indicates the standardized mean difference between Group 1 vs. Group 2, Group 2 vs. Group 3, and Group 1 vs. Group 3, respectively

divided by multiple comparison per group sample size) of approximately 0.70. Stated another way, in all cases where two groups had the same mean while a third group differed, the multiple comparisons required approximately 1.4 times more cases than the omnibus test did in order to achieve power of 0.80. For example, in the condition where the pattern of means was 0.0, 0.0 and 0.5, the multiple comparisons required 1.35 times more cases than did the overall test. For 0.0, 0.0 and 0.8, the multiple comparisons resulted in 1.38 times more cases. Complete relative efficiency results from the studied conditions can be reviewed in Table 2.

Two Equal Mean Differences

In conditions where two of the three mean differences were the same and the third mean difference was twice as large, the two smaller mean comparisons required a much larger sample size than the overall test, while the third comparison required roughly the same

sample size as the omnibus test. For example, in the case where the pattern of means was 0.0, 0.3 and 0.6 (therefore a pattern of mean differences of 0.3, 0.3 and 0.6, respectively), the smaller mean comparisons required approximately 228 cases per group, while the third mean comparison required 57 cases per group. These values were compared to the omnibus test sample size of 55 cases per group for a power rate of 0.80.

Like the two similar means pattern described above, the relative efficiencies of the two similar mean differences pattern were consistent across results. In all cases where two mean differences were the same, the multiple comparison tests required approximately 4.2 times more cases than the omnibus test. For the third, different comparison, approximately 1.1 times more cases were needed. For example, in the 0.0, 0.4, 0.8 condition, the two equal multiple comparison tests (i.e., group 1 vs. group 2 and group 2 vs. group 3) required

## MCP SAMPLE SIZES

Table 2: Sample Size Results for the Tukey HSD Multiple Comparison Procedure for the Primary Monte Carlo Design at Statistical Power of 0.80

Group 1 Mean	Group 2 Mean	Group 3 Mean	Comparison Tested	Total Sample Size	Sample Size per Group	Relative Efficiency <sup>a</sup>
0	0	0.2	Omnibus	1080	360	
			G1 v G2	*	*	
			G2 v G3	1521	507	1.41
			G3 v G1	1524	508	1.41
0	0	0.3	Omnibus	483	161	
			G1 v G2	*	*	
			G2 v G3	678	226	1.40
			G3 v G1	681	227	1.41
0	0	0.4	Omnibus	276	92	
			G1 v G2	*	*	
			G2 v G3	375	125	1.36
			G3 v G1	381	127	1.38
0	0	0.5	Omnibus	180	60	
			G1 v G2	*	*	
			G2 v G3	243	81	1.35
			G3 v G1	246	82	1.37
0	0	0.6	Omnibus	123	41	
			G1 v G2	*	*	
			G2 v G3	171	57	1.39
			G3 v G1	174	58	1.41
0	0	0.7	Omnibus	93	31	
			G1 v G2	*	*	
			G2 v G3	126	42	1.35
			G3 v G1	126	42	1.35
0	0	0.8	Omnibus	72	24	
			G1 v G2	*	*	
			G2 v G3	99	33	1.38
			G3 v G1	99	33	1.38

Notes: \* indicates that the Null Hypothesis was true for the given comparison, thus no sample size analysis was performed; <sup>a</sup>Relative efficiency is calculated as the total sample size for the particular comparison divided by the total sample size for the omnibus test for the condition

BROOKS & JOHANSON

Table 2 (continued): Sample Size Results for the Tukey HSD Multiple Comparison Procedure for the Primary Monte Carlo Design at Statistical Power of 0.80

Group 1 Mean	Group 2 Mean	Group 3 Mean	Comparison Tested	Total Sample Size	Sample Size per Group	Relative Efficiency <sup>a</sup>
0	0.2	0.4	Omnibus	366	122	
			G1 v G2	1524	508	4.16
			G2 v G3	1527	509	4.17
			G3 v G1	378	126	1.03
0	0.2	0.5	Omnibus	231	77	
			G1 v G2	1524	508	6.60
			G2 v G3	690	230	2.99
			G3 v G1	246	82	1.06
0	0.2	0.6	Omnibus	156	52	
			G1 v G2	1527	509	9.79
			G2 v G3	384	128	2.46
			G3 v G1	171	57	1.10
0	0.2	0.7	Omnibus	114	38	
			G1 v G2	1515	505	13.29
			G2 v G3	246	82	2.16
			G3 v G1	126	42	1.11
0	0.2	0.8	Omnibus	87	29	
			G1 v G2	1527	509	17.55
			G2 v G3	171	57	1.97
			G3 v G1	99	33	1.14
0	0.3	0.6	Omnibus	165	55	
			G1 v G2	684	228	4.15
			G2 v G3	684	228	4.15
			G3 v G1	171	57	1.04
0	0.3	0.7	Omnibus	120	40	
			G1 v G2	675	225	5.63
			G2 v G3	384	128	3.20
			G3 v G1	126	42	1.05

Notes: \* indicates that the Null Hypothesis was true for the given comparison, thus no sample size analysis was performed; <sup>a</sup>Relative efficiency is calculated as the total sample size for the particular comparison divided by the total sample size for the omnibus test for the condition

## MCP SAMPLE SIZES

Table 2 (continued): Sample Size Results for the Tukey HSD Multiple Comparison Procedure for the Primary Monte Carlo Design at Statistical Power of 0.80

Group 1 Mean	Group 2 Mean	Group 3 Mean	Comparison Tested	Total Sample Size	Sample Size per Group	Relative Efficiency <sup>a</sup>
0	0.3	0.8	Omnibus	93	31	
			G1 v G2	678	226	7.29
			G2 v G3	246	82	2.65
			G3 v G1	99	33	1.06
0	0.4	0.8	Omnibus	93	31	
			G1 v G2	381	127	4.10
			G2 v G3	378	126	4.06
			G3 v G1	99	33	1.06

Notes: \* indicates that the Null Hypothesis was true for the given comparison, thus no sample size analysis was performed; <sup>a</sup>Relative efficiency is calculated as the total sample size for the particular comparison divided by the total sample size for the omnibus test for the condition

approximately 4.10 times more cases than the omnibus test (i.e., 127 vs. 31), while the third different mean comparison (i.e., group 1 vs. group 3) required just 33 cases, for a relative efficiency of 1.06. Very much the same results occurred for the (0.0, 0.2, and 0.4) and (0.0, 0.3, and 0.6) conditions of two similar mean differences (see Table 2).

### Absolute Mean Difference Effect Sizes

There were also consistent required sample sizes for absolute standardized group mean difference effect sizes regardless of the pattern of means, that is, regardless of the pattern of means across the three groups, the same sample size was required for any given absolute mean difference (see Table 3). For example, when examining the specific results for a comparison-of-most-interest absolute standardized mean difference of 0.3, no matter whether the pattern of means was (0.0, 0.0, 0.3) or (0.0, 0.3, 0.6) or (0.0, 0.3, 0.8), results indicated that a total sample size of approximately 681 cases (227 per group) was required to achieve a statistical power rate of 0.80 for the comparison with a standardized mean difference effect size of 0.3. Thus, when researchers have a comparison-of-most-interest

expected to be approximately 0.3, regardless of the expected effect sizes for the other possible comparisons, they would choose a total sample size of approximately 681 cases. Alternatively, if there are multiple comparisons-of-interest, then researchers in this example would choose 0.3 as the smallest among the set of most interesting comparisons and therefore choose sample sizes based on that smallest comparison-of-interest.

### Conclusion

Perhaps even more important than the sample size tables produced for this study is the notion that when a researcher is considering sample size, it may not be sufficient to set sample size for the omnibus test being performed. Clearly, researchers should consider post hoc multiple comparisons in the same way they consider different sources of effects in factorial ANOVA: that is, the most important effects under study must be considered a priori so that adequate sample sizes may be obtained for the tests of those effects. With group comparison procedures such as ANOVA, these comparisons-of-most-interest are very frequently performed using post hoc comparison procedures.

Table 3: Sample Sizes Required for Statistical Power of .80 for the Tukey HSD Multiple Comparison Procedure Given Specific Absolute Standardized Mean Differences (regardless of the pattern of group means)

Standardized Mean Difference Effect Size	Total Sample Size	Per Group Sample Size
0.2	1521	507
0.3	681	227
0.4	381	127
0.5	246	82
0.6	171	57
0.7	126	42
0.8	99	33

These results clearly show that adequate statistical power for the omnibus ANOVA  $F$  test does not guarantee adequate statistical power for given pairwise MCPs performed post hoc. This condition may result in overall statistical significance for the omnibus  $F$  test, but no pairwise comparisons showing statistical significance. Although this will occur at times because the omnibus test is reflecting that a non-pairwise comparison is significant (e.g., one group compared to an average of two other groups in an experimental study where one control group is compared to an average of two experimental treatment groups), it will happen sometimes because there is not enough power for the adjusted-alpha MCP being performed by the researcher. In the end, researchers must determine whether they wish to have sufficient power for the overall test or for the often-more-informative post hoc pairwise comparisons. The comparison-of-most-interest approach to sample size selection may be useful for the latter situation.

Results of this study suggest that it may be inappropriate to select a sample size for ANOVA based only on the omnibus test. Clearly the expected pattern among the means has an impact on the usually important post hoc pairwise multiple comparisons. This may be analogous to situations involving other statistical methods, such as principal components analysis

and MANOVA, where the pattern of correlations has an important impact on the power of the analyses, and therefore also sample size determination. Additionally, it is clear that the absolute size of the given comparison is also important. Both of these findings could be useful to researchers as they plan studies that will use ANOVA.

#### Sample Size Recommendations

Based on the results generated, certain specific recommendations can be made concerning sample sizes that researchers should use with ANOVA with three groups. It should be remembered that these results were limited to Tukey HSD comparisons performed using statistical power of 0.80. In particular, these recommendations follow from the three cases identified in the results.

#### Case 1: Two Equal Means

A researcher may be using two control groups and a single treatment group; alternatively, the researcher might expect two treatment groups each to be equally different from the single control group. In such cases, the researcher should determine the sample size required for the omnibus ANOVA test and then multiply that sample size by 1.4 to obtain the sample size required for the Tukey comparisons between the differing groups. For example, in a

## MCP SAMPLE SIZES

case where a single treatment group is expected to differ from two control groups by 0.6 (i.e., means of 0.6, 0.0 and 0.0 for the three groups, respectively), the researcher would determine that approximately 123 total cases are needed for the omnibus test to have statistical power of 0.80. If the researcher wants statistical power of 0.80 for the post hoc multiple comparisons, however, approximately  $(123 * 1.4)$  or 173 cases are needed.

### Case 2: Two Equal Mean Differences

A researcher may expect one treatment to have twice the effect of the second treatment when each is compared to the third group (e.g., a control group). In such cases, the researcher should calculate the sample size required for the omnibus test and then multiply that sample size by 4.1 to obtain the sample size required for the Tukey comparisons between the differing groups. For example, in a case where the expected pattern of means across groups is 0.0, 0.3 and 0.6, the researcher would determine that approximately 165 total cases are needed for the omnibus test to have statistical power of 0.80. If the researcher wants statistical power close to 0.80 for the post hoc multiple comparisons, however, approximately  $(165 * 4.1)$  or 677 cases are needed.

### Case 3: Absolute Mean Difference Effect Sizes

A researcher may expect that a certain pair of groups will differ by a given amount – no matter how they each differ from the third group. For example, a researcher may consider the comparison between group 1 and group 2 to be the most important and expect them to differ by a standardized mean difference of 0.5. In such a case, how much group 1 or group 2 differs from group 3 is irrelevant. Table 3 shows that 246 total cases are needed for the specific Tukey comparison between group 1 and group 2, given the expected mean difference of 0.5. In such a case, the sample size required for the omnibus test is also irrelevant, because in all cases the recommended sample sizes for the Tukey comparisons are larger than those required for the omnibus ANOVA test.

If however, the researcher expects a pattern of means that does not fit into Case 1 or Case 2 above, the absolute size of the expected

mean differences can be used with Table 3. For example, if the means for group 1, group 2, and group 3 are expected to be 0.0, 0.3 and 0.8, respectively, then (a) 681 total cases are needed for the Tukey comparison between groups 1 and 2, where the standardized mean difference is expected to be 0.3, (b) a total sample size of 99 is needed for the expected standardized difference of 0.8 between group 1 and group 3, and (c) 246 total cases are needed for the Tukey comparison between group 2 and group 3. If all three comparisons are considered equally important, the researcher would choose 681 total cases in order to have statistical power of at least 0.80 for all comparisons. However, if the comparison-of-most-interest is the group 2 versus group 3 comparison, then the 246 total cases may be the sample size selected.

### Pilot Studies and Monte Carlo Analyses

The results show that the sample size required for the omnibus  $F$  statistic to reach a given level of statistical power is frequently not sufficient for the non-null multiple comparisons to achieve the same power. In fact, it could be argued that using sample sizes chosen based on Cohen's  $f$  are inappropriate even when the study is completely exploratory and the researcher has absolutely no research hypothesis concerning the mean differences. When the work is completely exploratory, it may be even more critical to have enough statistical power to find non-null multiple comparisons, rather than simply finding that there is a difference among means somewhere.

An expected pattern of means might be available in relevant literature. However, when the relevant literature provides few clues about such effect sizes, another way to determine sample sizes for a multiple group comparison study might be to conduct a pilot study using a sampling strategy very similar to what will be used in the final study. That is, one cannot necessarily expect pilot study samples chosen conveniently to produce results similar to those obtained from representative random samples from a given population. A well-done pilot study sample, however, might provide clues to the pattern of means, the pattern of mean differences, or the absolute sizes of the mean differences the researcher might expect in the

population, thereby helping to determine what sample sizes might be necessary to have sufficient power for the post hoc comparisons. These standardized mean difference effect sizes could then be used in a Monte Carlo analysis, much as was performed for this study, to determine the necessary sample sizes for the post hoc MCPs. Because the results presented here are limited to only a few specific conditions with statistical power of 0.80, the use of Monte Carlo analyses for other circumstances may be critical because sample size tables do not exist for most multiple comparison procedures.

Finally, it is important to note that with enough evidence or knowledge about the groups, exploratory ANOVA may not be a good choice, that is, there may be times there exists enough information to estimate a group mean difference without being able to predict a directional difference between those means. In such cases, the comparison-of-most-interest approach may be useful. However, when enough information is available to make such a prediction, statistical power would be gained by using directional tests and planned contrasts in the analyses described herein.

#### Future Research

A variety of questions, both philosophical and practical, exist that might be posed for future research based on the results presented. A few suggestions are:

#### Other Procedures Designed to Control Alpha-Inflation when Multiple Tests are Performed

Although several ad hoc analyses suggested that these results might hold also for Tukey comparisons at other statistical power levels, this would need to be confirmed by further study. Similarly, some analyses performed for Bonferroni revealed the same three cases of results reported here, but would need to be examined with further study. Future research might also investigate whether similar results occur for other multiple comparison procedures (e.g., Fisher LSD, Scheffé, Dunnett). Similarly, additional research should investigate the impact of unequal sample sizes and unequal variances across groups on the total sample sizes required to achieve target levels of statistical power for specialized MCPs (e.g., Games-

Howell). Further, how this comparison-of-most-interest approach works within factorial ANOVA, as follow-up to statistically significant main effects, may also be worth investigating.

In light of other approaches that control the increase in Type I errors that occur when multiple null hypothesis tests are performed, it may be argued that perhaps MCPs should be abandoned altogether. For example, researchers could explore the effect on sample size when the Holm (1979) procedure is used (Green & Salkind, 2005; Lubrook, 1998) or when the Benjamini and Hochberg (1995) False Discovery Rate approach is used (Thissen, Steinberg & Kuang, 2002; Williams, Jones & Tukey, 1999), or perhaps no adjustment to alpha should be made for multiple comparison procedures, as is often the case when the statistical significance of regression coefficients is examined following a statistically significant regression model – this too, would impact sample size requirements.

#### Cross Validation

There are very different ways to think about how to determine required sample sizes for research; perhaps statistical power analyses are not the best way to determine sample size at all. Future research could investigate whether some adaptation of the cross-validity approaches recommended for multiple regression (e.g., Algina & Keselman, 2000; Brooks & Barcikowski, 1996; Park & Dudycha, 1974; Stevens, 2007) would be more useful for researchers in group comparison studies. The basic idea behind the cross-validation approaches is that researchers would be more likely to find results, especially effect sizes, that will replicate if sample sizes are large enough for cross-validation.

#### A Priori Contrasts and *t* Tests

Future researchers could compare these results to multiple individual *t* tests or other planned comparisons performed as a priori contrasts when using either an adjusted or unadjusted alpha. It may be that MCP sample sizes are functionally related to *t* test sample sizes using a relative efficiency approach similar to that done in this study. Future researchers might investigate whether the results change if



only a subset of more important pairwise comparisons are performed (e.g., simple or repeated contrasts), instead of all possible pairwise comparisons. Similar analyses might also be performed for common non-pairwise comparisons, such as Helmert or polynomial contrasts.

#### Relative Efficiency

Although no function emerged for some mean difference patterns in the three-group analyses, there may be a less obvious function at work. One could study how well relative efficiency works with larger numbers of groups, with effect sizes larger or smaller than those investigated here, and with different statistical power targets than 0.80. A similar study with four or more groups would involve many more possible mean difference patterns, but could help to provide answers to some of these questions. Such a study would also verify whether such results occur with more than three groups. Finally, the present study can be modified to include non-normal data and different sample sizes in each group.

#### References

- Algina, J., & Keselman, H. J. (2000). Cross-validation sample sizes. *Applied Psychological Measurement, 24*, 173–179.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57*, 289-300.
- Brooks, G. P. (2008). *MC4G: Monte Carlo Analyses for up to 4 Groups*. [Computer software and manuals]. Retrieved from {<http://www.ohio.edu/people/brooksg/software.htm>}.
- Brooks, G. P., & Barcikowski, R. S. (1999, April). *The precision efficacy analysis for regression sample size method*. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada (ERIC Document Reproduction Service No. ED 449 177).
- Carmer, S. G., & Swanson, M. R. (1973). An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association, 68*, 66-74.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2<sup>nd</sup> Ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Einot, I., & Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. *Journal of the American Statistical Association, 70*, 574-583.
- Green, S. B., & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analyzing and understanding data (4<sup>th</sup> Ed.)*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences (5<sup>th</sup> Ed.)*. Boston: Houghton Mifflin.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.
- Hopkins, K. D., & Hopkins, B. R. (1979). The effect of the reliability of the dependent variable on power. *Journal of Special Education, 13*, 463-466.
- L'Ecuyer, P. (1988). Efficient and portable combined random number generators. *Communications of the ACM, 31*, 742-749, 774.
- Levin, J. R. (1975). Determining sample size for planned and post hoc analysis of variance comparisons. *Journal of Educational Measurement, 12*, 99-108.
- Light, R. J., Singer, J. D., & Willett, J. B. (1990). *By design: Planning research on higher education*. Cambridge, MA: Harvard University.
- Ludbrook, J. (1998). Multiple comparison procedures updated. *Clinical and Experimental Pharmacology and Physiology, 25*, 1032–1037.
- Pan, X., & Dayton, C. M. (2005). Sample Size Selection for Pair-Wise Comparisons Using Information Criteria. *Journal of Modern Applied Statistical Methods, 4*(2), 601-608.
- Park, C. N., & Dudycha, A. L. (1974). A Cross-Validation Approach to Sample Size Determination for Regression Models. *Journal of the American Statistical Association, 69*, 214-218.

Park, S. K., & Miller, K. W. (1988). Random number generators: Good ones are hard to find. *Communications of the ACM*, *31*, 1192-1201.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical recipes in Pascal: The art of scientific computing*. New York: Cambridge University.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing (2<sup>nd</sup> Ed.)*. New York: Cambridge University.

Stevens, J. (2007). *Intermediate statistics: A modern approach (3<sup>rd</sup> Ed.)*. New York: Lawrence Erlbaum Associates.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, *27*, 77-83.

Toothaker, L. E. (1991). *Multiple comparisons for researchers*. Newbury Park: Sage.

Williams, V. S. L., Jones, L. V., & Tukey, J.W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, *24*, 42-69.

## Weighting Large Datasets with Complex Sampling Designs: Choosing the Appropriate Variance Estimation Method

Sara Mann  
University of Guelph,  
Guelph, Ontario Canada

James Chowhan  
McMaster University,  
Hamilton, Ontario Canada

---

Using the Canadian Workplace and Employee Survey (WES), three variance estimation methods for weighting large datasets with complex sampling designs are compared: simple final weighting, standard bootstrapping and mean bootstrapping. Using a logit analysis, it is shown - depending on which weighting method is used - different predictor variables are significant. The potential lack of independence inherent in a multi-stage cluster sample design, as in the WES, results in a downward bias in the variance when conducting statistical inference (using the simple final weight), which in turn results in increased Type I errors. Bootstrap methods can account for the survey's design and adjust the variance so that it is inference appropriate and corrected for downward bias. The WES provides mean, as opposed to standard, bootstrap weights with the data; thus, a further adjustment to account for the reduced variation inherent when information is grouped is required. Failure to use mean bootstrap weights appropriately leads to biased standard errors and inappropriate inference.

Key words: Bootstrap, variance estimation, complex sampling design.

---

### Introduction

Choosing the appropriate variance estimation method when weighting large datasets with complex sampling designs has important implications for researchers. Using the Canadian Workplace and Employee Survey (WES), three variance estimation methods for weighting large datasets with complex sampling designs are compared: simple final weight, standard bootstrapping and mean bootstrapping. This study

uses a logit analysis to show that, depending on which weighting method was used, different predictor variables are significant. Failure to use the mean bootstrap weights appropriately can lead to both biased standard errors and inappropriate inference.

### Survey Instrument

A national workplace survey, the WES, conducted over a five year period from 1999-2003 by Statistics Canada was used in this study. Stratified sampling was used for WES and up to twenty four employees were surveyed within each workplace, depending on the establishment's size. In-person interviewers collected the workplace survey data and telephone interviews were conducted with the employees. The WES is unique in that employers and employees are linked at the micro data level and employees are selected from within sampled workplaces (Statistics Canada, 2003). The number of employers included in the sample was 6,322 in 1999, 6,068 in 2000, 6,207 in 2001, 5,818 in 2002 and 6,565 in 2003. The number of employees in the sample was 23,540 in 1999, 20,167 in 2000, 20,352 in 2001, 16,813

---

Sara Mann is an Assistant Professor of Strategic Human Resource Management and Organizational Behavior in the Department of Business. Her research interests include HRM practices such as performance appraisal and selection, specifically for lower wage workers and those in the agriculture sector. Email her at: [smann@uoguelph.ca](mailto:smann@uoguelph.ca). James Chowhan is a Ph.D. Candidate in the DeGroote School of Business at McMaster University, and has a Master's degree in Economics. Email him at: [chowhan@mcmaster.ca](mailto:chowhan@mcmaster.ca).

in 2002 and 20,834 in 2003 (with a survey response rate of 95% for locations and 83% of workers in 1999). Employers are followed for five years and employees are followed for two years.

The WES uses a multi-stage cluster design to select a sample of respondents. This results in respondents being sampled from the same cluster implying that they are not necessarily independent. This potential lack of independence results in a downward bias in the variance when conducting statistical inference (using the simple final weight). Downward bias results in an increase of Type I errors, rejecting the null when it is true. Bootstrap methods can account for the survey's design and adjust the variance so that it is inference appropriate and corrected for the downward bias. The WES provides mean bootstrap weights with the data, as opposed to standard bootstrap weights; thus, a further adjustment to account for the reduced variation inherent when information that is grouped is required.

The results presented in this article are from a study conducted by Mann & Latham (2008), which examined the predictors of the receipt of a performance appraisal. The variables that were significant predictors differed depending on which variance estimation method was used.

#### Variables

Five organization-level predictor variables were included in the analysis: size (operationalized as the number of employees), industry (service = 1, manufacturing = 0), for-profit (for-profit = 1, not-for-profit = 0), unionized (yes=1, no=0) and an in-house HR department (yes=1, no=0).

Several job-level predictor variables were also included in the analysis: hourly wage, four dummy variables representing whether the job is full-time or part-time and permanent or temporary (full-time/permanent, full-time/temporary, part-time/permanent, part-time/temporary), dummy variables representing occupation (professional, manager, technical/trades, marketing, clerical and production) and a dichotomous variable measuring the use of a computer in the job (yes=1, no=0).

Individual-level variables were also included in the analysis: age, gender (1=female, 0=male), recent immigrant (within the last 5 years) (1=yes, 0=no), and disability (1=yes, 0=no). One dependent variable was used, the receipt of a performance appraisal (1=yes, 0=no). To reduce common method bias, the organization variables were drawn from the employer survey, while the job- and individual-level variables, as well as the dependent variables were drawn from the employee survey.

#### Methodology

Descriptive statistics and correlations were presented in the Mann & Latham (2008) study but, because they are not relevant to this study, they are not discussed. The stepwise logit regression that was conducted (with the organization-level variables included in the first step, and the job-level and individual-level variables added in the next two steps) is of particular interest to this study. Three different regressions were conducted using different types of variance estimation methods: simple final weighting, standard bootstrapping and mean bootstrapping. The choice of method has important implications for the inference of the significance of the predictor variables.

#### Mean Bootstrap Comparison

Bootstrap weights are used to make use of complex survey design information and to calculate reliable design-based variance estimates. Generally, Statistics Canada uses a multistage, stratified, randomly selected cluster sample or complex design to draw a representative sample of respondents. Within the WES, workplaces from business locations operating in Canada are selected from relatively homogeneous strata (industry, region and size groupings). In addition, employers must have paid employees; exceptions are locations in the Yukon, Nunavut, the Northwest Territories, agricultural operations, private households, religious organizations or public administration. This results in respondents not necessarily being independent – respondents in the same cluster may share similar economic characteristics as a group relative to the population as a whole. This is a disadvantage of cluster sampling and results

in less efficient estimates (Satin & Shastry, 1993).

To enable researchers to correct for this downward bias in the variance, Statistics Canada has included bootstrap weights with the WES. Bootstrap techniques have been used to generate a set of 5,000 bootstrap weights, which give a more reasonable estimate of variance than estimation that does not account for the complex design of the survey. Statistics Canada generates bootstrap weights by randomly drawing samples with replacement from each stratum of primary sampling units. The size of each sample drawn is equal to the sample size of the data set. Using the same clustering and sample design the weights are assigned to each unit in the selected random draws; selected units receive a positive bootstrap weight and units not selected receive a weight of zero.

For WES this sampling is replicated 5,000 times to generate a set of bootstrap weights large enough to be consistent and allow for the calculation of average bootstrap weights. Further, for the WES data, the bootstrap weights that have been provided are average bootstrap weights. In other words, a set of 100 (B) average bootstrap weights have been calculated over groups of 50 (C) from the original set of 5,000 bootstrap weights. Average bootstrap weights were calculated to preserve the confidentiality of workplace's responses. Using the WES mean bootstrap weights requires a further adjustment to account for the reduced variation inherent when using grouped information. The variance estimator used to calculate the design-based variance estimate with mean bootstrap weights is:

$$v_{\bar{B}}(\hat{\theta}) = \frac{C}{B} \sum_b (\hat{\theta}_{(b)}^* - \hat{\theta}_{(.)}^*)^2$$

where

$$\hat{\theta}_{(.)}^* = \left(\frac{1}{B}\right) \sum_b \hat{\theta}_{(b)}^* \quad (1)$$

Each  $b^{\text{th}}$  average bootstrap sample set of weights is equal to the means of C bootstrap weights. In this specification, the term  $\hat{\theta}_{(b)}^*$  is obtained using the  $b^{\text{th}}$  mean bootstrap weight variable (Buckley & Chowhan, 2005).

The analysis herein used Stata 9, and specifically, the survey suite of commands to estimate the results. The advantage of these commands is that the final weight (used to generate the point estimates or parameters), the bootstrap weights (used to generate the standard errors), and the variance estimation method (balanced repeated replication) can all be specified using the svyset command. When each piece of analysis is run the adjustment for the mean bootstrap is made by using the fay(.85857864376269) option. This adjustment comes from the Fay's variance estimator, where K could be set equal to

$$K = 1 - \sqrt{\frac{1}{C}},$$

which is a transformation of equation 1, given that the Fay's variance estimator is as follows

$$v_{Fay}(\hat{\theta}) = \frac{1}{T(1-K)^2} \sum_t (\hat{\theta}_{(t)}^* - \hat{\theta}_{(.)}^*)^2$$

where

$$\hat{\theta}_{(.)}^* = \left(\frac{1}{T}\right) \sum_t \hat{\theta}_{(t)}^*$$

The use of this adjustment re-introduces the variability that had been removed when the average bootstrap weights were generated.

### Results

Descriptive statistics are presented in Table 1 and the predictors of the receipt of a performance appraisal are presented in Table 2. For a discussion on the predictors of the receipt of a performance paper, see the Mann & Latham (2008) paper; this study is only concerned with the variance estimation method used to produce the results.

Comparing the three variance estimation methods in Table 2 different predictor variables were significant depending on which method was used. The results under column 3 should be used as they present the findings from the analysis using the mean bootstrapping method and produce the most accurate, unbiased standard errors. Due to this finding, this article adds a significant methodological contribution to

Table 1: Means and Standard Deviations

	n	Mean	SD
<b>Organization-Level Variables</b>			
Industry	20834	.67	.47
In-house HR Dept	20362	.39	.49
Number of Employees	20362	414.23	1085.35
For-Profit	20362	.80	.40
Unionized	20834	.26	.44
<b>Job-Level Variables</b>			
Wage	20619	20.60	12.75
Full-time/perm	20619	.59	.49
Full-time/temp	20619	.02	.13
Part-time/perm	20619	.34	.47
Part-time/temp	20619	.05	.22
Use Computer	20834	.65	.48
<b>Occupation</b>			
Professional	20834	.16	.37
Manager	20834	.13	.33
Technical/Trade	20834	.41	.49
Marketing	20834	.08	.27
Clerical	20834	.15	.35
Production	20834	.07	.25
<b>Individual-Level Variables</b>			
Gender	20834	.53	.50
Age	20834	40.24	11.52
Recent Immigrant	20834	.03	.15
Disability	20834	.09	.29
<b>Dependent Variables</b>			
Receipt of PA	20834	.60	.49

our field by comparing these different weighting methods.

Although the coefficients are the same for all three methods presented in Table 2, which predictors are significant differs depends on which method is presented. Three related approaches are shown: column (1) shows significance levels when only the final weight is used to generate the standard errors, column (2) illustrates how the standard errors are downward

bias when the mean bootstrap weights are not accounted for, and column (3) presents the reliable standard errors from the correct use of mean bootstrap weights and the appropriate adjustments.

It is important to note that column 2's standard errors are generally the lowest, followed by columns 1 and 3. Column 3 presents the reliable design-based variance

## WEIGHTING LARGE DATASETS

Table 2: LOGIT Results: Predictors of the Receipt of a Performance Appraisal\*

Variables	Coeff.	Column 1: Weighting Method		Column 2: Bootstrapping Method		Column 3: Mean Bootstrapping Method	
		Std. Error	Sig.	Std. Error	Sig.	Std. Error	Sig.
<b>Org Variables</b>							
In House HR Dept	.53	.0713	p<.001	.0147	p<.001	.0140	p<.001
Number of Employees	.00	.0000		.0000	p<.001	.0000	
Unionized	-.25	.0797	p<.05	.0148	p<.001	.1048	p<.05
Service Industry	.12	.0787		.0134	p<.001	.0949	
For-Profit	-.10	.0994		.0182	p<.001	.1284	
<b>Job Variables</b>							
Wage	.01	.0037	p<.05	.0006	p<.001	.0041	p<.05
Full-time/Permanent	.66	.1711	p<.001	.0217	p<.001	.1538	p<.001
Full-time/Temporary	.06	.2562		.0351	p<.10	.2480	
Part-time/Permanent	.61	.1721	p<.001	.0241	p<.001	.1705	p<.001
Use Computer	.64	.0820	p<.001	.0127	p<.001	.0901	p<.001
Professional	.41	.2084	p<.05	.0323	p<.001	.2282	p<.10
Manager	.09	.2124		.0295	p<.05	.2084	
Technical/Trades	.14	.1853		.0259	p<.001	.1828	
Marketing	-.16	.2433		.0290	p<.001	.2053	
Clerical	-.08	.2027		.0295	p<.05	.2085	
<b>Individual Variables</b>							
Gender (Female)	.07	.0761		.0113	p<.001	.0797	
Age	.01	.0190		.0033	p<.10	.0236	
Age <sup>2</sup>	.00	.0002		.0000	p<.001	.0003	
Recent Immigrant	.27	.2549		.0299	p<.001	.2114	
Disability	-.11	.1138		.0154	p<.001	.1089	
Constant	-1.07	.4593	p<.05	.0761	p<.001	.5378	p<.05
R <sup>2</sup>	.06						

\*N=20,834; Reference Groups: Part-time/Temporary and Production

estimates. In column 3, the explanatory variables (in-house HR department, unionization, hourly wage, being full-time/permanent, being part-time/permanent, the use of a computer and professional occupation) are statistically significant at the 95% level. Compared to column 1 the significance of professional occupation as a predictor is more accurate and,

without the appropriate variance estimation the conclusions drawn from the inference, would have been inaccurate. Further, when no adjustments are made for the WES provided mean bootstrap weights (column 2) all variances are underestimated by a factor of C resulting in output that leads to inappropriate inference for all variables. All predictors are significant using

this method, when only those predictors significant in column 3 should be interpreted as such. Thus, failing to use bootstrap weights and the mean bootstrap weights appropriately lead to biased standard errors and inappropriate inference.

#### Conclusion

The results of this study portend a significant methodological contribution with respect to choosing the appropriate variance estimation method when using a large dataset, such as the WES. Although the beta coefficients are the same for all three methods, which predictors are significant differs depending on the method used. When presenting findings from a large dataset and a complex sampling design, the variance estimation method that was used should be acknowledged. Readers should be aware that different results can be presented depending on the method selected. This suggests that researchers should be cautious when choosing a weighting method and be aware of the biased standard errors that are produced when the inappropriate method is used. This study showed the practical implication of choosing an appropriate, unbiased weighting method when analyzing a large dataset with complex sampling design.

#### References

- Buckley, N., & Chowhan, J. (2005). Using mean bootstrap weights in Stata: A BSWREG revision. *The Research Data Centres Information and Technical Bulletin, Spring (2)*, 23-37. Statistics Canada Catalogue No. 12-002-XIE.
- Mann, S. L. & Latham, G. P. (2008, May). –Who receives a performance appraisal and does it matter? –An empirical investigation of the determinants of the receipt of a performance appraisal and its effect on job satisfaction. Paper presented at the Annual Meeting of the Administrative Sciences Association of Canada. Halifax, Canada.
- Satin, A., & Shastry, W. (1993). *Survey sampling: A non-mathematical guide*. Ottawa: Ministry of Industry, Statistics Canada, Social Survey Methods Division. Catalogue No. 12-602-XPE.
- Statistics Canada. (2003). Guide to the analysis of workplace and employee survey 2001. *Business and Labour Market Analysis Division & Labour Statistics Division, June*. Ottawa: Canada.
- Yeo, D., Mantel, H., & Liu, T. (1999). Bootstrap variance estimation for the national population health survey. *American Statistical Association: Proceedings of the Survey Research Methods Section*. Baltimore, MD, August.



## Double Acceptance Sampling Plans Based on Truncated Life Tests for Marshall-Olkin Extended Lomax Distribution

G. Srinivasa Rao  
Dilla University,  
Dilla, Ethiopia

---

Double Acceptance Sampling Plans (DASP) is developed for a truncated life test when the lifetime of an item follows the Marshall-Olkin extended Lomax distribution. Probability of Acceptance (PA) is calculated for different consumer's confidence levels fixing the producer's risk at 0.05. Probability of acceptance and producer's risk are illustrated with examples.

Key words: Marshall-Olkin extended Lomax distribution, double acceptance sampling plan, probability of acceptance, consumer's risk, producer's risk, truncated life test.

---

### Introduction

The main goal of competitive enterprises in a global business market is how to maintain and improve the quality of their products. A high quality product has the high probability of acceptance. Two important tools for ensuring quality are statistical quality control and acceptance sampling (AS). Acceptance sampling plans are concerned with accepting or rejecting a submitted large sized lot of products on the basis of the quality of the products inspected in a small sample taken from the lot.

A single acceptance sampling plan (SASP) is a specified plan that establishes the minimum sample size to be used for testing. In most acceptance sampling plans for a truncated life test, the major issue is to determine the sample size from a lot under consideration. It is implicitly assumed in the usual sampling plan that only a single item is put in a tester. On the basis of information obtained from this first sample a lot is either accepted or rejected. If a

good lot is rejected on the basis of this information, its probability is called the type-1 error (producer's risk) and it is denoted by  $\alpha$ . The probability of accepting a bad lot is known as the type-2 error (consumer's risk) and it is denoted as  $\beta$ . If the product is electronic components or has failure mechanisms a random sample of the lot is tested and the entire lot is accepted if no more than  $c$  (acceptance sampling number) failures occur during the experiment time. Recently, Aslam (2007) proposed the double acceptance sampling plan based on a truncated life test when the lifetime of an item follows the Rayleigh distribution.

Single acceptance sampling plans based on truncated life tests for a variety of distributions are discussed by Epstein (1954), Sobel and Tischendorf (1959), Goode and Kao (1961), Gupta and Groll (1961), Gupta (1962), Kantam and Rosaiah (1998), Kantam, et al. (2001), Baklizi (2003), Baklizi and El Masri (2004), Rosaiah and Kantam (2005), Rosaiah, et al. (2006, 2007, 2007), Tsai and Wu (2006), Balakrishnan, et al. (2007), Aslam and Kantam (2008) and Rao, et al. (2008, 2009).

This article proposes a double acceptance sampling plan (DASP) based on truncated life tests when the lifetime of a product follows the Marshall-Olkin extended Lomax distribution with known shape parameter as introduced by Ghitany, et al. (2007) and to determine the probability of acceptance (PA).

---

G. Srinivasa Rao is a Professor of Statistics in the Department of Statistics, School of Mathematical & Computer Sciences, Dilla University, Dilla, P. O. Box: 419, Ethiopia. His research interests include statistical inference, quality control and reliability estimation Email him at: gaddesrao@yahoo.com.

The probability density function (pdf) and cumulative distribution function (cdf) of the Marshall-Olkin extended Lomax distribution are given by

$$g(t; \nu, \sigma, \theta) = \frac{\nu \theta (1+t/\sigma)^{\theta-1}}{\left[ (1+t/\sigma)^{\theta} - \bar{\nu} \right]^2},$$

$$t > 0, \nu, \sigma, \theta > 0, \bar{\nu} = 1 - \nu \quad (1.1)$$

and

$$G_T(t; \nu, \sigma, \theta) = \frac{(1+t/\sigma)^{\theta} - 1}{\left[ (1+t/\sigma)^{\theta} - \bar{\nu} \right]},$$

$$t > 0, \nu, \sigma, \theta > 0, \bar{\nu} = 1 - \nu \quad (1.2)$$

respectively, where  $\sigma$  is a scale parameter,  $\theta$  is a shape parameter and  $\nu$  is an index parameter. The mean of this distribution is given by  $\mu = 1.570796 \sigma$  when  $\nu = 2, \theta = 2$ . Rao, et al. (2008, 2009) studied single acceptance sampling plans based on the Marshall-Olkin extended Lomax distribution.

The Double Acceptance Sampling Plan (DASP) for Life Tests

DASP is used to minimize the producer's risk because it provides another opportunity for acceptance of the product. In DASP a sample of size  $n_1$  items is taken from a lot which is called the sample first; this sample first is then put on tests. Let  $c_1$  and  $c_2$  be the acceptance number for sample first and the sample second respectively. The experiment is terminated if no more than  $c_1$  failures occur during the experiment time ( $t_0$ ), i.e., the lot is accepted or rejected on the basis of sample first if more than  $c_2$  failures occur or if time of experiment ends (whichever is earlier). If  $(c_1 + 1)^{\text{th}}$  failures occur in sample first, then all possibilities for sample second are given as shown in Table 1.

Table 1: Possibilities for Sample Second Failure Based on Sample First Failure

Sample First	Sample Second
$(c_1 + 1)^{\text{th}}$ failures occur in sample 1	less than $(c_2 - 1)^{\text{th}}$ must be occurred in this sample 2 for acceptance
$(c_1 + 2)^{\text{th}}$ failures occur in sample 1	less than $(c_2 - 2)^{\text{th}}$ must be occurred in this sample 2 for acceptance
$(c_1 + n)^{\text{th}}$ failures occur in sample 1	less than $(c_2 - n)^{\text{th}}$ must be occurred in this sample for acceptance

Let  $\mu$  represent the true average life of a product and  $\mu_0$  denote the specified life of an item, under the assumption that the lifetime of an item follows the Marshall-Olkin extended Lomax distribution. A product is considered good and accepted for consumer use if the sample information supports the hypothesis  $H_0 : \mu \geq \mu_0$ ; if the sample does not support this hypothesis, the lot of product is rejected. In acceptance sampling schemes, this hypothesis is tested based on the number of failures from a sample in a pre-fixed time: If the number of failures exceeds the action limit  $c$  the lot is rejected. The lot will be only be accepted if there is enough evidence that  $\mu \geq \mu_0$  at a certain level of consumer risk, otherwise the lot will be rejected.

When determining the parameters of a proposed sampling plan the consumer's risk is used, and often, the consumer's risk is expressed by the consumer's confidence level. If the confidence level is  $p^*$ , then the consumer's risk will be  $\beta = 1 - p^*$ . In this study the consumer's risk was fixed not to exceed  $1 - p^*$  and it satisfied the inequality:

$$\sum_{i=0}^c \binom{n}{i} p^i (1-p)^{n-i} \leq 1 - p^* \quad (2.1)$$

where  $p$  is the probability that an item fails before the termination time.

Consider a life testing experiment having the measurements: number of items from sample first put on test ( $n_1$ ), acceptance number for sample first ( $c_1 = 0$ ), number of items from sample second put on test ( $n_2$ ) and accept the lots if no more than two failures occur in sample second ( $c_2 = 2$ ).

In this life experiment if no failure occurs when sample first of  $n_1$  items is put on test the lot is accepted. If the true - but unknown - life of the product deviates from the specified life of the product it should result in a considerable change in the probability of acceptance of the lot based on the sampling plan. Hence, the probability of acceptance (PA) can be regarded as a function of the deviation of a specified average from the true average. This function is called the operating characteristic (OC) function of the sampling plan. The PA for sample first using the Marshall-Olkin extended Lomax distribution with  $\nu = 2, \theta = 2$  is shown in Table 2. The probability of acceptance for  $L(p_1)$

and  $L(p_2)$  for the sampling plan  $(n_1, c_1, \frac{t}{\sigma_0})$

and  $(n_2, c_2, \frac{t}{\sigma_0})$  is calculated using equations

$$L(p_1) = \sum_{i=0}^{c_1=0} \binom{n_1}{i} p^i (1-p)^{n_1-i} \quad (2.2)$$

and

$$L(p_2) = \sum_{i=0}^{c_2=2} \binom{n_2}{i} p^i (1-p)^{n_2-i} \quad (2.3)$$

respectively, where

$$p = G_T(t; \sigma) = G_T\left(\frac{t}{\sigma_0} \frac{\sigma_0}{\sigma}\right),$$

is given in (1.2).

The probability of acceptance for DASP can be obtained by using (2.2) and (2.3) as:

$$\begin{aligned} P(A) &= P(\text{no failures occur in sample 1}) \\ &+ P(1 \text{ failure occurs in samples 1 and 0,} \\ &\quad \text{and 1 failure occurs in sample 2}) \\ &+ P(2 \text{ failures occur in sample 1,} \\ &\quad \text{and 0 failures occur in sample 2}). \end{aligned}$$

The values of probability of acceptance for DASP are determined at  $p^* = 0.75, 0.90, 0.95, 0.99$  and  $t/\sigma_0 = 0.628, 0.942, 1.257, 1.571, 2.356, 3.142, 3.927, 4.712$  with  $\nu = 2, \theta = 2$  and are shown in Table 3. It is important to note that in sample first and sample second,  $p$  is function of the cdf of the Marshall-Olkin extended Lomax distribution. These choices are consistent with Gupta and Groll (1961), Gupta (1962), Kantam, et al. (2001), Baklizi and EI Masri (2004), Balakrishnan, et al. (2007) and Rao, et al. (2008).

### Results

Suppose that the lifetime of a product follows the Marshall-Olkin extended Lomax distribution with  $\nu = 2, \theta = 2$  and an experimenter wants to establish that the true unknown mean life is at least 1,000 hours with confidence 0.75. The acceptance numbers for this experiment would be  $c_1 = 0$  and  $c_2 = 2$  with sample sizes  $n_1 = 6$  and  $n_2 = 8$ . The lot will be accepted if no failure is observed in a sample of 6 during 628 hours. The probability of acceptance for this single sampling from Table 2 is 0.1558. The PA for the same measurements using double acceptance sampling from Table 3 is 0.29276. In the DASP scheme as  $\sigma / \sigma_0$  increases PA also increases.

For the above sampling plan, PA is 0.97289 when the ratio of unknown average life to specified average life is 12. As the time of experiment increases, the probability of acceptance for the double acceptance sampling plan decreases. From Table 3, it is clear that when the time of experiment is 4,712 hours, the PA for ratio  $\sigma / \sigma_0 = 2$  is 0.04906. For this same experiment time, as  $\sigma / \sigma_0$  increases, PA also increases. It is important to note that the double acceptance sampling scheme minimizes the

Table 2: Operating Characteristics Values for Sample First for the Sampling Plan  $(n_1, c_1, \frac{t}{\sigma_0})$

when  $c_1 = 0$  Marshall-Olkin Extended Lomax Distribution with  $\nu = 2, \theta = 2$

$p^*$	$n_1$	$t / \sigma_0$	$\sigma / \sigma_0$					
			2	4	6	8	10	12
0.75	6	0.628	0.15576	0.39119	0.53422	0.62466	0.68622	0.73062
	4	0.942	0.15968	0.39271	0.53488	0.62500	0.68641	0.73074
	4	1.257	0.08995	0.28921	0.43485	0.53462	0.60556	0.65817
	3	1.571	0.10891	0.31489	0.45928	0.55663	0.62527	0.67589
	3	2.356	0.04222	0.18266	0.31496	0.41760	0.49594	0.55670
	2	3.142	0.06907	0.22806	0.36256	0.46285	0.53784	0.59527
	2	3.927	0.04180	0.16475	0.28637	0.38514	0.46289	0.52449
	2	4.712	0.02660	0.12125	0.22812	0.32193	0.39946	0.46292
0.90	8	0.628	0.08381	0.28610	0.43348	0.53398	0.60527	0.65804
	6	0.942	0.06381	0.24610	0.39119	0.49411	0.56869	0.62466
	5	1.257	0.04926	0.21209	0.35312	0.45715	0.53420	0.59282
	4	1.571	0.05201	0.21423	0.35435	0.45789	0.53467	0.59315
	3	2.356	0.04222	0.18266	0.31496	0.41760	0.49594	0.55670
	3	3.142	0.01815	0.10891	0.21831	0.31489	0.39444	0.45928
	3	3.927	0.00855	0.06687	0.15325	0.23902	0.31493	0.37985
	3	4.712	0.00434	0.04222	0.10896	0.18266	0.25247	0.31496
0.95	9	0.628	0.06147	0.24467	0.39047	0.49371	0.56845	0.62451
	7	0.942	0.04034	0.19482	0.33455	0.43933	0.51763	0.57755
	5	1.257	0.04926	0.21209	0.35312	0.45715	0.53420	0.59282
	5	1.571	0.02484	0.14575	0.27339	0.37666	0.45720	0.52054
	4	2.356	0.01470	0.10364	0.21430	0.31214	0.39256	0.45796
	3	3.142	0.01815	0.10891	0.21831	0.31489	0.39444	0.45928
	3	3.927	0.00855	0.06687	0.15325	0.23902	0.31493	0.37985
	3	4.712	0.00434	0.04222	0.10896	0.18266	0.25247	0.31496
0.99	12	0.628	0.02426	0.15303	0.28540	0.39020	0.47089	0.53381
	9	0.942	0.01612	0.12208	0.24467	0.34732	0.42886	0.49371
	7	1.257	0.01477	0.11406	0.23286	0.33426	0.41570	0.48094
	6	1.571	0.01186	0.09916	0.21093	0.30984	0.39096	0.45682
	5	2.356	0.00512	0.05880	0.14580	0.23331	0.31072	0.37673
	4	3.142	0.00477	0.05201	0.13145	0.21423	0.28927	0.35435
	4	3.927	0.00175	0.02714	0.08201	0.14833	0.21427	0.27509
	4	4.712	0.00071	0.01470	0.05204	0.10364	0.15957	0.21430

DASP FOR MARSHALL-OLKIN EXTENDED LOMAX DISTRIBUTED LIFE TESTS

Table 3: Operating Characteristics Values for the Double Sampling Plan  $(n_2, c_2, \frac{t}{\sigma_0})$

when  $c_1 = 0$  and  $c_2 = 2$  Marshall-Olkin Extended Lomax Distribution with  $\nu = 2, \theta = 2$

$p^*$	$n_1$	$n_2$	$t / \sigma_0$	$\sigma / \sigma_0$					
				2	4	6	8	10	12
0.75	6	8	0.628	0.29276	0.50105	0.86425	0.92779	0.95741	0.97289
	4	6	0.942	0.28692	0.49834	0.85658	0.92304	0.95436	0.97084
	4	5	1.257	0.18319	0.38854	0.79087	0.88279	0.92856	0.95349
	3	5	1.571	0.17606	0.39214	0.76899	0.86779	0.91832	0.94632
	3	4	2.356	0.07833	0.24442	0.63350	0.77292	0.85228	0.89938
	2	4	3.142	0.09425	0.27216	0.61281	0.75306	0.83604	0.88665
	2	3	3.927	0.08244	0.22838	0.59443	0.73854	0.82510	0.87845
	2	3	4.712	0.04906	0.16496	0.49061	0.65090	0.75619	0.82513
0.90	8	10	0.628	0.14865	0.36654	0.76759	0.86826	0.91912	0.94708
	6	8	0.942	0.10420	0.31013	0.70705	0.82753	0.89150	0.92779
	5	6	1.257	0.09101	0.28114	0.68727	0.81386	0.88208	0.92116
	4	6	1.571	0.07686	0.26266	0.63949	0.77821	0.85639	0.90249
	3	5	2.356	0.05706	0.21770	0.56209	0.71656	0.81006	0.86785
	3	4	3.142	0.02910	0.14104	0.46313	0.63339	0.74510	0.81789
	3	4	3.927	0.01211	0.08352	0.32825	0.50281	0.63346	0.72651
	3	4	4.712	0.00560	0.05094	0.22930	0.39104	0.52766	0.63350
0.95	9	12	0.628	0.09469	0.30286	0.69561	0.81978	0.88620	0.92407
	7	9	0.942	0.06148	0.24378	0.62870	0.77176	0.85233	0.89979
	5	7	1.257	0.07461	0.26273	0.64649	0.78411	0.86090	0.90588
	5	6	1.571	0.04055	0.18977	0.56103	0.71960	0.81391	0.87145
	4	5	2.356	0.02212	0.13247	0.45228	0.62671	0.74100	0.81528
	3	5	3.142	0.02157	0.12425	0.38835	0.56197	0.68478	0.76899
	3	4	3.927	0.01211	0.08352	0.32825	0.50281	0.63346	0.72651
	3	4	4.712	0.00560	0.05094	0.22930	0.39104	0.52766	0.63350
0.99	12	16	0.628	0.03081	0.18223	0.52412	0.69017	0.79171	0.85476
	9	11	0.942	0.02141	0.15000	0.48085	0.65448	0.76426	0.83390
	7	9	1.257	0.01890	0.13770	0.45168	0.62824	0.74308	0.81729
	6	8	1.571	0.01449	0.11744	0.40429	0.58423	0.70674	0.78834
	5	6	2.356	0.00648	0.07208	0.30745	0.48783	0.62337	0.71970
	4	6	3.142	0.00516	0.05738	0.23006	0.39485	0.53317	0.63949
	4	5	3.927	0.00200	0.03154	0.16512	0.31441	0.45222	0.56520
	4	5	4.712	0.00077	0.01645	0.09793	0.21434	0.33887	0.45228

producer's risk, but this scheme also exerts pressure on the producer to improve the quality level of the product. At 4,712 hours,  $\sigma / \sigma_0 = 12$ ,  $p^* = 0.75$  and the PA is 0.82513. The producer's risk for the sample first and double sampling are placed for  $p^* = 0.75$  in Table 4. For  $\sigma / \sigma_0 = 2$  (the unknown average life is twice that of the specified average life), the producer's risk when time of experiment is 628 hours and 4,712 hours are 0.70724 and 0.95094 respectively. Thus, the

producer's risk decreases as the quality level of the product increases with  $p^* = 0.75$ . (See Table 4 and Figure 1.)

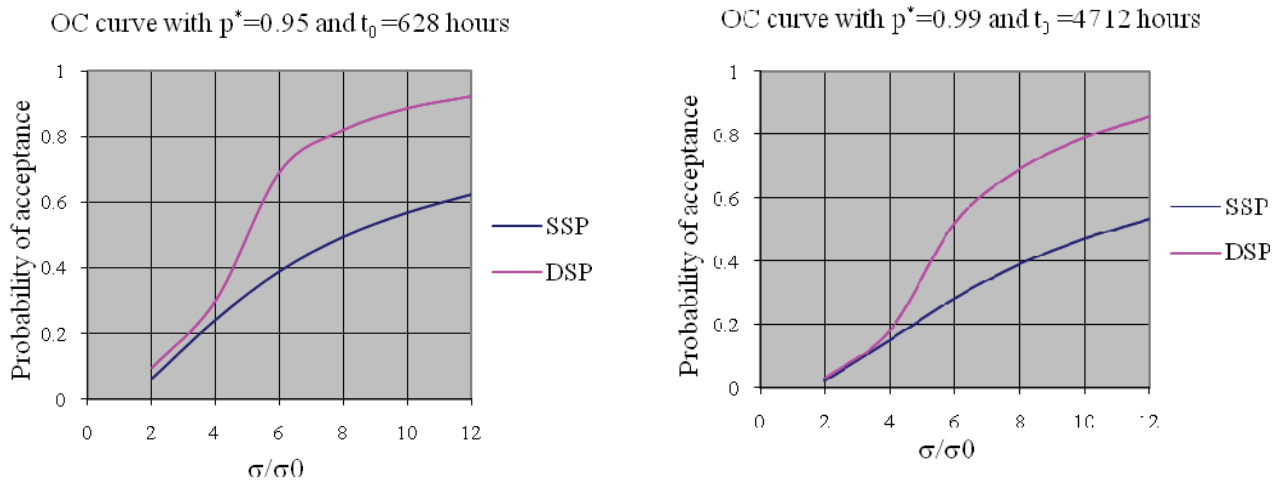
Conclusion

This study established the acceptance sampling plans for various values of  $\sigma / \sigma_0$  and an experiment time assuming a life test follows the Marshall-Olkin extended Lomax distribution. This distribution provides a high probability for  $\sigma / \sigma_0 > 6$ .

Table 4: Producer's Risk with Respect to Time of Experiment for Double Sampling ( $p^* = 0.75$ )

$\sigma / \sigma_0$	2	4	6	8	10	12
$(c_1 = 0, c_2 = 2, n_1 = 6, n_2 = 8, t / \sigma_0 = 0.628)$	0.70724	0.49895	0.13575	0.07221	0.04259	0.02711
$(c_1 = 0, c_2 = 2, n_1 = 2, n_2 = 3, t / \sigma_0 = 4.712)$	0.95094	0.83504	0.50939	0.34910	0.24381	0.17487

Figure 1: OC Curve with  $p^* = 0.95, t_0 = 628$  and  $p^* = 0.99, t_0 = 4712$  for Single and Double Acceptance Sampling Plans (SSP & DSP).



## DASP FOR MARSHALL-OLKIN EXTENDED LOMAX DISTRIBUTED LIFE TESTS

### References

- Aslam, M. (2007). Double acceptance sampling based on truncated life tests in Rayleigh distribution. *European Journal of Scientific Research*, 17(4), 605-611.
- Aslam, M., & Kantam, R.R.L. (2008). Economic reliability acceptance sampling based on truncated life tests in the Birnbaum-Saunders distribution. *Pakistan Journal of Statistics*, 24(4), 269- 276.
- Baklizi, A. (2003). Acceptance sampling based on truncated life tests in the Pareto distribution of the second kind. *Advances and Applications in Statistics*, 3(1), 33-48.
- Baklizi, A., & El Masri, A.E.K. (2004). Acceptance sampling based on truncated life tests in the Birnbaum-Saunders model. *Risk Analysis*, 24(6), 1453.
- Balakrishnan, N., Leiva, V., & Lopez, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum-Saunders distribution. *Communication in Statistics-Simulation and Computation*, 36, 643-656.
- Epstein, B. (1954). Truncated life tests in the exponential case. *Annals of Mathematical Statistics*, 25, 555-564.
- Ghitany, M. E., Al-Awadhi, F. A., & Al-khelfan, L. A. (2007). Marshal-Olkin extended Lomax distribution and its application to censored data. *Communications in Statistics Theory and Methods*, 36, 1855- 1866.
- Goode, H.P., & Kao, J.H.K (1961). *Sampling plans based on the Weibull distribution*. Proceeding of the Seventh National Symposium on Reliability and Quality Control, Philadelphia, PA, 24-40.
- Gupta, S. S. (1962). Life test sampling plans for normal and lognormal distribution. *Technometrics*, 4, 151- 175.
- Gupta, S. S., & Groll, P. A. (1961) Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association*, 56, 942-970.
- Kantam, R.R. L., & Rosaiah, K. (1998). Half logistic distribution in acceptance sampling based on life tests. *IAPQR Transactions*, 23(2), 117-125.
- Kantam, R.R. L., Rosaiah, K., & Rao, G. S. (2001). Acceptance sampling based on life tests: Log-logistic models. *Journal of Applied Statistics*, 28(1), 121-128.
- Rao, G. S., Ghitany, M. E., & Kantam, R. R. L. (2008). Acceptance sampling plans for Marshall-Olkin extended Lomax distribution. *International Journal of Applied Mathematics*, 21(2), 315-325.
- Rao, G. S., Ghitany, M. E., & Kantam, R. R. L. (2009). Acceptance sampling plans for Marshall-Olkin extended Lomax distribution. *International Journal of Applied Mathematics*, 22(1), 139-148.
- Rosaiah, K., & Kantam, R.R.L. (2005). Acceptance sampling based on the inverse Rayleigh distribution. *Economic Quality Control*, 20(2), 277-286.
- Rosaiah, K., Kantam, R.R.L., & Santosh Kumar, C. (2006). Reliability of test plans for exponentiated log-logistic distribution. *Economic Quality Control*, 21(2), 165-175.
- Rosaiah, K., Kantam, R.R. L., & Pratapa Reddy, J. (2007). Economic reliability test plan with Inverse Rayleigh Variate. *Pakistan Journal of Statistics*, 24(1), 57- 65.
- Rosaiah, K., Kantam, R.R. L., & Santosh Kumar, C. (2007). Exponentiated log-logistic distribution - an economic reliability test plan. *Pakistan Journal of Statistics*, 23(2), 147-156.
- Sobel, M., & Tischendorf, J. A. (1959). Acceptance sampling with new life test objectives, Proceedings of Fifth National Symposium on Reliability and Quality Control, Philadelphia, PA, 108-118.
- Tsai, T.R., & Wu, S.J. (2006). Acceptance sampling based on truncated life tests for generalized Rayleigh distribution. *Journal of Applied Statistics*, 33(6), 595-600.

## The Overall F-tests for Seasonal Unit Roots under Nonstationary Alternatives: Some Theoretical Results and a Monte Carlo Investigation

Ghassen El Montasser  
Manouba University,  
École Supérieure de Commerce de Tunis, Tunisia

---

In many empirical studies concerning seasonal time series, it has been shown that the whole set of unit roots associated with seasonal random walks are not present. This article focuses on the overall F-tests for seasonal unit roots under some nonstationary alternatives different from the seasonal random walk. The asymptotic theory of these tests is established for these cases using a new approach based on circulant matrix concepts. The simulation results joined to this theoretic analysis showed that the overall F-tests, as well as their augmented versions, maintained high power against the nonstationary alternatives.

Key words: Kunst test, nonstationary alternatives, Brownian motion, Monte Carlo Simulation.

---

### Introduction

The stochastic nature of seasonality appears to be gaining ground in empirical studies. Several aspects related to seasonal unit root tests are treated in the literature. In this respect, the power of these tests against nonstationary alternatives is an important issue that recently acquired some concern. To the best of our knowledge, Ghysels, Lee and Noh (1994) are the first authors who studied this question. Using a Monte Carlo study, they showed that, against a nonseasonal random walk, the power of the tests of Dickey, Hasza and Fuller (1984) is much lower than that of the tests introduced by Hylleberg, Engle, Granger and Yoo (1990).

Ghysels, et al. (1994) suggested that “the Dickey et al. test may not separate unit roots at each frequency” (p. 432). The restriction behind the Dickey, et al. procedure is that all unit roots (conventional and seasonal roots) are inseparably present with equal modulus; thus, it is clear that the conventional random walk does not fulfil this requirement. However, Rodrigues and Osborn (1999) showed that if this restriction

holds, the power of the Dickey, et al. tests would have a proper superiority in finite samples as opposed to that of the tests of Hylleberg, et al. (1990). Taylor (2003) analyzed the large sample behaviour of the seasonal unit root tests of Dickey, et al. when the data generating process (DGP) is a conventional random walk, that is, when the series only admits a zero frequency unit root. In such a case (and as shown by Taylor, 2003), the Dickey, et al. statistics have nondegenerate limiting distributions. These results theoretically explain the empirical findings of Ghysels, Lee and Noh (1994). Furthermore, Taylor (2005) showed that asymptotically the statistics of the Dickey, et al. augmented test will also do not diverge.

In a similar context, del Barrio Castro (2006) generalized the results of Taylor (2003) to a set of nonstationary alternatives which include the non seasonal random walk. He found that the Dickey, et al. statistics did not have standard limiting distributions and did not diverge. Based on the same methodology, del Barrio Castro (2007) established the limit theory of the Fisher and Student statistics originally developed by the Hylleberg, Engle, Granger and Yoo (1990) procedure. In that case, del Barrio Castro derived the effect that can have one unit root asymptotically on the others at different frequencies. Following the terminology of Buseti and Taylor (2003), this situation may be said to have “unattended unit roots” (p. 33).

---

Ghassen El Montasser is an Assistant Professor in the Ecole Supérieure de Commerce de Tunis. Email: g.elmontasser@gmail.com.



However, del Barrio Castro (2007), in a large sample analysis, did not directly consider the effects of nonstationary alternatives on the overall F-type statistic of seasonal integration which is complementarily specified for the Hylleberg, Engle, Granger and Yoo (1990) procedure by Ghysels, Lee and Noh (1994). This article adopts the seasonal integration definition of Ghysels and Osborn (2001, p. 43).

In a recent article, Osborn and Rodrigues (2002) developed an appealing approach for deriving asymptotic results for test statistics in seasonal models with unit roots. Such an approach is based on the use of circulant matrices which could, in seasonal context, retrieve the limit theory of the involved statistics as well as conveniently traducing the dynamics of time series and its evolution across different seasons. In a similar vein, Haldrup, Montanes and Sanso (2005) have used this approach to show the effects of outliers on the limit theory of seasonal unit root tests.

This article focuses on the large sample properties of the overall F-tests of the seasonal integration when the observed series is generated from nonstationary alternatives treated by del Barrio Castro (2006). This task is accomplished using the circulant matrix-based approach of Osborn and Rodrigues (2002).

The Kunst Test

The Kunst test for quarterly time series is based on the following regression

$$\Delta_4 y_t = \alpha_1 y_{t-1} + \dots + \alpha_3 y_{t-3} + \delta y_{t-4} + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

which is an F-type test of the form

$$F_{\hat{\alpha}_1, \dots, \hat{\alpha}_3, \hat{\delta}}^* = (T - 4)(\hat{\varepsilon}'_0 \hat{\varepsilon}_0 - \hat{\varepsilon}' \hat{\varepsilon}) / (\hat{\varepsilon}' \hat{\varepsilon}), \quad (2)$$

where  $\hat{\varepsilon}_0$  and  $\hat{\varepsilon}_1$  are vectors of residuals estimated under the null  $H_0 : \alpha_1 = \dots = \alpha_3 = \delta = 0$  and alternative hypotheses of the test. Assuming that, without any loss of generality, the initial values required by (1) are null. It should be noted that Kunst did not divide the numerator of the statistic (2) by 4

(the number of restrictions), as was done in this research to perform a conventional Fisher test.

The Hylleberg, Engle, Granger and Yoo (HEGY) Test

The basic regression for the HEGY test, without any augmentation and with no deterministic terms, is:

$$\Delta_4 y_t = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t, \quad t = 1, \dots, T, \quad (3)$$

where

$$y_{1t} = (1 + L + L^2 + L^3)y_t, \\ y_{2t} = -(1 - L + L^2 - L^3)y_t, \quad (4)$$

$$y_{3t} = -(1 - L^2)y_t,$$

with  $L$  as the lag operator.

Ghysels, Lee and Noh (1994) extended the HEGY approach with a joint test statistic  $F_{1234}$  for the null hypothesis,  $H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0$ , implying all unit roots in data are observed at quarterly frequency.  $H_0$  is an overall hypothesis for seasonal integration SI (1) in accordance with the notation of Ghysels and Osborn (2001). Note that:

$$\begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-2} \\ y_{3t-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ y_{t-4} \end{bmatrix}. \quad (5)$$

It may be deduced from (5) that the regressors of the Kunst test are non-singular linear transformations of those of the HEGY test. Consequently, the F-type statistics,  $F_{1234}$  and  $F_{\hat{\alpha}_1, \dots, \hat{\alpha}_3, \hat{\delta}}^* / 4$ , will have the same limit theory. Given that the two statistics are asymptotically related, the analysis is confined to that of Kunst in the sequel.

It can be observed that slight differences exist between the critical values of both statistics. In general, such critical values are tabulated assuming that the DGP of  $y_t$  is:

$$y_t = y_{t-4} + e_t. \quad (\text{A.0})$$

In this article it is assumed that the DGP of  $y_t$  is drawn from one of the following stochastic processes:

$$y_t = y_{t-1} + e_t, \quad (\text{A.1})$$

$$y_t = -y_{t-1} + e_t, \quad (\text{A.2})$$

$$y_t = y_{t-2} + e_t, \quad (\text{A.3})$$

$$y_t = -y_{t-2} + e_t, \quad (\text{A.4})$$

or

$$y_t = -y_{t-1} - y_{t-2} - y_{t-3} + e_t. \quad (\text{A.5})$$

Using the double subscript notation, the following annual vectors can be defined:

$$Y_n = (y_{1n}, y_{2n}, y_{3n}, y_{4n})',$$

and

$$E_n = (e_{1n}, e_{2n}, e_{3n}, e_{4n})',$$

where it is assumed that  $n = 1, \dots, N$  and in  $T$  observations there are  $N$  years, thus,  $T = 4N$ . To keep matters tractable, suppose that  $Y_0 = (y_{10}, y_{20}, y_{30}, y_{40})' = (0, 0, 0, 0)'$ .

The error processes in the alternatives (A.1)-(A.5) follow a stationary AR(p)

$$\phi(L)e_{sn} = v_{sn},$$

where  $\varphi(z)e_{sn} = 1 - \sum_{i=1}^p \phi_i z^i$  and  $s = 1, \dots, 4$ .

The roots of  $\varphi(z) = 0$  all lie outside the unit circle  $|z| = 1$ . As for the error sequence  $\{v_{sn}\}$ , it depicts an innovation process with constant conditional variance  $\sigma^2$  (see Spanos,

2003, p. 443). Similar to what has been conjectured by del Barrio Castro (2007) regarding the error structure in the nonstationary alternatives described above, suppose that the vector  $E_n$  has the following dynamics:

$$E_n = \sum_{j=0}^{\infty} \Gamma_j^* V_n,$$

where  $v_n = (v_{1n}, v_{2n}, v_{3n}, v_{4n})'$ , and the sequence of  $4 \times 4$  matrices are defined as:

$$\Gamma_0^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \gamma_1 & 1 & 0 & 0 \\ \gamma_2 & \gamma_1 & 1 & 0 \\ \gamma_3 & \gamma_2 & \gamma_1 & 1 \end{bmatrix}$$

$$\Gamma_0^* = \begin{bmatrix} \gamma_{j4} & \gamma_{j4-1} & \gamma_{j4-2} & \gamma_{j4-3} \\ \gamma_{j4+1} & \gamma_{j4} & \gamma_{j4-1} & \gamma_{j4-2} \\ \gamma_{j4+2} & \gamma_{j4+1} & \gamma_{j4} & \gamma_{j4-1} \\ \gamma_{j4+3} & \gamma_{j4+2} & \gamma_{j4+1} & \gamma_{j4} \end{bmatrix},$$

for  $j = 1, 2, \dots$ , with

$$\gamma(z) = 1 - \sum_{j=1}^{\infty} \gamma_j z^j$$

being the inverse of  $\varphi(z)$ . Finally,  $\Gamma^*(1)$  is defined as:

$$\Gamma^*(1) = \sum_{j=0}^{\infty} \Gamma_j^*.$$

del Barrio Castro (2006) used the vector of moving average representation to express the alternatives (A.i),  $i = 1, \dots, 5$ , in a vector of quarters representations where the observations of each year are stacked in the above defined vectors  $Y_n$  et  $E_n$ , let

$$(1 - B)Y_n = (\Theta_0^i + \Theta_1^i B)E_n, \quad i = 1, 2, \dots, 5, \quad (6)$$

where  $B$  is the annual backward operator. The  $4 \times 4$  matrices  $\Theta_0^i$  and  $\Theta_1^i$  (corresponding to the alternatives A.1-A.5) are defined as follows: for (A.1)

$$\Theta_0^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \Theta_1^1 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (7.1)$$

for (A.2)

$$\Theta_0^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 1 \end{bmatrix},$$

$$\Theta_1^2 = \begin{bmatrix} 0 & -1 & 1 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (7.2)$$

for (A.3)

$$\Theta_0^3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \Theta_1^3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (7.3)$$

for (A.4)

$$\Theta_0^4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix},$$

$$\Theta_1^4 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (7.4)$$

and for (A.5)

$$\Theta_0^5 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix},$$

$$\Theta_1^5 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (7.5)$$

The following result was established by del Barrio Castro (2007)

$$\frac{1}{\sigma \sqrt{N}} Y_{[rN]} \rightarrow_d B_i(r),$$

$$B_i(r) = C_i \Gamma^*(1) B(r),$$

$$C_i = \Theta_0^i + \Theta_1^i,$$

$$i = 1, 2, \dots, 5, \quad (8)$$

where the symbol  $\rightarrow_d$  denotes the convergence of probability measures,  $B_i(r)$  is a  $4 \times 1$  vector Brownian motion process with variance matrix  $\Omega_i = \sigma^2 C_i \Gamma^*(1) \Gamma^*(1)' C_i'$  and  $B(r)$  is a vector Brownian motion with variance matrix  $\sigma^2 I_4$ . The subscript  $i$  corresponds to the alternative (A.i),  $i = 1, \dots, 5$ .

Note that the rank of  $C_i$ ,  $i = 1, \dots, 5$ , is the number of (seasonal) unit roots implied by the process (A.i),  $i = 1, \dots, 5$ . In order to determine the number of cointegration relations between the quarters corresponding to every process (A.i),  $i = 1, \dots, 5$ , it is necessary to subtract from the periodicity of the quarterly data, that is 4, the rank of the matrix  $C_i$ ,  $i = 1, \dots, 5$ . Equation (8) may be rewritten more precisely by identifying the stochastic processes  $B_i(r)$ ,  $i = 1, 2, \dots, 5$ , on the grounds that there is always cointegration among the quarters of the time series (see del Barrio Castro, 2007, p.915).  
Limit Theory of the Kunst Test Under Nonstationary Alternatives

The following lemma can be directly deduced from the preceding result of del Castro Barrio (2007) and lemma A.1 of Osborn and Rodrigues (2002).

Lemma

Supposing that the DGP of  $y_t$  in (1) is given by the alternatives (A.1)-(A.5) and also that the vector  $(e_{1n}, \dots, e_{4n}), \forall n$ , satisfies assumption 1 of Phillips (1986, p.313), then under the null of the Kunst test is  $T \rightarrow \infty$

(a)

$$N^{-2} \sum_{n=1}^N Y_n Y_n' \rightarrow_d \sigma^2 \int_0^1 M_i B(r) B(r)' M_i' dr,$$

$$i = 1, 2, \dots, 5.$$

(b)

$$N^{-1} \sum_{n=1}^N Y_{n-1} \varepsilon_n' \rightarrow_d \sigma^2 M_i \int_0^1 B(r) dB(r)' M_i',$$

$$i = 1, 2, \dots, 5.$$

(c)

$$T^{-2} \sum_{t=1}^T y_{t-k}^2 \rightarrow_d \frac{\sigma^2}{16} \int_0^1 B(r)' M_i' M_i B(r) dr,$$

$$k = 1, \dots, 4, \quad i = 1, 2, \dots, 5.$$

(d)

$$T^{-2} \sum_{t=1}^T y_{t-k} y_{t-j} \rightarrow_d \frac{\sigma^2}{16} \int_0^1 B(r)' M_i' H_k' H_j M_i B(r) dr,$$

$$k \neq j, \quad i = 1, 2, \dots, 5.$$

(e)

$$T^{-1} \sum_{t=1}^T y_{t-k} \varepsilon_t \rightarrow_d \frac{\sigma^2}{4} \int_0^1 B(r)' M_i' H_k' M_i dB(r),$$

$$k = 1, \dots, 4, \quad i = 1, 2, \dots, 5,$$

where

$$\varepsilon_n = (\varepsilon_{1n}, \varepsilon_{2n}, \varepsilon_{3n}, \varepsilon_{4n})'$$

and

$$M_i = C_i \Gamma^*. \quad (1)$$

The matrix  $H_k, k = 1, 2, 3, 4$ , is a particular permutation matrix order 4 which produces the following elementary operations: let a matrix  $K$  have 4 lines, the operation  $H_1 K$  moves the last row of  $K$  to the top row of  $H_1 K$  and the other rows move down one place. More generally,  $H_i K$  shifts the final  $i^{\text{th}}$  rows to the top of the matrix while the remaining rows correspondingly move down; note that  $H_4 = I_4$  (see Golub & Van Loan, 1996, p. 109-112, for details). The OLS estimator  $\hat{\alpha}$  of the vector  $\alpha = (\alpha_1, \alpha_2, \alpha_3, \delta)'$  defined in Equation (1) satisfies under the null of Kunst test the asymptotic results represented by the following theorem.

Theorem

If that the DGP of  $y_t$  in (1) is given by one of the alternatives (A.1) - (A.5), then:

(a)

$$\frac{T}{4} (\hat{\alpha} - \alpha) \rightarrow_d F^{-1} f,$$

(matrices for  $F$  and  $f$  are shown in Figure 1);

(b) The Student statistic  $t_{\hat{\alpha}_i}$  corresponding to the  $i^{\text{th}}$  component of vector  $\hat{\alpha}$  satisfies the following result:

$$t_{\hat{\alpha}_i} \rightarrow_d \frac{(F^{-1} f)_i}{((F^{-1})_{ii})^{1/2}};$$

and

(c) The F-type statistic of Kunst test verifies:

$$F_{\hat{\alpha}_1, \dots, \hat{\alpha}_3, \delta}^* \rightarrow_d f' F^{-1} f.$$

Figure 1:  $F$  and  $f$  Matrices for  $\frac{T}{4}(\hat{\alpha} - \alpha) \rightarrow_d F^{-1}f$

$$F = \begin{bmatrix} \int_0^1 B(r)' M_i' H_1' H_1 M_i B(r) dr & \int_0^1 B(r)' M_i' H_1' H_2 M_i B(r) dr & \int_0^1 B(r)' M_i' H_1' H_3 M_i B(r) dr & \int_0^1 B(r)' M_i' H_1' M_i B(r) dr \\ \int_0^1 B(r)' M_i' H_2' H_1 M_i B(r) dr & \int_0^1 B(r)' M_i' H_2' H_2 M_i B(r) dr & \int_0^1 B(r)' M_i' H_2' H_3 M_i B(r) dr & \int_0^1 B(r)' M_i' H_2' M_i B(r) dr \\ \dots & \dots & \dots & \dots \\ \int_0^1 B(r)' M_i' H_3' H_1 M_i B(r) dr & \int_0^1 B(r)' M_i' H_3' H_2 M_i B(r) dr & \int_0^1 B(r)' M_i' H_3' M_i B(r) dr & \int_0^1 B(r)' M_i' M_i B(r) dr \end{bmatrix}$$

$$f = \begin{bmatrix} \int_0^1 B(r)' M_i' H_1' M_i dB(r) \\ \int_0^1 B(r)' M_i' H_2' M_i dB(r) \\ \dots \\ \int_0^1 B(r)' M_i' H_3' M_i dB(r) \\ \int_0^1 B(r)' M_i' M_i dB(r) \end{bmatrix} \forall i = 1, 2, \dots, 5.$$

Proof

Before showing the proof of result (a) of the theorem, consider the properties of the matrix  $F$ : the elements of the main diagonal of  $F$  are all equal and the elements of  $F$  along each diagonal line parallel to the principal diagonal are equal, thus,  $F$  is a Toeplitz matrix. Toeplitz matrices belong to the larger class of persymmetric matrices. A square matrix  $B$  of order  $n$  is persymmetric if it is symmetric about the northeast-southwest diagonal, that is,  $b_{ij} = b_{n-j+1, n-i+1}$  for all  $i$  and  $j$ . Moreover, from the properties of the matrices  $H_k$ ,  $k = 1, 2, 3, 4$ , it can be shown that the matrix  $F$  is also symmetric. Equation (1) can be written in matrix form:

$$Y = X\alpha + \varepsilon, \quad (9)$$

where

$$X = \begin{bmatrix} y_0 & y_{-1} & y_{-2} & y_{-3} \\ y_1 & y_0 & y_{-1} & y_{-2} \\ \dots & \dots & \dots & \dots \\ y_{T-1} & y_{T-2} & y_{T-3} & y_{T-4} \end{bmatrix}$$

and

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_T \end{bmatrix}.$$

In addition,  $\hat{\alpha} - \alpha = (X'X)^{-1}X'\varepsilon$ ,

where

$$X'X = \begin{bmatrix} \sum_1^T y_{t-1}^2 & \sum_1^T y_{t-1}y_{t-2} & \sum_1^T y_{t-1}y_{t-3} & \sum_1^T y_{t-1}y_{t-4} \\ \sum_1^T y_{t-2}y_{t-1} & \sum_1^T y_{t-2}^2 & \sum_1^T y_{t-2}y_{t-3} & \sum_1^T y_{t-2}y_{t-4} \\ \sum_1^T y_{t-3}y_{t-1} & \sum_1^T y_{t-3}y_{t-2} & \sum_1^T y_{t-3}^2 & \sum_1^T y_{t-3}y_{t-4} \\ \sum_1^T y_{t-4}y_{t-1} & \sum_1^T y_{t-4}y_{t-2} & \sum_1^T y_{t-4}y_{t-3} & \sum_1^T y_{t-4}^2 \end{bmatrix},$$

and

$$X'\varepsilon = \begin{bmatrix} \sum_1^T y_{t-1}\varepsilon_t \\ \dots \\ \sum_1^T y_{t-4}\varepsilon_t \end{bmatrix}.$$

Also:  $\frac{T}{4}(\hat{\alpha} - \alpha) = \frac{1}{4} \left( \frac{X'X}{T^2} \right)^{-1} \frac{X'\varepsilon}{T}$  and, due to parts (c), (d) and (e) of the preceding lemma and the fact that  $H_4 = I_4$ , the result of the theorem holds.

The asymptotic distributions of Student statistics corresponding to the parameters of Equation 1 can be deduced from result (a) of the theorem. To prove result (c), the F-type statistic of Kunst can be written as follows:  $F_{\hat{\alpha}_1, \dots, \hat{\alpha}_3, \hat{\delta}}^* = \hat{\alpha}' [(S^2)^{-1} X'X] \hat{\alpha}$ , where  $S^2$  is the OLS estimator of the residual variance in Equation (1). The (Toeplitz) circulant matrix F and its inverse are symmetric, consequently result (c) holds and the theorem is proved.

Empirical quantiles of the Kunst test for the processes (A.1)-(A.5) were generated and associated with nominal levels 90%, 95% and 99%. The sample size considered is 4,000 (1,000 years) with 20,000 replications; it has been shown that these empirical quantiles tend to be infinite (these results are not presented, but are available upon request). Consequently, it is possible to predict that in 100% of cases the null hypothesis will be rejected for the processes (A.1)-(A.5) for nominal levels of 5% and 1%. Table 1 shows the rejection frequencies for a sample size of 100 (25 years) and 20,000 replications. All simulations were conducted using the software Matlab.

In addition, augmented regression (1) corresponding to the Kunst test was carried out by lagged values of the independent variable, thus, this regression becomes

$$\Delta_4 y_t = \alpha_1 y_{t-1} + \dots + \alpha_3 y_{t-3} + \delta y_{t-4} + \sum_{i=1}^p \Delta_4 y_{t-i} + \varepsilon_t, \quad t = 1, \dots, T. \quad (10)$$

Table 2 reports the power of the augmented Kunst test against the nonstationary alternatives (A.1) - (A.5). Results in Table 2 show that perfect power is maintained across all the alternatives (A.1) - (A.5) even if the number of lagged terms of the dependent variable increases. At this level, a slight exception to this general finding was detected for the alternative (A.5) and for  $p = 4$  or  $p = 6$ . Particularly, and for this alternative, the exception is much clearer for  $p = 6$  and the nominal level 1%. In fact, the test power decreases and reaches a value of approximately 66%.

### Conclusion

A large amount of literature on testing for seasonal unit roots has appeared during the last two decades. However, the majority of econometricians treating this topic have bent over backwards to give the limit theory of the tests for unit roots at the zero, Nyquist and

Table 1: Empirical Rejection Frequencies of Kunst Test under Nonstationary Alternatives

Kunst Test	Processes					
	(A.0)	(A.1)	(A.2)	(A.3)	(A.4)	(A.5)
$F_{\hat{\alpha}_1, \dots, \hat{\alpha}_3, \hat{\delta}}^*$ Nominal Size 5%	0.095	1	1	1	1	1
$F_{\hat{\alpha}_1, \dots, \hat{\alpha}_3, \hat{\delta}}^*$ Nominal Size 1%	0.0158	1	1	1	1	1

Number of replications: 20,000; Sample size  $4N = 100$  observations

## F-TESTS FOR SEASONAL UNIT ROOTS UNDER NON-STATIONARY ALTERNATIVES

Table 2: Empirical Rejection Frequencies of the Kunst Augmented Test under Nonstationary Alternatives

		Processes					
		(A.0)	(A.1)	(A.2)	(A.3)	(A.4)	(A.5)
Nominal Size 5%	p = 2	0.0592	1	1	1	1	1
	p = 4	0.0549	1	1	1	1	0.9920
	p = 6	0.0522	1	1	0.9980	0.9976	0.9038
Nominal Size 1%	p = 2	0.0140	1	1	1	1	0.9998
	p = 4	0.0141	1	1	0.9992	0.9991	0.9271
	p = 6	0.0121	0.9990	0.9990	0.97770	0.97460	0.6632

harmonic seasonal frequencies by considering either an additional determinist component or a modified hypothesis set concerning the error terms which appear in the regression models associated with such tests. Seldom have works concerning this topic studied the power of seasonal unit roots against nonstationary alternatives. Ghysels, et al. (1994) studied this problem and, in a simulation study, they hypothesized that the DHF test may not separate unit roots at each frequency. Having enriched this analysis by a large sample investigation, Taylor (2003) found that the DHF statistics did not diverge to minus infinity when the DGP of the series is a conventional random walk. del Barrio Castro (2006, 2007) considered an extended set of nonstationary alternatives and studied their asymptotic effects on the DHF and HEGY statistics.

This article extended the problem treated by Taylor (2003) and del Castro Barrio (2007) to the overall F-type tests for seasonal integration. It has been realized that the most renowned tests, that is, those of HEGY (1990) and Kunst (1997), are asymptotically related. For this reason, this research focused on asymptotic effects of the nonstationary alternatives, (A.1) - (A.5) on Kunst F-type test. To reach this goal, I had the circulant-matrix-based approach introduced by Osborn and Rodrigues (2002) was chosen. Moreover, in a simulation study, it was found that the Kunst F-

type statistic maintained high power when the totality of unit roots implied by the filter  $(1-L^4)$  were not present. In addition, these high-power properties are preserved when the regression model of the test was augmented with lagged dependent variables. The approach adopted in this article can be applied to cases other than the quarterly one. To this aim, it is enough to write the adequate matrices  $C_i$  defined from Equation (5) following the data observation frequency.

### References

- Busetti, F., & Taylor, A. M. R. (2003). Testing against stochastic trend and seasonality in the presence of unattended breaks and unit roots. *Journal of Econometrics*, 117, 21-53.
- Dickey, D. A., Hasza, D. P., & Fuller, W. A. (1984). Testing for unit roots in seasonal time series. *Journal of American Statistical Association*, 79, 355-367.
- del Barrio Castro, T. (2007). Using the HEGY procedure when not all roots are present. *Journal of Time Series Analysis*, 28, 910-922.
- del Barrio Castro, T. (2006). On the performance of the DHF test against nonstationary alternatives. *Statistics and Probability Letters*, 76, 291-296.

- Ghysels, E., Lee, H. S., & Noh, J. (1994). Testing for unit roots in seasonal time series: Some theoretical extensions and a Monte Carlo investigation. *Journal of Econometrics*, 62, 415-442.
- Ghysels, E., & Osborn, D. R. (2001). *The econometric analysis of seasonal time series*. Cambridge, MA: Cambridge University Press.
- Golub, G. H., & van Loan, C. F. (1996). *Matrix computations*. Baltimore: The Johns Hopkins University Press.
- Haldrup, N., Montanés, A., & Sanso, A. (2005). Measurement errors and outliers in seasonal unit root testing. *Journal of Econometrics*, 127, 103-128.
- Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44, 215-238.
- Kunst, R. M. (1997). Testing for cyclical non-stationarity in autoregressive models. *Journal of Time Series Analysis*, 18, 123-135.
- Osborn, D. R., & Rodrigues, P. M. M. (2002). The asymptotic distributions of seasonal unit root tests: a unifying approach. *Econometric Reviews*, 21, 221-241.
- Phillips, P. C. B. (1986). Understanding Spurious Regressions in Econometrics. *Journal of Econometrics*, 33, 311-340.
- Rodrigues, P. M. M., & Osborn, D. R. (1999). Performance of seasonal unit root tests for monthly data. *Journal of Applied Statistics*, 26, 985-1004.
- Spanos, A. (2003). *Probability theory and statistical inference: econometric modeling with observational data*. Cambridge MA: Cambridge University Press (Virtual Publishing).
- Taylor, A. M. R. (2005). On the limiting behaviour of augmented seasonal unit root tests. *Economics Bulletin*, 3, 1-10.
- Taylor, A. M. R. (2003). On the asymptotic properties of some seasonal unit root tests. *Econometric Theory*, 19, 311-321.



## Fisher's Exact Test for Misclassified Data

Tze-San Lee  
Western Illinois University  
Macomb, IL USA

Fisher's exact test is adapted to handle the misclassified data arising from comparing two binomial populations. The bias-adjusted odds ratio is proposed to account for misclassification errors. Its expected power depends in a nonlinear way on the true sensitivity and specificity of the classification method. The data taken from the no conviction rate of criminality for two types of twin populations was used to illustrate how to calculate true sensitivity and specificity and the expected power of the adjusted odds ratio.

Key words: Fisher's exact test, misclassification, power function, odds ratio, sensitivity, specificity.

### Introduction

Fisher's (1946) exact test is used when the sample size is less than five. However, the issue on how to adapt Fisher's exact test if the data are misclassified has not been addressed. It is the aim of this article to adapt Fisher's exact test to account for misclassification errors.

### Methodology

Consider two independent binomial random variables  $X$  and  $Y$  with parameters  $(n_X, p_X)$  and  $(n_Y, p_Y)$ , respectively, where both  $n_X$  and  $n_Y$  are less than 5. A classical problem is to find an exact test for the null hypothesis  $H_0: p_X = p_Y$  against an alternative hypothesis  $H_1: p_X > p_Y$ , or equivalently,  $H_0: \gamma = 1$  against  $H_1: \gamma > 1$ , where  $\gamma$  is the odds ratio defined by (Fleiss, Levin & Paik, 2003)

$$\gamma = \frac{p_X q_Y}{p_Y q_X}, \quad (1)$$

Let the number of successes be  $x$  and  $y$ , respectively, among  $n_X$  and  $n_Y$  subjects. Assume that among  $x$  and  $y$  successes there were possible misclassified cases. Before showing how to adapt Fisher's exact test to deal with misclassified data, a depiction of Fisher's exact test is provided.

By conditioning that  $x + y = z$  is fixed, the [conditional] distribution of  $X = x$  is given by the extended (or non-central) hypergeometric distribution under the alternative hypothesis (Gart, 1971; Harkness, 1965; Levin, 1984)

$$\Pr(X = x | z; \gamma) = \frac{\binom{n_X}{x} \binom{n_Y}{z-x} \gamma^x}{\sum_{j=0}^z \binom{n_X}{j} \binom{n_Y}{z-j} \gamma^j}, \quad (2)$$

where  $x = \max(0, z - n_Y), \dots, \min(z, n_X)$ , or the [conditional] distribution of  $Y = y$  is given by

$$\Pr(Y = y | z; \gamma) = \frac{\binom{n_X}{z-y} \binom{n_Y}{y} \gamma^{-y}}{\sum_{j=0}^z \binom{n_X}{z-j} \binom{n_Y}{j} \gamma^{-j}}, \quad (3)$$

Tze-San Lee is presently working at the Centers for Disease Control and Prevention, Chamblee, GA. Email him at: [tjl3@cdc.gov](mailto:tjl3@cdc.gov).

where  $y = \max(0, z - n_x), \dots, \min(z, n_y)$ . In practical applications, equation 2 is used if  $x < y$ ; otherwise, equation 3 is used. Note that equations 2 and 3 were first introduced by Fisher (1935). Equation 2 (or 3) can be used to devise significance tests or confidence intervals on any value of  $\gamma$ .

If  $x > y$ , then the  $p$ -value of Fisher's exact test  $H_0: \gamma = 1$  against  $H_1: \gamma > 1$  is given by

$$p\text{-value} = \sum_{k=y}^z \Pr(Y = k | z; \gamma = 1), \quad (4)$$

where  $\Pr(Y = k | z; \gamma = 1)$  is given by equation 3 which yields an ordinary hypergeometric distribution as follows:

$$\Pr(Y = y | z; \gamma = 1) = \frac{\binom{n_x}{z-y} \binom{n_y}{y}}{\binom{n_x + n_y}{z}},$$

$$y = 0, 1, \dots, z, \quad (5)$$

on which the Fisher-Irwing exact test is based.

For small frequencies, the critical value of  $Y$  has been provided by choosing  $y_c$  for one-sided alternatives ( $p_x > p_y$ ) such that

$$\Pr(Y \leq y_c | z) = \sum_{y=0}^{y_c} \Pr(Y = y | z; \gamma = 1) \leq \alpha$$

and

$$\Pr(Y \leq y_c + 1 | z) > \alpha. \quad (6)$$

For nominal levels of significance  $\alpha = 0.05, 0.025, 0.01, 0.005$ ,  $x_c$  has been tabulated for  $n_y \leq n_x \leq 25$  (Bennett & Hsu, 1960). For two-sided alternatives ( $p_x \neq p_y$ ), the tabular exact probabilities are doubled accordingly.

If  $x > y$ , the (conditional) power function of the exact test is then given by

$$\beta(\gamma | z) = \Pr(y \leq y_c | H_1)$$

$$= \sum_{y=0}^{y_c} \Pr(Y = y | z; \gamma) \quad , \quad (7)$$

$$= \frac{\sum_{y=0}^{y_c} \binom{n_x}{z-y} \binom{n_y}{y} \gamma^{-y}}{\sum_{j=z_1}^{z_2} \binom{n_x}{z-j} \binom{n_y}{j} \gamma^{-j}}$$

where

$$z_1 = \max(0, z - n_x)$$

and

$$z_2 = \max(z, n_y).$$

Note that  $\beta(\gamma | z)$  of equation 7 is a rational function in  $\gamma$ , that is, a ratio of two polynomial functions in  $\gamma$ .

If  $x > y$ , the expected power of the exact test is given by

$$\beta(\gamma) = \sum_z \beta(\gamma | z) \Pr(Z = z)$$

$$= q_x^{n_x} q_y^{n_y} \left\{ \sum_z \left( \frac{p_x}{q_x} \right)^z \left[ \sum_{y=z_1}^{y_c} \binom{n_x}{z-y} \binom{n_y}{y} \gamma^{-y} \right] \right\}, \quad (8)$$

where  $\Pr(Z = z)$  representing the distribution of  $Z$  is given by

$$\Pr(Z = z) = \sum_{i=L}^U \binom{n_x}{z-i} p_x^{z-i} q_x^{n_x-z+i} \binom{n_y}{i} p_y^i q_y^{n_y-i}, \quad (9)$$

where  $L = \max(0, z - n_x)$ ,  $U = \min(z, n_y)$  and the summation in  $z$  is over all significance pairs of points on the diagonals,  $z = x + y$ , in the  $(x, y)$  sample space at level of significance equals to  $\alpha$  at most (Bennett & Hsu, 1960; Casagrange, Pike & Smith, 1978a). To facilitate a calculation of the expected power of equation 7 a FORTRAN program was written by Casagrange, Pike and Smith (1978b). If  $n_x = n_y$ , Conlon & Thomas (1993) presented an algorithm which was feasible for very large sample sizes.

## FISHER'S EXACT TEST FOR MISCLASSIFIED DATA

If classification errors exist on the number of successes for both  $X$  and  $Y$ , the question becomes how to test the hypotheses previously postulated. Let  $E^*$  be the surrogate classification variable for  $E$  and let  $\varphi_Z$  and  $\psi_Z$ ,  $Z = X, Y$ , be the sensitivity and specificity for classifying the status of the outcome among samples from the case and the control populations, respectively, that is, for  $Z = X$  or  $Y$ ,

$$\varphi_Z = \Pr(E^* = 1 | E = 1),$$

and

$$\psi_Z = \Pr(E^* = 0 | E = 0). \quad (10)$$

It is known that, for the unknown  $p_X$  and  $p_Y$ , the following maximum likelihood estimators are no longer unbiased:

$$\hat{p}_X = x/n_X \text{ and } \hat{p}_Y = y/n_Y, \quad (11)$$

assume that  $\hat{p}_Z \sim \text{Binomial}(n_Z, p_Z)$  for  $Z = X, Y$ . The crude odds ratio (COR) defined by

$$\hat{\gamma} = \frac{\hat{p}_X \hat{q}_Y}{\hat{p}_Y \hat{q}_X}, \quad (12)$$

as a point estimator for the true odds ratio  $\gamma$  of equation 1 can have substantial bias (Kleinbaum, et al., 1982).

To account for the misclassification bias, the bias-adjusted [point] estimators for the prevalence of success/failure  $p_Z$  and  $q_Z$  are given by (Lee, 2009)

$$\bar{p}_Z = (\psi_Z - \hat{q}_Z) / \Delta_Z$$

and

$$\bar{q}_Z = (\varphi_Z - \hat{p}_Z) / \Delta_Z, \quad (13)$$

where  $\hat{q}_Z = 1 - \hat{p}_Z$ ,  $Z = X, Y$ , and  $\Delta_Z$  is given by

$$\Delta_Z \equiv \varphi_Z + \psi_Z - 1. \quad (14)$$

Conditioned on that  $\varphi_Z$  and  $\psi_Z$  are given, it is easily shown that equation 13 is an unbiased estimator for  $p_Z$  and  $q_Z$ , respectively. The

bias-adjusted estimators  $\bar{p}_Z$  and  $\bar{q}_Z$  (equation 13) are said to be plausible if  $\bar{p}_Z$  and  $\bar{q}_Z$  lie between 0 and 1. In order for  $\bar{p}_Z$  and  $\bar{q}_Z$  to be plausible, the following constraints are imposed: for  $Z = X, Y$ ,

$$\varphi_Z > \hat{p}_Z, \psi_Z > \hat{q}_Z \text{ and } \Delta_Z > 0. \quad (15)$$

A set of  $\varphi_Z$  and  $\psi_Z$  is said to be feasible if equation 15 holds. Furthermore, a set of feasible  $\varphi_Z$  and  $\psi_Z$  is said to be admissible if for these feasible  $\varphi_Z$  and  $\psi_Z$ ,  $\bar{p}_Z$  and  $\bar{q}_Z$  are plausible.

The bias-adjusted odds ratio (BAOR) for  $\gamma$  which accounts for misclassification bias is then given by

$$\begin{aligned} \bar{\gamma} &= \frac{\bar{p}_X \cdot \bar{q}_Y}{\bar{p}_Y \cdot \bar{q}_X} \\ &= \frac{(\psi_X - \hat{q}_X)(\phi_Y - \hat{p}_Y)}{(\phi_X - \hat{p}_X)(\psi_Y - \hat{q}_Y)} \end{aligned} \quad (16)$$

and its asymptotic variance is given by

$$\text{var}(\ln(\bar{\gamma})) = \sum_{Z=X}^Y \frac{\text{var}(\hat{p}_Z)}{(\Delta_Z p_Z q_Z)^2}, \quad (17)$$

where  $\text{var}(\hat{p}_Z)$  is given by

$$\text{var}(\hat{p}_Z) = \frac{(p_Z \Delta_Z + 1 - \psi_Z)(q_Z \Delta_Z + 1 - \phi_Z)}{n_Z}. \quad (18)$$

Using equations 16-17 to find a 100% $\times(1 - \alpha)$  confidence interval (LCL, UCL) for the true  $\gamma$  (equation 1), where LCL and UCL are abbreviations denoting for lower and upper confidence limit, respectively, and  $0 < \alpha < 1$  as follows:

$$\text{(LCL, UCL):} = \exp(\ln(\bar{\gamma}) \mp z_{1-\frac{\alpha}{2}} \times \sqrt{\text{var}(\ln(\bar{\gamma}))}). \quad (19)$$

Note that Fisher's exact test, which accounts for misclassification errors, is exactly the same as

that shown previously, that is, the p-value for the significance test of  $H_0: \gamma = 1$  against  $H_1: \gamma > 1$  does not depend upon the sensitivity and specificity of the classification method at all. But the conditional power of equation 7 or the expected power of equation 8 at  $\gamma = \bar{\gamma}$  depends on the sensitivity and specificity of the classification method for both populations.

### Results

The Lange's data on criminality among twin brothers/sisters of criminals (Fisher, 1946) was used for analysis. Table 1 shows the numbers of twin brothers/sisters of criminals who have been convicted, separately for dizygotic (=  $X$ ) (but like-sexed) and monozygotic twins (=  $Y$ ). Because  $\hat{p}_X = 13/15 > 3/13 = \hat{p}_Y$ , the phrase not convicted is taken as success. Inspection of Table 1 shows  $x = 13$ ,  $y = 3$ ,  $n_X = 15$ , and  $n_Y = 13$ . The COR of equation 12 was obtained as  $\hat{\gamma} = 21.7$  with  $p = 0.001$  by using the SAS software with a specification to Fisher's exact test (Stokes, Davis & Koch, 2000). As a result, the null hypothesis is rejected. This means that the deviation from proportionality in Table 1 is significant to provide evidence that criminality is more frequent among monozygotic twins of criminals than among dizygotic twins of criminals (Finney, 1948).

Table 1: Lange's Data on Criminality among Twin Brothers/Sisters of Criminals

	Dizygotic (= $X$ )	Monozygotic (= $Y$ )	Row Total
Not convicted (Success)	13	3	16
Convicted (Failure)	2	10	12
Column Total	15	13	28

Suppose that classification error exists in the observed data shown in Table 1. Because

validation data was not available and because it was not possible to know what true classification is, all possible re-arrangements of Table 1 were considered and then each re-arranged table was treated as if it were a true table. Thus, it was possible to calculate true sensitivity and specificity for both of the two populations. For the population of dizygotic twins, there are a total of 13 possible truly classified tables. Hence, in theory, there are a total of 13 possible pairs of true sensitivities and specificities for this targeted population. However, after checking the feasibility constraints (equations 14-15), only four pairs of sensitivity and specificity were feasible (Table 2). Similarly, five out of a total of eleven possible pairs of sensitivity and specificity were feasible for the population of monozygotic twins (Table 2).

Because it was not possible to know which pair was the true sensitivity and specificity for either one of the two populations of twins, it was necessary to calculate the BAOR,  $\bar{\gamma}$  (equation 16), for all 20 (=  $4 \times 5$ ) possible combinations of feasible pairs of sensitivity and specificity for the two targeted populations of twins. The calculation was organized as follows. One pair of feasible sensitivity and specificity was fixed from the population of dizygotic twins and then combined with all five pairs of feasible sensitivity and specificity for the population of monozygotic twins in order to calculate  $\bar{\gamma}$  (equation 16). This procedure was then repeated by changing only the pair from the population of dizygotic twins until all four feasible pairs were used (Table 3). As shown in Table 3, only three BAORs from the 2<sup>nd</sup> to the 4<sup>th</sup> entries were found to be significant for cases i-ii, whereas none of the BAORs were significant for cases iii-iv. If the COR is credible, then this implies that to under-misclassify two or three pairs of dizygotic twins in the convicted category is implausible. If only one pair of dizygotic twins is over-misclassified (comparing case i with the correctly classified pair of dizygotic twins in Table 2), the COR ( $\hat{\gamma} = 21.7$ ) over-estimated the true  $\gamma$  because the BAORs were 17.2 and 19.5 when one pair of monozygotic twins was under- and over-misclassified in the convicted category, while under-estimated the true  $\gamma$  because the BAOR

## FISHER'S EXACT TEST FOR MISCLASSIFIED DATA

Table 2: Pairs of Feasible Sensitivity and Specificity for Two Types of Twins under the Assumption that True Classifications Are Known

Dizygotic [ $\hat{p}_x = 0.867$ , $\hat{q}_x = 0.133$ ]				Monozygotic [ $\hat{p}_y = 0.23$ , $\hat{q}_y = 0.77$ ]			
Convicted		$\varphi_x$	$\psi_x$	Convicted		$\varphi_y$	$\psi_y$
No	Yes			No	Yes		
14	1	0.963	0.667	1	12	0.5	0.909
12	3	0.960	0.800	2	11	0.8	0.952
11	4	0.917	0.667	4	9	0.857	0.947
10	5	0.870	0.571	5	8	0.75	0.889
				6	7	0.667	0.824

was 24.0 when two pairs of monozygotic twins was over-misclassified in the convicted category. Similarly, If only one pair of dizygotic twins is under-misclassified (case ii), the COR under-estimates the true  $\gamma$  because the BAORs were given by 22.2, 25.1, and 31.0 corresponding to when one pair was either under- or over-misclassified or two pairs of monozygotic twins were over-misclassified, respectively.

To calculate the expected power for either  $\hat{\gamma}$  (equation 12) or  $\bar{\gamma}$  (equation 16), the crude/bias-adjusted point estimator is substituted for  $p_i$  and  $q_i$ , namely, the COR or BAOR for  $\gamma$  in equation 8; thus the expected power of equation 8 (Table 4) were obtained. Note that the results shown in the first row of Table 4 correspond to the COR because the COR can be viewed as a special case of the BAOR with perfect classification, that is, both sensitivity and specificity equal to one. If  $n_x = 15$ ,  $n_y = 13$ , and  $x = 13$ , then  $y_c = 6$  from the table of critical values for  $y$  (Finney, 1948; p. 154); this is used in determining all possible  $z$ -values in equation 8.

The results of the expected power for the COR and six admissible BAORs are given in Table 4. It is not surprising to see that the COR has the highest expected power (= 0.45) because both the sensitivity and specificity equals one, whereas the expected power of the six BAORs varies. It seems that the expected power of the BAOR depends on the values of the sensitivity

and specificity: the higher the specificity across the board, the larger the expected power. Indeed, the expected power is higher for  $(\varphi_x, \psi_x) = (0.96, 0.8)$  than for  $(\varphi_x, \psi_x) = (0.963, 0.6667)$ , that is, 0.40, 0.36, and 0.29 compared to 0.33, 0.29, and 0.23, respectively (see Table 4). Also, the larger expected power 0.40 (or 0.33) corresponds to the highest specificity ( $\psi_y = 0.9524$ ) for the monozygotic population. In terms of the type of misclassification the highest power corresponds to that exactly one pair of twins is under-misclassified in the category of convicted for both populations.

### Conclusion

Fisher's exact test was adapted to handle a scenario where data are misclassified. The bias-adjusted odds ratio was proposed to account for the misclassification errors. Because a validation sample is not available, all possible pairs of true sensitivity and specificity were calculated from the observed data by assuming that a true table is known. Although the p-value is not affected by the true sensitivity and specificity of the classification method, the expected power of Fisher's exact test depends on these in a nonlinear way. The data regarding whether the no-conviction rate are the same between the dizygotic and monozygotic twins of brothers/sisters was used to illustrate how to calculate true sensitivity and specificity, the bias-adjusted odds ratio and their expected power accordingly.

Table 3: The Point Estimate, Standard Error of Logarithm of Adjusted Odds Ratio and 95% CI for 20 Possible Combinations of True Sensitivity and Specificity of Two Twin Populations

$(\varphi_Y, \psi_Y)$	$\bar{\gamma}$	$s.e.(\ln(\bar{\gamma}))$	LCL	UCL
(i) $(\varphi_X, \psi_X) = (0.963, 0.667)$				
(0.5, 0.909)	10.7	1.66	0.41	278.4
(0.8, 0.952)	17.2	1.37	1.18	251.1
(0.857, 0.947)	19.5	1.37	1.34	283.8
(0.75, 0.889)	24.0	1.61	1.02	567.7
(0.667, 0.824)	44.5	2.65	0.25	7993.5
(ii) $(\varphi_X, \psi_X) = (0.96, 0.8)$				
(0.5, 0.909)	13.8	1.66	0.53	357.0
(0.8, 0.952)	22.2	1.36	1.53	321.7
(0.857, 0.947)	25.1	1.36	1.74	363.6
(0.75, 0.889)	31.0	1.61	1.32	727.9
(0.667, 0.824)	57.3	2.65	0.32	10271.6
(iii) $(\varphi_X, \psi_X) = (0.917, 0.667)$				
(0.5, 0.909)	20.5	2.30	0.23	1869.6
(0.8, 0.952)	33.2	2.10	0.54	2020.8
(0.857, 0.947)	37.5	2.10	0.62	2284.9
(0.75, 0.889)	46.3	2.27	0.55	3921.5
(0.667, 0.824)	85.6	3.09	0.20	36497.9
(iv) $(\varphi_X, \psi_X) = (0.87, 0.571)$				
(0.5, 0.909)	291.0	30.5	$3.0 \times 10^{-24}$	$2.7 \times 10^{28}$
(0.8, 0.952)	469.8	30.5	$5.1 \times 10^{-24}$	$4.3 \times 10^{28}$
(0.8571, 0.947)	531.6	30.5	$5.8 \times 10^{-24}$	$4.9 \times 10^{28}$
(0.75, 0.889)	656.0	30.5	$7.0 \times 10^{-24}$	$6.1 \times 10^{28}$
(0.667, 0.824)	1213.6	30.6	$1.1 \times 10^{-23}$	$1.3 \times 10^{29}$

Table 4: The Expected Power of the Crude/Adjusted Odds Ratio

$(\varphi_X, \psi_X)$	$(\varphi_Y, \psi_Y)$	$p_X$	$p_Y$	$\gamma$	$\beta(\gamma   z)$
(1.0, 1.0)	(1.0, 1.0)	0.87	0.23	21.7	0.45
(0.963, 0.667)	(0.8, 0.952)	0.85	0.24	17.2	0.33
(0.963, 0.667)	(0.857, 0.947)	0.85	0.22	19.5	0.29
(0.963, 0.667)	(0.75, 0.889)	0.85	0.19	24.0	0.23
(0.96, 0.8)	(0.8, 0.952)	0.88	0.24	22.2	0.40
(0.96, 0.8)	(0.857, 0.947)	0.88	0.22	25.1	0.36
(0.96, 0.8)	(0.75, 0.889)	0.88	0.19	31.0	0.29

## FISHER'S EXACT TEST FOR MISCLASSIFIED DATA

### Acknowledgement

The author would like to acknowledge the use of the Microsoft's EXCEL spread sheet to facilitate numerical calculations for Tables 2-4.

### References

- Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, 7, 131-153.
- Agresti, A. (2002). *Categorical data analysis* (2<sup>nd</sup> Ed.). New York: Wiley.
- Bennett, B. M., & Hsu, P. (1960). On the power function of the exact test for the  $2 \times 2$  contingency table. *Biometrika*, 47, 393-398.
- Casagrande, J. T., Pike, M. C., & Smith, P. G. (1978a). The power function of the exact test for comparing two binomial distribution. *Applied Statistics*, 27, 176-180.
- Casagrande, J. T., Pike, M. C., & Smith, P. G. (1978b). Algorithm AS 129: The power function of the exact test for comparing two binomial distribution. *Applied Statistics*, 27, 212-219.
- Colon, M., & Thomas, R. G. (1993). Algorithm AS: 280: The power function for Fisher's exact test. *Applied Statistics*, 42, 258-260.
- Finney, D. J. (1948). The Fisher-Yates test of significance in  $2 \times 2$  contingency tables. *Biometrika*, 35, 145-156.
- Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, Series A*, 98, 39-82.
- Fisher, R. A. (1946). *Statistical methods for research workers* (10<sup>th</sup> Ed.). Edinburgh: Oliver and Boyd.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3<sup>rd</sup> Ed.). New York: Wiley.
- Gart, J. J. (1971). The comparison of proportions: a review of significance tests, confidence intervals and adjustments for stratification. *Review of the International Statistical Institute*, 39, 148-169.
- Harkness, W. L. (1965). Properties of extended hypergeometric distribution. *Annals of Mathematical Statistics*, 36, 938-945.
- Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research: principles and quantitative methods*. New York: Wiley.
- Lee, T-S. (2009). Adjusted odds ratio for the exposure misclassification in a case-control study. *The Internet Journal of Epidemiology*, 6(2). Accessed from <http://www.ispub.com/journal/the-internet-journal-of-epidemiology/volume-6-number-2/bias-adjusted-exposure-odds-ratio-for-misclassified-data-1.html>.
- Levin, B. (1984). Simple improvements on Cornfield's approximation to the mean of a noncentral hypergeometric random variable. *Biometrika*, 71, 630-632.
- Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system* (2<sup>nd</sup> Ed.). Cary, North Carolina: SAS Institute, Inc.

## Model Diagnostics for Proportional and Partial Proportional Odds Models

Ann A. O’Connell  
The Ohio State University,  
Columbus, OH USA

Xing Liu  
Eastern Connecticut State University,  
Willimantic, CT USA

---

Although widely used to assist in evaluating the prediction quality of linear and logistic regression models, residual diagnostic techniques are not well developed for regression analyses where the outcome is treated as ordinal. The purpose of this article is to review methods of model diagnosis that may be useful in investigating model assumptions and in identifying unusual cases for PO and PPO models, and provide a corresponding application of these diagnostic methods to the prediction of proficiency in early literacy for children drawn from the kindergarten cohort of the Early Childhood Longitudinal Study (ECLS-K; NCES, 2000).

Key words: Model diagnostics, proportional odds models, partial proportional odds models, residual analyses.

---

### Introduction

Although widely used to assist in evaluating the prediction quality of linear and logistic regression models, residual diagnostic techniques are not well developed for regression analyses where the outcome is treated as ordinal. For ordinal regression models, Hosmer and Lemeshow (2000) suggested recombining outcomes according to the ordinal structure of the data and applying residual strategies developed for binary logistic models, such as outlined in Pregibon (1981). This approach is useful in the investigation of the assumption of proportionality as well as for examination of unusual or extreme values via residual diagnostics and this article presents guidelines for proportional odds (PO) and non- or partial-proportional odds (PPO) models.

Residual analyses provide rich opportunities for researchers to examine model fit and misfit, and require going beyond the results obtained through a direct application of a statistical model, the interpretation of parameter estimates or the summary statistics obtained from that model. Studies of residuals are becoming an important analytic process in many research situations, for example, when high-performing or low-performing students or schools are selected for intensive investigation. Despite their importance, however, results are often presented in the research literature with little emphasis on or reference to the model residuals; readers are thus not always provided with a clear understanding of study findings. In the education field, it becomes particularly important to be able to reliably identify children (or schools, or teachers, or program participants, etc.) whose response or outcome may not be adequately represented by a particular derived model, because if such unusual cases can be discerned, attention may be directed to improve desired outcomes.

Extensive outlines of useful residual analyses and diagnostic measures have been provided for logistic (Pregibon, 1981) and linear (Fox, 1991) regression models. In addition, Bender and Benner (2000) suggested some graphical strategies that can be used to examine the feasibility of the proportional odds

---

Ann A. O’Connell is a Professor in the Program in Quantitative Research, Evaluation, and Measurement (QREM) in the School of Educational Policy and Leadership within the College of Education and Human Ecology. Email her at: [aconnell@ehe.osu.edu](mailto:aconnell@ehe.osu.edu). Xing Liu is an Associate Professor of Research and Assessment in the Education Department. Email him at: [liux@easternct.edu](mailto:liux@easternct.edu).



## MODEL DIAGNOSTICS FOR PO & PPO MODELS

assumption. However, analysis of residuals for PO and PPO models in ordinal logistic regression is not well established; thus, this study was designed to build on the collection of strategies available through logistic and linear approaches. Specifically, these include: Pearson residuals, Deviance residuals, Pregibon leverages, DFBeta's and the use of index plots and other graphical strategies to examine and isolate unusual cases within the logistic framework. The use of Mahalanobis' distance, leverages, SDResiduals, Cook's D and other statistics from the ordinary least-squares framework, when applied to ordinal data are also investigated.

This study contributes to the empirical literature on detection of extreme or unusual cases, investigation of statistical assumptions and validation of ordinal regression models by: reviewing methods of model diagnosis that may be useful in investigating model assumptions, identifying unusual cases for PO and PPO models and providing a corresponding application of these diagnostic methods to the prediction of proficiency in early literacy for children drawn from the kindergarten cohort of the Early Childhood Longitudinal Study (ECLS-K; NCES, 2000). The primary focus is on how outlying or influential cases in ordinal logistic regression models can be reliably detected and on how these strategies can be applied to

proportional and/or partial proportional odds models.

### Background

One of the most commonly used models for the analysis of ordinal data comes from the class of logistic models: the PO model. Consider a simple binary model; in a binary logistic model, the data represent two possible ordinal outcomes, success or failure, typically coded 0 for failure and 1 for success. For a  $K$ -level ordinal outcome, several different conceptualizations of success can be derived. Table 1 shows the ECLS-K ordinal outcome variable description and the data indicating the proportion of kindergarten children, drawn from a national random sample of kindergarteners followed through the third-grade, attaining mastery of five hierarchical early-literacy skills at the end of the kindergarten year. In this example,  $K=6$ , and the outcome values are scored as 0, 1, 2, 3, 4 and 5, to represent the highest level of proficiency attained on the ECLS-K literacy mastery test (0 = no mastery at any level; 5 = mastered all 5 levels). For these data, 26.9% of the children were not able to achieve beyond level 1 at the end of the kindergarten year and only 12.8% of these children mastered literacy skills beyond level 3, most students scored in levels 2 and 3 (60.3%).

Table 1: Proficiency Categories and Frequencies (Proportions) for the Study Sample, ECLS-K Measures for Early Literacy and  $N = 2687$  Public School Children, End of Kindergarten Year

Proficiency Category	Description	Frequency
0	Did not pass level 1	295 (11.0%)
1	Can identify upper/lowercase letters	427 (15.9%)
2	Can associate letters with sounds at the beginnings of words	618 (23.0%)
3	Can associate letters with sounds at the ends of words	1003 (37.3%)
4	Can recognize sight words	233 (8.7%)
5	Can read words in context	111 (4.1%)

A particular series of questions are of interest when analyzing ordinal outcome data; these involve predicting the likelihood that an observation is at or beyond each specific outcome level given a collection of explanatory variables. For the ECLS-K data, this involves estimating the probability that a child with particular background characteristics or a given set of explanatory variables is at or beyond level 0 (which would always be 1.0); then estimating the probability of that same child being at or beyond level 1, at or beyond level 2, etc., until reaching the probability of the child being at or beyond the last, or  $K^{\text{th}}$ , outcome category. This series of probabilities are referred to as cumulative probabilities.

The analysis that mimics this method of dichotomizing the outcome, in which the successive dichotomizations are used to form cumulative splits to the data, is referred to as the proportional or cumulative odds model (PO) (Agresti, 2000, 2007; Armstrong & Sloan, 1989; Long, 1997; Long & Freese, 2006; McCullagh, 1980; McCullagh & Nelder, 1989; O'Connell, 2006; Powers & Xie, 2000). The model is defined as:

$$\begin{aligned} \ln(Y_j') &= \text{logit} [\pi(x)] \\ &= \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) \\ &= \alpha_j + (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \end{aligned} \quad (1)$$

In this logistic model, the prediction represents the expected logit for being in category  $j$  or above, conditional on the collection of predictors, and  $Y_j'$  represents the odds of being in higher proficiency categories. These predicted logits can be transformed to odds and then to estimated probability:

$$P(Y \geq j) = \frac{\exp(\ln(Y_j'))}{1 + \exp(\ln(Y_j'))}$$

The intercept,  $\alpha_j$ , represents the threshold, or cutpoint, for each particular split to the data. Each person thus has  $K-1$  predicted values, representing their estimated likelihood of

scoring in category  $j$  or beyond, given their explanatory data. Note that in the PO model the effect of each predictor remains the same across each of these  $K-1$  prediction models: This means that for each predictor, its effect on the probability of being at or beyond any category is assumed to remain constant within the model; thus, the slope estimate provides a summary of each independent variable's relationship to the outcome across all cutpoints. In this model,  $\beta_1$ , for example, remains the same for all of the splits, although  $\alpha_j$  may change. This restriction is referred to as the assumption of proportional odds.

A model that relaxes the assumption of proportional odds is referred to as a partial-proportional odds (PPO) or non-proportional odds model. This model is given by:

$$\begin{aligned} \ln(Y_j') &= \ln \left( \frac{\pi_j(\underline{x})}{1 - \pi_j(\underline{x})} \right) \\ &= \alpha_j + (\beta_{1j} X_1 + \beta_{2j} X_2 + \dots + \beta_{pj} X_p). \end{aligned} \quad (2)$$

In this expression, all of the effects of the explanatory variables are allowed to vary across each of the cutpoints. If some of the effects are found to be stable, they can be held constant as in the PO model. Thus, partial-proportional odds refers to the case that at least one of the slopes for an explanatory variable varies across splits.

Due to its simplicity and natural correspondence to ordinary logistic regression, the proportional odds model is the most widely used ordinal regression model. Tests for the assumption of proportional odds can be very liberal (Peterson & Harrell, 1990), however, and are strongly affected by sample size and the number of covariate patterns - which will always be large if continuous covariates are used (Allison, 1999; Brant, 1990; Clogg & Shihadeh, 1994). Researchers have argued that if the assumption of proportional odds is rejected, good practice would dictate that the corresponding underlying binary models be fit and compared with the PO results to check for discrepancies or deviations from the general pattern suggested by the PO model (e.g., Allison, 1999; Bender & Grouven, 1998; Brant, 1990; Clogg & Shihadeh, 1994; Long, 1997;

## MODEL DIAGNOSTICS FOR PO & PPO MODELS

O'Connell, 2000, 2006). This strategy of considering the PO model as a collection of underlying binary models is an approach that has been found useful not only in qualifying the nature of the proportionality assumption, but also in assessing univariate proportionality, linearity in the logit, and the distribution of residuals from the PO model (O'Connell, Liu, Zhao & Goldstein, 2004).

### Methodology

#### Sample

The data were drawn from the public-use data base Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K) (NCES, 2000). The outcome variable of interest was proficiency in early reading, assessed at the end of the kindergarten year. Actual data were used rather than simulated data, because this could help create a realistic context for conducting residual analyses.

The final sample included  $n = 2687$  public school children sampled from within 198 public schools across the U.S. The sample contained first-time kindergarten children only, who remained in the same school between kindergarten and first-grade. Analyses intentionally ignored school-level effects in order to focus on residual diagnostics for single-level models (only public schools were selected for this study). Analyses were conducted for the full sample as well as for 10 different randomly selected subsamples of 10% each to facilitate use of casewise statistics and plots.

Residual analyses are often very intensive, thus, for demonstration purposes, two of the smaller subsets were selected to highlight interesting patterns and residual statistics within each data set. No attempt was made to draw inference to the overall sample or population; rather focus was placed on demonstration of diagnostic procedures and strategies for ordinal data.

The proficiency outcomes were obtained from the third-grade release of the ECLS-K data base (prior to that, researchers used a series of dichotomous variables to derive the ordinal proficiency scores). Proficiency is defined as mastery of a cluster of 4 items representing each of the domains outlined in Table 1. The domains are hierarchically structured and theoretically

assumed to follow the Guttman scale (NCES, 2000). Mastery is recognized as students passing 3 out of the 4 items representing each domain.

The selection of explanatory variables was theoretically driven and supported through prior research on factors affecting early childhood literacy. These included: gender (male = 1), minority status (minority = 1), whether the child attended half-day kindergarten (yes = 1), number of family risks (0 to 4), frequency with which parents read books to child (0 to 3), family socio-economic status (continuous), and assessment age (continuous). The data presented were from the end of the Kindergarten year. Tables 2a, 2b and 2c present descriptive statistics for the full-sample and the two subsamples, respectively.

#### Data and Models

The data were used to inspect the residuals from the PO model and test the assumptions of equal slopes. Residuals from an OLS regression of the same data as well as from the five corresponding cumulative binary logistic regression models (splits) underlying the proportional odds assumption (i.e., level 0 versus beyond level 0; levels 0 and 1 combined versus beyond level 1; levels 0, 1, and 2 combined versus beyond, etc.) were examined. Logistic regression diagnostics were investigated for each cumulative split to the data; these procedures were repeated for two of the 10% subsamples (referred to as Samples I and II). The study began by investigating the plausibility of the PO assumption in the full- and sub-samples. A PPO model was then fit where the effect of minority was allowed to vary across thresholds.

The SAS (V. 9.1.3), SPSS (V. 15.0) and Stata (V. 9.0) software packages were used for data analyses and graphing. The options for residual diagnostics in logistic regression models were also compared among these packages. SAS PROC LOGISTIC procedure was used for binary logistic models and ordinal logistic models. SPSS was used for descriptive statistics, casewise residual diagnostics in OLS and ordinal regression, index plotting and some scatterplots. Stata was used for residual diagnostics.

Table 2a: Descriptive Statistics at the End of Kindergarten, n = 2,687 (Full-Sample)

	Reading Proficiency Level						Total n = 2,687 100%
	0 n = 295 11.0%	1 n = 427 15.9%	2 n = 618 23.0%	3 n = 1,003 37.3%	4 n = 233 8.7%	5 n = 111 4.1%	
% Male	55%	57%	47%	50%	43%	42%	50%
% Minority	82.03%	51.05%	47.73%	33.40%	38.20%	32.43%	45.22%
Risknum Mean (sd)	1.48 (1.07)	.76 (.90)	.60 (.82)	.44 (.72)	.33 (.62)	.23 (.53)	.62 (.87)
Wksesl Mean (sd)	-0.7080 (0.64)	-0.2678 (0.67)	-0.1017 (0.72)	0.1278 (0.71)	0.2679 (0.72)	0.68 (0.75)	-0.0446 (0.77)
% Halfday	58.98%	47.07%	48.87%	49.55%	41.20%	54.05%	49.50%
% Readbk2	66.10%	74.47%	77.83%	84.85%	87.98%	98.20%	80.35%
r2_kage Mean (sd)	75.18 (4.15)	75.64 (4.69)	75.58 (4.47)	75.23 (4.37)	75.12 (4.43)	75.34 (4.20)	75.37 (4.42)

Table 2b: Descriptive Statistics at the End of Kindergarten, n = 244 (Sample I)

	Reading Proficiency Level						Total n = 244 100%
	0 n = 26 10.7%	1 n = 32 13.1%	2 n = 54 22.1%	3 n = 108 44.3%	4 n = 16 6.6%	5 n = 8 3.3%	
% Male	46%	66%	59%	53%	38%	38%	54%
% Minority	73.08%	50.00%	55.56%	33.33%	37.50%	37.50%	45.08%
Risknum Mean (sd)	1.38 (1.13)	0.75 (0.92)	0.69 (1.01)	0.50 (0.76)	0.25 (0.77)	0.25 (0.46)	0.64 (0.92)
Wksesl Mean (sd)	-0.8108 (0.55)	-0.2256 (0.69)	0.0057 (0.70)	0.1073 (0.76)	0.1019 (0.71)	0.2250 (0.56)	-0.0532 (0.76)
% Halfday	69.23%	46.88%	42.59%	49.07%	56.25%	50.00%	50.00%
% Readbk2	69.23%	71.88%	77.78%	81.48%	93.75%	87.50%	79.10%
r2_kage Mean (sd)	75.74 (4.53)	75.44 (3.24)	75.81 (4.93)	74.57 (4.30)	74.44 (3.94)	77.95 (4.64)	75.18 (4.37)

MODEL DIAGNOSTICS FOR PO & PPO MODELS

Table 2c: Descriptive Statistics at the End of Kindergarten, n = 278 (Sample II)

	Reading Proficiency Level						Total n = 278 100%
	0 n = 23 8.3%	1 n = 47 16.9%	2 n = 65 23.4%	3 n = 102 36.7%	4 n = 27 9.7%	5 n = 14 5.0%	
% Male	57%	68%	58%	46%	56%	29%	54%
% Minority	86.96%	44.68%	44.62%	25.49%	33.33%	42.86%	39.93%
Risknum Mean (sd)	1.39 (1.12)	0.66 (0.87)	0.41 (0.66)	0.45 (0.78)	0.19 (0.40)	0.07 (0.27)	0.51 (0.81)
Wksesl Mean (sd)	-0.8191 (0.50)	-0.2296 (0.53)	0.0534 (0.66)	0.1488 (0.70)	0.2315 (0.75)	0.9864 (0.81)	0.0327 (0.75)
% Halfday	56.52%	51.06%	55.38%	50.00%	40.74%	42.86%	50.72%
% Readbk2	65.22%	78.72%	78.46%	85.29%	85.19%	100.00%	81.65%
r2_kage Mean (sd)	74.62 (3.84)	75.12 (4.92)	75.20 (4.16)	75.81 (4.07)	74.94 (4.83)	75.20 (3.77)	75.34 (4.27)

Results

Interpretation of Proportional Odds for Full-Sample (n=2678)

The model including all seven predictors identified in Tables 2a, 2b, and 2c is referred to the full model. Table 3a provides a results summary for the fitted PO model with seven explanatory variables for the full-sample. The results were obtained using SAS with the descending option (see O’Connell (2006) for details on fitting ordinal regression models). The score test yielded  $\chi^2_{32} = 131.53$  ( $p < 0.0001$ ), indicating that the proportional odds assumptions for the full-model was not upheld. This suggested that the effect of one or more of the explanatory variables was likely to differ across separate binary models fit to the cumulative cutpoints. The Cox & Snell  $R^2 = 0.210$ , Nagelkerke  $R^2 = 0.219$ , and the likelihood ratio  $R^2_L = 0.074$  all suggested that the relationship between the response variable, proficiency and the seven predictors is small.

However, the model fit statistic  $\chi^2_7 = 633.55$  ( $p < 0.0001$ ), indicated that the full model provided a better fit than the null model with no independent variables in predicting cumulative probability for proficiency.

Because the proficiency was measured through six categories with outcomes as 0, 1, 2, 3, 4 or 5, with the descending option in SAS,  $\alpha_5$  corresponds to the intercept for the cumulative logit model for  $Y \geq 5$ ,  $\alpha_4$  corresponds to the intercept for the cumulative logit model for  $Y \geq 4$ , and so on. The effects of the seven independent variables can be interpreted as how variables contribute to the log of the odds of being at or beyond a particular category. In terms of odds ratios, boys were less likely than girls to be at or beyond a particular category (OR=0.73). Being in a minority category (OR=0.649), the presence of any family risk factor (OR=0.649) and attending half-day kindergarten rather than full-day kindergarten (OR=0.695) all had significant negative coefficients in the model and corresponding

Table 3a: Proportional Odds Model and OLS Regression Model for Full-Sample Data, n = 2687 Public School Children (Y &gt; cat.j)

Variable	Proportional Odds Model	OR	OLS Model
	b (se(b))		b (se(b))
$\alpha_5$	-2.88 (0.62)		
$\alpha_4$	-1.59 (0.61)		
$\alpha_3$	0.58 (0.61)		
$\alpha_2$	1.74 (0.61)		
$\alpha_1$	3.02 (0.61)		2.64 (0.38)
Gender <sup>δ</sup>	-0.31 (0.07)**	0.730	-0.21 (0.04)**
Minority	-0.43 (0.08)**	0.649	-0.26 (0.05)**
RiskNum	-0.36 (0.05)**	0.695	-0.23 (0.03)**
Halfday	-0.48 (0.07)**	0.619	-0.27 (0.04)**
Readbk2	0.35 (0.09)**	1.422	0.22 (0.06)**
Wksesl	0.78 (0.06)**	2.183	0.48 (0.03)**
R2_kage	-0.001 (0.01)	0.999	0 (0.01)
$R^2$	$R^2_L = .074$		$R^2 = .218$
Cox & Snell $R^2$	.210		
Nagelkerke $R^2$	.219		
Somer's D	.386		
Model Fit <sup>a</sup>	$\chi^2_7 = 633.55$ (p < 0.0001)		F(7, 2679) = 106.76**
Deviance	7869.82 (df = 13398)	0.5874 <sup>c</sup>	
Pearson $X^2$	13032.47 (df = 13398)	0.9727 <sup>c</sup>	
Score Test <sup>b</sup>	$\chi^2_{32} = 131.53$ (p < 0.0001)		

Notes: <sup>δ</sup>gender: male=1; <sup>a</sup>Likelihood ratio test; <sup>b</sup>For the proportional odds assumption; <sup>c</sup>Value/df; \*Significant at p < 0.05; \*\* p < 0.01

OR's were significantly less than 1.0. These characteristics were associated with a child being in lower proficiency categories rather than in higher categories. Conversely, increasing frequency of parents reading to their children (OR=1.422) and family SES (OR=2.183) had positive effects on children being in higher

proficiency categories. The slopes for both variables were positive and significantly different from zero in the model; child's assessment age was not associated with proficiency in this model because the slope was almost 0 and the OR is close to 1.0.

## MODEL DIAGNOSTICS FOR PO & PPO MODELS

Interpretation of Proportional Odds Models for Sub-samples I and II

Table 3b shows the results summary for the fitted PO model for sub-samples I ( $n = 244$ ) and II ( $n = 278$ ): both results were similar to that of the full-sample. An  $\alpha = 0.05$  was used to assess the hypothesis of proportionality. The score test for sub-sample I,  $\chi^2_{28} = 46.67$  ( $p = 0.0148$ ), and the score test for sub-sample II  $\chi^2_{28} = 44.41$  ( $p = 0.0253$ ), were both statistically significant, indicating that the proportional odds assumptions for both sub-models were violated.

The model fit statistic for sub-sample I,  $\chi^2_7 = 38.61$  ( $p < 0.0001$ ), and the model fit statistic for sub-sample II,  $\chi^2_7 = 71.67$  ( $p < 0.0001$ ) indicated that the models with seven predictors provided a better fit than the null model with no independent variables. The Odds Ratios (OR) for the seven explanatory variables for sub-sample I and II looked similar, and were also similar to those of the proportional odds model for the full-sample. However, in the PO model for sub-sample I, it was noticeable that the effects of gender, minority and frequency of being read to by parents were not statistically significant; and the  $p$  value of the slopes of number of family risks and attendance at half-day kindergarten were slightly larger than the 0.05 level. In the CO model for sub-sample II, the effects of minority, half-day kindergarten and having parents read to their children were not significant.

### Assumption of Proportional Odds

Table 4a shows the results of five separate binary logistic regression analyses for the full sample, where the data were dichotomized according to the cumulative probability pattern described earlier for the proportional odds model. Each logistic regression model estimates the probability of being at or beyond proficiency level  $j$ . In the data set, the grouping of categories coded 1 corresponded to children who were at or beyond each proficiency category and 0 was coded for children below each successive category. The model  $\chi^2$  for each separate logistic model was statistically significant, indicating that each model fit well compared to its corresponding null model. The Hosmer-Lemeshow tests were

all not statistically significant, indicating that observed and predicted probabilities were consistent.

Examining the patterns of slopes and ORs for each explanatory variable across these five logistic regression models, it was found that the effects of gender, after adjusting for the other predictors directionally and on average, were similar across the five separate logistic regressions. This was also true for family risk, family SES, half-day kindergarten, being read to by parents and the child's assessment age. However, the effect of minority did appear to present a dissimilar pattern across the five separate logistic regressions. The direction of the effect of minority changed between the first three regressions and the last two. For the other explanatory variables, the direction and average magnitude of the slopes and the ORs from the logistic models were similar to those of the PO model.

Because the proportional odds assumption for the full-sample ordinal model was violated, separate score tests unadjusted for the presence of other covariates in the cumulative odds model were examined for each of the explanatory variables, in order to illuminate where non-proportionality might lie. The univariate score tests for the assumption of proportional odds were upheld for gender and child's assessment age. However, the univariate score tests were violated for minority, family risk, family SES, being read to by parents and half-day kindergarten at the 0.05 level of significance. The  $p$ -values for these unadjusted tests are presented in the final column of Table 4a; it should be noted that these score tests are simply descriptive, given their univariate nature.

Table 4b presents the results of five separate binary logistic regression analysis for sub-sample I,  $n = 244$ . The univariate score tests for the assumption of proportional odds were upheld for separate PO models for all the variables, except for the continuous variable of family SES (*wksesl*). The  $p$  values for these unadjusted tests are presented in the final column of Table 4b. However, minority as well as gender, half-day kindergarten attendance and frequency of being read to by parents all exhibited inconsistencies in the directional patterns across the binary splits.

Table 3b: Subsamples I (n = 244) and II (n = 278): Proportional Odds Models (Y &gt; cat.j)

Variable	Sample I		Sample II	
	b (se(b))	OR	b (se(b))	OR
$\alpha_5$	-1.78 (2.11)		-5.68 (2.00)	
$\alpha_4$	-0.60 (2.09)		-4.41 (1.99)	
$\alpha_3$	1.96 (2.09)		-2.35 (1.97)	
$\alpha_2$	3.08 (2.10)		-1.14 (1.97)	
$\alpha_1$	4.16 (2.11)		0.40 (1.98)	
Gender <sup>δ</sup>	-0.29 (0.24)	0.750	-0.45 (0.22)*	0.638
Minority	-0.40 (0.28)	0.670	-0.38 (0.24)	0.681
RiskNum	-0.29 (0.15) <sup>a</sup>	0.752	-0.31 (0.16) <sup>*</sup>	0.733
Halfday	-0.45 (0.25) <sup>a</sup>	0.638	-0.33 (0.21)	0.716
Readbk2	0.27 (0.30)	1.312	0.24 (0.29)	1.266
Wksesl	0.58 (0.20)**	1.793	0.94 (0.18)**	2.552
R2_kage	-0.02 (0.03)	0.983	0.04 (0.03)	1.039
$R^2_L$	0.053		0.081	
Cox & Snell $R^2$	0.146		0.227	
Nagelkerke $R^2$	0.154		0.237	
Somer's D	0.322		0.393	
Model Fit <sup>b</sup> $\chi^2_7$	38.61 (p < 0.0001)		71.67 (p < 0.0001)	
Deviance	688.63 (df = 1208)	0.5701 <sup>d</sup>	813.11 (df = 1378)	0.5901 <sup>d</sup>
Pearson $X^2$	1190.51 (df = 1208)	0.9855 <sup>d</sup>	1221.86 (df = 1378)	0.8867 <sup>d</sup>
Score Test <sup>c</sup> $\chi^2_{28}$	46.67 (p = 0.0148)		44.41 (p = 0.0253)	

Notes: <sup>δ</sup>gender: male=1; <sup>a</sup>p = 0.06 for risknum; p = 0.07 for halfday; <sup>b</sup>Likelihood ratio test; <sup>c</sup>For the proportional odds assumption; <sup>d</sup>Value/df; \*Significant at p < 0.05; \*\* p < 0.01

Table 4c shows the results of five separate binary logistic regression analysis for sub-sample II, n = 278. Based on  $\alpha = 0.05$ , the univariate score tests for the assumption of proportional odds were upheld for separate PO models for these variables: gender, attending half-day kindergarten (halfday), having parents

read to their children (readbk2) and child's assessment age (r2\_kage). However, the score tests for the assumptions of proportional odds were violated for these three variables: being in a minority category (minority), number of family risks (risknum) and family SES (wksesl). The p-values for these unadjusted tests were presented in the final column of Table 4c.



MODEL DIAGNOSTICS FOR PO & PPO MODELS

Table 4a: Associated Cumulative Binary Models for the CO Analysis (Descending) for the Full-sample, Where CUMSPj Compares  $Y < \text{cat. } j$  to  $Y > \text{cat. } j$ ,  $n = 2687$

Variable	CUMSP1 b (se(b)) OR	CUMSP2 b (se(b)) OR	CUMSP3 b (se(b)) OR	CUMSP4 b (se(b)) OR	CUMSP5 b (se(b)) OR	Score Test <sup>b</sup> P value
Constant	2.53 (1.22)	1.88 (0.82)	1.10 (0.72)	-1.77 (1.04)	-5.84 (1.86)	
Gender <sup>δ</sup>	-0.33 (0.14) 0.72*	-0.44 (0.10) 0.64**	-0.23 (0.08) 0.80**	-0.40 (0.12) 0.67**	-0.41 (0.20) 0.67*	0.10
Minority	-1.21 (0.18) 0.30**	-0.49 (0.10) 0.61**	-0.47 (0.09) 0.62**	0.11 (0.13) 1.12	0.17 (0.23) 1.18	0.001
Risknum	-0.50 (.08) 0.61**	-0.33 (0.06) 0.72**	-0.23 (0.06) (0.79)**	-0.35 (0.11) 0.70**	-0.32 (0.21) 0.73	0.001
Wksesl	0.94 (0.13) 2.57**	0.83 (0.09) 2.29**	0.73 (0.07) 2.08**	0.79 (0.09) 2.19**	1.12 (0.14) 3.08**	0.001
Halfday	-0.85 (0.14) 0.43**	-0.42 (0.10) 0.65**	-0.37 (0.09) 0.69**	-0.42 (0.12) 0.65**	-0.12 (0.20) 0.89	0.004
Readbk2	0.19 (0.15) 1.21	0.25 (0.11) 1.28*	0.37 (0.11) 1.45**	0.60 (0.21) 1.83**	2.00 (0.72) 7.40**	0.043
r2_kage	0.19 (0.15) 1.02	0.00 (0.01) 1.00	-0.01 (0.01) 0.99	0.00 (0.01) 1.00	0.01 (0.02) 1.01	0.47
$R^2_L$	0.237	0.134	0.105	0.088	0.134	
Naglekerke $R^2$	0.30	0.21	0.18	0.12	0.16	
Model $\chi^2_7$	441.26**	418.37**	389.97**	180.89**	124.18**	
H-L <sup>a</sup> $\chi^2_8$	5.80	8.07	7.94	3.81	13.09	

Notes: <sup>δ</sup> gender: male=1; \*Significant at  $p < 0.05$ ;  $p < 0.01$ ; <sup>a</sup>Hosmer-Lemeshow test; <sup>b</sup>Score test for each IV, unadjusted (no other covariates in the model)

Table 4b: Associated Cumulative Binary Models for the CO Analysis (Descending) for Sub-sample I, Where CUMSP<sub>j</sub> Compares Y < cat. j to Y > cat. j, n = 244

Variable	CUMSP1 b (se(b)) OR	CUMSP2 b (se(b)) OR	CUMSP3 b (se(b)) OR	CUMSP4 b (se(b)) OR	CUMSP5 b (se(b)) OR	Score Test <sup>b</sup> P value
Constant	3.43 (4.29)	2.52 (.2.84)	4.26 (2.45)	-4.46 (4.05)	-15.79 (6.68)	
Gender <sup>δ</sup>	0.37 (0.48) 1.45	-0.22 (0.33) 0.80	-0.40 (0.28) 0.67	-0.71 (0.46) 0.49	-0.69 (0.78) 0.50	0.38
Minority	-0.61 (0.58) 0.54	-0.23 (0.38) 0.79	-0.65 (0.31) 0.52*	0.18 (0.50) 1.20	0.27 (0.84) 1.31	0.36
Risknum	-0.24 (0.26) 0.78	-0.16 (0.19) 0.85	-0.22 (0.18) (0.80)	-0.76 (0.42) 0.47	-0.71 (0.71) 0.49	0.47
Wksesl	1.74 (0.53) 5.71**	1.09 (0.32) 2.98**	0.38 (0.23) 1.47	0.13 (0.36) 1.24	0.55 (0.59) 1.73	0.01
Halfday	-1.66 (0.54) 0.19**	-0.74 (0.34) 0.48*	-0.30 (0.28) 0.74	0.11 (0.46) 1.11	0.01 (0.77) 1.01	0.26
Readbk2	-0.11 (0.59) 0.89	0.18 (0.40) 1.19	0.23 (0.35) 1.26	0.78 (0.78) 2.18	0.07 (1.14) 1.07	0.90
r2_kage	0.01 (0.06) 1.01	-0.01 (0.04) 0.99	-0.05 (0.01) 0.96	0.03 (0.05) 1.03	0.17 (0.08) 1.18	0.13
R <sup>2</sup> <sub>L</sub>	0.265	0.127	0.082	0.072	0.104	
NaglekerkeR <sup>2</sup>	0.33	0.19	0.14	0.10	0.12	
Model $\chi^2_7$	43.83**	33.89**	27.73**	11.34	7.29	
H-L <sup>a</sup> $\chi^2_8$	14.31	14.59	11.25	8.89	4.14	

Notes: <sup>δ</sup> gender: male=1; \*Significant at p < 0.05; <sup>a</sup>Hosmer-Lemeshow test; <sup>b</sup>Score test for each IV, unadjusted (no other covariates in the model)

MODEL DIAGNOSTICS FOR PO & PPO MODELS

Table 4c: Associated Cumulative Binary Models for the CO Analysis (Descending) for Sub-Sample II, Where CUMSPj Compares Y < cat. j to Y > cat. j, n = 278

Variable	CUMSP1 b (se(b)) OR	CUMSP2 b (se(b)) OR	CUMSP3 b (se(b)) OR	CUMSP4 b (se(b)) OR	CUMSP5 b (se(b)) OR	Score Test <sup>b</sup> P value
Constant	-4.16 (4.50)	-2.03 (2.79)	-2.09 (2.26)	-0.66 (3.20)	-13.56 (284.9)	
Gender <sup>δ</sup>	-0.02 (0.53) 0.98	-0.63 (0.32) 0.53	-0.62 (0.26) 0.54	-0.26 (0.36) 0.77	-0.97 (0.67) 0.38	0.33
Minority	-1.41 (0.71) 0.24	-0.42 (0.34) 0.66	-0.71 (0.29) 0.49*	0.42 (0.39) 1.53	0.81 (0.64) 2.24	0.00
Risknum	-0.14 (0.29) 0.87	-0.23 (0.21) 0.80	-0.08 (0.20) (0.93)	-1.00 (0.44) 0.37*	-1.65 (1.10) 0.19	0.04
Wksesl	2.60 (0.73) 13.43**	1.24 (0.32) 3.44**	0.71 (0.21) 2.04**	0.82 (0.25) 2.23**	1.59 (0.42) 4.91**	0.00
Halfday	-0.49 (0.54) 0.61	-0.23 (0.31) 0.98*	-0.37 (0.26) 0.69	-0.53 (0.36) 0.59	-.47 (0.62) 0.62	0.93
Readbk2	0.17 (0.57) 1.19	-0.02 (0.38) 0.98	0.25 (0.35) 1.28	0.32 (0.58) 1.38	11.60 (284.8) 1000.00	0.65
R2_Kage	0.12 (0.06) 1.13	0.05 (0.04) 1.06	0.04 (0.03) 1.04	-0.01 (0.04) 1.00	-0.01 (0.08) 1.00	0.77
$R^2_L$	0.351	0.163	0.112	0.121	0.297	
NaglekerkeR <sup>2</sup>	0.42	0.25	0.19	0.17	0.34	
Model $\chi^2_7$	55.70**	51.25**	43.08**	28.11**	32.94**	
H-L <sup>a</sup> $\chi^2_8$	4.17	5.07	6.99	2.63	2.22	

Notes: <sup>δ</sup> gender: male=1; \*Significant at p < 0.05; <sup>a</sup>Hosmer-Lemeshow test; <sup>b</sup>Score test for each IV, unadjusted (no other covariates in the model)

For comparison, Table 4d presents the p-values for univariate score tests of proportionality for each explanatory variable analyzed separately in a single variable model for the full-sample and for sub-samples I and II provided for SPSS, SAS and Stata. SPSS performs an approximation to the score test in their PLUM procedure (Nichols, 2004), so analysts should be aware of the possibility for discrepancies and differences in results between software packages.

Given the large sample size,  $\alpha = 0.01$  was used to evaluate the assumption of proportionality for these univariate tests for the full-sample. Consistent results were found across the three software packages for all explanatory variables except the frequency with which parents read books to children, for which  $p = 0.006$  (SPSS),  $p = 0.0426$  (SAS) and  $p = 0.078$  (Stata). However, for the two smaller sub-samples, and using  $\alpha = 0.05$ , it was found that the results of these univariate score tests (using SAS) varied across the sub-samples and were also inconsistent with the full-sample results. For example, the p-values of the score test for the minority variable was 0.3562 in the model for sub-sample I and 0.0031 for sub-sample II; the hypothesis of proportionality was rejected for the minority variable within the full sample ( $p < 0.0001$ ). In addition, the effect of attending

half-day kindergarten was found to deviate from proportionality in the full sample, but this assumption was upheld in both of the smaller samples.

Results of the Partial Proportional Odds (PPO) Model

The discrepancy of results of the score tests between the full-sample and sub-sample analyses and across statistical packages presents a disheartening situation for the analyst attempting to assess the plausibility of the proportional odds model through score tests alone. These results support the view that investigation of proportional odds may be more reasonably investigated through visual examination of the variable effects and odds ratios of the binary models underlying the ordinal progression of the outcome data. Consequently, it was decided to fit a PPO model that relaxes the assumption of proportionality for the minority variable because this effect changed direction across cutpoints in all three analyses (full-sample, I and II). Results using SAS PROC GENMOD are shown in Table 4e for the full-sample data.

The outcome being modeled in this PPO analysis was the probability that a child was at or beyond category  $j$ , with the effect of minority being allowed to vary across the  $K-1 = 5$

Table 4d: Score tests for Proportion Odds for each IV in the Model: Single Variable Models for the Full Sample and for Subsamples I and II.

IV	Full Sample, p-value SPSS	Full Sample, p-value SAS	Full Sample, p-value Stata	Sample I p-value SAS	Sample II p-value SAS
Gender <sup>δ</sup>	0.101	0.1013	0.102	0.3792	0.3313
Minority	0.000	< 0.0001	0.000	0.3562	0.0031
Risknum	0.000	< 0.0001	0.000	0.4698	0.0440
Halfday	0.004	0.0041	0.005	0.2643	0.8821
Readbk2	0.006	0.0426	0.078	0.8969	0.6471
Wksesl	0.000	< 0.0001	0.000	0.0119	0.0004
R2_Kage	0.467	0.4700	0.445	0.1263	0.7706

Notes: <sup>δ</sup>gender: male=1

MODEL DIAGNOSTICS FOR PO & PPO MODELS

Table 4e: Partial Proportional Odds (PPO) Model for Full-sample, SAS (Descending) ( $Y \geq \text{cat. } j$ ), N = 2687

Variable	b (se(b))	OR
Intercept	-3.56 (0.63)**	0.03
Gender (0)	0.34 (0.07)**	1.40
Minority (0)	0.06 (0.21)	
Risknum	-0.33 (0.05)**	0.72
Wksesl	0.83 (0.06)**	2.29
Halfday (0)	0.44 (0.07)**	1.55
Readbk2	-0.34 (0.09)**	0.71
R2_Kage	0.00 (0.00)	1.00
Split 1	5.46 (0.18)**	6.69
Split 2	4.42 (0.17)**	2.36
Split 3	3.26 (0.17)**	0.74
Split 4	1.39 (0.15)**	0.11
Split 5	0.00 (0.00)**	0.03
Split 1* Minority (0)	1.23 (0.26)**	
Split 2* Minority (0)	0.42 (0.22)	
Split 3* Minority (0)	0.35 (0.21)	
Split 4* Minority (0)	-0.15 (0.18)	
Split 5* Minority (0)	0.00 (0.00) (see score test below)	

\*p < 0.05; \*\*p < 0.01; Gender: female = 0; Minority: no = 0; Halfday: no = 0

Score Statistics for Type 3 GEE Analysis

Variable	Chi-Square	p
Gender	$\chi^2_1 = 22.54$	< 0.0001
Minority	$\chi^2_1 = 22.69$	< 0.0001
Risknum	$\chi^2_1 = 43.08$	< 0.0001
Wksesl	$\chi^2_1 = 158.61$	< 0.0001
Halfday	$\chi^2_1 = 35.83$	< 0.0001
Readbk2	$\chi^2_1 = 14.38$	0.0001
R2_kage	$\chi^2_1 = 0.00$	0.9734
Split	$\chi^2_4 = 2132.14$	< 0.0001
Split *Minority	$\chi^2_4 = 60.86$	< 0.0001

cutpoints while holding the other variable effects constant. Overall, only one intercept parameter was estimated (for the log-odds of a value being at or beyond proficiency level 5), but different estimates for each split were obtained which were used to modify the intercept value; and the split by minority interaction terms were used to determine how much the effect of minority changed across splits. In this analysis, the coding of categorical variables was internal to GENMOD; effects are shown in Table 4e for the lower value of the coded variables (for example, the effect for gender,  $I = 0.34$ , was the change in slope for females; this was the opposite of the slope in the PO model, where  $I = -0.31$  for the change in slope for males).

According to these results, the split by minority interaction terms were statistically significant for the first logit comparison (corresponding to  $P(Y \geq 1)$ ) and the last (corresponding to  $P(Y \geq 5)$ ), but not for the other splits. These results suggested that reliable differences existed in the effect of minority across the outcome levels of the ordinal model.

Overall, when coding was taken into consideration for the effect of minority, the predicted logits were similar to those obtained from the separate binary regressions shown in Table 4a for the full-sample. Additionally, the slope effects for the remaining variables, which were held proportional in the analysis, reflected the values obtained through the PO analysis in Table 3a (after accounting for coding reversals). Thus, using the separate binary models to investigate the presence and impact of extreme or unusual scores made sense for both the PO and PPO model, as these binary models reflect what was expected in the data for each of the separate splits.

#### Residuals in Ordinal Logistic Regression

SAS, SPSS and Stata do not provide residual diagnostics for ordinal models. Hosmer and Lemeshow (2000) suggested considering each of the underlying models separately and applying residual methods for these binary logistic models in order to identify unusual or extreme observations. This approach mirrors the aptness of model investigations for the proportional odds assumption presented previously.

Consider the residuals for these underlying cumulative binary models and in addition the OLS strategies that were used for preliminary analyses to examine whether that approach could assist in identifying unusual cases. Under the OLS framework, there are several commonly used measures to identify unusual cases. Mahalanobis' distance is the distance for each case to the centroid of remaining cases (multivariate outliers). Leverages are the diagonal elements of the hat matrix in OLS; they are a transformation of Mahalanobis distance. Leverages flag cases are considered extreme in the  $X$  or explanatory variable space, where two or three times the average leverage can be considered large.

However, cases with large leverage values may or may not be influential; that is, an observation may be unusual in terms of being outside an acceptable range relative to the other  $X$  values, but it may not affect the shape or direction of the regression function. Cook's  $D$  assesses how influential each case is to the fit of the overall model. This measure considers what happens to the model when each case is removed, one at a time, from the overall model. A large Cook's  $D$  value is determined in relation to values obtained from all the other cases. Generally Cook's  $D$  statistics are plotted to identify any large jumps in the measures. Finally, when assessing outliers on  $Y$ , the outcome variable, a common statistic is the studentized deleted residual, or SDRESID. For SDRESID the change in residuals was examined when each case is removed, one at a time, from the model.

Adjustments to the OLS statistics are required for logistic regression. Logistic models predict the probability that  $Y = 1$  for a dichotomous dependent variable. The residuals obtained through a logistic regression are heteroscedastic (variance =  $\pi_i(1 - \pi_i)$ ). Techniques similar to those used in OLS models have been developed for logistic regression in order to detect unusual or influential observations (Pregibon, 1981). The Pearson residual, deviance residual and Pregibon leverages are three main types of residual statistics commonly used for logistic regression diagnostics; however, many choices are available. Table 5 presents the types of residual

## MODEL DIAGNOSTICS FOR PO & PPO MODELS

diagnostics for logistic regression that are available in the three statistical software packages discussed here: SPSS, SAS and Stata. The corresponding mathematical form for each type of residual is also included in the table.

The Pearson residual is the standardized difference between the observed and fitted values. Pearson residuals,  $r_i$  are components of the summary  $\chi^2$ . SAS labels these residual components as *reschi*; SPSS labels it as *zresid* (normalized residuals); Stata provides residuals (Pearson residuals) and *rstandard* (standardized Pearson residuals). Pearson residuals can be computed using:

$$r_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}} \quad (3)$$

Where  $y_i$  is the observed number of success;  $n_i$  is the number of observations with explanatory variable  $x_i$ ;  $\hat{\pi}_i$  is estimated probability at  $x_i$ . When the number of observations is 1, that is,  $n_i = 1$  (assuming a Bernoulli rather than binomial model), the Pearson residuals can be simplified as:

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i (1 - \hat{\pi}_i)}} \quad (4)$$

Deviance residuals capture the difference between the maxima of the observed and fitted log likelihood functions.  $Dev_i$  are components of the summary model deviance,  $D = -2LL$ . SAS labels this as *resdev*; SPSS labels it as *dev*; and Stata labels it as *deviance*. Deviance residuals are defined as:

$$d_i = \pm \sqrt{2 \left[ y_i \log \frac{y_i}{n_i \hat{\pi}_i} + (n_i - y_i) \log \frac{(n_i - y_i)}{n_i (1 - \hat{\pi}_i)} \right]} \quad (5)$$

Also when  $n_i = 1$ , the formula can be simplified as:

$$d_i = \pm \sqrt{2 \left[ y_i \log \frac{y_i}{\hat{\pi}_i} + (1 - y_i) \log \frac{(1 - y_i)}{(1 - \hat{\pi}_i)} \right]} \quad (6)$$

Large values of  $r_i$  or  $d_i$  suggest that the model does not fit that case well. The  $r_i$  and  $dev_i$  are components of alternate tests for same null hypothesis tested based on the Chi-square distribution; that is, does the model fit as well as a saturated model of the data (i.e., one that perfectly reproduces the original data). Under the null hypothesis, the individual components are approximately normally distributed and it is expected that the summary value/df will be less than 1.0. However, neither the summary Pearson residual statistic nor the summary deviance residual statistic follows a Chi-square distribution when continuous explanatory variables are included in the model. Thus, the summary statistics are not appropriately used in that situation. When the data are sparse (i.e., with continuous IVs), this Chi-square distributional assumption is not upheld.

Pregibon leverages (*hat*) are the diagonal elements of the *hat* matrix in logistic regression. They are used to measure the relative influence of an observation on the model fit. SAS labels these as *h*; SPSS labels them as *lever*; and Stata labels them as *hat*. Pregibon hats tend to be small when the estimated probability of observations is outside of the 0.1 to 0.9 interval, because most extreme cases may also have small leverages (Hosmer & Lemeshow, 2000). Therefore, Pregibon hats may not provide a good assessment of influential cases when the estimated probability of an observation is too small or too large. Pregibon leverages are defined as:

$$H = W^{1/2} X (X' W X)^{-1} X' W^{1/2} \quad (7)$$

$$h_{ii} = \hat{\pi}_i (1 - \hat{\pi}_i) x_i' \hat{Var}(\hat{\beta}) x_i$$

where  $W$  is the diagonal weight matrix; and  $h_{ii}$  is the leverage or diagonal element of the *hat* matrix,  $H$ .

It is also often more informative to consider how each case affects fit of the overall model. There are several approaches in logistic regression to measure the change in Chi-square fit, deviance fit, and in the estimated parameters when a single observation is removed from the model. These measures are similar to Cook's  $D$  in ordinary least-squares regression. SPSS

provides a measure which is an analog of Cook's D, and labels it Cook; SPSS only provides change in estimated parameters and labels it as dfbeta. SAS and Stata provide all three options. SAS labels the standardized difference in estimated parameters dfbetas, which is different from SPSS. SAS also provides two measures of the change in the confidence interval for the regression estimates when an

observation is deleted, and labels them C and CBAR. SAS labels the change in Chi-square fit as difchisq and change in deviance fit as dfdev. Stata labels the standardized change in regression coefficient as dbeta, the change in Chi-square fit as dx2, and the change in deviance fit as ddeviance. Descriptions of several of these statistics in SPSS, SAS and Stata are listed in Table 5.

Table 5: Residual Diagnostics of Logistic Regression in SPSS, SAS and Stata

Types Of Residuals	Mathematical Formula	SPSS	SAS	Stata
Pearson Residuals	$r_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$	✓	✓	✓
Studentized Residuals	$\text{studentized } rs_i = \frac{r_i}{\sqrt{1 - h_{ii}}}$	✓		✓
Logit Residuals	$\text{logit } r_i = \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i (1 - \hat{\pi}_i)}$	✓		
Deviance Residual	$d_i = \pm \sqrt{2 \left( y_i \log \frac{y_i}{n_i \hat{\pi}_i} + (n_i - y_i) \log \frac{(n_i - y_i)}{n_i (1 - \hat{\pi}_i)} \right)}$	✓	✓	✓
Leverage (Hat)	$H = W^{1/2} X (X' W X)^{-1} X' W^{1/2}$ $h_{ii} = \hat{\pi}_i (1 - \hat{\pi}_i) x_i' \hat{V}ar(\hat{\beta}) x_i'$	✓	✓	✓
Cook's D	$C_i = \frac{r_i^2 h_{ii}}{(1 - h_{ii})^2}$	✓	✓	✓
Change in Pearson Chi-square Residuals	$\Delta \chi_i^2 = \frac{r_i^2}{1 - h_{ii}}$		✓	✓
Change in Deviance Residuals	$\Delta D_i = d_i^2 + C_i$		✓	
	$\Delta D_i = \frac{d_i^2}{(1 - h_{ii})}$			✓
Change in Betas	$\Delta b_i = \frac{r_i^2 h_{ii}}{1 - h_{ii}}$	✓	✓	✓



## MODEL DIAGNOSTICS FOR PO & PPO MODELS

The Hosmer-Lemeshow test does not yield a residual measure, but it does inform about the fit of a model, particularly when continuous explanatory variables are included. The H-L test attempts to compensate for the presence of continuous variables and resulting sparseness of the data by aggregating across deciles of risk, which are formed by collapsing over observations with similar covariate patterns. However, this same strategy of aggregating by covariate pattern for residual diagnostic purposes may mask influential or poorly fit cases; therefore, researchers tend to rely on visual assessment rather than specific concrete measures or approaches, for identification of unusual cases. In addition, given the lack of informative distributional theory regarding many of the residual statistics discussed above, it may be more valuable to apply graphical strategies for observing and identifying unusual or influential cases within logistic and ordinal models.

Many of these graphical approaches mirror what is available through OLS. For ordinary least squares regression, a graph of the observed dependent variable,  $y$ , versus the predicted dependent variable,  $\hat{y}$ , can be plotted. Observed and predicted outcomes in logistic regression are dichotomous; thus modifications are required – usually the resulting graphs are done at a casewise level. For logistic regression, index plots are enormously useful. Index plots display each residual diagnostic for a case against the case number. The resulting output can be unwieldy for samples with many observations and when multiple residual statistics are investigated; but visually, extreme or unusual cases can be readily detected.

As a minimum, Hosmer and Lemeshow (2000) recommended plotting the change in Chi-square fit versus the predicted probability ( $\hat{p}$ ) of the dependent variable; change in deviance fit versus  $\hat{p}$ ; and change in regression estimates versus  $\hat{p}$ . They pointed out that using the summary change statistic rather than the individual component values ( $r_i$  or  $d_i$ , above) for each case visually emphasized the poorly fit cases. Because not all statistics are available in each statistical package, choices among possible

graphs or plots have to be made depending on the options available.

### Residual Diagnostics Results

With large data sets, the amount of output involved in graphical displays for diagnostic statistics can become unworkable very quickly. Thus, two smaller data sets are employed to demonstrate the use and interpretation of regression diagnostic statistics. OLS regression of the ordinal outcome was used on the collection of explanatory variables as an initial strategy to identification of unusual or poorly fit cases for both Samples I and II. A series of logistic regression models were then run for Samples I and II corresponding to the cumulative splits; residual statistics obtained from these five logistic models were used to identify unusual cases. Next, the use of index plots to visually display the residual statistics for a collection of diagnostic values derived from the logistic splits for Sample I (one sample was selected for demonstration of the index plots) was demonstrated. After reviewing the residual strategies and identifying the poorly fit cases, those cases identified as unusual by inspecting the original data were explored along with characteristics of the children in order to better understand who is potentially being poorly fit by the model.

### Ordinal Residual Diagnostics for Sample I

Table 6a presents casewise statistics based on unusual cases identified through an OLS scatterplot of observed versus predicted values, OLS casewise diagnostics, and additional cases identified through the five sequential (cumulative) logistic models and corresponding index plots for Sample I ( $n = 244$ ). Figure 1 displays the OLS scatterplot of observed versus predicted values: several potentially unusual cases or outliers were observed in proficiency level 0, 3 and 5. From the scatterplot, nine cases were identified as unusual or poorly fit. The absolute values of the studentized deleted residuals (SDRESID) of these cases were mostly larger than or close to 2. For case number 1698 (ID = 0731010C) its Mahalanobis Distance was 39.23, which is very large, and the corresponding leverage value and Cook's D for this case are also the largest among

those cases identified as outliers. This child was predicted to be in proficiency level 0.28 (via OLS), yet had actually scored in proficiency level 3. Visually, this case can be clearly identified as an outlier in Figure 1 for proficiency level 3.

Table 6a shows statistics for two additional cases identified through the OLS

residual diagnostics options as having standardized deleted residuals (ZRESID) larger than our setting of 2.5 (2.5 was used instead of the default value of 3.0 to maintain consistency with the values used in the logistic regression procedure). Both children were in proficiency level 5 but were predicted to be in levels 2.16 and 2.15 respectively.

Table 6a: Casewise Diagnostics for OLS Model Sample I

From Residual Diagnostics

Case No.	Child ID	ProfrdK2	Predicted	Residual	SDRESID	Lever	Mahal	Cook's D
2103	0936002C	5	2.16	2.84	2.60	.02	4.86	.02
3414	3113018C	5	2.15	2.85	2.62	.03	7.71	.03

From Scatterplot

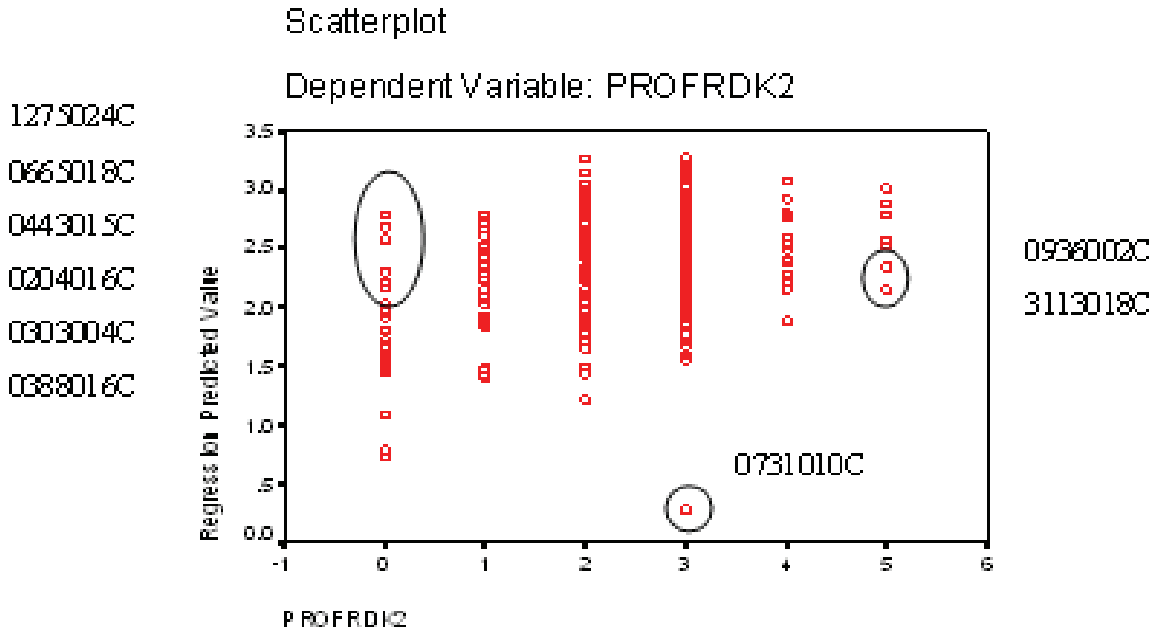
Case No.	Child ID	ProfrdK2	Predicted	Residual	SDRESID	Lever	Mahal	Cook's D
351	0204016C	0	2.30	-2.30	-2.12	.04	10.87	.03
665	0303004C	0	2.23	-2.23	-2.02	.02	4.81	.01
946	0388016C	0	2.21	-2.21	-1.99	.02	3.83	.01
1059	0443015C	0	2.58	-2.58	-2.35	.02	4.36	.02
1571	0665018C	0	2.68	-2.68	-2.44	.02	3.77	.01
1698	0731010C	3	0.28	2.72	2.69	.16	39.23	.17
2343	1049011C	0	2.17	-2.17	-1.97	.02	4.82	.01
2889	1271005C	0	2.04	-2.04	-1.85	.03	6.48	.01
2952	1275024C	0	2.78	-2.78	-2.53	.02	4.20	.02

Additional from Logistic Diagnostics and Plots

Case No.	Child ID	ProfrdK2	Predicted	Residual	SDRESID	Lever	Mahal	Cook's D
1917	0833003C	2	3.06	-1.06	-0.96	.02	4.15	.00
3392	3105020C	4	1.89	2.11	1.96	.06	15.23	.03
3114	3002012C	2	2.90	-0.90	-0.81	.02	5.39	.00
1384	0609022C	3	1.70	1.30	1.21	.08	18.84	.02
3078	2115003C	5	2.88	2.12	1.92	.02	3.88	.01

## MODEL DIAGNOSTICS FOR PO & PPO MODELS

Figure 1: OLS Observed Versus Predicted Values for Sample I, n = 244 Public School Children



For comparison, the summary statistics of Table 6a include OLS for cases that were identified through the binary logistic models as being potential outliers: five additional cases were detected. The OLS statistics for these cases, however, do not indicate that these cases are unusual, with the exception of one large Mahalanobis' distance value of 18.84 for Case No. 1384 (ID = 0609022C).

Table 6b presents casewise diagnostics for the five cumulative binary logistic regression models (splits) for Sample I. Four cases (Case No. 1059, 1571, 1698 and 2952) were flagged as outliers for the first cumulative split based on SRESID greater than 2.5 (set greater than the 2.0 default setting); one (Case No. 1698) was flagged as an outlier for the second cumulative split; no cases were identified in the analysis for the third cumulative split; two (Case No. 1917, and 3414) were flagged as outliers for the fourth cumulative split; and three (Case No. 2103, 3392, and 3414) were flagged as outliers for the fifth cumulative split.

### Ordinal Residual Diagnostics for Sample II

Table 7a presents casewise statistics based on unusual cases identified through an

OLS scatterplot of observed versus predicted values, OLS casewise diagnostics, and additional cases identified through the five sequential (cumulative) logistic models and corresponding index plots for Sample II (n = 278). Figure 2 displays the OLS scatterplot of observed versus predicted values; eleven potentially unusual cases or outliers can be observed in proficiency level 0, 4 and 5. Referring to Table 7a, only one of these eleven cases had a fairly large OLS Mahalanobis' distance of 11.85; this child scored in proficiency level 4 but was predicted into level 1.95. The statistics for the other 10 cases identified did not seem unusual. The absolute values of the studentized deleted residuals (SDRESID) of these cases were slightly larger than or close to 2.

After running the OLS regression to request residual diagnostics, no additional unusual observations were identified. Only one case, Case 5, had already been identified through the review of the scatterplot; this case had a ZRESID value greater than 2.5.

Table 7b presents casewise diagnostics for the five cumulative binary logistic regression models (splits) for Sample II. One case (Case No. 19) was flagged as an outlier for the first

Table 6b: Casewise Diagnostics for Cumulative Splits, Sample I, n = 244

CUMSP1: P(Y ge 1)					CUMSP2: P(Y ge 2)	
Case	1059	1571	1698	2952	Case	1698
CUMSP4	0**	0**	1**	0**	CUMSP5	1**
Resid	-0.959	-0.990	0.993	-0.978	Resid	0.973
SResid	-2.546	-3.044	3.197	-2.770	SResid	2.746
Dev	-2.525	-3.037	3.155	-2.756	Dev	2.689
ZResid	-4.823	-9.980	11.997	-6.607	ZResid	6.014
DFB0	-0.401	-0.009	2.442	-0.706	DFB0	0.932
DFB1	0.101	-0.137	-0.165	0.098	DFB1	-0.057
DFB2	0.119	0.168	-0.270	0.185	DFB2	-0.113
DFB3	-0.064	0.017	-0.055	0.025	DFB3	-0.017
DFB4	0.002	0.258	0.026	0.241	DFB4	0.002
DFB5	-0.008	-0.061	-0.060	-0.037	DFB5	-0.040
DFB6	-0.204	-0.104	-0.970	-0.021	DFB6	-0.369
DFB7	0.003	-0.004	-0.034	0.004	DFB7	-0.012

CUMSP4: P(Y ge 4)			CUMSP5: P(Y ge 5)			
Case	1917	3414	Case	2103	3392	3414
CUMSP1	1**	1**	CUMSP2	1**	1**	1**
Resid	0.982	0.954	Resid	0.961	0.966	0.985
SResid	2.883	2.521	SResid	2.595	2.628	2.943
Dev	2.844	2.480	Dev	2.552	2.596	2.907
ZResid	7.493	4.547	ZResid	4.996	5.296	8.204
DFB0	0.671	0.594	DFB0	-0.018	3.052	2.649
DFB1	-0.073	0.091	DFB1	-0.215	-0.207	0.331
DFB2	-0.280	-0.081	DFB2	0.396	-0.353	-0.244
DFB3	0.456	-0.066	DFB3	0.305	-0.043	-0.173
DFB4	-0.145	0.115	DFB4	0.369	-0.388	0.313
DFB5	0.136	-0.540	DFB5	0.146	0.183	-1.050
DFB6	-0.043	-0.065	DFB6	0.101	-0.114	-0.229
DFB7	-0.009	-0.001	DFB7	-0.004	-0.034	-0.022

a S = Selected, U = Unselected cases, \*\* = Misclassified cases; b Cases with studentized residuals greater than 2.500 are listed; For CUMSP3: P(Y ge 3), the casewise plot is not produced because no outliers were found

## MODEL DIAGNOSTICS FOR PO & PPO MODELS

Figure 2: OLS Observed Versus Predicted Values for Sample II, n = 278 Public School Children

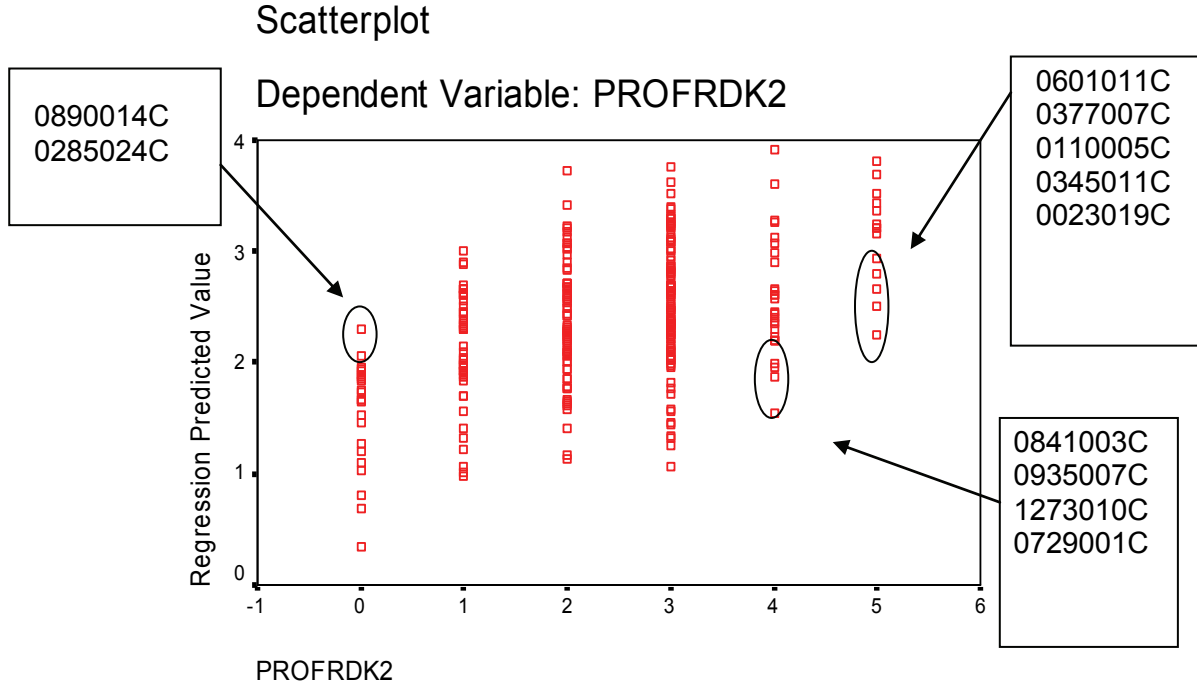


Table 7a: Casewise Diagnostics for OLS Model Sample II  
From Scatterplot

Case No.	Child ID	ProfrdK2	Predicted	Residual	SDRESID	Lever	Mahal	Cook's D
5	0023019C	5	2.25	2.75	2.56	.02	6.18	.02
23	0110005C	5	2.66	2.34	2.17	.02	6.82	.02
44	0285024C	0	2.06	-2.06	-1.91	.03	7.55	.01
53	0345011C	5	2.50	2.50	2.32	.02	5.45	.02
58	0377007C	5	2.80	2.20	2.03	.01	3.91	.01
93	0601011C	5	2.94	2.06	1.90	.01	4.10	.01
118	0729001C	4	1.55	2.45	2.27	.02	6.90	.02
142	0841003C	4	1.98	2.01	1.87	.03	8.08	.01
153	0890014C	0	2.29	-2.29	-2.11	.01	3.93	.01
155	0935007C	4	1.95	2.05	1.91	.04	11.85	.02
226	1273010C	4	1.87	2.13	1.97	.02	6.08	.01

From Residual Diagnostics: None

Additional from Logistic Diagnostics and Plots

Case No.	Child ID	ProfrdK2	Predicted	Residual	SDRESID	Lever	Mahal	Cook's D
19	0057012C	0	1.94	-1.94	-1.79	.03	7.66	.01

Table 7b: Casewise Diagnostics for Cumulative Splits, Sample II Public School Children, n = 278

CUMSP1: P(Y ge 1)		CUMSP4: P(Y ge 4)		CUMSP5: P(Y ge 5)	
Case	19	Case	118	Case	5
CUMSP1	0**	CUMSP4	1**	CUMSP5	1**
Resid	-0.977	Resid	0.957	Resid	0.983
SResid	-2.776	SResid	2.532	SResid	2.880
Dev	-2.750	Dev	2.507	Dev	2.862
ZResid	-6.547	ZResid	4.706	ZResid	7.690
DFB0	-0.984	DFB0	0.119	DFB0	1.714
DFB1	-0.129	DFB1	0.052	DFB1	-0.112
DFB2	0.370	DFB2	0.046	DFB2	-0.174
DFB3	-0.052	DFB3	0.113	DFB3	-0.071
DFB4	-0.127	DFB4	-0.055	DFB4	0.245
DFB5	0.275	DFB5	-0.249	DFB5	0.188
DFB6	-0.247	DFB6	-0.031	DFB6	-0.238
DFB7	0.006	DFB7	0.000	DFB7	-0.021

Notes: a S = Selected, U = Unselected cases, \*\* = Misclassified cases; b Cases with studentized residuals greater than 2.500 are listed; For CUMSP2: P(Y ge 2) and CUMSP3: P(Y ge 3) the casewise plots are not produced because no outliers were found

cumulative split; this case was overpredicted – that is, the observation was in level 0 but predicted to be at or beyond level 1. Case No. 118 was flagged as an outlier for the fourth cumulative split, and was underpredicted (i.e., child scored in a higher category but was predicted into a category below 4). The third outlier identified was Case No. 5 at the fifth cumulative split; this child was also underpredicted. For comparison purposes, the OLS statistics for the one new case, Case No. 19, is included in the bottom section of Table 7a. No additional unusual values were determined from the index plots for Sample II (demonstrated below for Sample I).

Index Plots

Using the separate binary logistic models to mimic the data patterns of an ordinal model yields five different regressions for a K = 6 level ordinal outcome variable. Plotting each

model’s residuals or diagnostic summary values against the case number can enhance the search for extreme or unusual values; however, there will necessarily be multiple plots for each binary split.

Figure 3 contains a display of index plots for 8 residual measures: normalized residual, residual, deviance residual, logit residual, Cook’s analog, leverages, difference in deviance statistic and difference in the Pearson Chi-square statistic. These displays contain the value of the residual statistic on the vertical axis, and the case number horizontally. Strong peaks indicate extreme value in the residual score. In the first plot, it is easy to detect unusual scores, as noted in the Figure. The three marked cases were previously identified in the residual analyses for Sample I.

Figure 4 contains two of the plots recommended by Hosmer and Lemeshow (2000), namely, the change in Chi-square and

## MODEL DIAGNOSTICS FOR PO & PPO MODELS

deviance fit statistics against  $\hat{p}$ . These graphs are a compilation of two curves, one representing  $Y = 1$  (the downward curve, so that outliers are in the top left corner), and one representing  $Y = 0$  (the upward curve, so outliers are in the top right corner). As illustrated in the first graph in Figure 4, previously identified cases are again indicated through

these index plots. The plots of change in the regression coefficients versus  $\hat{p}$  are not shown (with seven predictors and the intercept, there are eight graphs for each of the five logistic regression splits). Figures 5 through 12 provide the same plots as above for the remaining four logistic regression splits.

Figure 3: Index Plots for Diagnostic Statistics, Split 1 for Sample I

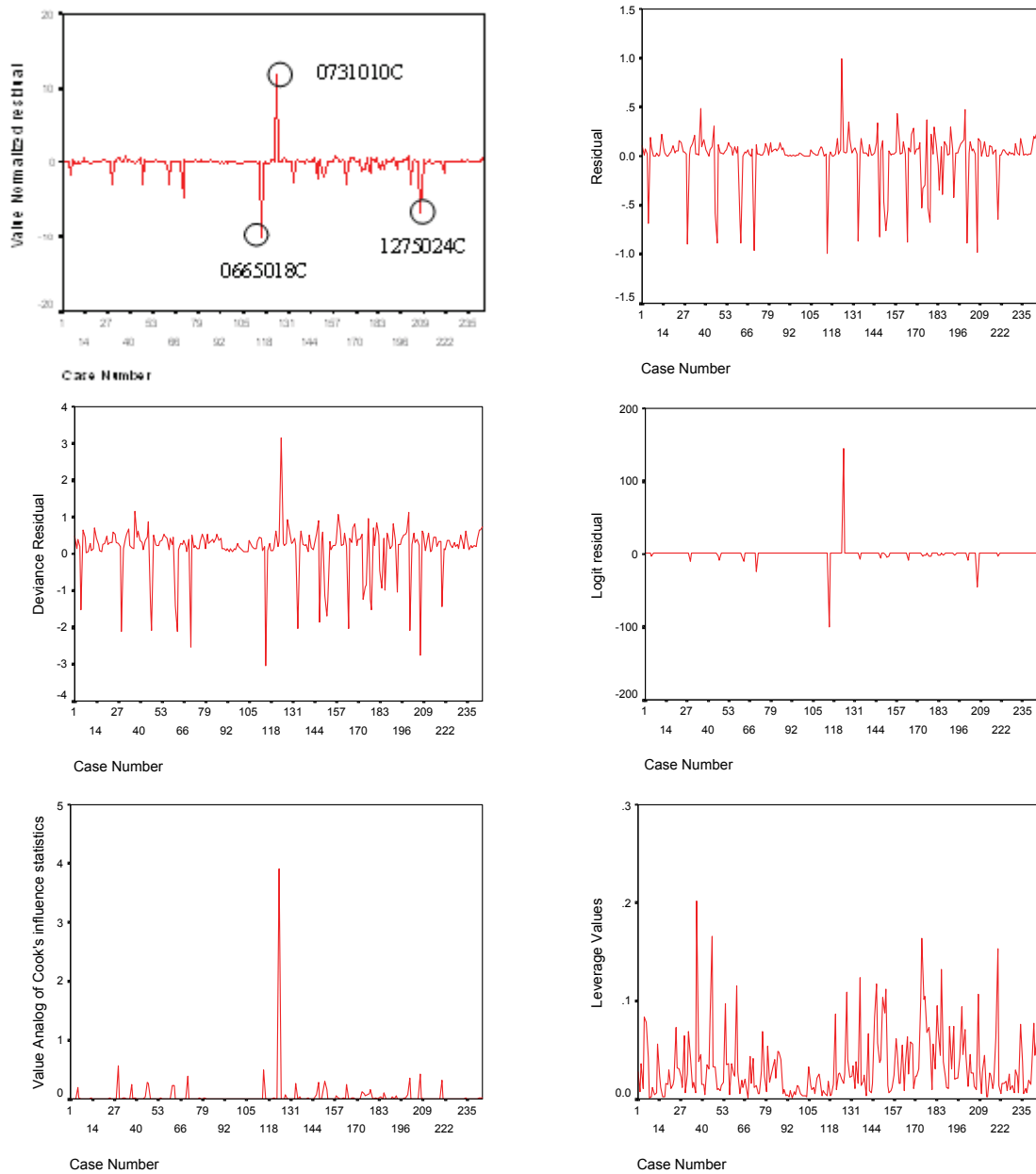


Figure 3 (continued): Index Plots for Diagnostic Statistics, Split 1 for Sample I

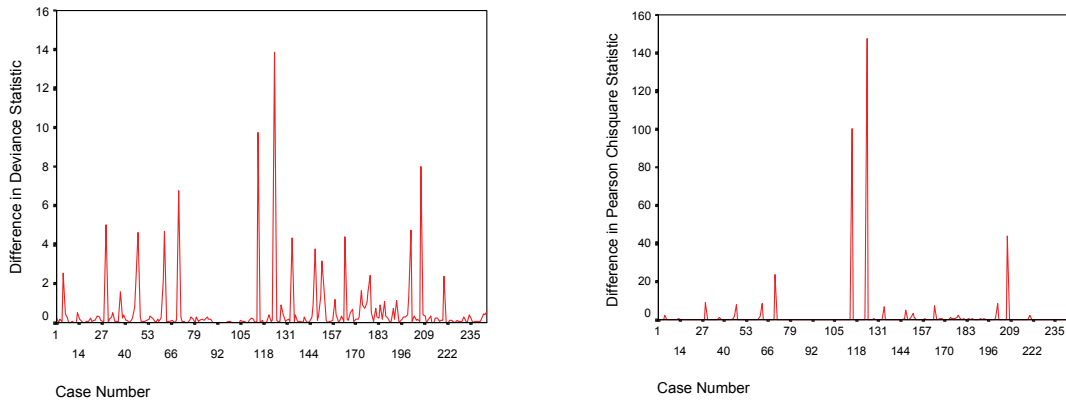
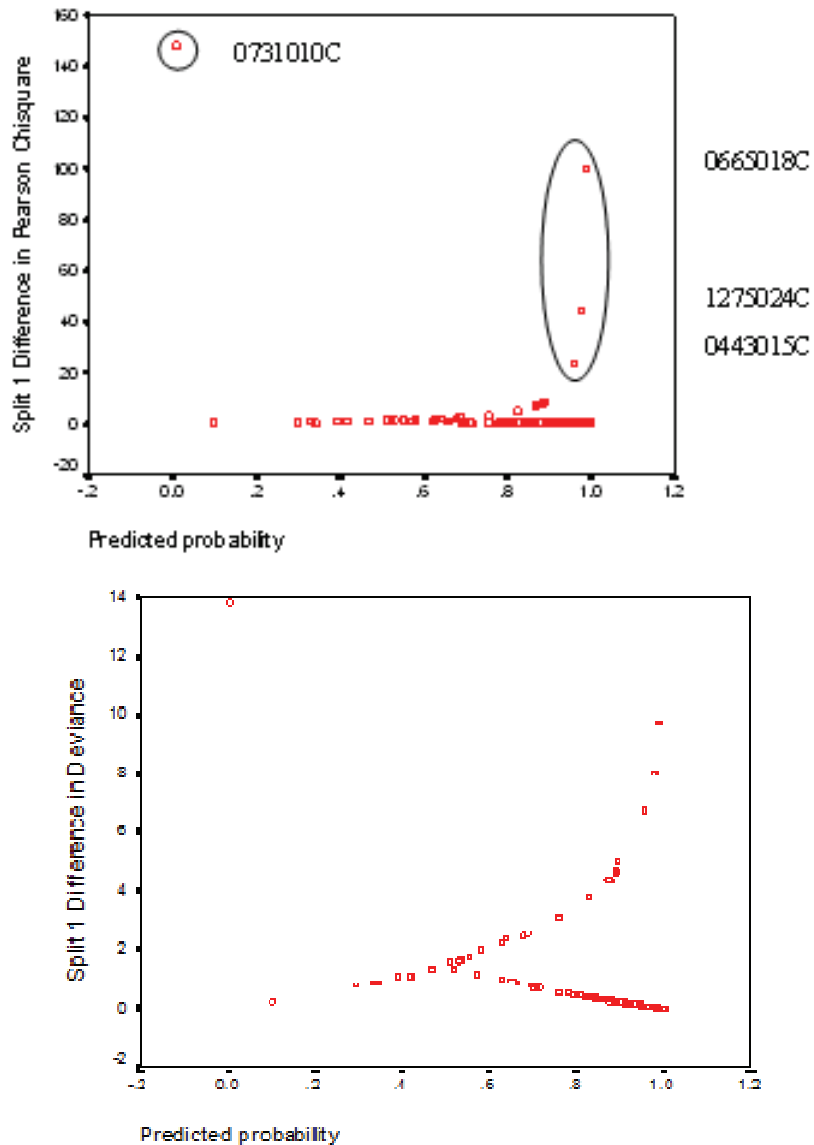


Figure 4: Change in Chi-Square and Deviance Fit against P-hat (Split 1), Sample I





# MODEL DIAGNOSTICS FOR PO & PPO MODELS

Figure 5: Index Plots for Diagnostic Statistics, Split 2 for Sample I

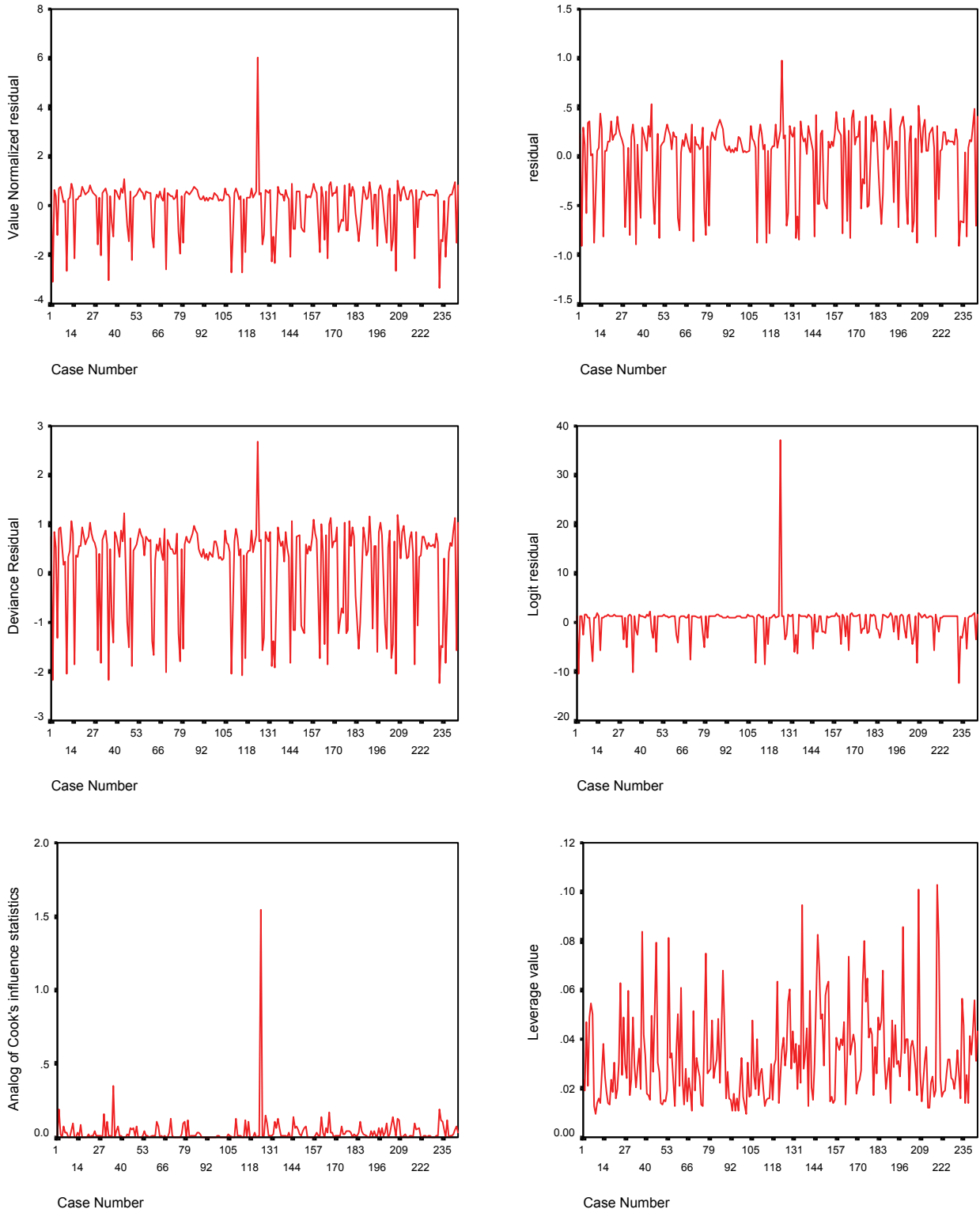
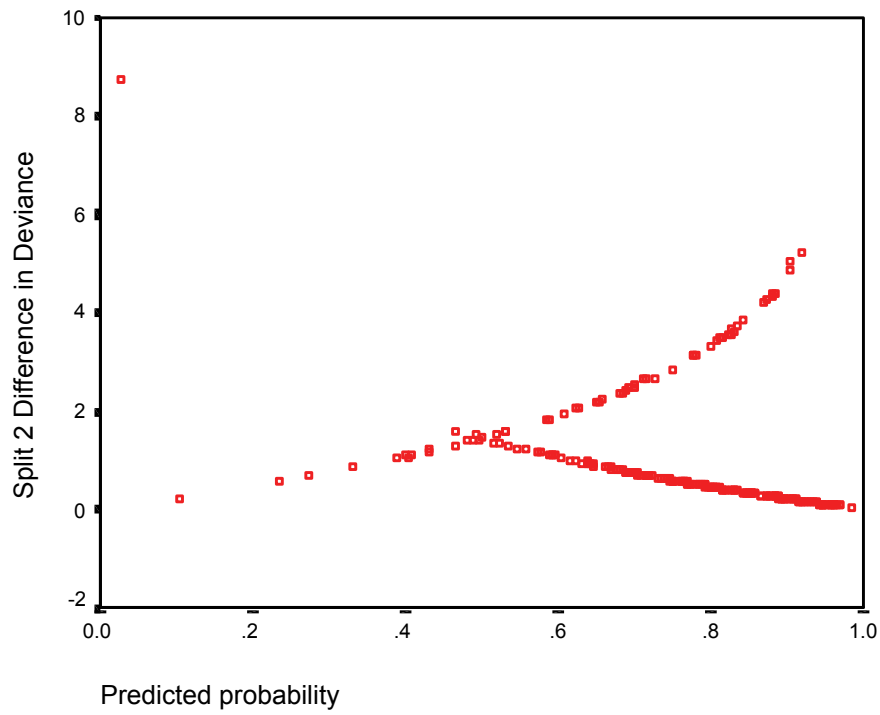
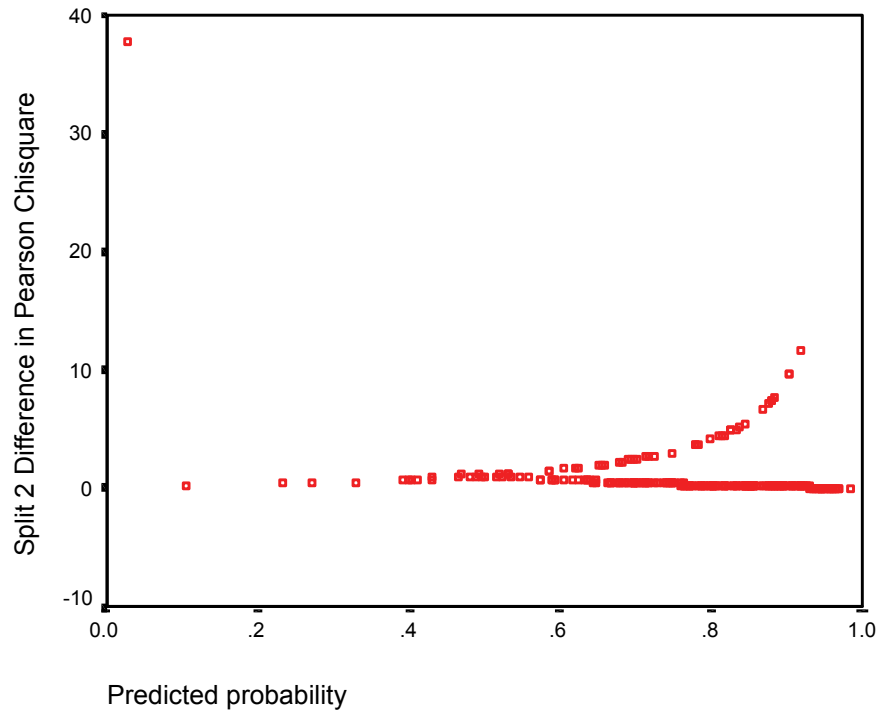


Figure 6: Change in Chi-Square and Deviance Fit against P-hat, Split 2, Sample I



# MODEL DIAGNOSTICS FOR PO & PPO MODELS

Figure 7: Index Plots for Diagnostic Statistics, Split 3 for Sample I

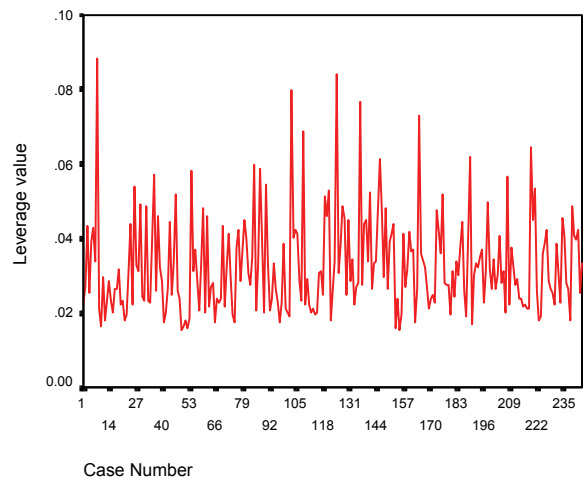
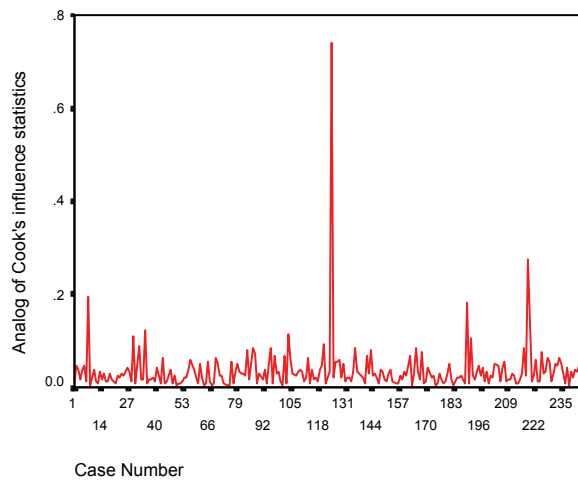
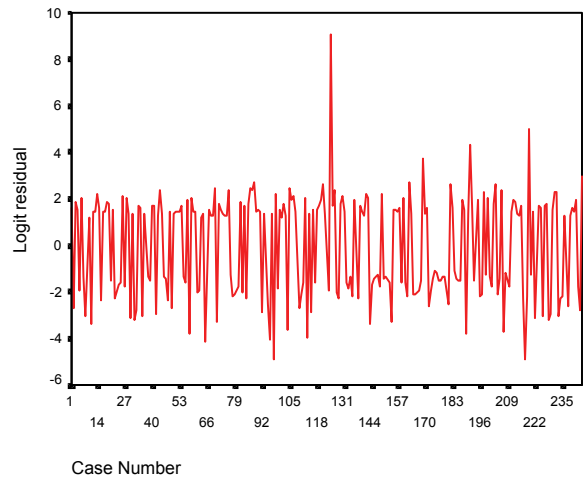
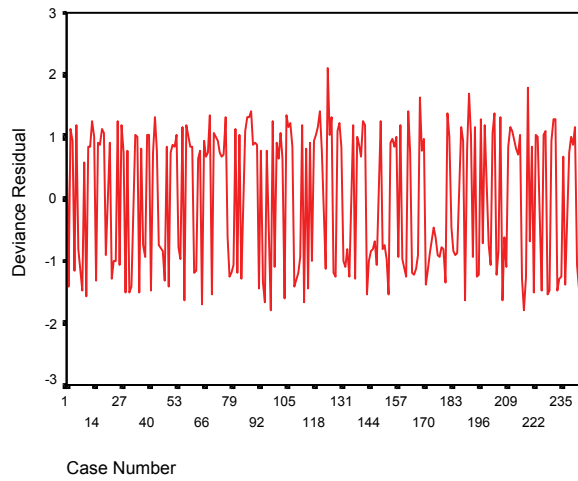
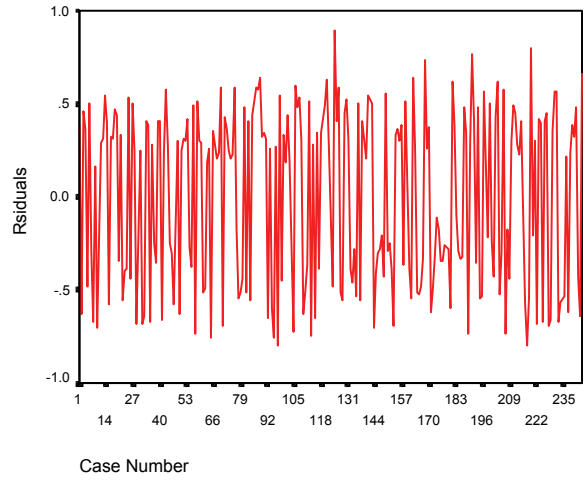
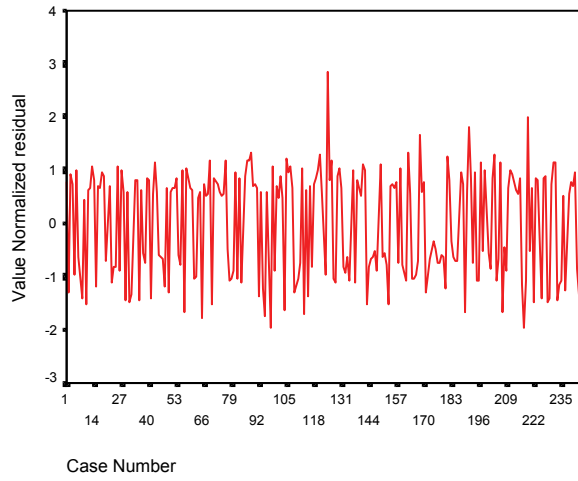
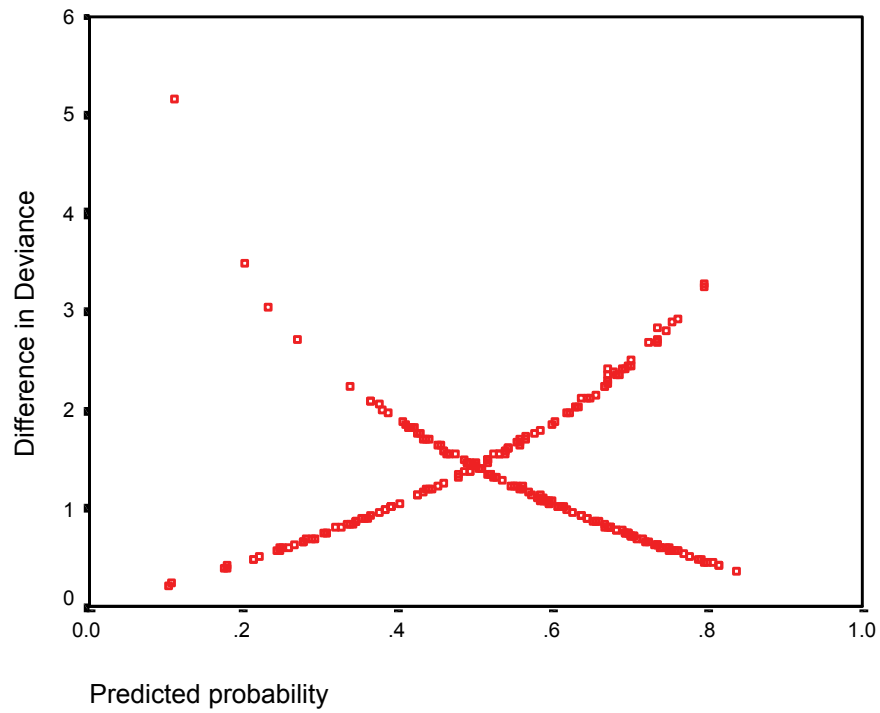
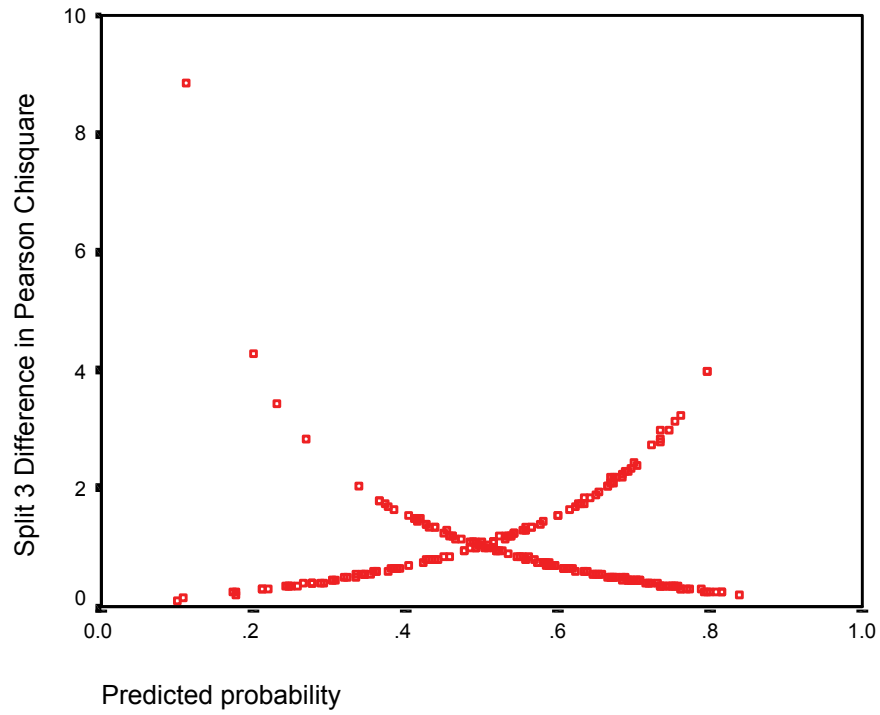


Figure 8: Change in Pearson and Deviance Fit against P-hat, Sample I, Split 3



# MODEL DIAGNOSTICS FOR PO & PPO MODELS

Figure 9: Index Plots for Diagnostic Statistics, Sample I Split 4

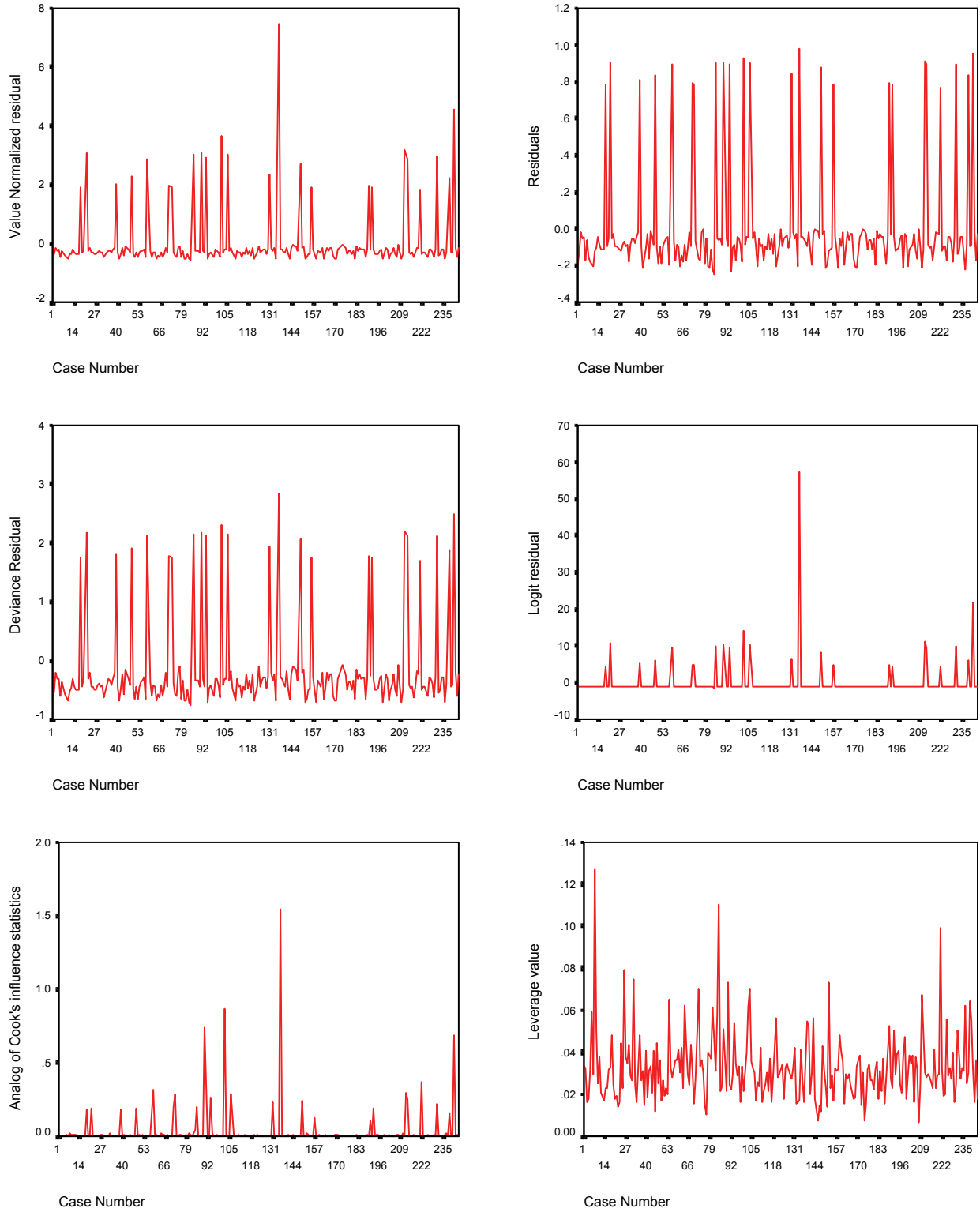
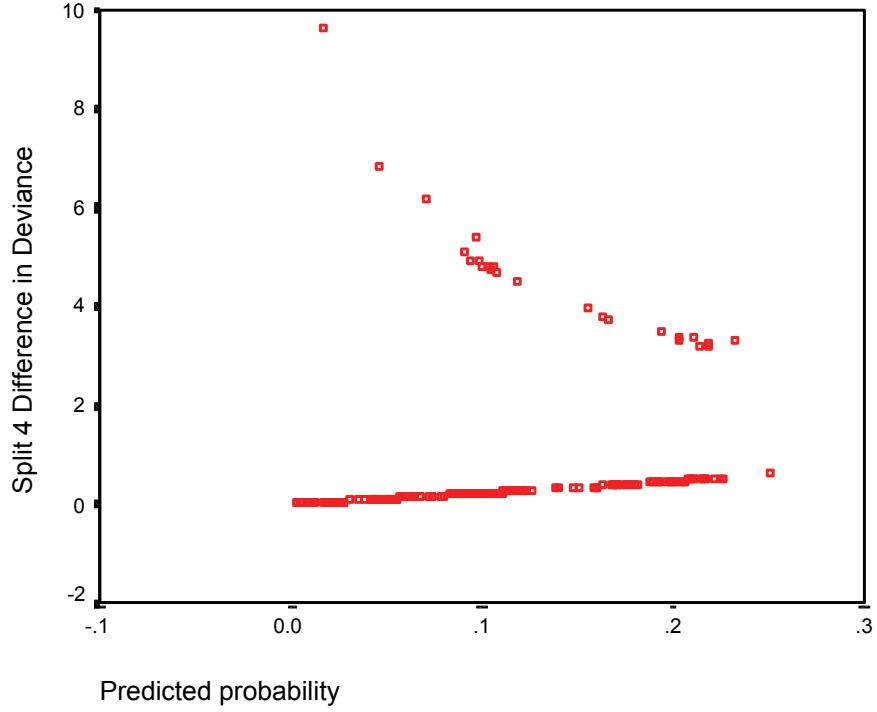
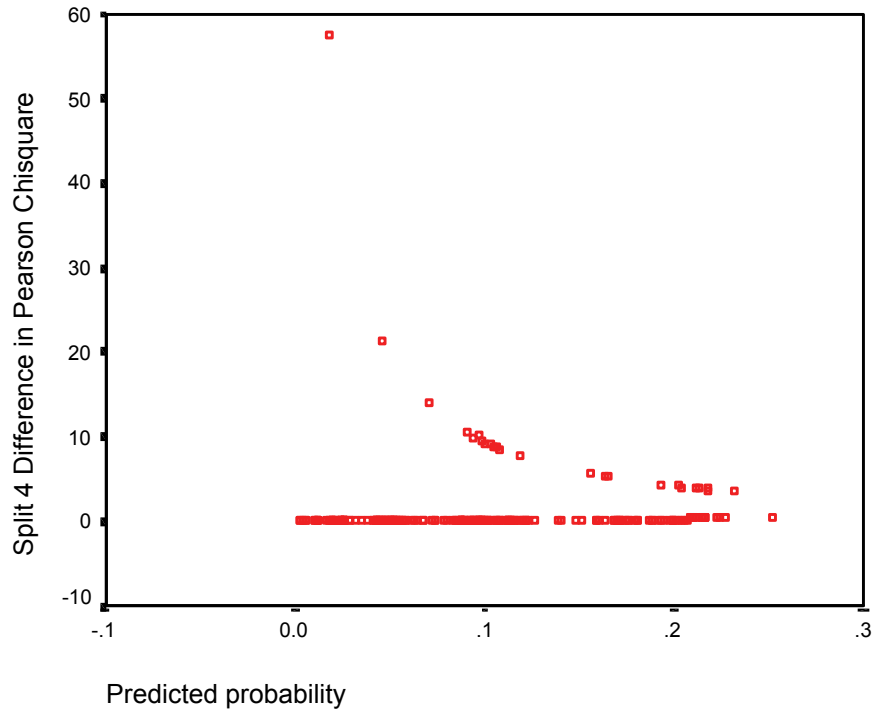


Figure 10: Change in Chi-Square and Deviance Fit, Sample I Split 4



# MODEL DIAGNOSTICS FOR PO & PPO MODELS

Figure 11: Index Plots for Diagnostic Statistics, Sample I Split 5

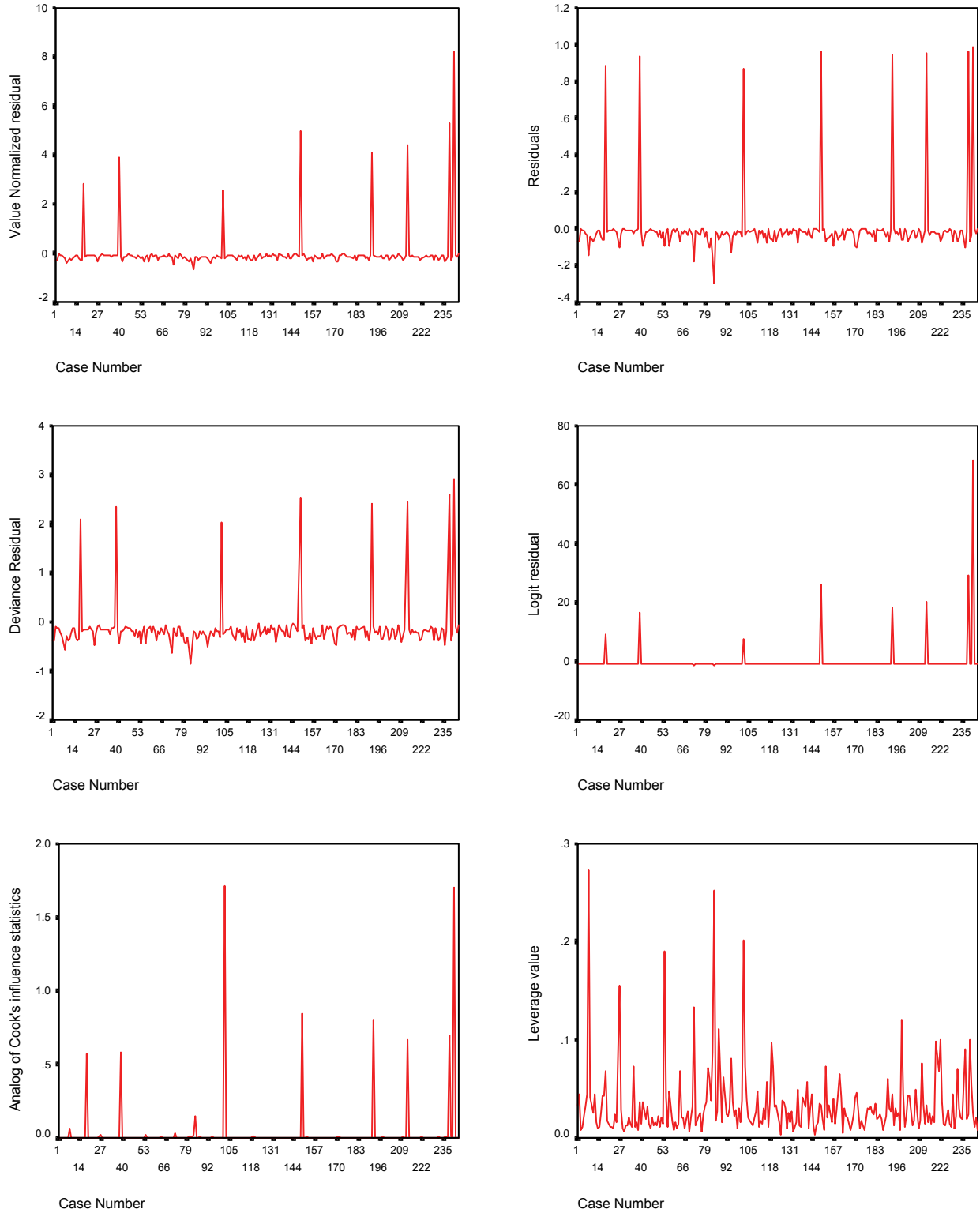
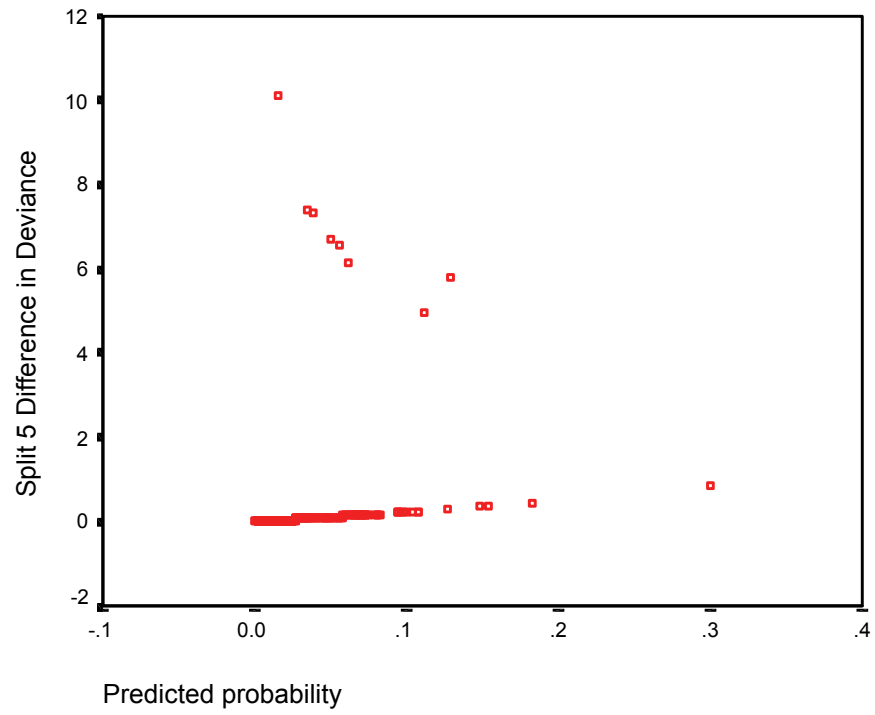
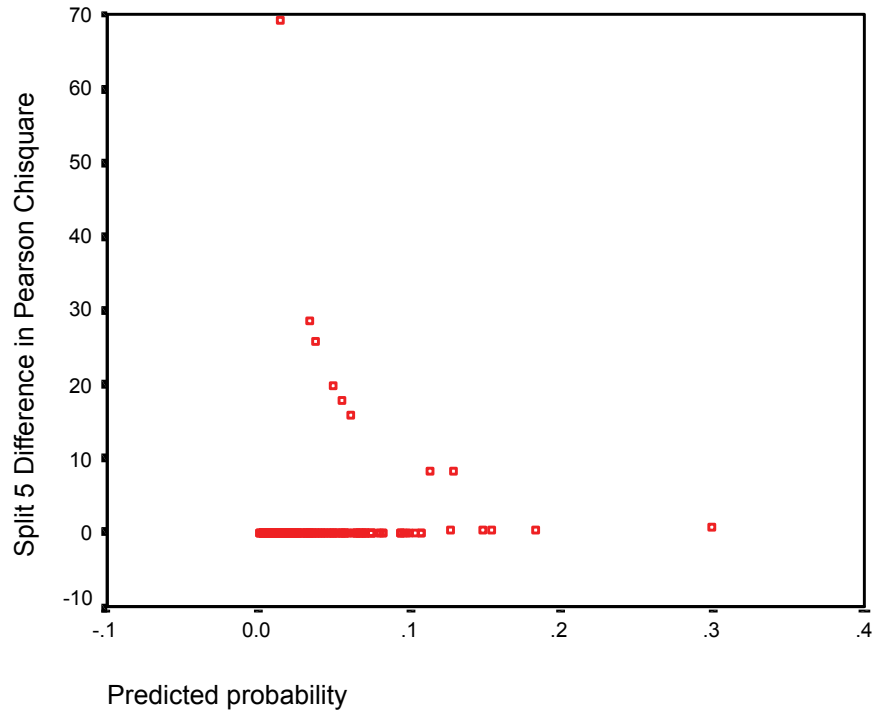


Figure 12: Change in Chi-Square and Deviance Fit, Sample I Split 5





## MODEL DIAGNOSTICS FOR PO & PPO MODELS

Reviewing plots for residual analyses can be informative, but to match extreme or unusual observations in a plot to the original data can be time-consuming. In identifying the outliers in plots, SPSS does not label the cases directly; the analyst must use a two-step procedure (isolating the extreme value and matching that back to the original case). However, Stata has an option of adding labels to the points using the `mlabel` command. Figure 13 displays the plots of change in Pearson Chi-square against  $\hat{p}$  for five cumulative splits for Sample I using Stata. This option can facilitate easy identification of unusual cases from the plots.

### Characteristics of Children Misfit by the Model

Investigation of the characteristics of the children who were identified as unusual, in terms of not being well represented by the model through at least one of the strategies employed, can ultimately help with understanding who the model is not providing a good fit for and why, and thus lead to development of better models that demonstrate stronger knowledge of the outcome for all persons. Tables 8 and 9 contain the values of the explanatory variables of interest for the collection of cases identified as unusual through the diagnostics applied to Sample I and Sample II, respectively. To summarize, for both samples, most identified children were female with SES below the standardized average of 0.

In general, the model tended to under-predict proficiency for some children who have theoretically assumed strikes against them, such as low SES or a large number of family risk characteristics. These cases tended to perform as well as or - in many cases - better than their peers. For example, child ID 0936002C (Case No. 2103) in sub-sample I is a female minority student who does not speak English at home and who attended half-day kindergarten rather than full-day kindergarten: this child's parents read to her at least three times per week and her actual reading proficiency level is 5. The logistic models predicted her to be in a lower category, as did the OLS model (level 2). Thus, some high achieving children do not have their reading proficiency adequately captured by the current model or current set of predictors.

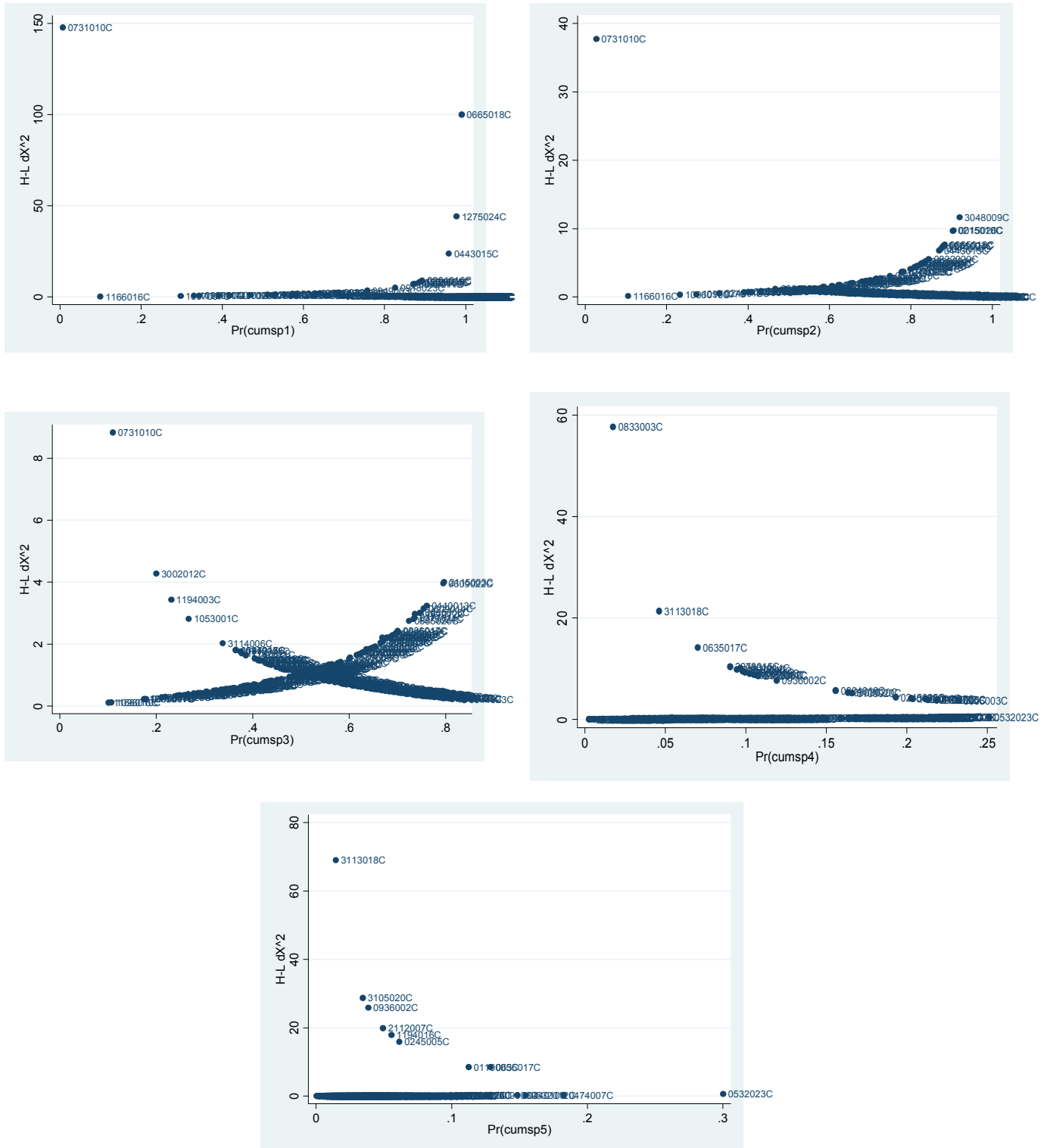
Conversely, some children who have perceived theoretical benefits in their favor, such as higher SES or no family risk characteristics, performed less well than the model predicts. For example, child ID 1275024C (Case No. 2952) in sub-sample I is a female non-minority student without any family risk factor, who attends full-day kindergarten, had parents read to her at least three times per week, and speaks English at home: this student's actual reading proficiency level is 0, but she is predicted to be in level 3.

### Conclusion

Although poor predictions are inevitable in any modeling situation, the concern is that the typically limited range of the dependent variable for ordinal (or logistic) regression models may lead to more systematic under- or over-predictions relative to what might be expected with a continuous outcome. Identification of unusual cases or cases that are poorly fit by a particular model is only the first step in a residual analysis. After the cases are identified, the process turns towards understanding what characteristics of the collection of identified cases are associated with their corresponding under- or over-prediction from the model. This study has only examined the collection of data for the variables included in the models; it could be that a variable external to the model would better explain why some children are so strongly under- or over-predicted by the model.

For the proportional odds and partial-proportional odds model, the analysis of residuals was split into two components, given that residual analyses for ordinal regression analyses are not directly available. These were the OLS analysis of the ordinal outcome (treated as interval/continuous) and the analysis of residuals from the separate binary regression models forming the progression of the cumulative logit model. The latter approach has been advocated by many researchers (Hosmer & Lemeshow; 2000; Harrell, 2001; Long & Freese, 2006), but it remains unclear how well these cases may be fit by a specific ordinal model. That particular model was only approximated in the approach taken herein. Other work has investigated quality of goodness-of-fit tests for ordinal and multinomial data (Fagerland, Hosmer & Bofin, 2008; Pulkstenis & Robinson,

Figure13: Change in Pearson versus P-hat using Stata, Sample I, 5 Cumulative Splits



2004). However, this study focused specifically on data diagnostics for ordinal outcomes, which constitutes a preliminary but necessary step in the determination of model fit.

This article reviewed proportional odds model and partial proportional odds model and the use of some valuable strategies were demonstrated for identifying influential or unusual cases after fitting ordinal regression models. In particular, OLS strategies for preliminary residual detection were applied, followed by application of logistic regression diagnostics approaches for each cumulative binary model in the cumulative odds ordinal series. Results of these methods confirm that investigation of casewise fit to the proportional odds model can be very intensive.

OLS strategies can be a good first step in the diagnostic process, but does it not capture some extreme or influential cases. Differences across statistical packages in terms of availability of options for residual diagnostics can limit the kinds of analyses a researcher has available, thus choice of statistical package should be made with full understanding of the procedures and statistics available. Further, it was demonstrated that index and change plots can yield important information about who is not being fit well by these model and these plots can augment findings from other residual strategies.

Overall, it may be concluded that reliance on one method or approach to understanding residuals from an ordinal regression model can be very misleading to the researcher. Results from this study clearly make a case for the need to consider multiple strategies in determining quality of model fit – not just overall, but for individuals as well. Further studies and extensions to this research should consider the residuals obtained from the PO model based on predicted category (i.e., classification accuracy). In addition, it would be worthwhile to pursue the use of Monte Carlo techniques to examine residuals and to control over the nature of departure from the assumption of proportional odds or other model characteristics; for example, one question not answered in the current study is the degree to which outliers or extreme values affects the determination of the assumption of proportional odds.

Substantively, it may be reasonable to consider separate models for cases that are not well-represented by a general population-type ordinal regression model. From a policy perspective, findings suggest that individual student performance or proficiency can be easily misunderstood, and in the current climate of accountability, the repercussions from such a situation can have great impact. The approaches and strategies presented here could be used to effectively argue for support for intervention programs, including gifted-education programs, or to support improved funding for special education or second-language acquisition programs. Only through residual diagnostics would the children who are left far behind, or who score far beyond their peers, be readily identified.

### Proposed Diagnostic Guidelines for Ordinal Regression Models

1. Residuals from both OLS and Binary Logistic Models provide a good first look at the potential for influential or unusual cases from an ordinal model.
2. OLS does not capture all the unique unusual values; neither do the corresponding binary logistic analyses. Thus, researchers need to be aware of the potential to miss important misfit cases and counteract this possibility by viewing/plotting as many different diagnostic statistics as possible, particularly for the binary models corresponding to a given ordinal approach (e.g., proportional odds)
3. Graphical strategies, while extensive and often time-consuming, can tell the researcher more about their data than a single summary statistic.
4. Researchers should make a commitment early on to include residual diagnostics in all their presented or published papers. It is easy to mislead oneself, one's audience, and various research stakeholders when residual diagnostic strategies are ignored.

## References

- Agresti, A. (2002). *Categorical data analysis (2<sup>nd</sup> Ed.)*. New York: John Wiley & Sons.
- Agresti, A. (2007). *An introduction to categorical data analysis (2<sup>nd</sup> Ed.)*. New York: John Wiley & Sons.
- Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*. Cary, NC: SAS Institute, Inc.
- Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology*, 26, 1323-1333.
- Armstrong, B. B., & Sloan, M. (1989). Ordinal regression models for epidemiological data. *American Journal of Epidemiology*, 129(1), 191-204.
- Bender, R., & Benner, A. (2000). Calculating ordinal regression models in SAS and S-Plus. *Biometrical Journal*, 42(6), 677-699.
- Bender, R., & Grouven, U. (1998). Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of Clinical Epidemiology*, 51(10), 809-816.
- Brant, (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46, 1171-1178.
- Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables*. Thousand Oaks, CA: Sage.
- Fagerland, M. W., Hosmer, D. W., & Bofin, A. M. (2008). Multinomial goodness-of-fit tests for logistic regression models. *Statistics in Medicine*, 27, 4238-4253.
- Fox, J. (1991). *Regression diagnostics*. Thousand Oaks, CA: Sage.
- Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer Verlag.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression (2<sup>nd</sup> Ed.)*. New York: John Wiley & Sons.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata (2<sup>nd</sup> Ed.)*. Texas: Stata Press.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models (2<sup>nd</sup> Ed.)*. London: Chapman and Hall.
- Menard, S. (1995). *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage.
- NCES (2000). *America's kindergarteners*. Retrieved August 10, 2009, from <http://www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000070>.
- O'Connell, A. A. (2000). Methods for modeling ordinal outcome variables. *Measurement and Evaluation in Counseling and Development*, 33(3), 170-193.
- O'Connell, A. A., Liu, X., Zhao, J., & Goldstein, J. (2004). *Understanding residual diagnostics of ordinal regression models*. Paper presented at the 35<sup>th</sup> Northeastern Educational Research Association annual meeting, October 2004, Kerhonkson, New York.
- O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables*. Thousand Oaks: SAGE.
- Peterson, B. L., & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39(3), 205-217.
- Powers, D. A., & Xie, Y. (2000). *Statistical models for categorical data analysis*. San Diego, CA: Academic Press.
- Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, 9, 705-724.
- Pulkstenis, E., & Robinson, T. J. (2004). Goodness-of-fit tests for ordinal response regression models. *Statistics in Medicine*, 23, 999-1014.

## Factors Influencing the Mixture Index of Model Fit in Contingency Tables Showing Independence

Xuemei Pan  
IBM Global Business Services,  
Springfield, VA USA

C. Mitchell Dayton  
University of Maryland,  
College Park, MD USA

---

Several competing computational techniques for dealing with sampling zeros were evaluated when estimating the two-point mixture model index,  $\pi^*$ , in contingency tables under an independence assumption. Also, the performance of the estimate and associated standard errors were studied under various combinations of conditions.

Key words: Mixture index, contingency tables, sampling zeros, standard error, Monte Carlo simulation.

---

### Introduction

Traditional methods for evaluating models for contingency table data based on Chi-square statistics or quantities derived from such statistics are not attractive in many applied research settings. According to Rudas (1998), “First, when the model is not true, a comparison of the data to what could only be expected if it were is of very little meaning; second, the actual distribution of the statistic may be very different from the reference distribution if some of the underlying assumptions are violated” (page 15). In addition, conventional methods are sensitive to sample size; often a model is rejected when fitted to a large data set even though the model

may represent a reasonable summary of the data for practical purposes.

In contrast to Chi-squared tests of fit methods, which rely heavily on size of the table, sample size and actual true probabilities (Rudas, 1998), the RCL mixture index of fit proposed by Rudas, Clogg and Lindsay (1994), provides a novel way of representing goodness-of-fit for contingency tables. In contrast to classical significance tests, this index has an intuitive rationale and it does not assume a simple model that describes the entire population; the RCL index is also not sensitive to sample size like Chi-square-related quantities. Specifically, two components (subgroups) are assumed in the population. One of size  $1-\pi$ , where some specified model H holds true, describes the fraction of the population that is consistent with model H (e.g., independence); the other component of size  $\pi$ , is completely unrestricted and represents the part of the population that is outside of model H. RCL also introduced an expectation-maximization (EM) algorithm to obtain maximum likelihood estimates of  $\pi^*$  and derived a way to construct a lower-bound confidence-interval estimate of  $\hat{\pi}^*$ . As summarized by Dayton (2003),  $\hat{\pi}^*$  possesses the following properties:

1.  $\hat{\pi}^*$  is always located on the 0, 1 interval;
2.  $\hat{\pi}^*$  is unique;

---

Xuemei Pan is currently a statistician at IBM Global Business Services. She earned her Ph.D. in Measurement and Statistics from University of Maryland, College Park. Her research interests include latent class modeling and model comparison procedures. Email her at: [xpan@us.ibm.com](mailto:xpan@us.ibm.com). Chan Dayton is a Professor Emeritus and past Chair in the Department of Measurement & Statistics of University of Maryland, College Park. For more than 20 years, he has pursued a research interest in latent class analysis which is a specialized field within the realm of discrete mixture models. Email him at: [cdayton@umd.edu](mailto:cdayton@umd.edu).

3.  $\hat{\pi}^*$  is invariant when frequencies in a contingency table are increased or decreased by a multiplicative constant.

Properties and applications of the mixture index of fit are further explored in Clogg, Rudas and Xi (1995), Xi (1996), and Clogg, Rudas and Matthews (1997). The two-point mixture index,  $\pi^*$ , can be applied when models are fitted to virtually any contingency table. For example, it has been applied in differential item functioning (Rudas & Zwick, 1997), latent class analysis (Dayton, 1999), regression models with normal and uniform error structures (Rudas, 1999) and logistic regression analysis (Verdes & Rudas, 2002).

Issues concerning  $\pi^*$  require further examination exist because they have not been studied in RCL or in other related research. In particular:

1.  $\hat{\pi}^*$  is positively biased in finite samples; that is, even if H holds so that, in theory,  $\pi^* = 0$ ,  $\hat{\pi}^*$  will have expectation greater than zero for finite samples.
2. Sampling 0's can greatly affect estimation so it is useful to study the effect of using flattening constants or redefining the model by regarding sampling zeros as structural zeros.
3. Although the estimated lower confidence bound of  $\hat{\pi}^*$  introduced by RCL gives inferential information that is independent of bias, it tends to be problematic when  $\pi^*$  is close to zero or sample size is small; thus a parametric simulation seems to be necessary to examine this measure of precision for  $\hat{\pi}^*$ . (As an aside, SAS code written for this study makes these analyses more accessible to researchers in various disciplines.)

#### Mixture Index of Fit

Suppose H represents a hypothesized probabilistic model for a frequency table and P is the true distribution for the cell proportions in the table. The two-point mixture model is defined as:

$$P = (1 - \pi)\Phi + \pi\psi \quad (1)$$

where  $\Phi$  is the probability distribution implied by H, and  $\psi$  is an arbitrary, unspecified probability distribution. The mixture parameter,  $\pi$ , defined on the 0, 1 interval, represents the proportion of the population that cannot be described by H. Note that  $\pi$  is not unique and that the representation of P in equation (1) is correct for any model for any frequency table. The index of fit,  $\pi^*$ , however, is defined as the smallest value of  $\pi$  for which equation (1) holds; that is:

$$\pi^* = \inf\{\pi \mid P = (1 - \pi)\phi + \pi\psi, \phi \in H\}$$

(Rudas and Zwick, 1997). Consequently, as shown by RCL,  $\pi^*$  is unique and represents the minimum proportion of cases that must be excluded from the frequency table in order for P to be fitted exactly by the model.

#### EM Algorithm and Interval Estimation

The procedure to estimate  $\hat{\pi}^*$  is as follows:

1. Set the initial estimate,  $\hat{\pi}^*$  to zero;
2. Obtain maximum likelihood estimates of the parameters in the components of the two-point mixture using an expectation-maximization (EM) algorithm as above, and,
3. Successively increase  $\hat{\pi}^*$  by some small increment with re-estimation of the parameters at each step (e.g., .01 is been used the example below).

The value of the likelihood ratio Chi-square fit statistic,  $G^2$ , converges to zero (e.g., less than a convergence criterion set to  $<10^{-5}$ ) and the step at which this first occurs provides the final estimate of the fit index,  $\hat{\pi}^*$ . (Dayton, 2003; RCL). In addition, RCL implemented this approach in their FORTRAN program, Mixit, as described in detail by Xi (1994). As shown by RCL, an appropriate lower confidence 95% bound,  $\hat{\pi}_L$ , is given by the value of  $\hat{\pi}$  that is associated with a  $G^2$  fit statistic equal to 2.71

(i.e., the 90<sup>th</sup> percentage point of the one-degree-of-freedom Chi-square distribution).

#### Sampling Zeros

According to RCL, the effect of sampling zeros on  $\hat{\pi}^*$  will depend on the structure of the data as well as the suitability of model H for the data. In general,  $\hat{\pi}^*$  will tend to be overestimated by a fraction that is directly related to the smaller of the observed row marginal proportion and the observed column marginal proportion pertinent to the cell with a sampling zero. Rudas and Zwick (1997) replaced zero frequencies with small positive flattening values in data from a study by Zwick, Thayer and Wingersky (1994) to investigate the sampling zero effect on the performance of  $\pi^*$ . Although they concluded that increasing the flattening value resulted in reducing overestimation for estimates of  $\pi^*$ , the effects were very small.

Structural zeros, also called logical zeros (Knoke and Burke, 1980), arise when it is logically impossible to observe positive cell counts for particular combinations of row and column variables. To demonstrate structural zeros, a typical example of the logical impossibility of observing male obstetrical patients was presented by Fienberg (1980). In practice, researchers could evaluate the variation in  $\pi^*$  by setting cells with no frequency to structural zeros.

#### Methodology

##### Research Design

The following aspects of the simulation were implemented:

1. Sizes of two-way contingency tables were selected: 2×2, 2×3, 2×4, 2×6, 3×3, 4×4 and 6×6. These table sizes were chosen because they provided a reasonable range of contingency table sizes in real data settings and are typical of what is found in practice.
2. Marginal distribution: evenly distributed, slightly dispersed and extremely dispersed distribution for each different table size. (Row and column total proportions for the various sized tables are shown in Figure 1.) These marginal distributions were chosen

because they represented a reasonable range of different values, and the extreme marginal values were used to ensure zero cell frequencies in the observed tables.

3. Sample size for simulated contingency table: 5, 10, 20 and 30 per cell were chosen because they entailed a practical variety of sample sizes and were large enough to demonstrate a sample size effect on the mixture index of fit.
4. Techniques for zeros cells: (A) treating as sampling zeros, (B) replacing with small flattening constants (0.1, 0.5 and 1 were used to represent extremely small, moderately small and small flattening constants range), and (C) redefining model H by regarding the sampling zeros as structural zeros.
5. In each of the above scenarios, a 95% lower confidence limit based on empirically simulated  $\hat{\pi}^*$  s was calculated and compared with the lower limit estimate presented by RCL.

For each table size, sample size and marginal distribution, 1,000 frequency tables were randomly generated based on the specified cumulative distribution. For example, for a 2×2 table with sample size of 10 per cell and marginal distribution {P<sub>1+.</sub>=. 9, P<sub>2+.</sub>=. 1, P<sub>+1.</sub>=. 9, P<sub>+2.</sub>=. 1}, the theoretical cumulative distribution is {0.81, 0.90, 0.99, 1}.

To generate each of the 1,000 simulated data tables, SAS code (SAS Institute, 2005) was used to generate 40 uniform random numbers on the 0, 1 and to locate them into appropriate cumulative categories (e. g., numbers less than or equal to 0.81 were placed in cell 1, 0.81; numbers between 0.81 and 0.90 in cell 2; numbers between 0.90 and 0.99 in cell 3 and the remainder in cell 4.) The value of  $\hat{\pi}^*$  and associated 95% lower bound  $\hat{\pi}_L$  following RCL was obtained for each generated data table; thus for each scenario, 1,000  $\hat{\pi}^*$  values and 1,000, 95% lower bound  $\hat{\pi}_L$  values were generated using RCL methods. This was repeated for each of the 96 scenarios. Also for each scenario, four techniques for sampling zeros cells were

Figure 1: Row and Column Total Proportions for the Various Sized Tables

2×2 Table

$$\begin{aligned} &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.5, P_{+2}=.5\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.9, P_{+2}=.1\}, \\ &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.9, P_{+2}=.1\}. \end{aligned}$$

2×3 Table

$$\begin{aligned} &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.8, P_{+2}=.1, P_{+3}=.1\}, \\ &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.33, P_{+2}=.33, P_{+3}=.33\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.8, P_{+2}=.1, P_{+3}=.1\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.33, P_{+2}=.33, P_{+3}=.33\}. \end{aligned}$$

2×4 Table

$$\begin{aligned} &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.25, P_{+2}=.25, P_{+3}=.25, P_{+4}=.25\}, \\ &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.4, P_{+2}=.4, P_{+3}=.1, P_{+4}=.1\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.25, P_{+2}=.25, P_{+3}=.25, P_{+4}=.25\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.4, P_{+2}=.4, P_{+3}=.1, P_{+4}=.1\}. \end{aligned}$$

2×6 Table

$$\begin{aligned} &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.167, P_{+2}=.167, P_{+3}=.167, P_{+4}=.167, P_{+5}=.167, P_{+6}=.167\}, \\ &\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.3, P_{+2}=.3, P_{+3}=.1, P_{+4}=.1, P_{+5}=.1, P_{+6}=.1\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.167, P_{+2}=.167, P_{+3}=.167, P_{+4}=.167, P_{+5}=.167, P_{+6}=.167\}, \\ &\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.3, P_{+2}=.3, P_{+3}=.1, P_{+4}=.1, P_{+5}=.1, P_{+6}=.1\}. \end{aligned}$$

3×3 Table

$$\begin{aligned} &\{P_{1+}=.4, P_{2+}=.4, P_{3+}=.2, P_{+1}=.4, P_{+2}=.4, P_{+3}=.2\}, \\ &\{P_{1+}=.33, P_{2+}=.33, P_{3+}=.33, P_{+1}=.33, P_{+2}=.33, P_{+3}=.33\}, \\ &\{P_{1+}=.33, P_{2+}=.33, P_{3+}=.33, P_{+1}=.4, P_{+2}=.4, P_{+3}=.2\}. \end{aligned}$$

4×4 Table

$$\begin{aligned} &\{P_{1+}=.25, P_{2+}=.25, P_{3+}=.25, P_{4+}=.25, P_{+1}=.25, P_{+2}=.25, P_{+3}=.25, P_{+4}=.25\}, \\ &\{P_{1+}=.4, P_{2+}=.4, P_{3+}=.1, P_{4+}=.1, P_{+1}=.4, P_{+2}=.4, P_{+3}=.1, P_{+4}=.1\}, \\ &\{P_{1+}=.25, P_{2+}=.25, P_{3+}=.25, P_{4+}=.25, P_{+1}=.4, P_{+2}=.4, P_{+3}=.1, P_{+4}=.1\}. \end{aligned}$$

6×6 Table

$$\begin{aligned} &\{P_{1+}=.167, P_{2+}=.167, P_{3+}=.167, P_{4+}=.167, P_{5+}=.167, P_{6+}=.167, P_{+1}=.167, P_{+2}=.167, P_{+3}=.167, P_{+4}=.167, P_{+5}=.167, \\ &\quad P_{+6}=.167\}, \\ &\{P_{1+}=.3, P_{2+}=.3, P_{3+}=.1, P_{4+}=.1, P_{5+}=.1, P_{6+}=.1, P_{+1}=.3, P_{+2}=.3, P_{+3}=.1, P_{+4}=.1, P_{+5}=.1, P_{+6}=.1\}, \\ &\{P_{1+}=.167, P_{2+}=.167, P_{3+}=.167, P_{4+}=.167, P_{5+}=.167, P_{6+}=.167, P_{+1}=.3, P_{+2}=.3, P_{+3}=.1, P_{+4}=.1, P_{+5}=.1, P_{+6}=.1\}. \end{aligned}$$

compared: treating zero cells as sampling zeros, replacing with different small flattening constant (i.e., 0.1, 0.5 and 1), and redefining model H by regarding a sampling zero as a structural zero.

The mean of the 1,000  $\hat{\pi}^*$  values for

each scenario was calculated and served as the final parameter estimate; the mean of the 1,000  $\hat{\pi}_L$  values was also computed to be the estimate 95%  $\hat{\pi}_L$  using the RCL method. Because the empirical distribution of  $\hat{\pi}^*$  is notably skewed



for the generated sets of 1,000  $\hat{\pi}^*$  values, the regular normal assumption cannot be used to compute the standard error and confidence interval for  $\hat{\pi}^*$ . Instead, 50<sup>th</sup>  $\hat{\pi}^*$  value among the 1,000 values (i.e., 5<sup>th</sup> percentage point) was adopted and treated as true 95% lower bound based on empirical simulations.

Typically,  $\hat{\pi}^*$  will tend to be overestimated by a fraction that is directly related to the smaller of the observed row marginal proportion and the observed column marginal proportion related to the cell with a sampling zero (RCL). As noted above, in practice, researchers could test the  $\pi^*$  variation by setting some to-be-ignored cells to structural zeros and resolve. This study focused this issue on any frequency tables with only one structural zero and the procedure using EM based methodology to obtain  $\hat{\pi}^*$ . The two-point mixture using an expectation-maximization (EM) algorithm proposed by RCL could still be applied to structural zero conditions with minor modification as follows:

1. Obtain  $\hat{\pi}^*$  treating zero cell as sampling zero utilized the same procedure in EM Algorithm and Interval Estimation; in this step the entire row or column with which smaller of observed row marginal proportion and the observed column marginal proportion would result in zero in the first component,  $\Phi$ , which is defined as the probability distribution designated by H.
2. Pull the proportion back from the second component,  $\Psi$ , an unspecified probability distribution outside of model H for the entire row or column with zeros in component  $\Phi$  at step 1.
3. Temporally cross out the other column or row that contains the zero cell but has not been forced zero at step 1.
4. Apply the same EM based procedure in the remaining contingency table while fixing all cell proportions in component 1,  $\Phi$  and component 2,  $\Psi$  except the row or column has frequency pulled back in step 2.
5. After iterations converge to a preset criterion,, subtract original  $\hat{\pi}^*$  at step 1 with the sum of the proportion pulled back in  $\Phi$  from step 4 and the final value is the estimate of  $\hat{\pi}^*$  using structural zero technique.

For the other sampling zero techniques, procedures are same as sampling zeros, simply replace the zero cell with different small flattening constant (0.1, 0.5 and 1) and recall associated  $\hat{\pi}^*$  based on the EM based procedures in EM Algorithm and Interval Estimation.

#### Simulation Details

The simulation code was written in SAS/IML version 9.1 (SAS Institute, 2005). The EM algorithm was used to calculate the mixture index of fit. Each simulation consisted of 1,000 replications with convergence criterion set to  $10^{-5}$ . Data were randomly generated according to cumulative proportion resulting from the different combination scenarios.

The method proceeded in the following manner:

1. A sample contingency table was randomly generated based on cumulative proportion resulting from different factor combinations. (table size, sample size and marginal distribution).
2. An EM algorithm based method for mixture index of fit (RCL) was implemented.  $\hat{\pi}^*$  and 95% lower bound  $\hat{\pi}_L$  were generated and saved in a matrix.
3. Replicate steps 1 and 2 1,000 times, therefore 1,000  $\hat{\pi}^*$  and  $\hat{\pi}_L$  were obtained and exported into an external file. Additionally, if any of the 1,000 generated contingency tables contained zero cell(s), they were replaced with different small flattening constants 0.1, 0.5 and 1, respectively, when evaluating the performance of  $\hat{\pi}^*$  using flattening constants techniques.

The only difference between the structural zero and other sampling zero technique procedure is in the above-mentioned step 1. If the frequency tables generated by UNIFORM contained 1 or less than 1 frequency zero, it would proceed to step 2 otherwise it would regenerate the table until it met the requirement.

### Results

#### Parameter Estimates and Bias

For the conditions studied,  $\hat{\pi}^*$  was significantly ( $p < 0.05$ ) positively, biased from its expected value of zero by an amount ranging from 0.02298 ( $2 \times 2$  table, slightly dispersed row and column marginals with sample size equals to 30 per cell) to 0.4086 ( $6 \times 6$  table, evenly dispersed row and column marginals with sample size equal to 5 per cell). As shown in Figures 2 and 3, for  $2 \times 2$ ,  $2 \times 3$ ,  $2 \times 4$ ,  $2 \times 6$  tables, as table size increases,  $\hat{\pi}^*$  consistently increased (with only two exceptions) for constant sample size (5, 10, 20 and 30 per cell) and marginal distribution (evenly, slightly and extremely dispersed).

The same conclusion applies to symmetric tables:  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ ,  $6 \times 6$ . In particular, for sample sizes 5, 10, 20 and 30 per cell in evenly dispersed tables,  $\hat{\pi}^*$  increased on average from 0.1252 to 0.4086; 0.096 to 0.3031; 0.0775 to 0.2242 and 0.0668 to 0.1867 for  $2 \times 2$  to  $6 \times 6$  tables, respectively. For sample sizes 5, 10, 20 and 30 per cell in extremely dispersed tables,  $\hat{\pi}^*$  increases on average from 0.0598 to 0.03629; 0.0568 to 0.2593; 0.0476 to 0.1942 and 0.0396 to 0.1626 for a  $2 \times 2$  table to a  $6 \times 6$  table, respectively.

Moreover, with few exceptions, for each frequency table, as sample size increases, the bias in  $\hat{\pi}^*$  significantly decreased ( $p < 0.05$ ). For each size contingency table,  $\hat{\pi}^*$  is, on average, smallest for extremely dispersed row and column marginal distributions, and largest on average for evenly distributed row and column tables. The only exception is the  $2 \times 2$  table where a slightly dispersed table contains slightly smaller  $\hat{\pi}^*$  values on average than an extremely dispersed frequency table; this is in

part due to a convergence problem (using less than 0.001 instead of otherwise 0.00001).

For all two-way tables, replacing zero with larger flattening values results in smaller average values of  $\hat{\pi}^*$ . For all extremely dispersed and most slightly dispersed (4 out of 6 scenarios) row and column marginal distributions with small sample size (5 per cell) and small table size ( $2 \times 2$ ,  $2 \times 3$ ,  $2 \times 4$ ,  $3 \times 3$ ) tables, the value of  $\hat{\pi}^*$  was smaller using structural zeros compared to using sampling zeros or any other replacement with positive flattening constants. Note that the techniques of replacing zero cell with flattening constants includes virtually any number of simulated zero cells for each table while the structural zero technique used in this study can only accommodate one zero cell per frequency table. Because the number of zero counts and patterns are somewhat different among these techniques, especially when encountering small sample sizes such as 5 per cell and 10 per cell, it might influence the comparison results between structural zero and using sampling zero or any other replacing with small positive flattening constants techniques.

#### Lower Bound Comparisons of RCL and True Estimates

The 95% lower bound estimate for  $\hat{\pi}^*$  using the RCL method is generally close to the so-called true estimate based on empirical simulations. When, under some circumstances, the RCL method underestimates the lower bound value, the magnitude of underestimation is relatively small and the difference from the true estimate decreases as the sample size increases.

Similar to parameter estimators for  $\hat{\pi}^*$ , the true (empirical) 95% lower bound estimates of  $\hat{\pi}^*$  consistently increased as table size increased within constant sample size per cell and constant marginal distribution (Figures 6 and 7). There are exceptions for  $2 \times 3$  and  $2 \times 4$  extremely dispersed tables with sample size 5 for which estimates remain nearly unchanged over conditions. Also in general for each frequency table, as sample size increases the 95% lower bound decreases.

Figure 2:  $\hat{\pi}^*$  for Evenly Distributed Marginals

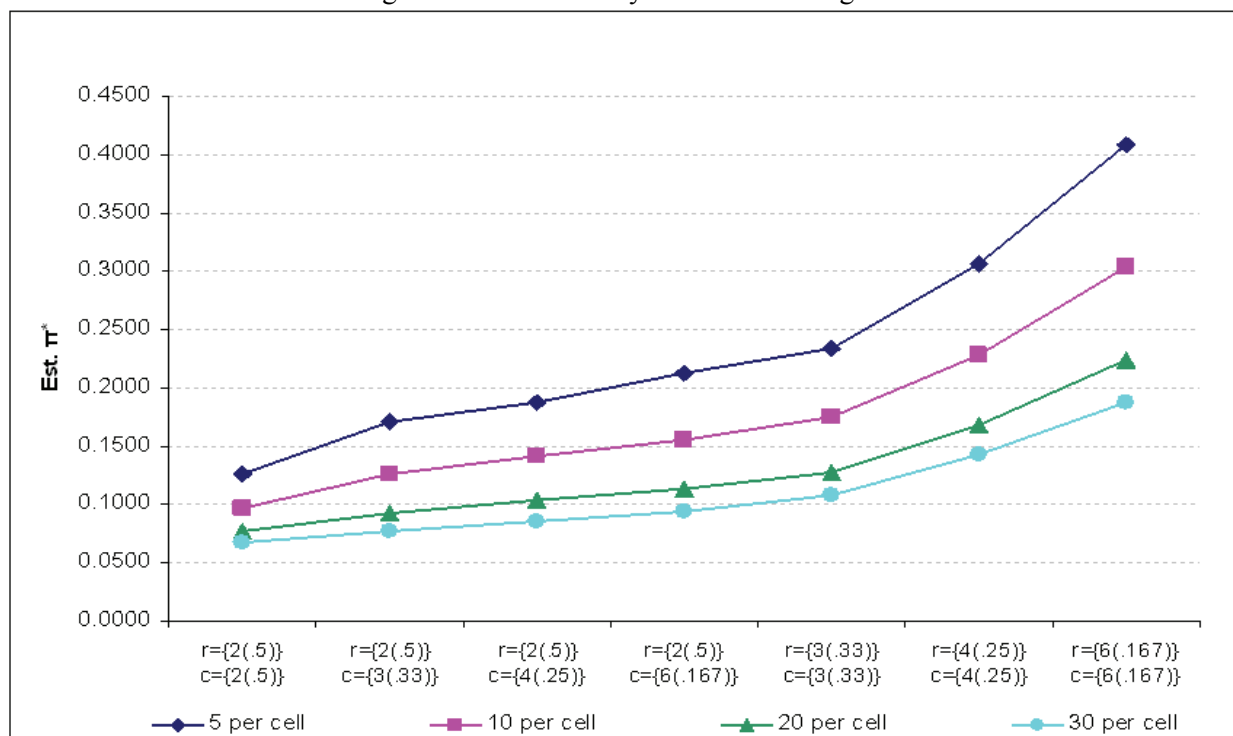


Figure 3:  $\hat{\pi}^*$  for Extremely Distributed Marginals

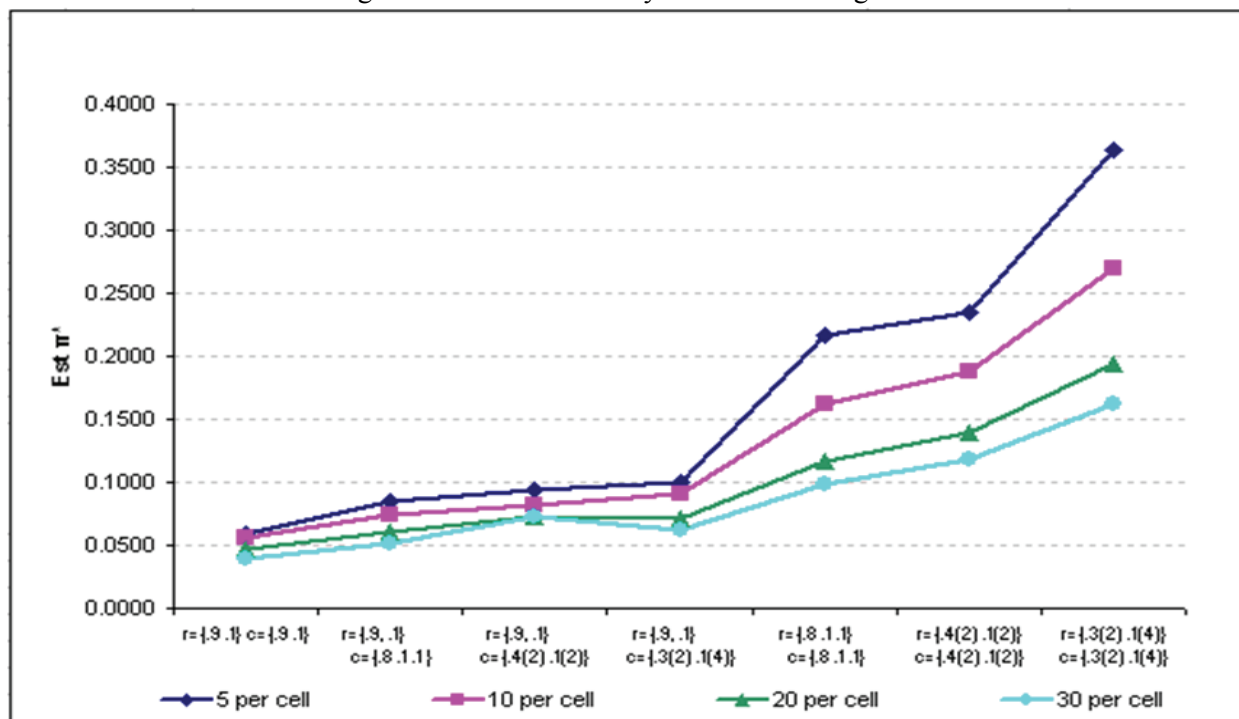


Figure 4:  $\hat{\pi}^*$  Comparison in  $\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.8, P_{+2}=.1, P_{+3}=.1\}$

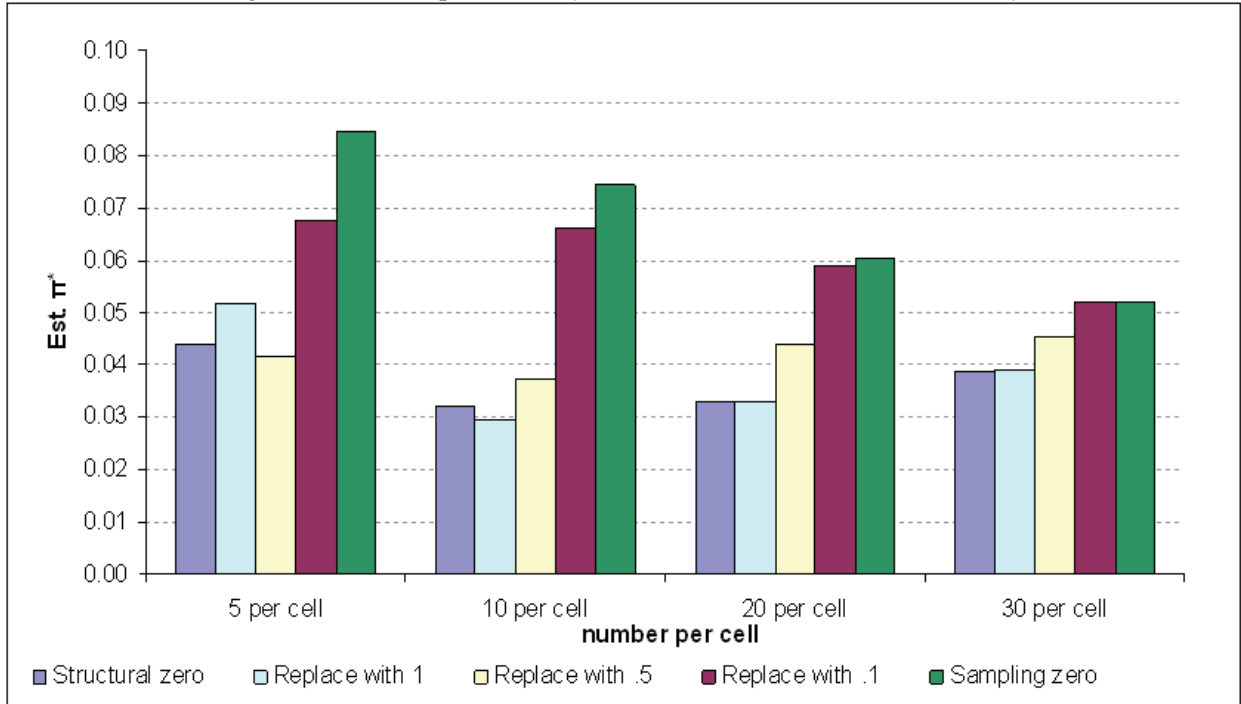


Figure 5:  $\hat{\pi}^*$  Comparison in  $\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.4, P_{+2}=.4, P_{+3}=.1, P_{+4}=.1\}$

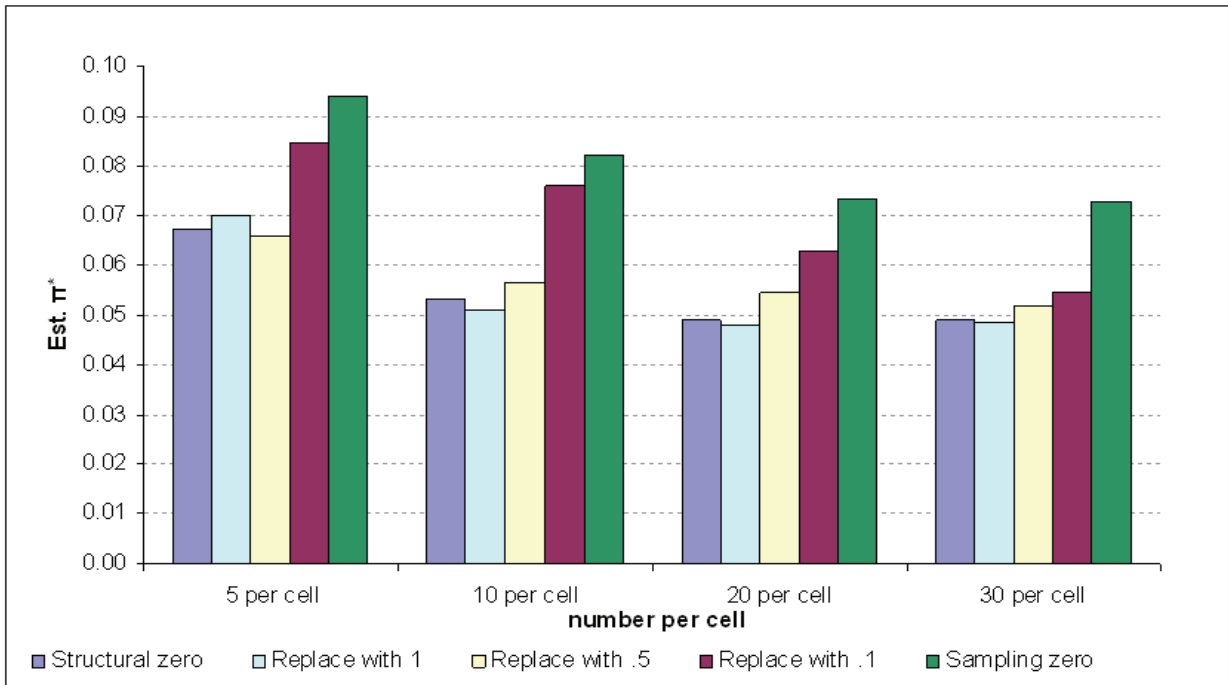


Figure 6: Empirical Simulation Based  $\hat{\pi}_L$  with Evenly Distributed Marginals

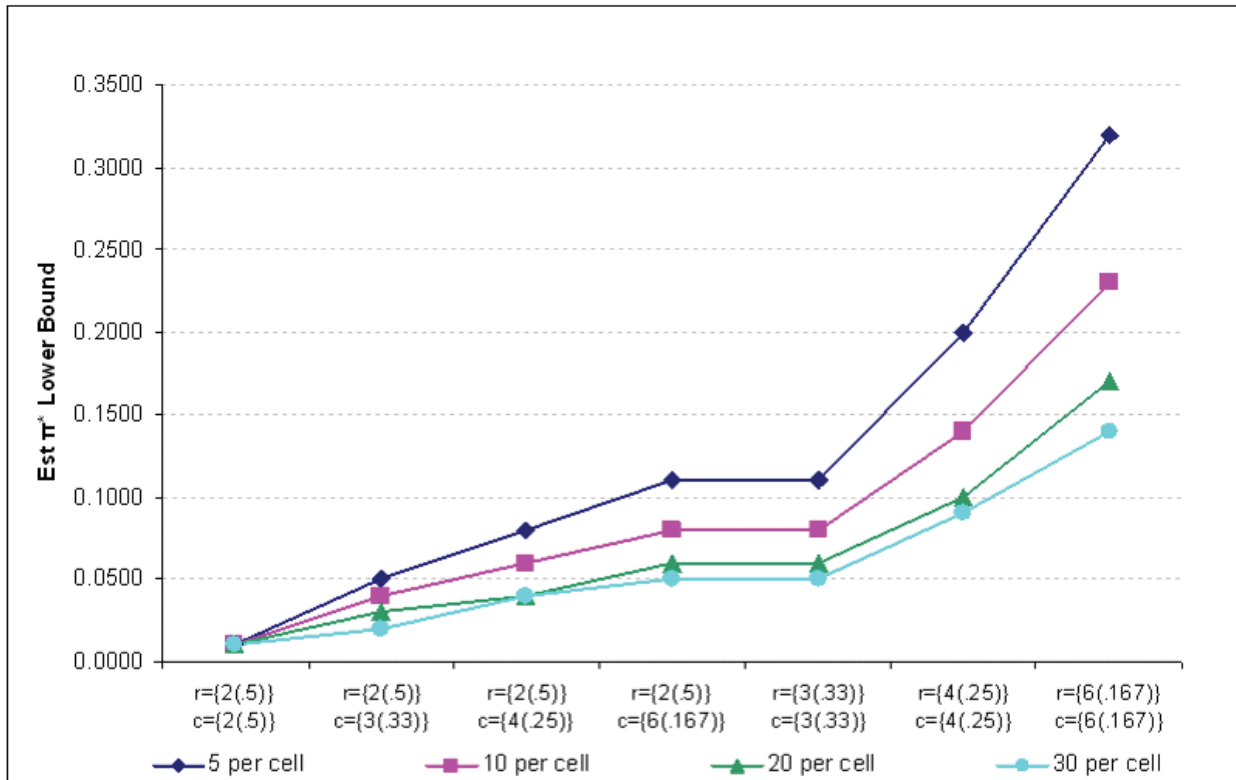
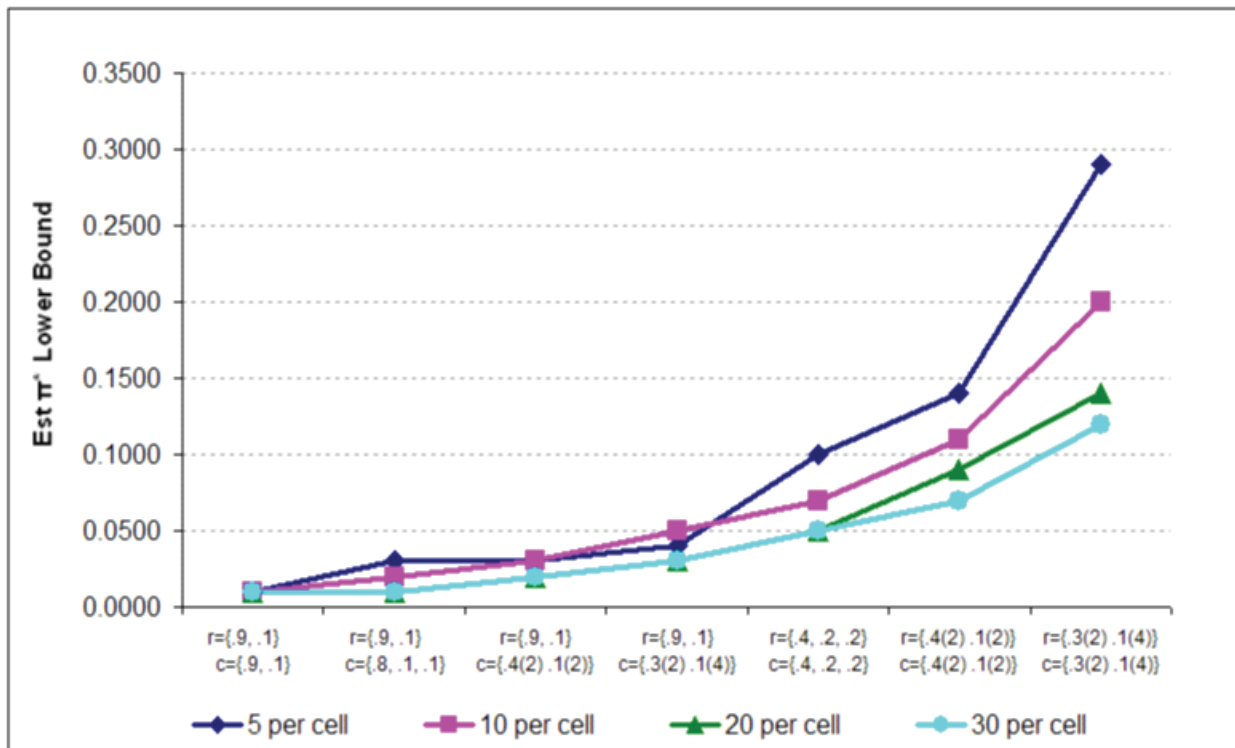


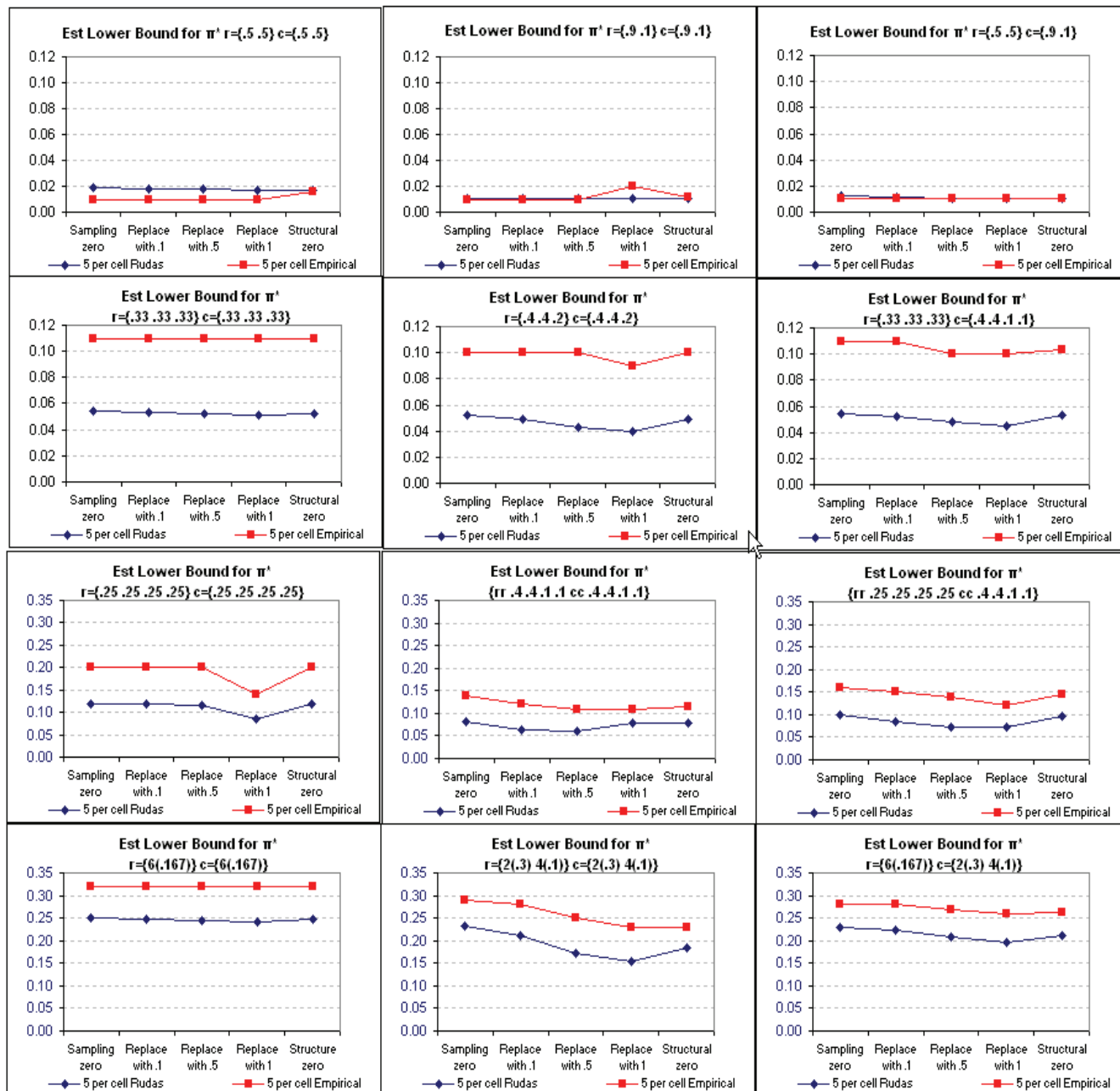
Figure 7: Empirical Simulation Based  $\hat{\pi}_L$  with Extremely Distributed Marginals



As shown in Figures 8 and 9, for each size of contingency table, the lower bound estimate of  $\hat{\pi}^*$  is generally smallest for extremely dispersed row and column marginal distributions, followed by slightly dispersed row and column marginal distributions; while largest

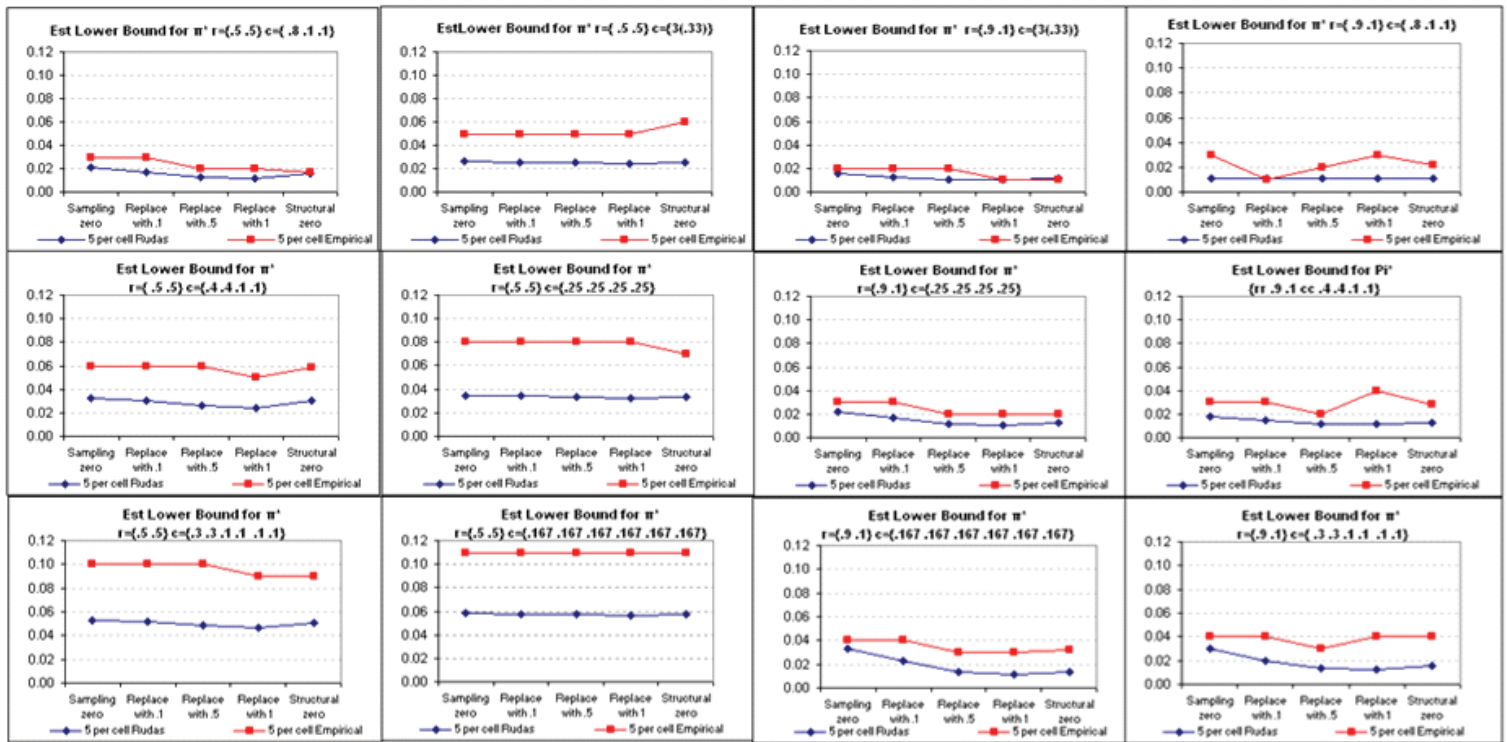
for evenly distributed row and column tables. Different techniques for dealing with sampling zeros seem to have no effect on the lower bound estimate of  $\hat{\pi}^*$  for either the RCL method or the true lower bound estimate based on empirical simulations.

Figure 8:  $\hat{\pi}_L$  Comparison between the RCL Method and Empirical Simulation Method



# MIXTURE INDEX OF MODEL FIT IN CONTINGENCY TABLES WITH INDEPENDENCE

Figure 9:  $\hat{\pi}_L$  Comparison between the RCL Method and Empirical Simulation Method (continued)



As shown in Figures 8 and 9, for each size of contingency table, the lower bound estimate of  $\hat{\pi}^*$  is generally smallest for extremely dispersed row and column marginal distributions, followed by slightly dispersed row and column marginal distributions; while largest for evenly distributed row and column tables. Different techniques for dealing with sampling zeros seem to have no effect on the lower bound estimate of  $\hat{\pi}^*$  for either the RCL method or the true lower bound estimate based on empirical simulations.

### Confidence Interval and Standard Errors

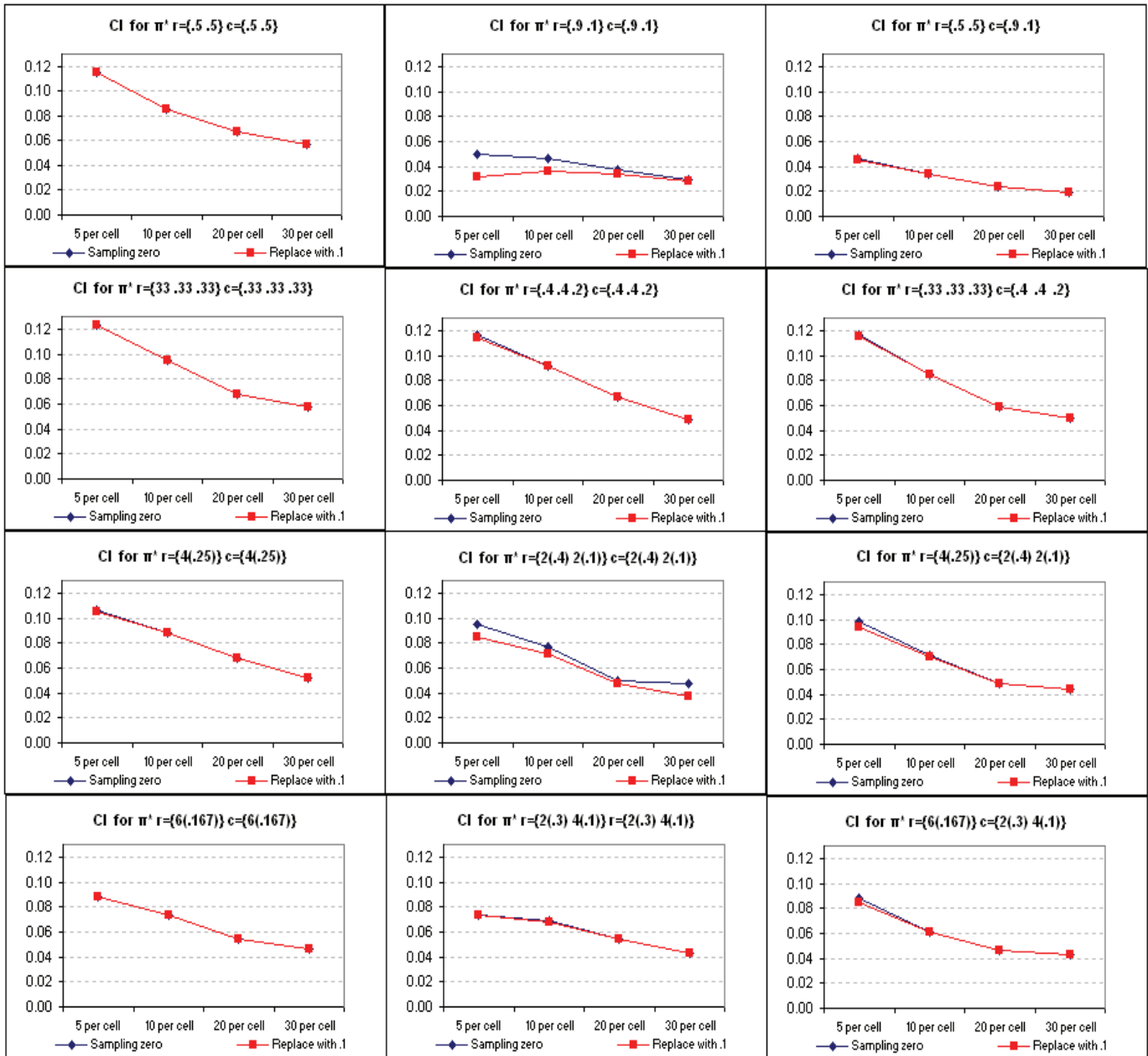
Figures 10 and 11 show that, given the same table size, extremely dispersed row and column marginal distributions consistently provide narrower confidence intervals ( $\hat{\pi}^* - \hat{\pi}_L$ ) than evenly dispersed row and column tables using both the RCL method and empirical true estimates. Also, when sample size

increases, confidence intervals become narrower for each table size and shrink approximately to the same confidence intervals for different marginal distribution for the same table size using both estimation methods. It is apparent that the RCL method underestimates the lower bound of  $\hat{\pi}^*$  in many cases and, thus, leads to a higher standard error compared with empirical true lower bound estimates.

### Example 1: Fatal Crashes by Speed Limit

Table 1 presents fatal crashes by speed limit and land use in the United States in 2004 from Traffic Safety Facts 2004, a compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. There are three categories in the land use variable (rural, urban and unknown), and six categories in the speed limit variable (30 mph or less, 35 or 40 mph, 45 or 50 mph, 55mph, 60 mph or higher and no statutory limit). This data table was used to compare the conclusion using traditional Chi-square and

Figure 10: Confidence Interval of  $\hat{\pi}^*$  Following Empirical Simulation Method



related model fit methods and the mixture index of fit introduced by RCL. More specifically, compare different sampling zero techniques impact on  $\hat{\pi}^*$  because there is one zero cell in the contingency table.

The value of the Pearson Chi-Square statistic is 7200.090, and the likelihood ratio,  $G^2$  statistic is 7600.54 both with degrees of freedom

equal to 10 ( $P < 0.01$ ). Thus, an independence model is not tenable based on these Chi-squared tests of fit. As displayed in Table 3, the mixture index of fit  $\hat{\pi}^*$  is 0.294, indicating that about 29.4% of the total of 37,295 cases (or, 10,965 cases) must be removed in order to attain perfect model fit. The mixture index of fit provides an interpretation consistent with traditional Chi-



MIXTURE INDEX OF MODEL FIT IN CONTINGENCY TABLES WITH INDEPENDENCE

Figure 11: Confidence Interval of  $\hat{\pi}^*$  (continued)

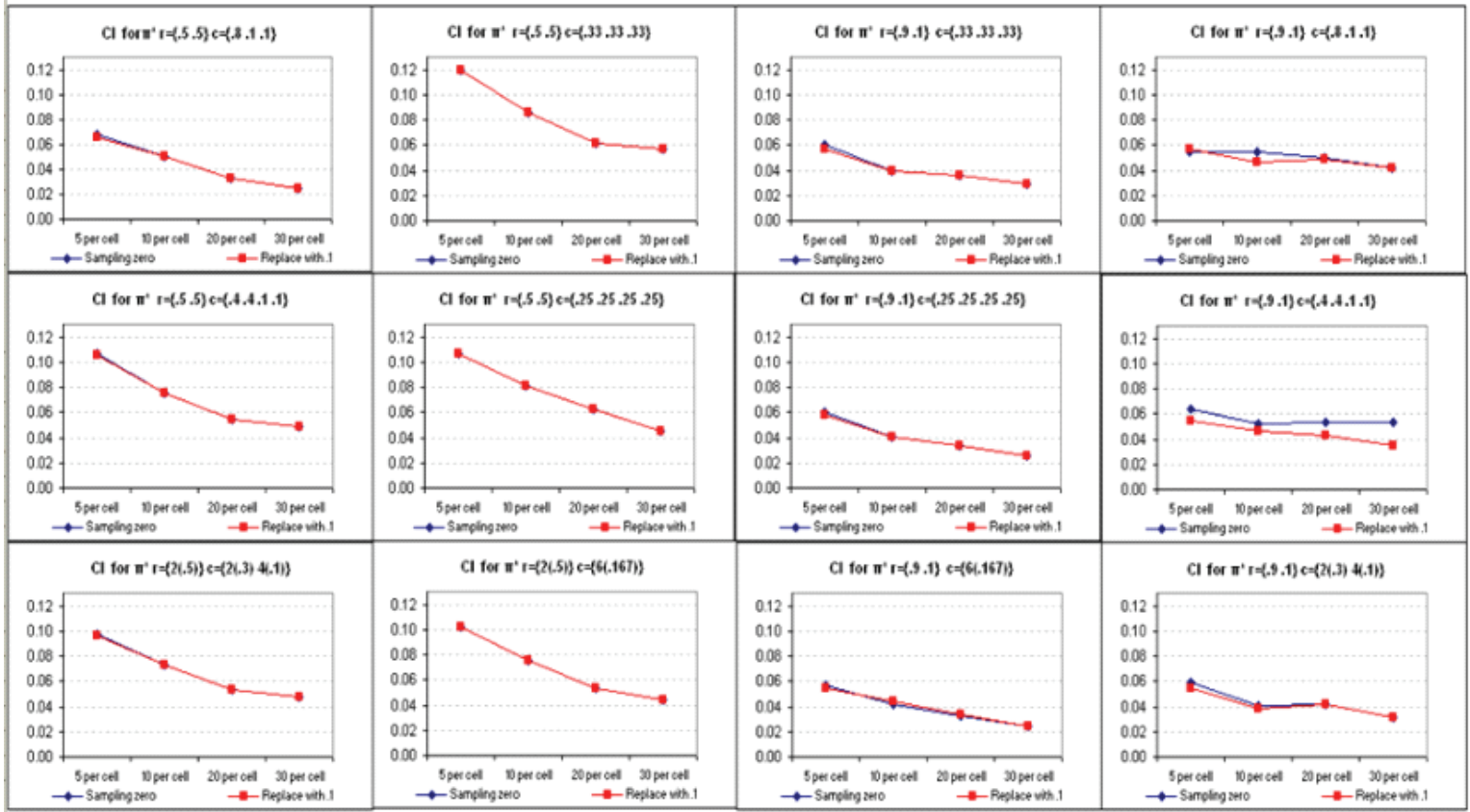


Table 1: Fatal Crashes by Speed Limit and Land Use

Speed Limit	Land Use		
	Rural	Urban	Unknown
30 mph or less	944	2929	27
35 or 40 mph	1951	4463	41
45 or 50 mph	3496	3559	46
55 mph	9646	2121	91
60 mph or higher	5484	2347	27
No statutory limit	92	31	0

Source: USDOT Traffic Safety Facts 2004 (Fatality Analysis Reporting System). Note: Omit 958 cases for the Unknown Speed Limit category.

Square analyses. Furthermore,  $\hat{\pi}^*$  only decreases to 0.293 when replacing sample zero with the flattening constant 0.1 and further reduces to 0.291 when replacing with 0.5 and 1 as well as using the structural zero method. The amount of change in  $\hat{\pi}^*$ , as well as its 95% lower bound using different sampling zero techniques, is extremely small in this example. This occurs due to the very small percentage (0.62%) of unknown land. In fact, it would not substantially effect  $\hat{\pi}^*$  even if the entire column were zeros.

Example 2: Eye Color and Hair Color

Table 2 presents a cross-classification of

eye color and hair color table (Snee, 1974), a  $4 \times 4$  table with total sample size of 592.

RCL utilized these data to study the properties of the mixture index of fit. In this study, these data were used to compare the differences in estimates that result from using sampling zeros and structural zeros. The 16 cells were set to zero one-by-one and the results are shown in Table 4. The percentage differences between use of sampling zero and structural zero techniques range from 11.1% to 40.0%. Note that 6 of these differences are statistically significant ( $p < 0.05$ ) using conventional z tests for proportions. The largest reductions in  $\hat{\pi}^*$  using structural zero occurs when black hair and hazel eye color is set to zero.

Table 2: Cross-classification of Eye Color and Hair Color

Eye Color	Hair Color			
	Black	Brunette	Red	Blonde
Brown	68	119	26	7
Blue	20	84	17	94
Hazel	15	54	14	10
Green	5	29	14	16

Source: Snee (1974), Diaconis & Efron (1985).

Table 3: Fit Statistics for Fatal Crashes by Speed Limit and Land Use

$\pi$	Chi-Square	G-Square	$\pi$ Repl with .1	Chi-Square	G-Square	$\pi$ Repl with .5	Chi-Square	G-Square	$\pi$ Repl with 1	Chi-Square	G-Square
0	7200.090	7600.540	0.000	7287.906	7599.937	0.000	7287.426	7599.116	0.000	7287.426	7599.116
0.1	2947.100	3044.340	0.100	2946.904	3043.700	0.100	2946.353	3042.763	0.100	2946.353	3042.763
0.15	1527.210	1578.220	0.150	1527.010	1577.582	0.150	1526.444	1576.641	0.150	1526.444	1576.641
0.2	632.920	655.610	0.200	632.726	654.986	0.200	632.173	654.094	0.200	632.174	654.094
0.25	140.530	142.830	0.250	140.330	142.266	0.250	139.875	141.612	0.250	139.875	141.612
0.26	81.050	82.280	0.260	80.856	81.737	0.260	80.449	81.149	0.260	80.449	81.149
0.27	37.190	37.890	0.270	37.003	37.364	0.270	36.643	36.841	0.270	36.643	36.841
0.28	10.010	10.510	0.280	9.824	9.992	0.280	9.060	9.534	0.280	9.510	9.534
<i>0.286</i>	<i>2.700</i>	<i>2.560</i>	<i>0.286</i>	<i>1.923</i>	<i>2.058</i>	<i>0.285</i>	<i>2.498</i>	<i>2.507</i>	<i>0.285</i>	<i>2.498</i>	<i>2.507</i>
0.29	0.510	0.730	0.290	0.272	0.321	0.290	0.004	0.004	0.290	0.004	0.004
0.294	0.000	0.000	0.293	0.000	0.000	0.291	0.000	0.000	0.291	0.000	0.000
<b><math>\pi \geq 0.294</math></b>	<b>0.000</b>	<b>0.000</b>	<b><math>\pi \geq 0.293</math></b>	<b>0.000</b>	<b>0.000</b>	<b><math>\pi \geq 0.291</math></b>	<b>0.000</b>	<b>0.000</b>	<b><math>\pi \geq 0.291</math></b>	<b>0.000</b>	<b>0.000</b>

NOTE: 1. Italic figures denote a 95% lower bound for  $\hat{\pi}^*$  and associated Chi-Square and G-Square values.

NOTE: 2. Italic and bolded figures denote  $\hat{\pi}^*$  and the associated Chi-Square and G-Square values.

MIXTURE INDEX OF MODEL FIT IN CONTINGENCY TABLES WITH INDEPENDENCE

Table 4:  $\hat{\pi}^*$  Comparison of Sampling Zero and Structure Zero using Eye Color Data  
(Each cell manipulated to be zero in turn.)

Eye color	Hair color											
	Black			Brunette			Red			Blonde		
	Sampling zero	Structure zero	%	Sampling zero	Structure zero	%	Sampling zero	Structure zero	%	Sampling zero	Structure zero	%
Brown	0.32	0.27	18.52%	0.42	0.40	5.00%	0.37	0.32	15.63%	0.32	0.28	14.29%
Blue	0.37	0.31	<b>19.35%</b>	0.44	0.38	15.79%	0.38	0.31	<b>22.58%</b>	0.20	0.18	11.11%
Hazel	0.42	0.30	<b>40.00%</b>	0.38	0.32	<b>18.75%</b>	0.38	0.31	<b>22.58%</b>	0.32	0.30	6.67%
Green	0.35	0.29	<b>20.69%</b>	0.32	0.30	6.67%	0.34	0.30	13.33%	0.34	0.29	17.24%

NOTE: 1. Bold figures denote significant percentage difference ( $p < .05$ , conventional z test for proportions) between sampling zero and structural zero techniques.

Recommendations

Among all the sampling zero techniques compared in terms of parameter bias, replacing zeros with larger flattening constants such as 1 and the structural zero technique appear to perform better in the sense that, on average,  $\hat{\pi}^*$  is smaller. Between these two techniques, the structural zero technique is generally recommended for extremely and slightly dispersed row and column marginal distributions tables with small sample sizes and small table sizes while in other cases replacing with larger flattening constant (i.e., 1) is preferred.

Based on the current findings, RCL standard error estimates were comparatively conservative. In general, it is preferable in practice to use variance estimates that tend to be conservative (i.e., larger) rather than liberal (i.e., smaller). However, it would be valuable to investigate the standard error of  $\hat{\pi}^*$  using re-sampling methods to provide better guidance for users.

Implications for Future Research

1. Evenly distributed, slightly and extremely dispersed marginal distributions for each different size of tables were manipulated in the current study. It would be valuable to investigate more diversified marginal distribution in future studies.
2. As noted, the limitation of structural zero technique with number of zero cells might affect the results when compared with other

sampling zero techniques. It would be of interest to investigate structural zero technique applied in two-point mixture model index in contingency tables with more than one zero when the independence assumption holds.

3. In order to attain reasonable execution times for the simulation, in this study, an increment of .01 was adopted to successively increase  $\hat{\pi}^*$  when estimating  $\hat{\pi}^*$  using an EM algorithm. For very small true values of  $\pi^*$ , it would be necessary to use a value of .001 or even .0001 in order to obtain a more detailed picture, especially for the lower bound of  $\hat{\pi}^*$ .
4. In a future study, it would be beneficial to investigate the standard error of  $\hat{\pi}^*$  using other re-sampling methods (e.g., jackknife) and compare with RCL to provide a more concrete guide.
5. The larger value of flattening constants (e.g., 1) might affect the original data structural when sample size of a contingency table is small (e.g., 5 per cell) and thus the results could be slightly influenced. Alternative ways to define the flattening constants such as a percentage to total sample size is of interest in future study.
6. Finally, it would be valuable to evaluate the performance of  $\pi^*$  under conditions where the independence assumption does not hold.

## References

- Clogg, C. C., Rudas, T., & Xi, L. (1995). A new index of structure for the analysis of models for mobility tables and other cross-classifications. In P. Marsden (Ed.) *Sociological Methodology*, 197-222. Blackwell, Oxford.
- Clogg, C. C., Rudas, T., & Matthews, S. (1997). Analysis of model misfit, structure, and local structure in contingency tables using graphical displays based on the mixture index of fit. In J. Blasius, M. Greenacre (Eds.) *Visualization of Categorical Data*, 425-439. New York: Academic Press.
- Dayton, C. M. (1999). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.
- Dayton, C. M. (2003). Applications and computational strategies for the two-point mixture index of fit. *British Journal of Mathematical and Statistical Psychology*, 56, 1-13.
- Diaconis, P., & Efron, B. (1985). Testing for independence in a two-way contingency table: New interpretations of the Chi-square statistic. *The Annals of Statistics*, 13, 845-874.
- Fienberg, S. (1980). *The analysis of cross-classified categorical data* (2<sup>nd</sup> Ed.). Boston, MA: MIT Press.
- Knoke, D., & Burke, P. J. (1980). *Log-linear models*. Newberry Park, California: Sage Publications, Inc.
- National Highway Traffic Safety Administration. (2006). *Traffic safety facts, 2004*. Report No. DOT HS-809-919. Washington, DC: US Department of Transportation.
- Rudas, T. (1998). The mixture index of fit. In Ferligoj (Ed.) *Advances in methodology, data analysis and statistics*, 15-22. University of Ljubljana.
- Rudas, T. (1999). The mixture index of fit and minimax regression. *Metrika*, 50, 163-172.
- Rudas, T., Clogg, C. C., Lindsay, & B. G. (1994). A new index of fit based on mixture methods for the analysis of contingency tables. *Journal of the Royal Statistical Society, B*, 56(4), 623-639.
- Rudas, T., & Zwick, R. (1997). Estimating the importance of differential item functioning. *Journal of Educational and Behavioral Statistics*, 22, 31-45.
- SAS Institute Inc. (2005). *SAS User's Guide, Version 9*. Cary, NC: SAS Institute Inc.
- Snee, R. (1974). Graphical display of two-way contingency tables. *The American Statistician*. 28, 9-12.
- Verdes, E., & Rudas, T. (2002). The  $\pi^*$  index as a new alternative for assessing goodness of fit of logistic regression. In: Haitovsky, y., Lerche, H. R., Ritov, y. (Eds.) *Foundations of statistical inference*, 167-177. New York: Springer.
- Xi, L. (1994). *The mixture index of fit for the independence model in contingency tables*. Master of Arts paper, Department of Statistics, Pennsylvania State University
- Xi, L., & Lindsay, B. G. (1996). A note on calculating the  $\pi^*$  index of fit for the analysis of contingency tables. *Sociological Methods & Research*, 25, 248-259.
- Zwick, R., Rhayer, D. T., & Wingersky, M. (1994). A simulation study of the methods for assessing differential item functioning in computerized adaptive tests. *Applied Psychological Measurement*, 18(9), 121-140.

## Inference in Simple Regression for the Intercept Utilizing Prior Information on the Slope

Ayman Baklizi Adil E. Yousif  
Qatar University, Qatar

Shrinkage type estimators are developed for the intercept parameter of a simple linear regression model and the case when it is suspected a priori that the slope parameter is equal to some specific value is considered. Three different estimators of the intercept parameters are examined. The relative performances of the estimators are investigated based on a simulation study of the biases and mean squared errors. The associated bootstrap confidence intervals are also studied and their performance is evaluated.

Key words: Prior information, preliminary test estimator, shrinkage estimators, p-value, bootstrap intervals.

### Introduction

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

where the error terms  $\varepsilon_i$  are independent and identically distributed as  $N(0, \sigma^2)$ . The aim is to estimate the intercept parameter  $\beta_0$  when prior information that the slope parameter,  $\beta_1$ , is equal to some specific value,  $\beta_1^0$  is uncertain. In the absence of any prior information, the maximum likelihood (equivalent, least squares) estimator of the regression parameters are given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

If the slope parameter is known and is equal to  $\beta_1^0$ , then the estimator of the intercept parameter  $\beta_0$  is  $\hat{\beta}'_0 = \bar{y} - \beta_1^0 \bar{x}$ . The additional uncertain prior information about the slope parameter is utilized with a view to produce improved estimators. Khan and Saleh (1997) developed the preliminary test estimator and certain types of shrinkage estimators that utilize a so-called distrust factor and adopted this approach to estimate the slope parameters of two suspected parallel regression models. Bhoj and Ahsanullah (1993, 1994) considered estimating the conditional mean for simple regression model, and Khan and Saleh (1997) discussed the problem of shrinkage preliminary test estimation for the multivariate Student-t regression model.

### The Estimators

The idea of preliminary test estimation is based on utilizing the result of a preliminary test in choosing between two alternative estimators. In the case considered herein, the two estimators are the unrestricted estimator  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  and the restricted estimator  $\hat{\beta}'_0 = \bar{y} - \beta_1^0 \bar{x}$ . The preliminary test is for the hypothesis:

$$H_0 : \beta_1 = \beta_1^0 \text{ vs. } H_1 : \beta_1 \neq \beta_1^0.$$

Ayman Baklizi is a Professor in the Department of Mathematics in the Statistics and Physics College. Email: a.baklizi@qu.edu.qa. Adil E. Yousif is an Associate Professor in the Department of Mathematics in the Statistics and Physics College. Email: aeyousif@qu.edu.qa.

The null hypothesis is rejected for large values  $|T|$  where

$$T = \frac{S_{xx}^{1/2}(\hat{\beta}_1 - \beta_1^0)}{S_n}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

$$S_n^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

and

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2.$$

The preliminary test estimator PTE is defined as

$$\hat{\beta}_0^{PT} = \hat{\beta}_0 I(|T| > t_{\alpha/2, n-2}) + \hat{\beta}'_0 I(|T| \leq t_{\alpha/2, n-2})$$

where  $I(A)$  denotes an indicator function of the set  $A$ . The PTE is a convex combination of  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  and  $\hat{\beta}'_0 = \bar{y} - \beta_1^0 \bar{x}$  and depends on the random coefficient  $I(|T| \leq t_{\alpha/2, n-2})$  whose value is 1 when the null hypothesis is not rejected and is 0 otherwise. Thus, the PTE is an extreme compromise between  $\hat{\beta}_0$  and  $\hat{\beta}'_0$ . Moreover, the PTE does not allow a smooth transition between the two extremes. A possible remedy for this is to use an estimator with a continuous weight function. This function could be the P-value of the preliminary test.

The use of the P-value as a continuous weight function in preliminary test estimation was utilized by Baklizi (2004) and Baklizi and Abu-Dayyeh, W. (2003). If  $v = P\text{-value} = \Pr(|T| > t_{\alpha/2, n-2})$ , then a shrinkage estimator can be found as follows:

$$\hat{\beta}_0^{PV} = (1-v)\hat{\beta}_0 + v\hat{\beta}'_0.$$

Another possibility is given by Khan and Saleh (1997) who suggested the estimator

$$\hat{\beta}_0^{SH} = \hat{\beta}_0 + c^* \hat{\beta}_1 \bar{x} S_n / \sqrt{S_{xx}} |\hat{\beta}_1|$$

where  $c^*$  is the value that minimizes the mean squared error of  $\hat{\beta}_0^{SH}$ . This value is given by

$$c^* = \left( \frac{4}{\pi(n-2)} \right)^{1/2} \frac{\Gamma((n-1)/2)}{\Gamma((n-2)/2)}$$

and does not depend on the significance level of the preliminary test.

### The Confidence Intervals

Bootstrap intervals are computer intensive methods based on re-sampling with replacement from original data. Bootstrapping regression models can be constructed and run in several ways. The procedure adopted in this study was to resample with replacement from the pairs  $(x_i, y_i), i = 1, \dots, n$ . Several bootstrap based intervals are discussed in the literature.; the most common are the bootstrap- $t$  interval, the percentile interval and the bias corrected and accelerated (BCa) interval.

Let  $\tilde{\beta}_0$  be an estimator of  $\beta_0$  and let  $\tilde{\beta}_0^*$  be the estimator calculated from the bootstrap sample. Let  $z_\alpha^*$  be the  $\alpha$  quantile of the bootstrap distribution of  $Z^* = (\tilde{\beta}_0^* - \tilde{\beta}_0) / \hat{\eta}^*$ , where  $\hat{\eta}^*$  is the estimated standard deviation of  $\tilde{\beta}_0$  calculated from the bootstrap sample. The bootstrap- $t$  interval for  $\beta_0$  is given by  $(\tilde{\beta}_0 - z_{1-\alpha/2}^* \hat{\eta}^*, \tilde{\beta}_0 - z_{\alpha/2}^* \hat{\eta}^*)$  where  $z_\alpha^*$  is determined by simulation.

When a variance estimate of the estimator under consideration is unavailable or difficult to obtain, a modification of the bootstrap- $t$  interval is needed. Such a modification is based on using a further bootstrap sample from the original bootstrap sample to estimate the variance or the standard deviation  $sd^*(\tilde{\beta}_0^*)$  of  $\tilde{\beta}_0^*$ . The modified bootstrap- $t$  interval is thus given by;

$$(\tilde{\beta}_0 - z_{1-\alpha/2}^* sd^*(\tilde{\beta}_0^*), \tilde{\beta}_0 - z_{\alpha/2}^* sd^*(\tilde{\beta}_0^*))$$

The percentile interval may be described as follows, let  $\tilde{\beta}_0^*$  be an estimate of the intercept parameter calculated from the bootstrap sample. Here the bootstrap distribution of  $\tilde{\beta}_0^*$  is simulated by re-sampling repeatedly from the regression model based on the original data and calculating  $\tilde{\beta}_{0i}^*, i = 1, \dots, B$  where  $B$  is the number of bootstrap samples. If  $\hat{H}$  is the cumulative distribution function of  $\tilde{\beta}_0^*$ , then the  $1 - \alpha$  interval is given by

$$\left( \hat{H}^{-1}\left(\frac{\alpha}{2}\right), \hat{H}^{-1}\left(1 - \frac{\alpha}{2}\right) \right).$$

The bias corrected and accelerated interval is also calculated using the percentiles of the bootstrap distribution of  $\tilde{\beta}_0$ . The percentiles depend on two numbers,  $\hat{a}$  and  $\hat{z}_0$ , called the acceleration and the bias correction. The interval (BCa) is given by

$$\left( \hat{H}^{-1}(\alpha_1), \hat{H}^{-1}(\alpha_2) \right)$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right),$$

and

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right),$$

$\Phi(\cdot)$  is the standard normal cumulative distribution function,  $z_\alpha$  is the  $\alpha$  quantile of the standard normal distribution. The values of  $\hat{a}$  and  $\hat{z}_0$  are calculated as follows:

$$\hat{a} = \frac{\sum_{i=1}^n (\tilde{\beta}_0(\cdot) - \tilde{\beta}_0(i))^3}{6 \left\{ \sum_{i=1}^n (\tilde{\beta}_0(\cdot) - \tilde{\beta}_0(i))^2 \right\}^{3/2}}$$

where  $\tilde{\beta}_0(i)$  is the intercept estimator using the original data excluding the  $i^{\text{th}}$  pair and

$$\tilde{\beta}_0(\cdot) = \frac{\sum_{i=1}^n \tilde{\beta}_0(i)}{n}.$$

The value of  $\hat{z}_0$  is given by

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\tilde{\beta}_0^* < \tilde{\beta}_0\}}{B}\right).$$

### Methodology

#### Simulation Study

A simulation study was designed to evaluate the performance of the shrinkage estimators in terms of their biases and mean squared errors. Results for the preliminary test estimator are included for comparison purposes and the performance of the bootstrap intervals associated with the shrinkage estimators is also studied. The simulations used the sample sizes  $n = 15, 30$  and  $45$ . The slope parameter true value was chosen to be  $\beta_1 = 0, 1, 2, 3$  and  $4$ , the true value of the intercept parameter was set at  $\beta_0 = 0$ , and the guess value of the slope is set equal to zero in all cases. The predictor values are generated from the uniform distribution while the error terms are generated from  $N(0, \sigma^2)$  with  $\sigma^2 = 1$  or  $4$ .

For each combination of the simulation indices 1,000 pairs of  $(x, y)$  values were generated and the estimators were calculated. The level of the preliminary test is set to  $\alpha = 0.05$ . The biases and mean squared errors are calculated as:

$$bias(\tilde{\beta}_0) = \frac{1}{1000} \sum_{i=1}^{1000} (\tilde{\beta}_{0i} - \beta_0)$$

and

$$MSE(\tilde{\beta}_0) = \frac{1}{1000} \sum_{i=1}^{1000} (\tilde{\beta}_{0i} - \beta_0)^2,$$

where  $\beta_0$  is the true value of the intercept parameter and  $\tilde{\beta}_0$  is the shrinkage estimator under consideration.

The performance of the intervals is evaluated in terms of their coverage probabilities

(CP) and expected lengths (EL), which are calculated as follows: For the confidence interval  $CI$ ,

$$CP = \frac{1}{1000} \sum_{i=1}^{1000} I(\beta_0 \in (LB, UB))$$

and

$$EL = \frac{1}{1000} \sum_{i=1}^{1000} (UB - LB)$$

where  $LB$  and  $UB$  are the lower and upper bounds of the confidence interval. The nominal coverage probability of each interval is taken as 0.95%. The bootstrap calculations used 500 replications, and the second stage used 25 replications to estimate the variances of the estimators.

Results

The results for biases are shown in Table 1. It appears that  $\hat{\beta}_0^{PV}$  has the least bias among the

shrinkage estimators. The bias of all estimators increases as the initial guess moves further from the true value to a certain point and then decreases again, and the biases of all estimators decreases as the sample size increases. Results for the MSE performance are shown in Table 2; it appears that  $\hat{\beta}_0^{PV}$  also has the best overall performance among the shrinkage estimators. For the confidence intervals, it appears that intervals based on  $\hat{\beta}_0^{PV}$  perform better than intervals based on  $\hat{\beta}_0^{SH}$  in terms of the attainment of coverage probabilities (see Table 3). Results indicate that the BCa intervals and t-int intervals perform better than the PRC intervals among intervals based on  $\hat{\beta}_0^{PV}$ . Regarding interval widths, it appears that the t-int intervals are the shortest followed closely by the BCa intervals (see Table 4). The PRC intervals are very wide compared to the other intervals based on  $\hat{\beta}_0^{PV}$ .

Table 1: Biases of the Estimators

$\beta_1^0$	$\hat{\beta}_0$	$\hat{\beta}_0^{PT}$	$\hat{\beta}_0^{SH}$	$\hat{\beta}_0^{PV}$
$n = 15, \sigma = 1$				
0.0	-0.010	-0.002	-0.007	-0.009
1.0	-0.008	0.314	0.255	0.066
2.0	0.002	0.316	0.354	0.065
3.0	0.012	0.179	0.375	0.042
$n = 15, \sigma = 2$				
0.0	-0.008	-0.004	-0.021	-0.013
1.0	0.007	0.374	0.318	0.100
2.0	0.065	0.661	0.555	0.204
3.0	-0.009	0.715	0.639	0.142
$n = 30, \sigma = 1$				
0.0	-0.007	0.000	0.004	-0.003
1.0	-0.002	0.230	0.221	0.047
2.0	0.003	0.093	0.258	0.020
3.0	-0.004	0.001	0.252	-0.002
$n = 30, \sigma = 2$				
0.0	0.010	-0.004	-0.001	0.007
1.0	-0.019	0.311	0.278	0.062
2.0	-0.007	0.445	0.435	0.089
3.0	0.010	0.369	0.512	0.079



SIMPLE REGRESSION INTERCEPT INFERENCE USING PRIOR SLOPE INFORMATION

Table 2: MSEs of the Estimators

$\beta_1^0$	$\hat{\beta}_0$	$\hat{\beta}_0^{PT}$	$\hat{\beta}_0^{SH}$	$\hat{\beta}_0^{PV}$
$n = 15, \sigma = 1$				
0.0	0.297	0.127	0.154	0.224
1.0	0.296	0.371	0.273	0.281
2.0	0.316	0.695	0.425	0.355
3.0	0.290	0.627	0.438	0.322
$n = 15, \sigma = 2$				
0.0	1.169	0.485	0.606	0.872
1.0	1.200	0.789	0.801	0.979
2.0	1.238	1.467	1.154	1.164
3.0	1.194	2.211	1.406	1.267
$n = 30, \sigma = 1$				
0.0	0.138	0.063	0.073	0.105
1.0	0.142	0.259	0.165	0.152
2.0	0.139	0.269	0.207	0.153
3.0	0.143	0.153	0.208	0.145
$n = 30, \sigma = 2$				
0.0	0.564	0.250	0.300	0.434
1.0	0.566	0.529	0.441	0.504
2.0	0.580	1.049	0.662	0.620
3.0	0.546	1.239	0.786	0.621

Table 3: Coverage Probabilities of the Intervals

$\beta_1^0$	$\hat{\beta}_0^{SH}$			$\hat{\beta}_0^{PV}$		
	t-int	BCa	PRC	t-int	BCa	PRC
$n = 15, \sigma = 1$						
0.0	0.744	0.963	0.925	0.922	0.937	0.950
1.0	0.709	0.910	0.856	0.904	0.931	0.876
2.0	0.787	0.842	0.866	0.876	0.925	0.890
3.0	0.855	0.854	0.913	0.910	0.946	0.935
$n = 15, \sigma = 2$						
0.0	0.744	0.957	0.927	0.928	0.941	0.946
1.0	0.705	0.936	0.900	0.910	0.929	0.920
2.0	0.705	0.905	0.862	0.909	0.932	0.885
3.0	0.757	0.875	0.853	0.881	0.919	0.878
$n = 30, \sigma = 1$						
0.0	0.761	0.969	0.933	0.944	0.955	0.944
1.0	0.764	0.892	0.869	0.918	0.949	0.883
2.0	0.863	0.855	0.926	0.925	0.960	0.951
3.0	0.859	0.841	0.922	0.923	0.956	0.928
$n = 30, \sigma = 2$						
0.0	0.733	0.970	0.928	0.937	0.955	0.946
1.0	0.718	0.945	0.893	0.929	0.946	0.903
2.0	0.756	0.874	0.861	0.904	0.933	0.867
3.0	0.838	0.877	0.901	0.916	0.957	0.920

Table 4: Widths of the Bootstrap Intervals

$\beta_1^0$	$\hat{\beta}_0^{SH}$			$\hat{\beta}_0^{PV}$		
	t-int	BCa	PRC	t-int	BCa	PRC
$n = 15, \sigma = 1$						
0.0	1.031	1.744	2.680	2.005	1.849	2.591
1.0	1.452	1.780	2.762	2.035	2.075	2.867
2.0	1.884	2.028	2.962	2.117	2.298	3.213
3.0	2.048	2.173	3.068	2.155	2.424	3.236
$n = 15, \sigma = 2$						
0.0	2.070	3.481	5.228	3.979	3.688	5.016
1.0	2.392	3.431	5.399	4.000	3.876	5.347
2.0	2.813	3.560	5.529	4.056	4.123	5.699
3.0	3.461	3.784	5.698	4.115	4.366	6.141
$n = 30, \sigma = 1$						
-	0.734	1.210	1.586	1.364	1.271	1.504
1.0	1.148	1.301	1.733	1.430	1.517	1.853
2.0	1.410	1.478	1.829	1.490	1.678	1.909
3.0	1.455	1.359	1.796	1.474	1.721	1.786
$n = 30, \sigma = 2$						
0.0	1.356	2.429	3.169	2.733	2.526	3.001
1.0	1.832	2.419	3.272	2.772	2.782	3.258
2.0	2.286	2.604	3.441	2.855	3.031	3.654
3.0	2.659	2.838	3.596	2.943	3.236	3.871

Conclusion

In conclusion, it is recommended that  $\hat{\beta}_0^{PV}$  and the associated t- interval be employed for inference about the intercept parameter when there uncertain prior information exists regarding the slope.

References

Baklizi, A. (2004). Shrinkage estimation of the common location parameter of several exponentials. *Communications in Statistics, Simulation and Computation*, 33(2), 321-339.

Baklizi, A., & Abu-Dayyeh, W. (2003). Shrinkage estimation of P(X<Y) in the exponential case. *Communications in Statistics, Simulation and Computation*, 32(1), 31-42.

Bhoj, D. S., & Ahsanullah, M. (1993). Estimation of conditional mean in a linear regression model. *Biometrical Journal*, 35, 791-799.

Bhoj, D. S., & Ahsanullah, M. (1994). Estimation of a conditional mean in a regression model after a preliminary test on regression coefficient. *Biometrical Journal*, 36, 153-163.

Khan, S., & Saleh, A. K. Md. E. (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. *Biometrical Journal*, 29, 131-147.

## New Perspectives in Applying the Regression-Discontinuity Design for Program Evaluation: A Simulation Analysis

Sally A. Lesik  
Central Connecticut State University,  
New Britain, CT, USA

---

Evaluating educational programs is a core component of assessment. One challenge occurs because participants often enter into programs with diverse skills and backgrounds. The regression-discontinuity design has been used to evaluate programs amongst a diverse group, but noncompliance is a limitation. A simulation analysis illustrates the impact of noncompliance.

Key words: Education, program evaluation, Monte Carlo simulation, regression-discontinuity design, ordinary least squares regression, noncompliance.

---

### Introduction

Evaluating the effectiveness of educational programs can pose a challenge for researchers. One reason is because it is very difficult to isolate a program effect versus the effect attributed to differences between students who participate in such programs. Most evaluative studies compare different and diverse groups of students with respect to how well they perform by relying on observational versus experimental data. In order to estimate an actual program effect versus estimating differences between program participants, many studies have incorporated the regression-discontinuity design (Thistlethwaite & Campbell, 1960). The regression-discontinuity design allows for making causal inferences about program effects as the design has properties similar to a random experiment (Leake & Lesik, 2007; Lesik, 2006; Luyten, 2006; Moss & Yeaton, 2006).

The regression-discontinuity design has become a popular statistical tool for program evaluation because when a prescribed pre-treatment assignment measure is available, the regression-discontinuity design can emulate a

random experiment (Pettersson-Lidbom, 2003). Because the design resembles a random experiment, it can be used to estimate a program effect versus simply describing differences between different groups of individuals (Lesik, 2007).

Although the regression-discontinuity can be used to evaluate educational programs, the design has some major limitations that can make it difficult to apply in practice. One limitation of the design is that it assumes that the functional form of the model is specified correctly (Schumacker & Mount, 2007). Given the correct model specification, the regression-discontinuity design can provide an unbiased estimate of the program effect (Shadish, Cook & Campbell, 2002). However, if the functional form of the model is not known - and therefore not specified correctly - then any inferences made using a regression-discontinuity analysis may be biased and unreliable (Schumacker & Mount, 2007; Shadish, et al., 2002).

Another limitation is that the regression-discontinuity design requires approximately three times as many participants to achieve adequate power compared to a true random experiment (Cappelleri, Darlington & Trochim, 1994). Thus, a large sample is needed in order to have sufficient power to detect an effect if such an effect were to exist (Cappelleri, et al., 1994; Shadish, et al., 2002). Considering these two limitations together, it can be challenging for researchers in less-than-perfect situations to use

---

Sally A. Lesik is an Associate Professor in the Department of Mathematical Sciences. Her research interests are in applied statistical inference for educational research.  
Email her at: lesiks@ccsu.edu.

the regression-discontinuity design. However, one of the more serious limitations of the regression-discontinuity design is the issue of noncompliance. Noncompliance occurs if some of the participants do not adhere to their treatment assignment based on the score they received on the pre-treatment assignment measure.

The purpose of this study is to provide a simulation analysis using the regression-discontinuity design to address the issue of non-compliance within the framework of ordinary least squares regression. Using Monte Carlo techniques to simulate a regression-discontinuity analysis with different sample and effect sizes along with various degrees of noncompliance, this paper illustrates how noncompliance can bias the estimate of the treatment effect. A set of guidelines is developed that researchers can use to determine if the degree of noncompliance is severely biasing the estimate of the treatment parameter.

#### Theoretical Background

Many researchers have used the regression-discontinuity design to evaluate the effectiveness of educational programs (Jacob & Lefgren, 2002; Leake & Lesik, 2007; Lesik, 2006; Moss & Yeaton, 2006; Ross & Lacey, 1983). One of the core requirements for using the regression-discontinuity design is that a pre-treatment assignment measure with a defined cutoff score is used as the sole mechanism for assigning participants to either the treatment group or the control group. When given such a pre-treatment assignment measure, those individuals who score just below and just above a defined cutoff score will be similar to what a true random assignment would generate (Shadish, et al., 2002; Thistlethwaite & Campbell, 1960; Trochim, 1984; vanDerKlaauw, 2002).

If such an assignment measure is used, then an unbiased estimate of the program effect can be found by interpreting the estimated parameter ( $\hat{\beta}_1$ ) at the cutoff score that is associated with the dichotomous treatment variable (*TREATMENT*) in a baseline regression-discontinuity equation (Lesik, 2008; Shadish, et al., 2002). A baseline regression-

discontinuity equation requires that the assignment variable (*ASSIGNMENT*) and treatment indicator variable (*TREATMENT*) be included in a regression model as is illustrated in equation (1).

$$Y = \beta_0 + \beta_1 TREATMENT + \beta_2 ASSIGNMENT + \varepsilon \quad (1)$$

Although the baseline regression-discontinuity design involves a very simple model (there are only two predictor variables represented in the model), there are some major limitations of the regression-discontinuity design. One such limitation is that the design requires the correct model specification between the assignment and outcome measure (Shadish, et al., 2002). Another limitation is that a larger sample size is needed for the regression-discontinuity design as compared to a true random experiment (Cappelleri, et al., 1994). Furthermore, it is expected that participants will adhere to their assignment, and not enter the treatment group if they were assigned to the control group or vice-versa.

The regression-discontinuity design requires that the regression model under consideration has the correct functional form with respect to the relationship between the assignment variable and the outcome measure of interest. For example, if the relationship between the assignment and outcome variables is linear, then the baseline regression-discontinuity model as given in equation (1) would suffice because this equation describes a linear relationship. However, if there is a non-linear relationship between the assignment and outcome variables, then using equation (1) would generate a biased estimate of the treatment effect (Shadish, et al., 2002; Thistlethwaite & Campbell, 1960; Trochim, 1984). Thus, if a non-linear relationship represents the true functional form of the model, then higher order terms and their respective interactions would need to be added to the baseline regression-discontinuity model (Lesik, 2008; Schumacker & Mount, 2007; Shadish, et al., 2002; Trochim, 1984). Equation (2) is a regression-discontinuity model that includes higher-order terms and interactions in addition to the baseline regression-discontinuity model.

## REGRESSION-DISCONTINUITY DESIGN FOR PROGRAM EVALUATION

$$\begin{aligned} Y = & \beta_0 + \beta_1 TREATMENT + \beta_2 ASSIGNMENT \\ & + \beta_3 ASSIGNMENT^2 + \beta_4 TREATMENT \cdot ASSIGNMENT^2 \\ & + \beta_5 ASSIGNMENT^3 + \beta_6 TREATMENT \cdot ASSIGNMENT^3 \\ & + \dots + \beta_k ASSIGNMENT^j \\ & + \beta_{k+1} TREATMENT \cdot ASSIGNMENT^j + \varepsilon \end{aligned} \quad (2)$$

Many studies that have used the regression-discontinuity design have either employed exploratory techniques to empirically estimate the functional relationship between the assignment and outcome variables (Jacob & Lefgren, 2002; Lesik, 2006), or have added higher-order polynomial terms of the assignment variable along with possible interaction terms and then tested to determine if such terms are significant (Lesik, 2008; Moss & Yeaton, 2006). One concern with including higher-order terms and interactions in a regression-discontinuity design is that it can be difficult to estimate the correct functional form by visualization. Furthermore, adding extra terms to the regression-discontinuity model reduces the power of the design (Cappelleri & Trochim, 1994). Similarly, with the sample size limitations of the regression-discontinuity design, researchers often have to extrapolate more observations above and below the cutoff score in order to obtain a sufficiently large sample and then try to model the functional form appropriately (Lesik, 2006; Moss & Yeaton, 2006; Shadish, et al., 2002).

Clearly, a trade-off exists between the number of observations collected around the cutoff score and the functional form specification of the regression-discontinuity model. Collecting too few observations around the cutoff score would require including more observations that are further away from the cutoff score, thus making it more difficult to describe the appropriate functional form. Collecting a small number of observations that lie only within a very narrow range around the cutoff score may not provide enough power to detect a reasonably sized effect.

Although the functional form and sample size limitations of the regression-discontinuity design can be difficult to address in practice, one of the most serious limitations of

the design occurs when participants do not adhere to their assignment that is based only on the score received on the assignment measure. Typically, in order to address such non-compliance, researchers have modeled selection effects by using the probability of actually participating in the treatment program as an instrumental variable for the program assignment (Lesik, 2006).

The technique of instrumental variable estimation measures the effect of the treatment rather than just the assignment to the treatment group (Angrist & Krueger, 1991; Jacob & Lefgren, 2002). However, most research using instrumental variables estimation has the first stage probability of actually participating in the program described by a linear probability model (Lesik, 2006; 2007), and this introduces some of the limitations of the linear probability model, such as probabilities greater than one or less than zero (Wooldridge, 2003). Nonlinear first-stage instrumental variables estimation tends not to be used in practice because their behavior is generally not well known (Wooldridge, 2002). Another concern with using instrumental variables estimation is that this technique on its own relies on some very strong assumptions which can be difficult to assess in practice (Wooldridge, 2002, 2003).

A simpler strategy to assess the effect of non-compliance in a regression-discontinuity analysis is to determine how much of an effect keeping those observations that did not comply with their assignment in the model. A basic sensitivity analysis can be used to determine if the estimate of the treatment effect is different when including and excluding non-compliers in the model (Leake & Lesik, 2007). However, there remains the need to establish some general guidelines that can be used to assess whether the amount of non-compliance could be introducing bias in the estimate of the treatment effect if these observations are included in the analysis, as well as to determine what happens to the estimate of the treatment effect when non-compliers are removed from the analysis.

### Methodology

In order to address the impact of non-compliance on the estimate of the treatment effect as obtained through a regression-

discontinuity design, the results of a simulation analysis are presented within the framework of an ordinary least squares regression. The simulation analysis was executed using Version 10 of STATA® (STATA Corporation, 2007).

#### Simulation Analysis

A simulation analysis was performed to investigate how participant noncompliance can impact the estimate of the treatment effect in a regression-discontinuity design within the framework of ordinary least squares regression. The challenge with running a simulation analysis for a regression-discontinuity design revolves around simulating realistic regression-discontinuity data. Figure 1 illustrates a hypothetical regression-discontinuity design (Shadish, et al., 2002). A perfect regression-discontinuity indicates 100% compliance.

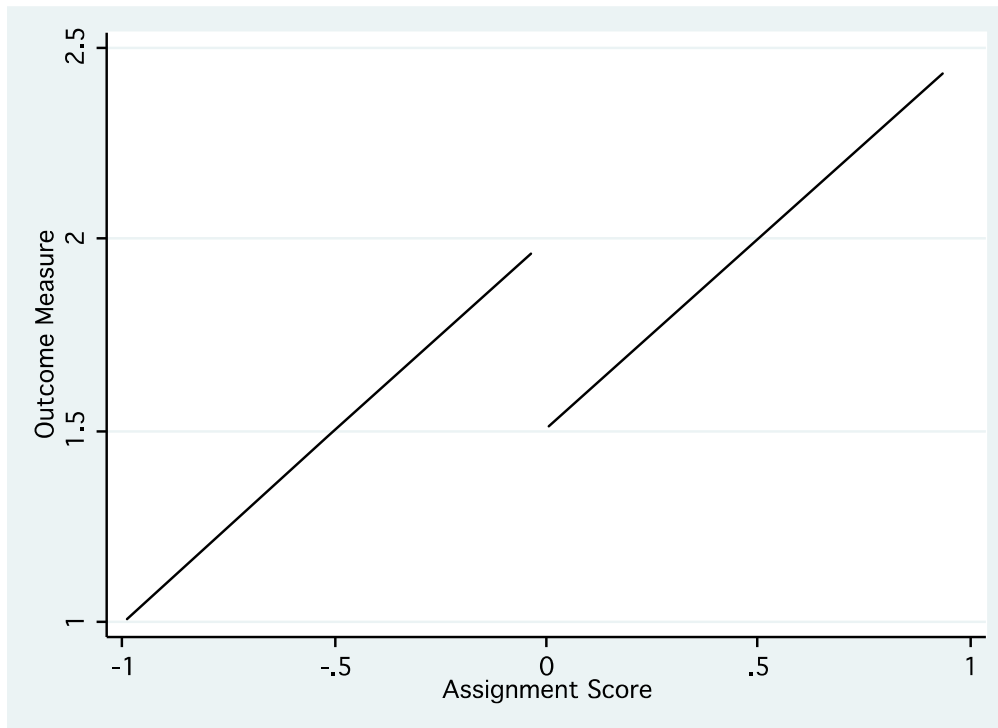
Perfect compliance is when those participants who were assigned to the treatment group actually did participate in the treatment, and those who were assigned to the control group did not participate in the treatment at all.

Notice that the discontinuity in Figure 1 represents a simulated treatment effect of  $-0.50$  at the cutoff score of 0.

In order to generate realistic regression-discontinuity data, a random component was introduced (see Figure 2). Figure 2 shows the scatter plot along with the lines of best fit and a lowess smoother estimating the relationship between the assignment variable and the outcome variable. The random component was fixed at one-half of the effect size to ensure that a linear model would be appropriate.

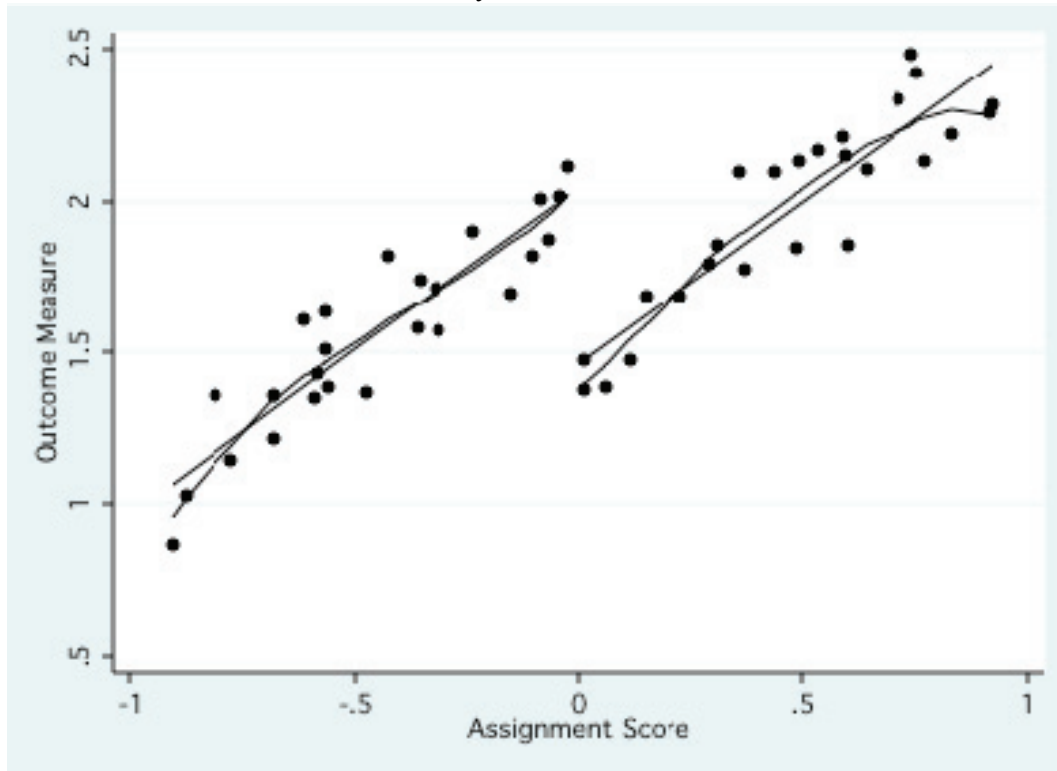
Appendix 1 contains an example of how Stata code can be written to simulate the regression-discontinuity data presented in Figure 2. This Stata code randomly generates 50 observations on the interval  $(-1, 1)$ , which centers about the cutoff score of 0. Observations that are greater than or equal to zero are assigned the value 0 to represent that they are assigned to the control group, whereas all other observations that fall below the cutoff score of 0 are assigned to the treatment program and given the value 1.

Figure 1: Simulated Regression-Discontinuity Data with an Effect Size of  $-0.50$  At the Centered Cutoff Score of 0



## REGRESSION-DISCONTINUITY DESIGN FOR PROGRAM EVALUATION

Figure 2: Regression-Discontinuity Data with Added Random Component of 0.25 and a  $-0.5$  Discontinuity at the Centered Cutoff Score of 0



This example illustrates a random component that is generated between  $-0.25$  and  $0.25$  from the original data: In other words, each observation is randomly assigned to be within  $\pm 0.25$  of its original value. The simulated treatment effect of  $0.50$  is generated for those observations that are assigned to the treatment group so around the cutoff score of  $0$ , the treatment effect remains at  $0.50$ . The choice of considering a small interval around the cutoff score and of having the random component equal to half of the effect size was introduced to generate what realistic regression-discontinuity data would look like and to ensure a linear model specification is appropriate.

In order to generate non-compliers, a random number on the interval  $[0,1)$  was generated and, for each observation, this random number was compared to the given percent of non-compliance. If the observation was designed as a crossover because the random number was

less than the given percentage of non-compliance, then the treatment indicator was changed for these observations.

### Results

A series of Monte Carlo simulations were run to investigate the impact of sample size, non-compliance and the effect size on the estimate of the treatment effect when non-compliers were kept in the analysis, when there was perfect compliance, and when non-compliers were deleted from the analysis. A total of  $10,000$  simulations were run for each of the different combinations of effect sizes, sample sizes, and percentages of non-compliance. Table 2 shows the estimate of the treatment effect and standard error for including non-compliers in the analysis, with perfect compliance and for deleting non-compliers from the analysis.

The results of the simulation analysis indicate that for any sample size, any effect size and any degree of non-compliance, keeping non-

compliers in the analysis will generate a biased estimate of the treatment effect. Also interesting is that larger sample sizes show greater bias as compared to smaller samples even for the same percent of non-compliance. In addition, as the degree of non-compliance increases, the estimate of the treatment effect becomes more biased. Furthermore, for any degree of non-compliance less than 10%, running the regression-discontinuity analysis without the non-compliers generates similar estimates of the treatment effect than was found assuming perfect compliance. This is consistent with Judd and Kenny (1981) and Trochim (1984), who suggest that excluding no more than 5% of the non-compliers should provide reasonable estimates of the treatment effect.

This simulation analysis illustrates that - if non-compliance is random - then excluding no more than 10% of the observations provides a reasonable estimate of the treatment effect. It is also interesting to note that analyses with greater effect sizes show a similar bias compared to smaller effect sizes. For example, for a sample of 500 with an effect size of 0.500, if 10% of the observations are non-compliers and if they are kept in the analysis, then the estimated effect size becomes 0.193, a reduction of approximately 60%. For a sample of 500 with an effect size of 2.000, if 10% of the observations are non-compliers and they are kept in the analysis, then the estimated effect size becomes 0.775, again a reduction of approximately 60%. Similarly for a sample size of 20 with 1% of the non-compliers kept in the analysis, then the estimated effect size becomes 0.458 for an effect size of 0.500, a reduction of approximately 8%, and for a sample of size 20 with 1% of the non-compliers kept in the analysis, then the estimated effect size becomes 1.834 for an effect size of 2.000, again a reduction of approximately 8%.

#### Conclusion

Although this study contributes to the literature by providing some guidelines for dealing with the noncompliance limitation of the regression-discontinuity design, this study did not address other issues such as attrition or numerous other threats to validity that are inherent with virtually every type of analyses (Shadish, et al., 2002).

Furthermore, because the range of data collected around the cutoff score was fixed to only include those observations within one point of the centered cutoff score of 0 the issue of correct functional form for this study was not relevant. However, as is often the case in practice, more observations need to be collected around a greater range of the cutoff score, and thus functional form specification becomes more of a concern. In this simulation, the functional form was forced to be approximately linear in order to avoid trying to model functional form and non-compliance together.

Results of the simulation analysis suggest that if non-compliance is essentially a random phenomenon, then removing the non-compliers from the analysis does not appear to bias the estimate of the treatment effect if the percentage of non-compliers is relatively small, such as less than 10%. However, if non-compliance is not random, then this may not be the case. For example, if only those participants who were on one side of the cutoff score chose not to comply with their assignment, then deleting them from the analysis will likely produce a biased estimate of the treatment effect. In cases where non-compliance is not random, instrumental variables estimation may be a better strategy to use, even given the relatively strong assumptions of instrumental variables estimation.

#### References

- Angrist, J., & Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106, 979-1014.
- Cappelleri, J., Darlington, R., & Trochim, W. (1994). Power analysis of cutoff-based randomized clinical trials. *Evaluation Review*, 18, 141-152.
- Cappelleri, J., & Trochim, W. (1994). An illustrative statistical analysis of cutoff-based randomized clinical trials. *Journal of Clinical Epidemiology*, 47, 261-270.
- Jacob, B., & Lefgren, L. (2002). *Remedial education and student achievement: A regression-discontinuity analysis* (Working paper 8918). Cambridge, MA: National Bureau of Economic Research.



## REGRESSION-DISCONTINUITY DESIGN FOR PROGRAM EVALUATION

Judd, C., & Kenny, D. (1981). *Estimating the Effects of Social Interventions*. London, England: Cambridge University Press.

Leake, M., & Lesik, S. (2007). Do remedial English programs impact first-year success in college? An illustration of the regression-discontinuity design. *International Journal of Research and Method in Education*, 30(1), 89-99.

Lesik, S. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *The Review of Higher Education*, 30(1), 1-19.

Lesik, S. (2007). Do developmental mathematics programs have a causal impact on student retention? An application of discrete-time survival and regression-discontinuity analysis. *Research in Higher Education*, 48(5), 583-608.

Lesik, S. (2008). Studying the Effectiveness of Programs and Initiatives in Higher Education Using the Regression-Discontinuity Design. In J. C. Smart (Ed.), *Higher Education Handbook of Theory and Research, Volume 23*. The Netherlands: Springer Business Media.

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. *Oxford Review of Education*, 32(3), 397-429.

Moss, B., & Yeaton, W. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, 28(3), 215-229.

Pettersson-Lidbom, P. (2003). *Do parties matter for fiscal policy choices? A regression-discontinuity approach*. Stockholm, Sweden: Stockholm University.

Ross, A., & Lacey, B. (1983). A regression discontinuity analysis of a remedial education programme. *Canadian Journal of Higher Education*, 13(1), 1-15.

Schumacker, R., & Mount, R. (2007). *Regression discontinuity: Examining model misspecification*. Paper presented at the Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

STATA Corporation. (2007). *STATA Base Reference Manual (Vol. 1)*. College Station, TX: STATA Press.

Thistlethwaite, D., & Campbell, D. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51, 309-317.

Trochim, W. (1984). *Research Design for Program Evaluation: The Regression -Discontinuity Approach*. Newbury Park, CA: Sage.

vanDerKlaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review*, 43(4), 1249-1287.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Wooldridge, J. (2003). *Introductory Econometrics: A Modern Approach (2<sup>nd</sup> Ed.)*. Mason, Ohio: South-Western.

### Appendix 1: Example of Stata Code for Simulation Analysis

```
program rdgen, rclass
version 10.1
drop _all
set obs 50
generate x = -1 + (1-(-1))*uniform()
generate treat = 1
replace treat = 0 if x >= 0
generate y = x + 2
replace y = (y - 0.05) if treat == 0
generate z = -0.25 + (0.25-(-0.25))*uniform()
generate crossover = 0
generate w = uniform()
generate treatcr = treat
replace treatcr = treat + 1 if w < 0.05
replace crossover = 1 if w < 0.05
replace treatcr = 0 if treatcr == 2
replace y = z+y
regress y treatcr x
return scalar B1cr = _b[treatcr]
return scalar seB1cr = _se[treatcr]
regress y treat x
return scalar B1 = _b[treat]
return scalar seB1 = _se[treat]
drop if crossover == 1
regress y treat x
return scalar B1nocr = _b[treat]
return scalar seB1nocr = _se[treat]
end
```

SALLY A. LESIK

Table 2: Estimate of the Treatment Effect and Standard Error (in [ ]) for Simulated Regression-Discontinuity Data Including Non-Compliers (NC), with Perfect Compliance and Without Non-Compliers for Various Percentages of Non-Compliance (% NC)

% NC	Sample Size	With NC	Perfect	Without NC	With NC	Perfect	Without NC	With NC	Perfect	Without NC	With NC	Perfect	Without NC
	Effect Size	0.500			1.000			1.500			2.000		
1%	20	0.458 [0.135]	0.500 [0.137]	0.500 [0.138]	0.917 [0.270]	1.003 [0.273]	1.002 [0.275]	1.371 [0.405]	1.498 [0.410]	1.498 [0.412]	1.834 [0.541]	1.997 [0.548]	1.996 [0.551]
	30	0.453 [0.108]	0.500 [0.110]	0.500 [0.110]	0.901 [0.216]	0.996 [0.219]	0.996 [0.220]	1.363 [0.324]	1.504 [0.329]	1.504 [0.330]	1.807 [0.432]	1.995 [0.439]	1.995 [0.441]
	40	0.449 [0.093]	0.501 [0.094]	0.500 [0.094]	0.900 [0.186]	1.001 [0.187]	1.001 [0.188]	1.353 [0.278]	1.503 [0.281]	1.503 [0.283]	1.800 [0.371]	2.001 [0.375]	2.001 [0.377]
	50	0.450 [0.083]	0.501 [0.084]	0.501 [0.084]	0.894 [0.165]	0.998 [0.167]	0.998 [0.168]	1.346 [0.248]	1.500 [0.250]	1.499 [0.251]	1.796 [0.330]	2.003 [0.333]	2.002 [0.335]
	100	0.443 [0.058]	0.500 [0.058]	0.500 [0.059]	0.885 [0.116]	0.997 [0.117]	0.997 [0.117]	1.328 [0.174]	1.499 [0.175]	1.499 [0.176]	1.775 [0.231]	2.002 [0.233]	2.003 [0.234]
	200	0.441 [0.041]	0.501 [0.041]	0.500 [0.041]	0.879 [0.082]	1.000 [0.082]	0.999 [0.082]	1.323 [0.122]	1.500 [0.123]	1.501 [0.124]	1.765 [0.163]	2.001 [0.164]	2.001 [0.165]
	300	0.440 [0.033]	0.500 [0.033]	0.500 [0.034]	0.880 [0.066]	1.000 [0.067]	1.000 [0.067]	1.321 [0.100]	1.502 [0.100]	1.502 [0.101]	1.760 [0.133]	1.998 [0.134]	1.998 [0.134]
	500	0.439 [0.026]	0.500 [0.026]	0.500 [0.026]	0.880 [0.051]	1.000 [0.052]	1.000 [0.052]	1.317 [0.077]	1.501 [0.078]	1.501 [0.078]	1.756 [0.103]	2.000 [0.104]	2.000 [0.104]
5%	20	0.327 [0.127]	0.496 [0.137]	0.496 [0.143]	0.658 [0.255]	1.002 [0.274]	1.004 [0.282]	0.988 [0.382]	1.498 [0.411]	1.499 [0.423]	1.335 [0.511]	2.010 [0.548]	2.010 [0.564]
	30	0.315 [0.102]	0.501 [0.109]	0.500 [0.113]	0.633 [0.204]	1.002 [0.219]	1.002 [0.226]	0.948 [0.305]	1.502 [0.328]	1.502 [0.338]	1.273 [0.407]	2.007 [0.436]	2.008 [0.449]
	40	0.310 [0.087]	0.500 [0.094]	0.500 [0.097]	0.617 [0.175]	1.002 [0.187]	1.002 [0.193]	0.931 [0.262]	1.502 [0.281]	1.501 [0.289]	1.235 [0.350]	1.998 [0.375]	1.999 [0.386]
	50	0.304 [0.079]	0.500 [0.084]	0.501 [0.086]	0.608 [0.156]	1.002 [0.167]	1.002 [0.172]	0.911 [0.233]	1.498 [0.250]	1.498 [0.257]	1.216 [0.311]	2.001 [0.334]	2.001 [0.343]
	100	0.295 [0.054]	0.500 [0.058]	0.500 [0.060]	0.588 [0.109]	1.000 [0.117]	1.000 [0.120]	0.887 [0.163]	1.501 [0.175]	1.502 [0.180]	1.184 [0.218]	2.003 [0.233]	2.003 [0.240]
	200	0.291 [0.038]	0.500 [0.041]	0.500 [0.042]	0.580 [0.077]	0.990 [0.082]	0.999 [0.084]	0.876 [0.115]	1.500 [0.123]	1.500 [0.126]	1.164 [0.153]	1.999 [0.164]	1.999 [0.169]
	300	0.290 [0.031]	0.500 [0.033]	0.500 [0.034]	0.580 [0.062]	1.000 [0.067]	1.000 [0.069]	0.868 [0.094]	1.499 [0.100]	1.499 [0.103]	1.158 [0.125]	2.001 [0.134]	2.002 [0.137]
	500	0.288 [0.024]	0.500 [0.026]	0.500 [0.027]	0.577 [0.048]	1.000 [0.052]	1.000 [0.053]	0.865 [0.072]	1.500 [0.078]	1.500 [0.080]	1.152 [0.096]	1.999 [0.104]	1.999 [0.106]

## REGRESSION-DISCONTINUITY DESIGN FOR PROGRAM EVALUATION

Table 2 (continued): Estimate of the Treatment Effect and Standard Error (in [ ]) for Simulated Regression-Discontinuity Data Including Non-Compliers (NC), with Perfect Compliance and Without Non-Compliers for Various Percentages of Non-Compliance (% NC)

% NC	Sample Size	With NC	Perfect	Without NC	With NC	Perfect	Without NC	With NC	Perfect	Without NC	With NC	Perfect	Without NC
	Effect Size	0.500			1.000			1.500			2.000		
10%	20	0.229 [0.119]	0.502 [0.137]	0.503 [0.146]	0.450 [0.236]	1.000 [0.274]	1.001 [0.292]	0.671 [0.354]	1.504 [0.410]	1.503 [0.436]	0.903 [0.474]	2.002 [0.548]	2.005 [0.584]
	30	0.214 [0.094]	0.500 [0.110]	0.500 [0.116]	0.429 [0.189]	1.002 [0.218]	1.001 [0.232]	0.645 [0.283]	1.505 [0.328]	1.504 [0.348]	0.853 [0.376]	1.999 [0.437]	2.001 [0.464]
	40	0.206 [0.081]	0.500 [0.094]	0.500 [0.099]	0.416 [0.161]	1.002 [0.187]	1.002 [0.198]	0.619 [0.242]	1.495 [0.281]	1.497 [0.297]	0.830 [0.323]	1.995 [0.375]	1.998 [0.397]
	50	0.205 [0.072]	0.500 [0.083]	0.500 [0.088]	0.405 [0.143]	0.996 [0.167]	0.997 [0.177]	0.610 [0.215]	1.496 [0.251]	1.496 [0.265]	0.817 [0.287]	1.999 [0.333]	2.000 [0.353]
	100	0.197 [0.050]	0.500 [0.058]	0.501 [0.062]	0.395 [0.100]	1.002 [0.117]	1.002 [0.123]	0.595 [0.150]	1.502 [0.175]	1.503 [0.185]	0.791 [0.200]	2.002 [0.233]	2.002 [0.246]
	200	0.195 [0.035]	0.500 [0.041]	0.500 [0.043]	0.388 [0.070]	1.000 [0.082]	1.000 [0.087]	0.582 [0.105]	1.498 [0.123]	1.498 [0.130]	0.781 [0.141]	2.001 [0.164]	2.001 [0.173]
	300	0.194 [0.029]	0.500 [0.033]	0.500 [0.035]	0.389 [0.057]	1.000 [0.067]	1.000 [0.071]	0.582 [0.086]	1.498 [0.100]	1.498 [0.106]	0.773 [0.114]	1.998 [0.134]	1.998 [0.141]
	500	0.193 [0.022]	0.500 [0.026]	0.500 [0.027]	0.387 [0.044]	1.000 [0.052]	1.000 [0.055]	0.579 [0.066]	1.499 [0.078]	1.499 [0.082]	0.775 [0.088]	2.000 [0.103]	2.001 [0.109]

## Information Technology for Increasing Qualitative Information Processing Efficiency

S. N. Martyshenko                      E. A. Egorov  
Vladivostok State University of Economics and Services  
Vladivostok, Russia

---

The problem of qualitative information processing in questionnaires is considered and a solution for this problem is offered. The computer technology developed by the authors to automate the offered decision is described.

Key words: Open question, typology consumers, qualitative information processing technology.

---

### Introduction

Questionnaire survey methods are becoming more widespread. This is facilitated in part by the development of democratic processes and a market economy in Russia and other countries. Moreover, institutions of power have shown increased interest in establishing public feedback using the Internet.

Most researchers use simple methods for questionnaire data collection and processing. However, qualitative and poorly structured information processing techniques are common. Martyshenko and Kustov have been engaged in questionnaire data processing computer technology development for more than six years. In their previous work (Martyshenko & Kustov, 2007) they envisaged a computer technology for open questionnaire processing. The technology is based on a type-design practice.

Researchers have achieved considerable results in qualitative information processing technology. Some new elements of the technology are considered here. Qualitative information processing efficiency has been improved through the introduction of a new

block, namely, knowledge base represented by three types of dictionaries.

### Qualitative Information Processing

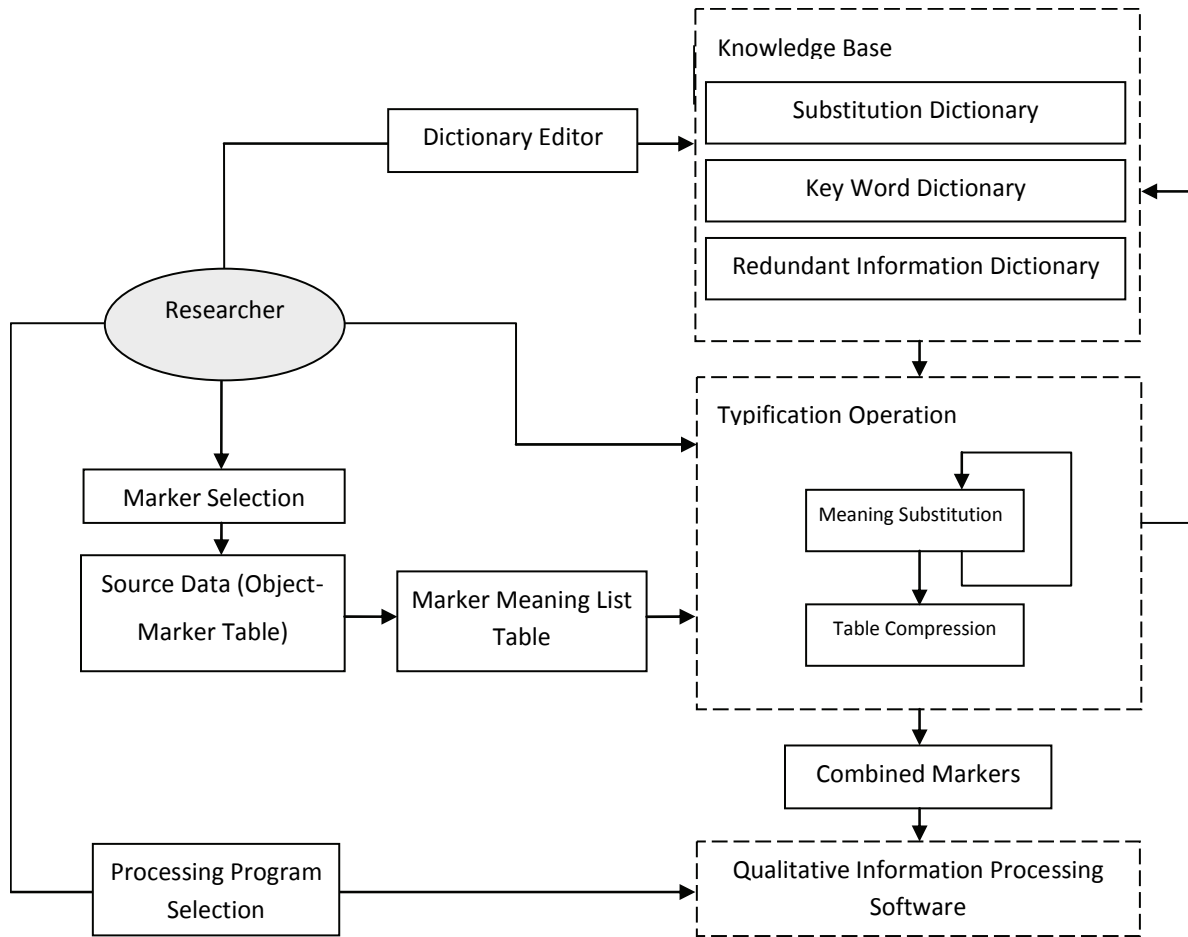
Consider a general scheme of qualitative information processing technology (see Figure 1). Source data is considered to be represented in an object-marker schedule, regardless of the collection technique used. Besides common properties with only one meaning of a marker under investigation, it is assumed that combined properties can be included into the object-marker table. A combined marker appears when a respondent can choose several answers for the same question. For example, a respondent may mention several cities when asked: *What large cities have you visited within the last three years?* Thus, a combined marker consists of at least more than one simple answer. To identify a combined marker it is necessary to introduce a single divider symbol because a simple answer may consist of several words or even an entire sentence.

Typification, which allows for the quantitative processing, is a practice used for the shift from non-structured data representation to structured representation. A typification operation is a substitution of a source simple statement (in the form of text) by a generalizing statement (in the form of text) with similar or close meaning. The typification operation is performed with the assistance of a tabulated marker meaning list, that is, one of the columns in the table contains all the unique meanings of the source marker.

---

S. N. Martyshenko is a Professor and Department Chair in Mathematics and Modeling. Email: sergey.martishenko@vvsu.ru. E. A. Egorov is a post-graduate student in Mathematics and modelling. Email: egorov.e.a@gmail.com.

Figure 1: Qualitative Data Processing Computer Technology Scheme



Although a combined marker is subject to a typification operation, all simple statements in a complex or combined statement are included into the marker meaning list. A marker meaning list table has an automatic filter and contains a column in which meaning frequency is calculated; a typification operation is used for marker meaning list table data rather than object-marker table data. In a typification operation, simple situations are processed first. For example, different spellings of the same word or word order. The best (or correct) form of the statement is selected from among several similar statements and is then copied into marker meaning list table cells containing similar statements. By substituting some unique statements for those existing in the meaning list the number of lines in the marker meaning list

table are reduced. After several substitutions have been made, a compression operation can be reasonably used to recalculate the marker meaning list table. Gradually, using this process, the marker meaning list table shrinks to become more demonstrative.

After simple situations have been processed, more complex statements are considered. A group of low-frequency statements related to the same topic should be found in a marker meaning list table. For this group of simple statements a researcher can select a generalizing statement or, if no statement can be found, a new generalizing statement may be introduced to render the general meaning or topic of a group of simple statements. For example, in response to the question, *what else do you like to do during a*

*vacation at the seaside besides sunbathing and swimming?*, respondents mentioned *raising children, raising grandchildren, playing with children and/or teaching children to swim*. However, these statements did not account for a high percentage of response (less than 0.1%), thus, the generalizing statement *taking care of children* found in the marker meaning list table was substituted. In general, this substitution preserved the meaning of the original statements.

However, it is important not to lose information, particularly when a repeated questionnaire survey is used, for this reason, similar statements for the generalizing ones using a specification in parenthesis are substituted. Specification or nuance is indicated in brackets. For example, in the above case the original meanings were substituted for:

- taking care of children (raising grandchildren),
- taking care of children (raising children),
- taking care of children (playing with grandchildren), and
- taking care of children (teaching children to swim).

The nature of the response, which helps to determine the respondent's (consumer's) personality, is more important than the word-by-word content of the answer. An initial marker meaning list table can contain several thousand meanings, but after processing (typification operation) the table normally has up to three hundred meanings, including specifications. Creation of the table is the final stage in typification (level one). Even if automated, the process is time-consuming and requires the skills and concentration of a researcher. After each session the results are stored and the next session starts from the point where the previous session was stopped.

If a newly designed marker contains two or more meanings it is subject to analysis and requires additional processing (level two). At this stage the specification is excluded and another column, called a subclass, is placed in the marker meaning list table with the number of unique statements reduced to 30 or 50.

Thirty to fifty variants of meaning is a large number for analysis on a nominal scale. Therefore, upon formation of an acceptable list of clearly different meaning variants, a researcher must group the answers in order to consider their properties of separate classes, types or topics depending on the informative meaning of the markers and tasks set developed for the typification operation. In our example determination of personality type would suit better. Combining simple statements into classes is a third level of typification practice. A researcher will introduce the name for each class in accordance with the nature of the statements to be combined. In practice, grouping results achieved by different researchers are very similar. Any differences may be explained by the transitory nature of some statements, which can be attributed to several classes and different researchers can give different names to each class.

Thus, as a result of open question data processing the following output is obtained:

- Three new representations of marker (property) included into the source data table, which can be subject to subsequent processing for obtaining informative conclusions;
- A marker meaning list table which can be used for a repeated questionnaire survey or typification practice with any other questionnaires developed for the process investigation; and
- Knowledge base in the form of three dictionaries: a substitution dictionary, a key word dictionary and a redundant information dictionary.

#### Dictionaries

It should be noted that, as a result of a typification operation with combined markers, other combined markers are formed. Special processing techniques have been developed for analysis of the latter. The efficiency of qualitative data processing computer technology can be increased by knowledge base creation and use. Computer technologies allowing for knowledge base use belong to an expert system class. The main distinctive feature of an expert system is that it is capable of correct forecasting.

# INCREASING QUALITATIVE INFORMATION PROCESSING EFFICIENCY

By giving various hints to a user during operation, specialized software can save much time. User hints are generated by special dictionaries, which are formed during typification operations. These dictionaries store user experiences gained through solving qualitative marker typification tasks. Consider the structure and functions of these dictionaries.

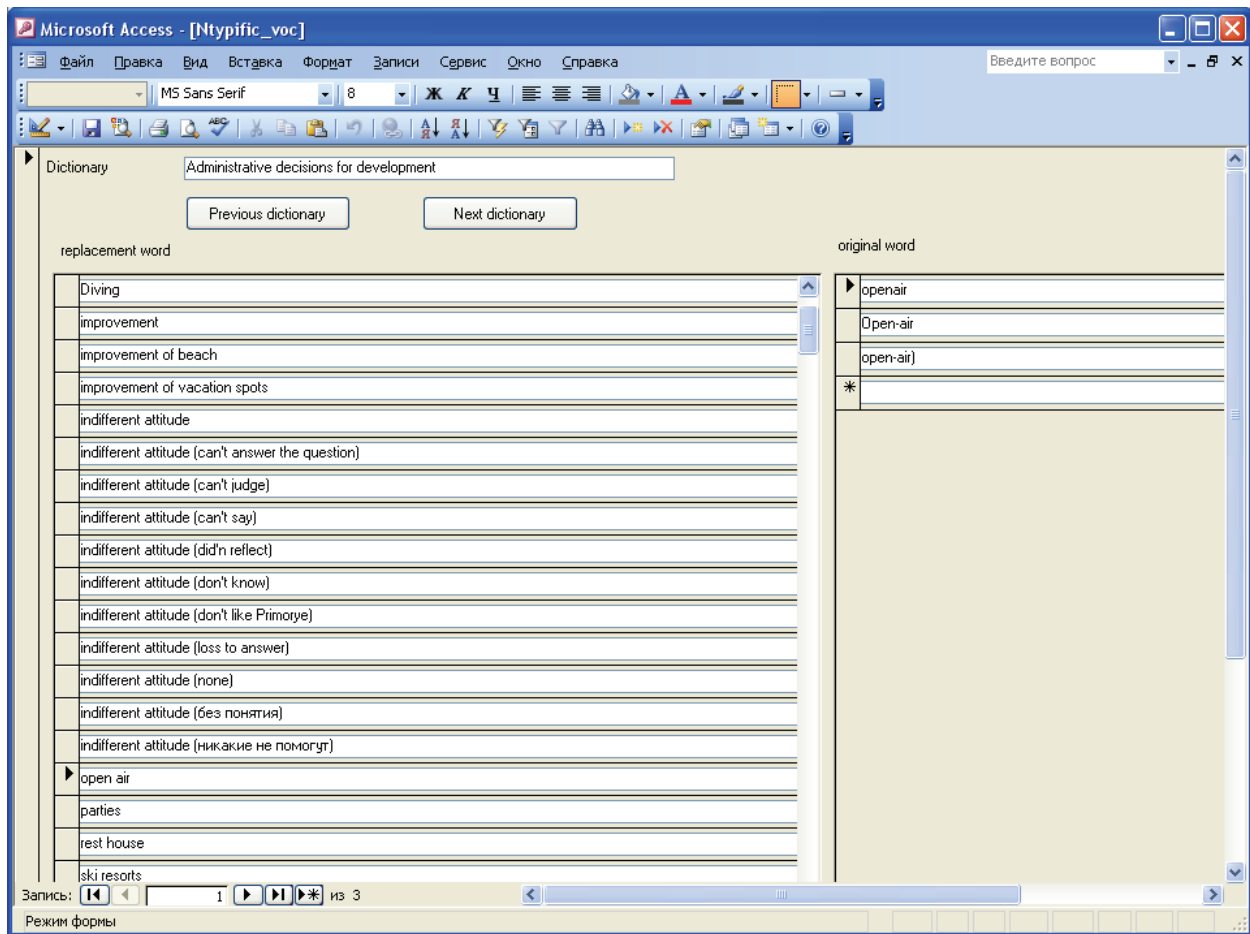
## Substitution Dictionary

A substitution dictionary is formed automatically when simple statements are substituted for by other simple statements. The substitution dictionary is supplemented when a user works with a typification program; the dictionary stores all substitutions made by the user. All dictionaries are stored in a single Access data base and new questionnaire data

requires new a typification operation. This means that a user must choose analogues for the new data. Situations processed by users at previous stages prove to repeat in most cases for the repeated data collection. By activating the substitution dictionary the user can acquire hints for substitution operations; this leaves only situations that have not been recorded before to be processed.

Dictionary support does not require an excess of time, but the dictionary should be reviewed occasionally because, over time, irrelevant variants of substitution collect. Any irrelevant records which accumulate in the dictionary should be deleted because excessive volume of the dictionary can reduce software operation speed. An example of a substitution dictionary is shown in Figure 2.

Figure 2: Substitution Dictionary Example



Key Word Dictionary

Next consider the functions of a key word dictionary. This dictionary is very useful for phrases containing many words. With long phrases the search for proper synonyms in the marker meaning list table can be difficult due to the large number of meanings in the beginning of operation. Because exact matching is a rare occasion, EXCEL filters are often used to identify shorter lists containing certain combinations of phrase elements (see Figure 3). Such phrases are termed key words, although phrase elements can be conditional key words only. The list shown in Figure 3 contains phrases identified by the key word exc. In the reduced list the phrase, *organize tours to preserved areas*, can be substituted for the phrase, *develop a preserved area excursion tour*. The substitution does not alter the meaning of the phrase.

Dictionary operation is based on the extended filter principle, and a key word dictionary includes high-frequency phrases from a marker meaning list table. These phrases are

provided with a list of key words for a similar phrase search. Users can perform these operations with the help of commands in a special dialog window titled show key word dictionaries (see Figure 4), which represents all functions for key word dictionary creation, supplement and review. Figures 5 and 6 show dialog windows used for supplementing key word dictionaries.

The difference between the dictionary and the extended filter operation is that the dictionary stores key words previously used for a synonym search. The dictionary can be created and supplemented by user only and is intended for future operations facilitation. Once the dictionary contains enough data it can be used as a database. When the dictionary is activated before a typification program is started an additional column is formed in the marker meaning list table, this is called a substitutions from key word dictionary. This column offers phrase substitutions from a key word dictionary that contains phrases chosen by maximum match between source phrase key words and key word dictionary phrases.

Figure 3: Example of the EXCEL User Filter Application for a Typification Operation

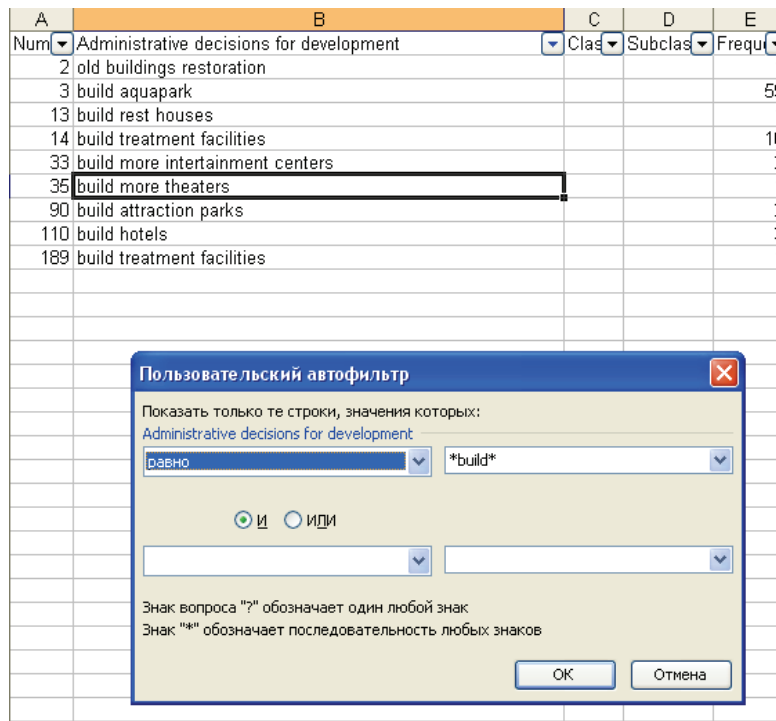




Figure 4: Toolbox for Operating a Key Word Dictionary

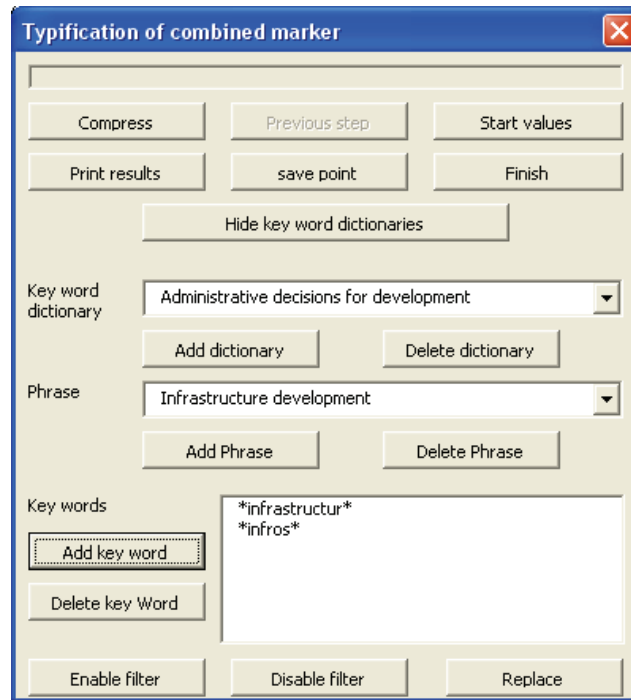
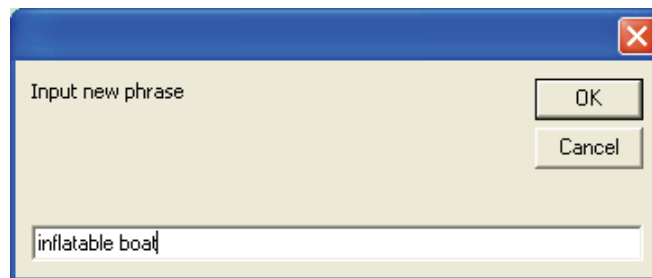


Figure 5: New Phrase Input for a Key Word Dictionary



#### Redundant Information Dictionary

The third dictionary in the database is the redundant information dictionary. Unlike the two previously mentioned dictionaries developed for each qualitative marker, the redundant information dictionary works with all the qualitative markers and also with different questionnaires. This dictionary is used at qualitative text information processing level one and it helps to delete or edit any statements containing redundant and/or irrelevant information.

For example, the dictionary can be used to exclude such phrases as *I think that, in*

with typical generalizations, for instance, *isl* or *isl.* in data can be substituted with *island*. The redundant information dictionary has proven to be very useful for long phrase and sentence processing. Using the dictionary the marker meaning list table used in the first stage of operation can be reduced considerably.

#### Conclusion

The above technology has been tested for over a dozen different questionnaires and has received approval for use. Some practical results of qualitative marker processing can be found in the works of Martyshenko, Martyshenko, and Starkov (2007) and in Martyshenko (2008).

References

Martyshenko, S. N., Martyshenko, N. S., & Kustov, D. A. (2007). EXCEL toolbox for questionnaire data processing. *Information Technologies in Modeling and Management, 1*, 112-120.

Martyshenko, N. S., & Starkov, A. S. (2007). Analysis of regional consumer tourist market structure. *Territory of Science, 4(5)*, 468-478.

Martyshenko, N. S. (2008). Tourist cluster formation and development management in Primorsky region. *Region: Systems, Economy, & Management, 2*, 122-132.

## An Exact Test for the Equality of Intraclass Correlation Coefficients Under Unequal Family Sizes

Madhusudan Bhandary  
Columbus State University,  
Columbus, GA USA

Koji Fujiwara  
North Dakota State University,  
Fargo, ND USA

---

An exact test for the equality of two intraclass correlation coefficients under unequal family sizes based on two independent multi-normal samples is proposed. This exact test consistently and reliably produced results superior to those of the Likelihood Ratio Test (LRT) and the large sample Z-test proposed by Young and Bhandary (1998). The test generally performed better in terms of power (for higher intraclass correlation values) for various combinations of intraclass correlation coefficient values and the exact test remained closer to the significance level under the null hypothesis compared to the other two tests. For small sample situations, sizes of the LRT and large-sample Z-tests are drastically higher than alpha-level, but the size of the exact test is close to the alpha-level. The proposed exact test is computationally simple and can be used for both small and large sample situations.

Key words: Likelihood ratio test, Z-test, F-test, intraclass correlation coefficient.

---

### Introduction

The intraclass correlation coefficient  $\rho$  is widely used to measure the degree of intrafamily resemblance with respect to characteristics such as blood pressure, cholesterol, weight, height, stature, lung capacity, etc. Several authors have studied statistical inference concerning  $\rho$  based on single multinormal samples (Scheffe, 1959; Rao, 1973; Rosner, et al., 1977, 1979; Donner & Bull, 1983; Srivastava, 1984; Konishi, 1985; Gokhale & SenGupta, 1986; SenGupta, 1988; Velu & Rao, 1990).

Donner and Bull (1983) discussed the likelihood ratio test (LRT) for testing the equality of two intraclass correlation coefficients based on two independent multinormal samples under equal family sizes. Konishi and Gupta (1987) proposed a modified LRT and derived its

asymptotic null distribution; they also discussed another test procedure based on a modification of Fisher's Z-transformation following Konishi (1985). Huang and Sinha (1993) considered an optimum invariant test for the equality of intraclass correlation coefficients under equal family sizes for more than two intraclass correlation coefficients based on independent samples from several multinormal distributions. For unequal family sizes, Young and Bhandary (1998) proposed a LRT, a large sample Z-test and a large sample  $Z^*$ -test for the equality of two intraclass correlation coefficients based on two independent multinormal samples.

For several populations and unequal family sizes, Bhandary and Alam (2000) proposed a LRT and a large sample ANOVA test for the equality of several intraclass correlation coefficients based on several independent multinormal samples. Donner and Zou (2002) proposed an asymptotic test for the equality of dependent intraclass correlation coefficients under unequal family sizes. Bhandary and Fujiwara (2006) proposed an F-max test for several populations and under unequal family sizes. None of these studies, however, derived an exact test under unequal family sizes. It is an important practical problem to consider an exact test for the equality of

---

Madhusudan Bhandary is an Associate Professor in the Department of Mathematics. Email him at: [bhandary\\_madhusudan@columbusstate.edu](mailto:bhandary_madhusudan@columbusstate.edu). Koji Fujiwara is a graduate student in the Department of Statistics. Email him at: [koji.fujiwara@ndsu.edu](mailto:koji.fujiwara@ndsu.edu).

intraclass correlation coefficients under unequal family sizes.

This article proposes an exact test for the equality of two intraclass correlation coefficients based on two independent multinormal samples under fixed but unequal family sizes. Developing an exact test for the equality of intraclass correlation coefficients under unequal family sizes would allow, for example, the determination of whether blood pressure, cholesterol, lung capacity, etc., among families in Native American or Caucasian races differ from the same among families in Asian races.

Tests of  $H_0 : \rho_1 = \rho_2$  versus  $H_1 : \rho_1 \neq \rho_2$ : Likelihood Ratio Test

Let  $\tilde{X}_i = (x_{i1}, x_{i2}, \dots, x_{ip_i})'$  be a  $p_i \times 1$  vector of observations from  $i^{th}$  family;  $i = 1, 2, \dots, k$ . The structure of mean vector and the covariance matrix for the familial data is given by the following (Rao, 1973):

$$\tilde{\mu}_i = \mu \tilde{1}_i$$

and

$$\tilde{\Sigma}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{pmatrix} \quad (2.1)$$

where  $\tilde{1}_i$  is a  $p_i \times 1$  vector of 1's,  $\mu (-\infty < \mu < \infty)$  is the common mean and  $\sigma^2 (\sigma^2 > 0)$  is the common variance of members of the family and  $\rho$ , which is called the intraclass correlation coefficient, is the coefficient of correlation among the members of the family and  $\max_{1 \leq i \leq k} \left( -\frac{1}{p_i - 1} \right) \leq \rho \leq 1$ .

It is assumed that  $\tilde{x}_i \sim N_{p_i}(\tilde{\mu}_i, \tilde{\Sigma}_i); i = 1, \dots, k$ , where  $N_{p_i}$

represents a  $p_i$ -variate normal distribution and  $\tilde{\mu}_i, \tilde{\Sigma}_i$ 's are defined in (2.1). Let

$$\tilde{u}_i = (u_{i1}, u_{i2}, \dots, u_{ip_i})' = Q \tilde{X}_i \quad (2.2)$$

where  $Q$  is an orthogonal matrix. Under the orthogonal transformation (2.2), it can be observed that  $\tilde{u}_i \sim N_{p_i}(\tilde{\mu}_i^*, \tilde{\Sigma}_i^*); i = 1, \dots, k$ , where

$$\tilde{\mu}_i^* = (\mu, 0, 0, \dots, 0)'$$

$$\tilde{\Sigma}_i^* = \sigma^2 \begin{pmatrix} \eta_i & 0 & \dots & 0 \\ 0 & 1 - \rho & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 - \rho \end{pmatrix},$$

and

$$\eta_i = p_i^{-1} \{1 + (p_i - 1)\rho\}.$$

The transformation used on the data from  $\tilde{x}$  to  $\tilde{u}$  is independent of  $\rho$  and can be accomplished using Helmert's orthogonal transformation.

Srivastava (1984) describes estimators of  $\rho$  and  $\sigma^2$  under unequal family sizes which are good substitutes for the maximum likelihood estimator and are given by the following:

$$\hat{\rho} = 1 - \frac{\hat{\gamma}^2}{\hat{\sigma}^2}$$

where

$$\hat{\sigma}^2 = (k-1)^{-1} \sum_{i=1}^k (u_{i1} - \hat{\mu})^2 + k^{-1} \hat{\gamma}^2 \left( \sum_{i=1}^k a_i \right)$$

$$\hat{\gamma}^2 = \frac{\sum_{i=1}^k \sum_{r=2}^{p_i} u_{ir}^2}{\sum_{i=1}^k (p_i - 1)}$$

$$\hat{\mu} = k^{-1} \sum_{i=1}^k u_{i1}$$

(2.3)

and

$$a_i = 1 - p_i^{-1}.$$

Next, consider the two sample problem with  $k_1$  and  $k_2$  families from each population. Let  $\tilde{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip_i})'$  be a  $p_i \times 1$  vector of observations from  $i^{th}$  family;  $i = 1, \dots, k_1$  and  $\tilde{x}_i \sim N_{p_i}(\tilde{\mu}_{1i}, \tilde{\Sigma}_{1i})$ , where

$$\tilde{\mu}_{1i} = \mu_1 \mathbf{1}_i, \tilde{\Sigma}_{1i} = \sigma_1^2 \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \dots & \rho_1 \\ \dots & \dots & \dots & \dots \\ \rho_1 & \rho_1 & \dots & 1 \end{pmatrix}$$

(2.4)

and  $\max_{1 \leq i \leq k_1} \left( -\frac{1}{p_i - 1} \right) \leq \rho_1 \leq 1$ .

Let  $\tilde{y}_j = (y_{j1}, y_{j2}, \dots, y_{jq_j})'$  be a  $q_j \times 1$  vector of observations from  $j^{th}$  family in the second population ;  $j = 1, \dots, k_2$  and, the distribution of  $\tilde{y}_j$  is same as  $\tilde{x}_i$  in (2.4) except

$\mu_1$  is replaced by  $\mu_2$ ,  $\sigma_1^2$  is replaced by  $\sigma_2^2$  and  $\rho_1$  is replaced by  $\rho_2$  and

$\max_{1 \leq j \leq k_2} \left( -\frac{1}{q_j - 1} \right) \leq \rho_2 \leq 1$ :

$$\tilde{\mu}_{2i} = \mu_2 \mathbf{1}_i, \tilde{\Sigma}_{2i} = \sigma_2^2 \begin{pmatrix} 1 & \rho_2 & \dots & \rho_2 \\ \rho_2 & 1 & \dots & \rho_2 \\ \dots & \dots & \dots & \dots \\ \rho_2 & \rho_2 & \dots & 1 \end{pmatrix}$$

(2.5)

Using orthogonal transformation, the data vector can be transformed from  $\tilde{x}_i$  to  $\tilde{u}_i$

and  $\tilde{y}_j$  to  $\tilde{v}_j$  as follows:

$$\tilde{u}_i = (u_{i1}, u_{i2}, \dots, u_{ip_i})' \sim N_{p_i}(\tilde{\mu}_{1i}^*, \tilde{\Sigma}_{1i}^*); i = 1, \dots, k_1$$

and

$$\tilde{v}_j = (v_{j1}, v_{j2}, \dots, v_{jq_j})' \sim N_{q_j}(\tilde{\mu}_{2j}^*, \tilde{\Sigma}_{2j}^*); j = 1, \dots, k_2$$

where

$$\tilde{\mu}_{1i}^* = (\mu_1, 0, 0, \dots, 0)', \tilde{\Sigma}_{1i}^* = \sigma_1^2 \begin{pmatrix} \eta_i & 0 & \dots & 0 \\ 0 & 1 - \rho_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 - \rho_1 \end{pmatrix}$$

$$\eta_i = p_i^{-1} \{1 + (p_i - 1)\rho_1\}$$

(2.6)

$\tilde{\mu}_{2j}^*$  and  $\tilde{\Sigma}_{2j}^*$  are the same as  $\tilde{\mu}_{1i}^*$  and  $\tilde{\Sigma}_{1i}^*$  respectively except that  $\mu_1$  is replaced by  $\mu_2$ ,  $\sigma_1^2$  is replaced by  $\sigma_2^2$ ,  $\rho_1$  is replaced by  $\rho_2$  and  $\eta_i$  is replaced by  $\xi_j = q_j^{-1} \{1 + (q_j - 1)\rho_2\}$ . The transformations used on the data from  $\tilde{x}$  to  $\tilde{u}$  and  $\tilde{y}$  to  $\tilde{v}$  are independent of  $\rho_1$  and  $\rho_2$ ; note it is assumed that  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ .

Under the above setup, Young and Bhandary (1998) derived likelihood ratio test

statistic for testing  $H_0 : \rho_1 = \rho_2$  vs.  
 $H_1 : \rho_1 \neq \rho_2$  which is given by the following:

$$\begin{aligned}
 -2 \log \Lambda = & \sum_{i=1}^{k_1} \log [p_i^{-1} \{1 + (p_i - 1) \hat{\rho}\}] + \sum_{i=1}^{k_1} (p_i - 1) \log(1 - \hat{\rho}) \\
 & + \sum_{j=1}^{k_2} \log [q_j^{-1} \{1 + (q_j - 1) \hat{\rho}\}] + \sum_{j=1}^{k_2} (q_j - 1) \log(1 - \hat{\rho}) \\
 & + \sum_{j=1}^{k_2} \left\{ q_j (v_{j1} - \hat{\mu}_2)^2 / [1 + (q_j - 1) \hat{\rho}] \right\} + \sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2 / (1 - \hat{\rho}) \\
 & + \sum_{j=1}^{k_2} \left\{ q_j (v_{j1} - \hat{\mu}_2)^2 / [1 + (q_j - 1) \hat{\rho}] \right\} + \sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2 / (1 - \hat{\rho}) \\
 & - \sum_{i=1}^{k_1} \log [p_i^{-1} \{1 + (p_i - 1) \hat{\rho}_1\}] - \sum_{i=1}^{k_1} (p_i - 1) \log(1 - \hat{\rho}_1) \\
 & - \sum_{j=1}^{k_2} \log [q_j^{-1} \{1 + (q_j - 1) \hat{\rho}_2\}] - \sum_{j=1}^{k_2} (q_j - 1) \log(1 - \hat{\rho}_2) \\
 & - \frac{1}{\hat{\sigma}^2} \left[ \sum_{i=1}^{k_1} \{ p_i (u_{i1} - \hat{\mu}_1)^2 / [1 + (p_i - 1) \hat{\rho}_1] \} + \sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2 / (1 - \hat{\rho}_1) \right] \\
 & + \sum_{j=1}^{k_2} \left\{ q_j (v_{j1} - \hat{\mu}_2)^2 / [1 + (q_j - 1) \hat{\rho}_2] \right\} + \sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2 / (1 - \hat{\rho}_2) \Big] \quad (2.7)
 \end{aligned}$$

where,  $\Lambda$  = likelihood ratio test statistic,

$$\hat{\rho} = 1 - \frac{\hat{\gamma}^2}{\hat{\sigma}^{*2}},$$

$$\hat{\gamma}^2 = \frac{\sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2 + \sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2}{\sum_{i=1}^{k_1} (p_i - 1) + \sum_{j=1}^{k_2} (q_j - 1)},$$

and

$$\begin{aligned}
 \hat{\sigma}^{*2} = & (k_1 + k_2 - 1)^{-1} \left[ \sum_{i=1}^{k_1} (u_{i1} - \hat{\mu}_z)^2 + \sum_{j=1}^{k_2} (v_{j1} - \hat{\mu}_z)^2 \right] \\
 & + (k_1 + k_2)^{-1} \hat{\gamma}^2 \left( \sum_{i=1}^{k_1} a_i + \sum_{j=1}^{k_2} b_j \right)
 \end{aligned}$$

which equals the estimate of  $\sigma^2$  under  $H_0$ ,

$$\begin{aligned}
 a_i &= 1 - p_i^{-1} \\
 b_j &= 1 - q_j^{-1}
 \end{aligned}$$

$$\hat{\mu}_z = (k_1 + k_2)^{-1} \left[ \sum_{i=1}^{k_1} u_{i1} + \sum_{j=1}^{k_2} v_{j1} \right]$$

$$\hat{\rho}_1 = 1 - \frac{\hat{\gamma}_1^2}{\hat{\sigma}_1^2},$$

where

$$\hat{\gamma}_1^2 = \frac{\sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2}{\sum_{i=1}^{k_1} (p_i - 1)},$$

$$\hat{\sigma}_1^2 = (k_1 - 1)^{-1} \sum_{i=1}^{k_1} (u_{i1} - \hat{\mu}_1)^2 + k_1^{-1} \hat{\gamma}_1^2 \left( \sum_{i=1}^{k_1} a_i \right)$$

and

$$\hat{\rho}_2 = 1 - \frac{\hat{\gamma}_2^2}{\hat{\sigma}_2^2},$$

where

$$\hat{\gamma}_2^2 = \frac{\sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2}{\sum_{j=1}^{k_2} (q_j - 1)},$$

$$\hat{\sigma}_2^2 = (k_2 - 1)^{-1} \sum_{j=1}^{k_2} (v_{j1} - \hat{\mu}_2)^2 + k_2^{-1} \hat{\gamma}_2^2 \left( \sum_{j=1}^{k_2} b_j \right)$$

$$\hat{\sigma}^2 = \frac{1}{2} (\hat{\sigma}_1^2 + \hat{\sigma}_2^2) = \text{estimate of } \sigma^2 \text{ under } H_1$$

$$\hat{\mu}_1 = k_1^{-1} \sum_{i=1}^{k_1} u_{i1},$$

$$\hat{\mu}_2 = k_2^{-1} \sum_{j=1}^{k_2} v_{j1}.$$

It is known from asymptotic theory that  $-2 \log \Lambda$  has an asymptotic Chi-square distribution with 1 degree of freedom. Here (2.7) is not exact  $-2(\log \text{likelihood ratio})$  because Srivastava's (1984) estimator of parameters was substituted instead of exact likelihood estimator (which are not closed form in this situation).

## EXACT TEST FOR THE EQUALITY OF INTRACLASS CORRELATION COEFFICIENTS

However, Srivastava's (1984) estimators are strongly consistent and hence the asymptotic behavior of (2.7) may not be accurately as Chi-square distribution. It is a close approximation to the Chi-square distribution.

Young and Bhandary (1998) also proposed large sample Z-test as follows:

$$Z = \frac{\hat{\rho}_1 - \hat{\rho}_2}{S \sqrt{\frac{1}{k_1} + \frac{1}{k_2}}} \quad (2.8)$$

where,  $\hat{\rho}_1$  = estimator of  $\rho_1$  from the first sample using Srivastava (1984),  $\hat{\rho}_2$  = estimator of  $\rho_2$  from the second sample using Srivastava (1984) and  $S^2$  = pooled estimator of variance under  $H_0$  using the formula given by Srivastava and Katapa (1986) as follows:

$$Var = 2(1-\rho)^2 \left\{ (\bar{p}-1)^{-1} + c^2 - 2(1-\rho)(\bar{p}-1)^{-1} k^{-1} \left( \sum_{i=1}^k a_i \right) \right\} \quad (2.9)$$

where,  $k$  = number of families in the sample,

$$\bar{p} = k^{-1} \sum_{i=1}^k p_i$$

$$c^2 = 1 - 2(1-\rho)^2 k^{-1} \sum_{i=1}^k a_i + (1-\rho)^2 \left[ k^{-1} \sum_{i=1}^k a_i + (\bar{p}-1)^{-1} \bar{a}^2 \right]$$

$$\bar{a} = k^{-1} \sum_{i=1}^k a_i$$

and  $a_i = 1 - p_i^{-1}$ .

It is clear that under  $H_0$ , the test statistic  $Z$  given by (2.8) has an asymptotic  $N(0,1)$  distribution. The statistic  $Z$  in (2.8) under  $H_0$  may not be exactly  $N(0,1)$  asymptotically, but it is a close approximation for large sample situation. A rough sketch of proof is as follows:

- i.  $\hat{\rho}_1 \sim N(\rho, V(\hat{\rho})/k_1)$  asymptotically under  $H_0$  by Srivastava and Katapa (1986), where  $V(\hat{\rho})$  is given by (2.9). (\*)

- ii. Similarly,  $\hat{\rho}_2 \sim N(\rho, V(\hat{\rho})/k_2)$  asymptotically under  $H_0$  by Srivastava and Katapa (1986), where  $V(\hat{\rho})$  is given by (2.9). (\*\*)

- iii. The approximate asymptotic distribution of  $Z$  is obtained in (2.8) as  $N(0,1)$  using (\*) and (\*\*), where  $S^2$  is  $V(\hat{\rho})$  with its estimate given by (2.9) where  $\rho$  is replaced by  $\hat{\rho}$  under  $H_0$ .

Young and Bhandary (1998) showed through simulation that the likelihood ratio test given by (2.7) consistently produced results superior to those of the large sample Z-test given by (2.8). It can be observed that likelihood ratio test given by (2.7) is computationally complex and is also used asymptotically, that is, when family sizes are large (at least 30). However, a researcher may have a situation in which only a small sample is available; thus, an exact F-test is proposed which is computationally simple and can be used for both small sample and large sample situations.

### Proposed Exact Test

The new exact test is described as follows:

$$F = \frac{\sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2 / \left\{ \sum_{i=1}^{k_1} (p_i - 1) \right\}}{\sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2 / \left\{ \sum_{j=1}^{k_2} (q_j - 1) \right\}} \quad (2.10)$$

It can be shown using (2.6) that

$$\frac{\sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2}{\sigma^2 (1 - \rho_1)} \sim \chi^2_{\sum_{i=1}^{k_1} (p_i - 1)} \quad (2.11)$$

and if the following replacements are made,  $u_{ir}$  by  $v_{js}$ ,  $k_1$  by  $k_2$ ,  $r$  by  $s$ ,  $p_i$  by  $q_j$  and  $\rho_1$

by  $\rho_2$ , in (2.11) expression another Chi-square distribution results with

$$\sum_{j=1}^{k_2} (q_j - 1) \text{ d.f.} \quad (2.12)$$

where,  $\chi_n^2$  denotes a Chi-square distribution with n degrees of freedom (d.f.).

Hence, under  $H_0$ , the exact distribution of the F-test statistic given by (2.10) is an F-distribution with degrees of freedom  $\sum_{i=1}^{k_1} (p_i - 1)$  and  $\sum_{j=1}^{k_2} (q_j - 1)$  respectively (using (2.11) and (2.12)). Hence, the critical region (or rejection region) for testing  $H_0 : \rho_1 = \rho_2$  vs.  $H_1 : \rho_1 \neq \rho_2$  can be written as follows:

$$\begin{aligned} F &> F_{\frac{\alpha}{2}; \sum_{i=1}^{k_1} (p_i - 1), \sum_{j=1}^{k_2} (q_j - 1)} \\ \text{or} & \\ F &< F_{1 - \frac{\alpha}{2}; \sum_{i=1}^{k_1} (p_i - 1), \sum_{j=1}^{k_2} (q_j - 1)} \end{aligned} \quad (2.13)$$

where F is the test statistic given by (2.10) and  $F_{\gamma; a, b}$  is the upper  $100\gamma\%$  point of an F-distribution with degrees of freedom a and b respectively.

The exact power function of the F-test given by (2.10) can be derived as follows:

$$\begin{aligned} \varphi(\rho_1, \rho_2) &= P \left( F > \frac{1 - \rho_2}{1 - \rho_1} \cdot F_{\frac{\alpha}{2}; \sum_{i=1}^{k_1} (p_i - 1), \sum_{j=1}^{k_2} (q_j - 1)} \right) \\ &+ P \left( F < \frac{1 - \rho_2}{1 - \rho_1} \cdot F_{1 - \frac{\alpha}{2}; \sum_{i=1}^{k_1} (p_i - 1), \sum_{j=1}^{k_2} (q_j - 1)} \right) \end{aligned} \quad (2.14)$$

where

$$F = \frac{\frac{\sum_{i=1}^{k_1} \sum_{r=2}^{p_i} u_{ir}^2}{\sigma^2(1 - \rho_1)}}{\frac{\sum_{j=1}^{k_2} \sum_{s=2}^{q_j} v_{js}^2}{\sigma^2(1 - \rho_2)}} \bigg/ \frac{\sum_{i=1}^{k_1} (p_i - 1)}{\sum_{j=1}^{k_2} (q_j - 1)} \sim$$

F-distribution with degrees of freedom  $\sum_{i=1}^{k_1} (p_i - 1)$  and  $\sum_{j=1}^{k_2} (q_j - 1)$  respectively (using (2.11) and (2.12)) under  $H_1$ .

It can be observed that the test statistic F given by (2.10) is very simple to compute and that the distribution of F is exact and hence can be used for both small sample and large sample situations.

#### Simulation Study

Multivariate normal random vectors were generated using the R program in order to evaluate the power of the F statistic as compared to the LRT statistic and Z-statistic. Five, 15 and 30 vectors of family data were created for each of the two populations. The family size distribution was truncated to maintain the family size at a minimum of 2 and a maximum of 15 siblings. Previous research in simulating family sizes (Rosner, et al., 1977; Srivastava & Keen, 1988) determined the parameter setting for FORTRAN IMSL negative binomial subroutine with a mean = 2.86 and a success probability = 0.483. In this study, the mean was set to equal 2.86 and theta equals 41.2552. All parameters were set the same for each population, except the values of  $\rho_1$  and  $\rho_2$  which took on all combinations possible over the range of values from 0.1 to 0.9 at increments of 0.1.

The R program produced estimates of  $\rho_1$  and  $\rho_2$  along with the F statistic, the LRT statistic and the Z- statistic 3,000 times for each particular combination of population parameters ( $\rho_1$  and  $\rho_2$ ). The frequency of rejection for each test statistic at  $\alpha=0.05$  was noted and the proportion of rejections are reported in Table 1 for various combinations of  $\rho_1$  and  $\rho_2$ . Table 2 shows the size comparison for the LRT statistic,



## EXACT TEST FOR THE EQUALITY OF INTRACLASS CORRELATION COEFFICIENTS

the F statistic and the Z statistic for various combinations of  $\rho_1$  and  $\rho_2$ . Figures 1-4 present power estimates as well as size estimates for the three tests.

This study found that the exact F-test showed consistently better results for higher  $\rho_1$  and  $\rho_2$  values in power as well as in size

compared to the LRT test and large sample Z-test. For the small sample situation, the LRT test and large sample Z-test have sizes drastically higher than alpha-level whereas the exact F test is close to the alpha-level. Based on these results, the F test is strongly recommended for use in practice.

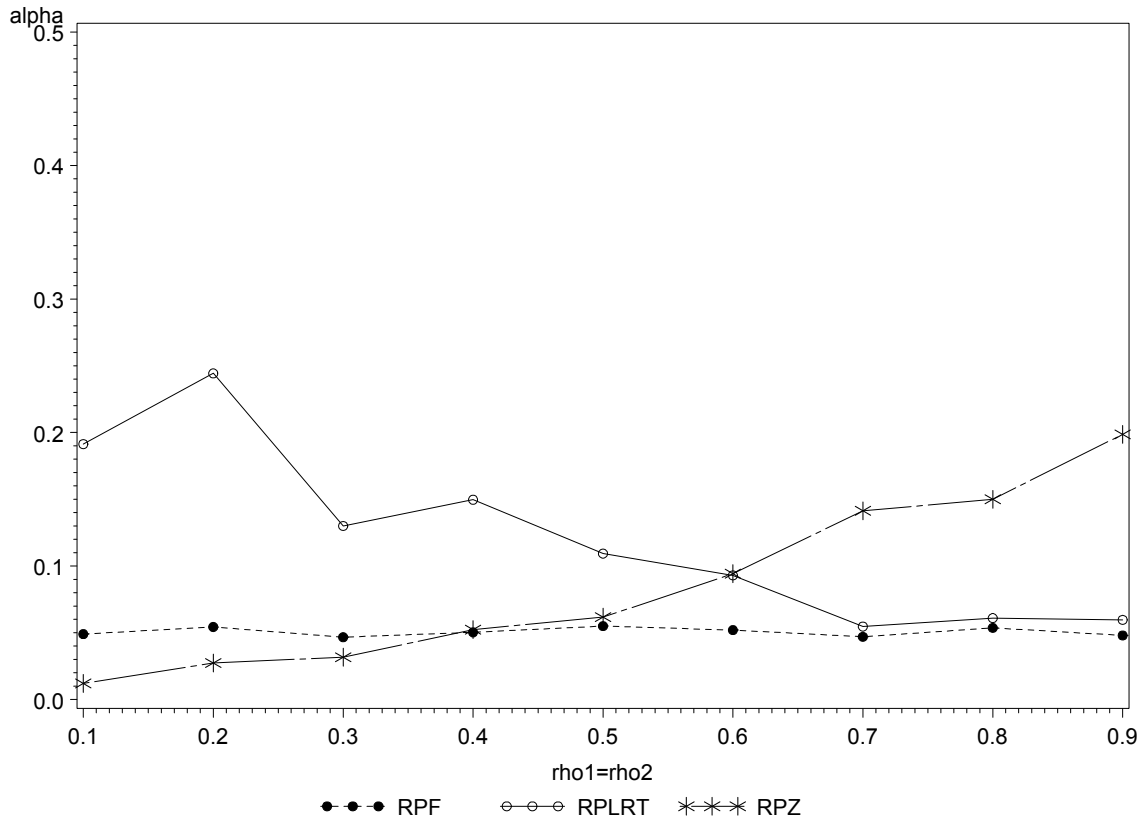
Table 1: Rejection Proportions for  $\alpha = 0.05$

rho1	rho2	K = 5			K = 15			K = 30		
		LRT	F	Z	LRT	F	Z	LRT	F	Z
0.1	0.2	0.1947	0.0673	0.0123	0.1377	0.0600	0.0233	0.0910	0.0780	0.0447
0.1	0.4	0.1660	0.0857	0.0607	0.2477	0.2103	0.1647	0.5423	0.4330	0.3633
0.1	0.6	0.3873	0.2520	0.1413	0.7470	0.6923	0.5497	0.9723	0.9557	0.8717
0.1	0.8	0.7067	0.7500	0.4257	0.9870	0.9943	0.9237	1.0000	1.0000	0.9987
0.3	0.2	0.1930	0.0597	0.0383	0.0807	0.0700	0.0470	0.0953	0.0817	0.0650
0.3	0.4	0.1347	0.0540	0.0493	0.0813	0.0783	0.0557	0.0997	0.1073	0.0667
0.3	0.6	0.1803	0.1570	0.1467	0.3477	0.3630	0.2593	0.6667	0.6633	0.4803
0.3	0.8	0.5410	0.5917	0.2763	0.8870	0.9147	0.7397	0.9990	1.0000	0.9600
0.5	0.2	0.2093	0.1303	0.0977	0.3013	0.2847	0.2180	0.5863	0.5310	0.4213
0.5	0.4	0.1113	0.0607	0.0723	0.0740	0.0880	0.0680	0.1033	0.1163	0.1010
0.5	0.6	0.1013	0.0663	0.1190	0.0737	0.0843	0.0877	0.1367	0.1663	0.1037
0.5	0.8	0.3403	0.3997	0.2493	0.6977	0.7747	0.4183	0.9327	0.9603	0.7173
0.7	0.2	0.3910	0.3757	0.2557	0.8183	0.8247	0.6033	0.9867	0.9873	0.9080
0.7	0.4	0.2647	0.2740	0.1717	0.5623	0.6080	0.3400	0.7983	0.8413	0.5783
0.7	0.6	0.0877	0.0730	0.1320	0.1080	0.1303	0.1070	0.1813	0.2227	0.1327
0.7	0.8	0.1107	0.1143	0.1523	0.1827	0.2610	0.1623	0.3373	0.4393	0.2277
0.9	0.2	0.8723	0.9390	0.5140	0.9963	1.0000	0.9667	1.0000	1.0000	1.0000
0.9	0.4	0.8527	0.9317	0.4453	0.9967	0.9993	0.8607	1.0000	1.0000	0.9983
0.9	0.6	0.4703	0.5330	0.2457	0.9710	0.9897	0.6973	0.9983	1.0000	0.9277
0.9	0.8	0.1680	0.1930	0.1797	0.3827	0.5087	0.2173	0.7263	0.8420	0.4337

Table 2: Checking the Alpha Level ( $\alpha = 0.05$ )

rho1	rho0	K = 5			K = 15			K = 30		
		LRT	F	Z	LRT	F	Z	LRT	F	Z
0.1	0.1	0.1913	0.0490	0.0120	0.0943	0.0480	0.0157	0.0770	0.0497	0.0153
0.2	0.2	0.2443	0.0543	0.0273	0.0757	0.0500	0.0220	0.0477	0.0490	0.0220
0.3	0.3	0.1300	0.0467	0.0317	0.0497	0.0480	0.0383	0.0387	0.0503	0.0330
0.4	0.4	0.1497	0.0503	0.0523	0.0403	0.0513	0.0457	0.0410	0.0470	0.0410
0.5	0.5	0.1093	0.0550	0.0617	0.0420	0.0540	0.0540	0.0363	0.0487	0.0433
0.6	0.6	0.0930	0.0520	0.0943	0.0367	0.0457	0.0633	0.0393	0.0540	0.0547
0.7	0.7	0.0547	0.0470	0.1413	0.0363	0.0573	0.0760	0.0337	0.0473	0.0540
0.8	0.8	0.0610	0.0537	0.1500	0.0290	0.0457	0.0827	0.0307	0.0533	0.0663
0.9	0.9	0.0597	0.0480	0.1987	0.0230	0.0430	0.0853	0.0327	0.0557	0.0623

Figure 1: Alpha Level (k = 5, alpha = 0.05)



# EXACT TEST FOR THE EQUALITY OF INTRACLASS CORRELATION COEFFICIENTS

Figure 2: Alpha Level ( $k = 15$ ,  $\alpha = 0.05$ )

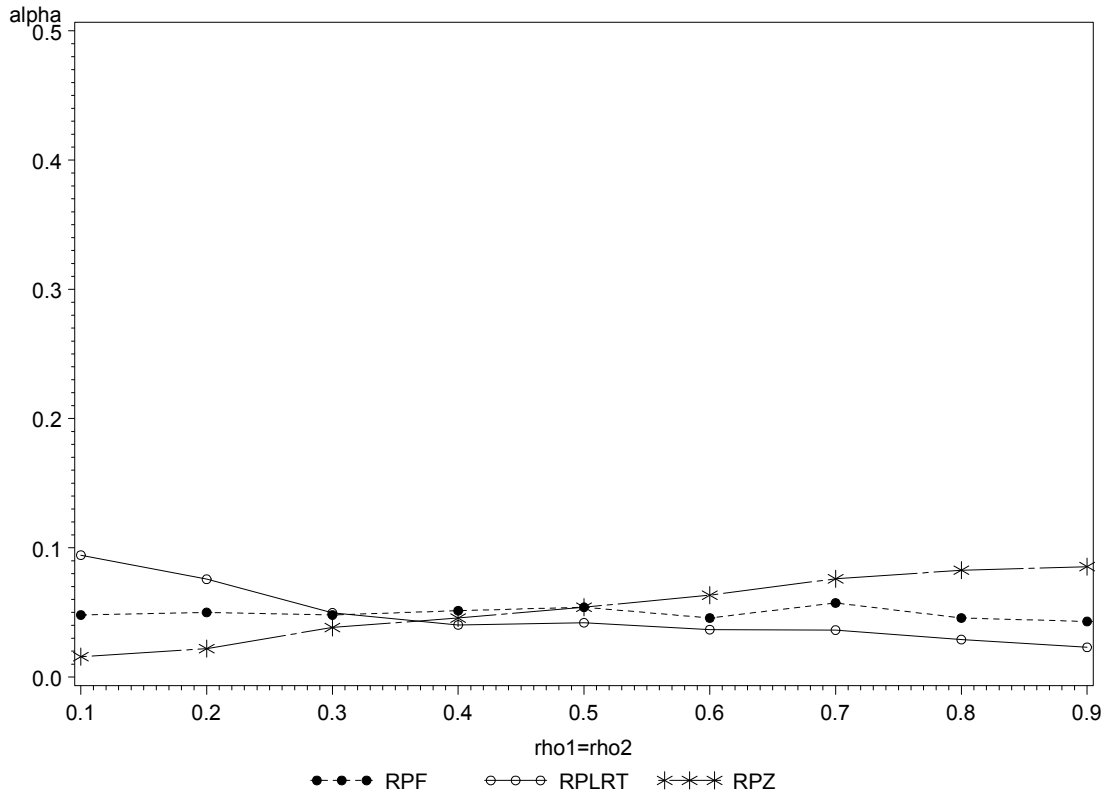


Figure 3: Power ( $\rho_1 = 0.9$ ,  $k = 15$ ,  $\alpha = 0.05$ )

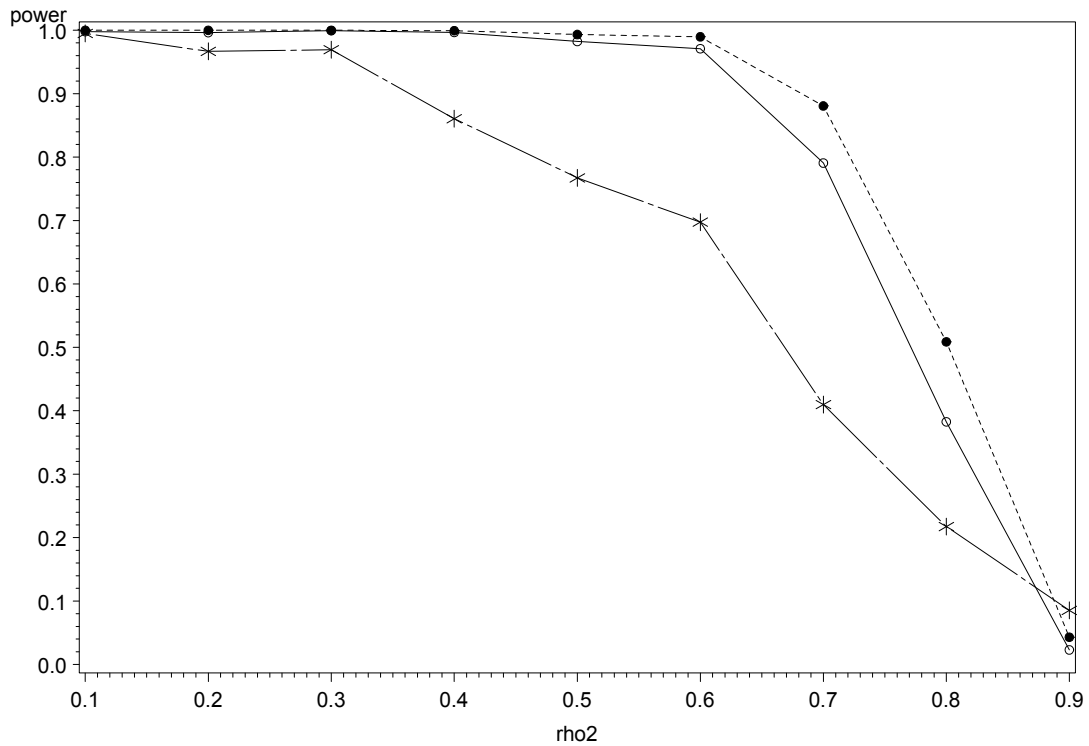
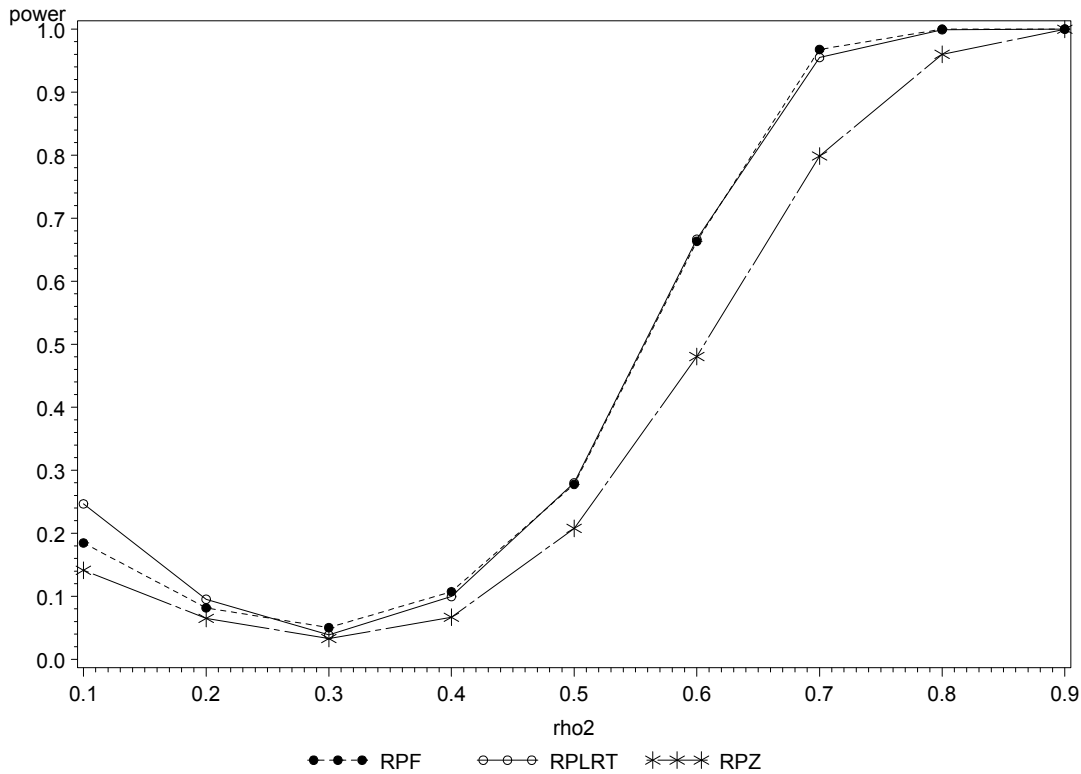


Figure 4: Power (rho1 = 0.3, k = 30, alpha = 0.05)



Example with Real Life Data

The main goal of this study is to find a better and easier way to compute the proposed exact F-test compared to the test given by Young and Bhandary (1998), the aim is not to find a test in situations such as  $\sigma_1^2 = \sigma_2^2$  and  $\rho_1 \neq \rho_2$  or  $\sigma_1^2 \neq \sigma_2^2$  and  $\rho_1 = \rho_2$  or for non-normal data, etc. For this reason, the same example - values of pattern intensity on soles of feet in fourteen families - as used by Young and Bhandary (1998) is employed for this example. In this example two tests using real life data collected from Srivastava and Katapa (1986) are compared. First the data is randomly split into two samples as shown in Table 3.

First, the data is transformed by multiplying each observation vector by Helmert's orthogonal matrix Q, where

$$Q = \begin{bmatrix} \frac{1}{\sqrt{p_i}} & \frac{1}{\sqrt{p_i}} & \frac{1}{\sqrt{p_i}} & \dots & \frac{1}{\sqrt{p_i}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \dots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \frac{1}{\sqrt{p_i(p_i-1)}} & \frac{1}{\sqrt{p_i(p_i-1)}} & \frac{1}{\sqrt{p_i(p_i-1)}} & \dots & -\frac{(p_i-1)}{\sqrt{p_i(p_i-1)}} \end{bmatrix}$$

This results in transformed vectors  $u_i$  and  $v_j$  respectively for  $i = 1, 2, \dots, k_1$  and  $j = 1, 2, \dots, k_2$ ; here,  $k_1 = 7$  and  $k_2 = 7$ .

Srivastava's formula given by (2.3) is used to compute intraclass correlation coefficients. The computed values of intraclass correlation coefficients are  $\hat{\rho}_1 = 0.8708$ ,  $\hat{\rho}_2 = 0.8544$  and  $\hat{\rho} = 0.85847$ . The computed values of the LRT and F statistics are obtained

# EXACT TEST FOR THE EQUALITY OF INTRACLASS CORRELATION COEFFICIENTS

Table 3: Random Sample Split for Values of Pattern Intensity on Soles of Feet in Fourteen Families Data

Sample	Family#	Mother	Father	Siblings
A	12	2	4	2, 4
	10	5	4	4, 5, 4
	9	5	5	5, 6
	1	2	3	2, 2
	4	2	4	2, 2, 2, 2, 2
	5	6	7	6, 6
	8	3	7	2, 4, 7, 4, 4, 7, 8
B	3	2	3	2, 2, 2
	6	4	3	4, 3, 3
	14	2	3	2, 2, 2
	7	4	3	2, 2, 3, 6, 3, 5, 4
	2	2	3	2, 3
	11	5	6	5, 3, 4, 4
	13	6	3	4, 3, 3, 3

from formula (2.7) and (2.10) respectively are as follows: LRT statistic = 1.73374 and F statistic = 2.15690. Based on these results, the null hypothesis would not be rejected by either test at 1%, 5% or 10% levels. Intuitively, the test should not be rejected because the data is from one population and split into two samples. Considering the power and level of the two tests suggested in the simulation the proposed exact F test is recommended for use in practice.

## References

Bhandary, M., & Alam, M. K.(2000). Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. *Communications in Statistics-Theory and Methods*, 29(4), 755-768.

Bhandary, M., & Fujiwara, K. (2006). A small sample test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. *Communications in Statistics-Simulation and Computation*, 35(3), 765-778.

Donner, A., & Bull, S. (1983). Inferences concerning a common intraclass correlation coefficient. *Biometrics*, 39, 771-775.

Donner, A., & Zou, G. (2002). Testing the equality of dependent intraclass correlation coefficients. *The Statistician*, 51(3), 367-379.

Gokhale, D. V., & SenGupta, A. (1986). Optimal tests for the correlation coefficient in a symmetric multivariate normal population. *Journal of Statistical Planning and Inference*, 14, 263-268.

Huang, W., & Sinha, B. K. (1993). On optimum invariant tests of equality of intraclass correlation coefficients. *Annals of the Institute of Statistical Mathematics*, 45(3), 579-597.

Konishi, S. (1985). Normalizing and variance stabilizing transformations for intraclass correlations. *Annals of the Institute of Statistical Mathematics*, 37, 87-94.

Konishi, S., & Gupta, A. K. (1989). Testing the equality of several intraclass correlation coefficients. *J Journal of Statistical Planning and Inference*, 21, 93-105.

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Rosner, B., Donner, A., & Hennekens, C. H. (1977). Estimation of intraclass correlation from familial data. *Applied Statistics*, 26, 179-187.

Rosner, B., Donner, A., & Hennekens, C. H. (1979). Significance testing of interclass correlations from familial data. *Biometrics*, 35, 461-471.

SenGupta, A. (1988). On loss of power under additional information – an example. *Scandinavian Journal of Statistics*, 15, 25-31.

Scheffe, H. (1959). *The analysis of variance*. New York: Wiley.

Srivastava, M. S. (1984). Estimation of interclass correlations in familial data. *Biometrika*, 71, 177-185.

Srivastava, M. S., & Katapa, R. S. (1986). Comparison of estimators of interclass and intraclass correlations from familial data. *Canadian Journal of Statistics*, 14, 29-42.

Srivastava, M. S., & Keen, K. J. (1988). Estimation of the interclass correlation coefficient. *Biometrika*, 75, 731-739.

Velu, R., & Rao, M. B. (1990). Estimation of parent-offspring correlation. *Biometrika*, 77(3), 557-562.

Young, D., & Bhandary, M. (1998). Test for the equality of intraclass correlation coefficients under unequal family sizes. *Biometrics*, 54(4), 1363-1373.

## General Piecewise Growth Mixture Model: Word Recognition Development for Different Learners in Different Phases

Amery D. Wu    Bruno D. Zumbo    Linda S. Siegel  
University of British Columbia,  
Vancouver, B.C. Canada

---

The General Piecewise Growth Mixture Model (GPGMM), without losing generality to other fields of study, can answer six crucial research questions regarding children's word recognition development. Using child word recognition data as an example, this study demonstrates the flexibility and versatility of the GPGMM in investigating growth trajectories that are potentially phasic and heterogeneous. The strengths and limitations of the GPGMM and lessons learned from this hands-on experience are discussed.

Key words: Structural equation model, piecewise regression, growth and change, growth mixture model, latent class analysis, population heterogeneity, word recognition, reading development, trajectories, literacy development.

---

### Introduction

People learn and develop in different ways in different phases. A rich body of literature has documented the complexities in human development, among which the best known is probably Piaget's phasic theory about children's cognitive development. However, in statistical modeling, such complexities are often disguised by a primitive assumption about homogeneity and linearity of data. The purpose of this study is, in the context of children's reading development, to demonstrate the application of the General Piecewise Growth Mixture Model (GPGMM). GPGMM is a versatile modeling strategy that allows for the investigation of trajectories that are heterogeneous and phasic. GPGMM marries the general growth mixture

model (GGMM) articulated by Muthén (2004) with piecewise regression (Li, Duncan, Duncan, & Hops, 2001; McGee & Carleton, 1970; Muthén & Muthén, 1998-2007).

### Overview of Two Reading Development Theories

The debate over the developmental pathways of children's literacy achievement has not been resolved. Two major competing theories exist: the deficit and the lag models. The deficit model suggests that children who have a superior start in precursor linguistic and cognitive skills will improve their reading skills at a faster rate than those with a slower start (e.g., Bast & Reitsma, 1998; Francis, et al., 1996). The increasing difference in reading performance among poor, average and advanced readers observed in early development is believed to be a result of initial skill sets that never develop sufficiently in those who turn out to be poor readers.

An alternative view, the lag model, suggests that children with a poorer start in their cognitive skills will display a faster growth in their later development, whereas those with a superior start will display a slower growth (Leppänen, et al., 2004; Phillips, et al., 2002).

---

Bruno D. Zumbo is a Professor, Amery D. Wu is a Post-doctoral Research Scientist, and Linda S. Siegel is a Professor at the University of British Columbia. Send correspondence to Bruno D. Zumbo at [bruno.zumbo@ubc.ca](mailto:bruno.zumbo@ubc.ca) or Amery Wu at [ameryw@yahoo.com](mailto:ameryw@yahoo.com).

Protagonists of this view believe that children who differ in reading ability vary only in the rate at which cognitive skills develop so that lagging children will eventually catch up with their peers, and that the gap in the early development will eventually disappear.

Empirical evidence has not consistently confirmed either the deficit or lag model. Bast and Reitsma (1998) provided support for the deficit model based on the findings that the rank ordering of the six waves of word recognition scores remained stable and that the differences in the score increased from grade one to grade three. They concluded that differences in reading achievement of the 280 Dutch children were cumulative. In a longitudinal study, Francis, et al. (1996) studied the trajectories of 403 non-disabled and disabled children in Connecticut from grade one to grade nine using the Rasch-scaled composite score of the Word Identification, Word Attack and Passage Comprehension subtests (Woodcock-Johnson Psychoeducational Test Battery; Woodcock & Johnson, 1977). They used quadratic trajectories to model the non-linear growth pattern displayed in the data. The results showed that the disabled readers were unable to develop adequate reading skills and their problems persisted into adolescence. They concluded that a deficit model best characterized the enlarging gap and an intervention at an early age is essential in order to reduce the impact of early deficit.

Other studies, however, have reported that initially poor readers improved faster, and the early gap decreased over time (e.g., Aarnoutse, et al., 2001; Aunola, et al., 2002; Jordan, Kaplan & Hanich, 2002; Scarborough & Parker, 2003). For example, assuming linearity from grade two to grade eight, Scarborough and Parker (2003) reported decreasing gaps of 57 non-disabled and disabled children in both WJ-Word Identification and WJ-Passage Comprehension.

In a longitudinal study of 198 English readers in Canada from grade one to grade six, Parrila, et al. (2005) studied the development of word identification, word attack and passage comprehension separately. For each outcome measure, they fitted a latent growth quadratic curve using growth mixture modeling and found that children with lower starting performance

reduced the distance between themselves and children who had higher initial performance. Aarnoutse, et al. (2001) also failed to find the fan-spread pattern in reading comprehension, vocabulary, spelling or word decoding efficiency. Their results suggested that the initially low performers tended to show greater gains than did medium or high performers. Similarly, Aunola, et al. (2002) found a decrease in individual differences in a reading skill score (a composite of four different reading tasks) of Finnish children. Scarborough and Parker (2003) also reported that the difference between good and poor readers in their US sample were smaller in grade eight than grade two in a composite reading score made of word reading, decoding and passage comprehension.

Existing evidence has not provided conclusive support for either the deficit or the lag models, or for the relationship between early performance and subsequent growth rate. The incongruence in the empirical findings is palpable if careful attention is paid to the diversified and piecemeal approach to the research design and data analysis (Parrila, et al., 2005).

As is evident from this brief review, the research designs varied in the length and phase of the studied time interval (i.e., earlier or later development in the grade school), the statistical analyses (e.g., ANOVA, regression or latent growth model), measures used to represent reading ability, the population of children whose growth trajectories were compared (e.g., normative or children with learning difficulties), the hypothesized pattern of growth trajectory (e.g., linear or quadratic), outcome measure (e.g., word recognition or reading comprehension) and sample size, as well as the terminologies and their operational definitions. Parrila, et al. (2005) concluded that reading development could follow multiple pathways, only some of which are captured by the existing conceptualizations. Thus, researchers could benefit from a more integral and comprehensive data analytical framework that is capable of modeling the complex, intricate, and diversified developmental nature of children's reading development.



# GPGMM WORD RECOGNITION

## Methodology

### Data

The data consists of 1,853 elementary school children from the North Vancouver school district in British Columbia. These children were measured every year in the fall starting from kindergarten to grade six. The dependent variable, word recognition, had a maximum score of 57, which was measured by naming 15 alphabet letters and followed by the reading subtest of the Wide Range Achievement Test-3 (WRAT-3; Wilkinson, 1995), which has a list of 42 words ordered by difficulty. The measurement of word recognition remained the same across the seven waves of data collection; hence, measurement invariance that warranted temporal score comparability across grades was

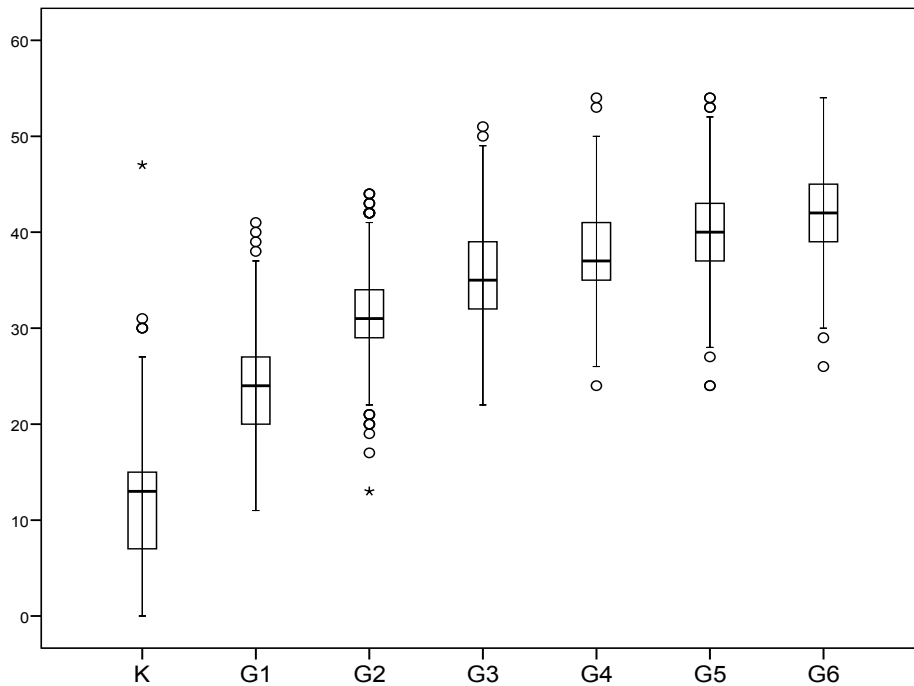
assumed. For demonstrative purposes, only the data of the 526 children who had all seven waves of data were included. Outliers were retained because this study aimed to model these cases through distinct latent classes so that – within each class – the distribution of reading performance was assumed to be normal.

Table 1 displays the mean (M), standard deviation (SD) and skewness of the seven waves of the data. Figure 1 displays the boxplots for the seven waves of word recognition scores. It can be observed that the distributions of the seven word recognition measures are, for the most part, symmetric. The overall performance in word recognition improved across time, with faster growth in the period between kindergarten and grade two, and relatively slower growth in

Table 1: Descriptive Statistics of the Word Recognition Scores

	K	G1	G2	G3	G4	G5	G6
Mean	11.60	23.86	31.55	35.63	37.62	40.27	41.89
SD	5.14	4.85	4.52	4.99	4.39	4.71	4.20
Skewness	-0.21	0.38	0.22	0.32	0.15	0.07	-0.14

Figure 1: Boxplots for the Word Recognition Scores from Kindergarten to Grade Six



the period between grade three and grade six. Figure 2 shows the individual trajectories. The overall pattern of the trajectories was consistent with those revealed in the boxplots and the literature, which showed a nonlinear trend (Francis, et al., 1996; Parrila, et al., 2005).

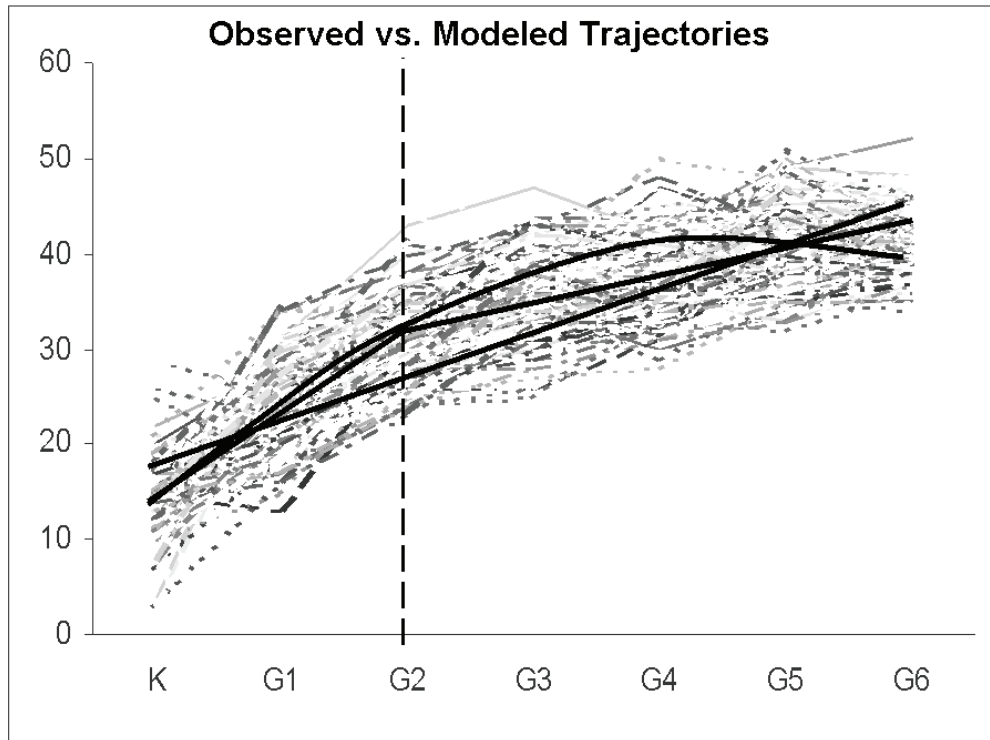
Given the observed non-linearity, it would seem inappropriate to impose a linear trajectory to the observed data as portrayed by the thick single straight line in Figure 2. Most previous studies fitted a quadratic curve to model this non-linear pattern as portrayed by the thick curve line in Figure 2, where the early development is assumed to improve with a faster growth, followed by a relatively slower growth, and then reach a peak with a possibility to decline near the end. Although a quadratic function is fairly accessible and widely used by applied researchers, it may be inappropriate for literacy development of school-age children, because it may portray a decline at the end of the developmental course, whereas reading development, at worst, is expected to plateau rather than decline, if not continue to grow.

Also, the meaning of a quadratic parameter is often hard to interpret conceptually for phenomena studied in the social and behavioral sciences, such as word recognition.

Another possibility for modeling the developmental pattern observed in Figure 2 is to fit a piecewise linear trajectory (Khoo, 1997; Li, et al., 2001; McGee & Carleton, 1970; Raudenbush & Bryk, 2002) as shown by the two thick segments connected at grade two in Figure 2. A piecewise trajectory allows different linear growth rates to be fitted to different developmental phases that are empirically observed or theoretically hypothesized.

Notice that despite the overall trend observed in Figure 2, a great deal of variation exists in individual's developmental pattern as demonstrated by the differences in the starting performance, the speed of learning over time and the ending performance at grade six. Imposing a homogeneous trajectory to these heterogeneous learning patterns may overlook the complexities and diversity of children's reading development.

Figure 2: Observed vs. Modeled Trajectories (Single Linear, Quadratic and 2-piece Linear) of Word Recognition Scores from Kindergarten to Grade Six



## GPGMM WORD RECOGNITION

### General Piecewise Growth Mixture Model (GPGMM): What Can It Do?

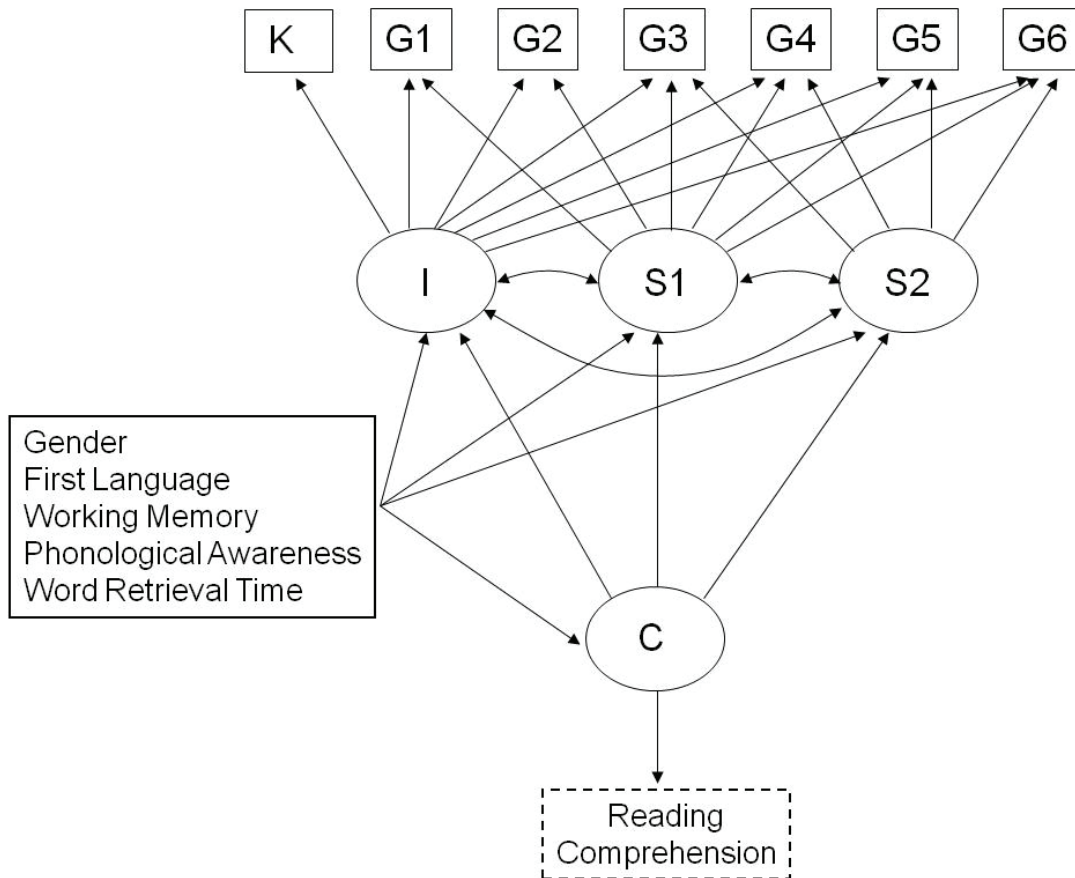
The GPGMM, at its foundation, is a structural equation model, a latent variable approach for investigating growth and change (Meredith & Tisak, 1990; Muthén, 2001; Muthén, 2008). GPGMM is a relatively new and fairly complex modeling framework for studying growth and change (Muthén, 2004). It combines the growth mixture model (GMM) that models population growth heterogeneity with the piecewise regression that models phasic growth rates.

The “mixture” of growth mixture modeling refers to the finite mixture modeling element; that is, modeling with categorical latent variables that represent subpopulations (classes) where population membership is unknown but is inferred from the data (McLachlan & Peel, 2000). The “piecewise” of the piecewise regression refers to the growth rates in different

developmental phases as reflected via the continuous latent growth factors. The GPGMM is an extension of the piecewise GMM by adding the covariates and the developmental outcome variable (Muthén, 2004). The combination of continuous and categorical latent variables of the GPGMM provides a very flexible analytical framework for investigating subpopulations showing distinct and phasic developmental patterns.

GMM has gained increased popularity in studying children’s reading development. Statisticians and methodologists have proposed growth mixture models other than the GPGMM demonstrated in this paper that are of great theoretical and practical significance. Examples of these developments include Muthén, et al. (2003) and Boscardin, et al. (2008). The GPGMM specified for this demonstration is depicted graphically in Figure 3.

Figure 3: GPGMM for Word Recognition Development



GPGMM attends to individual differences in developmental changes by allowing the growth factor to vary across individuals, resulting in individual varying trajectories over time. This individual heterogeneity in trajectories, in a conventional linear form, is captured by two continuous growth factors (a.k.a., random effects); one is a latent variable representing individual differences in the initial performance (i.e., intercept), and the other representing the individual differences in the growth rate (i.e., slope). Growth factors are created by summarizing the growth patterns observed in the repeated measures of the same individuals over time. The categorical latent variable C in Figure 3 models the population heterogeneity in the growth factors.

GPGMM can answer six crucial questions pertaining to children's reading developmental complexities. These research questions, shown in two sequential sets, are:

Set A:

- A1. Are there distinct phases where children differ in their speed of learning?
- A2. Are there unknown subpopulations (latent classes) that differ in their growth pattern?
- A3. How are the starting performance and growth rates related?

Set B:

- B1. What are the characteristics of the latent classes?
- B2. For each class, what explains children's starting performance and growth rates?
- B3. Do the latent classes differ in the reading developmental outcome?

Note that although these questions were posed and answered herein as two sequential sets (A1-A3 and B1-B3), the experience with the modeling procedures of the GPGMM in this study was non-linear; it required a recursive process of model specification, model selection, and meaning checking as would be the case with any other complex modeling. Nonetheless, two general modeling stages can be distinguished for a GPGMM. The first stage was the unconditional piecewise GMM (i.e., without the covariates and the proximal developmental

outcome variable), that answered questions A1, A2 and A3. The major modeling task of this stage was to choose the optimal growth trajectories and number of classes. The second stage entailed the full GPGMM by incorporating the covariates and the developmental outcome variable (i.e., reading comprehension measured at the last time point) into the unconditional piecewise GMM. The conditional piecewise GMM further answered questions B1, B2, and B3. The major modeling task of this stage was to understand the characteristic of the classes and explain class-specific variations in the growth factors and the developmental outcome variable.

Model Estimation and Fit

The following briefly describes the model estimation method and fit statistics used for this demonstration. As asserted at the outset, the focus of this article is to provide a conceptual account and modeling demonstration of the GPGMM instead of technical details. General model specification can be found in Technical Appendix 8 of *Mplus* (Muthén, 1998-2004) and *Mplus User's Guide* (Muthén & Muthén, 1998-2007). The technical details can be found in Muthén and Shedden (1999) and Muthén and Asparouhov (2008). To foster a wider use of the GPGMM, the *Mplus* syntax for the final model can be found in the Appendix.

In *Mplus*, three estimators are available for a GMM: (1) maximum likelihood parameter estimates with conventional standard errors (ML), (2) maximum likelihood parameter estimates with standard errors approximated by first-order derivatives (MLF), and (3) maximum likelihood parameter estimates with robust standard errors (MLR). The major difference among these estimators lies in the approach for approximating the Fisher information matrix.

The MLR is designed to be robust against non-normality and misspecification of the likelihood. Simulation studies have suggested that MLR standard errors perform slightly better than those of ML, and the standard errors of ML perform better than those of MLF (for details see Technical Appendix 8 of *Mplus*; Muthén & Muthén, 1998-2007). In this application, the GPGMM parameters were estimated in *Mplus* 5.21 with the default MLR estimation, since it is designed to model the

potential population non-normality due to the potentially unknown subpopulations. We also adopted the default number of 15 numerical integration points (Muthén & Muthén, 1998-2007), because increasing the integration points can substantially increase the time for estimating a complex model like the GPGMM.

When a mixture model is specified, *Mplus* uses random starts to guard against local maxima. The default starting values are automatically generated values that are used to create randomly perturbed sets of starting values for all parameters in the model except variances and covariances. Throughout the analyses, the number of initial stage random starts were, as a principle, first set to 1,000, and the final stage starts were set to 20 (e.g., the syntax reads `STARTS = 1,000 20`). If the log-likelihood values were not replicated as reported in the final 20 solutions, the number of the initial random starts was increased until the log-likelihood was replicated at least twice. For all analyses, the initial stage iterations are set to 200 and the maximum number of iterations for the EM algorithm was set to 3,000.

To speed up the estimation, *Mplus* allows user-specified starting values. In this application, four strategies were considered for specifying the starting values. The first and simplest strategy was to specify some or all of the starting values to zeros; this would significantly reduce the computing time. The second strategy was to use descriptive statistics obtained from the given data or reported in the literature (e.g., the mean of the WRAT-3 at the kindergarten year as the starting value of the intercept growth factor). The third strategy was to estimate a multi-class model with the variances and covariances of the growth factors fixed at zero. The estimates of the growth factor means from this analysis were then used as the starting values in the analysis where the growth factor variances and covariances were freely estimated. The fourth strategy was to use the estimates from a simpler model as the starting values for a more complex model. For example, the estimated means of the growth factors from the 1-class model were used for the 2-class model or the growth factor means estimated from the unconditional piecewise GMM were used as the starting values for the conditional piecewise

GMM. These methods for specifying starting values were used interchangeably and in concert to help the model estimation.

In the demonstration, the quality of a GMM model was assessed by several fit statistics and two alternative likelihood ratio tests (LRT). The conventional test of model fit based on the Chi-square likelihood ratio, comparing a compact model (K-1 classes) with an augmented model (K classes), does not function properly because it does not have the usual large-sample chi-square distribution. Two alternative likelihood-based tests have been developed to overcome this problem and have shown promise.

The first is the Lo, Mendell, and Rubin (2001) likelihood ratio test (LMR LRT; Lo, Mendell & Rubin, 2001; Nylund, Asparouhov, & Muthén, 2007). Assuming within class normality, this test proposes an approximation to the conventional distribution of likelihood ratio test and provides a *p*-value for testing K-1 classes against K class. A low *p*-value indicates that a K-1 class model has to be rejected in favor of a model with at least K classes. The second was the bootstrapped parametric likelihood ratio test (BLRT, described in McLachlan & Peel, 2000). As opposed to assuming that twice the difference between the two negative log-likelihoods follows a known distribution, the BLRT bootstraps samples to estimate the difference distribution based on the given data. The interpretation of the BLRT *p*-value is similar to that of the LMR LRT. Both LMR LRT and BLRT are available in *Mplus* in the Technical Output 11 and 14 respectively.

Another type of commonly used fit indices is the information criterion: Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Sample-Size Adjusted BIC (SBIC). These fit indices are scaled so that a small value corresponds to a better model with a large log-likelihood value and not too many parameters. The SBIC was found to give superior performance in a simulation study for latent class models in Yang (2006), and the BIC was found to give superior performance for mixture models including the GMM (Nylund, et al., 2007). Note that these indices do not address how well the model fit to

the data, but are relative fit measures comparing competing models.

When classification is a major modeling concern as with a GMM, the classifying quality is often assessed. Entropy assesses the degree to which the latent classes are clearly distinguishable by the data and the model. It is scaled to have a maximum value of 1 with a high value indicating a better classification quality. Entropy is calculated based on individual's estimated posterior probabilities of being in each of the  $K$  classes (analogous to factor scores in a factor analysis).

Consider each individual is classified into one of the  $K$  classes by the highest estimated posterior probability (i.e., most likely classes), entropy value will approach one if individuals' probabilities in the  $K$  classes approach one or zero, whereas the entropy value decreases if individuals' posterior probabilities of being in the  $K$  classes depart from zero or one (see Technical Appendix 8 for the calculation of Entropy (Muthén, 1998-2007; Clark & Muthén, 2009). In other words, Entropy reflects how much noise there is in the classification, hence, can be understood as an index for classification reliability.

There is not yet consensus upon the level of satisfactory entropy. Clark and Muthén (2009), in studying the effect of entropy on relating the latent classes to covariates, arbitrarily used the value of 0.8 as an indication of high entropy; thus, this is the minimum value that was used 0.8 for being considered as reliable class classification. All the aforementioned fit indices and LRTs were reported and examined in concert to choose an optimal number of classes.

Unconditional Piecewise GMM-Class Enumeration

The main modeling task of the unconditional model was to select the optimal growth trajectories and number of classes; recent simulation studies of mixture models have suggested that this unconditional model is the more reliable method for determining the number of classes is to run the class enumeration without the covariates. Class enumeration with covariates (i.e., the conditional model) could lead to poor decisions regarding the number of

classes, particularly when the entropy value is lower than 0.80. In some cases, researchers may not want the covariates to influence the determination of the class membership because the inclusion of covariates may potentially change the estimates of class distribution and growth factor means. For determining the number of classes using fit indices, recent simulation studies suggested that BIC performed best among the information criteria and BLRT was proved to be a consistent indicator for deciding on the number of classes (Chen & Kwok, 2009; Nylund-Gibson, 2009; Nylund, et al., 2007).

Following the suggestion of these current developments, the demonstration of the unconditional piecewise GMM herein was geared to optimize the number of classes while choosing a better-fitting growth function. If the fit indices point to inconsistent suggestions on the number of classes, BIC and BLRT will be used as the determinant rules. In addition, the first set of questions A1 through A3 were also addressed at this stage. Note that although the substantive research questions (A1- A3) were posed as distinct and sequential, the actual modeling was executed simultaneously in one single unconditional piecewise GMM. Also note that the variances and covariances structure of the growth factors was specified to the same across classes throughout class enumerations. This is because – when the class-specific variances are allowed – the likelihood function becomes unbound, and because when class-specific covariances between the growth factors are allowed, class separation and interpretation can be comprised.

Question A1: Are there distinct phases where children differ in their speed of learning?

A visual inspection of the observed data displayed in Figures 1 and 2 suggested that a 2-piece linear model summarizes the growth trend better than single linear and quadratic models. Figure 3 displays three latent continuous growth factors: (1) the intercept  $I$  representing the starting performance, (2)  $S1$  representing the first growth rate, and (3)  $S2$  representing the second growth rate. Two growth rate factors (i.e., two slopes), in contrast to the traditional one single linear growth rate, were specified to

## GPGMM WORD RECOGNITION

more aptly portray the two-phase growth pattern observed in the data. The two growth rates S1 and S2 depict the non-linear trend by assuming that, within each phase, the growth trajectory was linear. The three growth factors are indicated by part or all of the seven repeated word recognition scores from kindergarten to grade six as shown by the arrows going from the three growth factors in ovals to the seven word recognition scores in rectangles. Note that, at most, a 2-piece linear model was used because each piece requires a minimum of three waves of data, therefore a 3-piece model was not feasible with the study data which has 7 waves; a 3-piece model would require a minimum of 9 waves.

To specify the phasic trajectory, the loadings (i.e., time scores) of the seven measures must be fixed on the three growth factors using the coding scheme often seen in piecewise regression (See Table 2). For the starting performance, the loadings of the intercepts were all fixed at 1. In this demonstration, assuming a grade-2 transition, the loadings of the first growth phase from kindergarten to grade two were fixed at 0, 1, and 2, with an increment of one indicating a constant linear increase across each grade.

In the second phase, the first growth phase loading remained at 2 showing no incremental change to indicate no growth effect in the second phase. The loadings of the second growth phase were fixed at 1, 2, 3 and 4 from grade three to grade six with an increment of one indicating a constant linear increase across each grade. The loadings for S2 were all fixed at zero

with no increment indicating no growth effect in the first phase; note that the coding in Table 2 assumed a transition point at the end of grade two. The transition point should be justified by multiple sources of information, including the existing literature, the observed growth trend, the statistical model fit, and the interpretability of the results.

### Results for Question A1

First explored was which GMM – single linear, 2-piece linear, or quadratic – fit better by comparing the fit indices. Table 3 shows that the 2-piece models yielded better fit indices than the quadratic models, which in turn fit better than the single linear models irrespective of the number of classes. This indicates that fitting a 2-phase model not only captured the observed non-linearity better than a model merely ignoring the non-linearity but also did better than the commonly used quadratic model. (The default specification for estimation can be found in Chapter 13 of the *Mplus* User’s Guide.)

The transition point dissecting the two phases was specified at the end of grade two; this decision was made for three reasons. First, the observed pattern shown in Figures 1 and 2 indicated that the transition point occurred at either grade two or grade three. Second, Speece, et al. (2003) studied children from kindergarten to grade three and detected a non-linearity; this suggests that a turning point before grade three was necessary. Also, Francis, et al. (1996) argued that reading difficulty could not be defined at grade one or grade two because identifying reading difficulty often over-

Table 2: Codes for 2-piece Linear Growth Model with Seven Wave of Data

	First Phase			Second Phase			
	K	G1	G2	G3	G4	G5	G6
I	1	1	1	1	1	1	1
S1	0	1	2	2	2	2	2
S2	0	0	0	1	2	3	4

identifies children who have not had adequate educational exposure to reading and under-identifies children who demonstrate deficits in cognitive/linguistic skills. Third, the 2-piece grade-3 transition model yielded negative estimates of growth factor variances, which was undefined and counterintuitive, even with numerous trials of changing the starting values and increasing the number of starting sets up to 1,000 (although the log-likelihood values were replicated). This problem suggested the possibility of an incorrect model. The grade-2 transition models did not show these problems.

Question A2: Are there unknown subpopulations (latent classes) that differ in their growth pattern?

Different learners may display different learning patterns in their reading development. When different groups of learners are empirically observed or theoretically hypothesized, a statistical model must be able to aptly attend to this heterogeneity. Modeling population heterogeneity in growth trajectory is often carried out by a GMM.

GMM is the bedrock of a fully developed GPGMM. GMM relaxes the single population assumption to allow for differences in growth factors across unobserved subpopulations (Kreuter & Muthén, 2007; Muthén, 2004). This flexibility in identifying unobserved subpopulations of people (a.k.a., classes), who are distinct in the developmental pathways, is the cornerstone of the GMM model. The unobserved class membership is modeled by a latent categorical variable where individuals' developmental pathways are relatively similar within each class, yet distinct from one another across classes. As opposed to assuming that individuals vary around a single mean growth trajectory, GMM allows separate mean growth trajectories for each class. Individuals in each class are allowed to vary around the class mean of the growth factors. The variable C in Figure 3 represents such a categorical latent trajectory class.

Results for Question A2

Table 3 compares the results of the 2-piece grade-2 transition models with the number of classes ranging from two to six (rows in

bold). The 4-class model was supported by the BIC, the 5-class model was supported by the SBIC, LMR, LRT and BLRT, and the 6-class model was supported by the AIC. With the exception of the 6-class model, all models yielded high entropy values of greater than 0.8. The fit indices point to fairly inconsistent suggestions about the optimal number of classes to extract. The 6-class model was first eliminated from further consideration because it yielded an entropy value lower than 0.80 and because it was suggested only by AIC, which has been shown to be poorer criterion for choosing the correct number of classes (Nylund-Gibson, 2009).

The 4- and 5-class models were each supported by the determinant rule, BIC and BLRT, respectively. To compare the similarities and differences between the 4- and 5-class models, their growth factor means were tabulated and graphed on the first and second panel separately (see Figure 4). Figure 4 shows that the C3 class of the 4-class model branched into two classes of C3a and C3b resulting in 5 classes in total. The 4- and 5-class models were not entirely distinct models; how elaborate the class classification was their main difference. This phenomenon is also common in factor analyses where a model with a greater number of factors is often a more elaborate version of a model with a fewer number of factors.

For demonstrative purposes, it was necessary to select a model with which the conditional piecewise GMM could be demonstrated. In the trial runs of the conditional piecewise GMM, the 5-class model experienced a problem of non-identification and the problem that the log-likelihood could not be replicated – even with the number of starts increased to 10,000 and the assistance of user-specified starting values. For these reasons, the 5-class model was eliminated from further consideration.

The 4-class model, which still revealed a rich substantive story, was chosen because it yielded the smallest value of BIC, which has been shown to be superior in choosing the correct number of classes for GMMs. For real research contexts, choosing the number of classes to extract is not a simple technical task: a researcher must consider multiple factors such



## GPGMM WORD RECOGNITION

as the research purpose, statistical fit and the substantive and practical gains that different numbers of classes may bring about.

Question A3: How are the starting performance and growth rates related?

As discussed in the literature review, there has been a great deal of theoretical and practical interest in whether children with a better start will continue to learn faster and

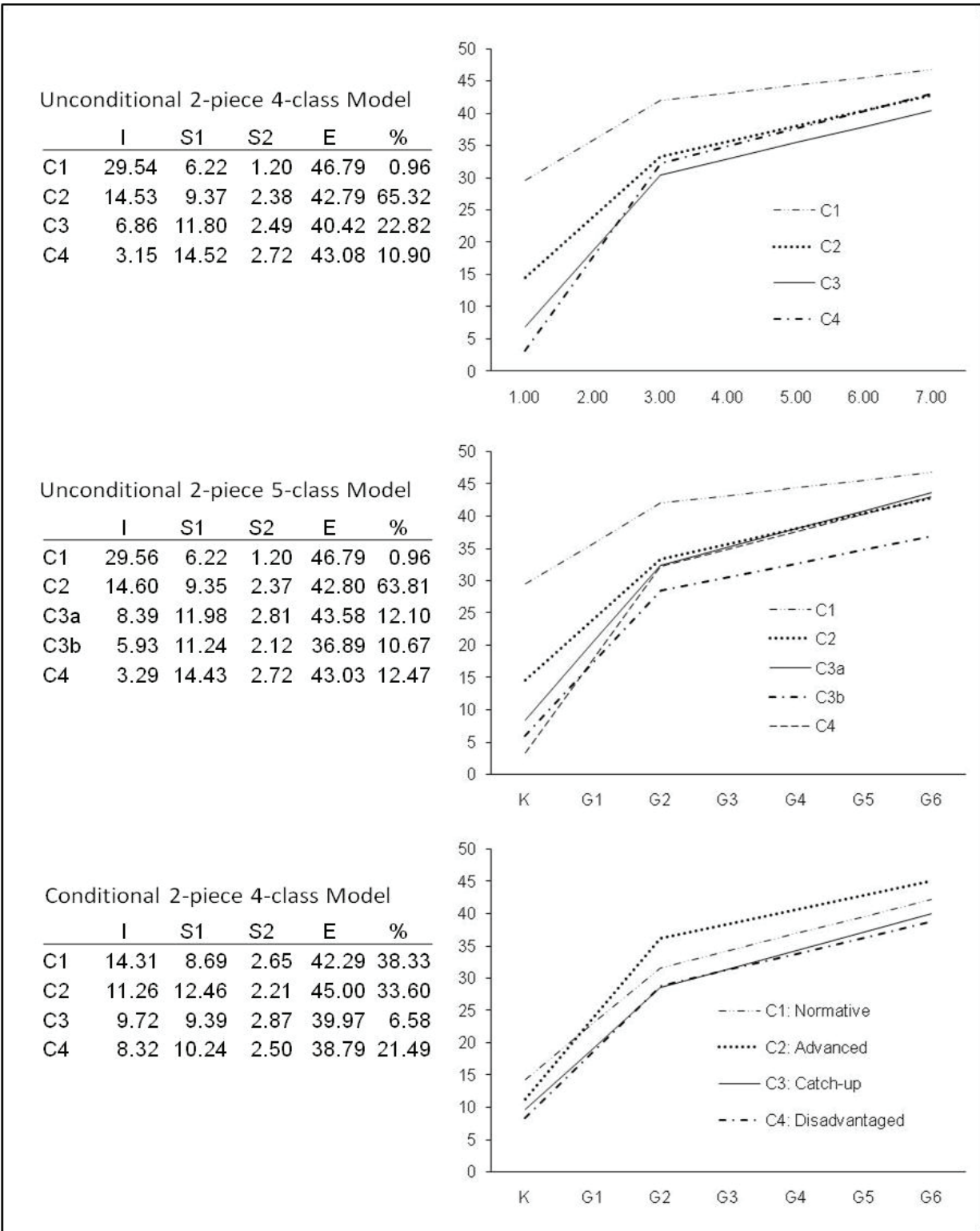
whether children who learn faster at an early age will continue to improve at a faster rate. The field of children’s reading development has not settled the debate over how earlier reading performance is related to the later development. In a GMM, these questions are answered by estimating the covariances among the growth factors, I, S1, and S2. These relationships are indicated by the curved arrows among the growth factor I, S1, and S2 in Figure 3.

Table 3: Fit Indices for Single Linear, 2-piece Linear and Quadratic Unconditional Models with 2-6 Classes

		df	L	AIC	BIC	SBIC	LMR LRT	BLRT	Entropy
Single Linear	2-class	15	-10915.215	21856.430	21911.879	21870.613	p=0.169	p=1.000	0.729
	3-class	18	-10996.744	22029.488	22106.399	22049.262	P=0.151	p=0.030	0.489
	4-class	21	-10904.422	21846.844	21927.884	21867.573	P=0.118	p=0.500	0.668
	5-class	25	-10903.928	21851.855	21945.692	21875.858	P=0.365	p=0.800	0.709
2-piece Linear	<b>2-class</b>	<b>20</b>	<b>-10015.847</b>	<b>20071.695</b>	<b>20157.001</b>	<b>20093.516</b>	<b>p= 0.349</b>	<b>p&lt; 0.001</b>	<b>0.811</b>
	<b>3-class</b>	<b>24</b>	<b>-9960.502</b>	<b>19969.004</b>	<b>20071.372</b>	<b>19995.189</b>	<b>p&lt; 0.001</b>	<b>p&lt; 0.001</b>	<b>0.931</b>
	<b>4-class</b>	<b>28</b>	<b>-9946.927</b>	<b>19949.854</b>	<u>20069.282</u>	<b>19980.403</b>	<b>p= 0.065</b>	<b>p&lt; 0.001</b>	<b>0.880</b>
	<b>5-class</b>	<b>32</b>	<b>-9936.602</b>	<b>19937.204</b>	<b>20073.693</b>	<u>19972.117</u>	<b>p= 0.043</b>	<b>p= 0.040</b>	<b>0.850</b>
	<b>6-class</b>	<b>36</b>	<b>-9930.838</b>	<u>19933.675</u>	<b>20087.226</b>	<b>19972.953</b>	<b>p= 0.692</b>	<b>p= 0.140</b>	<b>0.777</b>
Quadratic	2-class	20	-10227.122	20484.243	20548.223	20500.609	p=0.047	p=1.000	0.682
	3-class	24	-10200.583	20439.166	20520.206	20459.895	p=0.005	p=0.600	0.568
	4-class	28	-10249.835	20543.670	20637.507	20567.673	p=0.099	p<0.001	0.818
	5-class	32	-10211.386	20474.773	20585.671	20503.140	p<0.016	p<0.001	0.823

Notes: df: the number of free parameters of a specified model (when no parameters were fixing to zeros); L: log-likelihood; AIC: Akaike information criterion; BIC: Bayesian information criterion; SBIC: Sample size adjusted BIC; LRT: Lo-Mendell-Rubin adjusted likelihood ratio test; BLRT: bootstrapped parametric likelihood ratio test. The fit indices for the 2-piece models are in bold. The model with the lowest AIC, BIC, or SBIC is underlined. The model with K classes is underlined if the p-value of the LMR LRT or BLRT for the K+1 model was greater than 0.05. When the variance of a growth factor was estimated to be negative, the estimation proceeded with fixing it to zero.

Figure 4: Comparison among the 4- and 5-Class Unconditional and 4-Class Conditional Models



Notes: I denotes the initial performance at the kindergarten year, S1 denotes the growth rate in the first phase, S2 denotes growth rate in the second phase, E denotes the ending performance at grade six, and % denotes the proportions for the latent classes.

## Results for Question A3

By default, *Mplus* outputs estimate the covariances of growth factors. For interpretation ease, however, the growth factor correlations were reported by requesting the standardized command in the output. Recall that, for estimation reasons, the covariance structure was fixed to be the same across classes, that is, class-specific correlations among the growth factors were not allowed.

Based on the 4-class unconditional piecewise model, results show that the initial performance was not significantly correlated with the first growth rate ( $r = -0.06, p = 0.762$ ), nor was it significantly correlated with the second growth rate ( $r = -0.345, p = 0.053$ ). However, the two growth rates were significantly and negatively correlated ( $r = -0.599, p = 0.001$ ).

These findings suggest that word recognition performance at the beginning of the kindergarten year, as measured by WRAT-3, was not a good indicator of children's later speed of learning. However, the speed of learning in the first phase may be associated with children's development in the second phase. This suggests that early identification of advanced or disadvantaged children should not rely solely on children's starting performance. Rather, early identification of advanced or disadvantaged children should also look into children's early speed of learning. If a single linear trajectory had been modeled, the relationship between two growth rates would have been overlooked.

## Conditional Piecewise GMM with an Auxiliary Developmental Outcome Variable

The conditional piecewise GMM is the full version of GPGMM. It incorporates the covariates and an auxiliary developmental outcome variable into the unconditional piecewise GMM. In this demonstration, five covariates were included. Three were cognitive/linguistic variables that were measured prior to the first assessment of word recognition in the kindergarten year and were standardized scores of verbal working memory, phonological awareness, and word retrieval time. The other two were demographic background variables: gender (boy = 0; girl = 1, 50.2%) and first

language reported in the fall of kindergarten year (English = 0, ESL = 1, 15%).

Covariates can have direct and indirect effects on the growth factors. As shown in Figure 3, direct covariate effects explain the growth factor variations, as indicated by the arrow going from the covariates to the growth factors I, S1, and S2. Covariates can also have an indirect effect on the growth factors via their effects on the latent class as indicated by the arrow going from the covariates to the latent class C and then to the growth factors (see Figure 3). The developmental reading outcome variable used in this demonstration was the Stanford Diagnostic Reading Test (SDRT; Karlesen, Madden, & Gardener, 1966) measured at grade six. This developmental outcome variable served as an auxiliary dependent variable for checking the latent class validity, and was standardized for ease of interpretation.

Estimates of class distribution and growth factors means will change as a result of incorporating covariates information and how their effects are specified. Misspecification of the direct and/or indirect effects can lead to untrustworthy estimates. Because the correct population model is unknown and "all models are wrong, the practical question is how wrong they have to be to not be useful" (Box & Draper, 1987, p. 74), it is recommended that researchers experiment with various models. The choice of which model to select must rely heavily on the researchers' discretion borne on the model results (e.g., whether a model terminated normally) as well as their substantive knowledge, and common sense (e.g., checking the tenability of the direction and size of the covariate effects).

In this demonstration, all direct and indirect effects were first allowed on the growth factors for all classes. This model was not identified and the best log-likelihood value was not replicated after numerous trials of starting values and the number of starting values sets = 10,000. Based on the estimated posterior probabilities, this model distributed two very small classes ( $\approx 1\%$  and  $\approx 5\%$ ; size  $\approx 5$  and 26), leading to some difficulties in estimating the direct effects on the growth factors (e.g., empty cells in the joint distribution of the model variables). For these reasons, the direct covariate

effects of the two small classes were fixed to zeros. This model terminated normally, the log-likelihood values were replicated 3 times with STARTS = 1000 20 (the other 17 differed with their next best values only in the third decimal place). In addition, this model estimated 92 parameters with log-likelihood value = -9693.008. The information criteria of AIC = 19570.016, BIC = 19962.424, SBIC = 19670.392 all (except entropy = 0.768) suggested that this conditional model was a better fitting model than the unconditional model (compare the fit indices of the 2-piece 4-class unconditional model in Table 3).

Note that various models that allowed partial direct effects for the two smaller classes were also examined. Although the log-likelihood values of some of these models were replicated, their the BIC values were all greater than that of the model finally selected and their growth trajectories were harder to recognize and interpret, for these reasons they were not chosen and reported.

The bottom panel of Figure 4 shows the trajectories and estimated growth factor means of the selected conditional model. Two overall observations are pointed out here. First, a noticeable parameter shift was observed when being compared to the unconditional model. The cross-class differences in the growth trajectories diminished a great deal as a result of incorporating the covariates information. Second, it was observed that there was little difference in the estimates of the second growth rates (which only differed in the first decimal place). This result may be a true reflection of the small differences in the speed of recognizing new words among the subpopulations.

The potential ceiling effect of WRAT-3 on the lack of variation in the second growth rate should be considered, however. This instrument lists 42 words for recognition ordered in difficulty and was not originally designed for children. WRAT-3 is known to have a strong ceiling effect (Strauss, Sherman & Spreen, 2006, p. 388). The difficulty level elevates quickly as the words approach the end the list leading to few or no successes in word recognition. This ceiling effect may explain the small class differences in the second growth rates for the present child sample.

The results for specific classes (see the bottom table and graph in Figure 4) indicated that, on average, children in the first class recognized 14.31 words in the kindergarten year, learned 8.69 words per year in the first phase, and 2.65 words per year in the second phase with an ending performance (E) of 42.29 words. This class was referred to the normative class because it consisted of 38.33%, the largest proportion, of the sample, and because its growth trajectory was relatively more typical than those of the other classes.

Children in the second class (33.60%) initially recognized 3.05 fewer words on WRAT-3 than did the normative class. However, their first phase growth rate was 3.77 words faster than the normative class leading to a projection that this class would surpass the normative class at grade one and they would manage to stay ahead of the normative class thereafter despite the slightly slower second growth rate. This class was referred to as the advanced class.

Children in the third class (6.58%) initially recognized 4.59 fewer words on WRAT-3 than did the normative class, but were slowly catching up with the normative class with 0.70 words per year in the first phase and 0.22 words per year in the second phase. This class was referred to as the catch-up class.

The fourth class (21.49%) began with the lowest performance; initially recognizing 5.99 words fewer on WRAT-3 than did the normative class. Although this class appeared to catch up with the normative class with 1.55 words per year in the first phase, they slowed to a rate of 0.15 words per year slower than the normative class in the second phase. In grade six, they recognized 3.50 fewer words than did the normative class and 6.21 fewer words than did the advanced class. This class was referred to as the disadvantaged class.

Question B1: What are the characteristics of the latent classes?

Similar to factors in a factor analysis, the latent trajectory classes do not have inherent meanings (Bauer & Curran, 2003; Kreuter & Muthén, 2007; Muthén, 2003; Muthén, 2004). To understand the characteristics of the latent classes, the categorical latent class variable C is

regressed on to the covariates. Covariates play important roles in the GPGMM – they can aid in checking the interpretability and trustworthiness of the selected model. If classes are not statistically different with respect to the covariates, which theoretically or logically should characterize the classes, then there is weak support for the selected model. In Figure 3, the class characteristics regression is shown by the arrow going from the covariates to the latent trajectory class C. Recall that class characterization by covariates can have indirect effects on the growth factors.

Characterization of the latent classes by the set of covariates involves a multinomial logistic regression (or a binary logistic regression for the 2-class case). Coefficients of the covariates in a multinomial logistic regression are linear in the logit form; the increase in the log odds of being in a particular class versus the reference class. The reference class is normally the class with the largest size or the class or that is better recognized by the researchers. The exponent of the slope coefficient,  $\text{Exp}(\text{slope})$ , is the odds ratio for being in one particular class versus the reference class. For example, when comparing ESL (coded as 1) to non-ESL children (coded as 0); a slope = 1 implies that the odds of being in one particular class versus the reference class is  $\text{Exp}(1) = 2.72$  times higher for ESL children than non-ESL children.

#### Results for Question B1

To understand the characteristics of the classes, the results of the multinomial logistic regression were reported and interpreted using the normative class as the reference class. The normative class was used because of its estimated largest proportion and better-known growth pattern. Table 4 reports the slope coefficients (i.e., partial regression coefficient) for the five covariates and their corresponding standard errors and odds ratios. Bear in mind that the interpretation of the odds was based on per one unit change in the covariate. Because the ESL and gender variables were both coded as 0 and 1, their odds reflected the gender and first language differences, and because the cognitive/linguistic variables were all

standardized, their odds reflected per SD change.

Relative to the normative class, the odds of membership in the advanced class were significantly increased by being boys. The odds of being in the advanced class versus the normative class were 2.667 ( $1/\text{odds} = 1/0.375$ ) times higher for boys than girls. Relative to the normative class, the odds of membership in the catch-up class were significantly increased by word retrieval time. The odds of being in the catch-up class versus in the normative class were 4.289 times higher per SD increase in word retrieval time. Relative to the normative class, the odds of membership in the disadvantaged class were significantly increased by being a boy, being non-ESL, having poorer phonological awareness and longer retrieval time. The odds of being in the disadvantaged class versus the normative class were 3.106 ( $1/0.322$ ) times higher for boys than girls, 3.497 ( $1/0.286$ ) times higher for non-ESL students than ESL students, 2.165 ( $1/0.462$ ) times higher per SD decrease in phonological awareness, and 1.725 times higher per SD increase in word retrieval time.

Question B2: For each class, what explains children's starting performance and growth rates?

In a GPGMM, the growth factors' variations can also be explained by the same set of covariates. This relationship is analogous to a multiple regression where each of the continuous dependent variables, I, S1 and S2, is regressed onto the covariates. This relationship models the direct effects of the covariates on the growth factors as indicated by the arrows going from the covariates directly to the growth factors. As aforementioned, in the final model only direct effects were allowed on the two classes with larger class proportion, that is, the normative and advanced classes. Note that the indirect effect of covariates on the growth factors via the latent classes, as demonstrated in B2, is reflected by allowing class-varying regression coefficients of the covariates on the growth factors. Thus, the class-varying residual variances were also allowed.

Table 4: Indirect Covariate Effects: Multinomial Logistic Regression for Classes Characterization

	A vs. N	C vs. N	D vs. N
Gender (Girl)	<b>-0.981</b> <i>0.303</i> <u>0.375</u>	0.060 <i>0.810</i> <u>1.062</u>	<b>-1.134</b> <i>0.314</i> <u>0.322</u>
ESL (Yes)	0.633 <i>0.389</i> <u>1.833</u>	-0.442 <i>0.842</i> <u>0.643</u>	<b>-1.252</b> <i>0.633</i> <u>0.286</u>
Verbal Working Memory	0.183 <i>0.204</i> <u>1.201</u>	-0.133 <i>0.585</i> <u>0.875</u>	-0.083 <i>0.229</i> <u>0.920</u>
Phonological Awareness	-0.120 <i>0.187</i> <u>0.887</u>	-0.909 <i>0.502</i> <u>0.403</u>	<b>-0.773</b> <i>0.175</i> <u>0.462</u>
Word Retrieval Time	-0.077 <i>0.245</i> <u>0.926</u>	<b>1.456</b> <i>0.374</i> <u>4.289</u>	<b>0.545</b> <i>0.231</i> <u>1.725</u>

Notes: A: advanced class; N: normative class; D: disadvantaged class; C: catch-up class. Values in the first row of each covariate were the estimates of the slope coefficient, of which the standard errors were placed underneath in italic face, and the corresponding odds ratios were underlined. Significant slope coefficients at the 0.05 level were highlighted in bold face.

Results for Question B2

Results for the class-specific multiple regressions are shown in Table 5. The first row for each covariate reports the estimates of the slope coefficient (partial regression coefficient) and their standard errors were placed underneath in italic face. Significant slope estimates at  $\alpha=0.05$  level were highlighted in bold. For example, phonological awareness had a significant effect on all growth factors, except for the second growth rate of the normative class. Differential covariate effects in terms of size and direction were found across classes. For example, the initial growth factor I, gender and verbal working memory had significant effects only for the normative class, and word retrieval time had an effect only for the advanced class. Useful substantive information is revealed by comparing differential covariate effects across classes.

Question B3: Do the latent classes differ in the reading developmental outcome?

The GPGMM incorporates an auxiliary outcome variable that can either be proximal or distal. Note that this outcome variable is an auxiliary variable; it is not modeled as an observed dependent variable, nor was it part of the model. Its major role in a GPGMM is to assist in checking the validity of the latent classes by comparing and testing equalities in the class means of this variable (Masyn, 2009; Petras & Masyn, 2009). Because it is an auxiliary variable, the outcome variable is represented in Figure 3 as a dashed square to show that it not an actually modeled outcome variable. This part of the modeling is shown by the arrow going from the latent class variable to the reading comprehension outcome. Cross-class equality in the means of the reading comprehension was tested using the posterior

GPGMM WORD RECOGNITION

Table 5: Direct Covariate Effects: Class-specific Multiple Regression of Growth Factors

	Normative			Advanced		
	I	S1	S2	I	S1	S2
Gender (Girl)	<b>0.565</b> <i>0.224</i>	-0.009 <i>0.483</i>	-0.098 <i>0.197</i>	0.713 <i>0.867</i>	-0.611 <i>0.493</i>	-0.011 <i>0.161</i>
ESL (Yes)	0.190 <i>0.272</i>	<b>0.971</b> <i>0.449</i>	-0.377 <i>0.237</i>	-0.550 <i>1.125</i>	-0.813 <i>0.566</i>	0.290 <i>0.219</i>
Verbal working memory	<b>-0.258</b> <i>0.132</i>	0.282 <i>0.226</i>	-0.107 <i>0.115</i>	-0.676 <i>0.451</i>	0.148 <i>0.148</i>	0.130 <i>0.094</i>
Phonological awareness	<b>0.482</b> <i>0.120</i>	<b>0.553</b> <i>0.234</i>	-0.147 <i>0.084</i>	<b>4.167</b> <i>0.455</i>	<b>-1.345</b> <i>0.281</i>	<b>-0.300</b> <i>0.083</i>
Word retrieval time	-0.240 <i>0.234</i>	-0.183 <i>0.245</i>	0.146 <i>0.141</i>	<b>-1.714</b> <i>0.495</i>	<b>0.692</b> <i>0.289</i>	0.119 <i>0.114</i>

Notes: Values in the first row of each covariate were the estimates of the slope coefficient, of which the standard errors were placed underneath in italic face. Significant slope coefficients at the 0.05 level were highlighted in bold face.

probability-based multiple imputations method. Since the class means of the reading comprehension were not part of the models, *Mplus* needed to estimate means and their asymptotic variances/covariance using the pseudo-class draw technique (Wong, Brown & Bandeen-Roche, 2005). First, individuals' posterior class probabilities (conditional on the model and the observed data) were computed. Next, using this posterior distribution, L pseudo draws were generated for the latent class variable C for individuals – L denotes the number of pseudo draws. Class-specific sample means of the outcome variable then were obtained by taking the average of the L pseudo draws (see *Mplus* technical note at <http://www.statmodel.com/download/MeanTest1.pdf>).

As recommended in Wong, et al. (2005), the *Mplus* default number of pseudo draws of 20 was adopted. Equality in the class means were tested using Wald's Chi-square with degree of freedom = K-1 for the omnibus test and 1 degree of freedom for the pairwise tests; a statistically and theoretically/ practically

significant and meaningful mean difference should be detected for supporting the validity of the latent class variable. This validity check is analogous to the criterion validity in the traditional psychometrics literature.

Results for Question B3

The last two rows of Table 6 show the class means in the reading comprehension and their corresponding standard errors. First, note that the order of the size of the estimated class means were as expected (i.e., Advanced > Normative > Catch-up > Disadvantaged). The omnibus Wald  $\chi^2(3) = 80.094, p < 0.001$ . The Chi-square values for the paired tests were shown on the upper diagonal of the class matrix in Table 6, and their corresponding *p*-values were shown underneath in italic face. Significant mean differences, highlighted in bold, were found in four of the six paired tests. Mean differences between all non-neighboring classes were all found to be significant (e.g., the difference between the advanced and disadvantaged classes). Mean differences between two of the neighboring classes were found to be non-significant (differences between

the advanced and the normative classes and between the catch-up and disadvantaged classes). Judging by the order and size of the class mean estimates and the pattern of the significance tests, the results provided adequate criterion validity evidence for the latent trajectory class variable.

Conclusion

People learn and develop in different ways at different times. These developmental complexities and diversities are often overlooked or modeling tools are incapable of revealing them. This study demonstrated, with children’s word recognition development, that by taking into account the phasic learning speed and population heterogeneity, the GPGMM is able to point up evidence for both the deficit and lagging theoretical models reported in literature depending on which classes and developmental phases are being compared.

The advantages of the GPGMM, however, come with a price. To find the optimal model that is both statistically adequate and theoretically interpretable, the GPGMM requires fairly sophisticated modeling techniques that involve iterative and intricate trials of parameter specifications. For a complex model like the GPGMM, the parameter estimates can change remarkably in size and direction from one start set to another. Users are reminded to increase

the number of iterations and starting sets when necessary so as to ensure that the log-likelihood of the selected model is replicated. Also, due to the model complexity, the time taken for the estimation to terminate can be much longer than what is needed for simpler models. This is particularly the case when the random start sets are increased to a large number or when the bootstrapped likelihood ratio test is requested. It is suggested that, wherever possible, the GPGMM be run on a spare but fast computer.

To date there is no single agreed-upon best practice for choosing the optimal conditional model. The general statistical problem of choosing the optimal conditional model in latent class models shares a conceptual core in common with indeterminacy problems in factor analysis – note that there are several indeterminacies in factor analysis; for example, indeterminacy of common factors, and an indeterminacy in factor rotation. There may be something to be gained by noting this commonality between latent class and factor analysis. At this point, it is advisable that the unconditional model be used for class enumeration – i.e., for deciding the number of classes. Like the indeterminacy problem of factor rotation, estimates of class distribution and the growth factors of the conditional model may shift from those of the unconditional model

Table 6: Wald’s Chi-square Tests of Class Equality in the Means of the Reading Development Outcome

	A	N	C	D
A		0.049 <i>0.826</i>	17.688 <b>&lt;0.001</b>	72.170 <b>&lt;0.001</b>
N			16.377 <b>&lt;0.001</b>	60.159 <b>&lt;0.001</b>
C				0.161 <i>0.688</i>
M	0.273	0.250	-0.605	-0.694
SE	0.066	0.076	0.196	0.092

Notes: A: advanced class; N: normative class; D: disadvantaged class; C: catch-up class. The Chi-square values for the paired tests were shown on the upper diagonal of the class matrix; the corresponding *p*-values were shown underneath in italic face. Significant *p*-values were highlighted in bold. The class means and their corresponding standard errors were shown in the last two rows.



depending on how the direct and indirect effects are specified. Recent work by Nylund-Gibson (2009) suggests that first the indirect effects be added to the unconditional model followed by the direct effects. The parameter shift then is examined throughout the steps. Implicitly, this suggestion was used along with verifying the substantive interpretability, as a rough guide for checking and selecting a conditional model.

An intuitive, yet less than ideal solution, is to fix the growth factor parameters of the conditional model to those estimated by the unconditional model. By doing so, the covariate effects can be investigated without shifting the growth factor parameter estimates. This method could be problematic because it treats the fixed parameters as if they were true population values and ignore the sampling errors. Moreover, using the estimates of the unconditional model for the conditional model may be considered as cheating because both models are based on the same sample set. Hence, this strategy is not recommended if the purpose of the GPGMM is of an exploratory nature as demonstrated in this study. It may be more justified if the purpose is to cross-validate, that is, to verify growth trajectories suggested by other samples from the same population.

Traditionally, questions B1, B2, and B3, as addressed by the conditional model, are often answered by saving the likely class membership or the posterior probabilities for each individual in a new data file and running separate analyses. This method could also be problematic because the class membership or the posterior probabilities are treated as being observed, but they are, in fact, model estimates with errors. Recent studies have shown that these traditional approaches may yield incorrect parameter estimates and standard errors leading to incorrect conclusions about significance testing (Clark & Muthén, 2009; Masyn, 2009; Petras & Masyn, 2009). Answering these questions using a single GPGMM circumvents this problem, especially when the entropy is high (Clark & Muthén, 2009).

A trade-off between the number of classes extracted and the amount of variance of the growth factors (or residual variance after adding the covariates) was noticed. This phenomenon makes sense conceptually and

statistically because the mechanism behind the GMM is to extract  $K$  classes where people are relatively similar within each class, yet distinct from one another across classes.

In a highly hypothetical situation where  $K$  is equal to the sample size, there will be no within-class variation in the growth factors. The 4- and 5- class conditional models encountered scenarios where the variances and/or residual variances of the growth factors being estimated were negative and received warning messages such as non-positive definite latent variable covariance matrix. Fixing the negative residual variances to zero may solve these problems, however, these problems may be indicative of class over-extraction or misspecification of the covariate effects – this is conceptually similar to a Heywood case in factor analysis.

The balance between number of classes and the within-class variances/residual variances often dictates the number of classes one is able to interpret, especially for the conditional model. The maximum number of interpretable classes is often bounded by how much variance the growth factors are estimated to have and whether the variance is sufficient for the conditional model. Using the study data, difficulty in identifying the 5-class conditional model was experienced, although it is preferred for more richness in the substantive information.

With a full GPGMM, a large number of parameters are simultaneously estimated and the number of parameter estimates increases rapidly in multiples as the number of classes and covariates increase. The large set of the parameters is deemed to be the best solution for the data simply because it yields the least -2 log-likelihood value. The maximum likelihood algorithm cannot tell whether or not the parameter estimates, in term of size and direction, make sense for a real and specific research context. Valid interpretations of the GPGMM results rely heavily on the users' methodological and substantive knowledge of the study. This demonstration showed that the speed of learning new words slowed down in the second phase for all classes; however, it would be inappropriate to conclude that children learn fewer words annually after grade two than before grade two without some special caution. As mentioned, this finding may result from the

low floor effect but strong ceiling effect of the WRAT-3. As stated by Muthén (2003) and stressed throughout this article, a quality GPGMM should be guided not only by the statistical information, but also by the substantive interpretability of the results. GPGMM, in essence, is merely an analytical tool. Substantive expertise throughout the process of model specification, selection, and verification is the key to the success of a GPGMM.

#### References

- Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing, 14*, 61-89.
- Aunola, K., Leskinen, E., Onatsuarvilommi, T., & Nurmi, J-E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology, 72*, 343-64.
- Bast, J., & Reitsma, P. (1998). Analyzing the development of individual difference in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34*, 1373-1399.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumption of growth mixture models: Implications for over extraction of latent trajectory classes. *Psychological Methods, 8*, 338-63.
- Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*, 192-208.
- Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: Wiley.
- Chen, Q., & Kwok, O. (2009). Can the correct number of classes be detected in misspecified multilevel growth mixture models? A Monte Carlo study. Paper presented at the 2009 Annual Meeting of American Educational Research Association (AERA), San Diego, CA.
- Clark, S., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. *Submitted for publication*.
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., & Shaywitz, B. (1996). Development lag versus deficit models for reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology, 88*, 3-17.
- Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Finding of two-year longitudinal study. *Journal of Educational Psychology, 94*, 586-97.
- Karlesen, B., Madden, R. and Gardener, E.F. (1966). *Stanford diagnostic reading test*. New York: Harcourt, Brace & World.
- Khoo, S. T. (1997). *Assessing interaction between program effects and baseline: A latent curve approach*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Kreuter, F. & Muthén, B. (2007). Longitudinal modeling of population heterogeneity: Methodological challenges to the analysis of empirically derived criminal trajectory profiles. In *Advances in latent variable mixture models*, G. R. Hancock, & K. M. Samuelsen (Eds.). Charlotte, NC: Information Age Publishing.
- Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J-E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly, 39*, 72-93.
- Li, F., Duncan T. E., Duncan, S. C., & Hops, H. (2001). Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling, 8*, 175-204.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767-78.
- Masyn, K. E. (2009). *The consequences of (latent) classes: Results of a simulation study evaluating approaches to specifying and testing the effect of the latent class membership on distal outcomes*. Paper presented at the 2009 Annual Meeting of American Educational Research Association (AERA), San Diego, CA.

- McGee, V. E., & Carleton, W. T. (1970). Piecewise regression. *Journal of the American Statistical Association*, *65*, 1109-1124.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.
- Muthén, B.O. (1998-2004). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide (5<sup>th</sup> Ed.)*. Los Angeles, CA: Muthén & Muthén.
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using EM algorithm. *Biometrics*, *55*, 463-9.
- Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882-891.
- Muthén, B. (2001). Second generation structural equation modeling with a combination of categorical and continuous latent variable: New opportunities for latent class/latent growth modeling. In *New methods for analysis of change*, L. M. Collins & Sayer (Eds.), 291-232. Washington, DC: APA.
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 369-77.
- Muthén, B., Khoo, S. T., Francis, D., & Kim Boscardin, C. (2003). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In *Multilevel modeling: Methodological advances, issues and applications*, S.R. Reise & N. Duan (Eds.), 71-89. Mahaw, NJ: Lawrence Erlbaum Associates.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In *Handbook of quantitative methodology for the social sciences*, D. Kaplan (Ed.), 345-368. Newbury Park, CA: Sage Publications.
- Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In *Advances in latent variable mixture models*, G.R. Hancock & K. M. Samuelsen (Eds.), 1-24. Charlotte, NC: Information Age Publishing.
- Muthén, B., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In *Longitudinal data analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), 143-165. Boca Raton, FL: Chapman & Hall/CRC Press.
- Nylund, K. L., Asparouhov, T., Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535-569.
- Nylund-Gibson, K. L. (2009). *Covariates and mixture modeling: Results of a simulation studying exploring the impact of misspecified covariate effects*. Paper presented at the 2009 Annual Meeting of American Educational Research Association (AERA), San Diego, CA.
- Parrila, R., Aunola, K., Leskinen, E., Nurmi, J-E, & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finish. *Journal of Educational Psychology*, *97*, 299-319.
- Petras, H., & Masyn, K. (2009). General growth mixture analysis with antecedents and consequences of change. In *Handbook of quantitative criminology*, A. Piquero & D. Weisburd (Eds.), 69-100. New York: Springer.
- Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through six grades. *Journal of Educational Psychology*, *94*, 3-13.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2<sup>nd</sup> Ed.)*. Newbury Park, CA: Sage.
- Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychological problems from Grade 2 to Grade 8. *Annals of Dyslexia*, *53*, 47-71.

Speece, D. L., Ritchey, K. D., Cooper, D. H., Roth, F. P., & Schatschneider, C. (2003). Growth in early reading skills from kindergarten to third grade. *Contemporary Educational Psychology, 29*, 312-332.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests: administration, norms and commentary (3<sup>rd</sup> Ed.)*. New York: Oxford University Press.

Wilkinson, G. S. (1995). *The wide range achievement test-3*. Wilmington, DE: Jastak Associates.

Wong, C-P, Brown, C. H., & Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association, 100*, 1054-1076.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson psychoeducational test battery*. Boston: Teaching Resources.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*, 1090-1104.

Appendix:  
Mplus Syntax for the Final GPGMM  
Conditional Model

**TITLE:** GPGMM WORD RECOGNITION

**DATA:**

FILE IS wrat526.dat;  
FORMAT IS 418F22.0;

**VARIABLE:**

NAMES ARE K G1 G2 G3 G4 G5 G6  
Gender FirstLanguage  
WorkingMemory PhonoAwareness  
RetrievalTime ReadingComprehension;  
CLASSES = C(4);  
MISSING = K G1 G2 G3 G4 G5 G6  
ReadingComprehension (9999);  
AUXILIARY= (e)ReadingComprehension;  
USEVAR = K G1 G2 G3 G4 G5 G6 Gender  
FirstLanguage  
WorkingMemory PhonoAwareness  
RetrievalTime ReadingComprehension;

**ANALYSIS:**

TYPE = MIXTURE;  
STARTS = 1000 20;  
STITERATIONS = 100;  
MITERATIONS = 2000;

**MODEL:**

%OVERALL%  
I S1 | K@0 G1@1 G2@2 G3@2 G4@2  
G5@2 G6@2 ;  
I S2 | K@0 G1@0 G2@0 G3@1 G4@2  
G5@3 G6@4 ;  
C#1 C#2 C#3 ON Gender\*0 FirstLanguage\*0  
WorkingMemory\*0 PhonoAwareness\*0  
RetrievalTime\*0;  
I ON Gender\*0 FirstLanguage\*0  
WorkingMemory\*0 PhonoAwareness\*0  
RetrievalTime\*0;  
S1 ON Gender\*0 FirstLanguage\*0  
WorkingMemory\*0 PhonoAwareness\*0  
RetrievalTime\*0;  
S2 ON Gender\*0 FirstLanguage\*0  
WorkingMemory\*0 PhonoAwareness\*0  
RetrievalTime\*0;

%C#1%  
K@0 G1 G2 G3 G4 G5 G6  
I S1 S2@0;  
[I\*14.326 S1\*8.665 S2\*2.612;  
I with S1 @0;  
I ON  
Gender\*0.597  
FirstLanguage\*0.175  
WorkingMemory\*-0.221  
PhonoAwareness \*0.470  
RetrievalTime\*-0.185;  
S1 ON  
Gender\* -0.007  
FirstLanguage\* 0.907  
WorkingMemory\*0.300  
PhonoAwareness \*0.500  
RetrievalTime\*-0.162;  
S2 ON  
Gender\*-0.074  
FirstLanguage\*-0.375  
WorkingMemory\*-0.081  
PhonoAwareness\*-0.141  
RetrievalTime\*0.183;

%C#2%  
K@0 G1 G2 G3@0 G4 G5 G6 ;

## GPGMM WORD RECOGNITION

I S1 S2@0;  
 [\*8.436 S1\*10.289 S2\*2.911];  
 I with S1 @0;  
 I       ON  
 Gender@0  
 FirstLanguage@0  
 WorkingMemory@0  
 PhonoAwareness@0  
 RetrievalTime@0;  
 S1       ON  
 Gender@0  
 FirstLanguage@0  
 WorkingMemory@0  
 PhonoAwareness@0  
 RetrievalTime@0;  
 S2       ON  
 Gender@0  
 FirstLanguage@0  
 WorkingMemory@0  
 PhonoAwareness@0  
 RetrievalTime@0;

%C#3%  
 K\* G1 G2 G3 G4 G5 G6;  
 I S1 S2@0;  
 [\*11.329 S1\*11.352 S2\*2.156];  
 I with S1@0;

I       ON  
 Gender\*0.338  
 FirstLanguage\* 0.177  
 WorkingMemory\*-0.745  
 PhonoAwareness\*4.112  
 RetrievalTime\*-1.645;  
 S1       ON  
 Gender\*-0.524  
 FirstLanguage\*-0.914  
 WorkingMemory\*0.146  
 PhonoAwareness\*-1.250  
 RetrievalTime\*0.659;  
 S2       ON  
 Gender\*0.025  
 FirstLanguage\*0.215  
 WorkingMemory\*0.141  
 PhonoAwareness\*-0.304  
 RetrievalTime\*0.055;

%C#4%  
 K G1 G2 G3 G4 G5 G6@0;  
 I S1@0 S2;  
 [\*8.386 S1\*10.041 S2\*2.513];  
 I WITH S2@0;  
 I       ON  
 Gender@0  
 FirstLanguage@0  
 WorkingMemory @00  
 PhonoAwareness@0  
 RetrievalTime@0;  
 S1       ON  
 Gender@0  
 FirstLanguage@0  
 WorkingMemory@0  
 PhonoAwareness@0  
 RetrievalTime@0;  
 S2       ON  
 Gender@0  
 FirstLanguage@0  
 WorkingMemory @0  
 PhonoAwareness@0  
 RetrievalTime@0;

### OUTPUT:

TECH1 TECH4;  
 !STANDARDIZED;  
 SAVEDATA:File is final.sav;  
 SAVE = FSCORES;

## Using Finite Mixture Modeling to Deal with Systematic Measurement Error: A Case Study

Min Liu  
University of Hawaii,  
Manoa, HI USA

Gregory R. Hancock Jeffrey R. Harring  
University of Maryland,  
College Park, MD USA

---

Conventional methods and analyses view measurement error as random. A scenario is presented where a variable was measured with systematic error. Mixture models with systematic parameter constraints were used to test hypotheses in the context of general linear models; this accommodated the heterogeneity arising due to systematic measurement error.

Key words: Finite mixture models, systematic measurement error, general linear model.

---

### Introduction

In the social and behavioral sciences variables are frequently measured with error. A common approach is to treat measurement error as inherently unpredictable chance fluctuations, as opposed to something that can be caused by any factor that systematically affects measurement of a variable across the sample. This may be because systematic errors are – in general – difficult to detect, and because the estimation of the magnitude of such errors in practice is complex.

### Overview of Finite Mixture Models

Finite mixture modeling is an analytical paradigm used to analyze data sampled from a heterogeneous population with a different

probability density function (PDF) for each component population. While population heterogeneity could arise from multiple unrelated probability distributions (e.g., Cauchy and Student's  $t$  distributions), the more typical viewpoint is to assume that data come from a composite (i.e., mixture) of two or more distributions from the same parametric family, with the stipulation that parameters are permitted to differ across the unobserved components (see Titterington, Smith & Makov, 1985). In the general case, for data in  $\mathbf{x}$ , a finite mixture of  $K$  densities can be formulated as

$$f(\mathbf{x} | \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{x} | \boldsymbol{\theta}_k), \quad (1)$$

where  $\boldsymbol{\pi}$  contains mixing parameters  $\pi_k$  ( $k = 1, \dots, K$ ) reflecting prior probabilities of sampling from the  $k^{\text{th}}$  mixture component (class),  $f_k(\mathbf{x} | \boldsymbol{\theta}_k)$  represents the PDF for class  $k$ , and  $\boldsymbol{\theta}_k$  is a parameter vector for class  $k$ . In addition,  $\pi_k$  values are restricted to be nonnegative and sum to 1 over all  $K$  classes. The likelihood for the general mixture in (1) can be written as:

$$L(\boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^N f(x_i | \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^N \left[ \sum_{k=1}^K \pi_k f_k(x_i | \boldsymbol{\pi}, \boldsymbol{\theta}_k) \right]. \quad (2)$$

---

Min Liu is an assistant professor in the Department of Educational Psychology in the University of Hawaii at Manoa. Her primary research interest is in latent variable mixture models. Email her at [minliu@hawaii.edu](mailto:minliu@hawaii.edu). Gregory Hancock is a Professor and Chair of the Department of Measurement, Statistics and Evaluation. His primary research interest is in latent variable models. Email him at: [ghancock@umd.edu](mailto:ghancock@umd.edu). Jeffrey Harring is an Assistant Professor in the Department of Measurement, Statistics and Evaluation. Email him at: [harring@umd.edu](mailto:harring@umd.edu).

For finite mixtures, model parameter estimation is typically carried out by maximizing the logarithm of the likelihood function in (2) via the iterative expectation-maximization (EM) algorithm (Baum, Petrie, Soules & Weiss, 1970; Dempster, Laird & Rubin, 1977; Little & Rubin, 2002; McLachlan & Krishnan, 1997). In the current work, however, EM was not used in estimating parameters for the models of interest; instead, the natural logarithm of the likelihood in (2) was maximized directly using the Newton-Raphson algorithm (see Thisted, 1988).

This procedure uses a second-order Taylor series expansion of the gradient (first partial derivatives of the log-likelihood function) around the current estimate  $\theta^{(w)}$  to produce the next estimate  $\theta^{(w+1)}$ . At each iteration the Newton-Raphson optimization scheme requires both first and second partial derivatives of the log-likelihood function with respect to the parameters, which can be computationally demanding. This challenge notwithstanding, Newton-Raphson and its many variants remain popular choices due to the algorithm's quick convergence near the solution, and the fact that standard errors of the estimated parameters can be computed directly at convergence.

#### Mixture Models with Functionally Related Parameters

Applications of mixture models in the natural and biological sciences typically involve the blending of different natural groupings, such as visually similar species (e.g., crabs from the Bay of Naples; Pearson, 1894) or within-species age clusters (e.g., fish; Summerfelt & Hall, 1987). Within the social and behavioral sciences, subgroups often result from differential responses to stimuli or treatments. This includes invoking different problem-solving strategies (e.g., in spatial rotation tasks; Mislevy & Verhelst, 1990), different responses to test speededness (Bolt, Cohen & Wollack, 2002), or different responses to individual test items (i.e., differential item functioning; Samuelsen, 2008). Ding (2008) recommended regression mixture models as useful tools for modeling population heterogeneity, thus improving the accuracy of the regression function as evidenced by the

much lower error variance within each class or component population.

In the above examples there are typically no functional relations per se between the parameters governing distributions of the component populations, other than perhaps that one population should be higher on average than another (e.g., an older population of fish should be longer). In other cases, however, there might be a very specific relation between component populations' parameters. Oja, Koiranen and Rantakallio (1991), for example, examined birth weight data from Northern Finland for two cohorts: one from 1966 and one from 1985-1986. For the latter cohort the gestational age could be determined more accurately by reference to ultrasound measurements; for the earlier cohort, however, the gestational age could only be assumed based on each mother's self-report regarding her last menstrual period.

An examination of the data yielded an unexpected difference between the cohorts; specifically, the mean birth weight for the later cohort was higher than the earlier cohort, while its birth weight of (apparently) pre-term newborns was lower than those from the earlier cohort. To explain this difference, Oja, et al. (1991) hypothesized more frequent systematic measurement error in gestational assessment for the earlier cohort. As a result they modeled the two cohorts as being comprised of three subgroups: (1) newborns whose mothers correctly knew their last menstrual period, (2) newborns whose mothers mistakenly thought their last period was earlier, and (3) newborns whose mothers mistakenly thought their last period was later. The researchers successfully modeled both cohorts using these three classes, where relations among those classes' distributional parameters were constrained as a function of three gestational ages (i.e., correct age, 4 weeks earlier and 4 weeks later).

This last example demonstrates that, given specific knowledge or hypotheses about the origins of different classes, parameters may be functionally related across those classes. This study illustrates the case of such hypothesized relations across classes in the context of a systematic measurement problem as it relates to the accuracy of general linear model analyses.

Using Mixture Models to Accommodate Systematic Measurement Error

This study was inspired by a systematic measurement problem encountered when exploring and analyzing existing data for National Survey of Child and Adolescent Well-being (U.S. Department of Health and Human Services, Administration for Children, Youth and Families, 2003). Longitudinal data were collected from children who were subject to child abuse or neglect and included a wide variety of physiological and psychological variables.

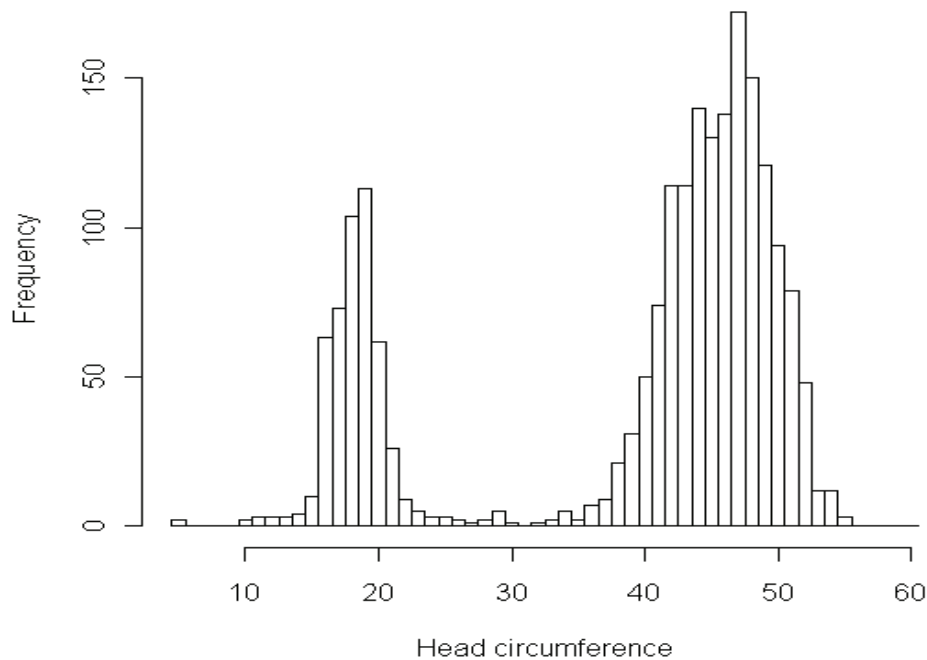
One key variable in the investigations of these data was head circumference for children up to 4 years old. Although members of the original research team had been instructed to measure head circumference in centimeters, an inspection of the data as shown in Figure 1 suggested that some of the researchers might have actually taken the measurements in inches (where 1 inch = 2.54 centimeters). [Unfortunately, the principal researchers were unable to contact individual field researchers to confirm our suspicions.] For example, the sample mean for the overall distribution was

38.99, which was inconsistent with prior published head circumference values for children in this age range (in either inches or centimeters).

Likewise, as is clear from Figure 1, the estimated standard deviation (12.29) was much larger than one would expect had only one set of units been employed. Thus, the apparently compromised first and second moments could not be used directly for even basic general linear model statistical analyses, such as a *t*-test or simple linear regression.

Fortunately, mixture modeling is a promising approach to accommodate systematic errors of this type. To this end, three studies were conducted using the problematic head circumference measure in the context of finite mixtures: a univariate analysis, a group means comparison and a simple regression. In each case competing models were constructed to analyze head circumference for a sample of 2,028 children. As noted, the Newton-Raphson algorithm was used and implemented using R to estimate all models (see Appendix for technical details; R code is available upon request from the first author).

Figure 1: Histogram of Head Circumference





Methodology

Univariate Analysis Model 1: 1-Class Model

Based on the physiological literature (e.g., Fok, et al., 2003; World Health Organization, 2007), head circumference measurements were considered to approximate a normal distribution. A one-class model was used as a baseline with which to compare the results of other models. The PDF for each observation may be written as:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}.$$

Assuming there were  $N$  independent observations, the log-likelihood was:

$$LL = \ln L(\mu, \sigma^2) = \sum_{i=1}^N \ln f(x_i; \mu, \sigma^2).$$

The two parameters to be estimated in this model were the distribution mean  $\mu$  and variance  $\sigma^2$ .

Univariate Analysis Model 2: 2-Class Mixture Model without Constraints

The second model was a composite of PDFs assuming that measurements of head circumference arose from two distributions distinguished by which units (centimeters and inches):

$$\begin{aligned} f_{i\_cent} &= f(x_i; \mu_{cent}, \sigma_{cent}^2) \\ &= \frac{1}{\sqrt{2\pi\sigma_{cent}^2}} \exp\left\{-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right\} \end{aligned}$$

and

$$\begin{aligned} f_{i\_inch} &= f(x_i; \mu_{inch}, \sigma_{inch}^2) \\ &= \frac{1}{\sqrt{2\pi\sigma_{inch}^2}} \exp\left\{-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right\} \end{aligned}$$

The corresponding log-likelihood function for the mixture was formulated as:

$$\begin{aligned} LL &= \ln L(\mu_{cent}, \sigma_{cent}^2, \mu_{inch}, \sigma_{inch}^2, \pi_{inch}) \\ &= \sum_{i=1}^N \left\{ \ln \left[ (1 - \pi_{inch}) f_{i\_cent} + \pi_{inch} f_{i\_inch} \right] \right\}. \end{aligned} \quad (3)$$

Five parameters were estimated in this model: two means  $\mu_{cent}$  and  $\mu_{inch}$ , two variances  $\sigma_{cent}^2$  and  $\sigma_{inch}^2$ , and the class mixing proportion parameter  $\pi_{inch}$ .

Univariate Analysis Model 3: 2-Class Mixture Model with Means Constrained

This model was identical to Model 2, with the exception that the means were constrained based on the fixed ratio of inch to centimeter ( $\mu_{cent} = 2.54\mu_{inch}$ ). Given this constraint, only four parameters were estimated in this model:  $\mu_{inch}$ ,  $\sigma_{cent}^2$ ,  $\sigma_{inch}^2$ , and  $\pi_{inch}$ .

Univariate Analysis Model 4: 2-Class Mixture Model with Variance Constrained

The fourth model hypothesized that the variances of two populations followed a fixed ratio of 6.45. This value came directly from the variance property

$$\begin{aligned} Var(cent) &= Var(2.54inch) \\ &= 2.54^2 Var(inch). \\ &= 6.45 Var(inch) \end{aligned}$$

Similar to the mean-constrained Model 3, the variance-constrained model also had four unknown parameters to be estimated:  $\mu_{cent}$ ,  $\mu_{inch}$ ,  $\sigma_{inch}^2$ , and  $\pi_{inch}$ .

Univariate Analysis Model 5: 2-Class Mixture with Both Means and Variances Constrained

The final model assumed that the means and variances of the two distributions differed by a function of 2.54. With the additional equality constraints placed on the means and variances ( $\mu_{cent} = 2.54\mu_{inch}$  and  $\sigma_{cent}^2 = 6.45\sigma_{inch}^2$ ), only three unknown

parameters needed to be estimated:  $\mu_{inch}$ ,  $\sigma_{inch}^2$ , and  $\pi_{inch}$ .

Model fit indices and corresponding parameter estimates for the univariate study are summarized in Tables 1 and 2, respectively. Three commonly used information criteria were determined for each analysis in this study. The Akaike's Information Criterion (AIC; Akaike, 1987) is based on the log-likelihood ( $LL$ ) of the hypothesized model and the number of parameters ( $p$ ), as follows:

$$AIC = -2LL + 2p.$$

An alternative to the AIC, the Bayesian information criterion (BIC; Schwarz, 1978), is obtained by modifying the penalty term based on sample size ( $N$ ) as follows:

$$BIC = -2LL + p \ln(N).$$

In the context of finite mixture modeling, BIC has been recommended for its consistency (Haughton, 1988; Leroux, 1992), meaning that it tends to select the correct model more frequently as sample size increases. Sclove (1987) suggested a further sample size adjustment for BIC (S-BIC) where  $N$  is replaced by  $N' = (N + 2) / 24$ , and Yang (2006) advocated this index, citing better performance when the model has either a large number of parameters or a small sample size. In this study, all three information criteria were used to compare both non-nested and nested models; when models being compared were nested, Chi-square difference (likelihood ratio) tests were also conducted.

Table 1: Model Fit Indices from Univariate Analyses

Model	# Parameters	lnL	AIC	BIC	S-BIC
1. 1 Class	2	-7964.97	15933.94	15945.17	15938.82
2. 2-Class Mixture	5	-6501.70	13013.40	13041.47	13025.59
3. 2-Class Mixture Mean Constrained	4	-6504.55	13017.10	13039.56	13026.85
4. 2-Class Mixture Variance Constrained	4	-6561.07	13130.14	13152.60	13139.89
5. 2-Class Mixture Both Constrained	3	-6563.38	13132.76	13149.60	13140.07

Table 2: Maximum Likelihood Estimates from Univariate Analyses

Model	$\hat{\mu}_{cent}$	$\hat{\sigma}_{cent}^2$	$\hat{\mu}_{inch}$	$\hat{\sigma}_{inch}^2$	$\hat{\pi}_{inch}$
1. 1 Class	38.99	150.98	–	–	–
2. 2-Class Mixture	45.66	14.93	18.25	5.67	0.24
3. 2-Class Mixture Mean Constrained	45.73	15.01	18.00	5.63	0.24
4. 2-Class Mixture Variance Constrained	45.53	19.67	18.15	3.05	0.24
5. 2-Class Mixture Both Constrained	45.72	19.74	18.00	3.06	0.24

## MEASUREMENT ERROR MIXTURES

As shown in Table 1, compared with Model 1 (which assumed population homogeneity), all other 2-class mixture models provided much better fit to the data; this is indicated by the larger (i.e., less negative) log-likelihood and the smaller fit indices (AIC, BIC and S-BIC). This finding provided initial support for the hypothesis that a systematic measurement error problem was present. Also noteworthy is the estimated mixing parameter (0.24), suggesting that roughly one-fourth of the subjects were improperly measured using inches rather than centimeters; this value was consistent across subsequent analyses as well.

Comparing the fit indices of the four mixture models, the unconstrained and mean-constrained models provided the best fit ( $BIC_2=13041.47$ ,  $BIC_3=13039.54$ ), whereas the last two models involving variance constraints did not fit as well. Looking at the two models without variance constraints shows that the estimated variances for the two classes did not follow the expected ratio of 6.45; this could be due to the different magnitudes of random measurement error caused by their different precisions of measurement.

Relatively speaking, measurement in centimeter units is more precise than that in inches; the latter thus introduces more random measurement error. Regarding the means, Table 1 shows the BIC favored the mean-constrained model (while the AIC and S-BIC were very close); Table 2 shows that the estimates from the mean-constrained and unconstrained models differed only at the decimal level. For these reasons, the functional relation of the two means was used in subsequent studies.

### Group Means Comparison

The second study was a group means comparison that examined whether there was a gender difference in head circumference. Due to the apparent unit mixture problem, a typical  $t$ -test comparing males and females could not be used; similarly, because cases' membership into a centimeter or inch group was latent, separate male-female comparisons could not be directly conducted within each unit group. Thus, with the inch/centimeter mean constraint in place, the values of the log-likelihood were compared under two multi-sample mixture models: one

assuming a common mean for both genders and one estimating a separate mean for males (M) and females (F).

In the model assuming a gender difference, there were four normal PDFs crossing units and gender:

1.  $f(x_i; \mu_{FemaleInch}, \sigma_{FemaleInch}^2)$  for females in inches with mixing proportion  $\pi_{FemaleInch}$ ;
2.  $f(x_i; 2.54\mu_{FemaleInch}, \sigma_{FemaleCent}^2)$  for females in centimeters with proportion of  $1-\pi_{FemaleInch}$ ;
3.  $f(x_i; \mu_{MaleInch}, \sigma_{MaleInch}^2)$  for males in inches with proportion of  $\pi_{MaleInch}$ ; and
4.  $f(x_i; 2.54\mu_{MaleInch}, \sigma_{MaleCent}^2)$  for males in centimeters with proportion of  $1-\pi_{MaleInch}$ .

It should be noted that a different mixing proportion was allowed for males and females ( $\pi_{FemaleInch}$  and  $\pi_{MaleInch}$ ), reflecting the possibility that researchers' erroneous use of inches rather than centimeters could have been related in some way to child gender. Overall, the log-likelihood for this model, which contained 8 parameters (4 per gender), could be written as:

$$LL = \ln L(\mu_{FemaleInch}, \sigma_{FemaleInch}^2, \sigma_{FemaleCent}^2, \pi_{FemaleInch}, \mu_{MaleInch}, \sigma_{MaleInch}^2, \sigma_{MaleCent}^2, \pi_{MaleInch}).$$

The second multi-sample mixture model, which assumed no gender difference in head circumference, included only one common mean for both gender groups to be determined. Thus, its log-likelihood, which contains 7 parameters, could be written as

$$LL = \ln L(\mu_{common}, \sigma_{FemaleInch}^2, \sigma_{FemaleCent}^2, \pi_{FemaleInch}, \sigma_{MaleInch}^2, \sigma_{MaleCent}^2, \pi_{MaleInch}).$$

Using the same estimation process previously described, model fit indices and estimated parameters for the mean comparison analysis are presented respectively in Tables 3

and 4. These results show that all model fit indices supported the model specifying a gender difference in head circumference over that which assumed gender equality ( $BIC_1=13036.50$ ,  $BIC_2=13057.74$ ). Additionally, because the second model assuming no gender difference was nested within the first model, a Chi-square difference test was used to compare the fit of the two models.

The observed Chi-square difference statistic ( $\chi^2_{df=1} = 28.86$ ,  $p < 0.001$ ) indicated the first model with gender difference was statistically significantly better than the constrained model. Therefore, it was inferred that male and female children who experienced abuse or neglect did have different head circumferences, with males being larger (18.21 inches vs. 17.82 inches). It is worth noting that if a traditional  $t$ -test had been used directly with the original unit-compromised variable, a non-significant test result would have been obtained ( $t_{(df = 2026)} = -1.706$ ,  $p = .088$ ); thus, the model would have failed to detect any difference between males and females.

Regression Analysis

The third study investigated if age in months was a useful predictor of head circumference by applying systematically constrained mixture regression models to these two variables. After the first baseline model, four regression mixture models were investigated, assuming different regression functions for the two unknown groups with different measurement units.

Model 1: Simple Regression of Head Circumference (y) on Age (x)

The simple bivariate regression model could be specified as  $y_i = \alpha + \beta x_i + e_i$ , where  $\alpha$  is the population intercept and  $\beta$  is the population slope. The residual  $e_i$  was assumed to be normally distributed, making the PDF for an individual observation

$$f(x_i, y_i; \alpha, \beta, \sigma_e^2) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_e^2}\right\}.$$

Table 3: Model Fit Indices from Mean Comparison Analyses

Model	# Parameters	lnL	AIC	BIC	S-BIC
Assuming Gender Difference	8	-6487.79	12991.58	13036.50	13011.08
Assuming Gender Equality	7	-6502.22	13018.44	13057.74	13035.50

Table 4: Maximum Likelihood Estimates from Mean Comparison Analyses

Model	$\hat{\mu}_{FemaleInch}$	$\hat{\sigma}^2_{FemaleInch}$	$\hat{\sigma}^2_{FemaleCent}$	$\hat{\pi}_{FemaleInch}$	$\hat{\mu}_{MaleInch}$	$\hat{\sigma}^2_{MaleInch}$	$\hat{\sigma}^2_{MaleCent}$	$\hat{\pi}_{MaleInch}$
Assuming Gender Difference	17.82	4.91	14.67	0.246	18.21	7.85	13.64	0.244
Assuming Gender Equality	18.02	4.93	14.96	0.246	18.02	7.47	14.18	0.243

## MEASUREMENT ERROR MIXTURES

The log-likelihood for the full sample was thus

$$LL = \ln L = \sum_{i=1}^N [\ln(f(x_i, y_i; \alpha, \beta, \sigma_e^2))].$$

As indicated by the log-likelihood function, 3 parameters needed to be determined.

**Model 2: Mixture Regression Models without Any Constraints**

In this model, all of the regression coefficients, residual variances, and class proportions were free to be estimated. Its equation could be expressed as  $y_{ik} = \alpha_k + \beta_k x_i + e_i$  for each  $k^{\text{th}}$  class, with  $k=1,2$  representing the inch and centimeter classes; thus, the two unit groups each had their own regression coefficients to be determined. Because each observation could be sampled from either of the two unit classes, there were two class-specific PDFs for an individual:

$$f_{i\_inch} = f(x_i, y_i; \alpha_{inch}, \beta_{inch}, \sigma_{e\_inch}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e\_inch}^2}} \exp\left\{-\frac{(y_i - \alpha_{inch} - \beta_{inch}x_i)^2}{2\sigma_{e\_inch}^2}\right\}$$

and

$$f_{i\_cent} = f(x_i, y_i; \alpha_{cent}, \beta_{cent}, \sigma_{e\_cent}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e\_cent}^2}} \exp\left\{-\frac{(y_i - \alpha_{cent} - \beta_{cent}x_i)^2}{2\sigma_{e\_cent}^2}\right\}.$$

The corresponding log-likelihood was

$$LL = \ln L \\ = \sum_{i=1}^N [\ln(\pi_{inch} f_{i\_inch} + (1 - \pi_{inch}) f_{i\_cent})];$$

in this case 7 parameters needed to be estimated.

**Model 3: Mixture Regression Model with Slope Constraint**

Given that the slope represents the degree of expected change in head

circumference per unit increase in age, the suspected unit problem would lead to the relation:  $\beta_{cent} = 2.54\beta_{inch}$ . Using this relation as a constraint, the PDF for centimeter group was

$$f_{i\_cent} = f(x_i, y_i; \alpha_{cent}, \beta_{inch}, \sigma_{e\_cent}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e\_cent}^2}} \exp\left\{-\frac{(y_i - \alpha_{cent} - 2.54\beta_{inch}x_i)^2}{2\sigma_{e\_cent}^2}\right\}.$$

The resulting log-likelihood was

$$LL = \ln L = \sum_{i=1}^N \left[ \ln(\pi_{inch} f_{i\_inch}(\alpha_{inch}, \beta_{inch}, \sigma_{e\_inch}^2)) \right. \\ \left. + (1 - \pi_{inch}) f_{i\_cent}(\alpha_{cent}, \beta_{inch}, \sigma_{e\_cent}^2) \right],$$

with 6 unknown parameters to estimate.

**Model 4: Mixture Regression Model with Intercept Constraint**

Based on the population relation for the intercept  $\alpha_k = \mu_y - \beta_k \mu_x$ , and assuming comparable ages for the two classes, it was not unreasonable to expect the familiar functional relation between the two classes' intercept terms  $\alpha_{cent} = 2.54\alpha_{inch}$ . After this replacement, the PDF for the centimeter group was

$$f_{i\_cent} = f(x_i, y_i; \alpha_{inch}, \beta_{cent}, \sigma_{e\_cent}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e\_cent}^2}} \exp\left\{-\frac{(y_i - 2.54\alpha_{inch} - \beta_{cent}x_i)^2}{2\sigma_{e\_cent}^2}\right\}.$$

The resulting log-likelihood is

$$LL = \ln L = \sum_{i=1}^N \left[ \ln(\pi_{inch} f_{i\_inch}(\alpha_{inch}, \beta_{inch}, \sigma_{e\_inch}^2)) \right. \\ \left. + (1 - \pi_{inch}) f_{i\_cent}(\alpha_{inch}, \beta_{cent}, \sigma_{e\_cent}^2) \right],$$

with 6 parameters to be estimated.

**Model 5: Mixture Regression Model with Constraints on Both Intercepts and Slopes**

Based on the rationales provided for Models 3 and 4, both relations  $\beta_{cent} = 2.54\beta_{inch}$  and  $\alpha_{cent} = 2.54\alpha_{inch}$  were applied in this

model. From these constraints, the PDF for the centimeter group was specified as

$$f_{i\_cent} = f(x_i, y_i; \alpha_{inch}, \beta_{inch}, \sigma_{e\_cent}^2) = \frac{1}{\sqrt{2\pi\sigma_{e\_cent}^2}} \exp\left\{-\frac{(y_i - 2.54\alpha_{inch} - 2.54\beta_{inch}x_i)^2}{2\sigma_{e\_cent}^2}\right\}$$

The corresponding log-likelihood function was

$$LL = \ln L = \sum_{i=1}^N \left[ \ln(\pi_{inch} f_{i\_inch}(\alpha_{inch}, \beta_{inch}, \sigma_{e\_inch}^2)) + (1 - \pi_{inch}) f_{i\_cent}(\alpha_{inch}, \beta_{inch}, \sigma_{e\_cent}^2) \right]$$

with 5 parameters to be estimated in total.

### Results

All estimated results and corresponding model fit indices are summarized in Tables 5 and 6, respectively. Table 5 shows that the mixture models (Models 2 - 5) were similar in fit and were strongly favored over the simple regression model (Model 1). Among the four mixture models, the choice depended on the information criterion measure used; specifically, while all mixture models were fairly close in information criteria values, the AIC and S-BIC favored the model with only the slope constrained (Model 3) while the BIC favored the model with intercept and slope constraints (Model 5).

A comparison of these two models using

a Chi-square difference test yielded a statistically significant difference ( $\chi_{df=1}^2 = 6.08, p < 0.01$ ), thus favoring the model without the intercept constraint. A potential reason for this constraint's failure is that it rests upon the assumption that the mean age for the inch and centimeter classes is the same (i.e., that a researcher's mistaken decision to use inches was unrelated to the child's age); however, the estimated mean age for the two classes (based on a posterior probability-based group assignment) was 16.77 months for the inch class and 15.65 for the centimeter class. This mean difference notwithstanding, the slope and intercept parameter estimates for all regression mixture models, constrained or unconstrained, are practically close in value. Thus, to assess the linear relation between age in months and head circumference, the slope would lead to an estimate roughly one-fifth (0.20) of a centimeter increase in head circumference for each one month increase in age.

These results are corroborated by the plot in Figure 2, where circles represent bivariate age/head circumference observations. It may be observed that the solid simple regression line does not capture the relation between the two variables, while the two dashed regression lines generated from the mixture model provided a much better approximation to the plotted observations.

Table 5: Model Fit Indices from Regression Analyses

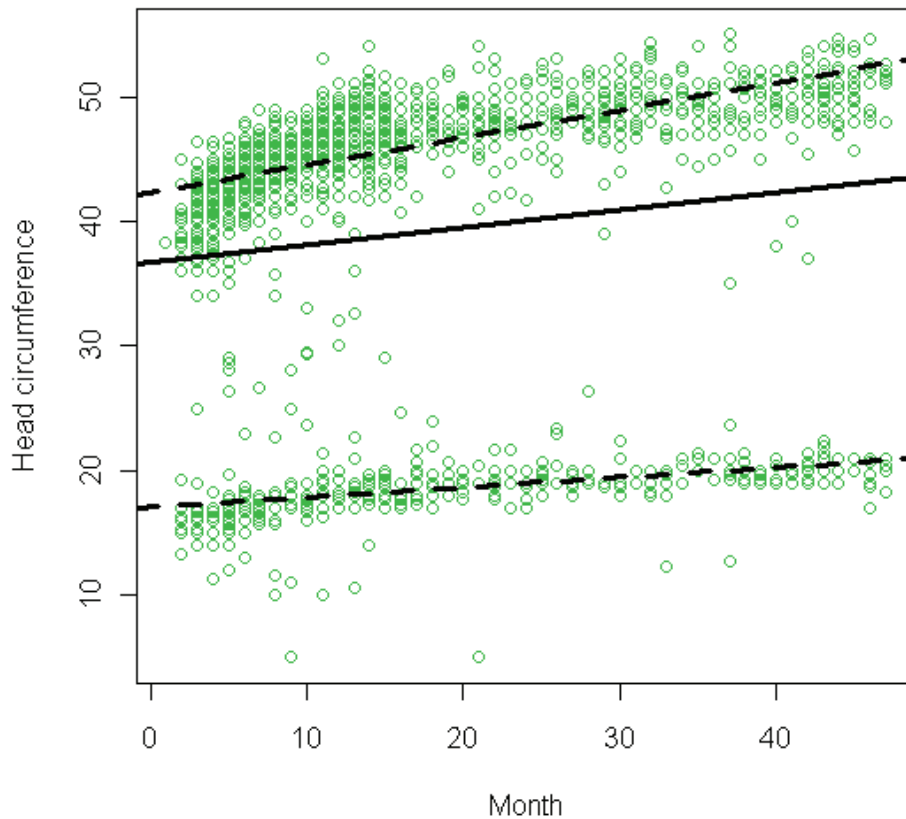
Model	# Parameters	lnL	AIC	BIC	S-BIC
1. Simple Regression	3	-7942.91	15908.66	15899.13	15899.13
2. 2-Class Regression Mixture No Constraint	7	-5918.70	11851.40	11890.70	11868.46
3. 2-Class Regression Mixture Constrained Slope	6	-5918.90	11849.80	11883.49	11864.43
4. 2-Class Regression Mixture Constrained Intercept	6	-5920.70	11853.40	11887.09	11868.03
5. 2-Class Regression Mixture Constrained Intercept and Slope	5	-5921.94	11853.88	11881.95	11866.07

## MEASUREMENT ERROR MIXTURES

Table 6: Maximum Likelihood Estimates from Regression Analyses

Model	$\hat{\alpha}_{inch}$	$\hat{\beta}_{inch}$	$\hat{\sigma}_{e\_inch}^2$	$\hat{\pi}_{inch}$	$\hat{\alpha}_{cent}$	$\hat{\beta}_{cent}$	$\hat{\sigma}_{e\_cent}^2$
1. Simple Regression	36.70	0.14	147.73	–	–	–	–
2. 2-Class Regression Mixture No Constraint	17.04	0.08	5.88	0.25	42.34	0.22	6.83
3. 2-Class Regression Mixture Constrained Slope	16.95	0.08	5.89	0.25	42.35	0.20	6.83
4. 2-Class Regression Mixture Constrained Intercept	16.69	0.09	5.92	0.25	42.39	0.21	6.83
5. 2-Class Regression Mixture Constrained Intercept and Slope	16.69	0.08	5.93	0.25	42.39	0.20	6.84

Figure 2: Simple Regression Line (Solid) vs. Unconstrained Mixture Regression Lines (Dashed)



### Conclusion

Based on the series of analyses examined in this study, mixture modeling appears to be an effective tool for investigating data consisting of variables with systematic measurement error. Systematic measurement errors have the potential to render data virtually useless: the

implementation of mixture models thus has the potential to salvage information regarding the key population relations in the data thereby avoiding the otherwise tremendous waste of time and expense associated with gathering such problematic data.

The studies presented may be viewed as a tip of the analytical iceberg. Although the only analyses demonstrated regarded mean comparisons and simple regression, extensions exist for the multivariate general(ized) linear model, as well as for latent variable models (e.g., structural equation models). The key to the implementation of such models is to have a specific hypothesis about the nature of the systematic measurement error and then translate that hypothesis into model constraints. It is hoped that the current case study has provided a useful illustration of the accommodation and adjustment for such measurement errors thereby bringing meaning to otherwise compromised data.

#### References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.
- Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, *41*, 164-171.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Applications of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, *39*, 331-348.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, *B39*, 1-38.
- Ding, C. S. (2008). Enhancing accuracy in research using regression mixture analysis. In J. Osborne (Ed.), *Best Practices in Quantitative Methods*, 410-421. Thousand Oaks, CA: Sage Publications
- Fok, T. F., et al. (2003). Updated gestational age specific birth weight, crown-heel length, and head circumference of Chinese newborns. *Archives of Disease in Childhood – Fetal and Neonatal Edition*, *88*, 229-236.
- Haughton, D. (1988). On the choice of a model to fit data from an exponential family. *The Annals of Statistics*, *16*, 342-355.
- Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *Annals of Statistics*, *20*, 1350-1360.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data (2<sup>nd</sup> Ed.)*. New York: John Wiley.
- McLachlan, G., & Krishnan, T. (1997). *The EM algorithm and extensions. Wiley series in probability and statistics*. New York: John Wiley & Sons.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item response when different subjects employ different solution strategies. *Psychometrika*, *55*, 195-215.
- Oja, H., Koironen, M., & Rantakallio, P. (1991). Fitting mixture models to birth weight data: A case study. *Biometrics*, *47*, 883-897.
- Pearson, K. (1894). Contributions to the theory of mathematical evolution. *Philosophical Transactions of the Royal Society of London*, *A185*, 71-110.
- Samuelsen, K. M. (2008). Examining differential item function from a latent class perspective. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in Latent Variable Mixture Models*, 177-197. Charlotte, NC: Information Age Publishing, Inc.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.
- Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333-343.
- Summerfelt, R. C., & Hall, G. E. (1987). *The age and growth of fish*. Ames, IA: Iowa State University Press.
- Thisted, R. A. (1988). *Elements of statistical computing: Numerical computation*. New York, NY: Chapman & Hall.
- Titterton, D. M., Smith, A. F. M., & Makow, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- U. S. Department of Health and Human Services, Administration for Children, Youth and Families (2003). *National Survey of Child and Adolescent Well-Being (NSCAW)*, Restricted Release, Waves 1-4. Washington, DC.



World Health Organization (2007). *WHO child growth standards: Head circumference-for-age, arm circumference-for-age, tricep skinfold-for-age and subscapula skinfold-for-age*. Retrieved from [http://www.who.int/childgrowth/standards/second\\_set/technical\\_report\\_2.pdf](http://www.who.int/childgrowth/standards/second_set/technical_report_2.pdf).

Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50, 1090-1104.

#### Appendix

The Newton-Raphson is a general optimization strategy based on a quadratic Taylor series expansion of the gradient (first partial derivatives of the log-likelihood function with respect to model parameters). To fashion the Newton-Raphson update, this quadratic function is maximized with respect to  $\boldsymbol{\theta}$  in order to generate the next iterate. Setting the gradient expression equal to zero and solving for  $\boldsymbol{\theta}$  provides the update

$$\boldsymbol{\theta}^{w+1} = \boldsymbol{\theta}^w - \mathbf{H}^{-1}(\boldsymbol{\theta}^w) \mathbf{g}(\boldsymbol{\theta}^w)$$

where  $\mathbf{H}$  is the Hessian, the matrix of partial second derivatives of the log-likelihood with respect to the parameter vector, and  $\mathbf{g}$  is the gradient.

Recall the log-likelihood function from Equation (3):

$$LL = \sum_{i=1}^N \left\{ \ln \left[ (1 - \pi_{inch}) f_{i\_cent}(\mu_{cent}, \sigma_{cent}^2) + \pi_{inch} f_{i\_inch}(\mu_{inch}, \sigma_{inch}^2) \right] \right\}$$

The gradient is defined as

$$\mathbf{g} = \begin{pmatrix} \frac{\partial LL}{\partial \mu_{inch}} \\ \frac{\partial LL}{\partial \sigma_{inch}^2} \\ \frac{\partial LL}{\partial \mu_{cent}} \\ \frac{\partial LL}{\partial \sigma_{cent}^2} \\ \frac{\partial LL}{\partial \pi_{inch}} \end{pmatrix},$$

with elements defined as shown in Appendix Figures. After elements of the gradient have been computed analytically, subsequent elements of the Hessian matrix can be obtained numerically.

#### Appendix Figures: Elements of Gradient

$$\frac{\partial LL}{\partial \mu_{inch}} = \sum_{i=1}^N \frac{\frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right) \left(\frac{x_i - \mu_{inch}}{\sigma_{inch}^2}\right)}{\left(\frac{1 - \pi_{inch}}{\sigma_{cent}}\right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

$$\frac{\partial LL}{\partial \sigma_{inch}^2} = \sum_{i=1}^N \frac{\frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right) \frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^4} - \frac{\pi_{inch}}{2\sigma_{inch}^3} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}{\left(\frac{1 - \pi_{inch}}{\sigma_{cent}}\right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

## Appendix Figures: Elements of Gradient (continued)

$$\frac{\partial LL}{\partial \mu_{cent}} = \sum_{i=1}^N \frac{\left( \frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) \left( \frac{x_i - \mu_{cent}}{\sigma_{cent}^2} \right)}{\left( \frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

$$\frac{\partial LL}{\partial \sigma_{cent}^2} = \sum_{i=1}^N \frac{\left( \frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) \frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^4} - \left( \frac{1-\pi_{inch}}{2\sigma_{cent}^3} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right)}{\left( \frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

$$\frac{\partial LL}{\partial \pi_{inch}} = \sum_{i=1}^N \frac{\frac{-1}{\sigma_{cent}} \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{1}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}{\left( \frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

## Bayesian Threshold Moving Average Models

Mahmoud M. Smadi  
Jordan University of Science and Technology,  
Irbid, Jordan

M. T. Alodat  
Yarmouk University,  
Irbid, Jordan

---

A Bayesian approach in threshold moving average model for time series with two regimes is provided. The posterior distribution of the delay and threshold parameters are used to examine and investigate the intrinsic characteristics of this nonlinear time series model. The proposed approach is applied to both simulated data and a real data set obtained from a chemical system.

Key words: Threshold time series, moving average model, Bayesian estimation, simulation, chemical data.

---

### Introduction

One class of nonlinear time series models is the threshold time series models which are extensively reported in literature. Among these, Tong and Lim (1980) introduced threshold autoregressive (TAR) models with statistical inference and applications. Bayesian inference for threshold autoregressive models have been investigated by different authors.

Geweke and Terui (1991) derived an exact posterior distribution of the delay and threshold parameters. Cathy, et al. (1995) used Monte Carlo Markov chain (MCMC) methods to implement a Bayesian inference on TAR models, and Broemeling and Cook (1992) performed a Bayesian analysis on TAR models. However, most of the literature emphasizes the threshold autoregressive models. Wang, et al. (1984) introduced the threshold autoregressive

moving average (TARMA) models and considered the estimation of the model parameters. De Gooijer (1998) studied various problems associated with the identification, estimation and testing of threshold moving average models. Ling and Tong (2005) considered a quasi-likelihood ratio test for the threshold in moving average models. Amendola, et al. (2009) discussed the stochastic structure of the self-exiting TARMA model; they specified sufficient conditions for weak stationarity and showed that the self-exiting TARMA model belongs to the class of the random coefficients autoregressive models. Smadi (1997) used the Bayesian approach for exploration of the joint posterior distribution for TARMA models using MCMC methods: he assumed noninformative priors, fixing the delay parameter  $d$ . In addition, he used a modified Gibbs sampling scheme, which is a hybrid strategy of Gibbs sampler, random walk Metropolis, and importance sampling. Safadi and Morettin (2000) considered a Bayesian analysis for threshold autoregressive moving average models and a hierarchical prior to perform Bayesian analysis using a rearranged procedure with MCMC methods.

The objective of this study is to provide a Bayesian approach in a threshold moving average model for time series with two regimes. The posterior distribution of the delay and the threshold parameters are used to examine and investigate the characteristics which are intrinsic to this nonlinear time series model. The

---

Mahmoud M. Smadi is an Associate Professor of Statistics in the Department of Mathematics and Statistics. His research interests include time series analysis, Bayesian statistics, and environmental statistics. Email him at: smadi@just.edu.jo. M. T. Alodat is an Associate Professor in the Department of Statistics. His research interests include the geometry of random fields, analysis medical images, approximating probability distributions, prediction and estimation using Bayesian methods. Email him at: alodatmts@yahoo.com.

proposed approach is applied to both simulated data and a real data set obtained from a chemical system.

Methodology

The Threshold Models

Let  $\{Y_t, t \geq 1\}$  be a time series, the threshold autoregressive moving average models with two regimes. Wang, et al. (1984) symbolized  $TARMA(2, (p_1, q_1), (p_2, q_2))$ , given by:

$$Y_t = \begin{cases} \sum_{i=1}^{p_1} \phi_i^{(1)} Y_{t-i} + \sum_{i=1}^{q_1} \theta_i^{(1)} Z_{t-i}^{(1)} + Z_t^{(1)}, & Y_{t-d} \leq r, \\ \sum_{i=1}^{p_2} \phi_i^{(2)} Y_{t-i} + \sum_{i=1}^{q_2} \theta_i^{(2)} Z_{t-i}^{(2)} + Z_t^{(2)}, & Y_{t-d} > r, \end{cases} \quad (1)$$

where  $\phi_1^{(j)}, \dots, \phi_{p_j}^{(j)}, \theta_1^{(j)}, \dots, \theta_{q_j}^{(j)}, (j=1,2)$  are the model coefficients of the two ARMA subsystems,  $r$  is called the threshold parameter and  $d$  is the delay parameter; assume the innovations  $Z_t^{(j)} : N(0, \sigma_j^2), j=1, 2$ .

A special case of equation (1), is the threshold autoregressive model  $TAR(2; p_1, p_2)$ :

$$Y_t = \begin{cases} \sum_{i=1}^{p_1} \phi_i^{(1)} Y_{t-i} + Z_t^{(1)} & Y_{t-d} \leq r, \\ \sum_{i=1}^{p_2} \phi_i^{(2)} Y_{t-i} + Z_t^{(2)} & Y_{t-d} > r. \end{cases} \quad (2)$$

Another special case of equation (1) is the threshold Moving Average model  $TMA(2; q_1, q_2)$ :

$$Y_t = \begin{cases} \sum_{i=1}^{q_1} \theta_i^{(1)} Z_{t-i}^{(1)} + Z_t^{(1)} & Y_{t-d} \leq r, \\ \sum_{i=1}^{q_2} \theta_i^{(2)} Z_{t-i}^{(2)} + Z_t^{(2)} & Y_{t-d} > r. \end{cases} \quad (3)$$

Posteriors Distribution

The approximate posterior distribution of the delay and threshold parameters  $(d, r)$  for threshold moving average models (3) is based on using estimated residuals instead of the true

innovations. Broemeling and Shaarawy (1986) implemented the estimated innovations for Bayesian analysis of ARMA models. Smadi (1997) and Safadi and Morettin (2000) have used this estimated innovation approach to explore the posterior distributions of the threshold autoregressive moving average models.

Defining  $\phi = (d, r)$ , and the set  $B_j = (\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{q_j}^{(j)})$ , and under the normality assumption, that is,  $Z_t^{(j)} i.i.d N(0, \sigma_j^2)$ , a prior  $\pi(\phi)$  could be any form as long as

$$\sum_{d=1}^R \int_{-\infty}^{\infty} \pi(d, r) dr = 1.$$

Conditional on  $\phi$ , independent priors on  $B_j$  and  $\sigma_j$  of standard Jeffreys prior can be expressed as:

$$\prod \{B_j, \sigma_j (j=1,2), \phi\} = \pi_c(B_j, \sigma_j (j=1,2), \phi) \pi(\phi) \quad (4)$$

where  $\pi_c(B_j, \sigma_j, j=1,2 | \phi) = (\sigma_1 \sigma_2)^{-1}$ . It is also assumed that  $Z_0^{(1)} = Z_{-1}^{(1)} = \dots = Z_{1-q_1}^{(1)} = 0$  and  $Z_0^{(2)} = Z_{-1}^{(2)} = \dots = Z_{1-q_2}^{(2)} = 0$ . Conditioning on  $\phi = (d, r)$ , estimates of the innovations  $Z_t^{(1)}$  and  $Z_t^{(2)}$  can be obtained using least squares estimates. In this case, the following model is obtained:

$$Y_t = \begin{cases} \sum_{i=1}^{q_1} \theta_i^{(1)} \widehat{Z}_{t-i}^{(1)} + Z_t^{(1)} & Y_{t-d} \leq r, \\ \sum_{i=1}^{q_2} \theta_i^{(2)} \widehat{Z}_{t-i}^{(2)} + Z_t^{(2)} & Y_{t-d} > r. \end{cases} \quad (5)$$

Derivation of the approximate posterior of the delay and the threshold parameters of the threshold moving average model is similar to the threshold autoregressive model (2) reported by Geweke and Terui (1993). After estimating the

innovations and using the model (3), conditional on  $\phi = (d, r)$ , let  $W_1$  be a vector consisting of  $N_1$  ordered observations on  $\{Y_t\}$  such that  $Y_{t-d} \leq r$ , and let  $W_2$  be a vector consisting of  $N_2$  ordered observations on  $\{Y_t\}$  such that  $Y_{t-d} > r$ . Let  $X_j$  be a  $N_j \times q_j$  matrix of lagged variables on the estimated innovations  $\{\hat{Z}_t\}$  corresponding to  $W_j$  ( $j = 1, 2$ ). Then, the approximate posterior density of  $(B_1, \sigma_1, B_2, \sigma_2)$  conditional on  $d$  and  $r$  is the product of two posterior densities, that is:

$$P(B_j, \sigma_j, j = 1, 2 | \phi, Y) = \prod_{j=1}^2 \sigma_j^{-(N_j+1)} (2\pi)^{-(N_j/2)} \exp \left\{ -\frac{(W_j - X_j B_j)' (W_j - X_j B_j)}{2\sigma_j^2} \right\} \quad (6)$$

The posterior distribution of  $\phi$  can be derived by integrating this expression with respect to  $(B_1, \sigma_1, B_2, \sigma_2)$ . The problem is to integrate the following expression with respect to  $\sigma$ :

$$\sigma^{-(T+1)} \exp \left\{ -\frac{(W - XB)' (W - XB)}{2\sigma^2} \right\} = \sigma^{-(T+1)} \exp \left\{ -\frac{v s^2 + (B - \hat{B})' X' X (B - \hat{B})}{2\sigma^2} \right\} \quad (7)$$

where

$$B = (X' X)^{-1} X' W, \quad s^2 = (W - X\hat{B})' (W - X\hat{B}) / v, \quad v = N - K - 1.$$

Integrating over  $\sigma_1$  and  $\sigma_2$  on the right hand side of equation (7), it is possible to obtain:

$$P\{B_j, j = 1, 2 | \phi, Y\} = \prod_{j=1}^2 (2\pi)^{-(N_j/2)} \int_0^\infty \sigma_j^{-(N_j+1)} \exp \left\{ -\frac{v_j s_j^2 + (B_j - \hat{B}_j)' W_j' W_j (B_j - \hat{B}_j)}{2\sigma_j^2} \right\} d\sigma_j$$

$$P\{B_j, j = 1, 2 | \phi, Y\} = \prod_{j=1}^2 (2\pi)^{-(N_j/2)} 2^{(N_j/2-1)} \Gamma\left(\frac{N_j}{2}\right) \{v_j s_j^2 + (B_j - \hat{B}_j)' W_j' W_j (B_j - \hat{B}_j)\}^{-(N_j/2)} \quad (8)$$

multiplying equation (8) by  $\pi(\phi)$  and integrating  $B_j$ , for  $j = 1, 2$  out results in the posterior distribution of  $\phi$  (Geweke & Terui, 1993):

$$P(\phi | Y) \propto \pi(\phi) \prod_{j=1}^2 (2\pi)^{-(N_j/2)} 2^{(N_j/2-1)} \Gamma\left(\frac{N_j}{2}\right) (s_j^2)^{-(N_j/2)} \times \int_{R^{(k_j+1)}} \left\{ v_j + (B_j - \hat{B}_j)' \frac{W_j' W_j}{s_j^2} (B_j - \hat{B}_j) \right\}^{-(N_j/2)} dB_j = \prod_{j=1}^2 2^{-(v_j/2+1)} (\pi)^{-(v_j/2)} \Gamma\left(\frac{v_j}{2}\right) \left(\frac{v_j s_j^2}{2}\right)^{-(v_j/2)} |W_j' W_j|^{-1/2} \quad (9)$$

where

$$s_j^2 = (W_j - X_j \hat{B}_j)' (W_j - X_j \hat{B}_j) / v_j, \quad \hat{B}_j = (X_j' X_j)^{-1} X_j' W_j, \quad v_j = N_j - K_j - 1.$$

### Results

The characteristics of the posterior distribution of  $\phi = (d, r)$  were investigated for simulated data and a real data set obtained from a chemical system. As a set of possible values of the threshold parameter  $r$ , the order statistics  $[Y^{(i)}, i = 1, \dots, q(\leq N)]$  of observations was used.

### Simulation Examples

Simulation results were based on both one realization and 100 realizations. The TMA(2;1,1) model was considered, where  $\theta^{(1)} = -0.4, \theta^{(2)} = 0.4, \sigma_1^2 = 1, \sigma_2^2 = 1, d = 1$ , and  $r = 0$ . A one realization was generated with series length of 50. As a set of possible values of the threshold parameter  $r, [r_0, r_L]$  was chosen as large as possible and the delay parameter  $d$  was selected as  $d = 1, 2, 3, 4$  and  $5$ .

Simulation results demonstrated that the posterior mass was concentrated at  $d=1, 2$  and  $3$ . Summary results of the joint posterior distribution of  $\phi = (d, r)$  are presented in Table 1. The marginal posterior distribution of  $d = 1, 2, 3$  have probabilities of 0.5566, 0.2500 and 0.1800 respectively. The posterior probability concentrates predominantly on few points, namely  $(d, r) = (1, 0.0036)$  and  $(1, 0.0350)$  with respective probabilities of 0.05582 and 0.06987.

Simulation results based on 100 realizations with series length of 100 were analyzed. For each realization,  $\phi = (d, r)$  was estimated based on modal value of the posterior distribution. The results yield relative frequencies of 87%, 7%, and 6% for  $d = 1, 2,$  and  $3$  respectively. The marginal posterior of  $r$  is shown in Figure 1; as expected, the model value is concentrated around the true threshold value  $r = 0$ .

#### Real Data Example

Series A, which consists of 197 observations and represents the concentration of a chemical process, was considered (Box & Jenkins, 1976). The differenced time series was considered. Fitting MA(1) model yields

$$X_t = -0.69Z_{t-1} + Z_t \quad (15)$$

Smadi (1997) used the MCMC technique for exploration of the posterior distribution of the threshold parameter  $r$ . The methodology proposed herein is applied to the differenced series in order to examine the posterior distribution of the threshold and the delay parameter  $\phi = (d, r)$ . The number of threshold points is reduced from 196 to 22 points because some differences have the same values. Values of  $[-0.4, 0.4]$  were assigned for  $[r_0, r_L]$ . For the delay parameter  $d$ , the set  $d = 1, 2, 3, 4$  and  $5$  were selected. It was found that the posterior mass was concentrated at  $d = 1, 2$  and  $3$ .

Summary results of the joint posterior distribution of  $\phi = (d, r)$  are presented in Table 2. It can be seen that the marginal posterior distribution of  $d = 1, 2, 3$  have probabilities of 0.8344, 0.13163 and 0.03398, respectively. Also, the posterior probability concentrates on

$(d, r) = (1, 0.0)$  with probability of 0.17; this corresponds to the largest mode of the posterior density. Conditioning on  $(d, r) = (1, 0.0)$ , the fitted TMA(2;1,1) model is

$$X_t = \begin{cases} -0.72Z_{t-1}^{(1)} + Z_t^{(1)} & \text{if } X_{t-1} \leq 0, \\ -0.66Z_{t-1}^{(2)} + Z_t^{(2)} & \text{if } X_{t-1} > 0. \end{cases}$$

where  $\hat{\sigma}_1^2 = 0.15$  and  $\hat{\sigma}_2^2 = 0.12$ .

#### Conclusion

From the proposed methodology and numerical results it can be concluded that the threshold moving average models are tractable from a Bayesian point of view. The nonlinearity threshold-type for moving average models can be detected by examining the marginal posterior distribution of the threshold parameter.

#### References

- Amendola, A., Niglio, M., & Vitale C. (2009). Statistical properties of threshold models. *Communication in Statistics-Theory and Methods*, 38, 2479-2497.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden-Day.
- Shaarawy, S., & Bromeling, L. (1984). Bayesian inferences and forecasts with moving average processes. *Communication in Statistics: Theory and Methods*, 13(15), 1871-1888.
- Cathy, W. S., Chen, ,& Lee, J. C. (1995). Bayesian inference of threshold models, *Journal of Time Series Analysis*, 16(5), 483-491.
- Geweke, J., & Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14(5), 441-54.
- Gooijer, J. D. (1998). On threshold moving-average models. *Journal of time series analysis*, 19(1), 1-18.
- Ling, S., & Tong, H. (2005). Testing for a linear MA model against threshold MA models. *Annals of Statistics*, 3, 2529-52.
- Safadi, T., & Morettin, P. A. (2000). Bayesian Analysis of threshold autoregressive moving average models. *Sankhya*, 62, Series B(3), 353-371.

BAYESIAN THRESHOLD MOVING AVERAGE MODELS

Table 1: Summary of Joint Posterior Densities of  $(r, d)$

$(r, d)$	Integrated Density	$(r, d)$	Integrated Density	$(r, d)$	Integrated Density
(-1.1832, 1)	0.02234	(-1.1832, 2)	0.00345	(-1.1832, 3)	0.01233
(-1.0101, 1)	0.02278	(-1.0101, 2)	0.00263	(-1.0101, 3)	0.00771
(-0.8562, 1)	0.02282	(-0.8562, 2)	0.00223	(-0.8562, 3)	0.00513
(-0.7346, 1)	0.01532	(-0.7346, 2)	0.00192	(-0.7346, 3)	0.00510
(-0.6573, 1)	0.01058	(-0.6573, 2)	0.00181	(-0.6573, 3)	0.00321
(-0.6326, 1)	0.00839	(-0.6326, 2)	0.00181	(-0.6326, 3)	0.00321
(-0.5574, 1)	0.00777	(-0.5574, 2)	0.00163	(-0.5574, 3)	0.00462
(-0.5478, 1)	0.00752	(-0.5478, 2)	0.00171	(-0.5478, 3)	0.00303
(-0.5237, 1)	0.00287	(-0.5237, 2)	0.00172	(-0.5237, 3)	0.00281
(-0.4982, 1)	0.00255	(-0.4982, 2)	0.00148	(-0.4982, 3)	0.00196
(-0.4631, 1)	0.00195	(-0.4631, 2)	0.00166	(-0.4631, 3)	0.00400
(-0.4336, 1)	0.00170	(-0.4336, 2)	0.00187	(-0.4336, 3)	0.00272
(-0.3709, 1)	0.00164	(-0.3709, 2)	0.00197	(-0.3709, 3)	0.00193
(-0.2798, 1)	0.00154	(-0.2798, 2)	0.00218	(-0.2798, 3)	0.00259
(-0.1939, 1)	0.00159	(-0.1939, 2)	0.00298	(-0.1939, 3)	0.00199
(-0.1918, 1)	0.00143	(-0.1918, 2)	0.00326	(-0.1918, 3)	0.00289
(-0.1837, 1)	0.00205	(-0.1837, 2)	0.00556	(-0.1837, 3)	0.00211
(-0.1450, 1)	0.00173	(-0.1450, 2)	0.00760	(-0.1450, 3)	0.00170
(-0.1430, 1)	0.00161	(-0.1430, 2)	0.00568	(-0.1430, 3)	0.00128
(-0.1151, 1)	0.00423	(-0.1151, 2)	0.00220	(-0.1151, 3)	0.01389
(-0.0884, 1)	0.00351	(-0.0884, 2)	0.00175	(-0.0884, 3)	0.02066
(-0.0560, 1)	0.00514	(-0.0560, 2)	0.00176	(-0.0560, 3)	0.01154
(-0.0124, 1)	0.00380	(-0.0124, 2)	0.00214	(-0.0124, 3)	0.01050
(0.0000, 1)	0.00296	(0.0000, 2)	0.00284	(0.0000, 3)	0.00881
(0.0036, 1)	0.05584	(0.0036, 2)	0.00492	(0.0036, 3)	0.00569
(0.0350, 1)	0.06987	(0.0350, 2)	0.00677	(0.0350, 3)	0.00355
(0.0879, 1)	0.03947	(0.0879, 2)	0.00821	(0.0879, 3)	0.00258
(0.1222, 1)	0.03186	(0.1222, 2)	0.01139	(0.1222, 3)	0.00189
(0.2192, 1)	0.02311	(0.2192, 2)	0.01434	(0.2192, 3)	0.00215
(0.2223, 1)	0.01459	(0.2223, 2)	0.00523	(0.2223, 3)	0.00155
(0.2748, 1)	0.01459	(0.2748, 2)	0.00523	(0.2748, 3)	0.00155
(0.3154, 1)	0.01221	(0.3154, 2)	0.00334	(0.3154, 3)	0.00129
(0.4112, 1)	0.01151	(0.4112, 2)	0.00458	(0.4112, 3)	0.00129
(0.4434, 1)	0.01184	(0.4434, 2)	0.00738	(0.4434, 3)	0.00115
(0.4484, 1)	0.00883	(0.4484, 2)	0.01028	(0.4484, 3)	0.00105
(0.4677, 1)	0.00626	(0.4677, 2)	0.01178	(0.4677, 3)	0.00102
(0.5505, 1)	0.01574	(0.5505, 2)	0.01142	(0.5505, 3)	0.00159
(0.5662, 1)	0.02498	(0.5662, 2)	0.02024	(0.5662, 3)	0.00177
(0.7154, 1)	0.01499	(0.7154, 2)	0.03736	(0.7154, 3)	0.00173
(0.8502, 1)	0.02341	(0.8502, 2)	0.0229	(0.8502, 3)	0.00286
(1.1930, 1)	0.01985	(1.1930, 2)	0.00789	(1.1930, 3)	0.01769

Figure 1: Marginal Posterior Density of  $r$

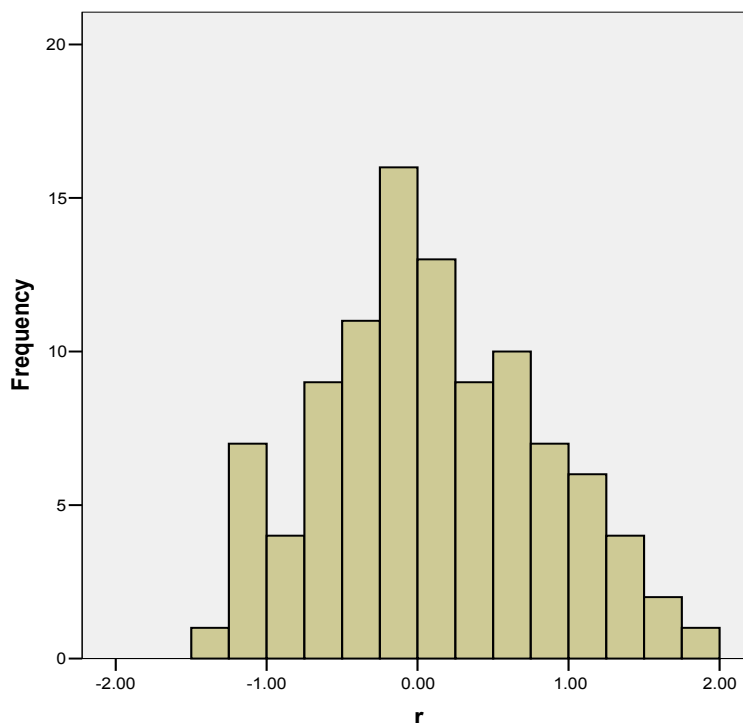


Table 2: Summary of Joint Posterior Densities of  $(d, r)$  for Chemical Data

$(d, r)$	Integrated Density	$(d, r)$	Integrated Density	$(d, r)$	Integrated Density
(1, -0.4)	0.03665	(2, -0.4)	0.01669	(2, -0.4)	0.00088
(1, -0.3)	0.15757	(2, -0.3)	0.01319	(2, -0.3)	0.00061
(1, -0.2)	0.15757	(2, -0.2)	0.01319	(2, -0.2)	0.00061
(1, -0.1)	0.09583	(2, -0.1)	0.00858	(2, -0.1)	0.00031
(1, 0.0)	0.17483	(2, 0.0)	0.01402	(2, 0.0)	0.00030
(1, 0.1)	0.04738	(2, 0.1)	0.02424	(2, 0.1)	0.00026
(1, 0.2)	0.03386	(2, 0.2)	0.02145	(2, 0.2)	0.00064
(1, 0.3)	0.03344	(2, 0.3)	0.01352	(2, 0.3)	0.00304
(1, 0.4)	0.09726	(2, 0.4)	0.00675	(2, 0.4)	0.02733

Smadi, M. M. (1997). *Bayesian inference of threshold autoregressive moving average models using sample-based methods*. Unpublished Ph. D. Dissertation, Colorado State University, Fort Collins, CO, U.S.A.

Tong H. (1990). *Nonlinear Time series models: A dynamical system approach*. London: Oxford University Press.

Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, B42*, 245-92.

Wang, W. Y., Du, J. G., & Xiang, J. T. (1984). Threshold autoregressive moving average model. *Journal of Computational Mathematics, 4*, 414-19 (in Chinese).



## Bayesian Regression Analysis with Examples in S-PLUS and R

S. P. Ahmad    A. A. Khan  
University of Kashmir,  
Srinagar, India

A. Ahmed  
Aligarh Muslim University,  
U.P, India

---

An extended version of normal theory Bayesian regression models, including extreme-value, logistic and normal regression models is examined. Methods proposed are illustrated numerically; the regression coefficient of  $pH$  on electrical conductivity (EC) of soil data is analyzed using both *S-PLUS* and *R* software.

Key words: Bayesian regression, extreme-value model, S-PLUS, R.

---

### Introduction

In statistics, regression analysis includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. In practice, many situations involve a heterogeneous population and it is important to consider the relationship of response variable  $y$  on concomitant variable  $x$  which is explicitly recognized.

One method to examine the relationship of a concomitant variable (or regressor variable) to a response variable  $y$  is through a regression model in which  $y$  has a distribution that depends upon the regressor variables. This involves specifying a model for the distribution of  $y$  given  $x$ , where  $x = (x_1, x_2, \dots, x_p)$  is a  $1 \times p$  vector of the regressor variables for an individual.

Let the distribution of  $y$  given  $x$  be

$$f(y|x, \beta, \sigma) = \frac{1}{\sigma} f\left(\frac{y - x\beta}{\sigma}\right), \quad (1.1)$$

where  $\beta$  is a  $p \times 1$  vector of regression

Coefficients,  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$   
and  $E[y|x, \beta] = x\beta$ . The alternative form of (1.1) is

$$y = x\beta + \sigma z \quad (1.2)$$

where

$$z = \frac{y - x\beta}{\sigma}$$

has the standardized distribution with density function  $f(z)$ . The family of models for which  $f(z)$  has a standard normal distribution is common in statistical literature (Searle, 1971; Rao, 1973; Seber, 1977; Draper & Smith, 1981; Weisberg, 1985 ) but models in which  $z$  has other distributions belonging to location-scale family (1.2) are also important. For example, extreme value regression models are employed in applications ranging from accelerated life testing (Lawless, 2003; Zelen, 1959) to the analysis of survival data on patients suffering from chronic diseases (Prentice, 1973; Feigl & Zelen, 1965; Krall, et al., 1975).

Furthermore, if data is contaminated with outliers, then the normal distribution can be replaced with Student's  $t$  distribution (with small degrees of freedom) to have a better fit (e.g., Lange, et al., 1989). Model (1.2) has the ability to accommodate linear as well as non-linear models for the various functional forms of  $x\beta$ . None of the above authors present a Bayesian approach. Box and Tiao (1973) and Gelman, et al. (1995) discuss this approach of regression

---

S. P. Ahmad is an Assistant Professor in the Department of Statistics. Email: sprvz@yahoo.com. A. A. Khan is a Professor in the Department of Statistics & O. R. Email: atharkhan1962@gmail.com. A. Ahmed is a Professor in the Department of Statistics. Email: aqlstat@yahoo.co.in.

analysis to deal with normal linear as well as non-linear non-normal models. Zellener (1971) describes Bayesian inference in reference to econometrics, but the discussion is mainly confined to normal linear models. The general framework used for casual inference is presented in Rubin (1974b, 1978a). Bayesian approaches to analyzing regression residuals appear in Zellener (1976), Chaloner and Brant (1988) and Chaloner (1991).

Joint inference for  $\beta$  and  $\log\sigma$  with Non-informative Prior  $p(\beta, \log\sigma)$

Suppose  $(y_i, x_i)$  for  $i = 1, 2, \dots, n$  is assumed to be a random sample from location-scale family of models in (1.1) and likelihood is

$$\prod_{i=1}^n f(y_i | x_i, \beta, \sigma)$$

This implies that

$$l(\beta, \log\sigma) = \sum_{i=1}^n \log f(z_i) - n \log \sigma \tag{2.1}$$

where

$$z_i = \frac{y_i - x_i \beta}{\sigma}$$

Consider the non-informative prior

$$p(\beta, \log\sigma) = 1 \tag{2.2}$$

The joint posterior density of  $\beta$  and  $\log\sigma$  given data vector  $y^T = (y_1, y_2, \dots, y_n)$  is

$$p(\beta, \log\sigma | x, y) \propto \prod_{i=1}^n f(y_i | x_i, \beta, \log\sigma) p(\beta, \log\sigma) \tag{2.3}$$

where  $x = (x_1, x_2, \dots, x_n)^T$  is a  $n \times p$  matrix of covariates (or regressors) corresponding to response vector  $y$ . Now joint inference for  $\beta$  and  $\log\sigma$  can be made from posterior (2.3).

Posterior mode  $(\hat{\beta}, \log \hat{\sigma})^T$  of  $p(\beta, \log\sigma | x, y)$  serves as a point estimate of  $\beta$  and  $\log\sigma$ . Its calculations require partial derivatives of log posterior

$$\begin{aligned} l^*(\beta, \log\sigma) &= l(\beta, \log\sigma) + \log p(\beta, \log\sigma) \\ &= l(\beta, \log\sigma) \end{aligned} \tag{2.4}$$

Defining partial derivatives as  $l_{\beta}^* = \frac{\partial l^*}{\partial \beta}$ , a vector of  $(p \times 1)$  partial derivatives,

$$l_{\phi}^* = \frac{\partial l^*}{\partial \phi}, \text{ a scalar and } \phi = \log\sigma$$

$$l_{\beta\phi}^* = \frac{\partial^2 l^*}{\partial \beta \partial \phi}, \text{ a } (p \times 1) \text{ vector,}$$

$$l_{\phi\beta}^* = \frac{\partial^2 l^*}{\partial \phi \partial \beta}, \text{ a } (1 \times p) \text{ vector,}$$

$$l_{\beta\beta}^* = \frac{\partial^2 l^*}{\partial \beta \partial \beta^T}, \text{ a } (p \times p) \text{ matrix, and}$$

$$l_{\phi\phi}^* = \frac{\partial^2 l^*}{\partial \phi \partial \phi^T}.$$

These derivatives can be defined more explicitly as:

$$l_{\beta}^* = l_{\beta}$$

$$l_{\phi}^* = l_{\phi}$$

$$l_{\beta\phi}^* = l_{\beta\phi}$$

$$l_{\phi\beta}^* = l_{\phi\beta}$$

$$l_{\beta\beta}^* = l_{\beta\beta} \text{ and}$$

$$l_{\phi\phi}^* = l_{\phi\phi}.$$

Consequently, score vector  $U(\beta, \phi)$  and Hessian matrix  $H(\beta, \phi)$  are a  $(p+1) \times 1$  vector

$$U(\beta, \phi) = \begin{bmatrix} l_{\beta}^* \\ l_{\phi}^* \end{bmatrix},$$

and a  $(p+1) \times (p+1)$  matrix

$$H(\beta, \phi) = \begin{bmatrix} l_{\beta\beta}^* & l_{\beta\phi}^* \\ l_{\phi\beta}^* & l_{\phi\phi}^* \end{bmatrix},$$

therefore, making use of Newton-Raphson iteration scheme, results in posterior mode vector  $(\hat{\beta}, \hat{\phi})^T$  as

$$\begin{bmatrix} \hat{\beta} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \phi_0 \end{bmatrix} - H^{-1}(\beta_0, \phi_0) \begin{bmatrix} l_{\beta}^* \\ l_{\phi}^* \end{bmatrix} \quad (2.5)$$

where  $\hat{\phi} = \log \hat{\sigma}$ .

The asymptotic posterior covariance matrix of (2.3) can be obtained as

$$\begin{aligned} I^{-1}(\hat{\beta}, \hat{\phi}) &= -H^{-1}(\hat{\beta}, \hat{\phi}) \\ &= \sum (\hat{\beta}, \hat{\phi}). \end{aligned}$$

More clearly, posterior density

$$\begin{aligned} p(\beta, \phi | x, y) &\cong \\ N_{p+1} \left( (\hat{\beta}, \hat{\phi})^T, I^{-1}(\hat{\beta}, \hat{\phi}) \right) &\left( 1 + O(n^{-\frac{1}{2}}) \right) \end{aligned} \quad (2.6)$$

where  $N_r(a, b)$  is the  $r$ -variate normal distribution with mean vector  $a$  and a covariance vector  $b$ . This is a first order approximation of the posterior density (e.g., Berger, 1985). An

equivalent version of this approximation is the Chi-square approximation, i.e.,

$$W(\beta, \phi) = -2 \left[ l^*(\beta, \phi) - l^*(\hat{\beta}, \hat{\phi}) \right] \approx \chi_{p+1}^2.$$

A more accurate approximation, Laplace's approximation (Tierney & Kadane, 1986; Reid, 1988) can be also used, i.e.,

$$\begin{aligned} p(\beta, \phi | x, y) &\cong \\ (2\pi)^{-\frac{p+1}{2}} |I(\hat{\beta}, \hat{\phi})|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} W(\hat{\beta}, \hat{\phi}) \right] &\left( 1 + O(n^{-1}) \right) \end{aligned} \quad (2.7)$$

Any of the approximations can be used both for hypothesis testing and construction of credible regions.

The Marginal Inference for  $\beta$  and  $\phi$  ( $\phi = \log \sigma$ )  
The marginal densities for  $\beta$  and  $\phi$  are

$$p(\beta | x, y) = \int p(\beta, \phi | x, y) d\phi. \quad (3.1)$$

Similarly, marginal posterior of  $\phi$  can be obtained by

$$p(\phi | x, y) = \int p(\beta, \phi | x, y) d\beta. \quad (3.2)$$

Bayesian analysis is to be based on these two posteriors. For the normal model,  $p(\beta | x, y)$  and  $p(\phi | x, y)$  can be obtained in closed form (e.g., Zellener, 1971). However, for non-normal members of location-scale family, these marginals can be obtained through numerical integration only (e.g. Naylor & Smith, 1982). The alternative approach is to deal with asymptotic theory approach (e.g., Tierney, Kass & Kadane, 1989a; Leonard, et al., 1989). Normal and Laplace's approximations can be written directly for posterior densities  $p(\beta | x, y)$  and  $p(\phi | x, y)$  as under:

a) Normal Approximation:

Marginal posterior density of  $\beta$  can be approximated by normal distribution, i.e.,

$$p(\beta | x, y) \cong N_p(\hat{\beta}, I_{11}^{-1}) \quad (3.3)$$

where  $\hat{\beta}$  is the posterior mode and  $I_{11}^{-1}$  is a  $p \times p$  matrix defined as

$$I^{-1}(\hat{\beta}, \hat{\phi}) = \begin{bmatrix} I_{11}^{-1} & I_{12}^{-1} \\ I_{21}^{-1} & I_{22}^{-1} \end{bmatrix}$$

where  $\hat{\phi} = \log \hat{\sigma}$  and suffixes 1 and 2 to  $I$  stand for  $\hat{\beta}$  and  $\hat{\phi}$ , respectively. This approximation is equivalent to the Chi square approximation defined as:

$$(\beta - \hat{\beta})^T I_{11} (\beta - \hat{\beta}) \approx \chi_p^2.$$

Corresponding approximations for  $p(\phi | x, y)$  can be written as:

$$p(\phi | x, y) = N_1(\hat{\phi}, I_{22}^{-1}) \quad (3.4)$$

This is equivalent to the Chi square approximation, i.e.,

$$(\phi - \hat{\phi})^T I_{22} (\phi - \hat{\phi}) \approx \chi_1^2.$$

b) Laplace's Approximation:

Laplace's approximation can also be used to approximate marginal density of  $\beta$ , i.e.,

$$p(\beta | y) \cong \left[ \frac{|I(I(\hat{\beta}, \hat{\phi}))|}{2\pi |I(\beta, \hat{\phi}(\beta))|} \right]^{\frac{1}{2}} \exp[l^*(\beta, \hat{\phi}(\beta)) - l^*(\hat{\beta}, \hat{\phi})] \quad (3.5)$$

where  $\hat{\phi}(\beta)$  is the posterior mode of  $\phi$  for a fixed  $\beta$ .

Corresponding approximation for  $p(\phi | x, y)$  can also be written as

$$p(\phi | x, y) \cong$$

$$(2\pi)^{-\frac{p}{2}} \left[ \frac{|I(\hat{\beta}, \hat{\phi})|}{|I(\beta, \hat{\phi}(\beta))|} \right]^{\frac{1}{2}} \exp[l^*(\hat{\beta}(\phi), \phi) - l^*(\hat{\beta}, \hat{\phi})] \quad (3.6)$$

where  $\hat{\beta}(\phi)$  is the posterior mode of  $\beta$  for a fixed  $\phi$ .

Bayesian Regression Analysis of the Extreme-Value Model

Let  $y$  be the response vector and  $x_i$  be the vector for the  $i^{\text{th}}$  observation. Assume that

$$z_i = \frac{y_i - x_i^T \beta}{\sigma} \sim f \quad (4.1)$$

for some  $f$  (extreme value distribution). Consequently, in terms of general notation  $\theta = (\beta, \sigma)^T$ , a vector of length  $(\beta + 1)$  and likelihood is given by:

$$\prod_{i=1}^n f(y_i | x_i, \beta, \sigma).$$

This implies that

$$\begin{aligned} l(\beta, \sigma) &= \log \prod_{i=1}^n f(y_i | x_i, \beta, \sigma) \\ &= \sum_{i=1}^n \log f(z_i) - n \log \sigma \end{aligned} \quad (4.2)$$

where  $z_i$  is defined in (4.1).

Taking partial derivatives with respect to  $\mu$  and  $\sigma$  results in

$$\begin{aligned} l_{\beta} &= \frac{\partial l}{\partial \beta} \\ &= -\frac{1}{\sigma} \sum_{i=1}^n (1 - e^{z_i}) x_i^T \end{aligned}$$

$$\begin{aligned}
 l_\sigma &= \frac{\partial l}{\partial \sigma} \\
 &= -\frac{1}{\sigma} \sum_{i=1}^n z_i (1 - e^{z_i}) - \frac{n}{\sigma} \\
 l_{\beta\sigma} &= \frac{\partial^2 l}{\partial \beta \partial \sigma} \\
 &= -\frac{1}{\sigma^2} \sum_{i=1}^n (z_i e^{z_i} + e^{z_i} - 1) x_i^T \\
 l_{\sigma\beta} &= \frac{\partial^2 l}{\partial \sigma \partial \beta} \\
 &= -\frac{1}{\sigma^2} \sum_{i=1}^n (z_i e^{z_i} + e^{z_i} - 1) x_i^T \\
 l_{\beta\beta} &= \frac{\partial^2 l}{\partial \beta \partial \beta^T} \\
 &= -\frac{1}{\sigma^2} \sum_{i=1}^n e^{z_i} x_i^T x_i
 \end{aligned}$$

and

$$\begin{aligned}
 l_{\sigma\sigma} &= \frac{\partial^2 l}{\partial \sigma \partial \sigma} \\
 &= -\frac{1}{\sigma^2} \sum_{i=1}^n (z_i^2 e^{z_i} + 2z_i e^{z_i} - 2z_i) + \frac{n}{\sigma^2}.
 \end{aligned}$$

Following the standard approach of Box and Tiao (1973) and Gelman, et al. (1995), assuming the prior

$$p(\beta, \sigma) \cong p(\beta)p(\sigma) \quad (4.3)$$

where  $p(\beta)$  and  $p(\sigma)$  are priors for  $\beta$  and  $\sigma$ . Using Bayes theorem obtain the posterior density  $p(\beta, \sigma | y)$  is obtained as

$$p(\beta, \sigma | x, y) \propto \prod_{i=1}^n f(y_i | x_i, \beta, \sigma) p(\beta, \sigma) \quad (4.4)$$

The log-posterior is given by

$$\begin{aligned}
 \log p(\beta, \sigma | x, y) &= \\
 &= \log \prod_{i=1}^n p(y_i | x_i, \beta, \sigma) + \log p(\beta) + \log p(\sigma)
 \end{aligned}$$

or

$$l^*(\beta, \sigma) = l(\beta, \sigma) + \log p(\beta) + \log p(\sigma). \quad (4.5)$$

For a prior  $p(\beta, \sigma) \cong p(\beta)p(\sigma) = 1$ ,  $l_\beta^* = l_\beta$ ,  $l_\sigma^* = l_\sigma$ ,  $l_{\beta\sigma}^* = l_{\beta\sigma}$ ,  $l_{\sigma\beta}^* = l_{\sigma\beta}$ ,  $l_{\beta\beta}^* = l_{\beta\beta}$  and  $l_{\sigma\sigma}^* = l_{\sigma\sigma}$ . The posterior mode is obtained by maximizing (4.5) with respect to  $\beta$  and  $\sigma$ . The score vector of the log posterior is given by

$$U(\beta, \sigma) = (l_\beta^*, l_\sigma^*)^T$$

and the Hessian matrix of log posterior is

$$H(\beta, \sigma) = \begin{bmatrix} l_{\beta\beta}^* & l_{\beta\sigma}^* \\ l_{\sigma\beta}^* & l_{\sigma\sigma}^* \end{bmatrix}.$$

Posterior mode  $(\hat{\beta}, \hat{\sigma})$  can be obtained from iteration scheme

$$\begin{bmatrix} \hat{\beta} \\ \hat{\sigma} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \sigma_0 \end{bmatrix} - H^{-1}(\beta_0, \sigma_0) \begin{bmatrix} l_\beta^* \\ l_\sigma^* \end{bmatrix}. \quad (4.6)$$

Consequently, the modal variance  $\Sigma$  can be obtained as

$$I^{-1}(\hat{\beta}, \hat{\sigma}) = -H^{-1}(\hat{\beta}, \hat{\sigma}).$$

Using the normal approximation, a bivariate normal approximation of  $p(\beta, \sigma | x, y)$  can be directly written as:

$$p(\beta, \sigma | y) \cong N_{p+1} \left( (\hat{\beta}, \hat{\sigma})^T, I^{-1}(\hat{\beta}, \hat{\sigma}) \right) \left( 1 + O(n^{-\frac{1}{2}}) \right). \quad (4.7)$$

Similarly, a Bayesian analog of likelihood ratio criterion can be written as:

$$W(\beta, \sigma) = -2[l^*(\mu, \sigma) - l^*(\hat{\mu}, \hat{\sigma})] \approx \chi_{p+1}^2 \quad (4.8)$$

Using Laplace's approximation,  $p(\beta, \sigma | x, y)$  can be written as:

$$p(\beta, \sigma | x, y) \cong (2\pi)^{-\frac{p-1}{2}} |I(\hat{\beta}, \hat{\sigma})|^{-\frac{1}{2}} \exp[l^*(\mu, \sigma) - l^*(\hat{\mu}, \hat{\sigma})] (1 + O(n^{-1})). \quad (4.9)$$

The marginal Bayesian inference about  $\beta$  and  $\sigma$  is based on the marginal posterior densities of these parameters. Marginal posterior for  $\beta$  can be obtained after integrating out  $p(\beta, \sigma | x, y)$  with respect to  $\sigma$ ,

$$p(\beta | x, y) = \int p(\beta, \sigma | x, y) d\sigma. \quad (4.10)$$

Similarly, the marginal posterior of  $\sigma$  can be obtained by:

$$p(\sigma | x, y) = \int p(\beta, \sigma | x, y) d\beta. \quad (4.11)$$

The normal approximation for marginal posterior  $p(\beta | x, y)$  can be written as:

$$p(\beta | x, y) = N_p(\hat{\beta}, I_{11}^{-1}) \quad (4.12)$$

where  $\hat{\beta}$  is the posterior mode and  $I_{11}^{-1}$  is a  $(p \times p)$  matrix defined as

$$I^{-1}(\hat{\beta}, \hat{\sigma}) = \begin{bmatrix} I_{11}^{-1} & I_{12}^{-1} \\ I_{21}^{-1} & I_{22}^{-1} \end{bmatrix}.$$

The Bayesian analog of likelihood ratio criterion can also be defined as a test criterion as:

$$(\beta - \hat{\beta})^T I_{11} (\beta - \hat{\beta}) \approx \chi_p^2. \quad (4.13)$$

Laplace's approximation of marginal posterior density  $p(\beta | x, y)$  can be given by:

$$p(\beta | x, y) \cong \left[ \frac{|I(\hat{\beta}, \hat{\sigma})|}{2\pi |I(\beta, \hat{\sigma}(\beta))|} \right]^{\frac{1}{2}} \exp[l^*(\beta, \hat{\sigma}(\beta)) - l^*(\hat{\beta}, \hat{\sigma})]. \quad (4.14)$$

Similarly,  $p(\sigma | x, y)$  can be approximated and results corresponding to normal and Laplace's approximation can be written as

$$p(\sigma | x, y) = N_1(\hat{\sigma}, I_{22}^{-1}) \quad (4.15)$$

or equivalently,

$$(\sigma - \hat{\sigma})^T I_{22} (\sigma - \hat{\sigma}) \approx \chi_1^2 \quad (4.16)$$

$p(\sigma | x, y) \cong$

$$(2\pi)^{-\frac{p}{2}} \left[ \frac{|I(\hat{\beta}, \hat{\sigma})|}{|I(\hat{\beta}(\sigma), \sigma)|} \right]^{\frac{1}{2}} \exp[l^*(\hat{\beta}(\sigma), \sigma) - l^*(\hat{\beta}, \hat{\sigma})]. \quad (4.17)$$

### Numerical Illustrations

Numerical illustrations are implemented in S-PLUS software for Bayesian regression analysis. These illustrations are show the strength of Bayesian methods in practical situations. Soil samples were collected from rice growing areas as well as fruit orchids of Kashmir valley and were analyzed for some relevant parameters. In our work, we studied pH and E.C in the soil of Kashmir valley. The functions *survReg* and *condnsorReg* were used for Bayesian analysis of various regression models with non-informative prior. S-PLUS has a function *condnsorReg* for regression analysis; this has a very substantial overlap with *survReg* but is more general in that it allows truncation as well as censoring (Venables & Ripley, 2002). The usage of *survReg* and *condnsorReg* are:

## BAYESIAN REGRESSION ANALYSIS WITH EXAMPLES IN S-PLUS AND R

*survReg(formula, data, dist)*

*sensorReg(formula, data, dist)*

where

- *formula*: a formula expression as for other regression models;
- *data*: optional data frame in which to interpret the variable occurring in the formula; and
- *dist*: assumed distribution for y variable.

### References

- Berger, J. O. (1985). *Statistical decision theory and bayesian analysis, 2<sup>nd</sup> Edition*. New York: Springer-Verlag.
- Box, G. E. P., & Tiao, G. C. (1973). *r*
- Chaloner, K. (1991). Bayesian residual analysis in the presence of censoring. *Biometrika*, 78, 637-644.
- Chaloner, K., & Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika*, 75, 651-659.
- Draper, N. R., & Smith, H., Jr. (1981). *Applied regression analysis*. New York: Wiley.
- Feigl, P., & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21, 826-836.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.
- Krall, J., Uthoff, V., & Harley, J. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics*, 31, 49-57.
- Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84, 881-896.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. New York: Wiley.
- Leonard, T., Hsu, J. S. J., & Tsui, K. W. (1989). Bayesian marginal inference. *Journal of the American Statistical Association*, 84, 1051-1058.
- Naylor, J. C., & Smith, A. F. M. (1982). Applications of a method for efficient computation of posterior distributions. *Applied Statistics*, 31, 214-225.
- Prentice, R. L. (1973). Exponential survival with censoring and explanatory variables. *Biometrika*, 60, 279-288.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
- Reid, N. (1988). Saddle point methods and statistical inference. *Statistical Science*, 3, 213-238.
- Rubin, D. B. (1974b). Estimating casual effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1978a). Bayesian inference for casual effects: the role of randomization. *The Annals of Statistics*, 6, 34-58.
- Searle, S. R. (1971). *Linear models*. New York: Wiley.
- Seber, G.A.F. (1977). *Linear regression analysis*. New York: Wiley.
- Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82-86.
- Tierney, L., Kass, R. E., & Kadane, J. B. (1989a). Fully exponential Laplace approximations to the expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84, 710-716.
- Venables, W. N., & Ripley, B. D. (2002). *S programming*. New York: Springer-Verlag.
- Weisberg, S. (1985). *Applied linear regression*. New York: Wiley.
- Zelen, M. (1959). Factorial experiments in life testing. *Technometrics*, 1, 269-288.
- Zellener, A. (1971). *An introduction to bayesian inference in econometrics*. New York: Wiley.
- Zellener, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *Journal of the American Statistical Association*, 71, 400-405.

Table 1: A Summary of Derivatives of Log-Likelihoods

Derivatives	Distributions		
	Normal	Extreme-Value	Logistic
$l_\beta$	$\frac{1}{\sigma} \sum_{i=1}^n z_i x_i^T$	$-\frac{1}{\sigma} \sum_{i=1}^n (1 - e^{z_i}) x_i^T$	$\frac{1}{\sigma} \sum_{i=1}^n \left( \frac{e^{z_i} - 1}{e^{z_i} + 1} \right) x_i^T$
$l_\sigma$	$\frac{1}{\sigma} \sum_{i=1}^n z_i^2 - \frac{n}{\sigma}$	$-\frac{1}{\sigma} \sum_{i=1}^n z_i (1 - e^{z_i}) - \frac{n}{\sigma}$	$\frac{1}{\sigma} \sum_{i=1}^n z_i \left( \frac{e^{z_i} - 1}{e^{z_i} + 1} \right) - \frac{n}{\sigma}$
$l_{\beta\sigma}$	$-\frac{2}{\sigma^2} \sum_{i=1}^n z_i x_i^T$	$-\frac{1}{\sigma^2} \sum_{i=1}^n (z_i e^{z_i} + e^{z_i} - 1) x_i^T$	$-\frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{e^{2z_i} + 2z_i e^{z_i} - 1}{(e^{z_i} + 1)^2} \right) x_i^T$
$l_{\sigma\beta}$	$-\frac{2}{\sigma^2} \sum_{i=1}^n z_i x_i^T$	$-\frac{1}{\sigma^2} \sum_{i=1}^n (z_i e^{z_i} + e^{z_i} - 1) x_i$	$-\frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{e^{2z_i} + 2z_i e^{z_i} - 1}{(e^{z_i} + 1)^2} \right) x_i$
$l_{\beta\beta}$	$-\frac{2}{\sigma^2} \sum x_i^T x_i$	$-\frac{1}{\sigma^2} \sum_{i=1}^n e^{z_i} x_i^T x_i$	$-\frac{2}{\sigma^2} \sum_{i=1}^n \left( \frac{e^{z_i} x_i^T x_i}{(e^{z_i} + 1)^2} \right)$
$l_{\sigma\sigma}$	$-\frac{3}{\sigma^2} \sum_{i=1}^n z_i^2 + \frac{n}{\sigma^2}$	$-\frac{1}{\sigma^2} \sum_{i=1}^n (z_i^2 e^{z_i} + 2z_i e^{z_i} - 2z_i) + \frac{n}{\sigma^2}$	$-\frac{2}{\sigma^2} \sum_{i=1}^n z_i \left( \frac{e^{2z_i} + z_i e^{z_i} - 1}{(e^{z_i} + 1)^2} \right) + \frac{n}{\sigma^2}$

where  $z_i = \frac{y_i - x_i \beta}{\sigma}$ ,  $i = 1, 2, \dots, n$ .

Table 2: A Summary of Prior Densities for Location Parameter  $\beta$

Name of Density	$p(\mu)$	$\frac{p'(\mu)}{p(\mu)}$	$\left[ \frac{p'(\mu)}{p(\mu)} \right]'$
Non-Informative	Constant	Zero	Zero
Normal	$c \exp\left(-\frac{1}{2} D^T D\right)$	$-\frac{D}{\sigma_0}$	$-\frac{1}{\sigma_0^2} I$

Where  $D^T = (D_1, D_2, \dots, D_p)$  a  $(p \times 1)$  vector,  $D_i = \frac{\beta_i - \beta_{0i}}{\sigma_{oi}}$  for  $i = 1, 2, \dots, p$ ;  $I$  stands for identity  $(p \times p)$  matrix and  $c$  is the normalizing constant.



## BAYESIAN REGRESSION ANALYSIS WITH EXAMPLES IN S-PLUS AND R

Table 3: Regression Coefficient of pH on EC for Various Models

Regression Model	$\hat{\beta}$ (Intercept)		Posterior Std.	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
Extreme-Value	6.71	2.32	0.0447	0.2881
Logistic	6.29	2.41	0.0365	0.2332
Normal	6.33	2.00	0.0335	0.1890

Table 4: Approximate Normal Posterior Quantiles for Regression Coefficient of Various Models

Model	Posterior	Posterior Quantile					
		0.025	0.25	0.50	0.75	0.95	0.975
Extreme-Value	Normal	6.63	6.68	6.72	6.75	6.79	6.80
Logistic	Normal	6.16	6.20	6.23	6.25	6.29	6.30
Normal	Normal	6.26	6.31	6.33	6.35	6.39	6.40

## Estimating Internal Consistency Using Bayesian Methods

Miguel A. Padilla  
Old Dominion University

Guili Zhang  
East Carolina University

---

Bayesian internal consistency and its Bayesian credible interval (BCI) are developed and Bayesian internal consistency and its percentile and normal theory based BCIs were investigated in a simulation study. Results indicate that the Bayesian internal consistency is relatively unbiased under all investigated conditions and the percentile based BCIs yielded better coverage performance.

Key words: Bayesian internal consistency, coefficient alpha, confidence interval, Bayesian confidence interval, coverage probability.

---

### Introduction

Psychological constructs are the building blocks of psychological/behavioral research. Indeed, one can easily argue that constructs are the foundation of these two sciences. A typical way of measuring a construct is through a questionnaire containing items that are purported to indirectly measure the construct of interest; thus, it becomes important that the items be consistent or reliable so that the questionnaire itself is consistent or reliable. Although there are several methods of measuring or estimating the reliability of a questionnaire, by far the most commonly used is coefficient alpha.

Coefficient alpha has remained popular since its introduction in Cronbach's (1951) article based on the work of Guttman and others in the 1940s (Guttman, 1945). Coefficient alpha is a measure of internal consistency for a group of items (i.e., questions) that are related in that they measure the same psychological/behavioral construct (Cortina, 1993). There are three main reasons for coefficient alpha's popularity. First,

coefficient alpha is computationally simple. The only required quantities for its computation are the number of items, variance for each item and the total joint variance for all the items; quantities that can easily be extracted from the item covariance matrix. Second, coefficient alpha can be computed for continuous or binary data: this is a significant advantage when working with right/wrong, true/false, etc. items. Third, it only requires one test administration: Most other forms of reliability require at least two test administrations, which come at a cost of time and resources. For these reasons coefficient alpha's power to assess the psychometric property of the reliability of a measurement instrument is widely used and it has remained relatively unchanged for over 60 years.

The advent of Bayesian methodology has brought about exciting and innovating ways of thinking about statistics and analyzing data. Bayesian methods have several advantages over traditional statistics, sometimes referred as frequentist statistics (Gelman, 2004; Lee, 2004), but two advantages stand out. First, researchers can now incorporate prior knowledge or beliefs about a parameter by specifying a prior distribution for the parameter in the model; thus, the analysis is now composed of data and prior knowledge and/or beliefs. By contrast, traditional or frequentist analyses are composed only of data. Through this combination of data and prior knowledge, more can be learned about the phenomenon under study and knowledge about the phenomenon can be updated accordingly. Second, Bayesian methods provide

---

Miguel A. Padilla is an Assistant Professor of Quantitative Psychology in the Department of Psychology. His research interests are in applied statistics and psychometrics. Email him at: mapadill@odu.edu. Guili Zhang is an Assistant Professor of Research Methodology in the Department of Curriculum and Instruction. Her research interests are in applied statistics. Email her at: ZHANGG@ecu.edu.

credible intervals (BCIs), the Bayesian analog to confidence intervals (CIs). However, credible intervals have a different interpretation from confidence intervals. A confidence interval allows one to make statements, such as “we are 95% confident that the interval captures” the true population parameter. By contrast, a BCI allows one to say that “we are 95% confident that the true population parameter lies between the bands of the credible interval,” a simpler and more powerful statement. This is, in fact, the interpretation most researchers would like to make with confidence intervals.

A Bayesian coefficient alpha retains the simplicity and power of the original coefficient alpha, but it also has the advantages of Bayesian methodology. By incorporating prior internal consistency information into the current estimation of coefficient alpha, more can be learned about the internal consistency of a measurement instrument and knowledge about the instrument can be updated accordingly. Additionally, credible intervals are generated for the Bayesian coefficient alpha. The bootstrap is the common method for generating confidence intervals for coefficient alpha; however, the bootstrap confidence interval has the same interpretation as the confidence interval from traditional statistical methods. With credible intervals direct statements can be made about where the true population coefficient alpha lies, which is a clear advantage over the standard confidence interval.

#### Prior Research on Coefficient Alpha CIs

As with all statistics, coefficient alpha is a population parameter and must be estimated from samples; thus, it is subject to sampling error that contributes to the variability around the true population parameter. Due to this, current statistical thinking and practice point to the need for providing confidence intervals to supplement point estimates and statistical tests (Duhachek, Coughlan & Iacobucci, 2005; Duhachek & Iacobucci, 2004; Iacobucci & Duhachek, 2003; Maydeu-Olivares, Coffman & Hartmann, 2007). Many professional publications are beginning to require authors to provide CIs in addition to point estimates, standard errors and test statistics. For example, the American Psychological Association Task

Force on Statistical Inference (Wilkinson, 1999) emphasizes the obligation of researchers to provide CIs for all principal outcomes; however, generating CIs for coefficient alpha has remained somewhat elusive and rarely implemented in practice.

Confidence intervals for coefficient alpha were first introduced by Kristof (1963) and Feldt (1965). These CIs assume that items are normally distributed and strictly parallel (Allen & Yen, 1979; Crocker & Algina, 1986), which implies that the item covariance matrix is compound symmetric; i.e.,  $\sigma_1 + \sigma^2 I (i = j)$  where  $\sigma_1$  are the item variances, and  $\sigma^2$  are the item covariances, and  $I(\cdot)$  is the indicator function. These confidence intervals, however, do not perform well when items are not strictly parallel (Barchard & Hakstian, 1997). Given that the strictly parallel assumption is unreasonable in applied research and that these CIs do not perform well when this assumption is violated may be the reason why coefficient alpha CIs are rarely implemented in applied research (Duhachek & Iacobucci, 2004).

An improvement to the CIs proposed by Kristof (1963) and Feldt (1965) was introduced by van Zyl, Neudecker and Nel (2000) who showed that the standard method of estimating coefficient alpha is a maximum likelihood estimator (MLE) and derived its corresponding CIs. Although the coefficient alpha MLE assumes that items are normally distributed, a major advantage is that it does not require the compound symmetry assumption of the item covariance matrix. In a simulation study, Duhacket and Iacobucci (2004) compared the performance of the coefficient alpha CIs for the method proposed by Feldt (1965) and the MLE proposed by van Zyl, et al. (2000) under a non-parallel measurement model. Their results indicate that the MLE method consistently outperformed the competing methods across all simulation conditions, but because the MLE method assumes that items are normally distributed, when the assumption is violated, the results can be untrustworthy.

Normally distributed items are not a completely realistic assumption in psychological/behavioral research. Most items in measurement instruments are dichotomous

(yes/no, true/false, etc.) or Likert-type items with several ordinal items: for these item types, normality is an unrealistic assumption. From this perspective, Yuan and Bentler (2002) extended the results of the coefficient alpha MLE to a wider range of distributions, pointing out that it is robust to some violations of normality. However, they point out that it is difficult to verify conditions to which the coefficient alpha MLE is robust to item non-normality. Thus, if the conditions cannot be verified theoretically then they are even more difficult to verify in applied work.

Yuan and Bentler (2003) built on this by introducing what Maydeu-Olivares and colleagues (2007) call asymptotically distribution-free (ADF) CIs for coefficient alpha. In this study the authors compared the ADF, MLE, and bootstrap coefficient alpha CIs estimated from the Hopkins Symptom Checklist (HSCL; Derogatis, Lipman, Rickels, Uhlenhuth & Covi, 1974). The results of Yuan and Bentler suggest that the ADF CIs are between the MLE and bootstrap methods in terms of their accuracy. However, they point out that the ADF CIs cannot describe the tail behavior of coefficient alpha of the HSCL due to the small sample ( $n = 419$ ); they suggest that with a larger sample size the ADF CIs could better describe the distribution of coefficient alpha.

Maydeu-Olivares, et al. (2007) extended the work by Yuan, et al. (2003) by simplifying the computation of ADF CIs and investigating its performance under several simulation conditions. Of particular interest was the comparison of the ADF CIs to the MLE CIs under various conditions of skewness and kurtosis. In general, they concluded that - with approximately normal items - the MLE CIs perform well even with a sample size as small as 50. However, once the items begin to deviate from normality, the ADF CIs begin to outperform the MLE CIs. In particular, the ADF CIs outperform MLE CIs with as little a sample size of 100. When the sample size gets larger than 100 the ADF CIs perform well regardless of the skewness and kurtosis investigated by the researchers.

Recent research has thus been fruitful in investigating the properties of coefficient alpha CIs; however, these CIs are based on traditional

frequentist statistics. As such, they have the traditional interpretation of CIs and cannot be updated with prior information. The primary purpose of this study is to develop a Bayesian internal consistency estimate and to evaluate its performances by investigating some of its properties through simulation.

### Coefficient Alpha

Consider a measurement instrument containing  $p$  items,  $y_1, y_2, \dots, y_p$ , that indirectly measure a single dimension, attribute, or construct. A useful and common computation in the psychological/behavioral sciences is the composite  $Y = y_1 + y_2 + \dots + y_p$ . This composite is placed in statistical models such as ANOVA and regression when conducting research using the corresponding attribute as a variable. Therefore, it is important to know the reliability of the composite and hence the construct being measured.

A popular way to estimate the composite reliability is through coefficient alpha. Coefficient alpha for the population is defined as

$$\alpha_c = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^p \sigma_{ii}}{\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij}} \right). \quad (1)$$

where  $\sum_{i=1}^p \sigma_{ii}$  is the sum of all item variances

and  $\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij}$  is the sum of all item variances

and covariances. For a sample of size  $n$ , population parameters are replaced by sample estimates to obtain a coefficient alpha estimate as

$$\hat{\alpha}_c = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^p \hat{\sigma}_{ii}}{\sum_{i=1}^p \sum_{j=1}^p \hat{\sigma}_{ij}} \right). \quad (2)$$

Note that coefficient alpha is being subscripted with  $c$  to distinguish it from the other forms that

will shortly be introduced. Recall that Zyl, et al. (2000) showed that  $\hat{\alpha}_c$  is the MLE for  $\alpha_c$ .

Coefficient alpha has three interesting properties implied from the classical true score model (Allen & Yen, 1979; Crocker & Algina, 1986). First, when all items have equal true scores that relate equally to the observed scores along with equal measurement error variance, the items are said to be parallel. In this case the covariance matrix for the items has a compound symmetric structure; i.e.,  $\sigma_1 + \sigma^2 I (i = j)$ .

Second, when the measurement error variances are not equal, the items are said to be tau-equivalent. In both of these conditions coefficient alpha is equal to the reliability of a measurement instrument. Lastly, when the true scores do not relate equally to the observed scores and measurement error variances are not equal, the items are congeneric. This last condition is the more general and in this case coefficient alpha underestimates the reliability of a measurement instrument. It is from these three conditions that the conclusion  $\hat{\alpha}_c \leq \rho_{xx'}$  is made, where  $\rho_{xx'}$  is the reliability coefficient of a measurement instrument.

### Bayesian Internal Consistency

The cornerstone of Bayesian methodology is Bayes' theorem. Through Bayes' theorem all unknown parameters are considered random variables. Due to this, prior distributions must be initially defined, which is a way for researchers to express prior beliefs or available information before data are involved in the statistical analysis. Using the observed data  $\mathbf{y}$  and prior distribution  $p(\boldsymbol{\theta})$ , a posterior distribution  $\pi(\boldsymbol{\theta}|\mathbf{y})$  of the parameters  $\boldsymbol{\theta}$  can be constructed. The posterior distribution can be fully expressed through Bayes' theorem as

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{L(\boldsymbol{\theta} | \mathbf{y}) p(\boldsymbol{\theta})}{\int_{\Theta} L(\boldsymbol{\theta} | \mathbf{y}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto L(\boldsymbol{\theta} | \mathbf{y}) p(\boldsymbol{\theta}) \quad (3)$$

where  $L(\boldsymbol{\theta} | \mathbf{y})$  is the data likelihood function and

$$\int_{\Theta} L(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

is a normalizing constant. One can directly see that the posterior is composed of both actual data and prior beliefs or knowledge about the parameter. After the posterior  $\pi(\boldsymbol{\theta}|\mathbf{y})$  is constructed it can be summarized by the mean and SD (or SE) along with other summarizing quantities. For example, the mean and variance can be computed as

$$E(\boldsymbol{\theta} | \mathbf{y}) = \int_{\Theta} \boldsymbol{\theta} \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (4)$$

and

$$\text{var}(\boldsymbol{\theta} | \mathbf{y}) = \int_{\Theta} [\boldsymbol{\theta} - E(\boldsymbol{\theta} | \mathbf{y})]^2 \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad (5)$$

where the  $SD = \sqrt{\text{var}(\boldsymbol{\theta} | \mathbf{y})}$  is also the SE for  $E(\boldsymbol{\theta} | \mathbf{y})$ . At times, the median (or 50<sup>th</sup> percentile) computed as

$$P(\boldsymbol{\theta} | \mathbf{y} \leq m) = P(\boldsymbol{\theta} | \mathbf{y} \geq m) \geq \frac{1}{2} \quad (6)$$

is of interest as it is less influenced by extreme values.

For the Bayesian coefficient alpha (Balpha), first start with the multivariate normal distribution. The posterior of a multivariate normal can be described by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) \propto L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) p(\boldsymbol{\mu} | \boldsymbol{\Sigma}) p(\boldsymbol{\Sigma}) \quad (7)$$

On the far right of (6), note that the prior for the mean is directly dependent on the prior covariance, in addition, this indicates that a different prior is specified for the covariance matrix and mean vector. By choosing the following conjugate priors for both the covariance matrix

$$\boldsymbol{\Sigma} \sim W^{-1}(d_0, \boldsymbol{\Lambda}) \quad (8)$$

and mean vector

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim N\left(\boldsymbol{\mu}_0, \frac{1}{n_0} \boldsymbol{\Sigma}\right), \quad (9)$$

Anderson (1984) and Schafer (1997) showed that the posterior distribution for the covariance matrix and mean vector is

$$\boldsymbol{\Sigma} | \mathbf{y} \sim W^{-1}\left(\begin{matrix} n + d_0, \\ (n-1)\mathbf{S} + \boldsymbol{\Lambda} + \frac{nn_0}{n+n_0}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)' \end{matrix}\right) \quad (10)$$

$$\boldsymbol{\mu} | (\boldsymbol{\Sigma}, \mathbf{y}) \sim N\left(\frac{1}{n+n_0}(n\bar{\mathbf{y}} + n_0\boldsymbol{\mu}_0), \frac{1}{n+n_0} \boldsymbol{\Sigma}\right) \quad (11)$$

where  $W^{-1}()$  denotes an inverted Wishart distribution and  $d_0$ ,  $\boldsymbol{\Lambda}$ ,  $\boldsymbol{\mu}_0$ , and  $n_0$  are hyperparameters chosen by the analyst, and  $\bar{\mathbf{y}}$  and  $\mathbf{S}$  are the mean vector and covariance matrix estimated from the data. Thus, the posterior of the multivariate normal is described by two distributions which jointly are called the normal-inverse Wishart distribution. Note that a prior needs be specified for  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ . If no prior is available a generic noninformative prior such as  $p(\boldsymbol{\theta}) \propto 1$  can be used. In this case the posterior is completely defined by the data. This parameterization fully describes the posterior and it can now be directly computed.

The coefficient alpha posterior can be difficult to obtain directly. However, by simulating  $t = 1, 2, \dots, T$  values from (9) and (10) as  $\boldsymbol{\Sigma}^{(t)} | \mathbf{y}$  and  $\boldsymbol{\mu}^{(t)} | (\boldsymbol{\Sigma}^{(t)}, \mathbf{y})$ , the estimation of the coefficient alpha posterior distribution can be obtained as

$$\alpha_c^{(t)} = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^p \sigma_{ii}^{(t)}}{\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij}^{(t)}} \right) \quad (12)$$

where  $\sigma_{ii}^{(t)}$  and  $\sigma_{ij}^{(t)}$  are elements of  $\boldsymbol{\Sigma}^{(t)} | \mathbf{y}$ . A Bayesian coefficient alpha (Balpha) can then obtained as

$$\alpha_b = E(\alpha_c | \mathbf{y}). \quad (13)$$

An alternative Bayesian coefficient alpha (BalphaM) can be obtained through

$$P(\alpha_c | \mathbf{y} \leq \alpha_{b,m}) = P(\alpha_c | \mathbf{y} \geq \alpha_{b,m}) \geq 1/2.$$

Bayesian credible intervals can then obtained by the lower  $\alpha/2$  and upper  $1-\alpha/2$  percentiles of the sample, where  $\alpha$  is the type I error rate. One can also obtain a normal theory based credible interval as  $\alpha_b \pm Z_{\alpha/2}SD$ . Other summary measures can also be computed as indicated above.

### Methodology

#### Simulation

A  $4 \times 3 \times 6$  Monte Carlo simulation design was utilized to investigate the properties of Bayesian coefficient alphas. First, the number of items was investigated: 5, 10, 15 and 20 and it was found that coefficient alpha increases as a function of the number of items, however, it is constrained to one. Although it is possible for tests and/or surveys to have more than 20 items, going beyond 20 items reaches a point of diminishing returns in terms of investigating coefficient alpha.

Second, the mean item correlation ( $\bar{r}$ ) was investigated: 0.173, 0.223, and 0.314. The mean item correlation is defined as

$$\bar{r} = \frac{2 \sum_{i < j}^p r_{ij}}{p(p-1)} \quad (14)$$

These mean items correlations were investigated because they generate coefficient alphas that range from 0.50 to 0.90, a sufficient range to investigate the properties of the Balpha.

Third, sample size was also explored: 50, 100, 150, 200, 250 and 300. As is the case for the number of items, going beyond a sample

## ESTIMATING INTERNAL CONSISTENCY USING BAYESIAN METHODS

size of 200 reaches a point of diminishing returns in terms of investigating coefficient alpha (Duhachek & Iacobucci, 2004); however, these are sample sizes typically found in psychological/behavioral research. Table 1 presents coefficient alpha as a function of mean item correlation and number of items and shows a reasonable range of coefficient alpha that may be found in psychological/behavioral research.

Table 1: Population Coefficient Alpha for Items by Mean Item Correlations

Items	Mean Item Correlations		
	.1667	.2208	.3103
5	.5001	.5862	.6923
10	.6667	.7392	.8182
15	.7500	.8095	.8709
20	.8000	.8500	.9000

Multivariate normal data were generated with mean vector zero and correlation matrix  $\mathbf{R}$  of dimensions defined by the number of items in the simulation.  $\mathbf{R}$  was chosen to have homogenous off-diagonal elements that generated the corresponding mean item correlation.

For each condition of the simulation study 1,000 replications were obtained. In each replication, Balpha was computed along with the SE and 95% BCIs. Relative bias for Balpha was computed as:

$$\hat{\alpha}_b = \frac{\bar{\hat{\alpha}}_b - \alpha}{\alpha} \quad (15)$$

The average of the estimated SE was computed as

$$\overline{SE} = \frac{\sum_{i=1}^B SE(\hat{\alpha}_{b,i})}{B} \quad (16)$$

where  $B$  is the number of replications. Lastly, two forms of BCI intervals were computed. The

first BCI was obtained by the lower  $\alpha/2$  and upper  $1-\alpha/2$  percentiles of the sample. The second BCI was obtained as  $\hat{\alpha}_b \pm Z_{\alpha/2} SE(\hat{\alpha}_b)$ . The coverage probability of the 95% BCIs were computed as the proportion of times the BCI contains the population parameter  $\alpha_c$ .

Coverage can be judged by forming confidence intervals around the coverage. Coverage should not fall approximately two standard errors (SEs) outside the nominal coverage probabilities ( $p$ ) (Burton, Altman, Royston & Holder, 2006). The standard error is defined as

$$SE(p) = \sqrt{\frac{p(1-p)}{B}} \quad (17)$$

where  $B$  is the number of simulations in the study. For the current study,  $p = .95$  with  $SE(p) = .006892$  and the CI is  $[.936, .964]$ . Thus, coverage that falls outside this CI is considered unacceptable.

For this study, Balpha and corresponding 95% BCIs were estimated from a total of 1,000 simulations from the posterior distribution. In addition, the prior for Balpha was set to be noninformative. A noninformative prior essentially lets the data essentially speak for themselves.

### Results

Relative bias for Balpha and corresponding standard errors are reported in Table 2. First, Balpha and BalphaM always tend to slightly underestimate the population coefficient alpha; however, both Balpha and BalphaM are relatively unbiased under all investigated conditions. Second, Balpha and BalphaM estimates get closer to the population coefficient alpha as sample size increases. Third, Balpha and BalphaM estimates get closer to the population coefficient alpha as the number of items increases. In addition, Balpha and BalphaM estimates get better as the mean item correlation increases. Lastly, BalphaM is consistently closer to the population coefficient than Balpha although the difference is nominal.

In terms of the standard error (SE), a few things should be pointed out. First, the SEs are smaller as the mean item correlation increases. Second, standard errors improve as sample size increases as should be expected. For samples sizes from 100 to 300, the SE difference is nominal when the number of items is between

5 and 10. When the number of items is between 15 and 20, the SE difference is nominal regardless of the sample size. Third, the SEs improve as the mean item correlation increase although the difference can be considered nominal; in most of these conditions, increasing the number of posterior samples should improve the estimation of the SEs.

Table 2: Balpha and BlaphaM Relative Bias with Standard Errors\*

Number of Items	Sample Size	Mean Item Correlation		
		.1667	.2208	.3103
5	50	-.0639, -.0458 (.1189)	-.0592, -.0434 (.1034)	-.0437, -.0328 (.0792)
	100	-.0384, -.0278 (.0835)	-.0304, -.0226 (.0699)	-.0173, -.0124 (.0518)
	150	-.0278, -.0205 (.0673)	-.0191, -.0141 (.0557)	-.0098, -.0066 (.0412)
	200	-.0198, -.0145 (.0577)	-.0128, -.0092 (.0477)	-.0109, -.0085 (.0357)
	250	-.0121, -.0079 (.0510)	-.0104, -.0074 (.0424)	-.0055, -.0037 (.0314)
	300	-.0167, -.0133 (.0467)	-.0103, -.0078 (.0387)	-.0070, -.0012 (.0285)
10	50	-.0537, -.0421 (.0860)	-.0325, -.0243 (.0666)	-.0228, -.0176 (.0468)
	100	-.0250, -.0203 (.0546)	-.0137, -.0103 (.0423)	-.0084, -.0062 (.0295)
	150	-.0100, -.0071 (.0426)	-.0090, -.0068 (.0335)	-.0062, -.0049 (.0235)
	200	-.0145, -.0122 (.0370)	-.0074, -.0059 (.0287)	-.0040, -.0030 (.0199)
	250	-.0073, -.0055 (.0324)	-.0075, -.0063 (.0256)	-.0034, -.0026 (.0177)
	300	-.0075, -.0061 (.0295)	-.0054, -.0044 (.0231)	-.0027, -.0020 (.0161)
15	50	-.0328, -.0245 (.0672)	-.0252, -.0192 (.0516)	-.0152, -.0115 (.0348)
	100	-.0136, -.0105 (.0410)	-.0104, -.0082 (.0314)	-.0052, -.0038 (.0211)
	150	-.0083, -.0064 (.0321)	-.0068, -.0055 (.0246)	-.0042, -.0034 (.0167)
	200	-.0082, -.0068 (.0275)	-.0056, -.0046 (.0209)	-.0036, -.0030 (.0142)
	250	-.0089, -.0078 (.0245)	-.0033, -.0025 (.0184)	-.0023, -.0018 (.0125)
	300	-.0048, -.0039 (.0219)	-.0037, -.0031 (.0167)	-.0017, -.0013 (.0113)
20	50	-.0276, -.0205 (.0580)	-.0213, -.0162 (.0440)	-.0129, -.0097 (.0292)
	100	-.0100, -.0076 (.0335)	-.0090, -.0073 (.0254)	-.0049, -.0037 (.0168)
	150	-.0065, -.0050 (.0260)	-.0041, -.0030 (.0194)	-.0033, -.0026 (.0130)
	200	-.0047, -.0036 (.0219)	-.0039, -.0032 (.0165)	-.0027, -.0022 (.0110)
	250	-.0050, -.0042 (.0194)	-.0028, -.0022 (.0145)	-.0013, -.0009 (.0096)
	300	-.0040, -.0033 (.0175)	-.0024, -.0019 (.0132)	-.0014, -.0011 (.0088)

\*Note: The first number is Balpha followed by BalphaM. Numbers in parentheses are standard errors.



## ESTIMATING INTERNAL CONSISTENCY USING BAYESIAN METHODS

The Bayesian credible intervals are displayed in Table 3 and are more interesting. In general, most of the credible intervals fall within the acceptable range of [.936, .964] based on 1,000 replications. In addition, the percentile based BCIs are consistently closer to 0.95 than the normal theory based BCIs. With 5 items, only two BCIs were not within the acceptable range. When the number of items shifts to 10, seven BCIs were not within the acceptable range, but most of the unacceptable BCIs are normal theory based.

As the number of items increases, more BCIs begin to fall outside the acceptable range, but once again, most of the unacceptable BCIs are normal theory based. However, the unacceptable BCIs occur when the numbers of items are between 15 to 20 and are paired with the smaller sample sizes. Specifically, when the numbers of items are 15, unacceptable BCIs occur at a sample size of 50. Also, when the numbers of items are 20, unacceptable BCIs occur at sample sizes of 50 to 100. In both cases, more normal theory BCIs become unacceptable as the item mean correlation increases. However, the percentile based BCIs tend to remain more stable and closer to 0.95.

### Conclusion

The building blocks of psychological/behavioral research are psychological constructs, which are typically indirectly measured through items on questionnaires. It is crucial to have items that are consistent or reliable in order for research results to be trustworthy and useful. A popular method for estimating a form of reliability is internal consistency via coefficient alpha (Cronbach, 1951; Guttman, 1945). However, coefficient alpha has remained unchanged for over 60 years. Many professional publications are encouraging and/or mandating researchers to supplement their parameter estimates with CIs. Although CIs for coefficient alpha have recently enjoyed fruitful research (Barchard & Hakstian, 1997; Duhachek & Iacobucci, 2004; Feldt, 1965; Kristof, 1963; Maydeu-Olivares, et al., 2007; van Zyl, et al., 2000; Yuan & Bentler, 2002; Yuan, et al., 2003), they are rarely implemented in applied research. In addition, all current coefficient alpha CIs are frequentist based and, as such, they have the traditional, less desirable

CI interpretation and cannot use prior information to stabilize inferences or update information.

This study developed a Bayesian coefficient alpha (Balpha or BalphaM) and its corresponding BCIs. The results from the Monte Carlo investigations indicate that Balpha and BalphaM are relatively unbiased under all investigated conditions of the simulation. However, Balpha and BalphaM have the added advantage of having the BCIs, which have the interpretation researchers really want to make with CIs. Again, BCIs allow one to make the following simpler and more powerful statement: results show 95% confidence that the true population parameter lies between the bands of the credible interval.

In terms of coverage, the percentile based BCIs performed better than the normal theory based BCIs. In particular, the normal theory BCIs begin to perform poorly when the mean item correlation is  $\bar{r} = .3101$ , and the condition worsens as the number of items increases. However, increasing the sample size offsets these conditions. In fact, having a sample size of 250 or more appears to provide protection against this breakdown of the normal theory BCIs. Conversely, the percentile based BCIs remain more consistent, but begin to become unacceptable with the smaller sample sizes and when the number of items is between 15 and 20. However, they remain acceptable as long as the sample size is at least 100. Thus, percentile based BCIs are recommended over the normal theory BCIs.

In general, this suggests that as the number of items increases a larger sample size is required to provide stable inferences. This is not a surprising result. In traditional frequentist statistics, this would be the only option. However, in Bayesian methodology there are two potential additional options to stabilize inferences. First, the number of posterior samples can be increased. This would increase the precision of the estimates. Second, a prior can be specified, which will stabilize inferences that, in turn, will provide better coverage.

It should be noted that the purpose of this study was to demonstrate how a Bayesian internal consistency can be estimated under the basic assumptions made of reliability (Allen &

Table 3: Balpha and BalpahM Bayesian Credible Interval Coverage

Number of Items	Sample Size	Mean Item Correlation		
		.1667	.2208	.3103
5	50	. <b>974</b> , .963	.945, .951	.949, <b>.968</b>
	100	.949, .953	.959, .958	.942, .950
	150	.944, .956	.952, .960	.951, .953
	200	.947, .955	.946, .953	.942, .949
	250	.938, .948	.944, .948	.950, .954
	300	.845, .949	.959, .960	.948, .948
10	50	.961, <b>.971</b>	.952, <b>.968</b>	<b>.966</b> , <b>.980</b>
	100	.949, .961	.961, <b>.965</b>	<b>.968</b> , <b>.976</b>
	150	.951, .951	.960, .961	.954, .959
	200	.937, .942	.953, .958	.959, .963
	250	.951, .958	.945, .951	.963, .965
	300	.949, .953	.947, .954	.947, .949
15	50	<b>.979</b> , <b>.992</b>	<b>.977</b> , <b>.988</b>	<b>.970</b> , <b>.981</b>
	100	<b>.968</b> , <b>.969</b>	<b>.969</b> , <b>.970</b>	.962, <b>.969</b>
	150	.955, .955	.958, .964	.965, <b>.966</b>
	200	.954, .966	.963, .964	.941, .947
	250	.941, .944	.946, .956	.952, .959
	300	.957, .957	.945, .948	.955, .953
20	50	<b>.978</b> , <b>.991</b>	<b>.979</b> , <b>.997</b>	<b>.984</b> , <b>.989</b>
	100	<b>.967</b> , <b>.967</b>	.958, <b>.967</b>	<b>.973</b> , <b>.974</b>
	150	.960, <b>.966</b>	.962, .958	.962, <b>.968</b>
	200	.953, .963	.955, .960	.957, <b>.968</b>
	250	.950, .953	.960, .964	.962, .960
	300	.964, .968	.953, .958	.955, .968

\*Note: The first number is the percentile BCIs followed by the normal theory based BCIs. Unacceptable coverage is bolded; acceptable coverage is within [.936, .964].

Yen, 1979; Crocker & Algina, 1986), thus, study provides a springboard from where future research on Bayesian coefficient alpha can be conducted. However, like any simulation study, this research is limited by the type and number of conditions investigated. In this study, only homogenous items were investigated.

Additionally, items were continuous and normally distributed. Further research is required to investigate the robustness of a Bayesian coefficient alpha to violations of the basic reliability assumptions and to establish its properties under binary or ordinal items.

## ESTIMATING INTERNAL CONSISTENCY USING BAYESIAN METHODS

As noted by Duhachek and Iacobucci (2004) and Maydeu-Olivares, et al. (2007), reporting only a point estimate of coefficient alpha is no longer sufficient. With inferential techniques reporting the SE and CIs provide more information as to the size and stability of the point estimate; in this case the point estimate is coefficient alpha. Within this context, a Bayesian internal consistency estimate may provide an attractive alternative to current coefficient alpha CIs because it provides researchers with BCIs that can be interpreted in a way researchers want and can make use of prior information to stabilize inferences.

### References

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks-Cole Publishing Company.
- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis (2<sup>nd</sup> Ed.)*. New York: Wiley.
- Barchard, K. A., & Hakstian, A. R. (1997). The robustness of confidence intervals for coefficient alpha under violation of the assumption of essential parallelism. *Multivariate Behavioral Research*, 32(2), 169-191.
- Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24), 4279-4292.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104.
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science*, 19(1), 1-15.
- Duhachek, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the standard error of the coefficient alpha index of reliability. *Marketing Science*, 24(2), 294-301.
- Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology*, 89(5), 792-808.
- Feldt, L. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, 30(3), 357-370.
- Gelman, A. (2004). *Bayesian data analysis (2<sup>nd</sup> Ed.)*. Boca Raton, Fla.: Chapman & Hall/CRC.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282.
- Iacobucci, D., & Duhachek, A. (2003). Advancing Alpha: Measuring reliability with confidence. *Journal of Consumer Psychology*, 13(4), 478-487.
- Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, 28(3), 221-238.
- Lee, P. M. (2004). *Bayesian statistics : An introduction (3<sup>rd</sup> Ed.)*. London: Arnold.
- Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, 12(2), 157-176.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall/CRC.
- van Zyl, J., Neudecker, H., & Nel, D. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, 65(3), 271-280.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.
- Yuan, K.-H., & Bentler, P. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, 67(2), 251-259.
- Yuan, K.-H., Guarnaccia, C. A., & Hayslip, B. (2003). A study of the distribution of sample coefficient alpha with The Hopkins Symptom Checklist: Bootstrap versus asymptotics. *Educational and Psychological Measurement*, 63(1), 5-23.

## One is Not Enough: The Need for Multiple Respondents in Survey Research of Organizations

Joseph L. Balloun  
Mercer University,  
Atlanta, GA USA

Hilton Barrett  
Elizabeth City State University,  
Elizabeth City, NC USA

Art Weinstein  
Nova Southeastern University,  
Fort Lauderdale, FL USA

---

The need for multiple respondents per organization in organizational survey research is supported. Leadership teams' ratings of their implementations of market orientation are examined, along with learning orientation, entrepreneurial management, and organizational flexibility. Sixty diverse organizations, including not-for-profit organizations in education and healthcare as well as manufacturing and service businesses, were included. The major finding was the large rating variance within the leadership teams of each organization. The results are enlightening and have definite implications for improved design of survey research on organizations.

Key words: Respondents, survey research, organizations.

---

### Introduction

The archetype 21<sup>st</sup> century organization accumulates knowledge throughout its management and teams (Fisher, 1998). Because the challenges, opportunities and problems facing today's organizations are complex (many interdependent variables), complicated (shades of gray instead of black and white) and require integration of various functions (e.g., marketing, finance, operations), it is difficult for a single leader or even a small group to manage an organization effectively. The concerted and integrated efforts of executives, managers and empowered goal-oriented teams are required for optimal performance (Özaralli, 2003). Many

well-run organizations have recognized the need for management diversity, with a myriad of different orientations within their management, for better decision-making (Roberson & Park, 2004). Peter Drucker stated that a different executive "sees a different reality and is concerned with a different problem" and the executive team "uses conflict of opinion as a...tool to make sure all major aspects of an important matter are looked at carefully" in the course of making a decision (Drucker, 1967, p. 155).

By virtue of society becoming more diverse, organizations are also becoming more diverse in race, gender and ethnicity (Cox, 1991) and in education and other background variables (Pitcher & Smith, 2001). Most often, the diversity cited is demographic and includes race, age and religion among others. Business professionals are also familiar with functional diversity which recognizes that accounting, marketing and operations managers tend to have different orientations and agendas (Hambrick & Mason, 1984).

Psychographic diversity (personality traits and lifestyles) is also important. A good example of this is the learning, decision-making and communication styles demonstrated by the Myers-Briggs Type Indicator (MBTI) (Leonard & Straus, 1997). Kilduff, Angelmar and Mehra (2000) demonstrate that these observable

---

Joseph L. Balloun is a Professor of Education. He has published extensively on social science research methods, statistical techniques, personnel management and developing information systems. Email him at: balloun\_jl@mercer.edu. Hilton Barrett is a Professor of Business with nearly thirty years of experience as an entrepreneur and a corporate executive in a Fortune 500 firm. Email him at: hbarrett@mail.ecsu.edu. Art Weinstein is a Professor of Marketing at Nova Southeastern. Email him at: art@huizenga.nova.edu.

sources of diversity are surrogates for cognitive diversity, something that is more difficult to measure. Rahe (2009) emphasized that the global platform for today's business decision-making makes it even more difficult due to the influences of local environments. The term interpretative ambiguity describes a leadership team whose individuals perceive reality (e.g., performance measures such as market orientation) in different ways because of their cognitive diversity. The resulting heterogeneity may be an impediment to successful marketing strategy implementation (Mengue & Auh, 2005). A development of greater leadership diversity may lead to more innovative decision-making, but a more diverse group of managers can also impede group congruence and unification for attaining strategic objectives.

Industrial (or business to business) marketing research commonly uses a single respondent per organization, also known as a key informant, in survey research on organizations. Researchers should encourage organizations to have only the most highly qualified informants respond to organizational surveys (R. J. Vandenberg, personal communication, June 28, 2010). Researchers target a particular position (e.g., CEO, Purchasing Director) that reflects the purpose of a research study and the need for specific information. Practical constraints on executives' time also suggest that using a single key informant may reduce the organization's cost of responding to surveys.

Traditional research methods typically use a single respondent or key informant to represent the entire organization in multi-organization studies. Using multiple respondents for such research is rare. A review of the *Journal of Marketing*, *Journal of Marketing Research* and *Journal of Marketing Theory and Practice* in recent years found no studies involving more than ten organizations that used multiple respondents per organization.

It is hoped that an especially informed person, the key informant, would be able to judge and report fully the issues affecting an organization. However, organizations choose the key informants who respond to survey research. A researcher does not have direct control of the qualifications of the respondent. For those who

have conducted research at the organizational level, there is a sense that a single respondent, whether a highly qualified key informant or not, may not be sufficiently representative of an organization. Members of an organization's top management team would be qualified informants, yet the degree of concurrence among the top management team is an empirical question.

Numerous authors in the management and marketing literature have called for using multiple, as opposed to single, respondents per organization (Dawes, 2000; Gray, Matear, Boshoff & Matheson, 1998; Tsai, 2002). Multiple respondents per organization may allow for an average measurement of the leadership team's response, but even more importantly, insights into the team's variation on specific topics. Prior organizational survey research has not identified what effects may be masked or distorted by using only one respondent per organization. For example, how do several leaders in different functions within a single organization perceive a specific product's capabilities or an organizational issue?

The purpose of this study is to describe effects that may be discovered when multiple respondents per organization are used in survey research on organizations. More specifically, the focus is on how perceptions of selected strategic management constructs vary within and among organizations. The study shows what information may be gained by having more than one informant per organization.

## Methodology

### Organizations and Participants

This study employed a snowball sampling technique, which consisted of soliciting the members of several organizations, contacting members of personal networks and targeting particular firms to build sectors and industries. The resulting non-probabilistic, convenience sample consisted of 696 usable individual responses within sixty organizations. Of these sixty, 37 organizations were in the business sector and 23 were not-for-profits. An effort was made to represent a variety of industries: banking (11), education (13), healthcare (10), manufacturing (10), real estate (6), retail (3) and all other services (7). Eighteen

organizations employed 500 or more employees, and forty-two employed fewer than 500. For each participating organization, a request was made for twenty of their top management team members to complete and return the survey. The participating managers were volunteers from their organizations.

#### Measurements

The measures of interest in such survey projects often are perceptually based. This research project specifically used measures of organizational market orientation, entrepreneurial management, organizational flexibility and learning orientation. These constructs have been major research topics for over a decade. They have been variously conceptualized with other variables and organizational performance as researchers attempt to develop better prescriptive models for executives (Frank, Kessler & Fink, 2010; Mokhtar & Yusoff, 2009).

Market orientation (MKT), as described by Jaworski and Kohli (1993) has three components: generation of market intelligence, sharing of this knowledge throughout the firm and a marketing response mechanism. Narver and Slater's (1990) work defined MKT as having three tenets: customer orientation, competitive orientation and inter-functional coordination.

Learning orientation (LRN), as popularized by Senge (1990), denotes that not only do individuals have and use the ability to do both adaptive (incremental) and generative (paradigm shift) learning, but also to keep an open mind to different perspectives and have a commitment to learning (Baker & Sinkula, 1999). When correctly practiced, the norm becomes collaborative learning. In their studies of company rejuvenation, Stopford and Baden-Fuller (1990) established that the development of a learning organization required flexibility and internal communication to achieve an effective market orientation. Slater and Narver stated that "a market orientation is inherently a learning orientation" (1995, p. 67).

Entrepreneurial management style and corporate entrepreneurship (ENT) are terms used to define an organization that acts entrepreneurially (Covin & Miles, 1999). ENT is an organizational process that encourages and

practices innovation, risk-taking and a proactive orientation toward customers, competition and opportunities (Miller & Friesen, 1982); thus, there is a relationship between the dimensions of ENT and the marketing activities of the organization. Hence, the organization: (a) is proactive in obtaining intelligence on customers and competitors, (b) is innovative by reconfiguring its resources to formulate a strategic response, and (c) implements the response, which, because it is different, entails some degree of risk and uncertainty.

Organizational flexibility (ORG) is defined as the degree in which an organization is adaptable in administrative relations and the authority vested in situational expertise. Khandwalla (1977) used the term organic to define such attributes. The management theorist Mary Parker Follett, in the 1920s, emphasized the need to match an organic structure to what is now considered an entrepreneurial management style (Graham, 1995).

Each of these organizational characteristics (MKT, LRN, ENT and ORG) has been found to be positively related to organizational performance (Zahra & Covin, 1995; Baker & Sinkula, 1999; Ellinger, Ellinger, Yang & Howton, 2002; Barrett, Balloun & Weinstein, 2004). However, these four characteristics and their relationship with organizational performance have not been analyzed in a single model. This study incorporated these organizational characteristics in a single model. Furthermore, in studying these four critical success variables and their relationships to organizational performance, the study addressed two noteworthy gaps in the literature:

1. Incorporating a multiple response methodology to assess the varying leadership team members' perspectives of how organizations are perceived on each of these four variables and organizational performance; and
2. Broadening the research base from the for-profit manufacturing sector to also include service industries and the non-profit sector.

## ONE IS NOT ENOUGH

This broader perspective recognizes the 21<sup>st</sup> century leadership team's diversity and the economic realities of our society.

Market orientation (MKT) was measured using the twenty-question construct developed by Kohli, Jaworski & Kumar (1993). Learning orientation (LRN) was measured using Yim-Teo's (2002) ten-question scale. Entrepreneurial management style (ENT) was measured using Covin & Slevin's (1989) nine-question construct for innovativeness, proactive approach to customers and competition, and risk-taking. Organizational flexibility (ORG) was measured using a seven-question Khandwalla (1977) instrument. For consistency, a seven point Likert scale was used for all questions. The resulting construct measures were the averages of the item ratings in each scale.

Given the difficulties in obtaining correct financial information that is of similar nature and time period among respondents, as well as the outright refusal by many organizations to release such information, a subjective measure of organizational performance is often more practical and useful than apparently objective financial information when the latter is available (Naman & Slevin, 1993; Sanberg & Hofer, 1987), and because financial measures would not be comparable or necessarily applicable across the diverse organizations included in a study. Due to these difficulties, a qualitative-based, two-question rating instrument developed by Jaworski & Kohli (1993) was used. This scale (PERF) assesses (a) how well the organization did this year versus last year, and (b) how well it did versus leading competitors or similar organizations (for businesses and non-profits, respectively). These two judgmental questions result in a subjective rating of financial performance.

### Results

#### Data Screening

An average of twelve managers per organization participated. The harmonic mean was 9.38 respondents per organization, and the range was from four to 31 respondents per organization. The data were screened for normality, outliers and non-response bias.

Twenty questionnaires out of 716 received were discarded due to excessive missing responses. The possibility of non-response bias was tested by a within organization chronological quartile comparison of returned questionnaires: Armstrong & Overton (1977) stated that late respondents (versus early respondents) are considered more similar to non-respondents. A set of ANOVA tests were conducted among quartile means on selected variables; these tests revealed no significant differences among earlier and later respondents.

#### Scale Reliability and Correlations among Scales

All of the Cronbach (1951) alphas exceed Nunnally's (1978) minimum criterion of 0.70 for reliability and all are significantly greater than zero at beyond the 0.001 level (Feldt, Woodruff & Salih, 1987). Table 1 contains the coefficient alpha reliabilities and the correlations among the scales within and among organizations. Based on related work (Barrett, et al., 2004) one-tailed tests for positive correlation were appropriate. The reliabilities and correlations all are significant at or well beyond the 0.05 level. This was expected as both theory and practice support the needed integration and interdependency among these constructs. These results support the use of all the scales and their constituent items in subsequent analyses.

Table 1: Pearson Correlations among Scales<sup>a</sup>

Rating Scale	MKT	LRN	ENT	ORG	PERF
MKT	0.92	0.78	0.65	0.47	0.62
LRN	0.45	0.91	0.64	0.54	0.54
ENT	0.47	0.29	0.90	0.37	0.46
ORG	0.26	0.20	0.42	0.82	0.21
PERF	0.49	0.33	0.33	0.30	0.82

Note: <sup>a</sup>Correlations above the diagonal are among the scale means of organizations

Sample size for correlations of scale means was considered to be 60. Correlations below the diagonal are within organizations. The

coefficient alpha reliabilities are shown on the diagonal. Sample sizes within organizations or on the diagonal were 696. All of the correlations and the coefficient alpha reliabilities were significant at or beyond the 0.01 level, with the exception of the correlation between PERF and ORG for the organization mean scores, which was significant at the 0.05 level. With two exceptions, the correlations among institutional means were greater than the correlations within organizations.

#### Partitioning Sources of Variance in Scale Items

An important first question is how scale item responses are related to differences among organizations, scales and respondents. Here the central question is on what percent of variance in item responses is attributable to respondent related effects versus organizational differences. For this purpose, the responses to the 48 items constituted the dependent variable. The independent variables for this analysis include: organizations, the five scales and respondents within organizations.

The sample of organizations is best considered a random effect in the analysis of variance (ANOVA) sense. If the study were to be repeated, an entirely new collection of similarly diversely selected organizations would be generated. Respondents (participating managers) are necessarily nested within their own organizations. The chosen rating scales are fixed in the sense that only in the results that apply to these specific rating scales are of interest. Items are nested within their respective scales. Each item within a scale provides a replication of what that scale measures for each respondent. The results reported in Table 1 support the use of scale items also as replicates across all the scales in the study. The overall analysis can be conducted using an ANOVA model including organizations, participants within those organizations, the rating scales and the scale items as replicates. Table 2 shows the means for each scale within the 60 participating organizations.

#### Analysis of Item Data with Multiple Respondents per Organization

The item data underlying the summary of Table 2 were analyzed by ANOVA. Table 3

displays the expected means squares, the observed mean squares and the significance test for each possible effect. The method of moments was used to estimate the variance components for each of the estimable effects. The percent of variance due to each effect in the intra-class correlation or omega squared sense also is shown in Table 3.

About 7% of the variance in item ratings is accounted for by organizational differences or the organization by scale interaction. Approximately 30% of the variance is due to differences among respondents within organizations or the respondent by scale interaction effect. The within organization variance due to respondent effects is likely underestimated because of the nature of cluster sampling of the organizations (R. J. Vandenberg, personal communication, June 28, 2010).

#### Analysis of Item Data with One Respondent per Organization

To illustrate what will happen when there is only one respondent per organization, the first respondent in each organization (by identification number in the file) was selected. The identification numbers were assigned by sequence of return of the surveys over the entire study. The identification numbers were assigned for convenience of coding, and do not have an a priori systematic relationship to the unknown expertise of the respondent. Data screening analyses supported the conclusion that response order was unrelated to scale means. The ANOVA with one respondent per organization was computed for these sixty respondents. The ANOVA shown in Table 4 assumed organizations as a random effect and scales as a fixed effect. Items within scales were replicates. Table 4 displays the expected means squares, the observed mean squares and the significance test for each possible effect. The method of moments was used to estimate the variance components for each of the estimable effects. The percent of variance accounted for by each effect also is shown in Table 4.

According to Table 4, 29% of the variance in item responses is attributable to differences among organizations or organization by scale interaction. The variance due to respondents or the respondent by scale



# ONE IS NOT ENOUGH

Table 2: Mean Scale Ratings by Organizations

Organization Number	Number of Respondents	Rating Scale					
		MKT	LRN	ENT	ORG	PERF	Row Mean <sup>a</sup>
1	15	3.37	3.33	3.25	3.86	4.83	3.73
2	4	5.99	5.13	3.22	5.68	6.38	5.28
3	10	4.72	4.47	3.75	4.46	4.85	4.45
4	12	4.22	4.17	3.63	3.52	4.46	4.00
5	14	4.48	4.59	4.45	4.31	5.75	4.72
6	16	5.22	5.17	4.84	4.89	5.84	5.19
7	6	4.18	4.63	4.30	4.60	4.42	4.42
8	12	5.43	4.75	4.00	4.72	6.00	4.98
9	11	5.20	4.97	5.15	4.17	4.82	4.86
10	10	4.63	4.21	3.98	4.07	5.50	4.48
11	17	5.11	4.75	4.48	3.70	5.79	4.76
12	11	4.78	4.77	4.26	3.88	5.18	4.57
13	18	5.46	5.07	5.56	4.84	5.47	5.28
14	14	5.81	4.76	4.32	3.69	6.75	5.07
15	31	4.57	4.24	3.80	4.32	4.68	4.32
16	9	4.03	4.20	4.14	4.05	4.11	4.11
17	14	3.66	3.54	2.77	3.07	4.75	3.56
18	14	4.25	4.64	2.74	3.82	4.54	4.00
19	5	4.80	4.76	4.47	3.42	6.00	4.69
20	16	4.78	4.17	4.13	4.28	5.09	4.49
21	13	4.60	4.52	4.22	4.12	4.31	4.35
22	11	3.91	3.20	3.16	3.74	3.59	3.52
23	9	4.48	4.63	3.77	4.30	4.50	4.34
24	4	4.16	4.25	3.47	3.80	3.50	3.84
25	13	5.37	4.62	4.43	4.63	4.73	4.75
26	17	3.52	3.89	2.82	3.44	5.56	3.85
27	21	4.57	4.23	3.56	3.90	4.10	4.07
28	12	4.16	3.98	3.37	4.02	4.88	4.08
29	22	5.23	4.60	4.65	2.90	5.55	4.58
30	8	5.35	5.29	5.13	5.41	5.44	5.32
31	14	3.87	3.77	3.25	4.01	3.93	3.77
32	9	3.41	3.49	3.70	3.46	4.61	3.73
33	10	3.78	4.28	4.13	3.91	4.00	4.02
34	5	3.87	4.42	4.16	3.49	4.60	4.11
35	4	5.01	4.78	5.83	4.08	6.13	5.16
36	15	4.53	4.07	4.38	4.15	5.10	4.45
37	9	4.97	4.20	4.74	3.33	5.67	4.58
38	13	3.74	3.58	3.97	3.22	3.77	3.66
39	12	3.35	3.62	3.26	3.81	4.83	3.77
40	7	5.28	4.89	5.13	3.89	6.21	5.08
41	15	4.87	4.92	4.10	4.65	6.17	4.94
42	9	5.39	4.26	4.27	4.34	5.78	4.81
43	15	4.19	4.34	3.19	3.82	4.90	4.09
44	15	5.50	4.75	4.79	4.66	5.50	5.04
45	7	5.04	4.53	3.35	3.21	4.71	4.17

Table 2: Mean Scale Ratings by Organizations (continued)

Organization Number	Number of Respondents	Rating Scale					
		MKT	LRN	ENT	ORG	PERF	Row Mean <sup>a</sup>
46	6	3.19	3.67	3.04	3.74	5.25	3.78
47	9	3.52	3.68	2.95	2.92	4.22	3.46
48	15	4.77	4.17	3.72	3.97	4.23	4.17
49	16	5.19	5.10	3.94	4.39	5.22	4.77
50	5	5.29	4.82	4.27	3.71	5.80	4.78
51	4	4.40	4.22	4.17	5.18	4.50	4.50
52	18	4.57	4.48	4.73	3.98	5.11	4.57
53	13	5.13	4.8	4.21	4.22	5.42	4.76
54	9	5.29	4.57	4.62	4.35	5.06	4.77
55	6	4.09	4.53	4.46	3.90	4.92	4.38
56	5	4.62	4.28	3.42	4.34	4.30	4.19
57	10	5.02	3.97	4.07	3.27	5.90	4.45
58	8	4.34	4.33	3.86	3.89	5.44	4.37
59	13	4.99	4.86	5.03	4.51	5.73	5.02
60	11	4.62	5.31	4.12	4.13	5.18	4.67
Column Means <sup>a</sup>	11.6	4.60	4.40	4.04	4.04	5.06	4.43

Note: <sup>a</sup>Unweighted means

Table 3: ANOVA with Multiple Respondents per Organization

Source	df <sup>a</sup>	EMS <sup>b</sup>	OMS <sup>c</sup>	F <sup>d</sup>	Variance Percentage <sup>e</sup>
Organization (A)	59	S+cdsB+bcdsA	74.42	3.37 <sup>***</sup>	3.65
Respondent (B(A))	635	S+cdsB	22.09	14.36 <sup>***</sup>	16.63
Scale (C )	4	S+dsBC+bdsAC+abdsC	520.28	35.14 <sup>***</sup>	2.94
AC	236	S+dsBC+bdsAC	14.81	3.02 <sup>***</sup>	3.45
BC(A)	2540	S+dsBC	4.91	3.19 <sup>***</sup>	13.62
S	29751	S	1.54		59.72
Total (Adjusted)	33225		2.63		

<sup>\*\*\*</sup> p < 0.001; Notes: <sup>a</sup>Degrees of freedom for the effect; <sup>b</sup>Expected mean squares; <sup>c</sup>Observed mean squares; <sup>d</sup>F ratio; <sup>e</sup>Percent of variance accounted for by each effect by the sample moment method

## ONE IS NOT ENOUGH

Table 4: Example ANOVA with One Respondent per Organization

Source	df <sup>a</sup>	EMS <sup>b</sup>	OMS <sup>c</sup>	F <sup>d</sup>	Variance Percentage <sup>e</sup>
Organization (A)	59	S+bsA	12.80	8.17***	10.28
Scale (B)	4	S+sAB+asB	37.73	6.73***	2.45
AB	236	S+sAB	5.61	3.58***	18.49
S(AB)	2580	S	1.57		68.78
Total (Adjusted)	2879		2.37		

\*\*\*  $p < 0.001$ ; Notes: <sup>a</sup>Degrees of freedom for the effect; <sup>b</sup>Expected mean squares; <sup>c</sup>Observed mean squares; <sup>d</sup>F ratio; <sup>e</sup>Percentages of variance accounted for by each effect were computed by the sample moment method

interaction is not estimable when there is one respondent per organization. By definition, with only one respondent per organization, there is no way to separate respondent effects from organization effects.

### Analysis of Item Data by the Mean of Respondents within Organizations

In lieu of other evidence, it would be expected that the average of a management team's judgments would be more accurate than those of a single respondent. That is, it is important to strive to simulate the judgments of a fully qualified or key informant. Therefore, the ANOVA with the average of the item responses within each organization was repeated. The ANOVA shown in Table 5 again assumed organizations as a random effect and scales as a fixed effect, and items within scales were replicates. Table 5 displays the expected means squares, the observed mean squares, the significance test and the percent of variance accounted for by each effect.

According to Table 5, 11% of the variance in item responses is attributable to differences among organizations or organization by scale interaction. Again, by definition, with only one respondent per organization, respondent effects cannot be separated from organization effects.

### Partitioning Sources of Variance in Perceived Organizational Performance

An applied researcher should be concerned about how much difference the research design would make in modeling a dependent variable. That is, will using a single respondent or multiple respondents make a difference in the percent of variance attributed to different effects? For this purpose, the performance rating (PERF) scale was the dependent variable. The independent variables for this analysis included industry groups, organization size, organizations within industry group and size classifications and respondents within organizations.

Seven categories of industries were included in this study. A concerted effort was made to have a variety of business and nonprofit sectors represented: banking (11), education (13), healthcare (10), manufacturing (10), real estate (6), retail (3) and all other services (7). Regarding the sizes of firms included in the study: 18 employed 500 or more employees and 42 employed less than 500. Almost all of the participating organizations were from five southeastern states: North Carolina, Tennessee, South Carolina, Georgia and Florida.

Table 6 shows the results when the ANOVA includes industry, size, organization and respondent effects. By this analysis method, the statistically significant effects are differences among organizations within industry by size subgroups and respondent differences within

organizations. The observed mean squares were decomposed according to their expected mean squares. After such decomposition, organization differences within subgroups, respondent differences within organizations, and unreliability of the dependent variable accounted respectively for about 3.5%, 59% and 37% of the total variance. The industry, size and industry by size effects accounted for no or very small components of total variance in the intra-class correlation or omega squared sense.

The same data were re-analyzed while ignoring the possible effects of organizations, respondents and unreliability of the dependent variable because this is the more common

analysis method used in practice (results are shown in Table 7). The industry and industry by size effects are now statistically significant. By this analysis, the industry and industry by size effects are statistically significant, and they account for about 3% of the variance each, however, about 93% of the variance is attributable to the residual variance. Using the common analysis shown in Table 7, a researcher would not discover the most important sources of variance in this study; namely the organization differences, respondent differences within organizations, and variance due to unreliability of the dependent variable.

Table 5: Example ANOVA with Organization Mean Responses

Source	df <sup>a</sup>	EMS <sup>b</sup>	OMS <sup>c</sup>	F <sup>d</sup>	Variance Percentage <sup>e</sup>
Organization (A)	59	S+bsA	6.66	18.05***	5.99
Scale (B)	4	S+sAB+asB	55.71	38.11***	4.30
AB	236	S+sAB	1.46	3.96***	5.19
S(AB)	2580	S	.37		84.52
Total (Adjusted)	2879		2.37		

\*\*\* p < 0.001; Notes: <sup>a</sup>Degrees of freedom for the effect; <sup>b</sup>Expected mean squares; <sup>c</sup>Observed mean squares; <sup>d</sup>F ratio; <sup>e</sup>Percent of variance accounted for by each effect by the sample moment method

Table 6: ANOVA with Organizations and Respondents Hierarchically Nested

Source	df <sup>a</sup>	EMS <sup>b</sup>	OMS <sup>c</sup>	F <sup>d</sup>	Variance Percentage <sup>e</sup>
Industry	6	S+sD+dsC+bcdaA	13.87	1.24	0.00
Size	1	S+sD+dsC+acdsB	0.00	0.00	0.00
Industry by Size	6	S+sD+dsC+cdaAB	5.64	0.51	0.00
Organizations within Industry by Size	46	S+sD+dsC	11.15	4.99***	3.50
Respondents within Organizations	636	S+sD	2.24	4.23***	59.44
Residual	687	S	0.53		37.06

\*\*\* p < 0.001; Notes: <sup>a</sup>Degrees of freedom for the effect; <sup>b</sup>Expected mean squares; <sup>c</sup>Observed mean squares; <sup>d</sup>F ratio; <sup>e</sup>Percent of variance accounted for by each effect by the sample moment method

# ONE IS NOT ENOUGH

Table 7: ANOVA with Organization Mean Responses Only

Source	df <sup>a</sup>	EMS <sup>b</sup>	OMS <sup>c</sup>	F <sup>d</sup>	Variance Percentage <sup>e</sup>
Industry	6	S+bsA	13.87	8.26 <sup>***</sup>	3.43
Size	1	S+asB	0.00	0.00	0.00
Industry by Size	6	S+sAB	5.64	3.36 <sup>**</sup>	3.18
Residual	1369	S	1.68		93.39

\*\* p < 0.01; \*\*\* p < 0.001; Notes: <sup>a</sup>Degrees of freedom for the effect; <sup>b</sup>Expected mean squares; <sup>c</sup>Observed mean squares; <sup>d</sup>F ratio; <sup>e</sup>Percent of variance accounted for by each effect by the sample moment method

## Conclusion

### Item Mean Differences

When multiple respondents are included in research on organizations the differences among them can be detected and evaluated for their magnitude. The effects on item means due to the respondents or due to their interactions with the rating scales accounted for several times as much variance as effects associated with organizational differences or the interaction of organizations and scales. By contrast, when only one respondent was used per organization, it appeared that 28% of the variance was due to effects related to organizational differences or the interaction between scales and organizations. However, the apparent organizational differences and organization by scale differences estimated in Table 4 are confounded with respondent-related effects.

### Perceived Performance Effects

Analogous results were obtained when ratings of organizational performance were modeled from industry classification and organization size. Again quite large effects are due to differences among respondents within organizations. Ignoring such possible individual differences among respondents makes little sense. These results support the conclusion that using only one respondent per organization in survey studies on organizational differences often will not detect nor appropriately estimate the size of effects of interest. It is time to move beyond survey studies using only one respondent per organization.

## Limitations

Only a small set of rating scales was used in this study, thus it is not certain that similar effects would emerge in other specific applications. However, as demonstrated, the variance among the multiple respondents' scale ratings within the organizations in this study was greater than the variance among the organizations' scale ratings for all five rating scales. In addition, the most important sources of variation in the dependent variable cannot be detected, and were not detected by one of the most common analytical methods used in such studies.

Estimating the magnitude of variance due to various effects in ANOVA often is ambiguous. Here a decomposition method was used that assumed a fully balanced design; yet that is not true in these data, and is it not likely to ever be true in real world surveys on organizations. As noted earlier, methods employed in this study probably understated the variance due to respondents within organizations. That likely downward bias in estimates of the within-organization variance re-emphasizes the point that individual differences within top management teams should be overtly assessed.

### Individual Differences among Top Executives

Organizational leaders and scholars should be concerned because almost all of the practical and academic research utilizes a single respondent, a key informant, in survey research involving many organizations. This variation

among the respondents seems less an indicator of different perspectives on the right way to solve a problem and more the question of a common recognition of reality. As has often been asked: Is everyone reading from the same page or even from the same book? The fable of six blind men describing the elephant also comes to mind. There are also anecdotal examples of senior corporate management who share neither the strategic plan (it's confidential) nor the ongoing operating results with their team members (the world is on a need-to-know basis). There are many organizations that do not have a proper information system in place to provide their managers the information needed for innovative decision-making. Regardless, the present method of using a single respondent in organizational research carries a large risk of providing misleading findings for decision-makers and researchers.

The diversity within leadership teams should be used to leverage individuals' perspectives to better understand what problems and opportunities exist and the possibilities to solve the former and make the most of the latter. Organizations need to leverage these same perspectives as a competitive advantage to conceptualize possible strategic alternatives and possible implementation tactics. While management may disagree on the proper objective and strategy, the leadership team should have some consistency and consensus as to the reality of actions taken and the results.

One immediate implication is that information is not being shared; this new reality in survey research data collection methodologies and in management practices needs to be recognized and corrected. The result will be better research studies and enhanced organizational decisions.

#### Recommendations

First, as researchers of organizations, multiple respondents must be incorporated into survey methodology. This will increase the difficulties and costs of obtaining participating organizations and it also begs the question of how many respondents within an organization. More may be better, but there is a trade-off between difficulties/costs of obtaining participants and feasible results of research

projects. Based on this study, requesting multiple respondents within each organization is a reasonable request. The appropriate number of respondents per organizations cannot be estimated with confidence over all possible applications, but at a minimum, the number of respondents requested should be sufficient to ensure detection of individual differences within executive teams and to detect salient differences among organizations.

Second, it must be recognized that our own organizations, business or non-profit, can display the same vulnerabilities as those surveyed. As previously mentioned, the 21<sup>st</sup> century organization needs to share the strategic analysis and plan within the leadership team. The organization requires an information system that provides team members access to pertinent information needed in understanding the realities of the internal and external environments. There is also a need to recognize that information is unique among the factors of production: It gains value through additional perspectives as it is shared for a common good or purpose.

No longer will the traditional business measures (revenues, profits, and market share) suffice. The challenge is selecting the right metrics to accurately capture business performance. Intuit, the manufacturer of Turbo Tax software, Enterprise Rent-a-Car, and GE now focus on a single item to gauge satisfaction based on customer's likelihood to recommend the product (Darlin, 2005). More typically, leading organizations now use marketing dashboards to understand their critical evaluative points (Clark, Abela & Ambler, 2006). A dashboard of the most vital metrics aids executives in managing their businesses. Farris, Bendle, Pfeifer & Reibstein stated "...no one metric is going to give a full picture. It is only when you can use multiple viewpoints that you are likely to obtain anything approaching a full picture" (2006, p. 334).

Zeithaml, et al. (2006), explain that even when batteries of items are used, the dashboard approach may yield inaccurate results because it largely reports past (rearview mirrors) or present (dashboard) data. They propose the need to develop headlight or forward-looking customer metrics such as customer lifetime value and customer equity to increase customer value.

Rearview mirrors, dashboards, and headlights may be viewed as the latest version of management information systems. Proper design, buy-in, use, and updating information systems allow the leadership team to have a shared reality. We believe that multiple respondent research can remedy some of the barriers to understanding marketing performance and answer the basic question once popularized by Ed Koch, former mayor of New York City "...how are we doing?"

#### Further Research

Additional research should go much further with the question of variation among respondents within an organization. Business and academic researchers need to be aware of this phenomenon. Expanded research using multiple respondents is required, but there is also a need to take the process several steps further. For example, future research using multiple respondents within an organization might segment the respondents by organizational variables such as function (e.g., marketing, finance), level (e.g., manager, director, vice-president), and years with the organization. In addition, research could also include demographic variables (e.g., race, gender, age) and psychographic variables (e.g., individualism vs. collectivism).

The sample size of organizations needs to be larger for in-depth statistical analyses to better seek and understand the nature of the respondent effects. The combinations of respondent and organizational variables are complex. The needed research to resolve these questions will be challenging. Hopefully, this innovative research on multiple respondents will aid in better understanding the subtle causes of business variation and inspire other researchers to pursue this stream of study. Hence, the performance objective is to accurately gauge the collective wisdom of management teams rather than relying solely on a single informed (or potentially uninformed) individual per organization.

#### References

- Armstrong, J. S., & Overton, T. S. (1977). Estimating non-response bias in mail surveys. *Journal of Marketing Research*, 14, 396-401.
- Baker, W. E., & Sinkula, J. M. (1999). The synergistic effect of market orientation and learning orientation on organizational performance. *Journal of the Academy of Marketing Science*, 27(4), 411-427.
- Barrett, H., Balloun, J. L., & Weinstein, A. (2004). Success factors for 21<sup>st</sup> century organizations: A multi-sector, multiple respondents approach. *International Journal of Business Disciplines*, 15(2), 79-88.
- Clark, B. H., Abela, A. V., & Ambler, T. (2006). Behind the wheel. *Marketing Management*, 30, 19-23.
- Covin, J. G., & Miles, M. P. (1999). Corporate entrepreneurship and the pursuit of competitive advantage. *Entrepreneurship Theory and Practice*, 23(3), 47-63.
- Covin, J. G., & Slevin, D. P. (1989). Strategic management of small firms in hostile and benign environments. *Strategic Management Journal*, 10, 75-87.
- Cox, T. (1991). The multicultural organization. *Academy of Management Executive*, 5(2), 34-47.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Darlin, D. (2005). The only question that matters. *Business 2.0*, September, 50-52.
- Dawes, J. (2000). Market orientation and company profitability: Further evidence incorporating longitudinal data. *Australian Journal of Management*, 25(2), 173-199.
- Drucker, P. F. (1967). *The effective executive*. New York: Harper Colophon.
- Ellinger, A. D, Ellinger, A. E., Yang, B., & Howton, S. W. (2002). The relationship between the learning organization concept and firms' financial performance: An empirical assessment. *Human Resource Development Quarterly*, 13(1), 3-21.
- Farris, P. W., Bendle, N. T., Pfeifer, P.E., & Reibstein, D. J. (2006). *Marketing metrics: 50+ metrics every executive should master*. Upper Saddle River, NJ: Wharton School Publishing.
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93-103.
- Frank, H., Kessler, A., & Fink, M. (2010). Entrepreneurial orientation and business performance – a replication study. *Schmalenbach Business Review*, 62, 175-198.

- Graham, P. (1995). *Mary Parker Follett-Prophet of Management*. Boston: Harvard Business School Press.
- Gray, B., Matear, S., Boshoff, C., & Matheson, P. (1998). *Developing a better measure of market orientation*. *European Journal of Marketing*, 32(9/10), 884-903.
- Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9, 193-206.
- Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: Antecedents and consequences. *Journal of Marketing*, 57, 53-70.
- Khandwalla, P. N. (1977). *The design of organizations*. New York: Harcourt Brace Jovanovich.
- Kilduff, M., Angelmar, R., & Mehra, A. (2000). Top management – team diversity & firm performance: Examining the role of cognitions. *Organization Science*, 11(1), 21-34.
- Kohli, A. K., Jaworski, B. J., & Kumar, A. (1993). MARKOR: A measure of market orientation. *Journal of Marketing Research*, 30, 467-477.
- Leonard, D., & Straus, S. (1997). Putting your company's whole brain to work. *Harvard Business Review*, 75(4), 111-121.
- Mengue, B., & Auh, S. (2005). A test of strategic orientation formation versus strategic orientation implementation: The influence of TMT functional diversity and inter-functional coordination. *Journal of Marketing Theory and Practice*, 13(2), 4-19.
- Miller, D., & Friesen, P. H. (1982). Innovation in conservative and entrepreneurial firms: Two models of strategic momentum. *Strategic Management Journal*, 3, 1-25.
- Mokhtar, S. S. M., & Yusoff, R. Z. (2009). Exploring the relationship of key strategic orientations towards sustainable organizational performance. *International Journal of Business and Management Science*, 2(1), 79-87.
- Naman, J. L., & Slevin, D. P. (1993). Entrepreneurship and the concept of fit: A model and empirical tests. *Strategic Management Journal*, 14, 137-153.
- Narver, J. C., & Slater, S.F. (1990). The effect of a market orientation on business profitability. *Journal of Marketing*, 54(4), 20-35.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Özaralli, N. (2003). Effects of transformational leadership on empowerment and team effectiveness. *Leadership & Organization Development Journal*, 24(5/6), 335-344.
- Pitcher, P., & Smith, A. D. (2001). Top leadership team heterogeneity: Personality, power, and proxies. *Organization Science*, 12(1), 1-18.
- Rahe, M. (2009). Subjectivity and cognition in knowledge management. *Journal of Knowledge Management*, 13(3), 102-117.
- Roberson, Q. M., & Park, H. J. (2004). Diversity reputation and leadership diversity as sources of competitive advantage for organizations. *Academy of Management Proceedings*, F1.
- Sanberg, W. R., & Hofer, C. W. (1987). Improving new venture performance: The role of strategy, industry structure, and the entrepreneur. *Journal of Business Venturing*, 2, 5-28.
- Senge, P. (1990). *The fifth discipline*. New York: Doubleday.
- Slater, S. F., & Narver, J. C. (1995). Market orientation and the learning organization. *Journal of Marketing*, 59, 63-74.
- Stopford, J. M., & Baden-Fuller, C. W. F. (1994). Creating corporate entrepreneurship. *Strategic Management Journal*, 15, 521-536.
- Tsai, W. (2002). Social structure of 'coopetition' within a multiunit organization: Coordination, competition, and intra-organizational knowledge sharing. *Organization Science*, 13(2), 179-190.
- Yim-Teo, T. H. (2002). Learning organization, a cultural breakthrough in the public sector. *The Journal of Applied Management & Entrepreneurship*, 7(3), 48-71.
- Zahra, S. A., & Covin, J. G. (1995). Contextual influences on the corporate entrepreneurship-performance relationship: A longitudinal analysis. *Journal of Business Venturing*, 10, 43-58.
- Zeithaml, V. A., Bolton, R. N., Deighton, J., Keiningham, T. L., Lemon, K. L., & Petersen, J. A. (2006). Forward-looking focus: Can firms have adaptive foresight? *Journal of Service Research*, 9(2), 168-183.



## A Simulation Study of the Relative Efficiency of the Minimized Integrated Square Error Estimator (L2E) For Phase I Control Charting

John N. Dyer  
Georgia Southern University,  
Statesboro, GA, USA

---

Parameter estimates used in control charting, the sample mean and variance, are based on maximum likelihood estimation (MLE). Unfortunately, MLEs are not robust to contaminated data and can lead to improper conclusions regarding parameter values. This article proposes a more robust estimation technique; the minimized integrated square error estimator (L2E).

Key words: Phase I control charting, SPC, L2E, MLE, parameter estimation.

---

### Introduction

Process monitoring using control charts is the most common method used in statistical process control (SPC). In the literature two phases of control charting are distinguished: Phase I and Phase II control charting. Phase I control charting consists of two stages: Stage 1, the retrospective stage, and Stage 2, the prospective stage (Koning & Does, 2000). During Phase I, the appropriate control charting methods must be determined, and the appropriate process parameters estimated (Jones, 2002).

The techniques associated with Phase I include analyzing sample data using gauge repeatability and reliability (GR&R) studies to investigate measuring system accuracy and variability, using capability indices to determine if a process is capable of producing within specification, using histograms and probability plots to verify distributional assumptions, using outlier detection tools (Ramsey & Ramsey, 2007) to detect and remedy special causes of variation in the process, and obtaining reliable estimates of the process parameters (Montgomery, 1997). Thus, part of Phase I can

be considered a data editing process wherein outlying or contaminated data are removed from the sample to enable estimation of the appropriate process parameters.

Phase II control charting is the actual use of the desired control chart to monitor and control a process in regards to changes in the process parameters (Woodall, 2000), distributional changes, and the randomness of the process. The construction of a Phase II control chart is based on the parameter estimates obtained in Phase I. Common Phase II control charts include the following (applied to either individual process observations or subgroups): the Shewhart-type, the exponential weighted moving average (EWMA), and the cumulative sum (CUSUM), among others (Dyer, Adams & Conerly, 2003).

It is crucial that the data collected in Phase I are good data, meaning, free from outliers (contaminated data) and representative of typical process data with no special causes of variability. Contaminated data can lead to unreliable parameter estimates which, in turn, lead to improper conclusions regarding distribution assumptions, process capability and control chart design. The use of most control charts requires the estimation of the mean,  $\mu$ , and standard deviation,  $\sigma$  (or a function thereof), of the in-control (IC) process. A process is said to be IC when only common cause variation is present, otherwise it is considered out-of-control (OC).

---

John N. Dyer is an Associate Professor of Information Systems in the College of Business Administration, with research streams in statistics, information systems, and finance. Email him at: [jdyer@georgiasouthern.edu](mailto:jdyer@georgiasouthern.edu).

The estimates used for the true process mean,  $\mu$ , and standard deviation,  $\sigma$ , are typically sample statistics, specifically, the sample mean,  $\bar{x}$ , and the sample standard deviation,  $s$ , obtained from the good data. The sample statistics used in Phase I control charting are based on the principle of maximum likelihood estimation, that is, the sample mean and sample variance are maximum likelihood estimates (MLEs) of  $\mu$  and  $\sigma^2$ , respectively.

Some of the practical deficiencies of MLEs are their lack of resistance to outliers and their general non-robustness with respect to model misspecification (Rudemo, 1982). For example, consider the following 5 data values: 4, 5, 6, 7 and 100, and estimates based on MLEs (the sample mean and standard deviation). The sample mean and variance of all five data values are 24.4 and 1,781, respectively. If the data value of 100 is identified as an outlier and removed, then the new MLEs for the mean and variance are 5.5 and 1.69, respectively. Although the magnitude of the outlier is absurdly large, it is obvious that the MLEs cannot resist the influence of the large value. The values of the new MLEs are dramatically different, but they are more representative of the true nature of the data values. Recall, one emphasis of Phase I control charting is to identify and remove outliers, hence providing reliable estimates of the true process parameters. It should also be noted that, although MLEs are nonresistant to outliers, they are typically preferred because of their constructive nature as well as their asymptotic optimality properties.

To overcome the deficiencies of MLEs and better enable the practitioner to obtain reliable parameter estimates, this article proposes the use of a specific nonparametric density estimation technique using a form of the integrated square error (ISE) estimator, also called L2E. Scott (2001) provides the theoretical construct of the L2E and the interested reader is encouraged to review the article.

In this study, the L2E technique is shown to provide parameter estimates that are robust to contaminated data and to be constructive in nature. For example, considering the full data set previously discussed, the L2E estimates of the mean and variance (obtained through a simply executed Excel spreadsheet

algorithm) are 5.5 and 2.25, respectively. Notice how the L2E estimates are robust to the inclusion of the outlier.

Although Scott (2001) introduced the L2E as an estimator of process parameters, evidences the estimator's robustness to outliers in large data sets, and shows its constructive nature, this research explores the properties of the L2E as an alternative estimator to MLE across a broad range of sample sizes and a broad range of data contamination affecting the mean alone, the variance alone, and the mean and variance together. This study also compares the absolute difference between MLE and L2E over the range of sample sizes and contaminations (mean, variance, and mean-variance), and shows that the L2E estimates are as good as MLE estimates in almost all cases. Additionally, the relative efficiency of MLE versus L2E estimates is compared across all cases and it is shown that the L2E estimates are more robust in most cases than MLE estimates.

The literature related to Phase I control charting for univariate processes is limited. Readers are referred to Chou & Champ (1995), Koning & Does (2000), Newton & Champ (1997), Sullivan & Woodall (1996), and Woodall (2000). Surprisingly, the focus of the majority of the literature is devoted to methods for multivariate Phase I SPC (Alt & Smith, 1988; Sullivan & Woodall, 1994; Sullivan, Barrett & Woodall, 1995; Woodall, 2000).

#### Overview of the Phase I Environment

During Phase I, process data are collected and analyzed to enable Phase II control charting. After the data are collected, the SPC method can be considered as the combination of Phase I and Phase II applications. The general SPC method can be thought of in terms of four design steps. The first three steps occur in the Phase I environment and step 4 occurs in the Phase II environment.

#### Step 1:

Identify the desired control chart (for monitoring individual observations or subgroup data), the required parameters, and the desired IC average run length (ARL). The IC ARL is the average number of samples taken until an IC

process indicates a statistic outside of the control limits.

Step 2:

Determine the subgroup size,  $n$ , and the number of subgroups,  $m$ , which will be used to estimate the parameters of the IC process. Obtain a reference sample of  $m$  subgroups of size  $n \geq 1$  observations.

Step 3:

Ensure that the reference sample is representative of the IC process, simultaneously estimating the required control chart parameters using a robust technique, such as, L2E (recommended herein) or an iteratively robust technique like MLE.

Step 4:

Apply the desired control chart to an ongoing process, monitoring, controlling and adjusting the process as it evolves.

In Step 1, the typical choice of control chart is related to the desire for quick detection of extreme changes in process parameters versus eventual detection of minor changes in process parameters (Dyer, Adams & Conerly, 2003; Lin & Adams, 1996). The Shewhart-type control charts are commonly used for the former, and the EWMA and CUSUM control charts are used for the latter. The choice of the IC ARL in Step 1 involves practical and economic considerations, depending largely on the costs associated with false alarms versus concealment of true process changes (Dyer, Adams & Conerly, 2003).

In Step 2, the subgroup size ( $n$ ) is a function of the sampling frequency, the process output rate, and practical considerations and limitations regarding time and costs. Marsaglia, Maclaren & Bray (1964) provide a discussion of the selection of an appropriate subgroup size ( $n$ ) and sampling frequency to design control charts. The choice of the number of subgroups ( $m$ ) is most likely an economic consideration (Jones, 2002). If contaminated data exist in the reference sample, the parameter estimates obtained can be adversely affected if MLEs are used to obtain parameter estimates (L2E to a lesser degree). Small reference samples tend to

magnify the adverse effects of estimation. A widely accepted heuristic is that  $m = 30$  subgroups from a process will provide reasonable estimates (Jones, 2002); Quesenberry (1993) suggests at least  $m = 100$  subgroups of size  $n = 5$  to estimate the parameters for the Shewhart-type control chart. Jones, Champ & Rigdon (2001) showed that an  $m$  much greater than 100, up to  $m = 400$ , is often required when designing an EWMA control chart.

In Step 3, the reference sample obtained in Step 2 is analyzed in order to estimate the unknown parameters and to determine the state of the process (IC versus OC). This is also the stage when distributional and randomness assumptions are verified, as well as when GR&R and capability studies are conducted. Concerning parameter estimation, if MLEs are used, the resulting values are the estimates used to construct an initial control chart with limits set according to the desired IC ARL in Step 1. In Stage 2, the control charts are used for prospective monitoring of the reference sample to determine departures from the estimated parameters. The control charts are primarily used to detect contaminated data or nonrandom process output, that is, data resulting from special cause variation.

Step 3 is often an iterative process, wherein contaminated data are identified (to the degree possible) and removed using a control chart based on the initial parameter estimates (MLEs). Any contaminated data identified are investigated and removed, new MLEs are obtained, a new control chart is constructed using the MLE values and more contaminated data are removed.

The process of parameter estimation and control chart removal of contaminated data continues until sufficient experience has been accumulated so that the IC parameters are effectively considered to be known through estimation. It should also be noted that if a large degree of contaminated data exist in the reference sample (as a percent of the sample size), or the magnitude of contaminated data is large (measured in terms of shifts in the process mean or variability), then the initial control limits may be inflated to a point where the contaminated data are hidden and unidentifiable. If this is the case, the Phase II parameter

estimates will be unreliable. If L2E estimates are used instead, it will be shown that the iterative process in Step 3 might be minimized by providing a more robust set of parameter estimates in the first iteration, which will lead to a more robust set of control limits, thus enabling more efficient detection and removal of contaminated data.

### Methodology

#### The L2E Estimation Technique

The L2E estimation criterion for the two-parameter normal density technique requires the minimization of the L2E function with respect to the parameters  $\mu$  and  $\sigma$ . (See Scott (2001) for the derivation of the general L2E criterion and specification of the two-parameter normal density.) Suppose a sample of size  $n \geq 1$  is drawn from a normal distribution with mean,  $\mu$ , and standard deviation,  $\sigma$ . Let the  $n$  sample data be represented by  $x_1, x_2, \dots, x_n$ , and let the univariate normal density be denoted by  $\phi(x|\mu, \sigma)$ . The minimization of the normal L2E function (equation 1) with respect to  $\mu$  and  $\sigma$  produces the L2E estimates,  $\hat{\mu}, \hat{\sigma}$ , that is, the estimation criterion is shown as:

$$\text{L2E}(\hat{\mu}, \hat{\sigma}) = \arg \min_{\mu, \sigma} \left[ \frac{1}{2\sigma\sqrt{\pi}} - \frac{2}{n} \sum_{i=1}^n \phi(x_i|\mu, \sigma) \right]. \quad (1)$$

Observe that the L2E minimizes a function of the sum of the densities; however, the MLE can be shown to maximize a function of the product of the densities. For values of  $x$  extremely distant from  $\mu$ , the density value approaches zero. As a result, the L2E utilizes only the largest portion of the data that matches the model (good data), that is,  $x$  values located within a reasonable distance of  $\mu \pm 3\sigma$ . In effect, the L2E criterion ignores contaminated data, hence generally providing more robust parameter estimates. Because MLE must account for all the data, the fits often blur the distinction between good data and contaminated data (Scott, 2001). In cases wherein there are no contaminated data, the L2E and MLE estimates are nearly equal. It can be shown through consistency theory that, for a large sample of

uncontaminated data, MLE is a very good estimator (Mood, Graybill & Boes, 1970); other estimators, such as the L2E may be just as good, but not better. In this study the L2E is shown to be just as good when the reference sample is uncontaminated and better in almost all simulated cases when contamination exists.

### Results

#### Comparison with MLEs

Unfortunately there are few example data sets that cover the range of samples sizes and contamination types and levels described herein. Montgomery (1997) provides some of the most referenced data sets in SPC research, but unfortunately none of these have sufficient examples required to cover the 96 cases of sample sizes and contamination types and levels described in this article. Simulation results are therefore used to investigate the behavior of the L2E estimates across a broad range of sample sizes as well as types and levels of data contamination. In lieu of borrowing an example data set, the simulation results are used to reveal the behavior of the L2E estimates over a broad range of cases and an example application is provided to assist the user in applying the L2E technique.

Regarding the simulation results, Tables 1a and 1b reveal average L2E and MLE estimates for  $\mu$  and  $\sigma^2$  ( $\sigma^2$  reported as  $\sigma$ ) based on averaging 10,000 simulations of  $n = 100$  normal pseudo-random variables representing differing levels and degrees of good versus contaminated data. (A complete description of the simulation design is provided in the Appendix.) The good data (IC process) are random variables representing a normal ( $\mu = 0, \sigma = 1$ ) process,  $N(0, 1)$ . The contaminated data are drawn from a normal process with parameters that vary from the IC process. Levels of contamination refer to the number of contaminated data values ( $cn$ ) in a sample of size  $n = 100$  and degrees of contamination refer to whether the contaminated data has experienced a mean shift alone, a shift in the standard deviation alone, or a shift in both the mean and standard deviation. Contamination levels in Tables 1a and 1b correspond to  $n = 5, 15, 25$  and  $45$ . Degrees of contamination

## EFFICIENCY OF MINIMIZED INTEGRATED SQUARE ERROR ESTIMATOR (L2E)

correspond to the following shifts (for  $cn = 5, 15, 25$  and  $45$ ):

- Mean shifts (alone) of  $\mu = 0.5, 1.0, 2.0,$  and  $3.0$  (16 cases)
- Standard deviation shifts (alone) of  $\sigma = 1.5, 2.0, 2.5, 3.0$  (16 cases)
- Simultaneous mean and standard deviation shifts representing combinations of all mean and standard deviation shifts alone (64 cases).

Tables 1a and 1b display simulation results providing 96 comparisons for average L2E versus MLE estimates of  $\mu$  and  $\sigma$ . For the IC process data ( $n = 100$  random variables generated from a  $N(0, 1)$  process, the resulting simulation based estimates are  $\hat{\mu}(L2E) = 0.0006$ ,  $\hat{\mu}(MLE) = 0.0007$ ,  $\hat{\sigma}(L2E) = 0.9988$ , and  $\hat{\sigma}(MLE) = 1.0015$ . In Tables 1a and 1b the estimates of  $\mu$  and  $\mu$  are shown as  $\hat{\mu}(L2E)$ ,  $\hat{\mu}(MLE)$ ,  $\hat{\sigma}(L2E)$  and  $\hat{\sigma}(MLE)$ . For all mean shifts and standard deviation shifts alone, the mathematical expectation and standard deviation (based on the levels and degrees of contamination) match the simulated MLE results.

In deriving the expected value, let  $X_n$  be the mixture of two normally distributed samples of size  $n$ , where  $X_{n-cn}$  is the uncontaminated distribution with  $E(X_{n-cn}) = \mu_{n-cn}$ , and  $X_{cn}$  is the contaminated distribution with  $E(X_{cn}) = \mu_{cn}$  (recall,  $cn$  is the number of contaminated data values in the combined sample of size  $n$ ). In this case, the  $E(X_n)$  is the weighted average expectation of each distribution of data, where the weights are the sample sizes from each distribution relative to the total sample size. Thus,

$$E(X_n) = \frac{n-cn}{n} E(X_{n-cn}) + \frac{cn}{n} E(X_{cn}).$$

In the case where the uncontaminated data distribution has  $E(X_{n-cn}) = 0$ ,

the  $E(X_n) = \frac{cn}{n} E(X_{cn})$ . For example, for  $X_{cn} \sim N(3, 1)$  where  $n = 100$  and  $cn = 45$ , the  $E(X_n) = \frac{45}{100}(3) = 1.35$ . This value matches

the simulated value given by  $\hat{\mu}$  (MLE) in Table 1b.

All simulated values for  $\hat{\mu}(MLE)$  (for both mean and standard deviation shifts alone) match the mathematical expectations. This is expected given that  $\hat{\mu}(MLE)$  is location invariant to distributional changes due to shifts in either the mean or standard deviation. The same can be observed for the standard deviation estimates,  $\hat{\sigma}(MLE)$ , where

$$\sigma(X_n) = \sqrt{\frac{n-cn}{n} Var(X_{n-cn}) + \frac{cn}{n} Var(X_{cn})}$$

when  $E(X_n) = 0$  and  $Var(X_{n-cn}) = 1$ . All simulated values for  $\hat{\sigma}(MLE)$  (for standard deviation shifts alone) match  $\sigma(X_n)$ . This is expected because  $\hat{\sigma}(MLE)$  is scale invariant to distributional changes due to shifts in the mean alone or the standard deviation alone. For cases where the mixed distribution has experienced both a mean shift and a standard deviation shift,  $\hat{\sigma}(MLE)$  is not scale invariant; hence, the variance is not the weighted average of mixed variance components.

### Simulation Result Comparison with MLEs

The simulation results reveal that in all cases

$$\{abs(\hat{\mu}(L2E) - \mu) \leq abs(\hat{\mu}(MLE) - \mu)\},$$

and in 95% of cases

$$\{abs(\hat{\sigma}(L2E) - \sigma) \leq abs(\hat{\sigma}(MLE) - \sigma)\}.$$

That is, the L2E estimates in almost all cases are as good (and often much better) as the MLE

estimates. This attests to the contention that the L2E estimators are as robust, or more robust, than MLE estimators.

Observe in Tables 1a and 1b that  $\hat{\mu}$  (L2E) is robust for most shifts in  $\mu_{cn}$ , for all  $\frac{cn}{n} \leq .45$ , and more robust than  $\hat{\mu}$  (MLE) in all cases. The relative efficiency measures in Tables 2a and 2b indicate that the worst cases are those with large  $\frac{cn}{n}$ , for  $\mu_{cn} \geq 2$ . When  $\mu_{n-cn} = 0$ , the relative efficiency for either mean estimator is defined as  $RE_{\mu} = 1 - \text{abs}(\hat{\mu})$  where  $\hat{\mu}$  is the estimate of  $\mu_{n-cn}$  and is the mean of the IC process. Table 3 displays the percent frequency distribution of relative efficiency measures for all cases simulated. Notice that, for mean shifts alone, 57% of  $\hat{\mu}$  (L2E) have  $RE_{\mu} > 0.80$  versus 44% of  $\hat{\mu}$  (MLE).

For shifts in the mean and standard deviation (simultaneously), the frequency of  $RE_{\mu} > 0.80$  is 81% for  $\hat{\mu}$  (L2E) and only 43% for  $\hat{\mu}$  (MLE). It appears that  $\hat{\mu}$  (L2E) is most robust when both a mean and standard deviation shift has occurred.

The relative efficiency for a standard deviation estimate is defined as

$$RE_{\sigma} = 1 - \left[ \frac{\text{abs}(\hat{\sigma} - \sigma_{n-cn})}{\sigma_{n-cn}} \right],$$

where  $\hat{\sigma}$  is the estimate of  $\sigma_{n-cn}$ , the standard deviation of the IC process. Because  $\sigma_{n-cn} = 1$  in all simulation cases,  $RE_{\sigma} = 1 - \text{abs}(\hat{\sigma} - 1)$ . Again, observe in Tables 1a and 1b that  $\hat{\sigma}$  (L2E) is robust for most shifts in  $\sigma_{cn}$ , for all  $\frac{cn}{n} \leq .45$ , and particularly when  $\mu_{n-cn} < 1$ . Notice also that  $\hat{\sigma}$  (L2E) is more robust than  $\hat{\sigma}$  (MLE) in 95%

of all cases. It appears that  $\hat{\sigma}$  (MLE) is less robust when all of  $\mu_{n-cn}$ ,  $\sigma_{n-cn}$ , and  $\frac{cn}{n}$  are large.

The relative efficiency measures in Tables 2a and 2b also indicates that these are the worst cases for  $\hat{\sigma}$  (L2E). Note in Table 3 that, for standard deviation shifts alone, 87% of  $\hat{\sigma}$  (L2E) have  $RE_{\sigma} > 0.80$  versus 50% of  $\hat{\sigma}$  (MLE). For shifts in both the mean and standard deviation (simultaneously), the frequency of  $RE_{\sigma} > 0.80$  is 69% for  $\hat{\sigma}$  (L2E) and only 31% for  $\hat{\sigma}$  (MLE). It appears that  $\hat{\sigma}$  (L2E) is more robust when only a shift in the standard deviation has occurred.

#### L2E Application Example

As noted, one advantage of using MLE is its constructive nature. In other words, it is simple to average a collection of data values or calculate the standard deviation. The L2E estimates are also constructive in nature, but require optimization techniques. Specifically, the L2E function given by equation 1 must be formulated and minimized subject to constraints. This can be readily accomplished in a spreadsheet environment with little or no knowledge of programming or minimization techniques. The authors suggest using Microsoft Excel and the spreadsheet add-in Solver. The data can be displayed in the spreadsheet, the L2E function can be formulated using the data and functions of the data as input, and the Solver function can be invoked to provide the L2E estimates via Solver's built-in optimization algorithm.

The data can represent individual observations or subgroup averages. If individual observations are used, then the resulting L2E estimates are those for process  $\mu$  and  $\sigma$ . If subgroup averages are used, the resulting L2E estimates are those for  $\mu$  and  $\sigma/\sqrt{n}$  (standard error of the mean, SE). In the latter case, multiplying the estimate of SE by  $\sqrt{n}$  yields the estimate for  $\sigma$ . For practitioners familiar with optimization, the L2E estimation problem can be

EFFICIENCY OF MINIMIZED INTEGRATED SQUARE ERROR ESTIMATOR (L2E)

Table 1a: L2E and MLE Estimates of  $\mu$  and  $\sigma$

CSS	$\mu$	0.0	0.0	0.0	0.0	0.5	0.5	0.5	0.5	0.5	1.0	1.0	1.0	1.0	1.0
	$\sigma$	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.04	0.02	0.01	0.01	0.00
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.03	0.05	0.05	0.05	0.05	0.05
	$\hat{\sigma}$ (L2E)	1.00	1.01	1.02	1.02	0.99	1.01	1.01	1.02	1.02	1.01	1.01	1.02	1.02	1.02
	$\hat{\sigma}$ (MLE)	1.03	1.07	1.12	1.17	1.00	1.03	1.07	1.12	1.18	1.02	1.05	1.09	1.14	1.19
15	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.07	0.04	0.02	0.01	0.01	0.13	0.07	0.04	0.03	0.02
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.08	0.08	0.08	0.08	0.08	0.15	0.15	0.15	0.15	0.15
	$\hat{\sigma}$ (L2E)	1.04	1.07	1.08	1.09	1.00	1.05	1.07	1.09	1.09	1.05	1.07	1.08	1.09	1.10
	$\hat{\sigma}$ (MLE)	1.08	1.20	1.33	1.47	1.01	1.10	1.21	1.34	1.48	1.06	1.14	1.25	1.37	1.51
25	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.12	0.07	0.04	0.03	0.02	0.23	0.14	0.08	0.05	0.03
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.12	0.12	0.12	0.12	0.12	0.25	0.25	0.25	0.25	0.25
	$\hat{\sigma}$ (L2E)	1.08	1.13	1.17	1.19	1.01	1.09	1.14	1.17	1.19	1.08	1.13	1.16	1.18	1.19
	$\hat{\sigma}$ (MLE)	1.14	1.31	1.51	1.72	1.02	1.16	1.33	1.52	1.73	1.09	1.22	1.38	1.57	1.77
45	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.23	0.15	0.10	0.07	0.05	0.44	0.30	0.21	0.15	0.11
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.22	0.22	0.22	0.22	0.22	0.45	0.45	0.45	0.45	0.45
	$\hat{\sigma}$ (L2E)	1.16	1.16	1.39	1.47	1.01	1.19	1.31	1.40	1.48	1.12	1.26	1.36	1.43	1.49
	$\hat{\sigma}$ (MLE)	1.24	1.53	1.82	2.13	1.02	1.26	1.54	1.84	2.14	1.11	1.34	1.60	1.89	2.19

viewed in the instructional form given by objective: minimize

$$L2E = \frac{1}{2\hat{\sigma}\sqrt{\pi}} - \frac{2}{n} \sum_{i=1}^n \phi\left(x_i | \hat{\mu}, \hat{\sigma}\right)$$

by changing the values  $\hat{\mu}, \hat{\sigma}$  subject to constraints:  $\hat{\sigma} > 0$ .

Figure 1 displays the author's spreadsheet in functional form, before using

Solver to minimize the L2E function. The data values 4, 5, 6, 7, 100 are input into column B, cells B11 to B15. The MLE sample mean and standard deviation, from the MLE variance, (24.4, 42.7) are calculated and displayed in column A, cells A5 and A6, respectively, using the built-in Excel function formulas shown in Figure 2. Figure 2 displays the same spreadsheet in formula/function view, allowing replication of cell formulas by the practitioner. Figure 1, column A, cells A11 to A15, contain the calculated normal probability density function

(Npdf) values resulting from the built-in Excel function shown in Figure 2. Because the Npdf function requires input values for the mean and standard deviation, the MLE estimates are initially used, and these values are temporarily input into the L2E estimate cells, column B, cells B5 and B6. Cells B5 and B6 will eventually be overwritten and contain the L2E estimates, as provided by Solver. Figure 1, cell A2, displays the L2E function value that is to be minimized, and Figure 2 displays the formula given by equation 1 as a function of both the

Table 1b: L2E and MLE Estimates of  $\mu$  and  $\sigma$

CSS	$\mu$	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
	$\sigma$	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	0.04	0.03	0.02	0.01	0.01	0.02	0.02	0.02	0.01	0.01
	$\hat{\mu}$ (MLE)	0.10	0.10	0.10	0.10	0.10	0.15	0.15	0.15	0.15	0.15
	$\hat{\sigma}$ (L2E)	1.04	1.02	1.02	1.02	1.02	1.04	1.03	1.03	1.02	1.02
	$\hat{\sigma}$ (MLE)	1.09	1.12	1.15	1.20	1.25	1.19	1.22	1.25	1.29	1.34
15	$\hat{\mu}$ (L2E)	0.16	0.10	0.07	0.04	0.03	0.09	0.08	0.07	0.05	0.03
	$\hat{\mu}$ (MLE)	0.30	0.30	0.30	0.30	0.30	0.45	0.45	0.45	0.45	0.45
	$\hat{\sigma}$ (L2E)	1.16	1.12	1.10	1.10	1.10	1.18	1.15	1.12	1.11	1.11
	$\hat{\sigma}$ (MLE)	1.23	1.30	1.39	1.51	1.63	1.47	1.53	1.61	1.71	1.82
25	$\hat{\mu}$ (L2E)	0.34	0.22	0.14	0.09	0.06	0.28	0.20	0.15	0.11	0.07
	$\hat{\mu}$ (MLE)	0.50	0.50	0.50	0.50	0.50	0.75	0.75	0.75	0.75	0.75
	$\hat{\sigma}$ (L2E)	1.30	1.24	1.21	1.21	1.21	1.46	1.33	1.26	1.24	1.23
	$\hat{\sigma}$ (MLE)	1.32	1.43	1.57	1.74	1.92	1.64	1.73	1.85	1.99	2.16
45	$\hat{\mu}$ (L2E)	0.86	0.61	0.42	0.30	0.21	1.24	0.95	0.66	0.44	0.31
	$\hat{\mu}$ (MLE)	0.90	0.90	0.90	0.90	0.90	1.35	1.35	1.35	1.35	1.35
	$\hat{\sigma}$ (L2E)	1.52	1.54	1.54	1.55	1.57	2.09	2.01	1.87	1.75	1.69
	$\hat{\sigma}$ (MLE)	1.41	1.59	1.82	2.08	2.35	1.80	1.95	2.14	2.36	2.61



EFFICIENCY OF MINIMIZED INTEGRATED SQUARE ERROR ESTIMATOR (L2E)

Table 2a: Relative Efficiency of L2E and MLE Estimates of  $\mu$  and  $\sigma$

CSS	$\mu$	0.0	0.0	0.0	0.0	0.5	0.5	0.5	0.5	0.5	1.0	1.0	1.0	1.0	1.0
	$\sigma$	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.98	0.99	0.99	1.00	1.00	0.96	0.98	0.99	0.99	1.00
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98	0.97	0.95	0.95	0.95	0.95	0.95
	$\hat{\sigma}$ (L2E)	1.00	0.99	0.98	0.98	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.98	0.98
	$\hat{\sigma}$ (MLE)	0.97	0.93	0.88	0.83	1.00	0.97	0.93	0.88	0.82	0.98	0.95	0.91	0.86	0.81
15	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.93	0.96	0.98	0.99	0.99	0.87	0.93	0.96	0.97	0.98
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.92	0.92	0.92	0.92	0.92	0.85	0.85	0.85	0.85	0.85
	$\hat{\sigma}$ (L2E)	0.96	0.93	0.92	0.91	1.00	0.95	0.93	0.91	0.91	0.95	0.93	0.92	0.91	0.90
	$\hat{\sigma}$ (MLE)	0.92	0.80	0.67	0.53	0.99	0.90	0.79	0.66	0.52	0.94	0.86	0.75	0.63	0.49
25	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.88	0.93	0.96	0.97	0.98	0.77	0.86	0.92	0.95	0.97
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.88	0.88	0.88	0.88	0.88	0.75	0.75	0.75	0.75	0.75
	$\hat{\sigma}$ (L2E)	0.92	0.87	0.83	0.81	0.99	0.91	0.86	0.83	0.81	0.92	0.87	0.84	0.82	0.81
	$\hat{\sigma}$ (MLE)	0.86	0.69	0.49	0.28	0.98	0.84	0.67	0.48	0.27	0.91	0.78	0.62	0.43	0.23
45	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.77	0.85	0.90	0.93	0.95	0.56	0.70	0.79	0.85	0.89
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.78	0.78	0.78	0.78	0.78	0.55	0.55	0.55	0.55	0.55
	$\hat{\sigma}$ (L2E)	0.84	0.84	0.61	0.53	0.99	0.81	0.69	0.60	0.52	0.88	0.74	0.64	0.57	0.51
	$\hat{\sigma}$ (MLE)	0.76	0.76	0.18	-0.13	0.98	0.74	0.46	0.16	-0.14	0.89	0.66	0.40	0.11	-0.19

sample size ( $n$ ) in cell B8 and the summed Npdf values. Prior to invoking the Solver function, the L2E function value (shown in Figure 1) is calculated using the MLE mean and standard deviation, but referencing the cells for the L2E mean and standard deviation. Figure 3 displays the Solver dialogue box referencing (1) the minimized L2E value cell (A2) as the target cell to minimize, (2) the cells to be changed to produce the minimum L2E value (B5 and B6), and (3) the constraint requiring the standard

deviation to be non-negative. Selecting the Solve button invokes Solver to produce the L2E estimates of  $\mu$  and  $\sigma$  whose values will overwrite the MLE values temporarily stored in cells B5 and B6. After solving for the L2E estimates, the actual value of the minimized L2E function is of no practical use and can be discarded. The L2E estimates of the mean and standard deviation (based on this example) are 5.5 and 1.5, respectively.

Table 2b: Relative Efficiency of L2E and MLE Estimates of  $\mu$  and  $\sigma$

CSS	$\mu$	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
	$\sigma$	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	0.96	0.97	0.98	0.99	0.99	0.98	0.98	0.98	0.99	0.99
	$\hat{\mu}$ (MLE)	0.90	0.90	0.90	0.90	0.90	0.85	0.85	0.85	0.85	0.85
	$\hat{\sigma}$ (L2E)	0.96	0.98	0.98	0.98	0.98	0.96	0.97	0.97	0.98	0.98
	$\hat{\sigma}$ (MLE)	0.91	0.88	0.85	0.80	0.75	0.81	0.78	0.75	0.71	0.66
15	$\hat{\mu}$ (L2E)	0.84	0.90	0.93	0.96	0.97	0.91	0.92	0.93	0.95	0.97
	$\hat{\mu}$ (MLE)	0.70	0.70	0.70	0.70	0.70	0.55	0.55	0.55	0.55	0.55
	$\hat{\sigma}$ (L2E)	0.84	0.88	0.90	0.90	0.90	0.82	0.85	0.88	0.89	0.89
	$\hat{\sigma}$ (MLE)	0.77	0.70	0.61	0.49	0.37	0.53	0.47	0.39	0.29	0.18
25	$\hat{\mu}$ (L2E)	0.66	0.78	0.86	0.91	0.94	0.72	0.80	0.85	0.89	0.93
	$\hat{\mu}$ (MLE)	0.50	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.25	0.25
	$\hat{\sigma}$ (L2E)	0.70	0.76	0.79	0.79	0.79	0.54	0.67	0.74	0.76	0.77
	$\hat{\sigma}$ (MLE)	0.68	0.57	0.43	0.26	0.08	0.36	0.27	0.15	0.01	-0.16
45	$\hat{\mu}$ (L2E)	0.14	0.39	0.58	0.70	0.79	-0.24	0.05	0.34	0.56	0.69
	$\hat{\mu}$ (MLE)	0.10	0.10	0.10	0.10	0.10	-0.35	-0.35	-0.35	-0.35	-0.35
	$\hat{\sigma}$ (L2E)	0.48	0.46	0.46	0.45	0.43	-0.09	-0.01	0.13	0.25	0.31
	$\hat{\sigma}$ (MLE)	0.59	0.41	0.18	-0.08	-0.35	0.20	0.05	-0.14	-0.36	-0.61

Table 3: Percent Frequency of L2E and MLE Estimates of  $\mu$  and  $\sigma$  within a Range of Relative Efficiency

Range of Relative Efficiency		$\mu$ Shifts Alone		$\sigma$ Shifts Alone		$\mu$ and $\sigma$ Shifts			
		$\hat{\mu}$ (L2E)	$\hat{\mu}$ (MLE)	$\hat{\sigma}$ (L2E)	$\hat{\sigma}$ (MLE)	$\hat{\mu}$ (L2E)	$\hat{\mu}$ (MLE)	$\hat{\sigma}$ (L2E)	$\hat{\sigma}$ (MLE)
0.90	1.00	38%	19%	56%	19%	67%	19%	47%	17%
0.80	0.90	19%	25%	31%	31%	14%	25%	23%	14%
0.70	0.80	19%	13%	6%	13%	6%	13%	9%	20%
0.60	0.70	6%	6%	6%	19%	5%	6%	13%	19%
0.50	0.60	6%	13%	0%	13%	3%	13%	5%	19%
<0.50		12%	18%	0%	6%	5%	24%	3%	11%

# EFFICIENCY OF MINIMIZED INTEGRATED SQUARE ERROR ESTIMATOR (L2E)

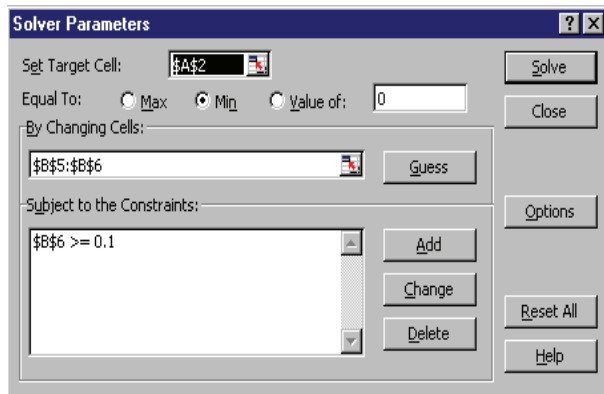
Figure 1: Functional Form Excel Spreadsheet  
(Prior to using Solver function)

	A	B	C
1	<b>Min L2E Value</b>		
2	-0.00772		
3	<b>Estimator</b>		
4	<b>MLE</b>	<b>L2E</b>	
5	24.4	24.4	<b>Mean</b>
6	42.3	42.7	<b>St Dev</b>
7			
8	<b>Sample Size: n =</b>	5	
9			
10	<b>Npdf</b>	<b>Data</b>	
11	0.00834	4	
12	0.00843	5	
13	0.00851	6	
14	0.00860	7	
15	0.00195	100	

Figure 2: Formula/Function View Excel Spreadsheet

	A	B	C
1	<b>Min L2E Value</b>		
2	$=(((2*B6*((PI())^0.5)^{-1}))-2*SUM(A11:A15)/B8)$		
3	<b>Estimator</b>		
4	<b>MLE</b>	<b>L2E</b>	
5	$=AVERAGE(B11:B15)$	24.4	<b>Mean</b>
6	$=STDEV(B11:B15)$	42.7	<b>St Dev</b>
7			
8	<b>Sample Size: n =</b>	$=COUNT(B11:B15)$	
9			
10	<b>Npdf</b>	<b>Data</b>	
11	$=NORMDIST(B11, \$B\$5, \$B\$6, FALSE)$	4	
12	$=NORMDIST(B12, \$B\$5, \$B\$6, FALSE)$	5	
13	$=NORMDIST(B13, \$B\$5, \$B\$6, FALSE)$	6	
14	$=NORMDIST(B14, \$B\$5, \$B\$6, FALSE)$	7	
15	$=NORMDIST(B15, \$B\$5, \$B\$6, FALSE)$	100	

Figure 3: Solver Dialogue Box Referencing the Minimized L2E Value Cell



## Conclusion

The importance of Phase I control charting was discussed, particularly the estimation of appropriate parameters to enable Phase II control charting. The general SPC method was shown to be a collection of steps that include both Phase I and Phase II control charting. For the Phase I environment, the minimized integrated square error estimator, L2E, was introduced as a robust parameter estimation technique and suggested as an alternative to MLEs.

Regarding managerial implications, the L2E estimation technique was described and shown to be easily constructed and applied in a spreadsheet environment. It was also shown to be a robust alternative to MLE estimation and just as simple to apply. The study also provided insights to the importance of clean data when constructing control charts based off of the Phase I processes and how the L2E estimator can facilitate robust parameter estimation required in SPC applications.

A simulation study revealed that the L2E estimates of  $\mu$  and  $\sigma$  for a normal distribution are as good, and in most cases better, than MLE estimates when the reference sample is contaminated by shifts in the mean, the variance, or both the mean and variance. Tables based on the simulation results compare the absolute and relative performance of both the L2E and MLE estimators. Finally, an example was provided to enable an SPC practitioner, with little or no knowledge of programming or optimization, to readily apply the L2E technique.

Although this article discussed the application of L2E estimators in the SPC environment (assuming a univariate normal distribution), the technique can also be adapted to enable robust parameter estimation when discrete (Poisson) or multivariate processes are to be monitored and controlled. Additionally, the L2E is only one of several nonparametric density estimators that can be considered in the Phase I environment. Other estimators that might be of research interest include M-Estimators and estimators based on Hellinger's distance criterion.

## References

- Alt, F. B., & Smith, N. D. (1988). Multivariate process control. In *Handbook of Statistics 7*, P. R. Krishnaiah & C. R. Rao (Eds.), 333-351. New York: Elsevier.
- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71, 353-360.
- Chou, S., & Champ, C. W. (1995). A comparison of two phase I control charts. *Proceeding of the Quality and Productivity Section of the American Statistical Association*, 31-35.
- Dyer, J. N., Adams, B. M., & Conerly, M. D. (2003). The reverse moving average control chart and comparisons of forecast based monitoring schemes. *Journal of Quality Technology*, 35(2), 135-155.
- Jones, L. A. (2002). The statistical design of EWMA control charts with estimated parameters. *Journal of Quality Technology*, 34, 277-288.
- Jones, L. A., Champ, C. W., & Rigdon, S. E. (2001). The performance of exponentially weighted moving average control charts with estimated parameters. *Technometrics*, 43, 156-167.
- Kinderman, A. J., & Ramage, J. G. (1976). Computer generation of normal random variables. *Journal of the American Statistical Association*, 71, 893-896.
- Koning, A. J., & Does, R. J. M. M. (2000). CUSUM charts for preliminary analysis of individual observations. *Journal of Quality Technology*, 32, 122-132.
- Lin, W. S. W., & Adams B. M. (1996). Combined control charts for forecast-based monitoring schemes. *Journal of Quality Technology*, 28, 289-301.
- Marsaglia, G. (1964). Generating a variable from the tail of a normal distribution. *Technometrics*, 6, 101-102.
- Marsaglia, G., & T. A. Bray (1964). A convenient method for generating normal variables. *SIAM Review*, 6, 260-264.
- Marsaglia, G., MacLaren, M. D., & Bray, T. A. (1964). A fast procedure for generating normal random variables. *Communications of the ACM*, 7, 4-10.
- Montgomery, D. C. (1997). *Introduction to statistical quality control*, 3<sup>rd</sup> Ed. New York, NY: John Wiley.
- Mood, M. M., Graybill, A. G., & Boes, D. C. (1970). *Introduction to the theory of statistics*, 3<sup>rd</sup> Ed. New York, NY: McGraw-Hill.
- Newton, P. B., & Champ, C. W. (1997). Probability limits for Shewhart Phase I xbar-charts. *Proceeding of the Southeast Decision Sciences Institute Annual Conference*, Atlanta, Georgia, 234-236.
- Quesenberry, C. P. (1993). The effect of sample size on estimated limits for  $\bar{X}$  and X control charts. *Journal of Quality Technology*, 25, 237-247.
- Ramsey, P. H., & Ramsey, P. P. (2007). Optimal trimming and outlier elimination. *Journal of Modern Applied Statistical Methods*, 6(2), 355-360.
- Rudemo, M. (1982). Empirical choice of histogram and kernel density estimators. *Scandinavian Journal of Statistics*, 9, 65-78.
- Scott, W. S. (2001). Parametric statistical modeling by minimum integrated square error. *Technometrics*, 43, 274-285.
- Sullivan, J. H., & Woodall, W. H. (1996). A control chart for preliminary analysis of individual observations. *Journal of Quality Technology*, 28, 265-278.
- Sullivan, J. H., & Woodall, W. H. (1994). A comparison of multivariate control charts for individual observations. *Technical Report, Applied Statistics Program*, University of Alabama, 265-278.
- Sullivan, J. H., Barrett, J. D., & Woodall, W. H. (1995). Interpreting the retrospective T2 control chart based on individual observations. *Indian Association for Productivity, Quality, and Reliability Transactions*, 20, 183-195.
- Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, 32, 341-350.

## Appendix: Simulation Description

The simulation program was designed and compiled using Microsoft Visual Basic 6.0, executed in Microsoft Excel 2000 using normal random variates generated and imported from

## EFFICIENCY OF MINIMIZED INTEGRATED SQUARE ERROR ESTIMATOR (L2E)

Microsoft FORTRAN PowerStation for Windows, Version 4.0, FORTRAN 90. Each simulation was conducted according to steps provided below. A series of 100  $N(0, 1)$  random variates was generated by FORTRAN MSIMSL subroutine RNNOA.

Routine RNNOA generates pseudorandom numbers from a standard normal (Gaussian) distribution using an acceptance/rejection technique due to (Kinderman & Ramage, 1976). In this method, the normal density is represented as a mixture of densities over which a variety of acceptance/rejection methods due to (Marsaglia, 1964), (Marsaglia & Bray, 1964), and (Marsaglia, Maclaren & Bray, 1964) are applied. The final parameter estimates for each of the 96 cases were based on 10,000 simulations, which provided a maximum margin of error of 0.02 in estimation of the MLE means, with 95% confidence. These variates were the simulated observations,  $X_i$ 's, for each of the cases investigated.

Step 1:

- a. For estimation of the mean (the 16 cases of a mean shift only), a shift in the mean was induced in the simulated observations affecting  $cn$  of the  $n = 100$  variates. The values of  $cn = 5, 15, 25$  and  $45$  (levels of contamination), and the magnitudes of shifts were  $\mu_{cn} = 1.50, 2.00, 2.50$  and  $3.00$  (degrees of contamination). Every combination of  $cn$  and  $\sigma_{cn}$  produced the 16 cases.
- b. For estimation of the standard deviation (the 16 cases of a standard deviation shift only), a shift in the standard deviation was induced in the simulated observations affecting  $cn$  of the  $n = 100$  variates. Again, the values of  $cn = 5, 15, 25$  and  $45$ , and the magnitudes of shifts were  $\sigma_{cn} = 1.50, 1.00, 2.00$  and  $3.00$ . Every combination of  $cn$  and  $\mu_{cn}$  produced the 16 cases.

- c. For estimation of the mean and standard deviation (the 64 cases of both a mean and standard deviation shift), a shift in each parameter was induced in the simulated observations affecting  $cn$  of the  $n = 100$  variates. Again, the values of  $cn = 5, 15, 25$  and  $45$ , and the magnitudes of shifts were  $\mu_{cn} = 1.50, 2.00, 2.50, 3.00$  and  $\sigma_{cn} = 1.50, 1.00, 2.00$  and  $3.00$ . Every combination of  $cn, \mu_{cn}$ , and  $\sigma_{cn}$  produced the 64 cases.

Step 2:

The individual L2E and MLE estimates of  $\mu$  and  $\sigma$  (10,000 for each estimate, per case) were calculated using the procedures described in the article.

Step 3:

The average L2E and MLE estimates of  $\mu$  and  $\sigma$  for each case was obtained by averaging over the 10,000 individual estimates for each estimator.

## Maximum Likelihood Solution for the Linear Structural Relationship With Three Parameters Known

Androulla Michaeloudis  
Middlesex University Business School  
London, England

---

A maximum likelihood solution is obtained for the simple linear structural relation model where the underlying incidental distribution and one error variance are assumed known. Expressions for the asymptotic standard errors of the maximum likelihood estimates are obtained and these are verified using a simulation study.

Key words: Maximum likelihood estimates, linear structural relation, errors-in-variables model, asymptotic standard errors, simulation.

---

### Introduction

A biochemical assay is a procedure used to measure an unknown quantity ( $\eta$ ) of a specified substance (analyte) present in a biological material, such as blood, obtained in the form of a test specimen. Biochemists are often faced with the problem of assessing the comparative performance of a new assay method with a well established reference assay method (method comparison study). An important aspect of this assessment is an examination of the degree of agreement between the results produced. Inaccuracy is unavoidable due to the complexities surrounding the measurement process. The so-called true value of the quantity of analyte can never be known in any absolute sense as the result of the test sample's composition. For example, non-analyte components present in a biological material can either enhance or inhibit the response of the analyte. These lead to what is referred to as interference biases (Strike, 1981). Different models and statistical methods have been employed as well as criticized in assessing

method comparison studies (Bland & Altman, 1986; Stockl, Dewitte, & Thienpont, 1998; Linnet, 1999). This article proposes a method comparison study for the linear structural relation of an errors-in-variables model which takes into account the presence of random errors in assays and in the recalibration effect, as well as interference effects in the biological test material. The model is complicated, but in simplified form is given by the simple errors-in-variables model as:

$$\begin{aligned} X &= U + \delta \\ Y &= V + \varepsilon \\ V &= \alpha + \beta U, \end{aligned} \tag{1}$$

where  $\alpha$  and  $\beta$  are constants defining a linear structural relation between the unobserved variables  $U$  and  $V$ . The latter are known functions of the unknown quantity of analyte of interest, that is  $U = f(\eta)$  and  $V = g(\eta)$ ,  $\delta$  and  $\varepsilon$  are the errors associated with the reference ( $X$ ) and new ( $Y$ ) assay methods respectively. It is assumed that  $\delta$  and  $\varepsilon$  are normally and independently distributed  $N(0, \sigma_\delta^2)$  and  $N(0, \sigma_\varepsilon^2)$  respectively, and are independent of  $U$ . The random variable  $U$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , that

---

Androulla Michaeloudis is a Senior lecturer in Statistics in the Department of Economics and Statistics. Email: A.Michaeloudis@mdx.ac.uk.

is  $N(\mu, \sigma^2)$ . Thus,  $(x_1, y_1), \dots, (x_n, y_n)$  are  $n$  independent observations of a bivariate normal variable  $(X, Y)$ ,  $N(\mu, \alpha + \beta\mu, \sigma^2 + \sigma_\delta^2, \beta^2\sigma^2 + \sigma_\epsilon^2, \rho)$ , where  $\rho = \beta\sigma^2 \{(\beta^2\sigma^2 + \sigma_\epsilon^2)(\sigma^2 + \sigma_\delta^2)\}^{-\frac{1}{2}}$ .

Birch (1964) and Barnett (1967) have obtained maximum likelihood solutions to model (1) for the cases where one  $(\sigma_\delta^2)$  and both error variances  $(\sigma_\delta^2, \sigma_\epsilon^2)$  are known. Note that in both cases the likelihood function has never been provided; this is provided in this article. The strengths and weaknesses of the reference method should be well-known to the analysts from their own direct experience and from nationally organized quality control schemes (Strike, 1981): thus, the distribution of  $U$  in the population under study should be known from extensive data for the reference method when this is used on the same population.

Under these conditions a maximum likelihood solution for the linear structural relation of the simple errors-in-variables model (1) with three parameters known, namely  $\mu, \sigma^2$

and  $\sigma_\delta^2$ , is considered herein. The information matrix for this case will be derived and, upon inverting this, expressions for the asymptotic standard errors of the derived maximum likelihood estimates will be obtained. These derived expressions will be verified by a simulation study. The effect, if any, of the knowledge of  $\mu$  and  $\sigma^2$  on the estimates, in particular the estimate of the slope of the linear structural relation, will be examined and will be compared with the derived maximum likelihood solution where only  $\sigma_\delta^2$  is known.

### The Problem

Assuming knowledge of  $\mu, \sigma^2$ , and  $\sigma_\delta^2$  the structural errors-in-variables model (1) has three unknown parameters and a set of minimal sufficient statistics of dimension five and as such the model is expected to be identifiable. For a given set of real observations  $X = (X, Y)$ , the likelihood function for all real  $\alpha, \beta$ , and  $\sigma_\epsilon^2 \geq 0$ , where the set of unknown parameters is  $\Psi = (\alpha, \beta, \sigma_\epsilon^2)$ . The likelihood function is a continuous function; it tends to zero as  $|\beta|$  or  $\sigma_\epsilon$  become infinite and is

Figure 1: Formulas (2), (3) and (4)

Formula 2:

$$l(\underline{X}, \underline{\Psi}) = \frac{\text{constant}}{[\sigma^2(\beta^2\sigma_\delta^2 + \sigma_\epsilon^2) + \sigma_\delta^2\sigma_\epsilon^2]^{n/2}} \times \exp. \left[ \frac{-1}{2[\sigma^2(\beta^2\sigma_\delta^2 + \sigma_\epsilon^2) + \sigma_\delta^2\sigma_\epsilon^2]} \times \left\{ (\beta^2\sigma^2 + \sigma_\epsilon^2) \sum_i^n (X - \mu)^2 - 2\beta\sigma^2 \left( nS_{xx} + (\bar{Y} - \alpha - \beta\mu) \sum_i^n (X - \mu) \right) \right\} \right]$$

Formula 3:

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \left( \frac{\sigma^2 \bar{X} + \mu \sigma_\delta^2}{\sigma^2 + \sigma_\delta^2} \right)$$

Formula 4:

$$\hat{\beta}^3 + \left( \frac{S_{xy}}{\sigma_\delta^2} \right) \hat{\beta}^2 + \left\{ \frac{\hat{\sigma}_\epsilon^2 S_{xx}}{\sigma_\delta^4} - \frac{\lambda(S_{yy} - \hat{\sigma}_\epsilon^2)}{\sigma_\delta^2} \right\} \hat{\beta} - \frac{\lambda \hat{\sigma}_\epsilon^2 S_{xy}}{\sigma_\delta^4} = 0$$

differentiable everywhere (see formula (2) in Figure 1). By partially differentiating the log-likelihood function with respect to the three unknown parameters and equating to zero, three equations are obtained which can be rearranged to give formulas (3) and (4) (also shown in Figure 1).

$$\hat{\sigma}_\varepsilon^2 = S_{yy} + \frac{\hat{\beta}^2 S_{xx}}{\lambda^2} - \frac{2\hat{\beta} S_{xy}}{\lambda} - \frac{\hat{\beta}^2 \sigma_\delta^2}{\lambda}, \quad (5)$$

where  $S_{xx}, S_{yy}, S_{xy}$  are the sample statistics and  $\lambda = \frac{(\sigma^2 + \sigma_\delta^2)}{\sigma^2}$ .

The monotonicity of the likelihood function (2), and the fact that the likelihood tends to zero as  $\sigma_\varepsilon^2$  tends to  $\pm\infty$ , implies that there is only one value for  $\hat{\sigma}_\varepsilon^2$  for which the likelihood function is a maximum. Therefore, the log-likelihood is maximized either when  $\hat{\sigma}_\varepsilon^2 > 0$  or when  $\hat{\sigma}_\varepsilon^2 = 0$ ; these cases are considered next, but the case  $\hat{\sigma}_\varepsilon^2 = 0$  is not a practical case in a method comparison study.

Case 1:  $\hat{\sigma}_\varepsilon^2 > 0$

In this case the maximum likelihood estimates of  $\alpha$ ,  $\beta$  and  $\sigma_\varepsilon^2$  are given by the solutions of likelihood equations (3) – (5). By substituting for  $\hat{\sigma}_\varepsilon^2$  in (4), the following cubic equation for  $\hat{\beta}$  is obtained

$$\hat{\beta}^3 - 3\lambda b_1 \hat{\beta}^2 + \lambda^2 \left( \frac{b_1}{b_2} + 2b_1^2 \right) \hat{\beta} - \lambda^3 \frac{b_1^2}{b_2} = 0, \quad (6)$$

which factorizes to

$$(\hat{\beta} - \lambda b_1) \left( \hat{\beta}^2 - 2\lambda b_1 \hat{\beta} + \lambda^2 \frac{b_1}{b_2} \right) = 0, \quad (7)$$

where  $b_1$  and  $b_2$  are the two sample regression coefficients, that is  $b_1 = \frac{S_{xy}}{S_{xx}}$  and  $b_2 = \frac{S_{xy}}{S_{yy}}$ . The cubic equation (6) yields one real root

$$\hat{\beta} = \lambda b_1, \quad (8)$$

and two complex roots

$$\hat{\beta} = \lambda \left[ b_1 \pm \left\{ \frac{b_1}{b_2} (r^2 - 1) \right\}^{\frac{1}{2}} \right], \quad (9)$$

where  $r$  is the sample correlation coefficient  $(b_1 b_2)^{\frac{1}{2}}$ . Substituting the real root for  $\hat{\beta}$  in (5) yields the following equation

$$\hat{\sigma}_\varepsilon^2 = S_{yy} - b_1^2 S_{xx} \left( 1 + \frac{\lambda \sigma_\delta^2}{S_{xx}} \right). \quad (10)$$

Case 2:  $\hat{\sigma}_\varepsilon^2 = 0$

Placing  $\hat{\sigma}_\varepsilon^2 = 0$  in cubic equation (4) leads to

$$\hat{\beta} \left( \hat{\beta}^2 + \hat{\beta} \frac{S_{xy}}{\sigma_\delta^2} - \frac{\lambda S_{yy}}{\sigma_\delta^2} \right) = 0: \quad (11)$$

this implies that either  $\hat{\beta} = 0$  or

$$\hat{\beta}^2 + \left( \frac{S_{xy}}{\sigma_\delta^2} \right) \hat{\beta} - \left( \frac{\lambda S_{yy}}{\sigma_\delta^2} \right) = 0. \quad (12)$$

The case  $\hat{\beta} = 0$  is excluded because, at this point, the likelihood function is undefined. Equation (12) factorizes to yield two real roots

$$\hat{\beta} = \frac{1}{2\sigma_\delta^2} \left[ (-S_{xy}) \pm \left\{ S_{xy}^2 + 4\lambda \sigma_\delta^2 S_{yy} \right\}^{\frac{1}{2}} \right], \quad (13)$$

where the one with same sign as  $S_{xy}$  is the maximum likelihood estimator of  $\beta$ .



## Maximum Likelihood Solution

The complete maximum likelihood solution of the linear structural errors-in-variables model for  $\mu$ ,  $\sigma^2$ , and  $\sigma_\delta^2$  known is as follows. If

1.

$$S_{yy} > \frac{S_{xy}^2}{S_{xx}} \left\{ 1 + \frac{\sigma_\delta^2 (\sigma^2 + \sigma_\delta^2)}{\sigma^2 S_{xx}} \right\}$$

then

$$\hat{\beta} = \frac{(\sigma^2 + \sigma_\delta^2) S_{xy}}{\sigma^2 S_{xx}},$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \left( \frac{\sigma^2 \bar{X} + \mu \sigma_\delta^2}{\sigma^2 + \sigma_\delta^2} \right),$$

and

$$\hat{\sigma}_\varepsilon^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}} \left\{ 1 + \frac{\sigma_\delta^2 (\sigma^2 + \sigma_\delta^2)}{\sigma^2 S_{xx}} \right\};$$

otherwise

2.

$$\hat{\beta} = \frac{1}{2\sigma_\delta^2} \left[ (-S_{xy}) \pm \left\{ S_{xy}^2 + \frac{4\sigma_\delta^2 (\sigma^2 + \sigma_\delta^2) S_{yy}}{\sigma^2} \right\}^{1/2} \right], \quad (14)$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \left( \frac{\sigma^2 \bar{X} + \mu \sigma_\delta^2}{\sigma^2 + \sigma_\delta^2} \right),$$

and

$$\hat{\sigma}_\varepsilon^2 = 0.$$

Because the sample statistics  $S_{xy}$ ,  $S_{xx}$ , and  $S_{yy}$  converge in probability to  $\beta\sigma^2$ ,  $(\sigma^2 + \sigma_\delta^2)$  and  $(\beta^2\sigma^2 + \sigma_\varepsilon^2)$  respectively, the derived maximum likelihood estimates (14) are consistent estimates of  $\alpha$ ,  $\beta$  and  $\sigma_\varepsilon^2$ . If  $\sigma_\delta^2$  is set equal to zero ( $\sigma_\delta^2 = 0$ ) so that the errors-in-variables model (1) reduces to the simple linear regression model, the derived results are in

agreement with the established results applicable to the latter model (that is,  $\hat{\alpha}_{OLS}$  and  $\hat{\beta}_{OLS}$ ).

It is also noted that further knowledge of the specific values of  $\mu$  and  $\sigma^2$  are relevant to the estimation of the scale parameter  $\alpha$ . This is in contrast to all other solutions obtained where  $\mu$  and  $\sigma^2$  were unknown, that is, in all previous solutions with  $\mu$  and  $\sigma^2$  unknown,  $\hat{\alpha} = f(\hat{\beta}, \hat{\mu})$  (Birch, 1964 & Barnett, 1967), while with  $\mu$  and  $\sigma^2$  known,  $\hat{\alpha} = f(\hat{\beta}, \mu, \sigma^2, \sigma_\delta^2)$  where  $f$  denotes a function.

It is worth noting that the derived solution can lead to the maximum likelihood solution when only  $\sigma_\delta^2$  is known, and when  $\mu$  and  $\sigma^2$  are substituted by their corresponding estimates. This establishes the compatibility of the derived solution with the maximum likelihood solution where only one error variance is known.

Note that condition (1) of (14) forces the estimate for  $\sigma_\varepsilon^2$  to be positive, that is, the first expression for  $\hat{\beta}$  applies if the likelihood does not reach its maximum in a boundary point owing to a positivity constraint of  $\sigma_\varepsilon^2$ . Because the probability of this to be true tends to one as the number of observations increases, it follows that  $\hat{\beta}$  is asymptotically equivalent to

$$\hat{\beta}_A = \frac{(\sigma^2 + \sigma_\delta^2) S_{xy}}{\sigma^2 S_{xx}}. \quad (15)$$

Hence, the maximum likelihood estimates of  $\beta$  and  $\hat{\beta}_A$  have the same limiting distribution and their asymptotic standard errors are identical.

## Asymptotic Variances

Expressions for the asymptotic variances of the maximum likelihood estimates of  $\underline{\psi} = (\alpha, \beta, \sigma_\varepsilon^2)$  can be obtained directly from the inverse information matrix,  $[I(\underline{\psi})]^{-1}$ . The information matrix is derived by calculating the

expected values of the second order derivatives of the log-likelihood function.

$$I(\underline{\psi}) = \frac{n}{T} \begin{pmatrix} \sigma^2 + \sigma_\delta^2 & \mu(\sigma^2 + \sigma_\delta^2) & 0 \\ \mu(\sigma^2 + \sigma_\delta^2) & \mu^2(\sigma^2 + \sigma_\delta^2) + M & \frac{\beta\sigma^2\sigma_\delta^2(\sigma^2 + \sigma_\delta^2)}{T} \\ 0 & \frac{\beta\sigma^2\sigma_\delta^2(\sigma^2 + \sigma_\delta^2)}{T} & \frac{(\sigma^2 + \sigma_\delta^2)^2}{2T} \end{pmatrix} \quad (16)$$

where

$$T = \sigma^2(\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2) + \sigma_\delta^2\sigma_\varepsilon^2$$

and

$$M = \frac{\sigma^4}{T^2} \left\{ \begin{array}{l} \sigma^2(\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2)^2(\sigma^2 + 2\sigma_\delta^2) \\ + \sigma_\delta^4\sigma_\varepsilon^2(\sigma_\varepsilon^2 + 2\beta^2\sigma_\delta^2) \end{array} \right\}.$$

The inverse of this (3×3) asymptotic covariance matrix of the maximum likelihood estimates  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\hat{\sigma}_\varepsilon^2$  is:

$$[I(\underline{\psi})]^{-1} = \frac{T}{n\sigma^4} \begin{pmatrix} \mu^2 + \frac{\sigma^4}{(\sigma^2 + \sigma_\delta^2)} & -\mu & \frac{2\beta\mu\sigma^2\sigma_\delta^2}{(\sigma^2 + \sigma_\delta^2)} \\ -\mu & 1 & -\frac{2\beta\sigma^2\sigma_\delta^2}{(\sigma^2 + \sigma_\delta^2)} \\ \frac{2\beta\mu\sigma^2\sigma_\delta^2}{(\sigma^2 + \sigma_\delta^2)} & -\frac{2\beta\sigma^2\sigma_\delta^2}{(\sigma^2 + \sigma_\delta^2)} & \frac{2MT}{(\sigma^2 + \sigma_\delta^2)^2} \end{pmatrix} \quad (17)$$

From (17), the asymptotic variances of  $\hat{\beta}$ ,  $\hat{\alpha}$  and  $\hat{\sigma}_\varepsilon^2$  are obtained as:

$$\text{var}(\hat{\beta}) = \frac{1}{n\sigma^4} \left\{ \sigma^2(\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2) + \sigma_\delta^2\sigma_\varepsilon^2 \right\}, \quad (18)$$

$$\text{var}(\hat{\alpha}) = \left\{ \mu^2 + \frac{\sigma^4}{(\sigma^2 + \sigma_\delta^2)} \right\} \text{var}(\hat{\beta}), \quad (19)$$

and

$$\text{var}(\hat{\sigma}_\varepsilon^2) = \left\{ \frac{2MT}{(\sigma^2 + \sigma_\delta^2)^2} \right\} \text{var}(\hat{\beta}). \quad (20)$$

A comparison of the above expressions (18 and 19) with the asymptotic variances of  $\hat{\beta}$  and  $\hat{\alpha}$ , where only  $\sigma_\delta^2$  is known, shows that the further knowledge of  $\sigma^2$  leads to smaller variances for the maximum likelihood estimates.

## Methodology

### Simulation Study

A simulation study was carried out using R statistical software to investigate the effect of sample size on the accuracy of the derived maximum likelihood estimates of  $\alpha$ ,  $\beta$  and  $\sigma_\varepsilon^2$  (14) and their corresponding asymptotic variances (18 – 20). Taking into account examples of data used for method comparison studies and the fact that, depending on the type of analyte considered, the sample size of a method comparison study will vary from a minimum of 17 to more than 500 (Bland & Altman, 1986; Stockl, Dewitte & Thienpont, 1998; Linnet, 1999), this simulation study considered sample sizes ranging from a minimum of 20 to a maximum of 1,000. This was also done in order to assess the effect of a sample size on the accuracy of the derived estimates. Ten thousand simulations have been considered in this study and particular attention was given to the estimates of  $\alpha$  and  $\beta$  because the values of these can allow for the estimation of possible constant and proportional interference biases in a biological test material. In all cases considered an interference bias of 10% was allowed so that  $\alpha = 0.10$  and  $\beta = 1.10$ .

Because there is a tendency for practitioners to use methods with which they are more familiar, such as the ordinary least square (OLS) estimation for the simple linear

regression model (Westgard & Hunt, 1973), the simulation study also compared the OLS estimates with the derived MLEs (14). The accuracy of these estimates is based on the mean squared error (MSE) criterion; some of the obtained results are presented in Tables 1 – 3 below.

### Results

The results are in agreement with what was expected, namely:

1. Increasing the sample size leads to a decrease in the bias of the maximum likelihood estimates and - as expected in such cases - the mean squared error reduces to the variance of the estimate.
2. The mean squared errors of the maximum likelihood estimates are less than the mean squared errors of the least squares estimates irrespectively of the sample size. It is clear that the OLS are inappropriate to use in a method comparison study where errors are assumed in both assays.
3. The accuracy of the maximum likelihood estimates particularly for  $\hat{\beta}$  and  $\hat{\alpha}$  can be achieved with samples as small as 20.
4. The expressions for the asymptotic variances have been verified for samples greater than 100 with biases less than 0.0001.

### Conclusion

Under the assumption that the parameters specifying the underlying incidental distribution  $(\mu, \sigma^2)$ , the maximum likelihood estimates of the unknown parameters  $\alpha$ ,  $\beta$  and  $\sigma_\varepsilon^2$  are obtained: these are consistent, asymptotically normal and efficient. The asymptotic variances of the estimates were obtained by the inversion of the information matrix. It has been shown that the asymptotically equivalent estimator of the slope is a function of  $\sigma^2$  and  $\sigma_\varepsilon^2$  thus utilizing the known information about the variances. The derived solution is in agreement with the case

where only  $\sigma_\varepsilon^2$  is known. In the latter case the asymptotically equivalent estimator of the slope is a function of the known variance  $\sigma_\varepsilon^2$  (Ketellapper, 1983). A simulation study verified the accuracy of the maximum likelihood estimates with samples as small as 20. This study also verified the accuracy of the asymptotic variances with biases less than 0.0001.

### Acknowledgements

The author would like to thank her colleague David Jarrett for his valuable comments on this article and Professor Bob Gilchrist for suggesting the use of the software R for the simulation work included in this study.

### References

- Barnett, V. D. (1967). A note on linear structural relationships when both residual variances are known. *Biometrika*, 54, 670-672.
- Birch, M. W. (1964). A note on the maximum likelihood estimation of a linear structural relationship. *Journal of the American Statistical Association*, 59, 1175-1178.
- Bland, J. M., & Altman, D. G. (1986). Comparing methods of measurement: Why plotting a difference against a standard method is misleading. *The Lancet*, 346, 1085-1087.
- Bland, J. M., & Altman D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, i., 307-310.
- Ketellapper, R. H. (1983). On estimating parameters in a simple linear errors-in-variables model. *Technometrics*, 25(1), 43-47.
- Linnet, K. (1999). Necessary sample size for method comparison studies based on regression analysis. *Clinical Chemistry*, 45, 882-894.
- Linnet, K. (1999). Limitations of the paired t-test for evaluation of method comparison data. *Clinical Chemistry*, 45, 314-315.
- Strike, P. W. (1981). *Medical Laboratory Statistics*. Bristol, England: John Wright-PSG.
- Stockl, D., Dewitte, K., & Thienpont, L. M. (1998). Validity of linear regression in method comparison studies: Is it limited by the statistical model or the quality of the analytical input data? *Clinical Chemistry*, 44, 2340-2346.
- Westgard J. O., & Hunt, M. R. (1973). Use and interpretation of common statistical tests in method comparison studies. *Clinical Chemistry*, 19, 49-57.

MICHAELOUDIS

Table 1: MLEs and OLS Estimates and Their Corresponding Mean Squared Errors; Simulated Variances  $\{.\}$  and Derived Asymptotic Variances of the MLEs  $[.]$  ( $\mu = 3$ ,  $\sigma = 0.45$ ,  $\sigma_\delta = 0.4$ ,  $\sigma_\varepsilon = 0.6$ )

Parameter	Sample Size ( $n$ )	Estimate		MSE	
		MLE {var} [var]	OLS	MLE	OLS
$\beta$	20	1.1017 {0.2381} [0.2069]	0.6201	0.2381	0.3095
	40	1.0956 {0.1106} [0.1035]	0.6122	0.1107	0.2727
	100	1.0999 {0.0435} [0.0414]	0.6144	0.0435	0.2493
	500	1.0989 {0.0086} [0.0083]	0.6139	0.0086	0.2390
	1,000	1.1009 {0.0042} [0.0041]	0.6150	0.0042	0.2366
$\alpha$	20	0.0965 {2.1723} [1.8857]	1.5414	2.1723	2.8173
	40	0.1127 {1.0052} [0.9429]	1.5627	1.0054	2.4631
	100	0.0997 {0.3958} [0.3771]	1.5562	0.3958	2.2472
	500	0.1036 {0.0778} [0.0754]	1.5587	0.0778	2.1526
	1,000	0.0975 {0.0380} [0.0377]	1.5551	0.0382	2.1297
$\sigma_\varepsilon^2$	20	0.2937 {0.0272} [0.0299]	-	0.0325	-
	40	0.3279 {0.0147} [0.0150]	-	0.0156	-
	100	0.3464 {0.0061} [0.0060]	-	0.0063	-
	500	0.3575 {0.0012} [0.0012]	-	0.0012	-
	1,000	0.3587 {0.0006} [0.0006]	-	0.0006	-

MLE FOR LINEAR STRUCTURAL RELATIONSHIP: THREE KNOWN PARAMETERS

Table 2: MLEs and OLS Estimates and their Corresponding Mean Squared Errors; Simulated Variances {·} and Derived Asymptotic Variances of the MLEs [·] ( $\mu = 3$ ,  $\sigma = 0.35$ ,  $\sigma_\delta = 0.4$ ,  $\sigma_\epsilon = 0.6$ )

Parameter	Sample Size ( $n$ )	Estimate		MSE	
		MLE {var} [var]	OLS	MLE	OLS
$\beta$	20	1.0890 {0.4603} [0.4179]	0.4792	0.4605	0.4799
	40	1.1017 {0.2274} [0.2089]	0.4782	0.2274	0.4299
	100	1.0958 {0.0844} [0.0836]	0.4752	0.0844	0.4063
	500	1.0992 {0.0168} [0.0167]	0.4766	0.0168	0.3918
	1,000	1.1013 {0.0083} [0.0084]	0.4776	0.0083	0.3890
$\alpha$	20	0.1318 {4.1787} [3.7831]	1.9609	4.1797	4.3437
	40	0.0950 {2.0645} [1.8916]	1.9656	2.0645	3.8840
	100	0.1124 {0.7643} [0.7566]	1.9743	0.7645	3.6603
	500	0.1026 {0.1517} [0.1513]	1.9701	0.1517	3.5267
	1,000	0.0961 {0.0749} [0.0757]	1.9674	0.0749	3.5017
$\sigma_\epsilon^2$	20	0.2880 {0.0268} [0.0294]	-	0.0323	-
	40	0.3236 {0.0147} [0.0147]	-	0.0164	-
	100	0.3451 {0.0059} [0.0059]	-	0.0061	-
	500	0.3570 {0.0012} [0.0012]	-	0.0012	-
	1,000	0.3581 {0.0006} [0.0006]	-	0.0006	-

MICHAELOUDIS

Table 3: MLEs and OLS Estimates and their Corresponding Mean Squared Errors; Simulated Variances  $\{.\}$  and Derived Asymptotic Variances of the MLEs  $[.]$  ( $\mu = 3$ ,  $\sigma = 0.4$ ,  $\sigma_\delta = 0.4$ ,  $\sigma_\epsilon = 0.6$ )

Parameter	Sample Size ( $n$ )	Estimate		MSE	
		MLE {var [var]}	OLS	MLE	OLS
$\beta$	20	1.0850 {0.3085} [0.2855]	0.5473	0.3087	0.3878
	40	1.1003 {0.1511} [0.1428]	0.5505	0.1511	0.3401
	100	1.0999 {0.0591} [0.0571]	0.5500	0.0591	0.3173
	500	1.0997 {0.0112} [0.0114]	0.5499	0.0112	0.3055
	1,000	1.0996 {0.0059} [0.0057]	0.5498	0.0059	0.3042
$\alpha$	20	0.1443 {2.7914} [2.5923]	1.7575	2.7934	3.5080
	40	0.0982 {1.3767} [1.2962]	1.7477	1.3767	3.0720
	100	0.0999 {0.5366} [0.5185]	1.7497	0.5366	2.8593
	500	0.1011 {0.1023} [0.1037]	1.7507	0.1023	2.7511
	1,000	0.1014 {0.0532} [0.0518]	1.7509	0.0532	2.7390
$\sigma_\epsilon^2$	20	0.2948 {0.0265} [0.0297]	-	0.0559	-
	40	0.3233 {0.0146} [0.0149]	-	0.0155	-
	100	0.3468 {0.0061} [0.0059]	-	0.0062	-
	500	0.3571 {0.0011} [0.0011]	-	0.0012	-
	1,000	0.3589 {0.0006} [0.0006]	-	0.0006	-

## Extension of Grizzle's Classic Crossover Design

James F. Reed III  
Christiana Care Hospital System,  
Newark, Delaware

---

The crossover design compares treatments A and B over two periods using sequences AB and BA (the AB|BA design) and is the classic design most often illustrated and critiqued in textbooks. Other crossover designs have been used but their use is relatively rare and not always well understood. This article introduces alternatives to a randomized two-treatment, two-period crossover study design. One strategy, which is to extend the classic AB|BA by adding a third period to repeat one of the two treatments, has several attractive advantages; an added treatment period may not imply a large additional cost but will allow carryover effects to be estimated and compared with the within-subject variability. Careful choice of treatment sequences will enable the first two trial periods to constitute a conventional two-period crossover trial if the third treatment period leads to excessive subject drop-outs. Four alternative designs that address the first-order carryover effect are presented. These designs have more statistical power than the classic design and allow the treatment effects to be estimated, even in the presence of a carryover effect.

Key words: Crossover design, Grizzle, carryover effect.

---

### Introduction

A crossover study is a longitudinal study in which subjects receive a sequence of different treatments; these designs are common in many scientific disciplines. In AB|BA crossover studies, subjects are randomly assigned to receive either treatment A in the first treatment period followed by treatment B in the second period or treatment B in the first period followed by treatment A in the second period. The crossover study allows for a within-subject comparison between treatments because each subject serves as his or her own control, the inter-patient variability is removed from the comparison between treatments and it can provide unbiased estimates for the differences between treatments. However, frequent misapplications of the design in clinical trials and even more frequent misanalysis of the data

have nearly doomed the crossover trial in clinical research (Freeman, 1989; Senn, 1994; Senn, 1996).

The most damning characteristic of a crossover study is the potential of a carryover effect of one treatment to the next period. To address this issue, researchers typically include washout periods in their study designs. These washout periods are thought to be of sufficient length to negate any lingering effect of one treatment into the next period. Unfortunately, what a sufficiently long washout period might be remains unclear. In this article, and in most of the literature on crossover designs, the persistence of a carryover effect is assumed to last for only a single period (a first-order carryover effect) and it is also assumed that the carryover effect is different for different treatments. If a carryover effect is suspected in any crossover trial, then a term for this effect must be included in the model and accounted for in the subsequent analysis. This article introduces three simple alternatives to Grizzle's classic AB|BA crossover design. These designs have more statistical power than the AB|BA design and allow unbiased treatment effects to

---

James F. Reed III, Ph.D., is a Senior Biostatistician. Email him at: JFReed3rd@gmail.com.

be estimated, even when a simple-order carryover effect is specified.

#### The Traditional Crossover Model with Continuous Data

The traditional design model assumes that each treatment has a simple first-order carryover effect that does not interact with the direct effect of the treatment in the subsequent period and that subject effects are either fixed or random. Although a variety of models are considered in the literature, virtually all of the work in crossover designs has the following traditional statistical model which assumes the following for the response of patient  $y_{ij}$ .

If  $y_{ijk}$  denotes the observed response of subject  $j$  ( $j = 1, \dots, n$ ) in period  $i$  ( $i = 1, \dots, p$ ), then

$$y_{ij} = \mu + \pi_i + \tau_{d(i,j)} + \lambda_{d(i-1,j)} + \beta_j + \varepsilon_{ij}.$$

where  $\pi_i$  is the effect of period  $i$ ;  $\tau_{d(i,j)}$  is the direct effect of treatment  $D$ ,  $\lambda_{d(i-1,j)}$  is the simple first-order carryover effect of treatment  $D$  and  $d(i,j)$  is the treatment allocated to patient  $j$  in period  $i$ ,  $\lambda_{d(0,j)} = 0$  for all  $j$ . It is assumed that all these effects are fixed effects.  $\beta_j$  is the effect of patient  $j$  and  $\varepsilon_{ij}$  is the error term. The random subject effect,  $\beta_j$ , and the experimental error,  $\varepsilon_{ij}$ , are assumed to be mutually independently distributed as  $N(0, \sigma_\beta^2)$  and  $N(0, \sigma_\varepsilon^2)$ .

#### The Classic AB|BA

The crossover design that compares treatments A and B over two periods using sequences AB and BA (the AB|BA design) is the classic design and is most often illustrated and critiqued in textbooks (Grizzle, 1965). Other crossover designs have been utilized but their use is relatively rare and is not always well understood. For example, when more than two treatments are to be compared, an extensive use of each subject may be desirable when the number of periods can be extended.

The primary purpose of an AB|BA crossover trial is to estimate the treatment contrast  $\tau_A - \tau_B$  (see Table 1). The period effects  $\pi_1$  and  $\pi_2$ , the first-order carryover effects  $\lambda_A$  and  $\lambda_B$ , and  $\mu$  are typically regarded as nuisance parameters that should be eliminated from any estimate. In sequence AB, the contrast  $c_1$ ,  $y_{11} -$

$y_{21}$ , has the expected value of  $E[c_1] = E[y_{11} - y_{21}] = (\pi_1 - \pi_2) + (\tau_A - \tau_B) - \lambda_A$ , while in sequence BA, the contrast  $c_2$ ,  $y_{21} - y_{22}$ , has the expected value of  $E[c_2] = E[y_{21} - y_{22}] = (\pi_1 - \pi_2) - (\tau_A - \tau_B) - \lambda_B$ .

The difference between contrasts  $c_1$  and  $c_2$  may be expressed is  $2(\tau_A - \tau_B) - (\lambda_A - \lambda_B)$ . It is then possible to generate a hypothesis by forming the differences in the two contrasts between responses for the two periods. That difference for the respective patients may be expressed by  $H_{\text{CROS}}: \{2(\tau_A - \tau_B) - (\lambda_A - \lambda_B) = 0\}$ .  $H_{\text{CROS}}$  is a combined null hypothesis tested by the difference (or crossover) test of equality of both the treatment effects and carryover effects of A and B  $\{\tau_A = \tau_B, \lambda_A = \lambda_B\}$ . The treatment effect and carryover effect are said to be aliased. The rejection of  $H_{\text{CROS}}$  is interpreted as demonstrating that the direct and/or carryover effects of A and B are different in the sense of a prevailing larger response for one treatment than the other across the two periods.

In sequence AB, the contrast  $c_3$ ,  $y_{11} + y_{21}$ , has the expected value of  $E[c_3] = E[y_{11} + y_{21}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_A$ , and in sequence BA, the contrast  $c_4$ ,  $y_{21} + y_{22}$ , has the expected value of  $E[c_4] = E[y_{21} + y_{22}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_B$ . The difference between  $c_3$  and  $c_4$  differ by  $\lambda_A - \lambda_B$ , and is a measure of the net carryover effect.

The hypothesis,  $H_{\text{SEQ}}: \lambda_A = \lambda_B$ , has been proposed for use when deciding whether the rejection of the hypothesis  $H_{\text{CROS}}$  is due mainly to differences between the direct treatment effects  $\tau_A = \tau_B$  or between the carryover effects  $\lambda_A = \lambda_B$ . A non-significant  $H_{\text{SEQ}}$  supports the contention that there is a difference between  $\tau_A$  and  $\tau_B$  as the primary contradiction of  $H_{\text{CROS}}$ . Alternatively, a significant  $H_{\text{SEQ}}$  is interpreted as indicating that the differences between  $\lambda_A$  and  $\lambda_B$  may account for the contradiction of  $H_{\text{CROS}}$ . When assessing the SEQ hypothesis, the type II error (falsely failing to reject the null hypothesis of no first-order carryover effect) is of some concern.

To reduce the probability of making a type II error, the recommendation has been to use larger than the usual  $\alpha$ , such as 25%. In bioequivalence studies, the commonly used significance level in a bioequivalence study for



EXTENSION OF GRIZZLE'S CLASSIC CROSSOVER DESIGN

Table 1: Design AB|BA

AB BA Design	Period 1 (k = 1)	Period 2 (k = 2)
Sequence AB (i = 1)	$\mu + \pi_1 + \tau_A$	$\mu + \pi_2 + \tau_B + \lambda_A$
Sequence BA (i = 2)	$\mu + \pi_1 + \tau_B$	$\mu + \pi_2 + \tau_A + \lambda_B$

Table 1 Notes:

Sequence AB (i = 1):  $E(y_{AB,1}) = \mu_{AB,1} = \mu + \pi_1 + \tau_A$ ,  $E(y_{AB,2}) = \mu_{AB,2} = \mu + \pi_2 + \tau_B + \lambda_A$

Sequence BA (i = 2):  $E(y_{BA,1}) = \mu_{BA,1} = \mu + \pi_1 + \tau_B$ ,  $E(y_{BA,2}) = \mu_{BA,2} = \mu + \pi_2 + \tau_A + \lambda_B$

In sequence AB, the contrast  $c_1$  has the expected value of  $E[c_1] = E[y_{11} - y_{21}] = (\pi_1 - \pi_2) + (\tau_A - \tau_B) - \lambda_A$

In sequence BA, the contrast  $c_2$  has the expected value of  $E[c_2] = E[y_{21} - y_{22}] = (\pi_1 - \pi_2) - (\tau_A - \tau_B) - \lambda_B$

In sequence AB, the contrast  $c_3$  has the expected value of  $E[c_3] = E[y_{11} + y_{21}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_A$

In sequence BA, the contrast  $c_4$  has the expected value of  $E[c_4] = E[y_{21} + y_{22}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_B$

SEQ is 25% (Chen & Tsong, 2007). This recommendation may be followed for any analyses.

A third hypothesis compares the two sequences with respect to the responses for the first period only. The prevailing strategy is to use this test if a significant carryover effect is identified. This test procedure is referred to as PAR. When data from the second period are ignored, an AB/BA crossover design has the same structure as a PARALLEL group trial. PAR addresses the hypothesis of equality of direct treatment effects of A and B in the presence of unequal carryover effects (Freeman, 1986; Willan & Pater, 1986). An unbiased estimator of the treatment effect can be found by means of a t-test applied to the measurements obtained in the first period only; however, unfortunately, when using data from the first period only, advantages of the crossover design are negated.

Balaam's Design

To solve the first-order crossover problem inherent in the traditional AB|BA design, an extension of the Grizzle design is needed. One alternative involves the use of additional treatment sequences in the two periods. For example, AA|AB|BA|BB (Balaam, 1968) could be utilized. This design is

universally optimal for estimating treatment effects regardless of whether baseline observations are available, and it is far more efficient than the classic AB|BA (Laska, Meisner & Kushner, 1983). However, in the absence of any carryover effect, this design is inefficient because many of the subjects will contribute little - if any - information to the estimate of treatment differences in the AA and BB sequences.

The schematic for this design is shown in Table 2. In sequence AB, the contrast,  $c_1 = (y_{11} - y_{12})$ , has an expected value of  $E[c_1] = E[y_{11} - y_{12}] = (\pi_1 - \pi_2) + (\tau_A - \tau_B) - \lambda_A$ , in sequence BA, the contrast  $c_2 = (y_{21} - y_{22})$ , has expected value of  $E[c_2] = E[y_{21} - y_{22}] = (\pi_1 - \pi_2) - (\tau_A - \tau_B) - \lambda_B$ , in sequence AA, the contrast  $c_3$  has an expected value of  $E[c_3] = E[y_{31} - y_{32}] = (\pi_1 - \pi_2) - \lambda_A$ , and in sequence BB, the contrast  $c_4$  has an expected value of  $E[c_4] = E[y_{41} - y_{42}] = (\pi_1 - \pi_2) - \lambda_B$ . A linear combination of  $c_1 - c_2 - c_3 + c_4$  yields an unbiased estimate of the treatment differences.

To derive an unbiased estimate of carryover effects,  $c_5$  is defined in sequence AB as  $y_{11} + y_{21}$ . The expected value of  $c_5$  is then  $E[c_5] = E[y_{11} + y_{21}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_A$ . In sequence BA,  $c_6$  is defined as  $y_{21} + y_{22}$

Table 2: Balaam's Design (AB|BA|AA|BB)

AB BA Design	Period 1 (k = 1)	Period 2 (k = 2)
Sequence AB (i = 1)	$\mu + \pi_1 + \tau_A$	$\mu + \pi_2 + \tau_B + \lambda_A$
Sequence BA (i = 2)	$\mu + \pi_1 + \tau_B$	$\mu + \pi_2 + \tau_A + \lambda_B$
Sequence AA (i = 3)	$\mu + \pi_1 + \tau_A$	$\mu + \pi_2 + \tau_A + \lambda_A$
Sequence BB (i = 4)	$\mu + \pi_1 + \tau_B$	$\mu + \pi_2 + \tau_B + \lambda_B$

Table 2 Notes:

- Sequence AB (i = 1):  $E(y_{AB,1}) = \mu_{AB,1} = \mu + \pi_1 + \tau_A$ ,  $E(y_{AB,2}) = \mu_{AB,2} = \mu + \pi_2 + \tau_B + \lambda_A$
- Sequence BA (i = 2):  $E(y_{BA,1}) = \mu_{BA,1} = \mu + \pi_1 + \tau_B$ ,  $E(y_{BA,2}) = \mu_{BA,2} = \mu + \pi_2 + \tau_A + \lambda_B$
- Sequence AA (i = 3):  $E(y_{AA,1}) = \mu_{AA,1} = \mu + \pi_1 + \tau_A$ ,  $E(y_{AA,2}) = \mu_{AA,2} = \mu + \pi_2 + \tau_A + \lambda_A$
- Sequence BB (i = 4):  $E(y_{BB,1}) = \mu_{BB,1} = \mu + \pi_1 + \tau_B$ ,  $E(y_{BB,2}) = \mu_{BB,2} = \mu + \pi_2 + \tau_B + \lambda_B$
- In sequence AB, the contrast  $c_1$  has the expected value of  $E[c_1] = E[y_{11} - y_{21}] = (\pi_1 - \pi_2) + (\tau_A - \tau_B) - \lambda_A$
- In sequence BA, the contrast  $c_2$  has the expected value of  $E[c_2] = E[y_{21} - y_{22}] = (\pi_1 - \pi_2) - (\tau_A - \tau_B) - \lambda_B$
- In sequence AA, the contrast  $c_3$  has the expected value of  $E[c_3] = E[y_{31} - y_{32}] = (\pi_1 - \pi_2) - \lambda_A$
- In sequence BB, the contrast  $c_4$  has the expected value of  $E[c_4] = E[y_{41} - y_{42}] = (\pi_1 - \pi_2) - \lambda_B$
- In sequence AB, the contrast  $c_5$  has the expected value of  $E[c_5] = E[y_{11} + y_{21}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_A$
- In sequence BA, the contrast  $c_6$  has the expected value of  $E[c_6] = E[y_{21} + y_{22}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_B$

and has the expected value of  $E[c_6] = E[y_{21} + y_{22}] = 2\mu + (\pi_1 + \pi_2) + (\tau_A + \tau_B) + \lambda_B$ . A linear combination of  $[\frac{1}{2}(c_5 - c_6 - c_3 + c_4)]$  yields an unbiased estimate of carryover effects  $(\lambda_A - \lambda_B)$ .

Two-Treatment, Three-Period Crossover Design

The second design strategy is to extend the AB|BA design by adding a third period and repeating one of the two treatments: This has several attractive advantages. For example, in clinical studies major costs are associated with planning and patient recruitment rather than routine follow-up, thus, an added period may not imply a large additional cost. The added treatment period will allow carryover effects to be estimated and compared with the within-subject variability. Finally, a careful selection of the treatment sequences to be used will insure that the first two trial periods constitute a conventional two-period crossover trial if the third treatment period leads to excessive subject drop-outs.

In three period crossover trial with two treatments, six possible treatment sequences can result when two treatments are applied in three

periods. Two of these sequences, AAB and BBA can be omitted because they do not enable carryover effects from A and B to be examined in the same subject and the first two periods do not constitute a conventional two-period crossover design.

The four remaining sequences ABB, BAA, ABA and BAB may be used in pairs to form two-treatment sequence three-period designs, three-treatment sequence three-period designs and one four-treatment sequence three-period design. Of the two-treatment sequence, three-period, the ABB|BAA is known to be the universally optimal design within the class of three periods (Cheng & Wu, 1980; Laska & Meisner, 1985; Hedayat & Stufken, 2003). In these designs half the subjects are randomly assigned to each sequence.

Two additional efficient two-treatment, three-period designs are the AAB|ABA and ABA|ABB designs. Another efficient two-treatment, three-period design is the ABB|BAA|ABA|BAB (Ebbutt, 1984). This set of designs with equal number of subjects per sequence is able to estimate all parameters in the

EXTENSION OF GRIZZLE'S CLASSIC CROSSOVER DESIGN

traditional model and provide a good estimate of the treatment contrast (Ebbutt, 1984; Heydat & Stufken, 2003, Liang & Carriere, 2010).

A balanced model for the two-treatment three-period crossover trial, ABB| BAA, is shown in Table 3. In sequence ABB, the contrast,  $c_1 = (2y_{11} - y_{21} - y_{31})$ , has the expectation  $\frac{1}{4}\{(2\pi_1 - \pi_2 - \pi_3) + 2(\tau_A - \tau_B) - \lambda_A - \lambda_B\}$ . In sequence BAA, the contrast  $c_2 = (2y_{21} - y_{22} - y_{32})$  has the expectation  $\frac{1}{4}\{(2\pi_1 - \pi_2 - \pi_3) + 2(\tau_B - \tau_A) - \lambda_A - \lambda_B\}$ . The difference between contrast  $c_1$  and  $c_2$  forms an unbiased estimator of  $\tau_A - \tau_B$ . It appears that the central problem of the AB|BA has been solved by simply extending the design by one period. An unbiased estimator of any carryover effect,  $\lambda_A - \lambda_B$  may also be constructed. Consider  $c_3 = (y_{11} - 2y_{21} + y_{31})$  and  $c_4 = (y_{21} - 2y_{22} + y_{23})$ . The expected value of  $E [c_3]$  is  $\frac{1}{3} \{(\pi_1 - 2\pi_2 + \pi_3) + (\tau_A - \tau_B) - 2\lambda_A + \lambda_B\}$  and the expected value of  $E [c_4]$  is  $\frac{1}{3} \{(\pi_1 - 2\pi_2 + \pi_3) + (\tau_B - \tau_A) - 2\lambda_B + \lambda_A\}$ . The difference

between  $c_3$  and  $c_4$  forms an unbiased estimate of  $\lambda_A - \lambda_B$ .

A second model for a two-treatment three-period crossover trial, ABA| BAB, is shown in Table 4. In sequence ABA, the expected value of  $E [c_1] = E [\frac{1}{2} (2y_{11} - y_{21} - y_{31})] = \frac{1}{2}\{(2\pi_1 - \pi_2 - \pi_3) + (\tau_A - \tau_B) - \lambda_A - \lambda_B\}$ . In sequence BAB, the expected value of  $E [c_2] = E [\frac{1}{2} (2y_{12} - y_{22} - y_{32})] = \frac{1}{2}\{(2\pi_1 - \pi_2 - \pi_3) - (\tau_A - \tau_B) - \lambda_A - \lambda_B\}$ . The difference between the means of the two contrasts  $c_1$  and  $c_2$  forms an unbiased estimator of  $\tau_A - \tau_B$ . In testing for carryover effect, let  $c_3 = y_{11} + 2y_{21} + y_{31}$  in sequence ABA and the expected value of  $E [c_3] = E [(y_{11} + 2y_{21} + y_{31})] = \{4\mu + (\pi_1 + 2\pi_2 + \pi_3) + 2(\tau_A + \tau_B) + 2\lambda_A + \lambda_B\}$ . In sequence BAB, define  $c_4 = y_{21} + 2y_{22} + y_{23}$  with the expected value of  $E [c_4] = E [(y_{21} + y_{22} - 2y_{23})] = \{4\mu + (\pi_1 + 2\pi_2 + \pi_3) + 2(\tau_A + \tau_B) + 2\lambda_B + \lambda_A\}$ . The difference between  $c_3$  and  $c_4$  then forms an unbiased estimate of  $\lambda_A - \lambda_B$ .

Table 3: ABB|BAA Design

Sequence	Period 1 (k = 1)	Period 2 (k = 2)	Period 3 (k = 3)
ABB (i = 1)	$\mu + \pi_1 + \tau_A$	$\mu + \pi_2 + \tau_B + \lambda_A$	$\mu + \pi_3 + \tau_B + \lambda_B$
BAA (i = 2)	$\mu + \pi_1 + \tau_B$	$\mu + \pi_2 + \tau_A + \lambda_B$	$\mu + \pi_3 + \tau_A + \lambda_A$

Table 3 Notes:

ABB (i = 1):  $E(y_{ABB,1}) = \mu + \pi_1 + \tau_A$ ,  $E(y_{ABB,2}) = \mu + \pi_2 + \tau_B + \lambda_A$ ,  $E(y_{ABB,3}) = \mu + \pi_3 + \tau_B + \lambda_B$

BAA (i = 2):  $E(y_{BAA,1}) = \mu + \pi_1 + \tau_B$ ,  $E(y_{BAA,2}) = \mu + \pi_2 + \tau_A + \lambda_B$ ,  $E(y_{BAA,3}) = \mu + \pi_3 + \tau_A + \lambda_A$

In sequence ABB, the expected value of  $E[c_1] = E[\frac{1}{4}(2y_{11} - y_{21} - y_{31})] = \frac{1}{4}\{(2\pi_1 - \pi_2 - \pi_3) + 2(\tau_A - \tau_B) - \lambda_A - \lambda_B\}$

In sequence BAA, the expected value of  $E[c_2] = E[\frac{1}{4}(2y_{21} - y_{22} - y_{32})] = \frac{1}{4}\{(2\pi_1 - \pi_2 - \pi_3) + 2(\tau_B - \tau_A) - \lambda_A - \lambda_B\}$

In sequence ABB, the expected value of  $E[c_3] = E[\frac{1}{3}(y_{11} - 2y_{21} + y_{31})] = \frac{1}{3}\{(\pi_1 - 2\pi_2 + \pi_3) + (\tau_A - \tau_B) - 2\lambda_A + \lambda_B\}$

In sequence BAA, the expected value of  $E[c_4] = E[\frac{1}{3}(y_{21} - 2y_{22} + y_{23})] = \frac{1}{3}\{(\pi_1 - 2\pi_2 + \pi_3) + (\tau_B - \tau_A) - 2\lambda_B + \lambda_A\}$

Table 4: ABA|BAB

Sequence	Period 1 (k = 1)	Period 2 (k = 2)	Period 3 (k = 3)
ABA (i = 1)	$\mu + \pi_1 + \tau_A$	$\mu + \pi_2 + \tau_B + \lambda_A$	$\mu + \pi_3 + \tau_A + \lambda_B$
BAB (i = 2)	$\mu + \pi_1 + \tau_B$	$\mu + \pi_2 + \tau_A + \lambda_B$	$\mu + \pi_3 + \tau_B + \lambda_A$

Table 4 Notes:

ABA (i = 1):  $E(y_{ABB,1}) = \mu + \pi_1 + \tau_A$ ,  $E(y_{ABA,2}) = \mu + \pi_2 + \tau_B + \lambda_A$ ,  $E(y_{ABA,3}) = \mu + \pi_3 + \tau_A + \lambda_B$   
 BAB (i = 2):  $E(y_{BAB,1}) = \mu + \pi_1 + \tau_B$ ,  $E(y_{BAB,2}) = \mu + \pi_2 + \tau_A + \lambda_B$ ,  $E(y_{BAB,3}) = \mu + \pi_3 + \tau_B + \lambda_A$   
 In sequence ABA,  $E[c_1] = E[\frac{1}{2}(2y_{11} - y_{21} - y_{31})] = \frac{1}{2}\{(2\pi_1 - \pi_2 - \pi_3) + (\tau_A - \tau_B) - \lambda_A - \lambda_B\}$   
 In sequence BAB,  $E[c_2] = E[\frac{1}{2}(2y_{21} - y_{22} - y_{32})] = \frac{1}{2}\{(2\pi_1 - \pi_2 - \pi_3) - (\tau_A - \tau_B) - \lambda_A - \lambda_B\}$   
 In sequence ABA,  $E[c_3] = E[(y_{11} + 2y_{21} + y_{31})] = \{4\mu + (\pi_1 + 2\pi_2 + \pi_3) + 2(\tau_A + \tau_B) + 2\lambda_A + \lambda_B\}$   
 In sequence BAB,  $E[c_4] = E[(y_{21} + 2y_{22} + y_{32})] = \{4\mu + (\pi_1 + 2\pi_2 + \pi_3) + 2(\tau_A + \tau_B) + 2\lambda_B + \lambda_A\}$

### Discussion

Advantages of a crossover trial are that each subject is used as their own control, optimal two treatment three-period crossover designs are statistically efficient and these designs require fewer subjects for the same number of observations than do non-crossover designs. This latter advantage is an important aspect, particularly in situations where the experimental subjects are scarce and are expensive to recruit and maintain in the study. Another advantage of crossover designs is that, by a defining a specific choice of treatment sequences, it is possible to estimate important treatment contrasts even when assuming a carryover effect in the overall model.

The major concern in a crossover design is the presence of carryover effects. In any given period, an observation from a subject is affected not only by the direct effect of a treatment in the period in which it is applied, but possibly by the effect of a treatment applied in the preceding period. In a clinical study, particularly a drug study, one way to avoid the impact of a carryover effect is to insert a rest period between two successive periods with the hope that the carryover effect would wash out during this period. This is the most common method of handling effects of drug studies. The insertion of rest periods effectively increases the interval

between the observed periods and may help in overcoming the carryover effect if the carryover effect is not expected to persist, however, inserting rest periods may not be feasible. The insertion of a rest period between each pair of successive periods increases the total duration of the experiment and there is no guarantee that the wash out period is sufficiently long enough to eliminate any carryover effect. An alternative is to design the experiment in such a manner that the difference in treatment effects may be estimated after adjusting for the presence of possible carryover effects.

Despite some of the problems associated with the use of a crossover design its advantages are attractive. Although crossover designs have been in use for several decades, issues relating to the finding optimal crossover designs have been addressed only in about the last 30 years. There has been a continuous effort in the general area of optimal crossover designs, often assuming different underlying models. The uniform consistency has been the inclusion of carryover effect. These models, in turn, may be regarded as an approximation to the real world relationship between the response and the effects included in the model. A caution worth noting: Any crossover design under an assumed model might not be the optimal if the model is incorrectly specified.

## EXTENSION OF GRIZZLE'S CLASSIC CROSSOVER DESIGN

### Conclusion

Although there are crossover models that specify higher order carryover effects, the two-treatment three-period designs described herein maintain their optimality characteristics. To address the potential of first-order carryover effect, the classic AB|BA crossover design could be extended to a three-period design using one of the designs outlined. In effect, the added treatment period permits any carryover effects to be estimated and compared with the within-subject variability. A careful selection of the treatment sequences would reduce to a classic two-treatment, two-period conventional crossover trial if the third treatment period leads to excessive subject drop-outs.

The statistical properties of two-treatment, three-period designs is well known but seldom used. When the traditional statistical model is acknowledged as being reasonable, these designs provide a framework to estimate treatment effects even in the presence of a carryover effect and effectively provide a way to address the impasse imposed by the classic AB|BA design.

### References

Balaan, L. N. (1968). A two-period design with  $t^2$  experimental units. *Biometrics*, *24*, 61-73.

Chen, L., & Tsong, Y. (2007). Design and analysis for drug abuse potential studies: Issues and strategies for implementing a crossover design. *Drug Information Journal*, *41*, 481-89.

Ebbutt, A. F. (1984). Three-period crossover designs for two treatments. *Biometrics*, *40*, 219-24.

Freeman, P. R. (1989). The performance of the two-stage analysis of two-treatment, two-period cross-over trials. *Statistics in Medicine*, *8*, 1421-1432.

Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, *21*, 467-480.

Hedayat, A. S., & Stufken, J. (2003). Optimal and efficient crossover designs under different assumptions about the carryover effects. *Journal of Biopharmaceutical Statistics*, *13*, 519-28.

Laska, E., Meisner, M., & Kushner, H. B. (1983). Optimal crossover designs in the presence of carryover effects. *Biometrics*, *39*, 1087-91.

Liang, Y., & Carriere, K. C. (2010). On the role of baseline measurements for crossover designs under the self and mixed carryover effects model. *Biometrics*, *66*, 140-148.

Senn, S. J. (1994). The AB/BA crossover design, past, present and future? *Statistical Methods in Medical Research*, *3*, 303-324.

Senn, S. J. (1996). The AB/BA crossover: how to perform the two-stage analysis if you can't be persuaded that that you shouldn't. In M. Hanses, Ed., *ed. Liber Amicorum Roel van Strik*, 93-100. Rotterdam, Netherlands: Erasmus University. Available at <http://www.ucl.ac.uk/~ucaksjs/roel.pdf>.

Willan, A. R., & Pater J. L. (1986). Carryover and the two-period crossover clinical trial. *Biometrics*, *42*, 593-599.

## Is Next Twelve Months Period Tumor Recurrence Free Under Restricted Rate Due to Medication? A Probabilistic Warning

Ramalingam Shanmugam  
Texas State University,  
San Marcos, TX USA

A methodology is formulated to analyze tumor recurrence data when its incidence rate is restricted due to medication. Analytic results are derived to make a probabilistic early warning of tumor recurrence free period of length  $\tau$ ; that is, the chance for a safe period of length  $\tau$  is estimated. The captured data are length biased. Expressions are developed to extract and relate to counterparts of the non-length biased data. Three data sets are considered as illustrations: (1) patients who are given a placebo, (2) patients who are given the medicine pyridoxine and (3) patients who are given the medicine thiotepea.

Key words: Targeted versus captured recurrence trend, order statistics, survival function.

### Introduction

Cancer is the second major cause of death after cardiovascular deaths in USA. Tumor, an abnormal growing of cells in the brain, is an important category in the cancer group and remains mysterious to medical researchers. Whether a tumor is a benign or malignant type, its recurrence time must be speculated for an efficient treatment and an early warning of its recurrence time is crucial. The early warning provides a basis for the decision of whether to continue the same medicine or change to another medicine. The medical community relies heavily on an answer to the question: what do the data suggest? The statistical community is obligated to devise an appropriate statistical methodology to analyze a patient's data to make a probabilistic early warning regarding whether a time period of length  $\tau$  will be tumor recurrence free.

The current literature does not contain information to answer these questions. To compensate for the gaps in this area, this study was undertaken to introduce a new statistical methodology to fulfill the need for making an early probabilistic warning of tumor recurrence during a tumor free period of length  $\tau$ .

Let  $T > 0$  represent a random time in which a tumor (malignant or benign) reoccurs in a patient. Given the collected data on his/her recurrence times, could a future period of length  $\tau$  be tumor recurrence free? Suppose that the uncertainty in  $T$  is governed by non-observable incidence parameters  $\theta < \beta$ , where  $\theta$  and  $\beta$  portray the tumor recurrence rate and its restriction level respectively. A restriction on the incidence rate exists due to the effect of the given medication. Let  $v > 0$  be the threshold time parameter connecting  $T$ ,  $\theta$  and  $\beta$ . To capture their intricacies, consider the probability density function:

$$f_{tumor}^{targeted}(t|\beta, \theta, v) = \left(\frac{\beta}{\theta} - 1\right) \phi^{\frac{\beta}{\theta}-1} / (t + \phi - v)^{\frac{\beta}{\theta}}; \\ t > v; 0 < \theta < \beta \quad (1)$$

for  $T > v > 0$  in terms of  $\theta$  and  $\beta$ . The expected and variance time of the probability pattern in (1) are, respectively,

Ramalingam Shanmugam is a Professor in the School of Health Administration. His research interests include multivariate data modeling, informatics, optimal decision support systems, and health/medical application of statistics. Email him at: rs25@txstate.edu.

$$\begin{aligned} \mu_{tumor}^{t \text{ arg eted}}(t|\beta, \theta, \nu) &= \\ &= E(t|\beta, \theta, \nu) \quad (2) \\ &= \frac{\phi}{\frac{\beta}{\theta} - 2} + \nu \end{aligned}$$

and

$$\begin{aligned} \sigma_{tumor}^{2, t \text{ arg eted}}(t|\beta, \theta, \nu) &= \\ &= Var(t|\beta, \theta, \nu) \quad (3) \\ &= \left(\frac{\phi + \mu_{tumor}^{t \text{ arg eted}}}{\phi - \mu_{tumor}^{t \text{ arg eted}}}\right) (\mu_{tumor}^{t \text{ arg eted}})^2 \end{aligned}$$

A recurrence is missed if the observation period is too short; a longer observation period increases the chance of recording another tumor recurrence. This concept is length biased sampling (see Zelen & Feinlieb, 1969 for details on length bias phenomenon). The length bias alters the statistical assessment of the recurrence trend. A caution is necessary regarding how the length-biased data are analyzed and interpreted. When the equal random sampling is replaced by proportional length biased sampling, the length-biased version of the model in (1) is appropriate for the recurrent tumor times as follows:

$$\begin{aligned} f_{tumor}^{captured}(t|\beta, \theta, \nu) &= \\ &= \frac{t f_{tumor}^{t \text{ arg eted}}(t|\beta, \theta, \nu)}{\mu_{tumor}^{t \text{ arg eted}}(t|\beta, \theta, \nu)} t \\ &= \frac{t \left(\frac{\beta}{\theta} - 1\right) \phi^{\frac{\beta}{\theta} - 1}}{\left[\frac{\phi}{\left(\frac{\beta}{\theta} - 2\right)} + \nu\right] (t + \phi - \nu)^{\frac{\beta}{\theta}}}; \quad (4) \\ & \quad t > \nu; 0 < \theta < \beta \end{aligned}$$

where  $\mu_{tumor}^{t \text{ arg eted}}$  denotes the actual average tumor recurrence time. The probability density functions in (1) and (4) are interrelated and their relations are used to forecast tumor recurrence in

this study by employing a methodology to analyze tumor recurrence data when its rate is restricted due to medication. Using the analytic results, an early warning of the next tumor recurrence free period of length  $\tau$  is made. Various pertinent information in the length-biased data are extracted and compared to their counterparts under non-biased data.

### Results

To identify whether a tumor recurrence free period of length  $\tau$  exists at the earliest time possible, understanding its patterns is prerequisite. Since the last major tumor recurrence, the probability that a next recurrence will happen within a selected time  $\tau > 0$  is

$$\begin{aligned} F_{tumor}^{captured}(\tau) &= \\ &= Pr_{tumor}^{captured}(t \leq \tau) = 1 - S_{tumor}^{captured}(\tau) \\ &= 1 - \left(\frac{\phi}{\tau + \phi - \nu}\right)^{\frac{\beta}{\theta} - 1} \left[1 + \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)\right] \\ &= F_{tumor}^{int ended}(\tau - \nu) \\ & \quad - \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right) [1 - F_{tumor}^{int ended}(\tau - \nu)]. \quad (5) \end{aligned}$$

The actual chance for tumor recurrence occurs within  $\tau$  units of time and requires an adjustment of the captured chance for tumor recurrence time. This is revealed in the relationship shown in (5); the adjustment is

$$F_{tumor}^{int ended}(\tau - \nu) = \frac{F_{tumor}^{captured}(\tau - \nu) + \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)}{1 + \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)}.$$

In an instantaneous time (that is,  $\tau \rightarrow \nu$ ), no adjustment is necessary, but after some time has elapsed, a finite amount of adjustment is necessary. Both a location shift and a scale change are warranted. The location and scale shifts are

$$\left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)$$

and

$$1 + \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)$$

respectively.

The captured survival chance without a tumor recurrence, that is

$$S_{tumor}^{captured}(\tau) = \Pr_{captured}[t > \tau]$$

with the length-biased data is

$$1 + \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)$$

times the actual survival chance, which is

$$S_{tumor}^{intended}(\tau) = \Pr_{intended}[t > \tau]$$

in a non-length biased data collection scenario without a tumor recurrence before that selected time  $\tau$ . That is,

$$S_{tumo}^{intended}(\tau) = \left[ \frac{\theta\phi}{\theta\phi + (\beta - \theta)(\tau - \nu)} \right] S_{tumor}^{captured}(\tau) \quad (6)$$

The cofactor

$$\left[ \frac{\theta\phi}{\theta\phi + (\beta - \theta)(\tau - \nu)} \right]$$

is the impact amount due to sampling bias. The captured survivability chance without a tumor recurrence before that selected time  $\tau$  is an over-estimate. The mean and variance time of the captured tumor recurrence patterns are expressed in terms of the actual average time  $\mu_{tumor}^{intended}$  between two tumor recurrences. They are

$$\begin{aligned} \mu_{tumor}^{captured} &= \\ &= \int_0^{\infty} t f_{tumor}^{captured}(t) dt \\ &= 2\phi\mu_{tumor}^{intended} / (\phi - \mu_{tumor}^{intended}) \end{aligned} \quad (7)$$

and

$$\begin{aligned} \sigma_{tumor}^{2,captured} &= \\ &= Var_{captured}(t) \\ &= \int_0^{\infty} (t - \mu_{tumor}^{captured})^2 f_{tumor}^{captured}(t) dt \\ &= \frac{2\phi^2 \mu_{tumor}^{2,intended}}{(\phi - \mu_{tumor}^{intended})(\phi - 2\mu_{tumor}^{intended})} \\ &\quad \text{if } \phi > 2\mu_{tumor}^{intended}. \end{aligned} \quad (8)$$

Expression (7) suggests a relation between the actual average tumor recurrence time  $\mu_{tumor}^{intended}$  and the captured average tumor recurrence time  $\mu_{tumor}^{captured}$ . Equivalently,

$$\mu_{tumor}^{intended} = \frac{\phi\mu_{tumor}^{captured}}{(2\phi + \mu_{tumor}^{captured})}.$$

The actual average tumor recurrence time is not linear with the captured recurrence time. Interestingly, according to expression (8), the captured variance of the recurrence times widens if the actual average tumor recurrence time is greater. Equivalently,

$$\begin{aligned} \sigma_{tumor}^{2,intended} &= \\ &= \frac{1}{3} \left(1 - \frac{2\mu_{tumor}^{intended}}{\phi}\right) \\ &\quad \left[ \sigma_{tumor}^{2,captured} + \frac{4\phi^2 \mu_{tumor}^{2,intended}}{(\phi - \mu_{tumor}^{intended})^2} \right] \\ &\quad - \mu_{tumor}^{2,intended} \end{aligned}$$

meaning that the actual tumor occurrence time variance increases proportionally with the captured variance.

Based on this, the incidence parameters of the tumor recurrence probability pattern in (4) are estimable with  $t_{(1)} = \min(t_1, t_2, \dots, t_n)$ ,  $S_{tumor}^{2,captured}$  and  $\bar{t}_{tumor}^{captured} - t_{(1)}$  denoting the captured threshold recurrence time, sample



TWELVE MONTH TUMOR RECURRENCE: A PROBABILISTIC WARNING

variance and mean of the tumor recurrence times respectively. The moment estimates are

$$\hat{\nu} = t_{(1)} \tag{9}$$

$$\hat{\phi} = \frac{[\bar{t}_{tumor}^{captured} - t_{(1)}][s_{tumor}^{2,captured} + (\bar{t}_{tumor}^{captured} - t_{(1)})^2]}{|s_{tumor}^{2,captured} - 2(\bar{t}_{tumor}^{captured} - t_{(1)})^2|}, \tag{10}$$

and

$$\hat{\beta} = \left\{ 2 + \frac{3[\bar{t}_{tumor}^{captured} - t_{(1)}]^2}{|s_{tumor}^{2,captured} - 2(\bar{t}_{tumor}^{captured} - t_{(1)})^2|} \right\} \theta \tag{11}$$

Thus, with estimates for (9), (10) and (11), the survival probability in (6) is estimable. When  $1 - \hat{S}_{tumor}^{captured}(\tau)$  is significantly large, it suggests that early warning is necessary to speculate whether the tumor recurrence is likely to happen within the selected time  $\tau$ . In addition, the factor

$$Restriction = \left\{ 1 + \frac{3[\bar{t}_{tumor}^{captured} - t_{(1)}]^2}{|s_{tumor}^{2,captured} - 2(\bar{t}_{tumor}^{captured} - t_{(1)})^2|} \right\} \tag{12}$$

signifies the restriction level on tumor recurrence due to medication.

Often the likelihood of noticing a quick tumor recurrence if it occurs repeatedly is of interest. This can be accomplished using the order statistics concept (see David, 2005 for details regarding order statistics definitions and tools). Suppose that tumor reoccurrence takes place  $n$  times for a patient and let the order statistics be

$$\nu \leq T_{(1)} < T_{(2)} < \dots < T_{(n)} < \infty;$$

then

$$Pr^{arg\ eted}[T_{(1)} > \tau] = [1 - F(\tau)]^n$$

describes the chance of surviving a period of length  $\tau$  without a tumor recurrence. Interestingly, it is observed that

$$\begin{aligned} Pr^{arg\ eted}[T_{(1)} > \tau] &= \\ &= \left[ \left( \frac{\phi}{\phi + \tau - \nu} \right)^{\frac{\beta}{\theta} - 1} \right]^n \\ &= \left[ \frac{\phi + (\frac{\beta}{\theta} - 2)\nu}{\phi + (\frac{\beta}{\theta} - 2)\nu + \tau - \nu} \right]^n Pr^{captured}[T_{(1)} > \tau]. \end{aligned} \tag{13}$$

The factor

$$\left[ \frac{\phi + (\frac{\beta}{\theta} - 2)\nu}{\phi + (\frac{\beta}{\theta} - 2)\nu + \tau - \nu} \right], \tag{14}$$

$\beta$

in (13) with a single incidence (that is,  $n = 1$ ), signifies the over-estimated proportion of the chance of having a minimal safe period  $\tau$ . This proportion of over-estimate diminishes as the patient experiences more recurrences (that is, as  $n$  increases).

Examples

To illustrate results, tumor data from Andrews and Herzberg (1990) were selected, these data show placebo, pyridoxine drug and thiotepa drug groups. The mean and variance of tumor recurrence times are calculated for each patient along with the captured survivability for each patient using (5) and the cofactor using (6). (See Tables 1A, 2A and 3A.) For example, the patient with ID# 13 in the placebo group has 0.71 captured chance of survivability without tumor recurrence in 12 months and the actual survivability chance is 0.03 times the captured chance of surviving without tumor recurrence.

Using expressions (9) through (11), the incidence parameters are estimated and are displayed in Tables 1B, 2B and 3B. For example, the parameter estimates  $\hat{\phi}$ ,

$\hat{\nu}$  and the restriction gap  $(.)$  in  $\theta < (.)\beta$  on the incidence rate for patient #13 in the placebo group are 0.53, 2.14 and 3 respectively; patient #15 is an anomaly. Results

presented in in Tables 1B, 2B and 3B also show the chance of a patient having a safe period without any tumor recurrence for the next twelve months is calculated using (13), and estimates for how much the over-estimation might have been in the length biased data are calculated using (14).

The restriction level in (12), location shift

$$\left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)$$

and scale shift

$$1 + \left(\frac{\beta}{\theta} - 1\right) \left(\frac{\tau - \nu}{\phi}\right)$$

due to length-biased sampling are calculated and displayed in Tables 1 B through 3 B. For example, patient #13 had 4.58, 19.2 and 20.2 as his/her restriction level for the incidence parameters, location and scale shifts due to length-biased sampling. Patient # 13 is in the placebo group and has 0.03 chance of having a safe next twelve months without any tumor recurrence. If the length biased captured data are used for this patient, the chance of having safe next twelve months without any tumor recurrence would have been over estimated by an amount 0.09.

Figures 1, 2 and 3 illustrate the pattern of actual survival chance in terms of the data captured for placebo, pyridoxine drug and thiotepa drug groups. Note that one outlier case (patient #15) is present in the placebo group. In addition, it is notable that the trend for the pyridoxine group is a reversed direction compared to the other two groups; patients in the pyridoxine drug group display an upward curve while the placebo group and the thiotepa drug group have downward curves.

Conclusion

If other predictor variables are available, a logistic regression can be built for patients. In addition, the more and less important predictors can be identified based on the estimate of survivability without a tumor recurrence in the next 12 months.

References

Andrews, D., & Herzberg, A. (1990). *Data: A collection of problems from many fields for the student and research worker*, Berlin: Springer Verlag.  
 David, H. (2005). *Order Statistics*. New York, NY: John Wiley and Sons.  
 Zelen, M., & Feinlieb, M. (1969). On the theory of screening for chronic diseases. *Biometrics*, 56. 601-614.

Table 1A: Recurrence Times (in Months) of Tumor Placebo Patients

ID	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	$t_{(1)}$	$\bar{t}$	$s_t^2$	Captured Survivability in 12 Months	Cofactor
13	3	13	7							3	7.7	25	0.71	0.03
15	7	3	6	8						3	6	4.67	0.03	0.15
16	3	12	10							3	8.33	22.3	0.65	0.05
24	28	2								2	15	338	0.98	0.003
34	16	3	4	5	6	5	6			3	6.42	18.95	0.66	0.04
44	3	12	31	5	2					3	10.6	145.3	0.96	0.005
48	2	6	4	1	4	4	12	16	15	1	7,2	34.44	0.81	0.02

TWELVE MONTH TUMOR RECURRENCE: A PROBABILISTIC WARNING

Table 1B: Parameter Estimates of Tumor Placebo Patients

ID	$\hat{\phi}$	$\theta < (.)\beta$	$\hat{\nu}$	Restriction Level	Location Shift	Scale Shift	Actual Chance of Safe 12 Months	Overestimated Proportion in the Chance of Safe 12 Months in the Length Biased Data
13	0.53	(2.140)	3	4.58	19.2	20.2	0.03	0.09
15	84.46	(39.38)	3	3.02	4.09	5.09	0.02	0.95
16	0.84	(2.23)	3	3.46	13.11	14.11	0.04	0.14
24	0.03	(2.00)	2	Infinity	270.69	271.69	0.003	0.004
34	0.58	(2.18)	3	8.73	18.16	19.16	0.03	0.11
44	0.06	(2.00)	3	85.69	154.9	155.9	0.006	0.008
48	0.29	(2.07)	1	3.64	40.3	41.35	0.02	0.03

Table 2A: Recurrence Times (in Months) of Tumor Patients Treated with Pyridoxine Drug

ID	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	$t_{(1)}$	$\bar{t}$	$s_t^2$	Captured Survivability in 12 Months	Cofactor
51	3	1								1	2	2	0.007	0.06
64	3	7	12	4	8					3	6.8	12.7	0.46	0.07
67	3	4	5	4	3	9	6	2	3	2	4.33	4.5	0.08	0.10
70	2	4	4	6	7	4	9	3	3	2	4,67	5	0.10	0.10

Table 2B: Parameter Estimates of Tumor Patients Treated with Pyridoxine Drug

ID	$\hat{\phi}$	$\theta < (.)\beta$	$t_{(1)}$	Restriction Level	Location Shift	Scale Shift	Actual Chance of Safe 12 Months	Overestimated Proportion in the Chance of Safe 12 Months in the Length Biased Data
51	6	8	1	Infinity	12.83	13.83	0.0006	0.52
64	1.52	2.52	3	3.67	9.03	10.03	0.05	0.25
67	5.87	5.27	2	3.55	7.29	8.29	0.01	0.55
70	5.38	4.81	2	3.31	7.08	8.08	0.01	0.52

Table 3A: Recurrence Times (in Months) of Tumor Patients Treated with Thiotepa Drug

ID	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	$t_{(1)}$	$\bar{t}$	$s_t^2$	Captured Survivability in 12 Months	Cofactor
89	18	0			0	13.3	40.5	0.97	0.003
104	24	2	3	11	2	10	103.3	0.94	0.007

Table 3B: Parameter Estimates of Tumor Patients Treated with Thiotepa Drug

ID	$\hat{\phi}$	$\theta < (.)\beta$	$t_{(1)}$	Restriction Level	Location Shift	Scale Shift	Actual Chance of Safe 12 Months	Overestimated Proportion in the Chance of Safe 12 Months in the Length Biased Data
89	0.04	2.00	0	infinity	288	289	0.003	0.003
104	0.09	2.01	2	8.78	105.5	106.5	0.008	0.01

Figure 1: Actual Survival Chance in Terms of Captured Survival Chance for Placebo Group

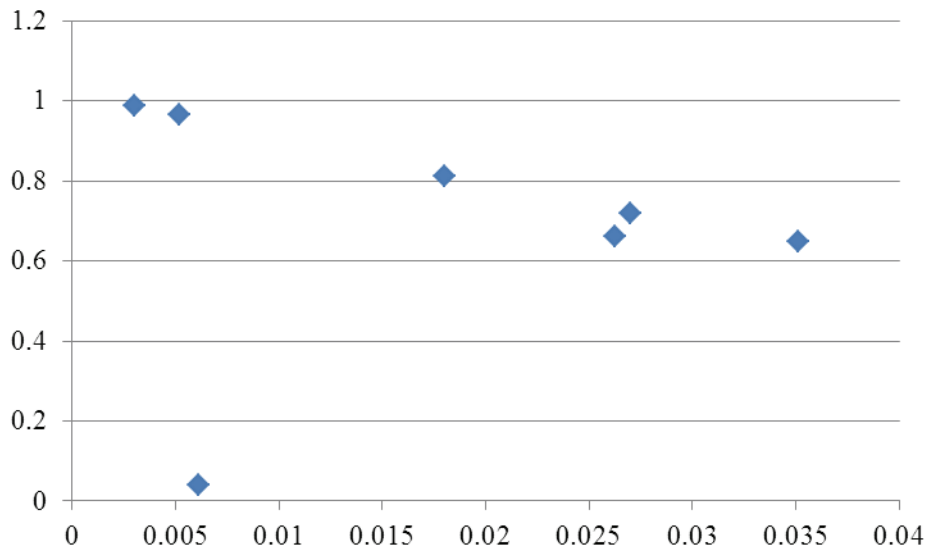
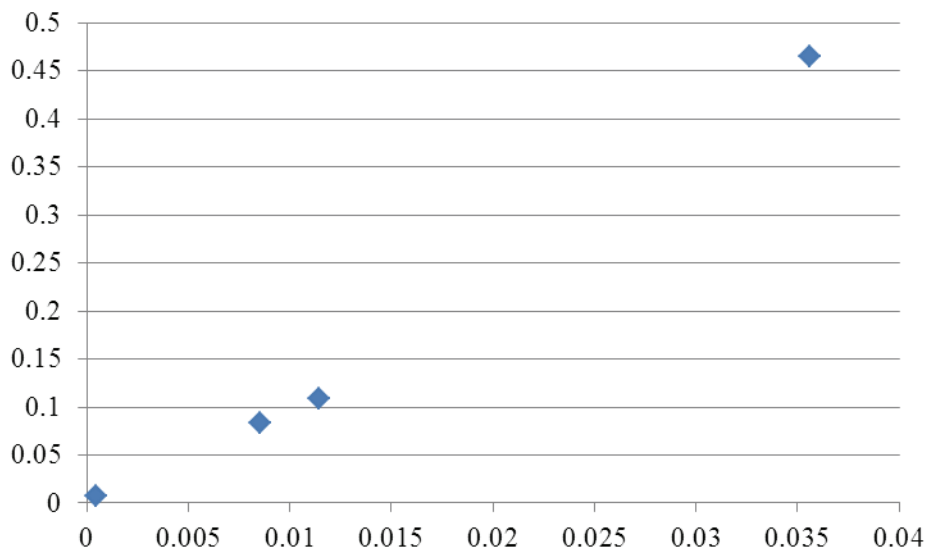
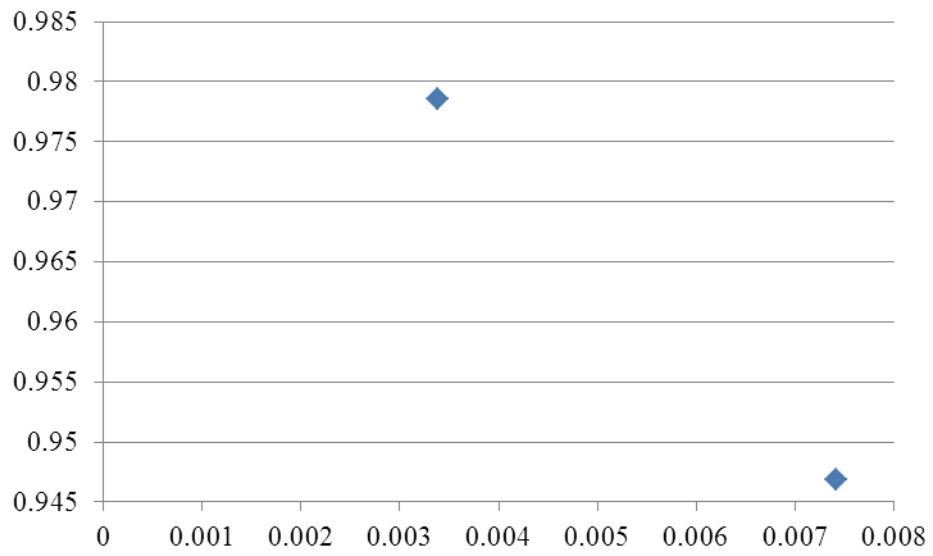


Figure 2: Actual Survival Chance in Terms of Captured Survival Chance for Pyridoxine Drug Group



# TWELVE MONTH TUMOR RECURRENCE: A PROBABILISTIC WARNING

Figure 3: Actual Survival Chance in Terms of Captured Survival Chance for Thiotepa Drug Group



## Logistic Regression Models for Higher Order Transition Probabilities of Markov Chain for Analyzing the Occurrences of Daily Rainfall Data

Narayan Chanra Sinha  
Ministry of Finance,  
Dhaka, Bangladesh

M. Ataharul Islam  
University of Dhaka,  
Dhaka, Bangladesh

Kazi Saleh Ahamed  
Foundation for Research on Educational  
Planning and Development,  
Dhaka, Bangladesh

---

Logistic regression models for transition probabilities of higher order Markov models are developed for the sequence of chain dependent repeated observations. To identify the significance of these models and their parameters a test procedure for a likelihood ratio criterion is developed. A method of model selection is suggested on the basis of AIC and BIC procedures. The proposed models and test procedures are applied to analyze the occurrences of daily rainfall data for selected stations in Bangladesh. Based on results from these models, the transition probabilities of first order Markov model for temperature and humidity provided the most suitable option to model forecasts for daily rainfall occurrences for five selected stations in Bangladesh.

Key words: Logistic regression, transition probabilities, Markov chain, ML estimation, LR test, AIC, BIC, daily rainfall occurrences data.

---

### Introduction

A Markov chain model is constructed for describing transition probabilities for time or chain dependent process under change or random process. A logistic regression model is used as probabilistic model for analyzing covariate dependent binary data. The logistic regression model may define covariate dependent transition probabilities of a Markov chain. Muenz and Rubinstein (1985) made an attempt to develop covariate dependent first order transition probabilities for Markov chain models. In their model two-health states, distress and no distress, recorded as binary responses 1

and 0 respectively were incorporated; they showed that healthy patients feel less distress than others at the time of biopsy as time proceeds.

To identify the pattern of daily rainfall occurrences Gabriel and Neumann (1962) developed a Markov chain model for Tel Aviv data. They showed that dry and wet spells follow a geometric distribution. For the same data, Green (1964) fitted the probability models better than that of Gabriel and Neumann's models assuming that dry and wet spells follow an exponential distribution. Parthasarathy and Dhar (1974) identified the negative trend for south Asian daily rainfall occurrences using regional rainfall over India for the period 1901 to 1960. Similar studies analyzing daily rainfall data have been conducted by Islam (1980), Stern (1980a), Stern (1980b), Stern, et al. (1982), Stern and Coe (1984), Sinha and Paul (1992), Sinha and Islam (1994), Shimizu and Kayano (1994), Sinha, et al. (2006), Sinha, et al. (2009) and others. However, these did not develop covariate dependent probability models for analyzing the patterns of daily rainfall occurrences. To identify the patterns and forecasting models for the occurrence or non-occurrence of rainfall, different types of covariate dependent transition

---

Narayan Chanra Sinha is a Financial Analyst in the Monitoring Cell, Finance Division of the Bangladesh Secretariat, Dhaka-1000, Bangladesh. Email: ncsinha2002@yahoo.com. M. Ataharul Islam is a Professor in the Department of Statistics. Email: mataharul@yahoo.com. Kazi Saleh Ahamed is President of the FREPD. Emails: ksahmed@bangla.net and frepd@dhaka.net.

# LOGISTIC REGRESSION MODELS FOR PROBABILITIES OF MARKOV CHAINS

probabilities of Markov chain models need to be developed for logistic regression.

## Transition Probabilities of Markov Chain for Logistic Regression Models

To develop logistic regression models for higher order transition probabilities (t. p.) of Markov chains, consider the chain dependent repeated observations  $x_1, x_2, \dots, x_n$  at time  $t$  ( $t = 1, 2, \dots, T$ ),  $X_n(t)$ . Here the assumption that observations occurring depend on different covariates,  $Z_n(t)$  is made. The first order transition count,  $n_{jk}(t)$  denotes the number of individuals in state  $j$  at time  $t-1$  and in state  $k$  at  $t$ . If the second order transition count,  $n_{ijk}(t)$  denotes the number of individuals in state  $i$  at time  $t-2$ , in state  $j$  at  $t-1$  and in state  $k$  at  $t$ , then the first and second order transition probabilities of the Markov chain are denoted by  $p_{jk}(t)$  and  $p_{ijk}(t)$  respectively, for all  $i, j, k = 1, 2, \dots, m$  and  $t = 1, 2, \dots, T$ .

For stationarity, these probabilities are denoted by  $p_{jk}$  and  $p_{ijk}$ , respectively. Similarly, higher order stationary or non-stationary transition probabilities of Markov chain  $p_{ij\dots krl}$  or  $p_{ij\dots krl}(t)$ , respectively, may be defined for transition count  $n_{ij\dots krl}$  or  $n_{ij\dots krl}(t)$ . The term  $p_{ij\dots krl}$  or  $p_{ij\dots krl}(t)$  indicates the transition probability of state  $l$  at time  $t$ , given the state  $r$  at time  $t-1, \dots$ , the state  $j$  at time  $t-s+1$  and state  $i$  at time  $t-s$ , where  $t = s, s+1, \dots, T$ , and for all  $i, j, \dots, k, r, l = 0, 1$ . The ML estimate (Anderson & Goodman, 1957; Muenz & Rubinstein, 1985; Sinha, et al., 2006; Sinha, et al., 2009) of higher order stationary or non-stationary transition probabilities for the transition probability matrices are

$$\hat{p}_{ij\dots krl} = \frac{n_{ij\dots krl}}{n_{ij\dots k}} = \frac{n_{ij\dots krl}(t)}{n_{ij\dots k}(t-1)}$$

where

$$n_{ij\dots k} = \sum_{l=1}^m n_{ij\dots krl}$$

To develop the covariate dependent two-state transition probabilities of the Markov chain, consider the parameters  $p_{01}$  and  $p_{11}$  for first order, and  $p_{001}, p_{101}, p_{011}$  and  $p_{111}$  for a second order Markov chain. Here  $p_{01}, p_{11}, p_{001},$

$p_{101}$ ,  $p_{011}$  and  $p_{111}$  specify the transition probabilities of  $0 \rightarrow 1, 1 \rightarrow 1, 0 \rightarrow 0 \rightarrow 1, 1 \rightarrow 0 \rightarrow 1, 0 \rightarrow 1 \rightarrow 1$  and  $1 \rightarrow 1 \rightarrow 1$  transitions respectively. Similarly,  $2^r$  parameters may be defined for  $r^{\text{th}}$  order two-state transition probabilities of the Markov chain. To formulate such a Markov chain, the following assumptions are made: (i) each observation of chain dependent process depends on different covariates; (ii) observations of the chain dependent process follow a logistic form; (iii) the counts  $n_j(0)$  and  $n_{jk}(1)$  are non-random; and (iv) each row of transition in the probability matrix is independent.

To estimate the covariate dependent transition probabilities for the Markov chain, consider logistic regression models for first, second and higher order transition probabilities (Muenze & Rubinstein, 1985) which are defined as

$$P_{ij\dots krl} = \frac{\exp(z_{ij\dots krl}(q, t))}{1 + \exp(z_{ij\dots krl}(q, t))}, \quad (2.1)$$

where

$$Z_{ij\dots krl}(q, t) = \alpha_{ij\dots krl} + \sum_{h=1}^n \beta_h(ij\dots krl) Z_h(ij\dots krl)q(t-r),$$

for all  $i, j, \dots, k, r = 0, 1, h = 1, 2, \dots, n, q = 1, 2, \dots, Q, t = 1, 2, \dots$ , where  $T$  and  $r$  are the order of the Markov chain. Here,  $Z_h(ij\dots krl)q(t-r)$  is the  $h^{\text{th}}$  covariate for  $i \rightarrow j \rightarrow \dots \rightarrow k \rightarrow r \rightarrow l(=1)^{\text{th}}$  state for  $(t-r)^{\text{th}}$  day of  $q^{\text{th}}$  year,  $\beta_h(ij\dots krl)$  is the parameter of  $h^{\text{th}}$  covariate for  $i \rightarrow j \rightarrow \dots \rightarrow k \rightarrow r \rightarrow l(=1)^{\text{th}}$  state and  $\alpha_{ij\dots krl}$  is the intercept term. Further for saturated model the term  $Z_{ij\dots krl}(q, t)$  for (2.1) can be defined as

$$\begin{aligned} Z_{ij\dots krl}(q, t) = & \alpha_{ij\dots krl} + \sum_{h=1}^n \beta_h(ij\dots krl) Z_h(ij\dots krl)q(t-r) \\ & + \sum_{(h<g)=1}^n \beta_{hg}(ij\dots krl) Z_h(ij\dots krl)q(t-r) Z_g(ij\dots krl)q(t-r) \\ & + \text{higher order interaction effect} \end{aligned} \quad (2.2)$$

where  $\beta_{h(ij \dots kr1)}$  is the main effect for  $Z_{h(ij \dots kr1)q(t-r)}$  covariate and  $\beta_{hg(ij \dots kr1)}$  is the interaction effect for the  $Z_{h(ij \dots kr1)q(t-r)}$  and  $Z_{g(ij \dots kr1)q(t-r)}$  covariates.

Estimation of Parameters for Covariate Dependent Transition Probabilities of Markov Chain Model

To identify the effect of different covariates for the changes of transition probabilities of Markov chain model the parameters of the models 2.1 and 2.2 are to be estimated. To estimate the parameters for transition probabilities of Markov chain model, Anderson and Goodman (1957), Muenz and Rubinstein (1985), Sinha, et al. (2006) and Sinha, et al. (2009) suggested the ML estimation method. Thus the method of MLE is used to estimate the parameters of model 2.1. The log likelihood function (Formula 2.3) is shown in Figure 1. To obtain the estimated value of parameters by ML estimation method under Newton-Raphson iteration procedure, the information matrix and information vector are denoted by I and U respectively, where  $I^{-1}$  is the variance covariance matrix with respect to parameters. Similarly, the parameters of model (2.2) may be estimated.

Test of Hypothesis

To test the significance of the parameters and models for logistic regression models for transition probabilities of a Markov chain, Wald (1943) suggested test statistic W as

consistent and asymptotically equivalent to the likelihood ratio test under the null hypothesis. This test statistic provides a significant result for the iterative nature of maximum likelihood estimate than that of likelihood ratio test. However, Rao (1965), Hauck and Donner (1977) and Jennings (1986) found that the test statistic W is less powerful compared to likelihood ratio test. Furthermore, for a large sample Hosmer and Lemeshow (1989) recommended the likelihood ratio test as opposed to Wald’s test, because often it fails to reject the co-efficient when it is significant. Due to these, the likelihood ratio test procedure is employed to test the significance of parameters and models for 2.1 and 2.2.

Likelihood Ratio Test

To identify the significance of the covariate dependent transition probabilities of Markov chain models and their parameters; consider hypotheses 1 and 2 for model 2.1, and 3 and 4 for model 2.2.

Hypotheses 1, Model 2.1:

$$H_0: \beta_{1(ij\dots r1)} = \beta_{2(ij\dots r1)} = \dots = \beta_{p(ij\dots r1)} = \beta_{(p+1)(ij\dots r1)} = \dots = \beta_{h(ij\dots r1)} = 0$$

vs.

$$H_1: \beta_{1(ij\dots r1)} = \beta_{2(ij\dots r1)} = \dots = \beta_{p(ij\dots r1)} = \beta_{(p+1)(ij\dots r1)} = \dots = \beta_{h(ij\dots r1)} \neq 0$$

Figure 1: Log Likelihood Function for the ML Estimation Method (Formula 2.3)

$$\log L = \sum_{q=1}^Q \sum_{t=1}^T n_{ij..r1}(q,t) \left( \alpha_{ij..r1} + \sum_{h=1}^n \beta_{h(ij..r1)} Z_{h(ij..r1)q(t-r)} \right) - \sum_{q=1}^Q \sum_{t=1}^T (n_{ij..r0}(q,t) + n_{ij..r1}(q,t)) \log \left( 1 + \exp \left( \alpha_{ij..r1} + \sum_{h=1}^n \beta_{h(ij..r1)} Z_{h(ij..r1)q(t-r)} \right) \right), \tag{2.3}$$

$n_{ij..r0}(q,t)$  and  $n_{ij..r1}(q,t)$  are the transition counts for the  $i, j, \dots, r, l(=0)^{th}$  state and  $i, j, \dots, r, l(=1)^{th}$  state respectively for the  $t^{th}$  day of the  $q^{th}$  year, where  $q = 1, 2, \dots, Q$  and  $t = 1, 2, \dots, T$



# LOGISTIC REGRESSION MODELS FOR PROBABILITIES OF MARKOV CHAINS

Hypotheses 2, Model 2.1:

$$H_0: \beta_{p(ij\dots r1)} = 0 \text{ and } H_0: \alpha_{ij\dots r1} = 0$$

vs.

$$H_1: \beta_{p(ij\dots r1)} \neq 0 \text{ and } H_1: \alpha_{ij\dots r1} \neq 0$$

Hypotheses 3, Model 2.2:

$$H_0: \beta_{1(ij\dots r1)} = \beta_{2(ij\dots r1)} = \dots = \beta_{h(ij\dots r1)} = \beta_{hg(ij\dots r1)} = \dots = 0$$

$h < g$

vs.

$$H_1: \beta_{1(ij\dots r1)} = \beta_{2(ij\dots r1)} = \dots = \beta_{h(ij\dots r1)} = \beta_{hg(ij\dots r1)} = \dots \neq 0$$

$h < g$

Hypotheses 4, Model 2.2:

$$H_0: \beta_{pk(ij\dots r1)} = 0$$

$$p < k$$

vs.

$$H_1: \beta_{pk(ij\dots r1)} \neq 0,$$

$$p < k$$

where  $h = 1, 2, \dots, p, p+1, \dots, n, g = 1, 2, \dots, k, k+1, \dots, n$ ,  $\beta_{h(ij\dots r1)}$  is the parameter of the  $h^{\text{th}}$  covariates for  $i, j, \dots, r, l(=1)$  transition and  $\beta_{hg(ij\dots r1)}$  is the interaction effect between  $h^{\text{th}}$  and  $g^{\text{th}}$  covariates. The likelihood ratio test statistic  $(-2\log\lambda_{ij\dots r1})$  is asymptotically distributed as  $\chi^2_{ij\dots r1}$  (Kendall & Stuart, 1973) with  $h, 1, (h+(h-1)/2) +$  number of higher order interaction effect) and 1 degree of freedom for the null hypotheses 1, 2, 3 and 4 respectively, where  $\lambda_{ij\dots r1}$  is the likelihood ratio for  $i, j, \dots, r, l(=1)^{\text{th}}$  (for all  $i, j, \dots, r, s = 0, 1$ ) transitions of the Markov chain. For the overall transition probabilities of the Markov chain this test statistic is defined as

$$\chi^2 = \sum_{ij\dots r1} \chi^2_{ij\dots r1} \quad (2.4)$$

for all  $i, j, \dots, r = 0, 1$  with  $2^h, 2^r, (h+(h-1)/2) +$  number of higher order interaction effect) and  $2^r$  degrees of freedom for null hypotheses 1, 2, 3 and 4 respectively, where  $r$  is the order of the Markov chain.

## Methods of Model Selection

To identify the best model among the significant models, several authors including McCullagh & Nelder (1983) and Agresti (1984) suggested various model selection procedures and they also identified some limitations and drawbacks. For example, these selection procedures sometimes provide almost equal emphasis for several possible models; often procedures do not provide the best model among the models sufficiently for a true alternative hypothesis. For overcoming these problems, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) procedures are employed for the selection of appropriate covariate dependent transition probabilities for the Markov model (Sakamoto, 1991; Shimizu, 1993).

Akaike (1972b) developed AIC by the utilization of a likelihood ratio criterion under the extension form of maximum likelihood. Akaike (1970) defined AIC on the basis of final prediction error (FPE) as the mean square prediction error of a predictor to identify the autoregressive model. Schwarz (1978) developed BIC as a more consistent and optimal procedure than the AIC. Sakamoto (1991) used a minimum of AIC (MAICE) and a minimum of BIC (MBICE) to identify the optimal explanatory variable for the model. For covariate dependent transition probabilities of Markov chain models, to develop a model selection procedure, the MAICE and MBICE are employed by utilizing the likelihood function and the ML estimate of parameters. For a large  $n$ , Bayes estimators are asymptotically equivalent to ML estimators (Kendall & Stuart, 1973) and the procedures are defined as

$$AIC(i) < AIC(i+1) < \dots < AIC(i+s), \quad (3.1)$$

and

$$BIC(i) < BIC(i+1) < \dots < BIC(i+s), \quad (3.2)$$

where  $i = 1, 2, \dots, \infty$ ,  $s = 0, 1, 2, \dots, \infty$ ,  $(i+s)$  indicate the number of models,  $AIC = -2(\text{maximum log likelihood}) + 2(\text{number of estimable parameters in the model})$  and  $BIC = -2(\text{maximum log likelihood}) + 2(\text{number of estimable parameters in the model}) \times \log n$ . The terms  $AIC(i)$  and  $BIC(i)$  indicate the best model among models  $AIC(i+1), \dots, AIC(i+s)$  and  $BIC(i+1), \dots, BIC(i+s)$  respectively.

#### Data

To identify the utility of the proposed models, the daily rainfall occurrence data during the rainy season for the period 1964-1990 for five selected stations, namely Chittagong, Mymensingh, Rajshahi, Faridpur and Satkhira of Bangladesh, were utilized. These data are collected by the Department of Meteorology, Government of People's Republic of Bangladesh. The period between the months of April and October is considered as the rainy season. The major agricultural crops (Aus and Aman rice) under the traditional system of this country, Bangladesh, are produced during this period and depend greatly on the occurrences of rainfall due to the shortage of sufficient irrigation facilities.

#### Logistic Regression Models for Transition Probabilities of Markov Chain for the Occurrence of Rainfall

A comprehensive idea regarding the probability of rainfall is essential in view of economic implications for crop production. The probabilities for the occurrences of rainfall are used in agricultural planning purposes, such as land-use, choice of crops and cropping system. Several researchers (Virmani, 1975; 1982; Dale, et al., 1981; Davy, et al., 1976) analyzed the occurrences of rainfall to identify the determinant of rainfall occurrences. They found that the occurrences of rainfall depend mainly on different climatic factors, such as temperature and humidity. Further, Shimizu (1993) developed a bivariate mixed lognormal distribution for assessing the probability of rainfall by using the Automated Meteorological Data Acquisition System (AMeDAS) data set of Japan.

In order to develop covariate dependent transition probabilities of a Markov chain model

for assessing and analyzing the occurrences of rainfall for the five selected areas of Bangladesh, consider probability models 2.1 and 2.2. The climatic variables temperature and humidity (Virmani, 1975, 1982) are employed to perform these models. For these variables, the term  $Z_{ij\dots kr1}(q,t)$  for model 2.1 may be defined as:

$$Z_{1(ij\dots kr1)}(q,t) = \alpha_{1(ij\dots kr1)} + \beta_{1(ij\dots kr1)} X_{1(ij\dots kr1)q(t-r)} \quad (5.1)$$

$$Z_{2(ij\dots kr1)}(q,t) = \alpha_{2(ij\dots kr1)} + \beta_{2(ij\dots kr1)} X_{2(ij\dots kr1)q(t-r)} \quad (5.2)$$

$$Z_{3(ij\dots kr1)}(q,t) = \alpha_{3(ij\dots kr1)} + \beta_{1(ij\dots kr1)} X_{1(ij\dots kr1)q(t-r)} + \beta_{2(ij\dots kr1)} X_{2(ij\dots kr1)q(t-r)} \quad (5.3)$$

$$Z_{4(ij\dots kr1)}(q,t) = \alpha_{4(ij\dots kr1)} + \beta_{1(ij\dots kr1)} X_{1(ij\dots kr1)q(t-r)} + \beta_{2(ij\dots kr1)} X_{2(ij\dots kr1)q(t-r)} + \beta_{12(ij\dots kr1)} X_{1(ij\dots kr1)q(t-r)} X_{2(ij\dots kr1)q(t-r)} \quad (5.4)$$

for all  $i, j, \dots, k, r = 0, 1$ ,  $q = 1, 2, \dots, Q$ ,  $t = 1, 2, \dots, T$  and  $r$  is the order of Markov model. Here  $i, j, \dots, k, r$  represent the transitions of the Markov model and  $q$  and  $t$  indicate the number of year and the number of days in the year respectively. The term  $(q, t)$  represents the  $t^{\text{th}}$  day of the  $q^{\text{th}}$  year. The variables  $X_{1(ij\dots kr1)q(t-r)}$  and  $X_{2(ij\dots kr1)q(t-r)}$  indicate the maximum temperature and average humidity respectively of the  $(t-r)^{\text{th}}$  day for the  $q^{\text{th}}$  year for  $(i, j, \dots, k, r, 1)^{\text{th}}$  transitions. The terms  $\beta_{1(ij\dots kr1)}$  and  $\beta_{2(ij\dots kr1)}$  indicate the effects of temperature and humidity respectively,  $\beta_{12(ij\dots kr1)}$  indicates the interaction effect between temperature and humidity on the

## LOGISTIC REGRESSION MODELS FOR PROBABILITIES OF MARKOV CHAINS

occurrences of rainfall and the terms 1 and 0 indicate wet and dry days respectively.

To test the significance of probability models 5.1-5.4 and their estimated parameters, the likelihood ratio test statistic is utilized. For performing this test statistic, the following four null hypotheses are considered.

a. For models (5.1) and (5.2):

$$\begin{aligned} H_0: \beta_{1(j1)} = 0 \text{ and} \\ H_0: \beta_{1(ij1)} = 0 \\ \text{vs.} \\ H_1: \beta_{1(j1)} \neq 0 \text{ and} \\ H_1: \beta_{1(ij1)} \neq 0. \end{aligned}$$

b. For model (5.3):

$$\begin{aligned} H_0: \beta_{1(j1)} = \beta_{2(j1)} = 0 \text{ and} \\ H_0: \beta_{1(ij1)} = \beta_{2(ij1)} = 0 \\ \text{vs.} \\ H_1: \beta_{1(j1)} = \beta_{2(j1)} \neq 0 \text{ and} \\ H_1: \beta_{1(ij1)} = \beta_{2(ij1)} \neq 0. \end{aligned}$$

c. For model (5.4):

$$\begin{aligned} H_0: \beta_{1(j1)} = \beta_{2(j1)} = \beta_{12(j1)} = 0 \text{ and} \\ H_0: \beta_{1(ij1)} = \beta_{2(ij1)} = \beta_{12(ij1)} = 0 \\ \text{vs.} \\ H_1: \beta_{1(j1)} = \beta_{2(j1)} = \beta_{12(j1)} \neq 0 \text{ and} \\ H_1: \beta_{1(ij1)} = \beta_{2(ij1)} = \beta_{12(ij1)} \neq 0. \end{aligned}$$

d. For models (5.1-5.4):

$$\begin{aligned} H_0: \alpha_{j1} = 0 \text{ and} \\ H_0: \alpha_{ij1} = 0 \\ \text{vs.} \\ H_1: \alpha_{j1} \neq 0 \text{ and} \\ H_1: \alpha_{ij1} \neq 0. \end{aligned}$$

e. For models (5.3) and (5.4) respectively:

$$\begin{aligned} H_0: \beta_{m(j1)} = 0 \text{ and} \\ H_0: \beta_{m(ij1)} = 0 \\ \text{vs.} \\ H_1: \beta_{m(j1)} \neq 0 \text{ and} \\ H_1: \beta_{m(ij1)} \neq 0, \end{aligned}$$

where  $m = 1, 2$ .

f. For model (5.3):

$$\begin{aligned} H_0: \beta_{12(j1)} = 0 \text{ and} \\ H_0: \beta_{12(ij1)} = 0 \\ \text{vs.} \\ H_1: \beta_{12(j1)} \neq 0 \text{ and} \\ H_1: \beta_{12(ij1)} \neq 0. \end{aligned}$$

To test the significance of transition probabilities for the occurrences of rainfall for first and second order Markov models 5.1-5.4, the values of  $\chi^2$  under the LR criterion for null hypotheses (a), (b) and (c) are identified. Further, to test the significance of parameters for transition probabilities of first and second order Markov models (5.1-5.4), the values of  $\chi^2$  under LR criterion for null hypotheses (d), (e) and (f) are also identified (the values of  $\chi^2$  for these null hypotheses are not shown herein, however). But based on these  $\chi^2$ -values, the significance of parameters and models are identified. To test the null hypothesis by the  $\chi^2$  statistic, it is always observed that the value of  $\chi^2$  increases as sample size increases. For overcoming this problem, although it is small, consider a p-value up to 0.001 as the cut-off point.

### Results

#### Significance of Estimated Parameters and Models

To estimate the parameters of models 5.1-5.4, consider the ML estimation method under the Newton-Raphson iteration procedure. To identify the order of the transition probabilities of a Markov chain for daily rainfall occurrences Sinha (1997) and Sinha, et al. (2009) showed that the Chittagong and Faridpur stations follow first order and the Mymensingh, Rajshahi and Satkhira stations follow second order transition probabilities of a Markov chain. To estimate the parameters of these models, consider  $t = 214$ ,  $Q = 27$  and  $r = 1$  for the Chittagong and Faridpur stations and  $r = 1$  and  $2$  for the remaining three stations. For transition probabilities of daily rainfall occurrences for first order Markov models 5.1-5.4 for the five selected stations and second order Markov models 5.1-5.4 for the Mymensingh, Rajshahi and Satkhira stations of Bangladesh, the

estimated values of parameters and their significance are shown in Table 1.

For the first order Markov models, the effect of temperature for model 5.1 and the effect of humidity for model 5.2 on the occurrences of transition probabilities (t.p.) of rainfall are found to be positive for the five selected stations (see Table 1). The exception to this result occurs for transition type Wet/Wet for model 5.1 for all selected stations and for model 5.2 for the Chittagong and Rajshahi stations.

The effect of humidity and temperature for model 5.3, and the effect of humidity and the interaction term between temperature and humidity for model 5.4 are also positive on the occurrences of t.p. of rainfall for all the selected stations. The exceptions to this result occur for transition type Wet/Wet for the Rajshahi, Faridpur and Satkhira stations and for all transitions of Chittagong station for temperature for model 5.3. Such an exceptional result is also observed for transition type Wet/Wet for the Chittagong station for humidity and for transition types Wet/Dry for the Chittagong station and Wet/Wet for the Satkhira station for the interaction term for model 5.4.

The positive effect of temperature and humidity and their interaction effect for the occurrences of rainfall transitions indicate that the probability of the occurrence of rainfall increases with increases of these variables for two consecutive days. The result for model 5.4 implies that the effect of temperature and humidity and their interaction effect on the occurrences of rainfall are inversely related.

The effect of temperature and of humidity on the occurrences of rainfall for different types of transition probabilities of first order Markov models 5.1-5.4 are significant ( $p$ -value  $< 0.001$ ) for the five selected stations. To assess the probability of rainfall occurrences for first order Markov models, the results of  $\chi_{ij}^2$  indicate that all transitions for model 5.1 are significant at the Chittagong, Rajshahi and Satkhira stations, for model 5.2 are significant at the Chittagong and Mymensingh stations, and for models 5.3 and 5.4 are significant at all selected stations. For the overall transition probability of rainfall occurrences, the  $\chi^2$  value indicates that the first order Markov models 5.1-5.4 are significant for all selected stations.

For second order Markov models, results show in Table 1 indicate that the effect of temperature for model 5.1 and the effect of humidity for model 5.2 are positive on the occurrences of transition probabilities of rainfall for the Mymensingh, Rajshahi and Satkhira stations. This result is an exception for transition type Wet/Dry/Wet for the Mymensingh and Rajshahi stations and Wet/Wet/Wet for the Satkhira station for model 5.1. Further, the effect of temperature and humidity for model 5.3, and the effect of humidity and interaction term (temperature and humidity) for model 5.4 are observed to be positive on the occurrences of t.p. of rainfall for these three stations. However, for the Rajshahi station is an exception for transition types Wet/Dry/Wet and Wet/Wet/Dry for temperature for model 5.3. This exceptional result is also found for model 5.4 for transition type Wet/Dry/Dry at the Satkhira station for the interaction term and for transition types Wet/Dry/Wet, Wet/Wet/Dry and Wet/Wet/Wet at Rajshahi, and Wet/Wet/Wet at Satkhira for humidity. This positive effect of temperature, humidity and their interaction effect for the occurrences of rainfall transitions imply that the probability of rainfall increases with increases of these variables for three consecutive days.

For different types of second order transition probabilities of Markov models, Table 1 shows that the effect of temperature for model 5.1, the effect of humidity for model 5.2, the effect of temperature and humidity for model 5.3, and the effect of temperature and humidity and their interaction effect for model 5.4 are nonsignificant ( $p$ -value  $< 0.001$ ) on the occurrences of rainfall for the maximum number of transitions for the Mymensingh, Rajshahi and Satkhira stations.

Further, to assess the probability of rainfall occurrences for second order Markov models, the results of  $\chi_{ijk}^2$  indicate that all transitions for models 5.3-5.4 are significant for these stations. The exceptions to this result occur for transition types Wet/Dry/Wet and Wet/Wet/Dry for the Mymensingh station and Wet/Dry/Wet and Wet/Wet/Wet for the Rajshahi station for model 5.3, and transition type Wet/Dry/Wet for the Mymensingh station for model 5.4. However, for overall transition probability of rainfall occurrences, the values of

# LOGISTIC REGRESSION MODELS FOR PROBABILITIES OF MARKOV CHAINS

Table 1: Estimated Parameters for Logistic Regression for Transition Probabilities of First and Second Order Markov Models 5.1-5.4 and their Significance for Five Selected Areas of Bangladesh

Station Name	Transition Type	Model 5.2		Model 5.2		Model 5.3			Model 5.4			
		$\alpha_{1(i)}$	$\beta_{1(i)}$	$\alpha_{2(i)}$	$\beta_{2(i)}$	$\alpha_{3(i)}$	$\beta_{1(i)}$	$\beta_{2(i)}$	$\alpha_{4(i)}$	$\beta_{1(i)}$	$\beta_{2(i)}$	$\beta_{12(i)}$
Chittagong	0 → 1	-1.66321* (0.20331)	0.02460* (0.00649)	-15.80485* (0.93470)	0.18630* (0.01155)	-13.87005* (1.2856)	-0.05096* (0.01173)	0.18216* (0.01161)	-34.61039* (8.95026)	0.61502 (0.28473)	0.44442* (0.11991)	-0.00842 (0.00360)
	1 → 1	8.55298* (0.74237)	-0.24899* (0.02383)	-17.11358* (0.83400)	-0.21235* (0.00995)	-10.07113 (1.07783)	-0.07826* (0.01855)	0.15733* (0.00924)	25.11332 (7.91970)	-1.22477* (0.25361)	-0.24801 (0.08778)	0.01323* (0.00282)
Mymeningh	0 → 1	-1.80030* (0.10932)	0.02794* (0.00360)	-4.91382* (0.33400)	0.05003* (0.00421)	-5.91503* (0.41529)	0.02303* (0.00405)	0.05448* (0.00466)	-4.21334* (0.47261)	-0.04123 (0.01928)	0.03434* (0.00586)	0.00077* (0.00024)
	1 → 1	0.85874* (0.15992)	-0.00711 (0.00540)	-7.71641* (0.57352)	0.09883* (0.00678)	-7.30683* (0.61389)	0.00455* (0.00551)	0.09249* (0.00674)	0.08709* (1.15585)	-0.26779* (0.04208)	0.00564 (0.01344)	0.00321* (0.00049)
	0→0→1	-2.27199* (0.13929)	0.03130* (0.00455)	-3.00491* (0.24970)	0.02145* (0.00321)	-3.48158* (0.30151)	0.02376* (0.00473)	0.01915* (0.00365)	-3.15954* (0.31601)	-0.01040 (0.01664)	0.01471 (0.00435)	0.00045 (0.00022)
	1→0→1	-0.11962 (0.23879)	-0.00206 (0.00797)	-2.31835 (0.79070)	0.02603 (0.00957)	-2.43567 (0.88411)	0.00243 (0.00818)	0.02661 (0.00977)	-1.09384 (2.88316)	-0.02267 (0.09513)	0.01102* (0.03384)	0.00028 (0.00112)
	0→1→1	0.34499 (0.22895)	0.00071 (0.00753)	-2.04344* (0.65770)	0.02968 (0.00805)	-2.11570 (0.72134)	0.00195 (0.00774)	0.02987 (0.00813)	-0.57187 (0.96734)	-0.02340 (0.03652)	0.01103 (0.01183)	0.00031 (0.00045)
	1→1→1	0.70411* (0.19879)	0.00401 (0.00680)	-1.30477 (0.67893)	0.02447 (0.00782)	-1.56163 (0.73602)	0.00643 (0.00682)	0.02534 (0.00792)	-0.51836 (1.28677)	-0.04736 (0.04874)	0.01326 (0.01487)	0.00063 (0.00057)
Rajshahi	0 → 1	-2.61113* (0.12574)	0.05089* (0.00392)	-9.73000* (0.47914)	0.11039* (0.00591)	-10.82544* (0.54733)	0.03616* (0.00539)	0.11045* (0.01958)	-5.35333* (0.59776)	-0.19978* (0.02632)	0.03782* (0.00838)	0.00307* (0.00035)
	1 → 1	1.43630* (0.26954)	-0.03344* (0.00860)	-12.59192* (0.78959)	-0.15132 (0.00917)	-9.44802* (0.82519)	-0.03370 (0.00971)	0.12696* (0.00274)	0.07798 (0.60304)	-0.50250* (0.03493)	0.00351 (0.00770)	0.00586* (0.00013)
	0→0→1	-2.91907* (0.14711)	0.05491* (0.00457)	-5.11202* (0.33424)	0.05136* (0.00436)	-5.80392* (0.39331)	0.04121* (0.00521)	0.04480* (0.00478)	-4.10119* (0.41056)	-0.04681 (0.01853)	0.01762 (0.00647)	0.00130* (0.00027)
	1→0→1	-0.13410 (0.34177)	-0.01329 (0.01076)	-2.97782* (0.88895)	0.02952 (0.01073)	-2.50378 (0.91911)	-0.01697 (0.01123)	0.030175 (0.01055)	-0.36481 (1.27667)	-0.14755 (0.05058)	-0.01027 (0.01720)	0.00176 (0.00067)
	0→1→1	0.01713 (0.29250)	0.00408 (0.00120)	-3.04452* (0.75410)	0.03864 (0.00908)	-2.96257* (0.78487)	-0.00344 (0.00964)	0.03896* (0.00909)	0.24531 (0.93122)	-0.15739* (0.03921)	-0.00762 (0.01285)	0.00210 (0.00016)
	1→1→1	0.31405 (0.29781)	0.00934 (0.00974)	-0.90133 (0.64385)	0.01694 (0.00726)	-1.04218 (0.69783)	0.00606 (0.01003)	0.01647 (0.00743)	1.15412 (0.81518)	-0.12338* (0.04037)	-0.01350 (0.00971)	0.00163* (0.00048)
Faridpur	0 → 1	-1.49605* (0.13717)	0.02027* (0.00433)	-7.23936* (0.45168)	0.08018 (0.00554)	-7.58175* (0.55765)	0.00712 (0.0064)	0.08165* (0.00578)	-1.94471* (0.20908)	-0.34008* (0.02180)	0.00022 (0.00492)	0.00465* (0.00031)
	1 → 1	1.21035* (0.23697)	-0.02112 (0.00771)	-12.9266* (0.76910)	0.15741* (0.00897)	-0.74976* (0.28198)	-0.15087* (0.01329)	0.07049* (0.00490)	0.07996 (0.26813)	-0.45381* (0.02845)	0.00639 (0.00515)	0.00529* (0.00038)
Satkhira	0 → 1	-2.99935* (0.15545)	0.05913* (0.00471)	-7.26723* (0.40541)	0.08084 (0.00510)	-7.63173* (0.48543)	0.01052 (0.00698)	0.08114* (0.00164)	-6.16671* (0.71258)	-0.04065 (0.02520)	0.06227* (0.01010)	0.00066 (0.00035)
	1 → 1	4.06126* (0.56772)	-0.10761* (0.01774)	-11.13848* (0.66308)	0.14027* (0.00789)	-9.66700* (0.83070)	-0.01306* (0.01140)	0.12768* (0.00790)	-13.75171* (4.44781)	0.11648 (0.13820)	0.17769* (0.05397)	-0.00159 (0.00534)
	0→0→1	-3.26893* (0.17891)	0.05882* (0.00540)	-4.47600* (0.27989)	0.04262 (0.00369)	-4.64660* (0.30980)	0.01328 (0.00677)	0.03926* (0.00419)	-4.92094* (0.42540)	0.02978 (0.01857)	0.04603* (0.00689)	-0.00031 (0.00028)
	1→0→1	-1.22253 (0.51918)	0.02539 (0.01589)	-3.32119* (0.80828)	0.03614 (0.00994)	-5.25520* (1.21875)	0.04439 (0.02007)	0.04233* (0.01051)	-2.45716 (4.97238)	-0.04177 (0.15117)	0.00693 (0.06278)	0.00109 (0.00193)
	0→1→1	0.13733 (0.43185)	0.00460 (0.00135)	-2.13771* (0.71428)	0.03007 (0.00881)	-2.46701 (0.91527)	0.00832 (0.01394)	0.03080* (0.00901)	-0.53421 (1.12239)	-0.07461 (0.04103)	0.00277 (0.01571)	0.00116 (0.00058)
	1→1→1	1.44444* (0.46885)	-0.01870 (0.00476)	-3.85287* (0.82413)	0.05447 (0.00952)	-4.38068* (1.09066)	0.01034 (0.01396)	0.05688* (0.01007)	4.64829 (5.81100)	-0.25704 (0.18482)	-0.04568 (0.06665)	0.00304 (0.00213)

Notes: The figure in parentheses indicates the standard deviation of estimated parameters. The transitions 0→1 and 1→1 indicates the transition of the type dry to wet and wet to wet respectively. The transitions 0→0→1, 1→0→1, 0→1→1 and 1→1→1 indicate the transition of the type dry to dry to wet, wet to dry to wet, dry to wet to wet and wet to wet to wet respectively. \*p < 0.001.

Table 2: Values of AIC and BIC for First and Second Order Transition Probabilities of Markov Models 5.1-5.4 for Five Selected Stations of Bangladesh

Station Name	Test Criteria	Order of Markov Models	Transition Types	Model 5.1	Model 5.2	Model 5.3	Model 5.4	
Chittagong	AIC	First Order	0→1 1→1	3496.83 3286.95	3085.17 2979.59	3067.95 2936.83	3064.85 2916.97	
	BIC	First Order	0→1 1→1	3498.67 3289.56	3087.01 2982.20	3070.71 2940.74	3068.53 2922.19	
Mymensingh	AIC	First Order	0→1 1→1	3700.92 3205.04	3511.57 2998.41	3477.76 2984.48	3464.93 2957.16	
		Second Order	0→0→1 1→0→1	2326.60 1178.27	2321.67 1172.80	2295.10 1173.51	2291.80 1175.33	
			0→1→1 1→1→1	1157.12 2027.04	1161.34 2017.16	1154.29 2018.28	1146.28 2018.41	
			BIC	First Order	0→1 1→1	3702.78 3210.70	3513.43 3000.85	3480.55 2988.14
	Second Order			0→0→1 1→0→1	2327.94 1179.45	2323.01 1173.98	2297.11 1175.28	2294.48 1177.69
		0→1→1 1→1→1	1158.52 2029.16	1162.74 2019.28	1156.45 2021.46	1149.08 2022.65		
AIC		First Order	0→1 1→1	3728.45 2839.46	3229.80 2529.84	3176.66 2528.50	3114.17 2384.42	
		Second Order	0→0→1 1→0→1	2539.36 1114.18	2498.56 1107.21	2419.86 1106.97	2399.60 1102.27	
	0→1→1 1→1→1		1173.59 1661.40	1152.07 1656.43	1153.94 1658.07	1140.69 1643.82		
	BIC		First Order	0→1 1→1	3730.31 2841.66	3231.66 2531.96	3179.45 2531.80	3117.89 2388.82
Second Order			0→0→1 1→0→1	2540.82 1115.16	2500.02 1108.19	2422.05 1108.44	2402.52 1104.23	
		0→1→1 1→1→1	1174.89 1663.22	1153.37 1658.25	1155.89 1660.80	1143.29 1647.46		
		AIC	First Order	0→1 1→1	3838.85 3365.90	3493.74 3147.79	3494.47 3115.20	3322.31 2944.74
	BIC		First Order	0→1 1→1	3840.79 3368.34	3495.68 3150.23	3497.38 3118.86	3326.19 2949.62
Satkhira	AIC	First Order	0→1 1→1	3449.47 2829.18	3044.67 2525.72	3044.27 2508.95	3044.16 2510.00	
		Second Order	0→0→1 1→0→1	2312.46 1027.00	2214.23 1016.03	2212.15 1011.64	2213.98 1013.33	
			0→1→1 1→1→1	1051.62 1796.37	1039.12 1764.78	1038.84 1766.25	1037.48 1768.22	
			BIC	First Order	0→1 1→1	3451.23 2831.52	3046.43 2528.06	3046.91 2512.46
	Second Order			0→0→1 1→0→1	2313.78 1027.96	2215.55 1016.99	2214.13 1013.08	2216.62 1015.25
		0→1→1 1→1→1		1052.90 1798.41	1040.40 1766.82	1040.76 1766.29	1040.04 1772.30	

Notes: For Table 2.7 the transitions 0→1 and 1→1 indicates the transition of the type dry to wet and wet to wet respectively. The transitions 0→0→1, 1→0→1, 0→1→1 and 1→1→1 indicate the transition of the type dry to dry to wet, wet to dry to wet, dry to wet to wet and wet to wet to wet respectively.

## LOGISTIC REGRESSION MODELS FOR PROBABILITIES OF MARKOV CHAINS

$\chi^2$  indicate that the second order Markov models 5.1-5.4 are significant for all the selected stations.

### Probability Model for Forecasting Rainfall

To select a forecasting model among the models for the occurrences of rainfall, AIC and BIC criteria were utilized. The values of AIC and BIC for covariate dependent transition probabilities of Markov models 5.1-5.4 for the occurrences of rainfall are shown in Table 2.

Table 2 indicates that the values of AIC and BIC are minimum for different types of transition probabilities of rainfall occurrences for first order Markov model 5.4 for all the selected stations. However, the effect of temperature and humidity for this model are not sufficiently effective Table 2.1 in explaining all the transition probabilities of rainfall occurrences for all stations. Therefore, results lack strong grounds to select this model as an appropriate forecasting model for daily rainfall occurrences.

To identify this model, consider next minimum value to the values of model 5.4 for these criteria. Table 2 shows that the values of AIC and BIC for all transition probabilities of first order Markov model 5.3 are the next minimum values to the values of model 5.4; therefore, the transition probabilities of first order Markov model 5.3 may be considered an appropriate forecasting model for daily rainfall occurrences for all selected stations. Although the effect of temperature for transition Wet/Wet for the Rajshahi station and Wet/Dry for the Faridpur and Satkhira stations is non-effective, overall transitions this effect are significant.

For second order transition probabilities of the Markov model, Table 2 shows that the values of AIC and BIC for models 5.3 and 5.4 for Mymensingh and Satkhira stations are approximately equal and these values are observed minimum compared to values of models 5.1 and 5.2. For the Rajshahi station these values are observed minimum for model 5.4 compared to models 5.1, 5.2 and 5.3. However, the effect of temperature and humidity for model 5.4 is not significant (Table 1) for maximum number of transitions. Therefore, this model is not selected as an appropriate forecasting model for daily rainfall occurrences

for the Rajshahi station. To select this model, consider next minimum values of these criteria rather than values of model 5.4. Table 2 indicates that the values of AIC and BIC for model 5.3 provide next minimum values compared to the values of model 5.4. Therefore, the transition probabilities of second order Markov model 5.3 may be selected for forecasting the occurrences of rainfall for the Mymensingh, Rajshahi and Satkhira stations. However, Table 1 shows that the effect of temperature and humidity for the transition probabilities of rainfall for first order Markov model 5.3 are significantly effective and for second order Markov model 5.3, but these effects are not sufficiently effective. Based on this logical view, it may be concluded that the transition probabilities of first order Markov model 5.3 make it an adequate choice for forecasting the occurrences of rainfall than that of second order Markov model 5.3 for all the selected stations of Bangladesh.

### Conclusion

Logistic regression models for higher order transition probabilities of Markov chains for the sequence of chain dependent repeated observations have been developed. An assumption is made that the sequence of repeated observations can be explained by certain covariates. These models are developed as an extension of the model proposed by Muenz and Rubinstein (1985). To identify the significance of covariate dependence in transition probabilities for higher order Markov models and also to identify the significance of parameters of these models, a test procedure under likelihood ratio criterion has been developed. Further, a method of model selection procedure is suggested in this study employing AIC and BIC procedures (Sakamoto; 1991).

The proposed models and test procedures have been used to analyze the occurrences of daily rainfall data for selected stations in Bangladesh. To apply these models, two climatic variables - temperature and humidity - were considered. These applications reveal that the proposed models and test procedures can be useful to identify the forecasting models for daily rainfall occurrences. From the results of these models

and test procedures, the effects of temperature and humidity on the occurrences of rainfall can be summarized for first order the Markov model 5.3 that provides statistically significant results.

From the analysis of models 5.1-5.4, positive results were observed for the effect of temperature for model 5.1 and the effect of humidity for model 5.2 on the occurrences of rainfall for maximum number of first and second order rainfall transitions of Markov models for all the selected stations. The effects of temperature and humidity for first and second order Markov models 5.3 on the occurrences of rainfall show similar results.

The first and second order Markov models 5.4 also provide positive effects for the humidity and interaction term (temperature and humidity) on the occurrences of rainfall for maximum number of rainfall transitions for all selected stations. These positive effects indicate that the probability of rainfall is positively associated with temperature and humidity. The effect of temperature and the effect of humidity on the occurrences of rainfall for first order Markov models 5.1 and 5.2 respectively, and the effect of these covariates for model 5.3 are observed to be significant for the maximum number of transitions for all selected stations. It is also demonstrated that the method of model selection procedure provides sufficient evidence that the first order Markov model 5.3 is the most suitable among the models investigated as the forecasting model for daily rainfall occurrences for the five selected stations of Bangladesh.

#### References

- Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: John Wiley and Sons.
- Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, 22, 203-217.
- Akaike, H. (1972b). *Information theory and an extension of the maximum likelihood principal*. Proceedings of the 2<sup>nd</sup> International Symposium on Information Theory, B. N. Petrov and F. Csaki, Eds., 267-281. Akademiai Kiado, Budapest.
- Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about Markov Chains. *The Annals of Mathematical Statistics*, 28, 89-110.
- Dale, I. C., Dennett, M. D., & Stern, R. D. (1981). Daily rainfall at Gaborone, Botswana. *Tropical Agricultural Meteorology Group*, University of Reading, Report-2.
- Davy, E. G., Mattei, F., & Solomon, S. I. (1976). *An evaluation of climate and water resources for development of agriculture in the Sudano-Sahelian zone of West Africa*. Special Environmental Report. No. 9, Secretariat of the World Meteorological Organization, Geneva, Switzerland.
- Gabriel, K. R., & Neumann, J. (1962). A Markov Chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of Royal Meteorological Society*, 88, 90-95.
- Green, J. R. (1964). A model for rainfall occurrence. *Journal of the Royal Statistical Society, Series B*, 26, 345-353.
- Hauck, Jr. W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logistical analysis. *Journal of the American Statistical Association*, 72(360), 851-853.
- Hosmer, Jr. D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley and Sons.
- Islam, M. A. (1980). Probability distribution of seasonal rainfalls of Pabna and some of its applications. *Chittagong University Studies, Part II*, 4, 111-119.
- Jennings, D. E. (1986). Judging inference adequacy in logistic regression. *Journal of the American Statistical Association*, 81(394), 471-476.
- Kendall, M. G., & Stuart, A. (1973). *The advance theory of statistics*, 3<sup>rd</sup> Ed., Vol. 2. London: Griffin.
- Mc Cullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London: Chapman and Hall.
- Muenz, L. R., & Rubinstein, L. V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics*, 41, 91-101.
- Parthasarathy, B., & Dhar, O. N. (1974). Secular variations of regional rainfall over India. *Quarterly Journal of Royal Meteorological Society*, 100, 245-257.



## LOGISTIC REGRESSION MODELS FOR PROBABILITIES OF MARKOV CHAINS

- Rao, C. R. (1965). *Linear statistical inference and its applications*. New York: John Wiley and Sons.
- Sakamoto, Y. (1991). *Categorical data analysis by AIC*. Kluwer Academic Publishers.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
- Shimizu, K. (1993). A bivariate mixed lognormal distribution with an analysis of rainfall data. *Journal of Applied Meteorology*, 32(2), 161-171.
- Shimizu, K., & Kayano, K. (1994). A further study of the estimation of area rain-rate moments by the threshold method. *Journal of the Meteorological Society of Japan*, 72(6), 833-839.
- Sinha, N. C. (1997). *Probability models in forecasting rainfall with application to agriculture*. Unpublished Ph. D. Thesis, Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh.
- Sinha, N. C.; Islam, M. A., & Ahmed, K. S. (2006). Chain dependence and stationarity test for transition probabilities of Markov chain under logistic regression model. *Journal of the Korean Statistical Society*, 35(4), 355-376.
- Sinha, N. C.; Islam, M. A., & Ahmed, K. S. (2010). Order determination for the transition probabilities Markov chain based on logistic regression model. *Journal of Applied Statistical Science*, 17(4), 519-539.
- Sinha, N. C.; & Islam, S. M. S. (1994). Impact of the patterns of rainfall on aus and aman crops: An application of Markov model. *Journal of Statistical Studies*, 14, 77-85.
- Sinha, N. C., & Paul, J. C. (1992). Analysis of daily rainfall occurrence for Sylhet station: An application of Markov model. *The Bangladesh Journal of Scientific Research*, 10(1), 95-102.
- Stern, R. D. (1980a). Analysis of daily rainfall at Samaru, Nigeria using a simple two-part model. *Archiv Fur Meteorologie Geophysik und Bioklimatologie, Series B*, 28, 123-135.
- Stern, R. D. (1980b). The calculation of probability distributions for models daily precipitation. *Archiv Fur Meteorologie Geophysik und Bioklimatologie, Series B*, 28, 137-147.
- Stern, R. D., & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of Royal Statistical Society, Series A, Part-I*, 147, 1-34.
- Stern, R. D., Dennett, M. D., & Dale, I. C. (1982). Methods for analyzing daily rainfall measurements to give useful agronomic results: I direct methods. *Experimental Agriculture*, 18, 223-236.
- Stern, R. D., Dennett, M. D., & Dale, I. C. (1982). Analyzing daily rainfall measurements to give agronomically useful results, II: A modeling approach. *Methodology of Experimental Agriculture*, 18, 237-253.
- Virmani, S. M. (1975). *The agricultural climate of the Hyderabad region in relation to crop planning - A sample analysis*. In-house Publication of ICRISAT, Hyderabad, India.
- Virmani, S. M. (1982). Rainfall probability estimates for selected locations of semi-arid India. *Research Bulletin No. 1. 2<sup>nd</sup> Ed.* ICRISAT, Andhra Pradesh, India.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426-482.

## The Likelihood of Choosing the Borda-Winner With Partial Preference Rankings of the Electorate

Ömer Egecioğlu  
University of California,  
Santa Barbara, CA

Ayça Ebru Giritligil  
Istanbul Bilgi University,  
Istanbul, Turkey

---

Given that  $n$  voters report only the first  $r$  ( $1 \leq r < m$ ) ranks of their linear preference rankings over  $m$  alternatives, the likelihood of implementing Borda outcome is investigated. The information contained in the first  $r$  ranks is aggregated through a Borda-like method, namely the  $r$ -Borda rule. Monte-Carlo simulations are run to detect changes in the likelihood of  $r$ -Borda winner(s) to coincide with the original Borda winner(s) as a function of  $m$ ,  $n$  and  $r$ . The voters' preferences are generated through the Impartial Anonymous and Neutral Culture Model, where both the names of the alternatives and voters are immaterial. It is observed that, for a given  $r$ , the likelihood of choosing the Borda winner decreases down to zero independent of  $n$  as  $m$  increases within the computed range of parameter values,  $1 \leq m, n \leq 30$ . For  $n = 30$ , this likelihood is given as an approximating function of  $r$  and  $m$  through least square fit method.

Key words: Borda rule,  $r$ -Borda rule, impartial anonymous, neutral culture.

---

### Introduction

A voting rule solves the collective decision problem where voters must jointly choose one among a number of possible candidates (alternatives) on the basis of reported ordinal preferences. The choice of a voting rule has been a major ethical question since the political philosophy of the Enlightenment. When only two alternatives are at stake, the ordinary majority voting is unambiguously regarded as the best method. For three and more alternatives, plurality voting at which each voter is asked to report exactly one alternative at her/his ballot and the alternative voted the most wins, has been historically the most popular voting rule. The two celebrated critiques of plurality voting, Borda (1781) and Condorcet (1785) noted that plurality voting may elect a poor candidate, namely, one that would lose in a simple pairwise majority comparison to every other candidate, or one 'disliked' by the strict majority of voters.

Borda and Condorcet are individually devised different rules to replace plurality voting. Borda introduced a scoring method; the Borda rule that assigns points to each candidate, increasing linearly with a candidate's ranking in a voter's opinion, and elects the alternative with the highest total score. Condorcet provided the voting principle which states that if a candidate defeats every other candidate in simple majority rule, then that candidate should be the winner in the election. These two approaches have generated most of the modern scholarly research in social choice literature.

As discussed by Niemi and Riker (1976), Fishburn (1984), Nurmi (1987) and Amy (2000), no voting rule is perfect in aggregating individual preferences into social decisions in a manner compatible with the fulfillment of a variety of positive and normative criteria. However, some procedures are clearly superior to others in satisfying these criteria. Saari (1987, 1989, 1990, 2001) show that the Borda rule is less susceptible than other positional scoring rules to some unsettling possibilities and paradoxes. Some of the theoretical and probabilistic results concerning the Borda rule are summarized in Brams and Fishburn (2002) and Pattanaik (2002). However, among its shortcomings, its vulnerability to strategic

---

Ömer Egecioğlu is at the Department of Computer Science. Email: omer@cs.ucsb.edu. Ayça Ebru Giritligil is at BELIS, Murat Sertel Center for Advanced Economic Studies. Email: ayca.ebru@bilgi.edu.tr.

manipulations and the practical difficulties of implementing it remain as the most criticized properties of the Borda rule. There are many studies theoretically and/or probabilistically considering the former issue. This article focuses on the latter which has not been studied in detail.

The implementation criterion is concerned with the complexity of information that a voting procedure requires voters to reveal concerning their preferences regarding the alternatives. Unlike non-ranked single-stage voting procedures (such as plurality voting, negative plurality voting and approval voting) and non-ranked multistage voting procedures (such as plurality with a run-off and plurality with successive elimination), the Borda rule is a ranked procedure. Asking the voters to provide their complete preference rankings over the set of all available alternatives is a difficult-to-fulfill requirement due to the associated complications both on the side of voters as well as administrators to collect the information.

This study investigates the likelihood of implementing Borda outcome when voters are asked to rank only a specified number of alternatives. The situations where  $n$  voters are required to report only the first  $r$  ( $1 \leq r < m$ ) ranks of their linear (i.e., full or total) preferences over  $m$  alternatives are considered. It is assumed that the partial individual preferences are aggregated through a Borda-like method, namely the  $r$ -Borda rule. The  $r$ -Borda rule assigns strictly positive points to each alternative appearing in the first  $r$ -ranks of a voter's total preference, linearly increasing with its rank, and assigns zero points to those that are not among the top  $r$ -ranks in the voter's decision. The alternative(s) that receive(s) the highest score aggregated over the electorate's preferences is (are) chosen as the  $r$ -Borda winner(s).

In this study, Monte Carlo simulations are run to ascertain the information content of only the first  $r$  ranks of the electorate's preferences from the perspective of implementing the (original) Borda outcome. The way the  $r$ -Borda rule aggregates the voters' preferences is different than the aggregation methods implemented by well-known single- and multi-stage non-ranked procedures which permit truncated ballots. In other words, the present study does not aim to detect the

likelihood of any of these rules to choose the Borda outcome.

The Borda-like aggregation of partially stated individual preferences is a popular method for sports and contests in real life. The Most Valuable Player of the National Basketball Association in the United States, the Eurovision Song Contest, The People's Remix Music Competition and the Formula 1 Car Race are well-known examples such cases. These contests require voters to rank a specified number of candidates. Each stated candidate is given a score depending on its rank in the preference ordering of a voter, and the candidate that receives the highest total score over the electorate is elected as the winner. The number of candidates to be ranked and the scores to be assigned to the ranks differs from one contest to another.

Given that the Borda rule can choose more than one alternative as winners, in this study, two types of probabilities are computed for triples of  $m$ ,  $n$  and  $r$  as the likelihood of choosing the Borda winner with partial individual preferences. The first type refers to the likelihood of the  $r$ -Borda rule choosing exactly the set of Borda winners. The second type of probability considers the likelihood of  $r$ -Borda winners to be included in the set of Borda winners. The changes in these values as a function of  $m$ ,  $n$  and  $r$  are investigated by considering all possible values of these parameters in an appropriate range.

For Monte Carlo simulations, the voters' preferences are generated via the Impartial Anonymous and Neutral Culture Model (IANC). As introduced by Egecioglu and Giritligil (2011), IANC treats voters' preferences through a class of preference profiles, namely root profiles, where the names of both voters and alternatives are immaterial.

#### Contribution and Relation to Literature

To our knowledge, this study is the first attempt in the literature to analyze the extent of difficulty in implementing Borda outcome when voters are asked to rank only a specified number of alternatives and where the underlying model is as structurally general as is possible. The contribution of this computational work and its relation to literature can be discussed based on

two grounds: aggregation of truncated preferences and sampling voters' preferences.

#### Aggregation of Truncated Preferences

A positional scoring rule assigns a score vector  $s = (s_1, s_2, \dots, s_m)$  with  $s_1 \geq s_2 \geq \dots \geq s_m$  and  $s_1 > s_m$  to a preference ranking over a set of  $m$  alternatives, and chooses the alternative(s) with the highest total score aggregated over the rankings of all voters. The Borda rule is defined by the scoring vector  $s(B_m) = (m, m-1, \dots, 1)$ . That is,  $s_i = m + 1 - i$  for all  $i$ , and the difference in scores  $s_i - s_j$  is proportional to  $j - i$  for all  $i$  and  $j$ . This article considers the situation where voters are asked to state only the first  $r$  ranks of their linear preferences over  $m$  alternatives and investigates the importance of the information revealed in the first  $r$  ranks of the electorate's preferences from the perspective of implementing the Borda outcome. This partial information is aggregated through  $r$ -Borda rule  $B_r$ , which assigns alternative  $i$  the score  $s_i(B_r) = r + 1 - i$  if  $i \leq r$ , and  $s_i(B_r) = 0$  otherwise.

It should be noted that the  $r$ -Borda rule aggregates the partial preferences unlike constant scoring rules. A constant scoring rule asks each voter to indicate a given (and constant) number of alternatives. Each indicated alternative receives one point whereas all others get zero, and the alternative with the most votes is elected. Hence, for  $m \geq 3$ , the scoring vector imposed by the  $r$ -Borda rule is not the same as the one assigned by a constant scoring procedure unless  $r = 1$  implying the scoring vector  $(1, 0, \dots, 0)$  which identifies the most popular constant scoring rule, namely the plurality rule. The probability of constant scoring voting rules to select the Borda outcome has been studied by Gehrlein (1981), Gehrlein and Lepelley (2000) and Vandercruyssen (1999). Gehrlein and Fishburn (1980) and Gehrlein, et al. (1982) provide the propensity of pairs of score vectors for a set  $A$  of alternatives and a non-empty proper subset of  $A$  to yield the same ranking over the subset for an arbitrary profile of linear orders on  $A$ .

The method adopted in this paper to aggregate truncated preferences is also different than the procedures that permit truncated ballots. Among these, approval voting has been widely considered in theoretical and practical grounds.

Brams and Fishburn (2002) provide a summary of the theoretical debate between approval voting and the Borda rule. Approval voting requires each voter to indicate the alternatives that she/he approves. Each approved alternative by a voter receives one point and the alternative(s) with the highest point summed over all voters' preferences is (are) chosen as the winner(s). Note that, in approval voting, the number of alternatives to be indicated or ranked by the electorate is not given and thus, is not homogeneous across voters.

Another rule which permits voters to submit truncated preference rankings is majoritarian compromise since it needs at most the first half of the voters' rankings over the entire set of alternatives. Introduced by Sertel (1987), majoritarian compromise selects the candidate(s) that has (have) the support of the majority in the best degree possible. Clearly, both approval voting and majoritarian compromise aggregate the truncated preferences in a different fashion than the  $r$ -Borda rule adopted herein.

Consider a social planner who believes that Borda rule is the 'best' voting rule to aggregate individual preferences into a social choice. Due to the complications about requiring voters to state their total preference orderings, the planner can ask the electorate to report the first  $r$  ( $1 \leq r < m$ ) ranks of their linear preference rankings instead of asking them to state only their first-best choices in the hope of increasing the probability of choosing the Borda winner. In such a case, it seems natural to aggregate the reported partial rankings via a Borda-like procedure for the sake of preserving some consistency in the aggregation method.

#### Sampling Voters' Preferences

An immense literature has been devoted to analyze the outcomes of various social choice rules through the use of computer simulations employing probability models to generate voters' preferences. The most commonly used probability models in the literature are Impartial Culture (IC) and Impartial Anonymous Culture (IAC) conditions. Introduced by Guilbaud (1952), IC is a multinomial equiprobable preference profiles model which assumes that each voter selects her/his preference according

to a uniform probability distribution. IAC, which was first introduced by Fishburn and Gehrlein (1978), also relies on an equiprobability assumption, but without taking the identity of the voters into account. Details about these assumptions and their use in the literature are presented in Berg and Lepelly (1994) and Gehrlein (1997).

In this paper, voters' preferences are sampled through IANC which is also an equiprobability assumption, however, neglecting the names of both alternatives and voters. IANC treats each 'root profile' (profile from which all preference profiles can be generated through renaming the alternatives and voters) equally probable. Because the number of root profiles is small relative to the number of profiles that can be generated for  $m$  alternatives and  $n$  voters, IANC enables the researchers to obtain accurate probabilities even for large parameter values. Based on Egecioglu and Giritligil (2011), it is known that the probabilities computed using the IAC and IANC models coincide only in the vanishingly small likelihood of  $m!$  and  $n$  being relatively prime.

In the social choice theory literature, anonymity and neutrality are among the most important ethical axioms which a voting rule is expected to fulfill. Anonymity requires all voters to be treated equally whereas neutrality calls for equal treatment of alternatives. A large group of voting rules including all scoring rules and pairwise majority relation rules fulfill these two axioms. The outcomes of anonymous and neutral voting rules are invariant under group symmetries of voter and alternative names. Among the probability models used for sampling voters' preferences, IC assumes no set of symmetries whereas IAC takes into account only the symmetry of voter names. On the other hand, IANC takes into account the symmetries of both voters' and alternatives' names. This paper is the first study in the literature that adopts IANC to sample electorate's preferences.

Through the preference sampling method developed by Falmagne and Regenwetter (1996), Regenwetter and Grofman (1998) analyzed seven three-candidate elections conducted under approval voting and constructed a distribution of preference rankings from subset choice data to compare the results

with potential winners of the Borda and Condorcet rules. Based on the method of generalized spectral analysis introduced by Lawson et al. (2006), Brams et al. (2006) compare the results of The Public Choice Society presidential elections in 2006, which was run through approval voting, with the possible outcomes that would have been obtained if plurality, Condorcet, Borda or a single transferable vote had been adopted.

Both Regenwetter and Grofman (1998) and Brams, et al. (2006) start with partial information on voter preferences and assign probabilities to each alternative to be the Borda winner, and based on these probabilities, check whether the possible Borda winner(s) coincide(s) with the actually elected alternative(s). The present study, on the other hand, generates the full orderings of the electorate over the set of alternatives and then considers the first  $r$  ranks of the preference profiles. The approaches of the former studies and the present one are clearly different from each other methodologically.

Preliminaries: Preference Profiles and the Borda rule

A preference on a set  $A$  means any function  $p: A \rightarrow 2^A$  which assigns to every  $a \in A$  a subset (lower contour set)  $p(a) \subseteq A$  such that, at all  $a, b \in A$ :

- (1)  $b \in p(a)$  or  $a \in p(b)$ : completeness;
- (2)  $p(b) \subset p(a)$  whenever  $b \in p(a)$ : transitivity
- (3)  $b \in p(a)$  and  $a \in p(b)$  only if  $a = b$ : anti-symmetry.

Such a preference clearly corresponds to a linear (or total) order on  $A$ .

$\mathbf{p}(A)$  denotes the set of all preferences on  $A$ , any positive integer  $n$  means  $[n] = \{1, 2, \dots, n\}$ , and a preference profile for a society of  $n$  voters on a set  $A$  means any family  $P_{m,n} = (p_i)_{i \in [n]} \in \mathbf{p}(A)^{[n]}$  of preferences  $p_i$  on  $A$  indexed by voters  $i \in [n]$ . Let  $\text{card}(p_i(a))$  be the cardinality of the lower counter set of  $a \in A$  for the voter  $i \in [n]$ . Note that the cardinality of the lower counter set of the top- and bottom- ranked alternatives are  $m$  and  $1$ , respectively.

The Borda score of  $a \in A$  for  $i \in [n]$  is defined as,

$$B_i^a = \text{card}(p_i(a)),$$

and the set of Borda winners at each  $P_{m,n} \in \mathcal{P}(A)^{[n]}$  is determined by setting

$$B(P_{m,n}) = \arg \max_{a \in A} \sum_{i \in [n]} B_i^a.$$

Thus, the Borda rule chooses the candidates who maximize the total Borda score aggregated over the set of all  $n$  voters.

Let  $P_{m,n}^r$  denote the portion of a preference profile  $P_{m,n}$  where only the first  $r$  ranks of the voters' preferences can be observed.

Viewing the profile  $P_{m,n}$  as a  $m \times n$  matrix,  $P_{m,n}^r$  corresponds to the  $r \times n$  submatrix of  $P_{m,n}$ . Note that, if  $r = m$  or  $r = m - 1$ ,  $B(P_{m,n})$  is detectable. However, if  $r < m - 1$ , the observable preference  $p_i^r$  of voter  $i$  corresponds to a partial strict ordering on  $A$ , which is transitive and anti-symmetric, however incomplete. Let  $A_i^r \subseteq A$  be the set of alternatives appearing at  $p_i^r$ .

Let  $\text{card}(p_i^r(a))$  be the cardinality of the observable lower counter set of  $a \in A_i^r$  for  $i \in [n]$ . Note that  $1 \leq \text{card}(p_i^r(a)) \leq r$ . The Borda score of  $a \in A$  for  $i \in [n]$  is redefined as:  $B_i^a = \text{card}(p_i^r(a))$ , if  $a \in A_i^r$ , and 0 otherwise, and the set of  $r$ -Borda winners at any  $P_{m,n}^r$  is given by:

$$B(P_{m,n}^r) = \arg \max_{a \in A} \sum_{i \in [n]} B_i^a.$$

In other words, if an alternative is among the first  $r$ -ranks in voter  $i$ 's ranking, then its associated  $r$ -Borda score is equal to the cardinality of its lower counter set in  $p_i^r$ . If it is not among the top  $r$ -ranked alternatives, it receives a score of zero. The  $r$ -Borda rule chooses the alternative(s) with the highest  $r$ -

Borda score aggregated over  $P_{m,n}^r$  as winner(s).

From this point on,  $B(P_{m,n}^r)$  and  $B(P_{m,n})$  will be denoted by  $B_r$  and  $B$ , respectively, for  $1 \leq r \leq m$ .

Root Profiles and IANC

Let  $\Omega(m, n)$  denote the set of all preference profiles that can be generated for  $m$  alternatives and  $n$  voters. As shown in Eğecioğlu and Giritligil (2011), a product permutation group on the names of alternatives and of voters 'acts' on  $\Omega(m, n)$ , and splits it into a disjoint union of subsets called orbits, that is:

$$\Omega(m, n) = \theta_1 + \theta_2 + \dots + \theta_\omega$$

where each  $\theta_i$  is an anonymous and neutral equivalence class (ANEC). All preference profiles within an ANEC can be generated from each other through re-labeling the alternatives and/or the voters. That is, all preference profiles in any ANEC are 'equivalent' in the sense that any anonymous and neutral voting rule (such as the Borda rule) yields the same outcome (under different names) for all of these profiles.

A root profile is any preference profile that represents an ANEC. That is, all other preference profiles within the same equivalence class can be generated from this root profile via permuting the names of the  $m$  alternatives and of the  $n$  voters. The collection of all root profiles for  $m$  alternatives and  $n$  voters is denoted by  $R = R(m, n)$ , and each element of this set represents an ANEC in  $\Omega(m, n)$ .

Consider a case with two alternatives,  $a$  and  $b$ , and three voters labeled  $v_1, v_2$  and  $v_3$  linearly ranking these alternatives. Note that, in this example, there are  $2!$  preference rankings over alternatives ( $a$  being strictly preferred to  $b$  and  $b$  strictly preferred to  $a$ ). Below are the  $(2!)^3 = 8$  possible preference profiles that can be generated:

$P_1$ :

$v_1$	$v_2$	$v_3$
$a$	$a$	$a$
$b$	$b$	$b$

$P_2$ :

$v_1$	$v_2$	$v_3$
$a$	$a$	$b$
$b$	$b$	$a$

$P_3$ :

$v_1$	$v_2$	$v_3$
$a$	$b$	$a$
$b$	$a$	$b$

$P_4$ :

$v_1$	$v_2$	$v_3$
$b$	$a$	$a$
$a$	$b$	$b$

$P_5$ :

$v_1$	$v_2$	$v_3$
$b$	$b$	$a$
$a$	$a$	$b$

$P_6$ :

$v_1$	$v_2$	$v_3$
$b$	$a$	$b$
$a$	$b$	$a$

$P_7$ :

$v_1$	$v_2$	$v_3$
$a$	$b$	$b$
$b$	$a$	$a$

$P_8$ :

$v_1$	$v_2$	$v_3$
$b$	$b$	$b$
$a$	$a$	$a$

If the group of permutations on the names of the voters acts on the above set of preference profiles, it partitions it into four anonymous equivalence classes (AECs):

$$AEC_1 = \{P_1\}, AEC_2 = \{P_2, P_3, P_4\},$$

$$AEC_3 = \{P_5, P_6, P_7\}, AEC_4 = \{P_8\}.$$

Note that there are two possible permutations on the names of the alternatives: one is the identity permutation which leaves the names of the alternatives intact and the other is the permutation which re-labels  $a$  as  $b$  and  $b$  as  $a$ . If this group of permutations act on the set of AECs, two ANECs are obtained:

$$ANEC_1 = \{AEC_1, AEC_4\},$$

$$ANEC_2 = \{AEC_2, AEC_3\}.$$

The root representing  $ANEC_1$  shows a preference structure at which all voters have the same preference ranking and the root representing  $ANEC_2$  exhibits a structure where one of the preference rankings is adopted by two voters and the other is adopted by one voter.

IANC uses root profiles to represent voters' preferences through an application of the Dixon-Wilf algorithm which enables the root profiles to be generated from the uniform distribution for  $m$  alternatives and  $n$  voters. That is, each root profile is generated uniformly with probability  $1/\text{card}(R(m, n))$ . The formula for  $\text{card}(R(m, n))$  and the details of the application of the Dixon-Wilf algorithm are given by Egecioğlu and Giritligil (2011).

Likelihood Measures: Types of Likelihood

Two types of probabilities are considered to measure the likelihood of implementing the Borda outcome with truncated preference orderings.

1.  $Pr_1 = Pr_1(m, n, r)$  refers to the likelihood of choosing the entire set of Borda winners when only the first  $r$  rows of a preference profile,  $P_{m,n}^r$ , are considered. In other words,  $Pr_1$  is the probability that  $B_r = B$ .

For a given preference profile  $P_{m,n}$ , consider the random variable:

$$f_1(P_{m,n}^r) = 1, \text{ if } B_r = B \text{ and} \\ 0, \text{ otherwise.}$$

Given the distribution of profiles to be generated for given  $m, n$  and  $r$ , the approximate  $Pr_1$  is computed through Monte Carlo integration based on the law of large numbers. The law of large numbers implies that the average of a random sample from a large population is likely to be close to the mean of the whole population. That is,  $Pr_1$  is defined through the random variables  $f_1$ :

$$Pr_1 = \frac{1}{\text{card}(R(m, n))} \sum_{P_{m,n} \in R(m, n)} f_1(P_{m,n}^r).$$

2.  $Pr_2 = Pr_2(m, n, r)$  is the likelihood that an  $r$ -Borda winner is among the Borda winners. Thus, it is the likelihood of an element of  $B_r$  to be also an element of  $B$ . For a given  $P_{m,n}$ , consider the random variable:

$$f_2(P_{m,n}^r) = \frac{\text{card}(B \cap B_r)}{\text{card}(B_r)}$$

(Note that if  $f_1(P_{m,n}^r) = 1$ , then  $B_r = B$ , and consequently,  $f_2(P_{m,n}^r) = 1$ .)

Then, through the above explanation on approximation,

$$Pr_2 = \frac{1}{\text{card}(R(m, n))} \sum_{P_{m,n} \in R(m, n)} f_2(P_{m,n}^r).$$

Below given some examples regarding the calculations of  $f_1(P_{m,n}^r)$  and  $f_2(P_{m,n}^r)$ :

Example 1

a	a	b	c	d
b	b	d	d	a
c	c	c	b	b
d	d	a	a	c

Note that  $B = \{b\}$ . For  $r = 1$ ,  $B_1 = \{a\}$ . Since  $\{a\} \neq \{b\}$ ,  $f_1(P_{4,5}^1) = 0$ . Also  $f_2(P_{4,5}^1) = 0$  because  $\{a\} \cap \{b\} = \emptyset$ . For  $r = 2$ ,  $B_2 = \{a\}$ . So,  $f_1(P_{4,5}^2) = 0$  and  $f_2(P_{4,5}^2) = 0$ .

Example 2

a	a	b	b	c
b	c	c	c	a
c	b	a	d	b
d	d	d	a	d

For the above profile,  $B = \{b, c\}$  and  $B_1 = \{a, b\}$ , so  $f_1(P_{4,5}^1) = 0$ . Since  $\{b, c\} \cap \{a, b\} = \{b\}$ ,  $f_2(P_{4,5}^1) = 1/2$ . For  $r = 2$ ,  $B_2 = \{a, b, c\}$ , so the profile yields  $f_1(P_{4,5}^2) = 0$  and  $f_2(P_{4,5}^2) = 2/3$  (since  $\{b, c\} \cap \{a, b, c\} = \{b, c\}$ ).

The tools provided by Eğecioğlu and Giritligil (2011) allow for the generation of roots profiles from  $R(m,n)$  with probability  $1/\text{card}(R(m, n))$ . The actual probability can then be approximated as follows: generate a large number of root profiles from  $R(m, n)$  with uniform probability  $1/\text{card}(R(m, n))$ , where each selection is independent of the others. If  $S(m, n)$  denote the set of these generated profiles, then the law of large numbers implies that

$$Pr_1 = \frac{1}{\text{card}(S(m, n))} \sum_{P_{m,n} \in S(m, n)} f_1(P_{m,n}^r) \quad (1)$$

Similarly,

$$Pr_2 = \frac{1}{\text{card}(S(m, n))} \sum_{P_{m,n} \in S(m, n)} f_2(P_{m,n}^r) \quad (2)$$

Note that for (1) and (2) to result in a valid Monte Carlo algorithm for the computation of  $Pr_1$  and  $Pr_2$  respectively, it is essential that each  $P_{m,n}$  in  $S(m,n)$  be drawn from the uniform probability on  $R(m, n)$ .

#### Monte Carlo Experiments

At the heart of the Monte Carlo experiments of this study is the Mathematica program `GenerateRoot[m, n]` (the Mathematica notebook containing this function can be accessed online for experimentation: see Eğecioğlu, 2004). The program `GenerateRoot[m, n]` takes two integers  $m$  and  $n$  as input parameters and generates a root profile in a matrix form  $m \times n$  as output. The preference profile generated each time by `GenerateRoot[m, n]` is guaranteed to be distributed over the  $R(m, n)$  roots uniformly. To be able to estimate the probabilities  $Pr_1$  and  $Pr_2$  through the formulations (1) and (2) by using the law of large numbers, the preference profiles generated must be uniform over the set of roots  $R(m, n)$ : `GenerateRoot[m, n]` does exactly that.

The design of the Monte Carlo experiments is as follows. One thousand root profiles are generated for each value of the parameters  $m, n$  under consideration. Thus,  $\text{card}(S(m, n)) = 1,000$ . The ranges  $1 \leq m \leq 30$  and  $1 \leq n \leq 30$  for most of the Monte Carlo experiments carried out. The basic steps followed for the computation of  $Pr_1$  and  $Pr_2$  in the symbolic algebra package Mathematica are:

1. Generate the values of  $m$  and  $n$  themselves,  $1 \leq m, n \leq 30$  iteratively by means of two nested loops.
2. For the given values of  $m$  and  $n$ , invoke the function `GenerateRoot[m, n]`, which generates a preference profile  $P_{m,n}$  from the uniform distribution on the set of root profiles  $R(m, n)$ .
3. Compute the set of Borda winners  $B$  for the profile  $P_{m,n}$  returned.





Figure 1: The Probability that  $r$ -Borda Winners are Identical to the Actual Borda Winners

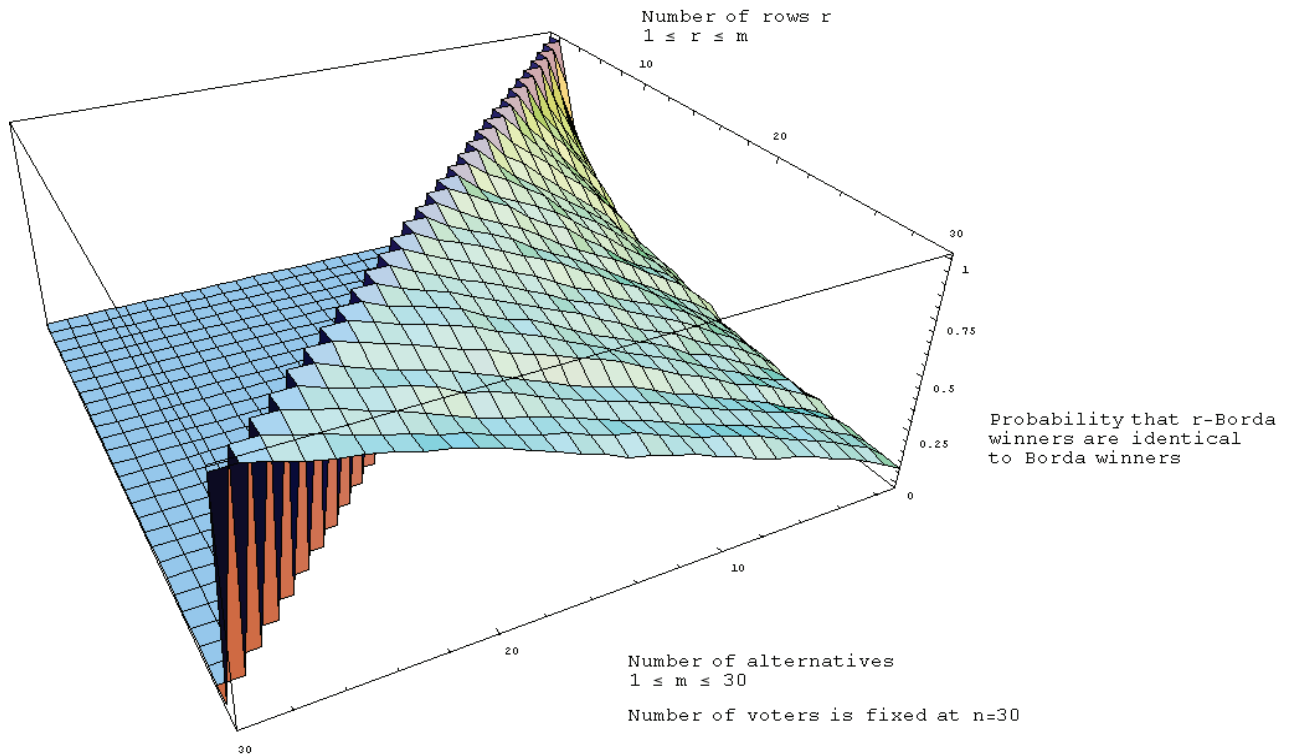


Figure 2: Approximating Function of the Probability that  $r$ -Borda Winners are Identical to Actual Borda Winners

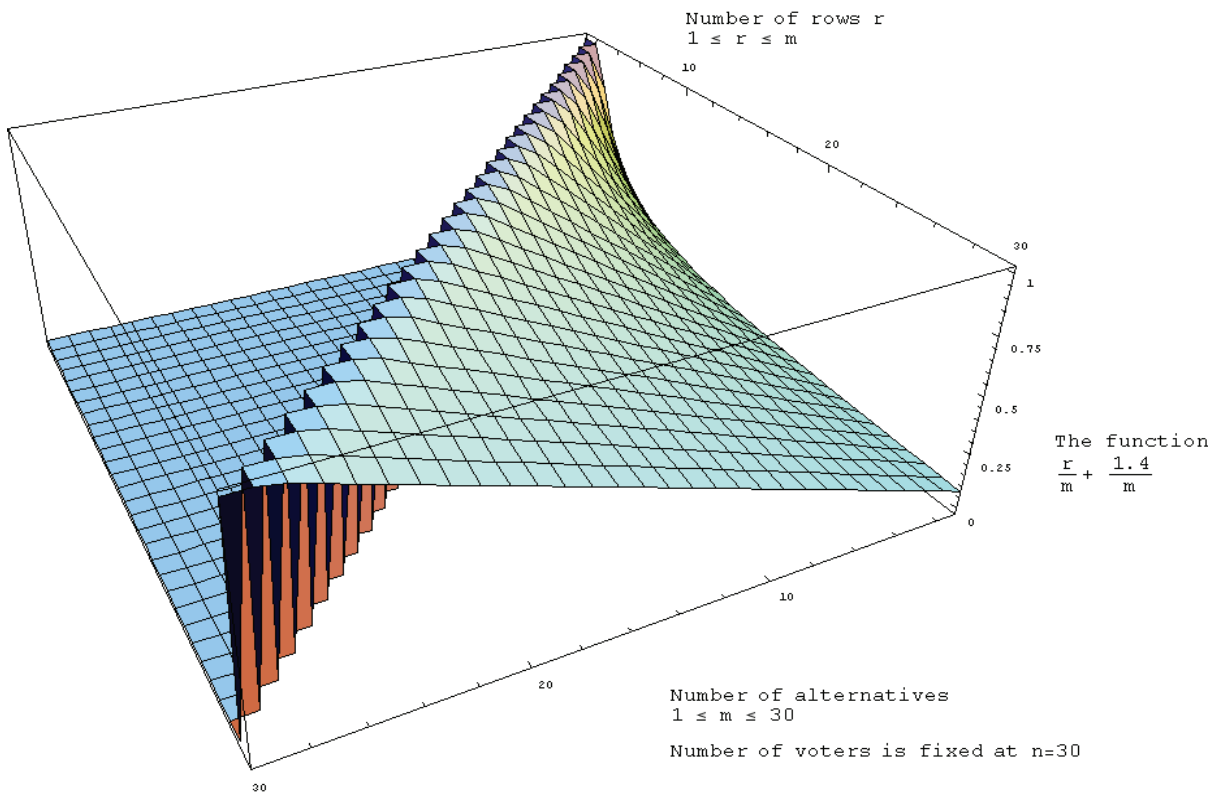




Figure 3: Probability that an  $r$ -Borda Winner is a Borda Winner

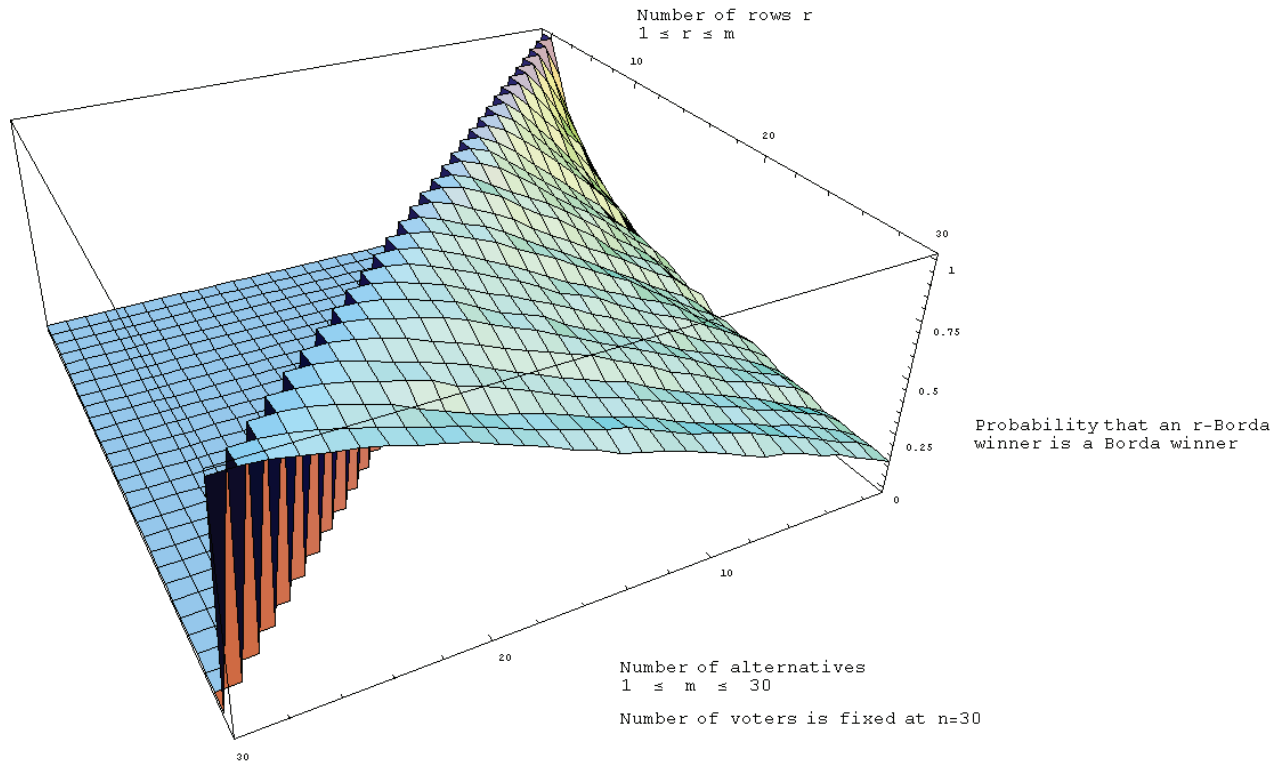
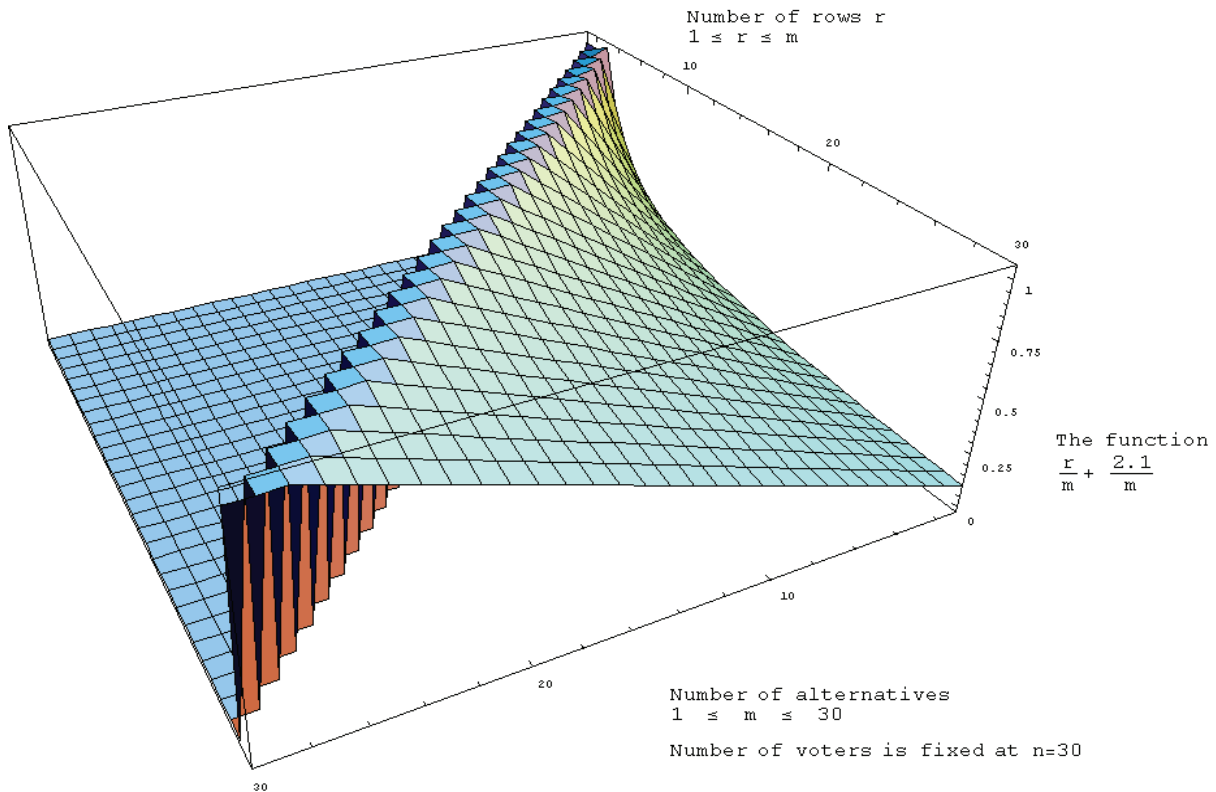


Figure 4: Approximating Function of the Probability that an  $r$ -Borda Winner is a Borda Winner



### Conclusion

The Borda rule is one of the most studied voting procedures in the social choice theory literature. However, despite its well-known superiorities concerning the fulfillment of important positive and normative it is very difficult to be implemented in practice since it requires voters to rank all the alternatives at stake.

This computational study investigates the likelihood of implementing Borda outcome when  $n$  voters are asked to report only the first  $r$  ( $1 \leq r < m$ ) ranks of their linear preferences over  $m$  alternatives. The truncated individual preferences are aggregated through a Borda-like method called the  $r$ -Borda rule.

The voters' preferences are sampled via IANC model which is an equiprobability assumption neglecting the names of both alternatives and voters.

The results of the Monte Carlo simulations indicate that, for large values of  $n$ , the likelihood of choosing exactly the set of Borda winners by considering only the first  $r$  ranks of voter preference orderings is independent of  $n$ , and approaches to zero as  $m$  gets large for a fixed  $r$ . Through the least square fit method, it is shown that, for any  $m$ , it is impossible to guarantee the exact Borda outcome with partial rankings over the alternatives.

It is observed that the likelihood that an  $r$ -Borda winner to be among the Borda winners is also independent of  $n$  and approaches to zero as  $m$  gets large for a fixed  $r$ . Our results show that for  $r = m - k$ ,  $k$  being fixed, this probability approaches to one as  $m$  increases.

Some immediate directions exist for further research on this topic. First, although the  $r$ -Borda rule, as an equal-distance scoring method, is an intuitive way of aggregating the truncated preferences, computational studies can be designed to compare the success of assigning different score vectors to the reported ranks from the perspective of implementing the Borda outcome. Second, given truncated preferences of voters, the likelihood of implementing other well-known ranked rules can be investigated. However, especially in the case of pair-wise majority rules (such as the Condorcet rule), it should be noted that the methods used for aggregating truncated preferences are not as

straightforward or intuitive as in the case of scoring methods. Hence, a similar study for such rules calls for theoretical and computational research.

### References

- Amy, D. C. (2000). *Behind the ballot box: A citizen's guide to voting systems*. Westport, CT: Praeger.
- Berg, S., & Lepelley, D. (1994). On probability models in voting theory. *Statistica Neerlandica*, 48,133-146.
- Borda, J. C. (1781). Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, Paris.
- Brams, S., & Fishburn, P. C. (2002). Voting procedures. In *Handbook of social choice and welfare*, Vol. 1, K. J. Arrow, A. K. Sen & K. Suzumura (Eds.), Amsterdam: Elsevier Science, 175-236.
- Brams, S., Hansen, M. W., & Orrison, M. E. (2006). Dead heat: The 2006 public choice society election. *Public Choice*, 128, 361-366.
- Condorcet, M. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris.
- Doğan, O. and Giritligil, A.E. (2011). Anonymous and neutral social choice: an existence result on resoluteness", Istanbul Bilgi University Murat Sertel Center for Advanced Economic Studies Working Papers, 2011:2.
- Eğecioğlu, Ö. (2004). GenerateRoot[m,n], mathematica function. Available online as a Mathematica notebook *randomvote.nb* in the directory <http://www.cs.ucsb.edu/~omer/>.
- Eğecioğlu, Ö., & Giritligil, A. E. (2011). The impartial, anonymous and neutral culture model: a probability model for sampling public preference structures, *Journal of Mathematical Sociology*, forthcoming.
- Falmagne, J., & Regenwetter, M. (1996). A random utility model for approval voting. *Journal of Mathematical Psychology*, 40, 152-159.
- Fishburn, P. C., Gehrlein, W. (1978). Condorcet paradox and anonymous preference profiles, *Public Choice*, 26, 1-18.

- Fishburn, P. C. (1984). Discrete mathematics in voting and group choice. *SIAM Journal of Algebraic and Discrete Methods*, 5, 263-275.
- Gehrlein, W. V., & Fishburn, P. C. (1980). Robustness of positional scoring over subsets of alternatives. *Applied Mathematics and Optimization*, 6, 241-255.
- Gehrlein, W. V. (1981). Single stage election procedures for large electorates. *Journal of Mathematical Economics*, 8, 263-275.
- Gehrlein, W. V., Gopinath, B., Lagarias, J. C., & Fishburn, P. C. (1982). Optimal pairs of score vectors for positional scoring rules. *Applied Mathematics and Optimization*, 8, 309-324.
- Gehrlein, W. V. (1997). Condorcet's paradox, the Condorcet efficiency of the voting rules. *Mathematica Japonica*, 45, 173-199.
- Gehrlein, W. V., & Lepelley, D. (2000). The probability that all weighted scoring rules elect the same winner. *Economic Letters*, 66, 191-197.
- Guilbaud, G. T. (1952). Les théories de l'intérêt général et le problème logique de l'agrégation. *Economie Appliquée*, 5, 501-584.
- Lawson, B. L., Orrison, M., & Uminsky, D. T. (2006). Spectral analysis of supreme court. *Mathematics Magazine*, 79, 340-346.
- Niemi, R. G., & Riker, W. H. (1976). The choice of voting systems. *Scientific American*, 234, 21-27.
- Nurmi, H. (1987). *Comparing voting systems*, Dordrecht: Reidel.
- Pattanaik, P. K. (2002). Positional rules of collective decision making, In *Handbook of Social Choice and Welfare, Vol.1*, K. J. Arrow, A. K. Sen and K. Suzumura (Eds.), Amsterdam: Elsevier Science, 361-394.
- Regenwetter, M., & Grofman, B. (1998). Approval voting, Borda winners, and Condorcet winners: Evidence from seven elections. *Management Science*, 44, 520-533.
- Saari, D. G. (1987). The source of some paradoxes from social choice and probability. *Journal of Economic Theory*, 41, 1-22.
- Saari, D. G. (1989). A dictionary for voting paradoxes. *Journal of Economic Theory*, 48, 443-475.
- Saari, D. G. (1990). Susceptibility to manipulation. *Public Choice*, 64, 21-41.
- Saari, D. G. (2001). *Chaotic elections! A mathematician looks at voting*. Providence, RI: American Mathematical Society.
- Sertel, M. R. (1987). A non-dictatorial compromise. *Social Choice and Welfare*, 4, 1-11.
- Vandercruyssen, D. (1999). Analysis of voting procedures in one-seat elections: Condorcet efficiency and Borda efficiency, Discussion Paper Series DPS 99.11. Katholieke Universiteit Leuven, Leuven: Center for Economic Studies.

### *Emerging Scholars*

## Type I Error Inflation of the Separate-Variiances Welch $t$ test with Very Small Sample Sizes when Assumptions Are Met

Albert K. Adusah    Gordon P. Brooks  
Ohio University,  
Athens, Ohio USA

---

This Monte Carlo study shows that the separate-variances Welch  $t$  test has inflated Type I error rates at very small sample sizes, especially when sample sizes are very small in one group and larger in the second group – even when all assumptions for the statistical test are met.

Key words: Small sample sizes, type I error inflation, Welch  $t$  test, Student  $t$  test.

---

#### Introduction

It is well known that violations of the homogeneity of variance assumption can severely diminish the confidence we have in the statistically significant results of our statistical tests—in particular, the pooled-variance independent  $t$  test. For example, the independent  $t$  test is relatively robust to violations of the homogeneity of variance assumption when sample sizes are equal, or perhaps even just relatively equal. Stevens (1999) indicated that “unequal variances will distort the Type I error rate appreciably *only if* the group sizes are sharply unequal (largest/smallest  $> 1.5$ )” (p. 9). But when the sample sizes are not relatively close, Type I error rates can be affected dramatically (Author, et al., 2004). As Mickelson and Ayers (2001) stated, “this implies a real risk of claiming to have generated a new

understanding, ostensibly corroborated by a statistical significance test, when in actuality the ‘finding’ is nothing more than an artifact of violating an assumption of the test” (p. 3).

Just as it is well known that the actual Type I error probability rate of the pooled-variance  $t$  test, or Student  $t$  test, is raised or depressed by unequal variances combined with unequal sample sizes, it is also fairly well known that the separate-variances version of the  $t$  test, often called the Welch  $t$  test, usually eliminates these effects (Hinkle, Wiersma & Jurs, 2003). That is, the Welch  $t$  test maintains the nominal Type I error rate (i.e., level of significance or  $\alpha$ ) no matter how unequal the variances. Because power differences between the tests are relatively small when assumptions are met, and because the Welch  $t$  test maintains the nominal  $\alpha$  even under violations of the homogeneity of variances assumption, some researchers have recommended abandoning both the Student  $t$  test and the commonly used preliminary tests of variances (e.g., Levene’s test of equality of variances) in favor of Welch  $t$  tests with no preliminary variance tests. For example, Zimmerman (2004a) suggested that “when sample sizes are unequal, it appears that the most efficient strategy is to perform the Welch  $t$  test or a related separate-variances test unconditionally, without regard to the variability of sample values” (p. 180).

Interestingly results reported – but not interpreted – by Zimmerman (2004a), Gibbons

---

Albert K. Adusah was a doctoral student in Educational Research and Evaluation at Ohio University. Gordon P. Brooks is an Associate Professor of Educational Research and Evaluation. His research interests include statistics education, power and sample size analysis, and Monte Carlo programming. Email: brooksg@ohio.edu.

Editorial Note: JMASM regrets the passing of Mr. Adusah. This article was accepted for issue 7(2), but was inadvertently omitted.

and Chakraborti (1991), and Penfield (1994) suggested a problem with Type I error rates for the Welch  $t$  test even when variances are equal. The Welch  $t$  test appears to exhibit inflated Type I error rates when sample sizes are very small and the homogeneity assumption is met (i.e., both groups have the same variance). For example, Zimmerman found that with  $n_1 = 5$  and  $n_2 = 25$ , actual Welch  $t$  test Type I error rates were approximately 0.058 in the equal variance condition; as the standard deviation ratio increased from 1.0 to 2.5, however, Type I error rates decreased toward 0.05. Gibbons and Chakraborti calculated a similar result with equal variances when  $n_1 = 4$  and  $n_2 = 16$ : an actual Type I error rate of 0.0582. Curiously, Penfield reported a too-conservative actual Type I error rate for the Welch  $t$  test with  $n_1 = 5$  and  $n_2 = 15$ .

Unfortunately, because none of these studies sought to examine this problem specifically, they did not include sufficient sets of conditions to confirm whether such results represented systematic bias or were simply artifacts of Monte Carlo sampling error (e.g., result of a particular random number generator seed or a particular random number generation process). For example, Zimmerman (2004a) only used conditions where the sample size combinations were (50, 10), (40, 20), (25, 5), and (20, 10). Penfield (1994) used combinations of (5, 5), (10, 10), (20, 20), (5, 15), and (10, 20). Gibbons and Chakraborti (1991) only used sample size combinations of (10, 10) and (4, 16). However, when taken together, these studies suggest that it may be fruitful to examine the matter further. Therefore, the purpose of this study is to investigate the Type I error rate behavior of the Welch  $t$  test under very small sample size conditions.

It is commonly understood that the Type I error rates of the Student  $t$  test and the Welch  $t$  test differ in respect to how these tests fare when both sample sizes and population variances are unequal across groups. These conditions alter both Type I error rates and power (Author, et al., 2004); that is, when the larger group has the smaller variance, the actual Type I error rate of the Student  $t$  test is inflated – or higher than the nominal Type I error rate. In other words, researchers would make more Type I errors than

they expect to make using their given level of significance. Recall that when Type I error is set to 0.05, a researcher expects to make Type I errors at a rate of 5%; when assumptions are violated and the actual Type I error rate becomes inflated, however, the expected number of actual Type I errors is higher than 5% over a hypothetically large number of samples.

For example, if a researcher conducts (hypothetically) 100 statistical tests where the null hypothesis is true and statistical assumptions are met, 5 of those 100 tests would be wrongly rejected using an actual Type I error rate roughly equal to nominal  $\alpha = 0.05$ ; but if the homogeneity of variance assumption is not met and the actual Type I error rate becomes inflated to 0.14, then roughly 14 of the 100 null hypotheses would be wrongly rejected, not 5 as expected when  $\alpha = 0.05$ . Conversely, when the larger group has the larger variance, the Type I error rate of the Student  $t$  test is conservative (i.e., lower than nominal  $\alpha$ ) and the null hypothesis is rejected less often than it should be (e.g., 2% of the time), which in turn reduces statistical power.

Much research has confirmed that these problematic properties of the Student  $t$  test can be eliminated by using the Welch  $t$  test (e.g., Gibbons & Chakraborti, 1991; Glass, Peckham & Sanders, 1974; Zimmerman, 2004a). Numerous studies have found that the Welch  $t$  method maintains Type I probabilities close to the nominal significance level and also eliminates spurious increases or decreases of Type II error rates and power (Zimmerman, 2004b). Although several studies have investigated unequal samples and unequal variances, no studies could be found that examined the impact of small sample sizes on such results. That is, Monte Carlo studies have included sample size as a variable (e.g., Gibbons and Chakraborti, 1991; Zimmerman, 2004a), but none could be found that systematically studied the effects of sample size itself on Type I error.

Gibbons and Chakraborti (1991) compared the Mann-Whitney  $U$  test, the Student  $t$  test, and the Welch  $t$  test. They used a total sample size of 20 for the two groups, sometimes equal (i.e.,  $n_1 = n_2 = 10$ ) and sometimes with  $n_1 = 4$  and  $n_2 = 16$ . Because their focus was on violations of assumptions, they paid little



## WELCH $t$ UNDER SMALL SAMPLES

attention to the inflated Type I error rates of the Welch  $t$  test for the equal variance but unequal groups condition, where "the largest difference of the average of the three runs was  $0.0596 - 0.0500 = 0.0096$  for the two-tailed [Welch  $t$ ] test" (p. 261). This is the summary of their results wherein actual Type I error for the equal variance but unequal sample size conditions were consistently beyond Bradley's (1978) fairly stringent criterion of  $\alpha \pm 0.1\alpha$  (i.e., 0.045 to 0.055). In the end, Gibbons and Chakraborti recommended that "if the populations can be assumed normal with equal variances, use Student's  $t$  test for any sample sizes" (p. 266), but "if the populations can be assumed normal but the variances cannot be assumed equal, use the alternate  $t$  test for any sample sizes" (p. 266). Gibbons and Chakraborti recommended the Mann-Whitney test for non-normal data and when either (or both) sample size is less than 30.

Also for example, Zimmerman (2004a) compared the unconditional Student  $t$  test (i.e., no preliminary test of equality of variances), the unconditional Welch  $t$  test, and the Conditional  $t$  test (i.e., Levene's test followed by the appropriate  $t$  test). Zimmerman reported – but did not comment on – the condition where  $n_1 = 25$ ,  $n_2 = 5$ , and  $\sigma_1/\sigma_2 = 1.0$ , in which actual Type I error was 0.058 for the Welch  $t$  test but a more accurate 0.051 for the Student  $t$  test. Because Gibbons and Chakraborti (1991) used on 5,000 replications per condition, their results may have been subject to Monte Carlo sampling error issues (e.g., a poor seed choice, a particularly odd set of 5,000 randomly drawn samples). However, Zimmerman's (2004a) results were based on 50,000 replications, thus producing results less likely to be due to Monte Carlo sampling error issues. Further, among the equal variance conditions in both studies, only these results with very small  $n$  in one group were outside the fairly stringent range (i.e., 0.045 to 0.055).

### Small Sample Sizes in Research

Although very small sample sizes are rare when  $t$  tests are used in actual research, several meta-analyses have been reported to suggest that researchers sometimes, in practice, do use very small sample sizes. For example, Reid, Kenaley and Colvin (2004) completed a

meta-analysis of 39 small-group interventions in social work. They found that 15 of these 39 studies (i.e., 38%) had a total sample size of 20 or less; only 10 had total sample sizes over 50. Similarly, Shadish and Baldwin (2005) performed a meta-analysis of marital therapy interventions and found 14 of 30 studies had total sample sizes of 20 or less, while only 2 had total sample sizes over 50. Unfortunately, these studies did not report individual sample sizes, so whether group sizes were equal is unknown without further investigation.

### Methodology

A Monte Carlo data generation and analysis program, called MC4G: Monte Carlo Analyses for up to 4 Groups (Author, 2005), was used to simulate data to obtain the appropriate Type I error rates. The rejection rates of both the Student  $t$  test and the Welch  $t$  test will be recorded for various combinations of sample sizes, especially with very small sample size in one group. That is, the specific conditions for the study were: (a) both Group 1 and Group 2 means remained constant at 0.0, (b) Group 1 sample size varied from 3 to 150 by 1, (c) Group 2 sample size varied from 3 to 30 by 1, (d) Group 1 standard deviation remained constant at 1.0, and (e) Group 2 standard deviation varied from 1.0 to 4.0 by 0.5.

For the primary research question, only the 3,738 conditions were analyzed where Group 1 sample sizes were larger than Group 2 sample sizes and both standard deviations were 1.0; however, some other conditions were analyzed for specific reasons. All data were generated to follow a univariate normal distribution. There were 100,000 replications performed for each condition in order to minimize the impact of Monte Carlo sampling problems. For each sample generated, appropriate standard error estimates and degrees of freedom were used to calculate both the Student  $t$  test (Hinkle et al., 2003, p. 240), the Welch  $t$  test (Hinkle et al., 2003, p. 252), and a Conditional  $t$  test (either the Student  $t$  test or the Welch  $t$  test was calculated appropriately depending on the results of Levene's test of equality of variances). Nominal level of significance was set at  $\alpha = 0.05$  for each test performed.

The MC4G program was developed (Brooks, 2005) to perform Monte Carlo analyses for  $t$  tests and ANOVA in a Windows environment. The MC4G program was written in Delphi Pascal and is available for download from the author's web site (see references). The program was used to create normally distributed data that met the conditions for the study. For these robustness analyses, the number of incorrect rejections of the null hypothesis (i.e., Type I error rate) was stored and reported by the program.

The MC4G program uses the L'Ecuyer (1988) uniform pseudorandom number generator. Specifically, the FORTRAN code of Press, Teukolsky, Vetterling, and Flannery (1992), was translated into Delphi Pascal. The L'Ecuyer generator was chosen because of its large period and because combined generators are recommended for use with the Box-Muller method for generating random normal deviates, as will be the case in this study (Park & Miller, 1988). The computer algorithm for the Box-Muller method used in this study was adapted for Delphi Pascal from the standard Pascal code provided by Press, Flannery, Teukolsky and Vetterling (1989). Extended precision floating point variables were used, providing the maximum possible range of significant digits. Simulated samples were chosen randomly to test program function by comparison with results provided by SPSS.

### Results

First, the Type I error rates of the Student  $t$  test are investigated across the full range of sample size conditions. These results confirmed that Type I error rates for the Student  $t$  test are robust to variation of all sample sizes tested. Specifically, every one of the 3,738 sample size conditions under equal variances (i.e., both group standard deviations are 1.0) was between 0.0446 and 0.0560, just beyond the most stringent criterion recommended by Bradley (1978). One would not expect Type I error rates of exactly 5% due to the sampling error inherent to the Monte Carlo process. Therefore, Bradley recommended a stringent criterion of  $\alpha \pm 0.1\alpha$  to be used for robustness studies; that is, results within 10% of  $\alpha$  are considered close enough to  $\alpha$  for the statistical test to be considered robust

to the conditions being investigated. These results are shown graphically in Figure 1.

A similar examination of the Welch  $t$  test was performed and an issue with robustness for these results was identified (see Figure 2). In particular, the actual Type I error rates across the 3,738 conditions (100,000 samples per condition) ranged from 0.0424 to 0.0793. Clearly, some of the Type I error rates for the Welch  $t$  test fell outside Bradley's (1978) stringent criterion range. Further comparison showed that 99% of all Student  $t$  test Type I error rates were less than 0.0536, but only 88% of the Welch  $t$  test Type I error rates were below 0.0551, at the top end of Bradley's range. Also, there were only 10 extreme Student  $t$  test Type I error rates beyond 0.0542 but there were 340 extreme Welch  $t$  test Type I error rates beyond 0.0569.

In order to investigate further the inflated Type I error rates for the Welch  $t$  test, an attempt was made to identify the patterns in Figure 2. Observe clear patterns among the scatter that represent Group 2 sample sizes. For example, at the top of the chart, there is a clear pattern of circles, representing a Group 2 sample size of  $n_2 = 3$ . Because a sample size of  $n_2 = 3$  is not practical, we examined further the  $n_2 = 5$  condition (while still not terribly practical, it is more reasonable than  $n_2 = 3$  and has been studied by several authors cited above). Table 1 shows these results for a subset of the data (only where  $n_1 < 45$ , but no important differences existed beyond  $n_1 = 45$ ). Figure 3 displays these data for equal variances, Figure 4 illustrates the data where variances were unequal (Group 1 SD = 1.0 and Group 2 SD = 2.0), Figure 5 shows the data where variances were unequal (Group 1 SD = 1.0 and Group 2 SD = 4.0).

The Welch  $t$  test clearly has inflated Type I error rates when sample sizes are small and unequal; however, note in Figure 3 that the inflation does not emerge until the sample size ratio increases beyond 2:1 (specifically, where  $n_1 = 13$  and  $n_2 = 5$ ). Although the inflation is not dangerously high, as is the case with the Student  $t$  test when both sample size and variances are unequal (e.g., where  $n_1 = 44$ ,  $n_2 = 5$ ,  $\sigma_1 = 1.0$ , and  $\sigma_2 = 4.0$ , as shown in Figure 5), it does exist. Interestingly, Figures 4 and 5 show that the Welch  $t$  test does indeed maintain nominal Type

## WELCH $t$ UNDER SMALL SAMPLES

I error rates when variances are unequal, but Figure 3 shows that when variances are equal the Type I error rates are biased upward. Further investigation beyond the conditions where  $n_2 = 5$  suggested that the problem is limited to very small sample sizes. Figure 6 shows that, although there is a clear, upward bias of Type I error beyond a smaller group size of  $n = 10$ , those rates do fall well within Bradley's (1978) stringent criterion range. Figure 6 also shows that the average inflation of Type I error reduces dramatically as the smaller group size increases. Further note in Figure 6 that the  $t$  test conditional on the result of Levene's test does not help the matter, because its Type I error rates are inflated even beyond the Welch  $t$  test once  $n_2 > 4$ .

### Conclusion

Results suggest that the Welch  $t$  test is indeed inflated, according to Bradley's (1978) fairly stringent criterion, when sample sizes are unequal – even when assumptions for the  $t$  test are met in the population. The inflation rate seems to be dependent more on the size of the smaller group than on the total sample size, but sample size ratio does seem to play a small role (i.e., with roughly equal sample sizes there was no apparent inflation). Although the Welch  $t$  test Type I error inflation exposed here is not dangerously high, it is high enough to be considered more than trivial, particularly with the smallest smaller group sample sizes examined. Specifically, Type I error rates are inflated beyond Bradley's stringent criterion

Figure 1: Type I Error Rates for the Student  $t$  test when Homogeneity of Variance Assumption Is Met in the Population

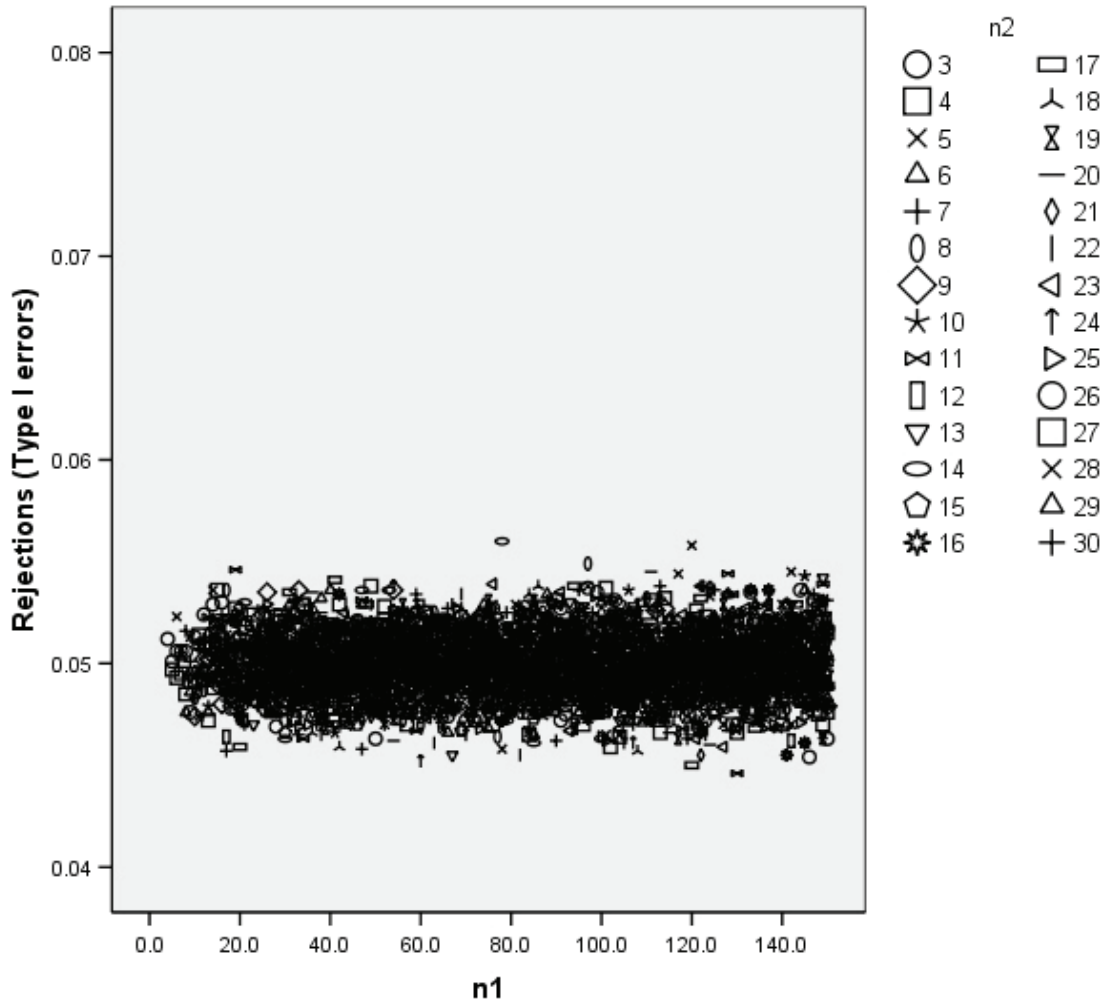


Figure 2: Type I Error Rates for the Welch  $t$  test when Homogeneity of Variance Assumption Is Met in the Population

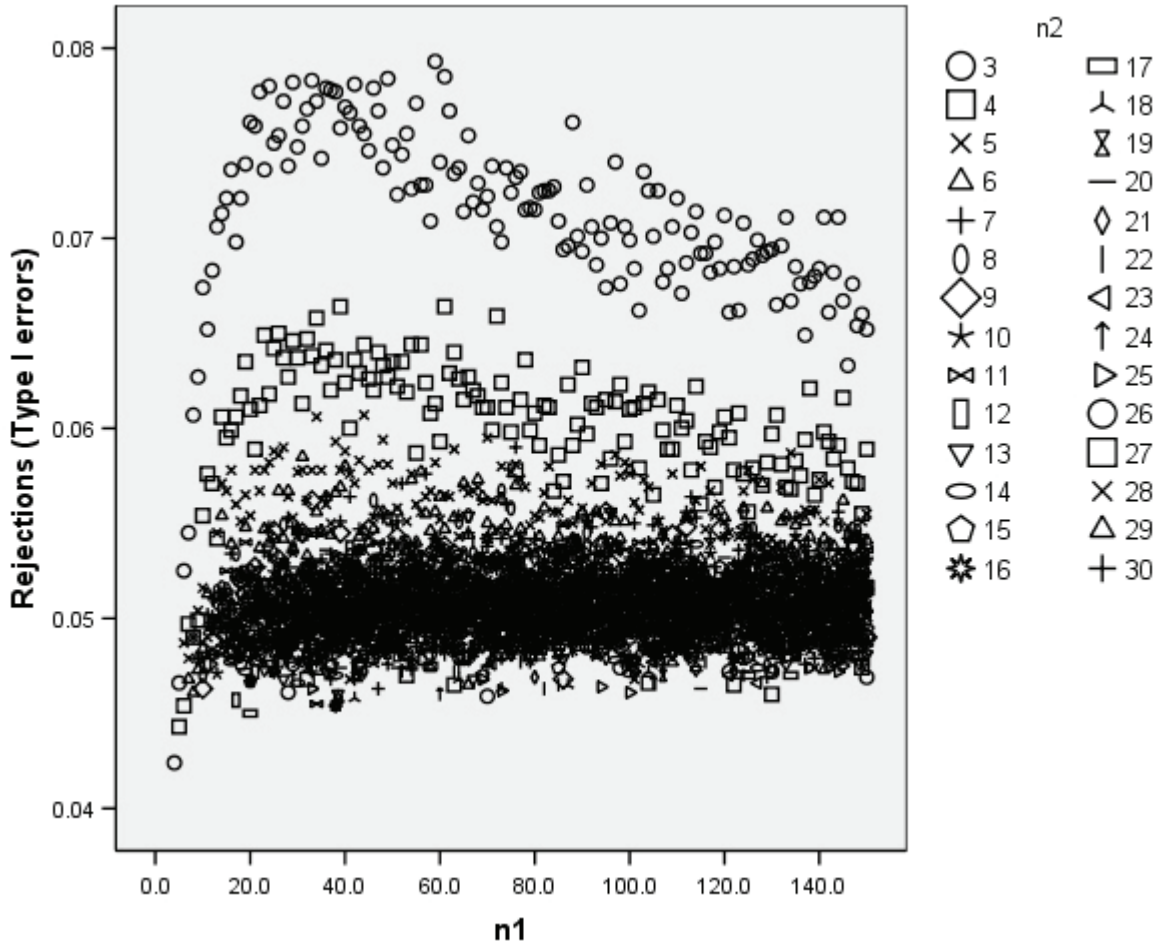


Table 1: Type I error rates of Student  $t$  test, Welch  $t$  test, and the Conditional  $t$  test at  $\alpha = 0.05$  Where  $n_2 = 5$ ,  $n_1 < 45$ ,  $n_1 > n_2$ , and Both Population Standard Deviations are 1.0

$n_1$	Student $t$	Welch $t$	Conditional $t$
6.0	0.0523	0.0487	0.0514
7.0	0.0507	0.0479	0.0499
8.0	0.0495	0.0490	0.0499
9.0	0.0513	0.0504	0.0519
10.0	0.0493	0.0516	0.0505

## WELCH $t$ UNDER SMALL SAMPLES

Table 1 (continued): Type I error rates of Student  $t$  test, Welch  $t$  test, and the Conditional  $t$  test at  $\alpha = 0.05$   
 Where  $n_2 = 5$ ,  $n_1 < 45$ ,  $n_1 > n_2$ , and Both Population Standard Deviations are 1.0

$n_1$	Student $t$	Welch $t$	Conditional $t$
11.0	0.0474	0.0490	0.0488
12.0	0.0492	0.0512	0.0518
13.0	0.0505	0.0545	0.0523
14.0	0.0536	0.0571	0.0557
15.0	0.0498	0.0567	0.0525
16.0	0.0519	0.0578	0.0556
17.0	0.0514	0.0560	0.0542
18.0	0.0508	0.0551	0.0551
19.0	0.0505	0.0565	0.0541
20.0	0.0525	0.0554	0.0554
21.0	0.0499	0.0578	0.0549
22.0	0.0493	0.0567	0.0532
23.0	0.0507	0.0578	0.0558
24.0	0.0501	0.0582	0.0553
25.0	0.0493	0.0588	0.0551
26.0	0.0527	0.0586	0.0586
27.0	0.0494	0.0590	0.0556
28.0	0.0507	0.0564	0.0554
29.0	0.0501	0.0553	0.0558
30.0	0.0512	0.0577	0.0569
31.0	0.0480	0.0578	0.0542
32.0	0.0503	0.0562	0.0561
33.0	0.0532	0.0578	0.0589
34.0	0.0490	0.0606	0.0573
35.0	0.0481	0.0578	0.0547
36.0	0.0485	0.0560	0.0549
37.0	0.0503	0.0550	0.0567
38.0	0.0490	0.0593	0.0567
39.0	0.0498	0.0578	0.0567
40.0	0.0490	0.0588	0.0560
41.0	0.0518	0.0576	0.0565
42.0	0.0485	0.0571	0.0558
43.0	0.0507	0.0583	0.0573
44.0	0.0501	0.0607	0.0578

Figure 3: Graphical Display of Results Where  $n_2 = 5$  Across All  $n_1 > 5$  and Both Standard Deviations were 1.0

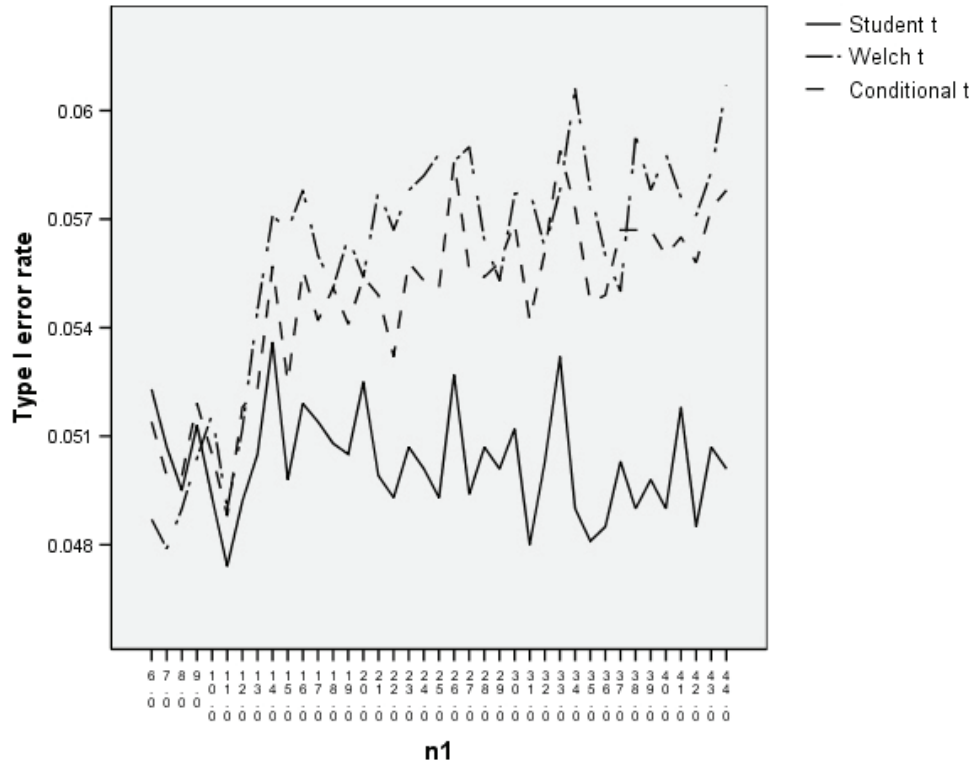
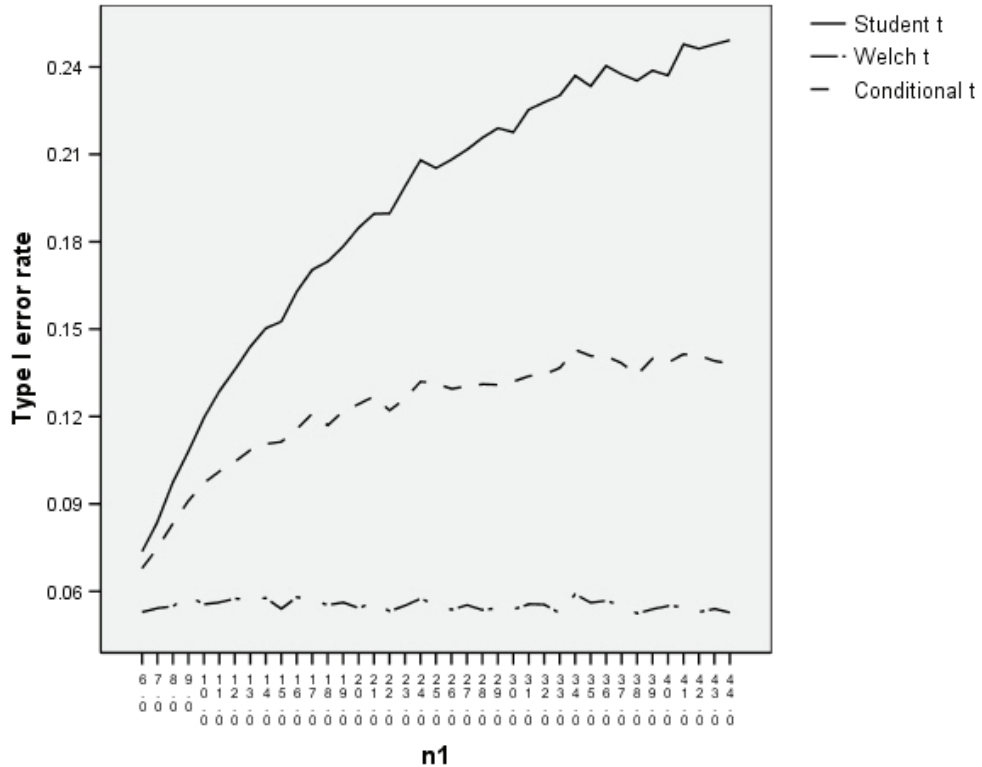


Figure 4: Graphical Display of Results where  $n_2 = 5$  Across All  $n_1 > 5$ , Group 1 Standard Deviation was 1.0, and Group 2 Standard Deviation was 2.0



## WELCH $t$ UNDER SMALL SAMPLES

Figure 5: Graphical Display of Results where  $n_2 = 5$  Across All  $n_1 > 5$ , Group 1 Standard Deviation was 1.0, and Group 2 Standard Deviation was 4.0

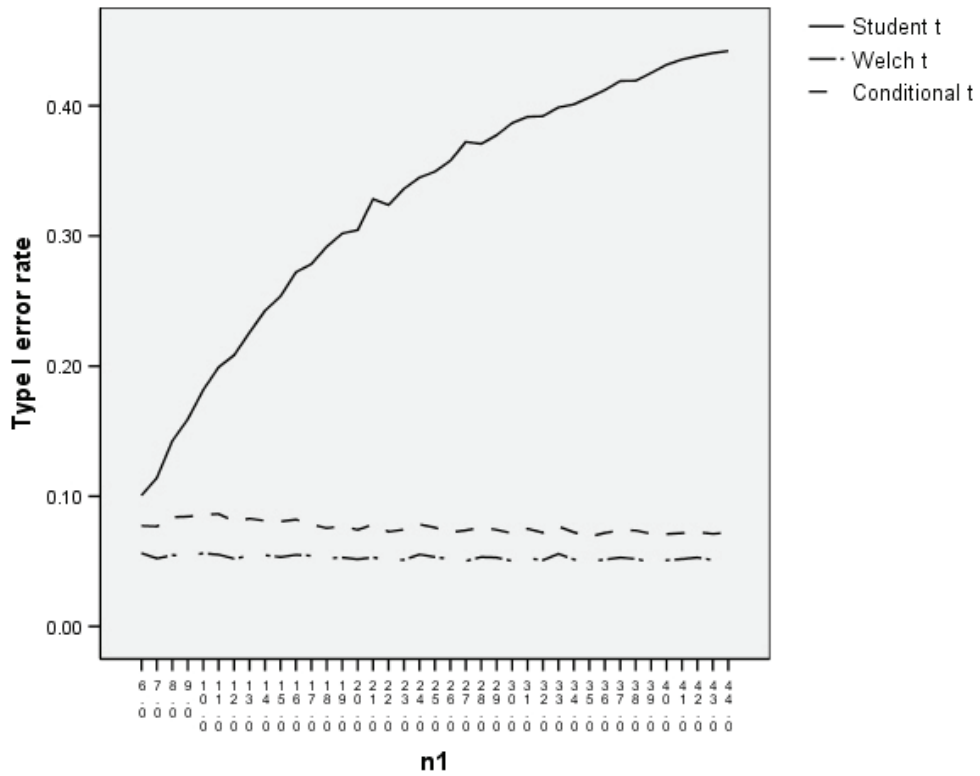
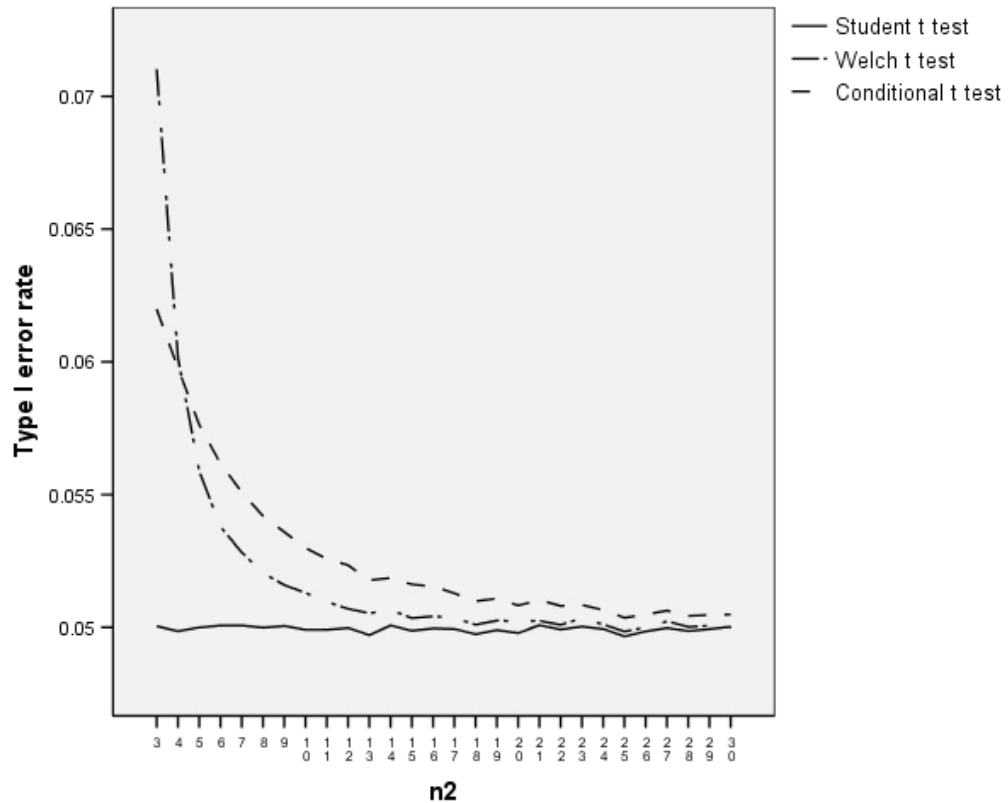


Figure 6: Average Type I Error Rates where Both Standard Deviations are 1.0



when the smaller sample size is less than  $n = 6$ . We also found that the inflation problem becomes relatively benign once the smaller sample size is greater than  $n = 10$ ; that is, the average actual Type I error rates for the Student  $t$  test and the Welch  $t$  test differ by no more than 0.002 when smaller  $n > 10$ . Finally, we confirmed that the Student  $t$  test did not exhibit any noticeable problems with Type I error when assumptions are met, no matter the sample size combinations.

There have been a number of studies to ascertain the best statistical test to use for two-group comparison studies under violations of assumptions. Such studies have often also showed that there is not a dramatic difference in statistical power between the Student  $t$  test and Welch  $t$  test under many conditions. Consequently, these results have led some scholars (e.g., Zimmerman, 2004a) to recommend using the Welch  $t$  test unconditionally, so as to minimize the impact of violations of assumptions on Type I error rates. Unfortunately, because it appears that the Welch  $t$  test may have unexpected problems when one group is very small, this recommendation may lead to problems in studies with very small sample sizes. Indeed, supplemental analyses performed here suggested that the Welch  $t$  test may be conservative for very small, equal sample sizes (less than 7 in each group) even when variances are equal.

Because the Conditional  $t$  test did not help the situation, there is no easy solution to the problem. That is, because one does not know whether the homogeneity of variance assumption has been violated, one cannot know which  $t$  test to choose with small sample sizes. More specifically, if one knew that the populations had unequal variances, one could choose to use the Welch  $t$  test with little concern for type I error, even with small sample sizes; conversely, if one knew that variances were equal, one could use the Student  $t$  test. However, the commonly recommended Conditional  $t$  test using Levene's test also appears to lead to inflated type I error rates with very small sample sizes in one group and with larger sample sizes in the other—even when variances are equal.

The most obvious recommendation, for a variety of reasons both statistical and

otherwise, is for researchers to use more than 10 participants per group when comparing means. In situations where there is no choice, based on Gibbons and Chakraborti's (1991) results, it appears that researchers should use the Mann-Whitney  $U$  test when sample sizes are very small to maintain nominal Type I error rates; their results do not hint at any inflation of Type I error rates at small sample sizes. However, future research must verify this recommendation. Further investigation into type I error rates should include examinations of Analysis of Variance and its alternatives (e.g., Brown-Forsythe, Welch, and Kruskal-Wallis). There is no reason to expect terribly different results when viewed from an ANOVA perspective; such similarities between the Type I error rate properties of the  $t$  test and ANOVA have been confirmed in the literature (e.g., Glass, Peckham, & Sanders, 1974). Finally, these results relied on the assumption of normality being met; future researchers may want to investigate the problem by violating the normality assumption. Based on work by Gibbons and Chakraborti, and others, there is reason to suspect that the nonparametric tests should be uniformly adopted as the tests of choice when the sample size of at least one group is very small.

#### References

- Brooks, G. P. (2005). MC4G: Monte Carlo analyses for up to 4 groups [computer software]. <http://www.ohio.edu/people/brooksg/software.htm>.
- Brooks, G. P., Fang, H., Heh, V. (2004, October). *Modeling true Type I error rates: A Monte Carlo analysis*. Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Brooks, G. P., Johanson, G. A., & Barcikowski, R. S. (2004, April). *Underappreciated factors that affect statistical power*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.



## WELCH $t$ UNDER SMALL SAMPLES

Gibbons, J. D., & Chakraborti, S. (1991). Comparisons of the Mann-Whitney, Student's  $t$ , and alternate  $t$  tests for means of normal distributions. *Journal of Experimental Education*, 59, 258-267.

Glass, G. V., Peckham, P. D., Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42, 237-288.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5<sup>th</sup> Ed.). Boston: Houghton Mifflin.

Mickelson, W. T., & Ayers, R. (2001, October). *Testing for group mean difference: What happens when population variances are unequal*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Reid, W. J., Kenaley, B. D., & Colvin, J. (2004). Do some interventions work better than others? A review of comparative social work experiments. *Social Work Research*, 28(2), 71-81.

Shadish, W. R., & Baldwin, S. A. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology*, 73, 6-14.

Stevens, J. P. (1999). *Intermediate statistics: A modern approach* (2<sup>nd</sup> Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Zimmerman, D. W. (2004a). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57, 173-181.

Zimmerman, D. W. (2004b). Conditional probabilities of rejecting null hypotheses by pooled and separate variances  $t$  tests given heterogeneity of sample variances. *Communications in Statistics - Simulation and Computation*, 33, 69-81.

## A Robust One-Sided Variability Control Chart

P. Borysov      Ping Sa  
University of North Florida  
Jacksonville, FL USA

---

A new control charting technique to monitor the variability of any distribution is proposed. The simulation study shows that the new method outperforms all the existing methods in controlling the Type I error rates and it also has good power performance for all distributions considered in the study.

Key words: Edgeworth expansion, Type I error rate, power performance.

---

### Introduction

A major objective of statistical quality control is to quickly detect any sustained shift of central tendency and variability of a process. The control chart, proposed by Shewhart (1931), is an on-line process-monitoring technique widely used for this purpose. The Shewhart chart contains three lines: the center line, which represents the average value of the quality characteristic, and two control limit lines, the UCL (Upper Control Limit) and LCL (Lower Control Limit). These lines are chosen in such a way that, if the process is in control, nearly all the sample points will fall between the lines. If a sample point plots between the two control limits, the process is assumed to be in control. If any point plots above UCL or below LCL, then it is reasonable to suspect that the process is out of control. In this case, investigations and corrective actions are required to find and to eliminate the assignable cause responsible for this behavior.

Much work has been done to develop and improve control charts that are able to detect small and large shifts in the process mean. However, less work has been done to control the process variability. One of the most widely used methods to control variability of a process is the Shewhart R-chart. The UCL and LCL for the standard three-sigma chart are as follows:

$$UCL = \bar{R} + 3\hat{\sigma}_R = \bar{R} + 3d_3 \frac{\bar{R}}{d_2},$$

and

$$LCL = \bar{R} - 3\hat{\sigma}_R = \bar{R} - 3d_3 \frac{\bar{R}}{d_2},$$

with a Center line =  $\bar{R}$ , where

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m},$$

and  $R_i$  is the range (difference between the largest and the smallest observation) of the  $i^{\text{th}}$  preliminary sample;  $d_2$  and  $d_3$  are the mean and the standard deviation of  $W = \frac{R}{\sigma}$ , respectively. Tables of  $d_2$  and  $d_3$  are available for various sample sizes (Montgomery, 1996). The Shewhart R-charts are constructed under the assumption that the underlying process is normally distributed.

An alternative to the R-chart is the S-chart. Rather than using range as a measure of

---

Ping Sa is a Professor of the Department of Mathematics and Statistics. She received her Ph.D. in Statistics from the University of South Carolina in 1990. She has published 20 papers. Her recent scholarly activities have involved research in multiple comparisons and quality control. Email: [psa@unf.edu](mailto:psa@unf.edu). P. Borysov received his Master's degree in Statistics from the University of North Florida in 2007.

## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

variability, the S-chart uses the standard deviation. The UCL and the LCL of an S-chart are formulated as follows:

$$UCL = \bar{s} + 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2},$$

and

$$LCL = \bar{s} - 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2},$$

with a Center line  $\bar{s}$  where

$$\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i, \quad s_i$$

is the sample standard deviation of the  $i^{\text{th}}$  preliminary sample,  $c_4$  is a constant such that the statistic  $s$  is an unbiased estimator of  $c_4 \sigma$ . Tables of  $c_4$  for various sample sizes are available in many statistical quality control books (Montgomery, 1996). Similar to the R-chart, all  $c_4$  tables are constructed under the assumption of normal process.

Another alternative charting technique recommended by many practitioners, is the Shewhart  $S^2$ -chart. In the construction of a  $S^2$ -chart, the fact that  $\frac{(n-1)S^2}{\sigma^2}$  has  $\chi^2_{(n-1)}$  distribution under normality is used. The control limits for this chart are:

$$UCL = \frac{\bar{s}^2}{n-1} \chi^2_{\frac{\alpha}{2}, n-1},$$

with Center line =  $\bar{s}^2$  and

$$LCL = \frac{\bar{s}^2}{n-1} \chi^2_{1-(\frac{\alpha}{2}), n-1}$$

where  $\chi^2_{\frac{\alpha}{2}, n-1}$  and  $\chi^2_{1-(\frac{\alpha}{2}), n-1}$  denote the upper

and lower  $\frac{\alpha}{2}$  percentage points of the Chi-

square distribution with  $n-1$  degrees of freedom, and  $\bar{s}^2$  is the average sample variances of  $m$  preliminary samples.

In many situations the underlying distribution of the process might not be normal. For example, the distributions of measurements from chemical processes and cutting tool wear processes are often skewed. Burr (1967) and Chan, Hapuarachchi and Macpherson (1988) have examined the effect of non-normality on R-charts. They found that, for skewed populations, Type I risk probabilities grow larger as the skewness of the distribution increases. The problem is in the “discrepancy between the variability pattern of the asymmetric distribution and the normality assumed in placing control limits on Shewhart R-chart.” (Bai & Choi, 1995, p. 120). The impact of non-normality on the S-chart and  $S^2$ -chart is also expected.

To remedy the non-normal problem, Bai and Choi (1995) proposed a heuristic method for controlling variability of the skewed distributions based on the Weighted Variance (WV) method. Their chart is an R chart with 3-sigma control limits:

$$UCL = \bar{R} \left[ 1 + 3 \frac{d'_3}{d'_2} \sqrt{2 \hat{P}_X} \right] = V_U \bar{R},$$

and

$$LCL = \bar{R} \left[ 1 - 3 \frac{d'_3}{d'_2} \sqrt{2(1 - \hat{P}_X)} \right] = V_L \bar{R},$$

where

$$\hat{P}_X = \frac{\sum_{i=1}^m \sum_{j=1}^n \delta(\bar{X} - X_{ij})}{m \cdot n}$$

with

$$\delta(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

$n$  is the sample size and  $m$  is the number of preliminary samples. Therefore,  $\hat{P}_X$  is the proportion of observations from pre-run stage that are less than or equal to the estimated process mean. Bai and Choi (1995) used  $\hat{P}_X$  as

an estimator of the measure of the degree of skewness of the process. They claimed that, if the underlying distribution is symmetric, then  $P_X = 0.5$ . If the population is positively or negatively skewed then  $P_X$  becomes greater than 0.5 or less than 0.5, respectively. The constants  $d'_2$  and  $d'_3$  are the mean and the standard deviation of  $W = \frac{R}{\sigma}$  for the given skewed population.

Bai and Choi investigated five different families of distributions: Weibull, lognormal and three different forms of distributions from Burr's family. For each of the five families, they selected the proper parameter values, such that  $P_X$  is equal to a fixed quantity. For each value of  $P_X$  considered, they computed  $d'_2$  and  $d'_3$  via numerical integration for each distribution. They found that the values of  $d'_2$  and  $d'_3$  are similar across the distributions for each given  $P_X$ , so they took the average of those  $d'_2$  and  $d'_3$  as constants to construct tables for  $V_L$  and  $V_U$  for selected values of  $n$  and  $P_X$ .

Although Bai and Choi only considered five families of distributions in the computation of  $d'_2$  and  $d'_3$ , they used the results to apply to all skewed distributions. Due to the limited choices of the skewed distributions, one may suspect that any distribution other than those considered, even with the same skewness but different kurtosis, may produce different constants  $d'_2$  and  $d'_3$ . Furthermore, Chan and Cui (2003) raised the question of using  $P_X$  as a measure of the degree of skewness in the WV method. They indicated that many skewed distributions may have a  $P_X$  value of 0.5, which leads to an incorrect control charting procedure.

To correct the skewness problem produced by WV method, Chan and Cui (2003) proposed the Skewness Correction (SC) method to construct R-control charts for skewed process distributions. The two control limits for SC R-chart are:

$$UCL = \bar{R} \left[ 1 + (3 + d_4^*) \frac{d_3^*}{d_2^*} \right] = D_4^* \bar{R}$$

and

$$LCL = \bar{R} \left[ 1 + (-3 + d_4^*) \frac{d_3^*}{d_2^*} \right]^+ = D_3^* \bar{R}$$

where chart constants  $d_2^*$  and  $d_3^*$ , as  $d_2$  and  $d_3$  in Shewhart control charts for the normal distribution, are defined as the mean and standard deviation of the relative range  $\frac{R}{\sigma}$ ,

$$a^+ = \begin{cases} a & \text{for } a \geq 0 \\ 0 & \text{for } a < 0 \end{cases},$$

$$d_4^* = \frac{\frac{4}{3} k_3(R)}{1 + 0.2 k_3^2(R)},$$

where  $k_3(R)$  is the coefficient of skewness of the sample range  $R$ . The values of  $d_2^*$ ,  $d_3^*$  and  $d_4^*$  can be obtained directly through numerical integration if the process distribution and sample size are specified.

In Chan and Cui's (2003) research, a collection of Weibull, lognormal, and four forms of distributions from the Burr's family are considered as representatives of all skewed distributions. The values of  $d_2^*$ ,  $d_3^*$  and  $d_4^*$  are computed for selected values of  $k_3$ , the skewness of the distribution, in each family of distributions. Due to the similar values of the six  $d_2^*$ ,  $d_3^*$  and  $d_4^*$  across the distributions for each given  $k_3$ , Chan and Cui took the averages of these  $d_2^*$ ,  $d_3^*$  and  $d_4^*$  as constants for all the skewed distribution with a given  $k_3$  to construct tables of  $D_4^*$  and  $D_3^*$  for various sample sizes. Skewness of the distribution  $k_3$  is estimated by the sample skewness

## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

$$k_3^* = \frac{1}{nm-1} \sum_{i=1}^m \sum_{j=1}^n \left( \frac{X_{ij} - \bar{X}}{S_{nm}} \right)^3,$$

where

$$\bar{X} = \frac{1}{nm} \sum_{i=1}^m \sum_{j=1}^n X_{ij},$$

and

$$S_{nm} = \sqrt{\frac{1}{nm-1} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X})^2}.$$

Although the authors introduced the skewness correction to resolve the problem with  $P_x$ , the other potential problem is still unsolved. The tables are constructed based on three families with coefficient of skewness ranging from  $-4$  to  $4$ . It would be problematic for the practitioner to determine the control limits if the estimated coefficient of skewness is outside of this range. For example, the Weibull distribution with unit scale parameter and  $0.5$  shape parameter has  $k_3 = 6.62$ .

In a real life situation, it is more important to detect upper sustained shift than the lower shift in the process variability because the goal of statistical process control is to reduce the variability in the process as much as possible, the upper limit becomes more critical. As noted, it is common that the data has a non-normal underlying distribution; hence, the goal of this study is to develop an upper control chart for controlling the variability of the process that will work for any non-normal distribution, including both skewed and symmetric distributions.

### Methodology

Long and Sa (2005) proposed a method that uses Edgeworth expansions to perform a hypothesis test for the variance for non-normally distributed populations. Their test controls type I error rates well and has power performance comparable to tests that have been developed in the past. The proposed control chart is an adaptation of their test.

Edgeworth expansion is an approximation to the distribution of the estimate

$\hat{\theta}$  of the unknown quantity  $\theta_0$ . If  $\sqrt{n}(\hat{\theta} - \theta_0)$  is asymptotically normally distributed with mean zero and variance  $\sigma^2$ , then the distribution function of  $\sqrt{n}(\hat{\theta} - \theta_0)$  may be expanded as a power series in  $\sqrt{n}$  (Hall, 1992):

$$P \left\{ \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sigma} \leq x \right\} = \Phi(x) + n^{-\frac{1}{2}} p_1(x) \phi(x) + \dots + n^{-\frac{j}{2}} p_j(x) \phi(x) + \dots,$$

where  $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$  is the standard normal density function and  $\Phi(x) = \int_{-\infty}^x \phi(u) du$  is the standard normal distribution function. The functions  $p_j(x)$  are polynomials with coefficients depending on cumulants of  $\hat{\theta} - \theta_0$ .

Hall (1992) provided the Edgeworth expansion for the sample variance

$$P \left\{ \frac{\sqrt{n}(S^2 - \sigma^2)}{\tau} \leq x \right\} = \Phi(x) + n^{-\frac{1}{2}} p_1(x) \phi(x) + \dots + n^{-\frac{j}{2}} p_j(x) \phi(x) + \dots,$$

where

$$p_1 = - \left( B_1 + B_2 \frac{x^2 - 1}{6} \right),$$

$$B_1 = -(\nu_4 - 1)^{\frac{1}{2}},$$

$$B_2 = (\nu_4 - 1)^{-\frac{3}{2}} (\nu_6 - 3\nu_4 - 6\nu_3^2 + 2),$$

$$\nu_j = E \left[ \frac{X - \mu}{\sigma} \right]^j,$$

and

$$\tau = \sqrt{E(X - \mu)^4 - \sigma^4},$$

where  $\tau/\sqrt{n}$  is the standard deviation of the estimator  $S^2$ .

The variable  $S^2$  admits the inversion of Edgeworth expansion as follows:

$$P\left\{\frac{\sqrt{n}(S^2 - \sigma^2)}{\tau} \leq x + n^{-\frac{1}{2}}\left(B_1 + B_2 \frac{x^2 - 1}{6}\right)\right\} = \Phi(x) + o(n^{-\frac{1}{2}}).$$

Long and Sa (2005) adapted the inversion formula of the Edgeworth Expansion to test  $H_0 : \sigma^2 = \sigma_0^2$  versus  $H_a : \sigma^2 > \sigma_0^2$ . An intuitive decision rule is to reject  $H_0$  if

$$Z > z_\alpha + n^{-\frac{1}{2}}\left(B_1 + B_2 \frac{z_\alpha^2 - 1}{6}\right), \quad (1)$$

where  $z_\alpha$  is the upper  $\alpha$  percentage point of the standard normal distribution and

$$Z = \frac{S^2 - \sigma_0^2}{\tau/\sqrt{n}}.$$

They first estimated  $B_1$  and  $B_2$  by

$$\hat{B}_1 = -\left(\frac{S^4}{\kappa_4 + 2S^4}\right)^{\frac{1}{2}}$$

and

$$\hat{B}_2 = \frac{\kappa_6 + 12\kappa_4 S^2 + 4\kappa_3^2 + 8(S^2)^3}{(\kappa_4 + 2S^4)^{\frac{3}{2}}},$$

where  $\kappa_3, \kappa_4$ , and  $\kappa_6$  are the third, fourth and sixth sample cumulants, respectively. They then investigated six different forms of  $Z$  and found that

$$Z6 = \frac{S^2 - \sigma_0^2}{\sqrt{\frac{\kappa_4 \sigma_0^2}{nS^2} + \frac{2\sigma_0^4}{n-1}}} \quad (2)$$

yields the best results for controlling the type I error rates as well as satisfying power performance; their final decision rule is to reject  $H_0$  if:

$$Z6 > z_\alpha + n^{-\frac{1}{2}}\left(\hat{B}_1 + \hat{B}_2 \frac{z_\alpha^2 - 1}{6}\right). \quad (3)$$

The decision rule (3) of Long and Sa (2005) can be used in the construction of the upper-sided control chart for variability with some modifications. Population variance  $\sigma_0^2$  can be estimated in the preliminary stage by the sample variance  $\tilde{S}^2$ .  $\hat{B}_1$  and  $\hat{B}_2$  can also be calculated based on the preliminary samples.

The upper control limit can then be set as:

$$UCL = z_\alpha + n^{-\frac{1}{2}}\left(\hat{B}_1 + \hat{B}_2 \frac{z_\alpha^2 - 1}{6}\right). \quad (4)$$

In the control charting stage, a sample is selected and

$$Z6 = \frac{S_{(i)}^2 - \tilde{S}^2}{\sqrt{\frac{\kappa_{4,(i)} \tilde{S}^2}{nS_{(i)}^2} + \frac{2\tilde{S}^4}{n-1}}}, \quad (5)$$

where  $S_{(i)}^2$  and  $\kappa_{4,(i)}$  are variance and fourth cumulant of the  $i^{th}$  sample, is calculated. An out-of-control signal occurs when:

$$Z6 > UCL. \quad (6)$$

The proposed Variability Control Chart can be constructed as follows:

1. Based on process requirements, select a significance level  $\alpha$  and find the corresponding critical point  $z_\alpha$ ;
2. Assuming the process is under statistical control, select  $m$  preliminary samples of size  $n$  to calculate all the process quantities ( $\tilde{S}^2, \kappa_3, \kappa_4, \kappa_6, \hat{B}_1, \hat{B}_2$ ). Two methods are employed to calculate these quantities. The

## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

first is called the combined sample method, which merges all  $m$  samples in the preliminary run as one sample with  $m \cdot n$  observations to compute the process quantities. The second is the not combined sample method in which all the process quantities are equal to the averages of the  $m$  corresponding preliminary sample values.

3. Calculate  $UCL$  using (4);
4. Start the quality control stage. Select samples of size  $n$  periodically. After the  $i^{th}$  sample is selected, calculate the sample variance  $S_{(i)}^2$  and sample cumulant  $\kappa_{4,(i)}$ ;
5. Plug them into (5) to get the sample point  $Z6$  for this sample;
6. Plot the sample point  $Z6$  on the chart and draw the conclusion about this sample (in-control or out-of-control);
7. If the process is in-control, then go back to step 4 to select next sample; otherwise quality control team should investigate and possibly remove the assignable causes.

### Simulation Study

In order to compare different control charts for variability of a process, a simulation study to investigate the type I error rates and power performance is performed. The methods compared include the Shewhart R-chart, S-chart,  $S^2$ -chart, WV R-chart, SC R-chart and the proposed control charts.

### Distributions Examined

A large collection of distributions with a wide range of skewness and kurtosis are investigated via a simulation study. Distributions considered are separated into two groups: skewed and symmetric.

The skewed family includes eight Weibull distributions with scale parameter  $\lambda = 1$  and shape parameter  $k$  from 0.5 to 3.5; exponential with  $\lambda = 1$ ; Gamma with scale parameter  $\beta = 1$  and shape parameters

$\alpha = 0.15, 1.2$  and  $4.0$ ; Chi-square with  $\nu$  degrees of freedom ( $\nu = 1$  to 24); lognormal with  $\mu = 0$  and  $\sigma^2 = 1$ ; and the Barnes2 distribution which is a polynomial function of the standard normal distribution (Fleishman, 1978). For comparison purposes, the standard normal distribution is also studied and reported.

The symmetric distributions considered include: Student's T ( $\nu = 5, 6, 8, 16, 32, 40$ ), JTB ( $\mu = 0, \sigma = 1, \alpha, \tau$ ) with various  $\alpha$  and  $\tau$  (Johnson, Tietjen & Beckman, 1980) and special designed distributions Barnes1 and Barnes3 with respective kurtosis 6.89 and 1049 (Fleishman, 1978). All the distributions in this group are symmetric with the exception of Barnes3 with a coefficient of skewness of 3.00, which is negligible in comparison to its kurtosis of 1049.

Random number generators from the Fortran 90 IMSL library are used to generate normal, Weibull, exponential, lognormal, Chi-square, Gamma and Student's T variates. In addition, the Barnes1, Barnes2, Barnes3 and JTB random variates were created with Fortran 90 program subroutines using IMSL library's random number generators for standard normal, gamma and uniform distributions in various parts of the program.

### Simulation Description

The simulation study includes two parts: (1). Process is in-control (type I error rate comparisons) and (2). Process is shifted (the power study). In both studies, the process parameters are assumed unknown and therefore need to be estimated. The number of samples used in the preliminary run is  $m = 30$ ; a relatively small sample size of 10 and a moderate sample size of 25 are used in the study. The steps of the simulation take place in two parts: steps 1 – 4 are preliminary runs and steps 5 – 9 are the quality control stages.

Preliminary Runs (assumes the process is in-control):

1. Set up the nominal level  $\alpha = 0.0027$  (which corresponds to the frequently used Average Run Length, ARL, = 370) and select the parent distribution;

2. Generate 30 samples of size  $n$  from the parent distribution;
3. Calculate the necessary quantities used in different methods:  
 $\bar{R}, \bar{S}, \bar{S}^2, \hat{P}_X, k_3, \hat{B}_1, \hat{B}_2, \tilde{S}^2, \kappa_3, \kappa_4, \kappa_6$ ;  
 both the combined sample method and not combined sample method are used to calculate the process quantities for the proposed methods.
4. Calculate the appropriate upper control limit for each of the control methods; the control limits of the Shewhart R-chart, S-chart, S<sup>2</sup>-chart, WV R-chart and SC R-chart are adjusted to meet the purpose of the comparisons. In order to achieve the desired nominal level of  $\alpha = 0.0027$  for a one-sided control chart,  $z_\alpha = 2.78215$  is used to construct the appropriate upper control limits for all the methods.

The Quality Control Stage: Steps (5) – (9)

5. Generate 1,000 samples of size  $n$  from the same parent distribution and calculate the statistic to be plotted for each of the control methods (sample range  $R$  for the Shewhart, WV and SC R-charts, sample standard deviation  $S$  for S-chart, sample variance for S<sup>2</sup>-char, and Z6 for the proposed method);
6. Compare the statistic with the corresponding control limits and tabulate the number of out-of-control signals;
7. Calculate type I error rate for each method by finding the proportion of out-of-control signals in the 1000 samples;
8. Repeat steps (2) – (7) 4,000 times;
9. Calculate the average of 4,000 type I error rates.

In the power study each generated variate is multiplied by a pre-determined  $\sqrt{k}$ , where  $k = 1, 2, 3, 4, 5, 6$ ; thus, a new set of observations is created with variance  $k$  times

larger than the variance of the original distribution. Steps (5) – (9) are then repeated for each value of  $k$  to investigate the power of each charting technique. The corresponding ARL can be calculated for an in-control or an out-of-control process by inverting the average type I error rate or power from step (9).

Results

The Study of Type I Error Rates

Tables 1 through 4 provide comparisons of type I error rates for skewed and symmetric distributions with sample sizes  $n = 10$  and  $n = 25$ . The first and the second columns are the type I error rates of the proposed method Z6 using the combined sample and the not combined sample methods in the calculations of the process quantities. Three critical points  $z_\alpha$ ,  $\frac{z_\alpha + t_{n-1, \alpha}}{2}$  and  $t_{n-1, \alpha}$  are used in construction of the upper control limits for the proposed method with sample size  $n = 10$ ; results shown are the first, second and third numbers in the respective column.

Skewed Distributions

Table 1 shows that all traditional control charts (the Shewhart R-chart, S-chart and S<sup>2</sup>-chart) fail to maintain the type I error rates under nominal level  $\alpha = 0.0027$  when the parent distribution is skewed. In general, the larger the degree of skewness, the bigger the type I error rate. For example, considering a  $\chi^2_{(24)}$  distribution with skewness 0.58, the type I error rates of the three traditional charts are 0.0235, 0.0224 and 0.00566 with corresponding ARLs 42.55, 44.64 and 176.68 for the Shewhart R-chart, S-chart and S<sup>2</sup>-chart respectively. Those rates change to 0.124, 0.132 and 0.0624 with respective ARLs 8.06, 7.58 and 16.03 when the parent distribution is  $\chi^2_{(1)}$ , with a more severe skewness of 2.83.

Among the three traditional charts, the S<sup>2</sup>-chart tends to outperform the other two, however, it still consistently yields inflated type I error rates which result in very short ARLs. It usually performs reasonably well for distributions with low skewness. The best cases



## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 1: Skewed Distributions, Comparisons of Type I Error Rates when  $n = 10$

Distribution (skewness) (kurtosis)	Combined Sample	Not Combined	R-chart	S-chart	S <sup>2</sup> -chart	WV-chart	SC-chart
Normal (0,1)	3.97E-03	1.09E-02	1.45E-02	1.24E-02	1.86E-03	1.24E-02	6.40E-03
(0.00)	1.23E-03	9.36E-03					
(0.00)	3.58E-04	9.77E-03					
Exponential (1)	5.82E-03	4.71E-02	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03
(2.00)	2.73E-03	5.00E-02					
(6.00)	1.31E-03	5.39E-02					
Lognormal (0,1)	5.80E-03	4.19E-02	1.37E-01	1.44E-01	7.61E-02	2.52E-02	2.33E-02
(6.18)	2.85E-03	4.23E-02					
(110.93)	1.48E-03	4.38E-02					
Weibull (0.5)	5.83E-03	4.10E-02	1.66E-01	1.75E-01	8.74E-02	3.11E-02	3.07E-02
(6.62)	2.82E-03	4.28E-02					
(84.72)	1.48E-03	4.48E-02					
Weibull (0.75)	5.50E-03	4.64E-02	1.27E-01	1.35E-01	6.62E-02	1.18E-02	1.30E-02
(3.12)	2.69E-03	4.85E-02					
(16.0)	1.38E-03	5.10E-02					
Weibull (0.85)	5.70E-03	4.70E-02	1.10E-01	1.17E-01	5.56E-02	1.14E-02	1.02E-02
(2.56)	2.76E-03	4.95E-02					
(10.35)	1.41E-03	5.26E-02					
Weibull (1)	5.82E-03	4.71E-02	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03
(2.00)	2.73E-03	5.00E-02					
(6.00)	1.31E-03	5.39E-02					
Weibull (1.2)	5.70E-03	4.56E-02	6.24E-02	6.61E-02	2.59E-02	1.06E-02	5.52E-03
(1.52)	2.53E-03	4.89E-02					
(3.24)	1.14E-03	5.34E-02					
Weibull (1.5)	5.10E-03	4.13E-02	3.73E-02	3.91E-02	1.22E-02	8.87E-03	3.90E-03
(1.07)	2.05E-03	4.48E-02					
(1.39)	8.18E-04	4.97E-02					
Weibull (2.0)	4.08E-03	2.85E-02	1.71E-02	1.77E-02	3.65E-03	7.41E-03	2.70E-03
(0.63)	1.41E-03	3.12E-02					
(0.25)	4.66E-04	3.57E-02					
Weibull (3.5)	3.36E-03	1.07E-02	6.53E-03	7.07E-03	6.93E-04	5.43E-03	2.10E-03
(0.03)	9.14E-04	9.33E-03					
(-0.29)	2.13E-04	1.00E-02					
Barnes 2	7.25E-03	4.86E-02	7.41E-02	7.82E-02	2.95E-02	4.77E-03	3.90E-03
(1.75)	3.39E-03	5.19E-02					
(3.75)	1.63E-03	5.69E-02					

Table 1 (Continued): Skewed Distributions, Comparisons of Type I Error Rates when  $n = 10$

Distribution (skewness) (kurtosis)	Combined Sample	Not Combined	R-chart	S-chart	S <sup>2</sup> -chart	WV-chart	SC-chart
Chi (1)	6.31E-03	5.13E-02	1.24E-01	1.32E-01	6.24E-02	7.89E-03	1.06E-02
(2.83)	3.13E-03	5.48E-02					
(12.0)	1.64E-03	5.84E-02					
Chi (2)	5.82E-03	4.71E-02	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03
(2.00)	2.73E-03	5.00E-02					
(6.00)	1.31E-03	5.39E-02					
Chi (3)	5.72E-03	4.46E-02	6.86E-02	7.28E-02	3.05E-02	1.27E-02	6.74E-03
(1.63)	2.60E-03	4.75E-02					
(4.00)	1.19E-03	5.10E-02					
Chi (4)	5.44E-03	4.30E-02	5.83E-02	6.14E-02	2.43E-02	1.32E-02	6.54E-03
(1.41)	2.42E-03	4.56E-02					
(3.00)	1.07E-03	4.91E-02					
Chi (8)	4.87E-03	3.81E-02	3.93E-02	4.03E-02	1.37E-02	1.30E-02	6.03E-03
(1.00)	1.99E-03	4.12E-02					
(1.50)	8.23E-04	4.51E-02					
Chi (10)	4.67E-03	3.48E-02	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03
(0.89)	1.88E-03	3.76E-02					
(1.20)	7.45E-04	4.15E-02					
Chi (12)	4.52E-03	3.37E-02	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03
(0.82)	1.77E-03	3.64E-02					
(1.00)	6.77E-04	4.05E-02					
Chi (16)	4.45E-03	2.62E-02	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03
(0.71)	1.71E-03	2.76E-02					
(0.75)	6.39E-04	3.01E-02					
Chi (24)	4.22E-03	2.51E-02	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03
(0.58)	1.54E-03	2.64E-02					
(0.50)	5.42E-04	2.95E-02					
Gamma (0.15)	6.00E-03	4.54E-02	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02
(5.16)	2.99E-03	4.71E-02					
(40.0)	1.61E-03	4.90E-02					
Gamma (1.2)	6.07E-03	5.12E-02	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03
(1.83)	2.87E-03	5.46E-02					
(5.00)	1.34E-03	5.88E-02					
Gamma (4.0)	4.95E-03	3.97E-02	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03
(1.00)	2.07E-03	4.22E-02					
(1.50)	8.29E-04	4.58E-02					

## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 2: Skewed Distributions, Comparisons of Type I Error Rates when  $n = 25$

Distribution	(skewness) (kurtosis)	Combined Sample	Not Combined	R-chart	S-chart	S2-chart	WV-chart
Normal (0,1)	(0.00) (0.00)	3.41E-03	3.50E-03	1.83E-02	1.64E-02	1.82E-03	1.70E-02
Exponential (1)	(2.00) (6.00)	2.59E-03	3.03E-03	1.09E-01	1.19E-01	5.32E-02	8.16E-03
Lognormal (0,1)	(6.18) (110.93)	1.72E-03	2.51E-03	1.68E-01	1.85E-01	1.05E-01	2.49E-02
Weibull (0.5)	(6.62) (84.72)	1.97E-03	2.98E-03	1.97E-01	2.20E-01	1.24E-01	2.84E-02
Weibull (0.75)	(0.12) (1.23)	1.67E-03	2.22E-03	1.53E-01	1.71E-01	9.10E-02	9.19E-03
Weibull (0.85)	(2.56) (10.35)	2.02E-03	2.51E-03	1.35E-01	1.49E-01	7.49E-02	8.71E-03
Weibull (1)	(2.00) (6.00)	2.59E-03	3.03E-03	1.09E-01	1.19E-01	5.32E-02	8.16E-03
Weibull (1.2)	(1.52) (3.24)	3.18E-03	3.61E-03	8.06E-02	8.47E-02	3.16E-02	8.08E-03
Weibull (1.5)	(1.07) (1.39)	3.50E-03	3.88E-03	4.98E-02	5.03E-02	1.35E-02	8.91E-03
Weibull (2.0)	(0.63) (0.25)	3.20E-03	3.45E-03	2.27E-02	2.30E-02	3.54E-03	9.28E-03
Weibull (3.5)	(0.03) (-0.29)	2.72E-03	2.77E-03	6.67E-03	9.91E-03	7.08E-04	6.08E-03
Chi (1)	(2.83) (12.0)	2.10E-03	2.60E-03	1.46E-01	1.65E-01	8.38E-02	5.01E-03
Chi (2)	(2.00) (6.00)	2.59E-03	3.03E-03	1.09E-01	1.19E-01	5.32E-02	8.16E-03
Chi (3)	(1.63) (4.00)	2.78E-03	3.24E-03	8.80E-02	9.35E-02	3.77E-02	9.73E-03
Chi (4)	(1.41) (3.00)	2.84E-03	3.31E-03	7.58E-02	7.88E-02	2.92E-02	1.17E-02
Chi (8)	(1.00) (1.50)	2.90E-03	3.26E-03	5.17E-02	5.15E-02	1.51E-02	1.47E-02
Chi (10)	(0.89) (1.20)	2.92E-03	3.27E-03	4.61E-02	4.49E-02	1.22E-02	1.56E-02
Chi (12)	(0.82) (1.00)	3.03E-03	3.32E-03	4.24E-02	4.12E-02	1.06E-02	1.63E-02
Chi (16)	(0.71) (0.75)	3.03E-03	3.28E-03	3.71E-02	3.56E-02	8.25E-03	1.69E-02
Chi (24)	(0.58) (0.50)	3.16E-03	3.43E-03	3.14E-02	2.95E-02	5.79E-03	1.70E-02
Gamma (0.15)	(5.16) (40.0)	1.52E-03	2.26E-03	2.04E-01	2.36E-01	1.30E-01	1.77E-02
Gamma (1.2)	(1.83) (5.00)	2.56E-03	3.06E-03	9.83E-02	1.06E-01	4.51E-02	8.31E-03
Gamma (4.0)	(1.00) (1.50)	2.97E-03	3.38E-03	5.19E-02	5.15E-02	1.51E-02	1.46E-02
Barnes 2	(1.75) (3.75)	4.09E-03	4.73E-03	8.85E-02	9.66E-02	3.55E-02	2.08E-03

Table 3: Symmetric Distributions, Comparisons of Type I Error Rates when  $n = 10$

Distribution	Combined Sample	Not Combined	R-chart	S-chart	S <sup>2</sup> -chart	WV-chart	SC-chart
(skewness)							
(kurtosis)							
Normal (0,1)	3.97E-03	1.09E-02	1.45E-02	1.24E-02	1.86E-03	1.24E-02	6.40E-03
(0.00)	1.23E-03	9.36E-03					
(0.00)	3.58E-04	9.77E-03					
JTB (2.0, 1.0)	4.58E-03	1.50E-02	6.60E-02	6.30E-02	2.39E-02	5.77E-02	4.23E-02
(0.00)	1.89E-03	1.39E-02					
(3.00)	7.78E-04	1.46E-02					
JTB (0.75, 0.5)	5.69E-03	1.37E-02	3.93E-02	3.62E-02	9.57E-03	3.33E-02	2.21E-02
(0.00)	2.22E-03	1.10E-02					
(1.20)	8.52E-04	1.06E-02					
JTB (4.0, 1.0)	4.33E-03	1.43E-02	3.31E-02	2.78E-02	7.42E-03	2.95E-02	1.88E-02
(0.00)	1.59E-03	1.31E-02					
(0.78)	5.71E-04	1.35E-02					
JTB (1.0, 0.5)	5.08E-03	1.30E-02	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02
(0.00)	1.83E-03	1.07E-02					
(0.60)	6.38E-04	1.09E-02					
JTB (1.25, 0.5)	4.47E-03	1.30E-02	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03
(0.00)	1.52E-03	1.13E-02					
(0.24)	4.69E-04	1.17E-02					
JTB (1.35, 0.5)	4.19E-03	1.12E-02	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03
(0.00)	1.38E-03	9.34E-03					
(0.13)	4.11E-04	9.76E-03					
JTB (1.5, 0.5)	4.00E-03	1.18E-02	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03
(0.00)	1.26E-03	1.01E-02					
(0.00)	3.70E-04	1.05E-02					
JTB (2.0, 0.5)	3.22E-03	1.26E-02	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03
(0.00)	9.07E-04	1.18E-02					
(-0.30)	2.43E-04	1.29E-02					
JTB (4.0, 0.5)	1.65E-03	1.22E-02	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04
(0.00)	3.43E-04	1.28E-02					
(-0.75)	6.33E-05	1.52E-02					
JTB (9.0, 0.5)	8.09E-04	1.86E-02	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06
(0.00)	1.28E-04	2.27E-02					
(-1.00)	1.33E-05	2.92E-02					

A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 3 (continued): Symmetric Distributions, Comparisons of Type I Error Rates when  $n = 10$

Distribution	Combined Sample	Not Combined	R-chart	S-chart	S <sup>2</sup> -chart	WV-chart	SC-chart
(skewness)							
(kurtosis)							
Barnes 3	2.04E-03	1.73E-02	1.10E-01	9.09E-02	6.07E-02	1.06E-01	7.76E-02
(3.00)	8.02E-04	1.71E-02					
(1049)	3.41E-04	1.77E-02					
Barnes 1	3.97E-03	1.09E-02	1.45E-02	1.24E-02	1.86E-03	1.24E-02	6.40E-03
(0.00)	1.23E-03	9.36E-03					
(6.89)	3.58E-04	9.77E-03					
Student (5)	3.27E-03	1.66E-02	6.24E-02	5.61E-02	2.52E-02	5.76E-02	4.17E-02
(0.00)	1.27E-03	1.58E-02					
(6.00)	5.04E-04	1.66E-02					
Student (6)	3.41E-03	1.43E-02	5.31E-02	4.67E-02	1.90E-02	4.87E-02	3.44E-02
(0.00)	1.31E-03	1.35E-02					
(3.00)	5.15E-04	1.40E-02					
Student (8)	3.79E-03	1.44E-02	4.23E-02	3.63E-02	1.27E-02	3.84E-02	2.62E-02
(0.00)	1.41E-03	1.35E-02					
(1.50)	5.15E-04	1.39E-02					
Student (16)	4.09E-03	1.32E-02	2.68E-02	2.22E-02	5.48E-03	2.38E-02	1.46E-02
(0.00)	1.42E-03	1.22E-02					
(0.50)	4.80E-04	1.27E-02					
Student (25)	4.20E-03	1.15E-02	2.22E-02	1.83E-02	3.86E-03	1.94E-02	1.13E-02
(0.00)	1.41E-03	9.63E-03					
(0.29)	4.68E-04	9.64E-03					
Student (32)	4.01E-03	9.66E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03
(0.00)	1.35E-03	7.51E-03					
(0.21)	4.24E-04	7.32E-03					
Student (40)	4.08E-03	1.26E-02	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03
(0.00)	1.34E-03	1.09E-02					
(0.17)	4.14E-04	1.11E-02					

Table 4: Symmetric Distributions, Comparisons of Type I Error Rates when  $n = 25$

Distribution	(skewness) (kurtosis)	Combined Sample	Not Combined	R-chart	S-chart	S2-chart	WV-chart
Normal (0,1)	(0.00) (0.00)	3.41E-03	3.50E-03	1.83E-02	1.64E-02	1.82E-03	1.70E-02
JTB (2.0, 1.0)	(0.00) (3.00)	2.31E-03	2.51E-03	8.16E-02	7.97E-02	2.84E-02	7.63E-02
JTB (0.75, 0.5)	(0.00) (1.20)	4.42E-03	4.57E-03	4.41E-02	4.45E-02	9.86E-03	4.03E-02
JTB (4.0, 1.0)	(0.00) (0.78)	3.26E-03	3.38E-03	4.47E-02	3.55E-02	7.69E-03	4.24E-02
JTB (1.0, 0.5)	(0.00) (0.60)	4.36E-03	4.48E-03	3.18E-02	3.05E-02	5.21E-03	2.97E-02
JTB (1.25, 0.5)	(0.00) (0.24)	3.81E-03	3.96E-03	2.36E-02	2.16E-02	2.86E-03	2.20E-02
JTB (1.35, 0.5)	(0.00) (0.13)	3.72E-03	3.83E-03	2.15E-02	1.97E-02	2.39E-03	2.00E-02
JTB (1.5, 0.5)	(0.00) (0.00)	3.36E-03	3.46E-03	1.80E-02	1.63E-02	1.77E-03	1.68E-02
JTB (2.0, 0.5)	(0.00) (-0.30)	2.42E-03	2.50E-03	1.11E-02	9.94E-03	7.17E-04	1.02E-02
JTB (4.0, 0.5)	(0.00) (-0.75)	9.33E-04	9.70E-04	2.29E-03	3.03E-03	8.20E-05	2.02E-03
JTB (9.0, 0.5)	(0.00) (-1.00)	3.23E-04	3.33E-04	1.01E-04	8.21E-04	8.25E-06	8.53E-05
Barnes 3	(3.00) (1.49)	6.14E-04	9.31E-04	1.91E-01	1.47E-01	9.66E-02	1.88E-01
Barnes 1	(0.00) (6.89)	3.41E-03	3.50E-03	1.83E-02	1.64E-02	1.82E-03	1.70E-02
Student (5)	(0.00) (6.00)	1.30E-03	1.42E-03	9.00E-02	7.69E-02	3.36E-02	8.69E-02
Student (6)	(0.00) (3.00)	1.64E-03	1.75E-03	7.67E-02	6.35E-02	2.45E-02	7.39E-02
Student (8)	(0.00) (1.50)	2.32E-03	2.46E-03	6.11E-02	4.82E-02	1.50E-02	5.86E-02
Student (16)	(0.00) (0.50)	3.16E-03	3.29E-03	3.76E-02	2.88E-02	5.56E-03	3.56E-02
Student (25)	(0.00) (0.29)	3.28E-03	3.43E-03	2.95E-02	2.32E-02	3.64E-03	2.78E-02
Student (32)	(0.00) (0.21)	3.42E-03	3.50E-03	2.72E-02	2.17E-02	3.17E-03	2.56E-02
Student (40)	(0.00) (0.17)	3.35E-03	3.44E-03	2.50E-02	2.04E-02	2.75E-03	2.35E-02

## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

other than the standard normal distribution produced by  $S^2$ -chart are for the Weibull(2.0) with skewness 0.63 and Weibull(3.5) with skewness 0.03 which result respective type I error rates 0.00365 and 0.000693 equivalent to ARLs 273.97 and 1443. However, when skewness increases, such as  $\chi_{(8)}^2$  with skewness of 1.00, the performance goes down dramatically with type I error rate 0.0137 and  $ARL = 72.99$ .

The WV R-chart is also unable to maintain the type I error rate for skewed distributions; although it works well for a few distributions, in general it produces false alarms too often. The SC R-chart has better performances among the existing variability control charts. It shows a degree of robustness when the coefficient of skewness is small, but if the skewness becomes somewhat severe, it fails to keep the type I error rates close to the nominal level. For example, the SC R-chart produces type I error rates of 0.0233 and 0.0307 with corresponding ARL of 42.92 and 32.57 for the standard lognormal with skewness = 6.18 and Weibull (0.5) with skewness = 6.62.

For the proposed method, results show that the combined sample method, which merges all the samples in the preliminary runs as one large sample to compute the process quantities, consistently outperforms all the other methods with very few exceptions. The worst case is for the Barnes2 distribution with skewness 1.75. It produces the highest type I error rate of 0.0075 with a corresponding ARL 137.93 when  $z_\alpha$  is used as the critical point. However, it drops to 0.00339 with ARL 294.99 when  $\frac{z_\alpha + t_{n-1,\alpha}}{2}$  is used. When  $t_{n-1,\alpha}$  is used as a critical point, the proposed method becomes too conservative, which is not recommended because it will become more difficult to detect shifts if present.

When a larger sample size  $n = 25$  is used in the simulation study (see Table 2), the performances of the Shewhart R-chart, S-chart,  $S^2$ -chart and WV R-chart do not change much. Type I errors rates for these charts are still inflated for distributions with high degrees of skewness such as the standard lognormal and

Weibull with shape = 0.5, etc. Conversely, the proposed method with combined sample produces type I error rates close to the nominal level even with  $z_\alpha$  as the critical point. The highest type I error rate produced by the proposed method is 0.00409 ( $ARL = 244.5$ ) for the Barnes2 distribution.

Table 1 shows that the proposed method with combined sample can also be used for the standard normal distribution. The type I error rates produced are smaller than those of all the charts except the  $S^2$ -chart, even though it is not designed for the normal distribution. This nice performance adds another desirable property to the proposed method.

Note that the SC R-chart is not used in the simulation study with sample size  $n = 25$  because Chan and Cui (2003) do not provide constants for calculations of the control limits for any sample size larger than 10. It is extremely difficult for the practitioners to implement this control chart if the situation requires collecting a sample size larger than 10.

### Symmetric Distributions

Table 3 provides type I error rate comparisons for the symmetric distributions with sample size 10. The proposed method is the only one that holds the type I error rates almost all the time. Although some of the type I error rates for the proposed method are a little higher than 0.0027, they are all within an acceptable range. The worst case found in the study is for the JTB distribution ( $\alpha = 0.75$ ,  $\tau = 0.5$ ) with kurtosis 1.2 using  $z_\alpha$  as a critical point producing the lowest  $ARL = 175.75$  with type I error rate 0.0057. However, once the critical point is changed to  $\frac{z_\alpha + t_{n-1,\alpha}}{2}$ , the ARL increases to 454.55 with type I error = 0.0022. Again when the critical point  $t_{n-1,\alpha}$  is used, the proposed method becomes unnecessarily conservative.

The two traditional methods, Shewhart R-chart and S-chart, are not robust at all, but the  $S^2$ -chart performs surprisingly well when the kurtosis of the distribution is either very close to zero or negative. However, the good performance soon disappears once the

distribution has a kurtosis larger than 0.5. It is expected that WV and SC methods will not perform very well, because they only try to correct the skewness of the distribution, not the kurtosis.

It can be observed that the type I error rates for all the existing charts are strongly affected by the kurtosis of the distributions. The type I error rate increases when the kurtosis increases. When Barnes3 with kurtosis 1049 is the parent distribution, all the other charting techniques fail. The type I error rates for Shewhart R-chart, S-chart, S<sup>2</sup>-chart, WV R-chart and SC R-chart are 0.11, 0.091, 0.0607, 0.106 and 0.0776 with corresponding ARL 9.09, 11, 16.47, 9.43 and 13.04, respectively.

Table 4 provides type I error rate comparisons for the symmetric distributions with sample size 25. Similar results to those shown in Table 3 are observed in this table. The proposed method is the only one with robust performance. The highest type I error rate is 0.00442 with ARL = 226.24 for JTB ( $\alpha = 0.75$ ,  $\tau = 0.5$ ) with kurtosis = 1.2. All other methods are not able to maintain type I error rates for distributions with kurtosis greater than 0.78. When the coefficient of kurtosis is in negative values, the type I error rates are generally much lower than the desired nominal level; this is observed in all the methods studied except in R-chart which generally fails in nearly all cases.

#### Power Study

The primary goal of the power study is to find the control charts with improved type I error rates and power performance comparable to other charts. It is reasonable to expect that more conservative charts might produce lower power than other charts because it is more difficult to detect an out-of-control state with these charts.

The results of the power study for skewed distributions are presented in Table 5 for sample size 10 with  $z_\alpha$  and  $\frac{z_\alpha + t_{n-1,\alpha}}{2}$  as critical points; results for symmetric distributions are reported in Table 6 for sample size 10. A power study was also conducted for cases with sample size 25. For complete

simulation results, please see Borysov and Sa (2010).

The following similarities in the power performances of all the control charting methods are observed: As sample size increases from 10 to 25, power increases; as  $k$  in  $k\sigma^2$  increases, the power increases; as the skewness of the skewed distribution increases, the power tends to decrease; and as kurtosis of the symmetric distributions increases, the power also tends to decrease.

It can be observed that the power performance of the proposed method is relatively good and is similar to other charts. In the cases of highly skewed distributions with large kurtosis (e.g., standard lognormal with skewness 6.18 and kurtosis 110, Weibull (0.5) with skewness 6.62 and kurtosis 84.72, Gamma (0.15) with skewness 5.16 and kurtosis 40), the power of the proposed method tends to be lower than those of other charts. However, recall that the proposed chart is the only one able to control the type I error rates for those distributions. When the shift in process variability increases

the proposed scheme with  $\frac{z_\alpha + t_{n-1,\alpha}}{2}$  becomes compatible to the WV and SC control charts.

Although the three Shewhart charts generally have higher power than the proposed control chart, it must be restated that power performance of the control chart is useless if it cannot preserve an appropriate type I error rate. Frequent false alarms can create more damage than quick shift detections can benefit. If sample size 25 is used, the proposed method has better power performance than the WV R-chart for almost all the distributions considered, even for small shifts of the variability.

#### Simulation Study Summary

The proposed Variability Control Chart which plots  $Z_6$  against  $UCL$  with combined sample should be used with decision rule (6) in order to achieve controllable type I error rates as well as to detect shifts in variability. It can be implemented for a process with any form of the underlying distribution consisting of skewed and and/or symmetric distributions including normal.



A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 5: Power Comparison Study for Skewed Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart
Normal (0,1)	(0.00)	k=1	3.97E-03	1.24E-02	1.86E-03	1.24E-02	6.40E-03	1.24E-02
	(0.00)		1.23E-03					
Exponential (1)	(2.00)		5.82E-03	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03
	(6.00)		2.73E-03					
Lognormal (0,1)	(6.18)		5.80E-03	1.37E-01	1.44E-01	7.61E-02	2.52E-02	2.33E-02
	(110)		2.85E-03					
Weibull (0.5)	(6.62)		5.83E-03	1.66E-01	1.75E-01	8.74E-02	3.11E-02	3.07E-02
	(84.72)		2.82E-03					
Weibull (0.85)	(0.12)		5.50E-03	1.27E-01	1.35E-01	6.62E-02	1.18E-02	1.30E-02
	(1.23)		2.69E-03					
Weibull (0.75)	(2.56)		5.70E-03	1.10E-01	1.17E-01	5.56E-02	1.14E-02	1.02E-02
	(10.35)		2.76E-03					
Weibull (1)	(2.00)		5.82E-03	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03
	(6.00)		2.73E-03					
Weibull (1.2)	(1.52)		5.70E-03	6.24E-02	6.61E-02	2.59E-02	1.06E-02	5.52E-03
	(3.24)		2.53E-03					
Weibull (1.5)	(1.07)		5.10E-03	3.73E-02	3.91E-02	1.22E-02	8.87E-03	3.90E-03
	(1.39)		2.05E-03					
Weibull (2.0)	(0.63)		4.08E-03	1.71E-02	1.77E-02	3.65E-03	7.41E-03	2.70E-03
	(0.25)		1.41E-03					
Weibull (3.5)	(0.03)	3.36E-03	6.53E-03	7.07E-03	6.93E-04	5.43E-03	2.10E-03	
	(-0.29)	9.14E-04						
Barnes2	(1.75)	7.25E-03	7.41E-02	7.82E-02	2.95E-02	4.77E-03	3.90E-03	
	(3.75)	3.39E-03						
Normal(0,1)	(0.00)	k=2	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01
	(0.00)		1.01E-01					
Exponential (1)	(2.00)		6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02
	(6.00)		3.53E-02					
Lognormal (0,1)	(6.18)		2.54E-02	2.78E-01	2.98E-01	1.75E-01	6.53E-02	6.12E-02
	(110)		1.41E-02					
Weibull (0.5)	(6.62)		2.16E-02	2.89E-01	3.07E-01	1.76E-01	7.48E-02	7.38E-02
	(84.72)		1.18E-02					
Weibull (0.85)	(0.12)		3.69E-02	3.01E-01	3.26E-01	1.99E-01	5.08E-02	5.48E-02
	(1.23)		2.10E-02					
Weibull (0.75)	(2.56)		4.56E-02	2.99E-01	3.27E-01	2.00E-01	5.74E-02	5.35E-02
	(10.35)		2.64E-02					
Weibull (1)	(2.00)		6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02
	(6.00)		3.53E-02					
Weibull (1.2)	(1.52)		8.02E-02	2.84E-01	3.20E-01	1.88E-01	8.99E-02	5.90E-02
	(3.24)		4.74E-02					
Weibull (1.5)	(1.07)		1.09E-01	2.69E-01	3.09E-01	1.73E-01	1.14E-01	7.05E-02
	(1.39)		6.45E-02					
Weibull (2.0)	(0.63)		1.48E-01	2.51E-01	2.96E-01	1.51E-01	1.58E-01	9.67E-02
	(0.25)		8.89E-02					
Weibull (3.5)	(0.03)	1.97E-01	2.45E-01	2.90E-01	1.35E-01	2.28E-01	1.62E-01	
	(-0.29)	1.23E-01						
Barnes2	(1.75)	8.32E-02	3.01E-01	3.35E-01	2.00E-01	5.57E-02	5.13E-02	
	(3.75)	5.13E-02						

Table 5 (continued): Power Comparison Study for Skewed Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Chi (1)	(2.83)	k=1	6.31E-03	1.24E-01	1.32E-01	6.24E-02	7.89E-03	1.06E-02																																																																																																																																																																																																																																	
	(12.0)		3.13E-03						Chi (2)	(2.00)	5.82E-03	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03	(6.00)	2.73E-03	Chi (3)	(1.63)	5.72E-03	6.86E-02	7.28E-02	3.05E-02	1.27E-02	6.74E-03	(4.00)	2.60E-03	Chi (4)	(1.41)	5.44E-03	5.83E-02	6.14E-02	2.43E-02	1.32E-02	6.54E-03	(3.00)	2.42E-03	Chi (8)	(1.00)	4.87E-03	3.93E-02	4.03E-02	1.37E-02	1.30E-02	6.03E-03	(1.50)	1.99E-03	Chi (10)	(0.89)	4.67E-03	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03	(1.20)	1.88E-03	Chi (12)	(0.82)	4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03	(1.00)	1.77E-03	Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)	(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02	Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01
Chi (2)	(2.00)		5.82E-03	8.71E-02	9.26E-02	4.10E-02	1.15E-02	7.58E-03																																																																																																																																																																																																																																	
	(6.00)		2.73E-03						Chi (3)	(1.63)	5.72E-03	6.86E-02	7.28E-02	3.05E-02	1.27E-02	6.74E-03	(4.00)	2.60E-03	Chi (4)	(1.41)	5.44E-03	5.83E-02	6.14E-02	2.43E-02	1.32E-02	6.54E-03	(3.00)	2.42E-03	Chi (8)	(1.00)	4.87E-03	3.93E-02	4.03E-02	1.37E-02	1.30E-02	6.03E-03	(1.50)	1.99E-03	Chi (10)	(0.89)	4.67E-03	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03	(1.20)	1.88E-03	Chi (12)	(0.82)	4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03	(1.00)	1.77E-03	Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02	Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02			
Chi (3)	(1.63)		5.72E-03	6.86E-02	7.28E-02	3.05E-02	1.27E-02	6.74E-03																																																																																																																																																																																																																																	
	(4.00)		2.60E-03						Chi (4)	(1.41)	5.44E-03	5.83E-02	6.14E-02	2.43E-02	1.32E-02	6.54E-03	(3.00)	2.42E-03	Chi (8)	(1.00)	4.87E-03	3.93E-02	4.03E-02	1.37E-02	1.30E-02	6.03E-03	(1.50)	1.99E-03	Chi (10)	(0.89)	4.67E-03	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03	(1.20)	1.88E-03	Chi (12)	(0.82)	4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03	(1.00)	1.77E-03	Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02												
Chi (4)	(1.41)		5.44E-03	5.83E-02	6.14E-02	2.43E-02	1.32E-02	6.54E-03																																																																																																																																																																																																																																	
	(3.00)		2.42E-03						Chi (8)	(1.00)	4.87E-03	3.93E-02	4.03E-02	1.37E-02	1.30E-02	6.03E-03	(1.50)	1.99E-03	Chi (10)	(0.89)	4.67E-03	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03	(1.20)	1.88E-03	Chi (12)	(0.82)	4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03	(1.00)	1.77E-03	Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																					
Chi (8)	(1.00)		4.87E-03	3.93E-02	4.03E-02	1.37E-02	1.30E-02	6.03E-03																																																																																																																																																																																																																																	
	(1.50)		1.99E-03						Chi (10)	(0.89)	4.67E-03	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03	(1.20)	1.88E-03	Chi (12)	(0.82)	4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03	(1.00)	1.77E-03	Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02		(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																														
Chi (10)	(0.89)		4.67E-03	3.49E-02	3.53E-02	1.14E-02	1.32E-02	6.10E-03																																																																																																																																																																																																																																	
	(1.20)		1.88E-03						Chi (12)	(0.82)	4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03	(1.00)	1.77E-03	Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02		(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01		8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																							
Chi (12)	(0.82)		4.52E-03	3.16E-02	3.16E-02	9.71E-03	1.29E-02	5.95E-03																																																																																																																																																																																																																																	
	(1.00)		1.77E-03						Chi (16)	(0.71)	4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03	(0.75)	1.71E-03	Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02		(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01		8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01		1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																
Chi (16)	(0.71)		4.45E-03	2.82E-02	2.75E-02	7.89E-03	1.35E-02	6.08E-03																																																																																																																																																																																																																																	
	(0.75)		1.71E-03						Chi (24)	(0.58)	4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03	(0.50)	1.54E-03	Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02		(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01		8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01		1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01		1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																									
Chi (24)	(0.58)		4.22E-03	2.35E-02	2.24E-02	5.66E-03	1.30E-02	5.71E-03																																																																																																																																																																																																																																	
	(0.50)		1.54E-03						Gamma (0.15)	(5.16)	6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02	(40.0)	2.99E-03	Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02		(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01		8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01		1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01		1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01		2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																		
Gamma (0.15)	(5.16)		6.00E-03	1.85E-01	1.97E-01	9.43E-02	2.91E-02	2.82E-02																																																																																																																																																																																																																																	
	(40.0)		2.99E-03						Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03	(5.00)	2.87E-03	Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)		4.24E-02	Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02		(3.00)	4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01		8.13E-02	(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01		1.43E-01	8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01		1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01		2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01		2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																											
Gamma (1.2)	(1.83)	6.07E-03	7.89E-02	8.38E-02	3.65E-02	1.23E-02	7.27E-03																																																																																																																																																																																																																																		
	(5.00)	2.87E-03						Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03	(1.50)	2.07E-03	Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)	(2.00)		6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02	Chi (3)		(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02		Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)		4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02		(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01		8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01		1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01		1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01		2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																						
Gamma (4.0)	(1.00)	4.95E-03	3.91E-02	3.99E-02	1.36E-02	1.29E-02	6.02E-03																																																																																																																																																																																																																																		
	(1.50)	2.07E-03						Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02	5.37E-02	(12.0)	2.55E-02	Chi (2)		(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02	Chi (3)		(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02		Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)		4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02		(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01		8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01		1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01		1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01		2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																
Chi (1)	(2.83)	k=2	4.35E-02	3.09E-01	3.35E-01	2.05E-01	4.31E-02		5.37E-02																																																																																																																																																																																																																																
	(12.0)		2.55E-02					Chi (2)			(2.00)	6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02	5.44E-02	(6.00)	3.53E-02		Chi (3)	(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02		Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)		4.72E-02	Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02		(1.50)	6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01		8.88E-02	(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01		1.51E-01	9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01		1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01		2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																										
Chi (2)	(2.00)		6.02E-02	2.94E-01	3.25E-01	1.97E-01	7.20E-02		5.44E-02																																																																																																																																																																																																																																
	(6.00)		3.53E-02					Chi (3)			(1.63)	7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02	5.90E-02	(4.00)	4.24E-02		Chi (4)	(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)	4.72E-02		Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)		6.01E-02	Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02		(1.20)	6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01		9.40E-02	(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01		1.66E-01	1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01		1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																					
Chi (3)	(1.63)		7.21E-02	2.84E-01	3.18E-01	1.89E-01	9.10E-02		5.90E-02																																																																																																																																																																																																																																
	(4.00)		4.24E-02					Chi (4)			(1.41)	8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01	6.44E-02	(3.00)	4.72E-02		Chi (8)	(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02		Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)		6.73E-02	Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02		(1.00)	7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01		1.04E-01	(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01		1.80E-01	1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																
Chi (4)	(1.41)		8.05E-02	2.79E-01	3.14E-01	1.84E-01	1.04E-01		6.44E-02																																																																																																																																																																																																																																
	(3.00)		4.72E-02					Chi (8)			(1.00)	1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01	8.13E-02	(1.50)	6.01E-02		Chi (10)	(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02		Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)		7.28E-02	Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01		(0.75)	6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01		1.15E-01	(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																											
Chi (8)	(1.00)		1.02E-01	2.67E-01	3.05E-01	1.70E-01	1.33E-01		8.13E-02																																																																																																																																																																																																																																
	(1.50)		6.01E-02					Chi (10)			(0.89)	1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01	8.88E-02	(1.20)	6.73E-02		Chi (12)	(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02		Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)		6.42E-02	Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01		(0.50)	7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																						
Chi (10)	(0.89)		1.09E-01	2.64E-01	3.02E-01	1.66E-01	1.43E-01		8.88E-02																																																																																																																																																																																																																																
	(1.20)		6.73E-02					Chi (12)			(0.82)	1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01	9.40E-02	(1.00)	7.28E-02		Chi (16)	(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02		Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)		7.81E-02	Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																																	
Chi (12)	(0.82)		1.15E-01	2.61E-01	3.00E-01	1.63E-01	1.51E-01		9.40E-02																																																																																																																																																																																																																																
	(1.00)		7.28E-02					Chi (16)			(0.71)	1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01	1.04E-01	(0.75)	6.42E-02		Chi (24)	(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02		Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																																												
Chi (16)	(0.71)		1.24E-01	2.60E-01	2.99E-01	1.60E-01	1.66E-01		1.04E-01																																																																																																																																																																																																																																
	(0.75)		6.42E-02					Chi (24)			(0.58)	1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01	1.15E-01	(0.50)	7.81E-02		Gamma (0.15)	(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																																																							
Chi (24)	(0.58)		1.32E-01	2.56E-01	2.95E-01	1.54E-01	1.80E-01		1.15E-01																																																																																																																																																																																																																																
	(0.50)		7.81E-02					Gamma (0.15)			(5.16)	2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02	7.84E-02	(40.0)	1.34E-02	Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																																																																		
Gamma (0.15)	(5.16)		2.42E-02	3.15E-01	3.34E-01	1.96E-01	8.04E-02		7.84E-02																																																																																																																																																																																																																																
	(40.0)		1.34E-02					Gamma (1.2)		(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02	(5.00)	3.94E-02	Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																																																																													
Gamma (1.2)	(1.83)	6.66E-02	2.90E-01	3.23E-01	1.95E-01	8.15E-02	5.72E-02																																																																																																																																																																																																																																		
	(5.00)	3.94E-02						Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02	(1.50)	6.05E-02																																																																																																																																																																																																																								
Gamma (4.0)	(1.00)	1.03E-01	2.66E-01	3.04E-01	1.70E-01	1.32E-01	8.13E-02																																																																																																																																																																																																																																		
	(1.50)	6.05E-02																																																																																																																																																																																																																																							

A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 5 (continued): Power Comparison Study for Skewed Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Normal (0,1)	(0.00)	k=3	4.36E-01	5.49E-01	6.17E-01	4.38E-01	5.28E-01	4.54E-01																																																																																																																																																																																																																																	
	(0.00)		3.29E-01						Exponential (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Lognormal (0,1)	(6.18)	5.36E-02	3.91E-01	4.19E-01	2.65E-01	1.08E-01	1.02E-01	(110)	3.19E-02	Weibull (0.5)	(6.62)	4.13E-02	3.78E-01	4.00E-01	2.48E-01	1.16E-01	1.14E-01	(84.72)	2.42E-02	Weibull (0.85)	(0.12)	8.54E-02	4.37E-01	4.72E-01	3.20E-01	1.01E-01	1.07E-01	(1.23)	5.28E-02	Weibull (0.75)	(2.56)	1.10E-01	4.53E-01	4.92E-01	3.40E-01	1.20E-01	1.14E-01	(10.35)	7.00E-02	Weibull (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)	5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01	(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01
Exponential (1)	(2.00)		1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01																																																																																																																																																																																																																																	
	(6.00)		9.96E-02						Lognormal (0,1)	(6.18)	5.36E-02	3.91E-01	4.19E-01	2.65E-01	1.08E-01	1.02E-01	(110)	3.19E-02	Weibull (0.5)	(6.62)	4.13E-02	3.78E-01	4.00E-01	2.48E-01	1.16E-01	1.14E-01	(84.72)	2.42E-02	Weibull (0.85)	(0.12)	8.54E-02	4.37E-01	4.72E-01	3.20E-01	1.01E-01	1.07E-01	(1.23)	5.28E-02	Weibull (0.75)	(2.56)	1.10E-01	4.53E-01	4.92E-01	3.40E-01	1.20E-01	1.14E-01	(10.35)	7.00E-02	Weibull (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)	5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01	(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01			
Lognormal (0,1)	(6.18)		5.36E-02	3.91E-01	4.19E-01	2.65E-01	1.08E-01	1.02E-01																																																																																																																																																																																																																																	
	(110)		3.19E-02						Weibull (0.5)	(6.62)	4.13E-02	3.78E-01	4.00E-01	2.48E-01	1.16E-01	1.14E-01	(84.72)	2.42E-02	Weibull (0.85)	(0.12)	8.54E-02	4.37E-01	4.72E-01	3.20E-01	1.01E-01	1.07E-01	(1.23)	5.28E-02	Weibull (0.75)	(2.56)	1.10E-01	4.53E-01	4.92E-01	3.40E-01	1.20E-01	1.14E-01	(10.35)	7.00E-02	Weibull (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)	5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01	(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01												
Weibull (0.5)	(6.62)		4.13E-02	3.78E-01	4.00E-01	2.48E-01	1.16E-01	1.14E-01																																																																																																																																																																																																																																	
	(84.72)		2.42E-02						Weibull (0.85)	(0.12)	8.54E-02	4.37E-01	4.72E-01	3.20E-01	1.01E-01	1.07E-01	(1.23)	5.28E-02	Weibull (0.75)	(2.56)	1.10E-01	4.53E-01	4.92E-01	3.40E-01	1.20E-01	1.14E-01	(10.35)	7.00E-02	Weibull (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01	(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																					
Weibull (0.85)	(0.12)		8.54E-02	4.37E-01	4.72E-01	3.20E-01	1.01E-01	1.07E-01																																																																																																																																																																																																																																	
	(1.23)		5.28E-02						Weibull (0.75)	(2.56)	1.10E-01	4.53E-01	4.92E-01	3.40E-01	1.20E-01	1.14E-01	(10.35)	7.00E-02	Weibull (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																														
Weibull (0.75)	(2.56)		1.10E-01	4.53E-01	4.92E-01	3.40E-01	1.20E-01	1.14E-01																																																																																																																																																																																																																																	
	(10.35)		7.00E-02						Weibull (1)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																							
Weibull (1)	(2.00)		1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01																																																																																																																																																																																																																																	
	(6.00)		9.96E-02						Weibull (1.2)	(1.52)	2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01	(3.24)	1.43E-01	Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																
Weibull (1.2)	(1.52)		2.12E-01	4.93E-01	5.44E-01	3.85E-01	2.14E-01	1.56E-01																																																																																																																																																																																																																																	
	(3.24)		1.43E-01						Weibull (1.5)	(1.07)	2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01	(1.39)	2.08E-01	Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01		5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																									
Weibull (1.5)	(1.07)		2.96E-01	5.17E-01	5.74E-01	4.07E-01	2.91E-01	2.09E-01																																																																																																																																																																																																																																	
	(1.39)		2.08E-01						Weibull (2.0)	(0.63)	4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01	(0.25)	2.99E-01	Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01		5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01		6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																		
Weibull (2.0)	(0.63)		4.03E-01	5.44E-01	6.07E-01	4.29E-01	4.12E-01	3.10E-01																																																																																																																																																																																																																																	
	(0.25)		2.99E-01						Weibull (3.5)	(0.03)	4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01	(-0.29)	3.88E-01	Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01		5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01		6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01		6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																											
Weibull (3.5)	(0.03)		4.93E-01	5.65E-01	6.30E-01	4.47E-01	5.45E-01	4.63E-01																																																																																																																																																																																																																																	
	(-0.29)		3.88E-01						Barnes 2	(1.75)	2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01	(3.75)	1.40E-01	Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01		5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01		6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01		6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)		6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																				
Barnes 2	(1.75)		2.02E-01	4.94E-01	5.40E-01	3.87E-01	1.46E-01	1.39E-01																																																																																																																																																																																																																																	
	(3.75)		1.40E-01						Normal (0,1)	(0.00)	k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(0.00)	5.30E-01	Exponential (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01		5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01		6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01		6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)		6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)		(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																													
Normal (0,1)	(0.00)		k=4	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01		6.60E-01																																																																																																																																																																																																																															
	(0.00)			5.30E-01					Exponential (1)			(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Lognormal (0,1)	(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)		5.37E-02	Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01		(84.72)	3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01		1.62E-01	(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01		1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01		5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01		6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01		6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)		6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)		(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01		Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																						
Exponential (1)	(2.00)			2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01		2.11E-01																																																																																																																																																																																																																															
	(6.00)			1.77E-01					Lognormal (0,1)			(6.18)	8.59E-02	4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01	(110)	5.37E-02		Weibull (0.5)	(6.62)	6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01	(84.72)		3.82E-02	Weibull (0.85)	(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01		(1.23)	9.13E-02	Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01		1.78E-01	(10.35)	1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01		2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01		5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01		7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01		7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)		6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2		(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																	
Lognormal (0,1)	(6.18)	8.59E-02		4.80E-01	5.13E-01	3.43E-01	1.49E-01	1.41E-01																																																																																																																																																																																																																																	
	(110)	5.37E-02							Weibull (0.5)	(6.62)		6.24E-02	4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01	(84.72)	3.82E-02	Weibull (0.85)		(0.12)	1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02		Weibull (0.75)	(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)		1.23E-01	Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01		(6.00)	1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01		2.67E-01	(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01		4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01		6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01		8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01		6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																													
Weibull (0.5)	(6.62)	6.24E-02		4.44E-01	4.70E-01	3.07E-01	1.53E-01	1.51E-01																																																																																																																																																																																																																																	
	(84.72)	3.82E-02							Weibull (0.85)	(0.12)		1.40E-01	5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01	(1.23)	9.13E-02	Weibull (0.75)		(2.56)	1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01		Weibull (1)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)		1.77E-01	Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01		(3.24)	2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01		3.64E-01	(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01		6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01		6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01		6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																																								
Weibull (0.85)	(0.12)	1.40E-01		5.38E-01	5.76E-01	4.21E-01	1.53E-01	1.62E-01																																																																																																																																																																																																																																	
	(1.23)	9.13E-02							Weibull (0.75)	(2.56)		1.83E-01	5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01	(10.35)	1.23E-01	Weibull (1)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Weibull (1.2)	(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)		2.54E-01	Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01		(1.39)	3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01		5.18E-01	(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01		7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01		5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																																																			
Weibull (0.75)	(2.56)	1.83E-01		5.67E-01	6.09E-01	4.56E-01	1.86E-01	1.78E-01																																																																																																																																																																																																																																	
	(10.35)	1.23E-01							Weibull (1)	(2.00)		2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Weibull (1.2)		(1.52)	3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01		Weibull (1.5)	(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)		3.63E-01	Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01		(0.25)	4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01		6.80E-01	(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01		2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																																																														
Weibull (1)	(2.00)	2.52E-01		6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01																																																																																																																																																																																																																																	
	(6.00)	1.77E-01							Weibull (1.2)	(1.52)		3.47E-01	6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01	(3.24)	2.54E-01	Weibull (1.5)		(1.07)	4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01		Weibull (2.0)	(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)		4.96E-01	Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01		(-0.29)	5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01		2.39E-01	(3.75)	2.37E-01																																																																																																																																																																									
Weibull (1.2)	(1.52)	3.47E-01		6.44E-01	6.94E-01	5.46E-01	3.41E-01	2.67E-01																																																																																																																																																																																																																																	
	(3.24)	2.54E-01							Weibull (1.5)	(1.07)		4.70E-01	6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01	(1.39)	3.63E-01	Weibull (2.0)		(0.63)	6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01		Weibull (3.5)	(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)		5.96E-01	Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01		(3.75)	2.37E-01																																																																																																																																																																																				
Weibull (1.5)	(1.07)	4.70E-01		6.90E-01	7.42E-01	5.97E-01	4.62E-01	3.64E-01																																																																																																																																																																																																																																	
	(1.39)	3.63E-01							Weibull (2.0)	(0.63)		6.04E-01	7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01	(0.25)	4.96E-01	Weibull (3.5)		(0.03)	6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01		Barnes 2	(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)		2.37E-01																																																																																																																																																																																															
Weibull (2.0)	(0.63)	6.04E-01		7.36E-01	7.87E-01	6.47E-01	6.20E-01	5.18E-01																																																																																																																																																																																																																																	
	(0.25)	4.96E-01							Weibull (3.5)	(0.03)		6.91E-01	7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01	(-0.29)	5.96E-01	Barnes 2		(1.75)	3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																																																																																																											
Weibull (3.5)	(0.03)	6.91E-01		7.60E-01	8.12E-01	6.80E-01	7.45E-01	6.80E-01																																																																																																																																																																																																																																	
	(-0.29)	5.96E-01							Barnes 2	(1.75)		3.19E-01	6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01	(3.75)	2.37E-01																																																																																																																																																																																																																						
Barnes 2	(1.75)	3.19E-01		6.28E-01	6.74E-01	5.33E-01	2.46E-01	2.39E-01																																																																																																																																																																																																																																	
	(3.75)	2.37E-01																																																																																																																																																																																																																																							

Table 5 (continued): Power Comparison Study for Skewed Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Chi (1)	(2.83)	k=3	9.98E-02	4.49E-01	4.85E-01	3.35E-01	9.30E-02	1.11E-01																																																																																																																																																																																																																																	
	(12.0)		6.36E-02						Chi (2)	(2.00)	1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01	(6.00)	9.96E-02	Chi (3)	(1.63)	1.90E-01	4.84E-01	5.33E-01	3.76E-01	2.08E-01	1.49E-01	(4.00)	1.27E-01	Chi (4)	(1.41)	2.18E-01	4.94E-01	5.47E-01	3.85E-01	2.42E-01	1.68E-01	(3.00)	1.47E-01	Chi (8)	(1.00)	2.86E-01	5.14E-01	5.73E-01	4.04E-01	3.20E-01	2.26E-01	(1.50)	1.98E-01	Chi (10)	(0.89)	3.07E-01	5.19E-01	5.79E-01	4.08E-01	3.46E-01	2.49E-01	(1.20)	2.15E-01	Chi (12)	(0.82)	3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01	(1.00)	2.26E-01	Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)	(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01
Chi (2)	(2.00)		1.53E-01	4.72E-01	5.17E-01	3.63E-01	1.59E-01	1.29E-01																																																																																																																																																																																																																																	
	(6.00)		9.96E-02						Chi (3)	(1.63)	1.90E-01	4.84E-01	5.33E-01	3.76E-01	2.08E-01	1.49E-01	(4.00)	1.27E-01	Chi (4)	(1.41)	2.18E-01	4.94E-01	5.47E-01	3.85E-01	2.42E-01	1.68E-01	(3.00)	1.47E-01	Chi (8)	(1.00)	2.86E-01	5.14E-01	5.73E-01	4.04E-01	3.20E-01	2.26E-01	(1.50)	1.98E-01	Chi (10)	(0.89)	3.07E-01	5.19E-01	5.79E-01	4.08E-01	3.46E-01	2.49E-01	(1.20)	2.15E-01	Chi (12)	(0.82)	3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01	(1.00)	2.26E-01	Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01			
Chi (3)	(1.63)		1.90E-01	4.84E-01	5.33E-01	3.76E-01	2.08E-01	1.49E-01																																																																																																																																																																																																																																	
	(4.00)		1.27E-01						Chi (4)	(1.41)	2.18E-01	4.94E-01	5.47E-01	3.85E-01	2.42E-01	1.68E-01	(3.00)	1.47E-01	Chi (8)	(1.00)	2.86E-01	5.14E-01	5.73E-01	4.04E-01	3.20E-01	2.26E-01	(1.50)	1.98E-01	Chi (10)	(0.89)	3.07E-01	5.19E-01	5.79E-01	4.08E-01	3.46E-01	2.49E-01	(1.20)	2.15E-01	Chi (12)	(0.82)	3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01	(1.00)	2.26E-01	Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01												
Chi (4)	(1.41)		2.18E-01	4.94E-01	5.47E-01	3.85E-01	2.42E-01	1.68E-01																																																																																																																																																																																																																																	
	(3.00)		1.47E-01						Chi (8)	(1.00)	2.86E-01	5.14E-01	5.73E-01	4.04E-01	3.20E-01	2.26E-01	(1.50)	1.98E-01	Chi (10)	(0.89)	3.07E-01	5.19E-01	5.79E-01	4.08E-01	3.46E-01	2.49E-01	(1.20)	2.15E-01	Chi (12)	(0.82)	3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01	(1.00)	2.26E-01	Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																					
Chi (8)	(1.00)		2.86E-01	5.14E-01	5.73E-01	4.04E-01	3.20E-01	2.26E-01																																																																																																																																																																																																																																	
	(1.50)		1.98E-01						Chi (10)	(0.89)	3.07E-01	5.19E-01	5.79E-01	4.08E-01	3.46E-01	2.49E-01	(1.20)	2.15E-01	Chi (12)	(0.82)	3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01	(1.00)	2.26E-01	Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01		(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																														
Chi (10)	(0.89)		3.07E-01	5.19E-01	5.79E-01	4.08E-01	3.46E-01	2.49E-01																																																																																																																																																																																																																																	
	(1.20)		2.15E-01						Chi (12)	(0.82)	3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01	(1.00)	2.26E-01	Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01		(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01		3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																							
Chi (12)	(0.82)		3.21E-01	5.22E-01	5.83E-01	4.11E-01	3.65E-01	2.66E-01																																																																																																																																																																																																																																	
	(1.00)		2.26E-01						Chi (16)	(0.71)	3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01	(0.75)	2.46E-01	Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01		(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01		3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01		5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																
Chi (16)	(0.71)		3.44E-01	5.29E-01	5.91E-01	4.17E-01	3.97E-01	2.94E-01																																																																																																																																																																																																																																	
	(0.75)		2.46E-01						Chi (24)	(0.58)	3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01	(0.50)	2.65E-01	Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01		(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01		3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01		5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01		6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																									
Chi (24)	(0.58)		3.66E-01	5.33E-01	5.96E-01	4.19E-01	4.29E-01	3.26E-01																																																																																																																																																																																																																																	
	(0.50)		2.65E-01						Gamma (0.15)	(5.16)	4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01	(40.0)	2.77E-02	Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01		(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01		3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01		5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01		6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01		7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																		
Gamma (0.15)	(5.16)		4.69E-02	4.00E-01	4.21E-01	2.73E-01	1.29E-01	1.26E-01																																																																																																																																																																																																																																	
	(40.0)		2.77E-02						Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01	(5.00)	1.14E-01	Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)		2.27E-01	Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01		(3.00)	2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01		3.84E-01	(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01		5.29E-01	4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01		6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01		7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01		7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																											
Gamma (1.2)	(1.83)	1.72E-01	4.79E-01	5.26E-01	3.71E-01	1.82E-01	1.39E-01																																																																																																																																																																																																																																		
	(5.00)	1.14E-01						Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01	(1.50)	2.00E-01	Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)	(2.00)		2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Chi (3)		(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01		Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)		2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01		(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01		4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01		5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01		6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01		7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																						
Gamma (4.0)	(1.00)	2.88E-01	5.13E-01	5.72E-01	4.03E-01	3.19E-01	2.27E-01																																																																																																																																																																																																																																		
	(1.50)	2.00E-01						Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01	1.71E-01	(12.0)	1.08E-01	Chi (2)		(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01	Chi (3)		(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01		Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)		2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01		(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01		4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01		5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01		6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01		7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																
Chi (1)	(2.83)	k=4	1.61E-01	5.52E-01	5.90E-01	4.40E-01	1.47E-01		1.71E-01																																																																																																																																																																																																																																
	(12.0)		1.08E-01					Chi (2)			(2.00)	2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01	2.11E-01	(6.00)	1.77E-01		Chi (3)	(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01		Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)		2.64E-01	Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01		(1.50)	3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01		4.19E-01	(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01		5.53E-01	4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01		6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01		7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																										
Chi (2)	(2.00)		2.52E-01	6.03E-01	6.50E-01	5.00E-01	2.51E-01		2.11E-01																																																																																																																																																																																																																																
	(6.00)		1.77E-01					Chi (3)			(1.63)	3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01	2.50E-01	(4.00)	2.27E-01		Chi (4)	(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)	2.64E-01		Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)		3.52E-01	Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01		(1.20)	3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01		4.43E-01	(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01		5.90E-01	4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01		6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																					
Chi (3)	(1.63)		3.15E-01	6.31E-01	6.81E-01	5.30E-01	3.27E-01		2.50E-01																																																																																																																																																																																																																																
	(4.00)		2.27E-01					Chi (4)			(1.41)	3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01	2.86E-01	(3.00)	2.64E-01		Chi (8)	(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01		Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)		3.78E-01	Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01		(1.00)	3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01		4.82E-01	(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01		6.25E-01	5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																
Chi (4)	(1.41)		3.60E-01	6.51E-01	7.03E-01	5.53E-01	3.80E-01		2.86E-01																																																																																																																																																																																																																																
	(3.00)		2.64E-01					Chi (8)			(1.00)	4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01	3.84E-01	(1.50)	3.52E-01		Chi (10)	(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01		Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)		3.96E-01	Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01		(0.75)	4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01		5.24E-01	(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																											
Chi (8)	(1.00)		4.61E-01	6.89E-01	7.44E-01	5.96E-01	4.94E-01		3.84E-01																																																																																																																																																																																																																																
	(1.50)		3.52E-01					Chi (10)			(0.89)	4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01	4.19E-01	(1.20)	3.78E-01		Chi (12)	(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01		Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)		4.23E-01	Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01		(0.50)	4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																						
Chi (10)	(0.89)		4.89E-01	6.98E-01	7.53E-01	6.06E-01	5.29E-01		4.19E-01																																																																																																																																																																																																																																
	(1.20)		3.78E-01					Chi (12)			(0.82)	5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01	4.43E-01	(1.00)	3.96E-01		Chi (16)	(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01		Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)		4.50E-01	Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																																	
Chi (12)	(0.82)		5.07E-01	7.03E-01	7.59E-01	6.13E-01	5.53E-01		4.43E-01																																																																																																																																																																																																																																
	(1.00)		3.96E-01					Chi (16)			(0.71)	5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01	4.82E-01	(0.75)	4.23E-01		Chi (24)	(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01		Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																																												
Chi (16)	(0.71)		5.35E-01	7.12E-01	7.69E-01	6.24E-01	5.90E-01		4.82E-01																																																																																																																																																																																																																																
	(0.75)		4.23E-01					Chi (24)			(0.58)	5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01	5.24E-01	(0.50)	4.50E-01		Gamma (0.15)	(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																																																							
Chi (24)	(0.58)		5.62E-01	7.19E-01	7.76E-01	6.33E-01	6.25E-01		5.24E-01																																																																																																																																																																																																																																
	(0.50)		4.50E-01					Gamma (0.15)			(5.16)	7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01	1.68E-01	(40.0)	4.34E-02	Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																																																																		
Gamma (0.15)	(5.16)		7.02E-02	4.60E-01	4.83E-01	3.32E-01	1.71E-01		1.68E-01																																																																																																																																																																																																																																
	(40.0)		4.34E-02					Gamma (1.2)		(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01	(5.00)	2.02E-01	Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																																																																													
Gamma (1.2)	(1.83)	2.83E-01	6.17E-01	6.66E-01	5.16E-01	2.87E-01	2.30E-01																																																																																																																																																																																																																																		
	(5.00)	2.02E-01						Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01	(1.50)	3.54E-01																																																																																																																																																																																																																								
Gamma (4.0)	(1.00)	4.63E-01	6.88E-01	7.43E-01	5.95E-01	4.94E-01	3.84E-01																																																																																																																																																																																																																																		
	(1.50)	3.54E-01																																																																																																																																																																																																																																							

A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 5 (continued): Power Comparison Study for Skewed Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Normal (0,1)	(0.00)	k=5	7.61E-01	8.44E-01	8.88E-01	7.95E-01	8.33E-01	7.88E-01																																																																																																																																																																																																																																	
	(0.00)		6.71E-01						Exponential (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Lognormal (0,1)	(6.18)	1.20E-01	5.52E-01	5.87E-01	4.10E-01	1.89E-01	1.79E-01	(110)	7.75E-02	Weibull (0.5)	(6.62)	8.36E-02	4.97E-01	5.25E-01	3.56E-01	1.87E-01	1.85E-01	(84.72)	5.29E-02	Weibull (0.85)	(0.12)	1.94E-01	6.15E-01	6.54E-01	5.02E-01	2.04E-01	2.15E-01	(1.23)	1.32E-01	Weibull (0.75)	(2.56)	2.53E-01	6.51E-01	6.93E-01	5.48E-01	2.50E-01	2.41E-01	(10.35)	1.78E-01	Weibull (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)	1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01	(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01
Exponential (1)	(2.00)		3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01																																																																																																																																																																																																																																	
	(6.00)		2.54E-01						Lognormal (0,1)	(6.18)	1.20E-01	5.52E-01	5.87E-01	4.10E-01	1.89E-01	1.79E-01	(110)	7.75E-02	Weibull (0.5)	(6.62)	8.36E-02	4.97E-01	5.25E-01	3.56E-01	1.87E-01	1.85E-01	(84.72)	5.29E-02	Weibull (0.85)	(0.12)	1.94E-01	6.15E-01	6.54E-01	5.02E-01	2.04E-01	2.15E-01	(1.23)	1.32E-01	Weibull (0.75)	(2.56)	2.53E-01	6.51E-01	6.93E-01	5.48E-01	2.50E-01	2.41E-01	(10.35)	1.78E-01	Weibull (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)	1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01	(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01			
Lognormal (0,1)	(6.18)		1.20E-01	5.52E-01	5.87E-01	4.10E-01	1.89E-01	1.79E-01																																																																																																																																																																																																																																	
	(110)		7.75E-02						Weibull (0.5)	(6.62)	8.36E-02	4.97E-01	5.25E-01	3.56E-01	1.87E-01	1.85E-01	(84.72)	5.29E-02	Weibull (0.85)	(0.12)	1.94E-01	6.15E-01	6.54E-01	5.02E-01	2.04E-01	2.15E-01	(1.23)	1.32E-01	Weibull (0.75)	(2.56)	2.53E-01	6.51E-01	6.93E-01	5.48E-01	2.50E-01	2.41E-01	(10.35)	1.78E-01	Weibull (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)	1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01	(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01												
Weibull (0.5)	(6.62)		8.36E-02	4.97E-01	5.25E-01	3.56E-01	1.87E-01	1.85E-01																																																																																																																																																																																																																																	
	(84.72)		5.29E-02						Weibull (0.85)	(0.12)	1.94E-01	6.15E-01	6.54E-01	5.02E-01	2.04E-01	2.15E-01	(1.23)	1.32E-01	Weibull (0.75)	(2.56)	2.53E-01	6.51E-01	6.93E-01	5.48E-01	2.50E-01	2.41E-01	(10.35)	1.78E-01	Weibull (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01	(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																					
Weibull (0.85)	(0.12)		1.94E-01	6.15E-01	6.54E-01	5.02E-01	2.04E-01	2.15E-01																																																																																																																																																																																																																																	
	(1.23)		1.32E-01						Weibull (0.75)	(2.56)	2.53E-01	6.51E-01	6.93E-01	5.48E-01	2.50E-01	2.41E-01	(10.35)	1.78E-01	Weibull (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																														
Weibull (0.75)	(2.56)		2.53E-01	6.51E-01	6.93E-01	5.48E-01	2.50E-01	2.41E-01																																																																																																																																																																																																																																	
	(10.35)		1.78E-01						Weibull (1)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																							
Weibull (1)	(2.00)		3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01																																																																																																																																																																																																																																	
	(6.00)		2.54E-01						Weibull (1.2)	(1.52)	4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01	(3.24)	3.58E-01	Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																
Weibull (1.2)	(1.52)		4.64E-01	7.47E-01	7.90E-01	6.64E-01	4.55E-01	3.73E-01																																																																																																																																																																																																																																	
	(3.24)		3.58E-01						Weibull (1.5)	(1.07)	6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01	(1.39)	4.94E-01	Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01		6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																									
Weibull (1.5)	(1.07)		6.05E-01	7.99E-01	8.40E-01	7.27E-01	5.99E-01	5.02E-01																																																																																																																																																																																																																																	
	(1.39)		4.94E-01						Weibull (2.0)	(0.63)	7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01	(0.25)	6.39E-01	Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01		6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01		8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																		
Weibull (2.0)	(0.63)		7.36E-01	8.45E-01	8.81E-01	7.83E-01	7.56E-01	6.73E-01																																																																																																																																																																																																																																	
	(0.25)		6.39E-01						Weibull (3.5)	(0.03)	8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01	(-0.29)	7.29E-01	Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01		6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01		8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01		8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																											
Weibull (3.5)	(0.03)		8.06E-01	8.63E-01	8.99E-01	8.14E-01	8.53E-01	8.09E-01																																																																																																																																																																																																																																	
	(-0.29)		7.29E-01						Barnes 2	(1.75)	4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01	(3.75)	3.26E-01	Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01		6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01		8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01		8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)		8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																				
Barnes 2	(1.75)		4.21E-01	7.20E-01	7.62E-01	6.40E-01	3.41E-01	3.35E-01																																																																																																																																																																																																																																	
	(3.75)		3.26E-01						Normal (0,1)	(0.00)	k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(0.00)	7.64E-01	Exponential (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01		6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01		8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01		8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)		8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)		(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																													
Normal (0,1)	(0.00)		k=6	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01		8.65E-01																																																																																																																																																																																																																															
	(0.00)			7.64E-01					Exponential (1)			(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Lognormal (0,1)	(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)		1.02E-01	Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01		(84.72)	6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01		2.64E-01	(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01		3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01		6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01		8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01		8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)		8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)		(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01		Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																						
Exponential (1)	(2.00)			4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01		3.67E-01																																																																																																																																																																																																																															
	(6.00)			3.26E-01					Lognormal (0,1)			(6.18)	1.54E-01	6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01	(110)	1.02E-01		Weibull (0.5)	(6.62)	1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01	(84.72)		6.77E-02	Weibull (0.85)	(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01		(1.23)	1.73E-01	Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01		3.01E-01	(10.35)	2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01		4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01		7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01		8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01		9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)		8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2		(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																	
Lognormal (0,1)	(6.18)	1.54E-01		6.10E-01	6.46E-01	4.67E-01	2.26E-01	2.16E-01																																																																																																																																																																																																																																	
	(110)	1.02E-01							Weibull (0.5)	(6.62)		1.04E-01	5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01	(84.72)	6.77E-02	Weibull (0.85)		(0.12)	2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01		Weibull (0.75)	(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)		2.32E-01	Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01		(6.00)	3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01		4.67E-01	(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01		7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01		8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01		9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01		7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																													
Weibull (0.5)	(6.62)	1.04E-01		5.40E-01	5.69E-01	3.99E-01	2.18E-01	2.15E-01																																																																																																																																																																																																																																	
	(84.72)	6.77E-02							Weibull (0.85)	(0.12)		2.47E-01	6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01	(1.23)	1.73E-01	Weibull (0.75)		(2.56)	3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01		Weibull (1)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)		3.26E-01	Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01		(3.24)	4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01		6.13E-01	(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01		8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01		8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01		8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																																								
Weibull (0.85)	(0.12)	2.47E-01		6.74E-01	7.12E-01	5.68E-01	2.52E-01	2.64E-01																																																																																																																																																																																																																																	
	(1.23)	1.73E-01							Weibull (0.75)	(2.56)		3.19E-01	7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01	(10.35)	2.32E-01	Weibull (1)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Weibull (1.2)	(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)		4.49E-01	Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01		(1.39)	5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01		7.78E-01	(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01		9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01		7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																																																			
Weibull (0.75)	(2.56)	3.19E-01		7.15E-01	7.54E-01	6.20E-01	3.10E-01	3.01E-01																																																																																																																																																																																																																																	
	(10.35)	2.32E-01							Weibull (1)	(2.00)		4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Weibull (1.2)		(1.52)	5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01		Weibull (1.5)	(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)		5.98E-01	Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01		(0.25)	7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01		8.82E-01	(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01		4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																																																														
Weibull (1)	(2.00)	4.28E-01		7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01																																																																																																																																																																																																																																	
	(6.00)	3.26E-01							Weibull (1.2)	(1.52)		5.60E-01	8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01	(3.24)	4.49E-01	Weibull (1.5)		(1.07)	7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01		Weibull (2.0)	(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)		7.38E-01	Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01		(-0.29)	8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01		4.20E-01	(3.75)	4.05E-01																																																																																																																																																																									
Weibull (1.2)	(1.52)	5.60E-01		8.16E-01	8.53E-01	7.49E-01	5.50E-01	4.67E-01																																																																																																																																																																																																																																	
	(3.24)	4.49E-01							Weibull (1.5)	(1.07)		7.02E-01	8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01	(1.39)	5.98E-01	Weibull (2.0)		(0.63)	8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01		Weibull (3.5)	(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)		8.13E-01	Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01		(3.75)	4.05E-01																																																																																																																																																																																				
Weibull (1.5)	(1.07)	7.02E-01		8.67E-01	8.97E-01	8.12E-01	7.01E-01	6.13E-01																																																																																																																																																																																																																																	
	(1.39)	5.98E-01							Weibull (2.0)	(0.63)		8.19E-01	9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01	(0.25)	7.38E-01	Weibull (3.5)		(0.03)	8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01		Barnes 2	(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)		4.05E-01																																																																																																																																																																																															
Weibull (2.0)	(0.63)	8.19E-01		9.05E-01	9.30E-01	8.64E-01	8.41E-01	7.78E-01																																																																																																																																																																																																																																	
	(0.25)	7.38E-01							Weibull (3.5)	(0.03)		8.73E-01	9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01	(-0.29)	8.13E-01	Barnes 2		(1.75)	5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																																																																																																											
Weibull (3.5)	(0.03)	8.73E-01		9.19E-01	9.43E-01	8.88E-01	9.12E-01	8.82E-01																																																																																																																																																																																																																																	
	(-0.29)	8.13E-01							Barnes 2	(1.75)		5.05E-01	7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01	(3.75)	4.05E-01																																																																																																																																																																																																																						
Barnes 2	(1.75)	5.05E-01		7.84E-01	8.21E-01	7.17E-01	4.25E-01	4.20E-01																																																																																																																																																																																																																																	
	(3.75)	4.05E-01																																																																																																																																																																																																																																							

Table 5 (continued): Power Comparison Study for Skewed Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Chi (1)	(2.83)	k=5	2.21E-01	6.28E-01	6.66E-01	5.23E-01	2.00E-01	2.29E-01																																																																																																																																																																																																																																	
	(12.0)		1.54E-01						Chi (2)	(2.00)	3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01	(6.00)	2.54E-01	Chi (3)	(1.63)	4.27E-01	7.33E-01	7.78E-01	6.47E-01	4.35E-01	3.48E-01	(4.00)	3.24E-01	Chi (4)	(1.41)	4.83E-01	7.58E-01	8.02E-01	6.76E-01	5.00E-01	3.97E-01	(3.00)	3.74E-01	Chi (8)	(1.00)	5.96E-01	8.00E-01	8.44E-01	7.28E-01	6.31E-01	5.21E-01	(1.50)	4.84E-01	Chi (10)	(0.89)	6.26E-01	8.09E-01	8.52E-01	7.40E-01	6.66E-01	5.62E-01	(1.20)	5.15E-01	Chi (12)	(0.82)	6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01	(1.00)	5.35E-01	Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)	(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01
Chi (2)	(2.00)		3.46E-01	6.97E-01	7.41E-01	6.05E-01	3.38E-01	2.92E-01																																																																																																																																																																																																																																	
	(6.00)		2.54E-01						Chi (3)	(1.63)	4.27E-01	7.33E-01	7.78E-01	6.47E-01	4.35E-01	3.48E-01	(4.00)	3.24E-01	Chi (4)	(1.41)	4.83E-01	7.58E-01	8.02E-01	6.76E-01	5.00E-01	3.97E-01	(3.00)	3.74E-01	Chi (8)	(1.00)	5.96E-01	8.00E-01	8.44E-01	7.28E-01	6.31E-01	5.21E-01	(1.50)	4.84E-01	Chi (10)	(0.89)	6.26E-01	8.09E-01	8.52E-01	7.40E-01	6.66E-01	5.62E-01	(1.20)	5.15E-01	Chi (12)	(0.82)	6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01	(1.00)	5.35E-01	Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01			
Chi (3)	(1.63)		4.27E-01	7.33E-01	7.78E-01	6.47E-01	4.35E-01	3.48E-01																																																																																																																																																																																																																																	
	(4.00)		3.24E-01						Chi (4)	(1.41)	4.83E-01	7.58E-01	8.02E-01	6.76E-01	5.00E-01	3.97E-01	(3.00)	3.74E-01	Chi (8)	(1.00)	5.96E-01	8.00E-01	8.44E-01	7.28E-01	6.31E-01	5.21E-01	(1.50)	4.84E-01	Chi (10)	(0.89)	6.26E-01	8.09E-01	8.52E-01	7.40E-01	6.66E-01	5.62E-01	(1.20)	5.15E-01	Chi (12)	(0.82)	6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01	(1.00)	5.35E-01	Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01												
Chi (4)	(1.41)		4.83E-01	7.58E-01	8.02E-01	6.76E-01	5.00E-01	3.97E-01																																																																																																																																																																																																																																	
	(3.00)		3.74E-01						Chi (8)	(1.00)	5.96E-01	8.00E-01	8.44E-01	7.28E-01	6.31E-01	5.21E-01	(1.50)	4.84E-01	Chi (10)	(0.89)	6.26E-01	8.09E-01	8.52E-01	7.40E-01	6.66E-01	5.62E-01	(1.20)	5.15E-01	Chi (12)	(0.82)	6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01	(1.00)	5.35E-01	Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																					
Chi (8)	(1.00)		5.96E-01	8.00E-01	8.44E-01	7.28E-01	6.31E-01	5.21E-01																																																																																																																																																																																																																																	
	(1.50)		4.84E-01						Chi (10)	(0.89)	6.26E-01	8.09E-01	8.52E-01	7.40E-01	6.66E-01	5.62E-01	(1.20)	5.15E-01	Chi (12)	(0.82)	6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01	(1.00)	5.35E-01	Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01		(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																														
Chi (10)	(0.89)		6.26E-01	8.09E-01	8.52E-01	7.40E-01	6.66E-01	5.62E-01																																																																																																																																																																																																																																	
	(1.20)		5.15E-01						Chi (12)	(0.82)	6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01	(1.00)	5.35E-01	Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01		(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01		6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																							
Chi (12)	(0.82)		6.45E-01	8.14E-01	8.58E-01	7.48E-01	6.90E-01	5.89E-01																																																																																																																																																																																																																																	
	(1.00)		5.35E-01						Chi (16)	(0.71)	6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01	(0.75)	5.64E-01	Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01		(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01		6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01		7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																
Chi (16)	(0.71)		6.71E-01	8.22E-01	8.65E-01	7.60E-01	7.23E-01	6.29E-01																																																																																																																																																																																																																																	
	(0.75)		5.64E-01						Chi (24)	(0.58)	6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01	(0.50)	5.93E-01	Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01		(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01		6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01		7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01		8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																									
Chi (24)	(0.58)		6.96E-01	8.29E-01	8.72E-01	7.69E-01	7.55E-01	6.71E-01																																																																																																																																																																																																																																	
	(0.50)		5.93E-01						Gamma (0.15)	(5.16)	9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01	(40.0)	5.93E-02	Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01		(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01		6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01		7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01		8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01		9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																		
Gamma (0.15)	(5.16)		9.29E-02	5.07E-01	5.30E-01	3.80E-01	2.09E-01	2.05E-01																																																																																																																																																																																																																																	
	(40.0)		5.93E-02						Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01	(5.00)	2.89E-01	Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)		4.11E-01	Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01		(3.00)	4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01		6.31E-01	(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01		7.62E-01	6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01		8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01		9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01		8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																											
Gamma (1.2)	(1.83)	3.86E-01	7.15E-01	7.59E-01	6.26E-01	3.84E-01	3.19E-01																																																																																																																																																																																																																																		
	(5.00)	2.89E-01						Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01	(1.50)	4.87E-01	Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)	(2.00)		4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Chi (3)		(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01		Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)		4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01		(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01		6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01		7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01		8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01		9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																						
Gamma (4.0)	(1.00)	5.99E-01	7.99E-01	8.43E-01	7.28E-01	6.30E-01	5.22E-01																																																																																																																																																																																																																																		
	(1.50)	4.87E-01						Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01	2.83E-01	(12.0)	1.99E-01	Chi (2)		(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01	Chi (3)		(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01		Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)		4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01		(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01		6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01		7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01		8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01		9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																
Chi (1)	(2.83)	k=6	2.77E-01	6.86E-01	7.23E-01	5.90E-01	2.51E-01		2.83E-01																																																																																																																																																																																																																																
	(12.0)		1.99E-01					Chi (2)			(2.00)	4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01	3.67E-01	(6.00)	3.26E-01		Chi (3)	(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01		Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)		4.69E-01	Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01		(1.50)	5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01		6.71E-01	(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01		7.83E-01	6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01		8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01		9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																										
Chi (2)	(2.00)		4.28E-01	7.65E-01	8.04E-01	6.84E-01	4.16E-01		3.67E-01																																																																																																																																																																																																																																
	(6.00)		3.26E-01					Chi (3)			(1.63)	5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01	4.37E-01	(4.00)	4.11E-01		Chi (4)	(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)	4.69E-01		Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)		5.89E-01	Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01		(1.20)	6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01		6.98E-01	(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01		8.11E-01	7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01		8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																					
Chi (3)	(1.63)		5.22E-01	8.04E-01	8.42E-01	7.32E-01	5.27E-01		4.37E-01																																																																																																																																																																																																																																
	(4.00)		4.11E-01					Chi (4)			(1.41)	5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01	4.95E-01	(3.00)	4.69E-01		Chi (8)	(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01		Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)		6.20E-01	Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01		(1.00)	6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01		7.35E-01	(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01		8.38E-01	7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																
Chi (4)	(1.41)		5.82E-01	8.29E-01	8.65E-01	7.63E-01	5.98E-01		4.95E-01																																																																																																																																																																																																																																
	(3.00)		4.69E-01					Chi (8)			(1.00)	6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01	6.31E-01	(1.50)	5.89E-01		Chi (10)	(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01		Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)		6.40E-01	Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01		(0.75)	6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01		7.72E-01	(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																											
Chi (8)	(1.00)		6.96E-01	8.68E-01	9.01E-01	8.16E-01	7.30E-01		6.31E-01																																																																																																																																																																																																																																
	(1.50)		5.89E-01					Chi (10)			(0.89)	7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01	6.71E-01	(1.20)	6.20E-01		Chi (12)	(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01		Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)		6.68E-01	Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01		(0.50)	6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																						
Chi (10)	(0.89)		7.24E-01	8.76E-01	9.08E-01	8.27E-01	7.62E-01		6.71E-01																																																																																																																																																																																																																																
	(1.20)		6.20E-01					Chi (12)			(0.82)	7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01	6.98E-01	(1.00)	6.40E-01		Chi (16)	(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01		Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)		6.95E-01	Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																																	
Chi (12)	(0.82)		7.41E-01	8.80E-01	9.13E-01	8.34E-01	7.83E-01		6.98E-01																																																																																																																																																																																																																																
	(1.00)		6.40E-01					Chi (16)			(0.71)	7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01	7.35E-01	(0.75)	6.68E-01		Chi (24)	(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01		Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																																												
Chi (16)	(0.71)		7.63E-01	8.87E-01	9.19E-01	8.44E-01	8.11E-01		7.35E-01																																																																																																																																																																																																																																
	(0.75)		6.68E-01					Chi (24)			(0.58)	7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01	7.72E-01	(0.50)	6.95E-01		Gamma (0.15)	(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																																																							
Chi (24)	(0.58)		7.85E-01	8.92E-01	9.24E-01	8.53E-01	8.38E-01		7.72E-01																																																																																																																																																																																																																																
	(0.50)		6.95E-01					Gamma (0.15)			(5.16)	1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01	2.37E-01	(40.0)	7.52E-02	Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																																																																		
Gamma (0.15)	(5.16)		1.15E-01	5.44E-01	5.67E-01	4.19E-01	2.42E-01		2.37E-01																																																																																																																																																																																																																																
	(40.0)		7.52E-02					Gamma (1.2)		(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01	(5.00)	3.68E-01	Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																																																																													
Gamma (1.2)	(1.83)	4.75E-01	7.85E-01	8.23E-01	7.08E-01	4.69E-01	4.00E-01																																																																																																																																																																																																																																		
	(5.00)	3.68E-01						Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01	(1.50)	5.92E-01																																																																																																																																																																																																																								
Gamma (4.0)	(1.00)	6.99E-01	8.67E-01	9.01E-01	8.15E-01	7.30E-01	6.32E-01																																																																																																																																																																																																																																		
	(1.50)	5.92E-01																																																																																																																																																																																																																																							

A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 6: Power Comparison Study for Symmetric Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Barnes 3	(3.00)	k=1	2.04E-03	1.10E-01	9.09E-02	6.07E-02	1.06E-01	7.76E-02																																																																																																																																																																																																																																	
	(1049)		8.02E-04						Barnes 1	(0.00)	3.97E-03	1.45E-02	1.24E-02	1.86E-03	1.24E-02	6.40E-03	(6.89)	1.23E-03	JTB (2.0, 1.0)	(0.00)	4.58E-03	6.60E-02	6.30E-02	2.39E-02	5.77E-02	4.23E-02	(3.00)	1.89E-03	JTB (0.75, 0.5)	(0.00)	5.69E-03	3.93E-02	3.62E-02	9.57E-03	3.33E-02	2.21E-02	(1.20)	2.22E-03	JTB (4.0, 1.0)	(0.00)	4.33E-03	3.31E-02	2.78E-02	7.42E-03	2.95E-02	1.88E-02	(0.78)	1.59E-03	JTB (1.0, 0.5)	(0.00)	5.08E-03	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02	(0.60)	1.83E-03	JTB (1.25, 0.5)	(0.00)	4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03	(0.24)	1.52E-03	JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01	JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)	3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01
Barnes 1	(0.00)		3.97E-03	1.45E-02	1.24E-02	1.86E-03	1.24E-02	6.40E-03																																																																																																																																																																																																																																	
	(6.89)		1.23E-03						JTB (2.0, 1.0)	(0.00)	4.58E-03	6.60E-02	6.30E-02	2.39E-02	5.77E-02	4.23E-02	(3.00)	1.89E-03	JTB (0.75, 0.5)	(0.00)	5.69E-03	3.93E-02	3.62E-02	9.57E-03	3.33E-02	2.21E-02	(1.20)	2.22E-03	JTB (4.0, 1.0)	(0.00)	4.33E-03	3.31E-02	2.78E-02	7.42E-03	2.95E-02	1.88E-02	(0.78)	1.59E-03	JTB (1.0, 0.5)	(0.00)	5.08E-03	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02	(0.60)	1.83E-03	JTB (1.25, 0.5)	(0.00)	4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03	(0.24)	1.52E-03	JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01	JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)	3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01			
JTB (2.0, 1.0)	(0.00)		4.58E-03	6.60E-02	6.30E-02	2.39E-02	5.77E-02	4.23E-02																																																																																																																																																																																																																																	
	(3.00)		1.89E-03						JTB (0.75, 0.5)	(0.00)	5.69E-03	3.93E-02	3.62E-02	9.57E-03	3.33E-02	2.21E-02	(1.20)	2.22E-03	JTB (4.0, 1.0)	(0.00)	4.33E-03	3.31E-02	2.78E-02	7.42E-03	2.95E-02	1.88E-02	(0.78)	1.59E-03	JTB (1.0, 0.5)	(0.00)	5.08E-03	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02	(0.60)	1.83E-03	JTB (1.25, 0.5)	(0.00)	4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03	(0.24)	1.52E-03	JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)	3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01												
JTB (0.75, 0.5)	(0.00)		5.69E-03	3.93E-02	3.62E-02	9.57E-03	3.33E-02	2.21E-02																																																																																																																																																																																																																																	
	(1.20)		2.22E-03						JTB (4.0, 1.0)	(0.00)	4.33E-03	3.31E-02	2.78E-02	7.42E-03	2.95E-02	1.88E-02	(0.78)	1.59E-03	JTB (1.0, 0.5)	(0.00)	5.08E-03	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02	(0.60)	1.83E-03	JTB (1.25, 0.5)	(0.00)	4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03	(0.24)	1.52E-03	JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																					
JTB (4.0, 1.0)	(0.00)		4.33E-03	3.31E-02	2.78E-02	7.42E-03	2.95E-02	1.88E-02																																																																																																																																																																																																																																	
	(0.78)		1.59E-03						JTB (1.0, 0.5)	(0.00)	5.08E-03	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02	(0.60)	1.83E-03	JTB (1.25, 0.5)	(0.00)	4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03	(0.24)	1.52E-03	JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																														
JTB (1.0, 0.5)	(0.00)		5.08E-03	2.71E-02	2.41E-02	5.18E-03	2.34E-02	1.40E-02																																																																																																																																																																																																																																	
	(0.60)		1.83E-03						JTB (1.25, 0.5)	(0.00)	4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03	(0.24)	1.52E-03	JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																							
JTB (1.25, 0.5)	(0.00)		4.47E-03	1.94E-02	1.68E-02	3.03E-03	1.67E-02	9.29E-03																																																																																																																																																																																																																																	
	(0.24)		1.52E-03						JTB (1.35, 0.5)	(0.00)	4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03	(0.13)	1.38E-03	JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01		2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																
JTB (1.35, 0.5)	(0.00)		4.19E-03	1.69E-02	1.45E-02	2.44E-03	1.46E-02	7.86E-03																																																																																																																																																																																																																																	
	(0.13)		1.38E-03						JTB (1.5, 0.5)	(0.00)	4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03	(0.00)	1.26E-03	JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01		2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01		1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																									
JTB (1.5, 0.5)	(0.00)		4.00E-03	1.45E-02	1.24E-02	1.88E-03	1.25E-02	6.44E-03																																																																																																																																																																																																																																	
	(0.00)		1.26E-03						JTB (2.0, 0.5)	(0.00)	3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03	(-0.30)	9.07E-04	JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01		2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01		1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01		2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																		
JTB (2.0, 0.5)	(0.00)		3.22E-03	8.52E-03	7.47E-03	8.68E-04	7.08E-03	3.27E-03																																																																																																																																																																																																																																	
	(-0.30)		9.07E-04						JTB (4.0, 0.5)	(0.00)	1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04	(-0.75)	3.43E-04	JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01		2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01		1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01		2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01		2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																											
JTB (4.0, 0.5)	(0.00)		1.65E-03	1.55E-03	2.00E-03	1.02E-04	1.20E-03	3.71E-04																																																																																																																																																																																																																																	
	(-0.75)		3.43E-04						JTB (9.0, 0.5)	(0.00)	8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06	(-1.00)	1.28E-04	Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01		2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01		1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01		2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01		2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)		1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																				
JTB (9.0, 0.5)	(0.00)		8.09E-04	6.28E-05	4.81E-04	9.75E-06	4.28E-05	5.75E-06																																																																																																																																																																																																																																	
	(-1.00)		1.28E-04						Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01	(1049)	4.20E-03	Barnes 1		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01		JTB (2.0, 1.0)	(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)		3.93E-02	JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01		(1.20)	6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01		1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01		2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01		1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01		2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01		2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)		1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)		(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																													
Barnes 3	(3.00)	k=2	9.49E-03	1.80E-01	1.76E-01	1.19E-01	1.76E-01	1.35E-01																																																																																																																																																																																																																																	
	(1049)		4.20E-03						Barnes 1	(0.00)		1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01	(6.89)	1.01E-01	JTB (2.0, 1.0)		(0.00)	7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)	3.93E-02		JTB (0.75, 0.5)	(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)		6.86E-02	JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01		(0.78)	6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01		2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01		2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01		1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01		2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01		2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)		2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																								
Barnes 1	(0.00)		1.66E-01	2.54E-01	2.95E-01	1.46E-01	2.37E-01	1.78E-01																																																																																																																																																																																																																																	
	(6.89)		1.01E-01						JTB (2.0, 1.0)	(0.00)		7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01	(3.00)	3.93E-02	JTB (0.75, 0.5)		(0.00)	1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)	6.86E-02		JTB (4.0, 1.0)	(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)		6.85E-02	JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01		(0.60)	8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01		1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01		2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01		1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01		2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01		2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																			
JTB (2.0, 1.0)	(0.00)		7.00E-02	2.90E-01	3.20E-01	1.87E-01	2.68E-01	2.22E-01																																																																																																																																																																																																																																	
	(3.00)		3.93E-02						JTB (0.75, 0.5)	(0.00)		1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01	(1.20)	6.86E-02	JTB (4.0, 1.0)		(0.00)	1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02		JTB (1.0, 0.5)	(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)		8.27E-02	JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01		(0.24)	9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01		1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01		2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01		1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01		2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																														
JTB (0.75, 0.5)	(0.00)		1.14E-01	2.85E-01	3.20E-01	1.78E-01	2.62E-01	2.15E-01																																																																																																																																																																																																																																	
	(1.20)		6.86E-02						JTB (4.0, 1.0)	(0.00)		1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01	(0.78)	6.85E-02	JTB (1.0, 0.5)		(0.00)	1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02		JTB (1.25, 0.5)	(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)		9.25E-02	JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01		(0.13)	9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01		1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01		2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01		1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																									
JTB (4.0, 1.0)	(0.00)		1.18E-01	2.68E-01	3.03E-01	1.63E-01	2.52E-01	1.98E-01																																																																																																																																																																																																																																	
	(0.78)		6.85E-02						JTB (1.0, 0.5)	(0.00)		1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01	(0.60)	8.27E-02	JTB (1.25, 0.5)		(0.00)	1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02		JTB (1.35, 0.5)	(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)		9.54E-02	JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01		(0.00)	1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01		1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01		1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																				
JTB (1.0, 0.5)	(0.00)		1.36E-01	2.73E-01	3.10E-01	1.64E-01	2.54E-01	2.00E-01																																																																																																																																																																																																																																	
	(0.60)		8.27E-02						JTB (1.25, 0.5)	(0.00)		1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01	(0.24)	9.25E-02	JTB (1.35, 0.5)		(0.00)	1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02		JTB (1.5, 0.5)	(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)		1.01E-01	JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01		(-0.30)	1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01		1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																															
JTB (1.25, 0.5)	(0.00)		1.52E-01	2.61E-01	3.00E-01	1.53E-01	2.44E-01	1.87E-01																																																																																																																																																																																																																																	
	(0.24)		9.25E-02						JTB (1.35, 0.5)	(0.00)		1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01	(0.13)	9.54E-02	JTB (1.5, 0.5)		(0.00)	1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01		JTB (2.0, 0.5)	(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)		1.14E-01	JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01		(-0.75)	1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																																										
JTB (1.35, 0.5)	(0.00)		1.57E-01	2.57E-01	2.96E-01	1.49E-01	2.40E-01	1.82E-01																																																																																																																																																																																																																																	
	(0.13)		9.54E-02						JTB (1.5, 0.5)	(0.00)		1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01	(0.00)	1.01E-01	JTB (2.0, 0.5)		(0.00)	1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01		JTB (4.0, 0.5)	(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)		1.44E-01	JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																																																					
JTB (1.5, 0.5)	(0.00)		1.66E-01	2.54E-01	2.95E-01	1.45E-01	2.38E-01	1.78E-01																																																																																																																																																																																																																																	
	(0.00)		1.01E-01						JTB (2.0, 0.5)	(0.00)		1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01	(-0.30)	1.14E-01	JTB (4.0, 0.5)		(0.00)	2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01		JTB (9.0, 0.5)	(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																																																																
JTB (2.0, 0.5)	(0.00)		1.87E-01	2.40E-01	2.85E-01	1.33E-01	2.22E-01	1.62E-01																																																																																																																																																																																																																																	
	(-0.30)		1.14E-01						JTB (4.0, 0.5)	(0.00)		2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01	(-0.75)	1.44E-01	JTB (9.0, 0.5)		(0.00)	2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																																																																											
JTB (4.0, 0.5)	(0.00)		2.34E-01	2.09E-01	2.67E-01	1.09E-01	1.89E-01	1.24E-01																																																																																																																																																																																																																																	
	(-0.75)		1.44E-01						JTB (9.0, 0.5)	(0.00)		2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02	(-1.00)	1.76E-01																																																																																																																																																																																																																						
JTB (9.0, 0.5)	(0.00)		2.80E-01	1.72E-01	2.54E-01	9.03E-02	1.50E-01	8.22E-02																																																																																																																																																																																																																																	
	(-1.00)		1.76E-01																																																																																																																																																																																																																																						

Table 6 (continued): Power Comparison Study for Symmetric Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																				
Student (5)	(0.00)	k=1	3.27E-03	6.24E-02	5.61E-02	2.52E-02	5.76E-02	4.17E-02																																																																																																																																																																																																				
	(6.00)		1.27E-03						Student (6)	(0.00)	3.41E-03	5.31E-02	4.67E-02	1.90E-02	4.87E-02	3.44E-02	(3.00)	1.31E-03	Student (8)	(0.00)	3.79E-03	4.23E-02	3.63E-02	1.27E-02	3.84E-02	2.62E-02	(1.50)	1.41E-03	Student (16)	(0.00)	4.09E-03	2.68E-02	2.22E-02	5.48E-03	2.38E-02	1.46E-02	(0.50)	1.42E-03	Student (25)	(0.00)	4.20E-03	2.22E-02	1.83E-02	3.86E-03	1.94E-02	1.13E-02	(0.29)	1.41E-03	Student (32)	(0.00)	4.01E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03	(0.21)	1.35E-03	Student (40)	(0.00)	4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03	(0.17)	1.34E-03	Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)	(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02	Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)	5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01	(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01	Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01
Student (6)	(0.00)		3.41E-03	5.31E-02	4.67E-02	1.90E-02	4.87E-02	3.44E-02																																																																																																																																																																																																				
	(3.00)		1.31E-03						Student (8)	(0.00)	3.79E-03	4.23E-02	3.63E-02	1.27E-02	3.84E-02	2.62E-02	(1.50)	1.41E-03	Student (16)	(0.00)	4.09E-03	2.68E-02	2.22E-02	5.48E-03	2.38E-02	1.46E-02	(0.50)	1.42E-03	Student (25)	(0.00)	4.20E-03	2.22E-02	1.83E-02	3.86E-03	1.94E-02	1.13E-02	(0.29)	1.41E-03	Student (32)	(0.00)	4.01E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03	(0.21)	1.35E-03	Student (40)	(0.00)	4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03	(0.17)	1.34E-03	Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)		(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02	Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)	5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01	(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01		
Student (8)	(0.00)		3.79E-03	4.23E-02	3.63E-02	1.27E-02	3.84E-02	2.62E-02																																																																																																																																																																																																				
	(1.50)		1.41E-03						Student (16)	(0.00)	4.09E-03	2.68E-02	2.22E-02	5.48E-03	2.38E-02	1.46E-02	(0.50)	1.42E-03	Student (25)	(0.00)	4.20E-03	2.22E-02	1.83E-02	3.86E-03	1.94E-02	1.13E-02	(0.29)	1.41E-03	Student (32)	(0.00)	4.01E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03	(0.21)	1.35E-03	Student (40)	(0.00)	4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03	(0.17)	1.34E-03	Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)		(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02		Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)	5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01	(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01		(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01										
Student (16)	(0.00)		4.09E-03	2.68E-02	2.22E-02	5.48E-03	2.38E-02	1.46E-02																																																																																																																																																																																																				
	(0.50)		1.42E-03						Student (25)	(0.00)	4.20E-03	2.22E-02	1.83E-02	3.86E-03	1.94E-02	1.13E-02	(0.29)	1.41E-03	Student (32)	(0.00)	4.01E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03	(0.21)	1.35E-03	Student (40)	(0.00)	4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03	(0.17)	1.34E-03	Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)		(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02		Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)		5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01	(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01		(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01		4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																		
Student (25)	(0.00)		4.20E-03	2.22E-02	1.83E-02	3.86E-03	1.94E-02	1.13E-02																																																																																																																																																																																																				
	(0.29)		1.41E-03						Student (32)	(0.00)	4.01E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03	(0.21)	1.35E-03	Student (40)	(0.00)	4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03	(0.17)	1.34E-03	Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)		(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02		Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)		5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01		(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01		(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01		4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01		5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																										
Student (32)	(0.00)		4.01E-03	1.97E-02	1.64E-02	3.21E-03	1.72E-02	9.79E-03																																																																																																																																																																																																				
	(0.21)		1.35E-03						Student (40)	(0.00)	4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03	(0.17)	1.34E-03	Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)		(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02		Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)		5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01		(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01		1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01		(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01		4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01		5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																		
Student (40)	(0.00)		4.08E-03	1.89E-02	1.57E-02	2.93E-03	1.65E-02	9.18E-03																																																																																																																																																																																																				
	(0.17)		1.34E-03						Student (5)	(0.00)	k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01	2.00E-01	(6.00)	3.24E-02	Student (6)		(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02		Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)		5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01		(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01		1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01		2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01		(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01		4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01		5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01		Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																										
Student (5)	(0.00)		k=2	6.15E-02	2.69E-01	2.98E-01	1.69E-01	2.54E-01		2.00E-01																																																																																																																																																																																																		
	(6.00)			3.24E-02					Student (6)			(0.00)	7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01	1.97E-01	(3.00)	4.02E-02		Student (8)	(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)		5.30E-02	Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01		(0.50)	7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01		1.86E-01	(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01		2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01		1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)	(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01		(3.00)	1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01		4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01		5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01		Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01		(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																		
Student (6)	(0.00)			7.43E-02	2.67E-01	2.97E-01	1.66E-01	2.51E-01		1.97E-01																																																																																																																																																																																																		
	(3.00)			4.02E-02					Student (8)			(0.00)	9.47E-02	2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01	(1.50)	5.30E-02		Student (16)	(0.00)	1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01	(0.50)		7.59E-02	Student (25)	(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01		(0.29)	8.55E-02	Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01		1.82E-01	(0.21)	8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01		2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01		5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01	Student (6)		(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)		1.45E-01	Student (8)	(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01		4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)	3.60E-01	5.29E-01	5.98E-01		4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01		5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)		(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)		2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01		4.48E-01	(0.17)	3.03E-01																																																											
Student (8)	(0.00)	9.47E-02		2.66E-01	3.00E-01	1.64E-01	2.50E-01	1.96E-01																																																																																																																																																																																																				
	(1.50)	5.30E-02							Student (16)	(0.00)		1.30E-01	2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01	(0.50)	7.59E-02	Student (25)		(0.00)	1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02		Student (32)	(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)		8.77E-02	Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01		(0.17)	9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01		4.62E-01	3.86E-01	(6.00)	1.17E-01	Student (6)	(0.00)		2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)	(0.00)		2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01		Student (16)	(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01		(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01		5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01		6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)		4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																								
Student (16)	(0.00)	1.30E-01		2.61E-01	2.99E-01	1.56E-01	2.45E-01	1.89E-01																																																																																																																																																																																																				
	(0.50)	7.59E-02							Student (25)	(0.00)		1.44E-01	2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01	(0.29)	8.55E-02	Student (32)		(0.00)	1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02		Student (40)	(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)		9.15E-02	Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01		3.86E-01	(6.00)	1.17E-01	Student (6)		(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)	(0.00)		2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)		3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01		Student (25)	(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01		(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01		5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01		6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																			
Student (25)	(0.00)	1.44E-01		2.60E-01	2.98E-01	1.53E-01	2.43E-01	1.86E-01																																																																																																																																																																																																				
	(0.29)	8.55E-02							Student (32)	(0.00)		1.47E-01	2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01	(0.21)	8.77E-02	Student (40)		(0.00)	1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02		Student (5)	(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01		(6.00)	1.17E-01	Student (6)		(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)		(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)		3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01		Student (32)	(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01		(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01		5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																														
Student (32)	(0.00)	1.47E-01		2.56E-01	2.95E-01	1.50E-01	2.39E-01	1.82E-01																																																																																																																																																																																																				
	(0.21)	8.77E-02							Student (40)	(0.00)		1.52E-01	2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01	(0.17)	9.15E-02	Student (5)		(0.00)	k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)		1.17E-01	Student (6)		(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)		(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)		(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)		3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01		Student (40)	(0.00)	4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01		(0.17)	3.03E-01																																																																																																									
Student (40)	(0.00)	1.52E-01		2.57E-01	2.96E-01	1.50E-01	2.41E-01	1.83E-01																																																																																																																																																																																																				
	(0.17)	9.15E-02							Student (5)	(0.00)		k=3	1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01	3.86E-01	(6.00)	1.17E-01		Student (6)		(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)		(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)		(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)		(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)		3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)		4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																					
Student (5)	(0.00)	k=3		1.88E-01	4.81E-01	5.45E-01	3.73E-01	4.62E-01		3.86E-01																																																																																																																																																																																																		
	(6.00)			1.17E-01					Student (6)				(0.00)	2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01	3.97E-01	(3.00)	1.45E-01	Student (8)		(0.00)	2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)		(0.00)	3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)		(0.00)	3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)		(0.00)	3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)		4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																																
Student (6)	(0.00)			2.24E-01	4.91E-01	5.57E-01	3.86E-01	4.72E-01		3.97E-01																																																																																																																																																																																																		
	(3.00)			1.45E-01					Student (8)		(0.00)		2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01	(1.50)	1.87E-01	Student (16)	(0.00)		3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)		3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)		4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																																												
Student (8)	(0.00)		2.77E-01	5.08E-01	5.77E-01	4.03E-01	4.88E-01	4.16E-01																																																																																																																																																																																																				
	(1.50)		1.87E-01						Student (16)	(0.00)	3.60E-01		5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01	(0.50)	2.58E-01	Student (25)	(0.00)	3.90E-01		5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01		5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01		5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																																																								
Student (16)	(0.00)		3.60E-01	5.29E-01	5.98E-01	4.23E-01	5.10E-01	4.38E-01																																																																																																																																																																																																				
	(0.50)		2.58E-01						Student (25)	(0.00)	3.90E-01		5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01	(0.29)	2.86E-01	Student (32)	(0.00)	3.97E-01		5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01		5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																																																																			
Student (25)	(0.00)		3.90E-01	5.37E-01	6.06E-01	4.29E-01	5.18E-01	4.45E-01																																																																																																																																																																																																				
	(0.29)		2.86E-01						Student (32)	(0.00)	3.97E-01		5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01	(0.21)	2.93E-01	Student (40)	(0.00)	4.07E-01		5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																																																																														
Student (32)	(0.00)		3.97E-01	5.37E-01	6.06E-01	4.29E-01	5.17E-01	4.44E-01																																																																																																																																																																																																				
	(0.21)		2.93E-01						Student (40)	(0.00)	4.07E-01		5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01	(0.17)	3.03E-01																																																																																																																																																																																									
Student (40)	(0.00)		4.07E-01	5.41E-01	6.10E-01	4.32E-01	5.22E-01	4.48E-01																																																																																																																																																																																																				
	(0.17)		3.03E-01																																																																																																																																																																																																									



A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 6 (continued): Power Comparison Study for Symmetric Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Barnes 3	(3.00)	k=3	3.00E-02	2.41E-01	3.62E-01	1.91E-01	2.33E-01	1.78E-01																																																																																																																																																																																																																																	
	(1049)		1.32E-02						Barnes 1	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.38E-01	5.28E-01	4.54E-01	(6.89)	3.29E-01	JTB (2.0, 1.0)	(0.00)	1.91E-01	4.93E-01	5.47E-01	3.88E-01	4.66E-01	4.10E-01	(3.00)	1.24E-01	JTB (0.75, 0.5)	(0.00)	2.95E-01	5.23E-01	5.82E-01	4.18E-01	4.97E-01	4.40E-01	(1.20)	2.09E-01	JTB (4.0, 1.0)	(0.00)	3.28E-01	5.22E-01	5.90E-01	4.17E-01	5.02E-01	4.34E-01	(0.78)	2.31E-01	JTB (1.0, 0.5)	(0.00)	3.55E-01	5.34E-01	5.97E-01	4.27E-01	5.13E-01	4.47E-01	(0.60)	2.59E-01	JTB (1.25, 0.5)	(0.00)	3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01	(0.24)	2.97E-01	JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01	JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)	2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01
Barnes 1	(0.00)		4.36E-01	5.49E-01	6.17E-01	4.38E-01	5.28E-01	4.54E-01																																																																																																																																																																																																																																	
	(6.89)		3.29E-01						JTB (2.0, 1.0)	(0.00)	1.91E-01	4.93E-01	5.47E-01	3.88E-01	4.66E-01	4.10E-01	(3.00)	1.24E-01	JTB (0.75, 0.5)	(0.00)	2.95E-01	5.23E-01	5.82E-01	4.18E-01	4.97E-01	4.40E-01	(1.20)	2.09E-01	JTB (4.0, 1.0)	(0.00)	3.28E-01	5.22E-01	5.90E-01	4.17E-01	5.02E-01	4.34E-01	(0.78)	2.31E-01	JTB (1.0, 0.5)	(0.00)	3.55E-01	5.34E-01	5.97E-01	4.27E-01	5.13E-01	4.47E-01	(0.60)	2.59E-01	JTB (1.25, 0.5)	(0.00)	3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01	(0.24)	2.97E-01	JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01	JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)	2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01			
JTB (2.0, 1.0)	(0.00)		1.91E-01	4.93E-01	5.47E-01	3.88E-01	4.66E-01	4.10E-01																																																																																																																																																																																																																																	
	(3.00)		1.24E-01						JTB (0.75, 0.5)	(0.00)	2.95E-01	5.23E-01	5.82E-01	4.18E-01	4.97E-01	4.40E-01	(1.20)	2.09E-01	JTB (4.0, 1.0)	(0.00)	3.28E-01	5.22E-01	5.90E-01	4.17E-01	5.02E-01	4.34E-01	(0.78)	2.31E-01	JTB (1.0, 0.5)	(0.00)	3.55E-01	5.34E-01	5.97E-01	4.27E-01	5.13E-01	4.47E-01	(0.60)	2.59E-01	JTB (1.25, 0.5)	(0.00)	3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01	(0.24)	2.97E-01	JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)	2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01												
JTB (0.75, 0.5)	(0.00)		2.95E-01	5.23E-01	5.82E-01	4.18E-01	4.97E-01	4.40E-01																																																																																																																																																																																																																																	
	(1.20)		2.09E-01						JTB (4.0, 1.0)	(0.00)	3.28E-01	5.22E-01	5.90E-01	4.17E-01	5.02E-01	4.34E-01	(0.78)	2.31E-01	JTB (1.0, 0.5)	(0.00)	3.55E-01	5.34E-01	5.97E-01	4.27E-01	5.13E-01	4.47E-01	(0.60)	2.59E-01	JTB (1.25, 0.5)	(0.00)	3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01	(0.24)	2.97E-01	JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																					
JTB (4.0, 1.0)	(0.00)		3.28E-01	5.22E-01	5.90E-01	4.17E-01	5.02E-01	4.34E-01																																																																																																																																																																																																																																	
	(0.78)		2.31E-01						JTB (1.0, 0.5)	(0.00)	3.55E-01	5.34E-01	5.97E-01	4.27E-01	5.13E-01	4.47E-01	(0.60)	2.59E-01	JTB (1.25, 0.5)	(0.00)	3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01	(0.24)	2.97E-01	JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																														
JTB (1.0, 0.5)	(0.00)		3.55E-01	5.34E-01	5.97E-01	4.27E-01	5.13E-01	4.47E-01																																																																																																																																																																																																																																	
	(0.60)		2.59E-01						JTB (1.25, 0.5)	(0.00)	3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01	(0.24)	2.97E-01	JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																							
JTB (1.25, 0.5)	(0.00)		3.99E-01	5.41E-01	6.06E-01	4.31E-01	5.21E-01	4.49E-01																																																																																																																																																																																																																																	
	(0.24)		2.97E-01						JTB (1.35, 0.5)	(0.00)	4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01	(0.13)	3.09E-01	JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01		6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																
JTB (1.35, 0.5)	(0.00)		4.13E-01	5.42E-01	6.09E-01	4.32E-01	5.22E-01	4.50E-01																																																																																																																																																																																																																																	
	(0.13)		3.09E-01						JTB (1.5, 0.5)	(0.00)	4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01	(0.00)	3.30E-01	JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01		6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01		6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																									
JTB (1.5, 0.5)	(0.00)		4.36E-01	5.49E-01	6.17E-01	4.37E-01	5.30E-01	4.54E-01																																																																																																																																																																																																																																	
	(0.00)		3.30E-01						JTB (2.0, 0.5)	(0.00)	4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01	(-0.30)	3.78E-01	JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01		6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01		6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01		7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																		
JTB (2.0, 0.5)	(0.00)		4.88E-01	5.60E-01	6.29E-01	4.44E-01	5.39E-01	4.59E-01																																																																																																																																																																																																																																	
	(-0.30)		3.78E-01						JTB (4.0, 0.5)	(0.00)	5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01	(-0.75)	4.84E-01	JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01		6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01		6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01		7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01		7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																											
JTB (4.0, 0.5)	(0.00)		5.92E-01	5.90E-01	6.57E-01	4.59E-01	5.66E-01	4.72E-01																																																																																																																																																																																																																																	
	(-0.75)		4.84E-01						JTB (9.0, 0.5)	(0.00)	6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01	(-1.00)	5.70E-01	Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01		6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01		6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01		7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01		7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)		6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																				
JTB (9.0, 0.5)	(0.00)		6.69E-01	6.24E-01	6.78E-01	4.72E-01	5.96E-01	4.89E-01																																																																																																																																																																																																																																	
	(-1.00)		5.70E-01						Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01	(1049)	3.47E-02	Barnes 1		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01		JTB (2.0, 1.0)	(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)		2.25E-01	JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01		(1.20)	3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01		6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01		6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01		6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01		7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01		7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)		6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)		(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																													
Barnes 3	(3.00)	k=4	7.68E-02	3.92E-01	6.06E-01	3.16E-01	3.67E-01	2.52E-01																																																																																																																																																																																																																																	
	(1049)		3.47E-02						Barnes 1	(0.00)		6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01	(6.89)	5.30E-01	JTB (2.0, 1.0)		(0.00)	3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)	2.25E-01		JTB (0.75, 0.5)	(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)		3.54E-01	JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01		(0.78)	4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01		6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01		7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01		6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01		7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01		7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)		7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																								
Barnes 1	(0.00)		6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.21E-01	6.60E-01																																																																																																																																																																																																																																	
	(6.89)		5.30E-01						JTB (2.0, 1.0)	(0.00)		3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01	(3.00)	2.25E-01	JTB (0.75, 0.5)		(0.00)	4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)	3.54E-01		JTB (4.0, 1.0)	(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)		4.02E-01	JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01		(0.60)	4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01		6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01		7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01		6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01		8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01		8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																			
JTB (2.0, 1.0)	(0.00)		3.20E-01	6.37E-01	6.97E-01	5.51E-01	6.12E-01	5.57E-01																																																																																																																																																																																																																																	
	(3.00)		2.25E-01						JTB (0.75, 0.5)	(0.00)		4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01	(1.20)	3.54E-01	JTB (4.0, 1.0)		(0.00)	5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01		JTB (1.0, 0.5)	(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)		4.30E-01	JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01		(0.24)	4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01		6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01		7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01		6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01		8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																														
JTB (0.75, 0.5)	(0.00)		4.59E-01	6.83E-01	7.42E-01	6.03E-01	6.60E-01	6.09E-01																																																																																																																																																																																																																																	
	(1.20)		3.54E-01						JTB (4.0, 1.0)	(0.00)		5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01	(0.78)	4.02E-01	JTB (1.0, 0.5)		(0.00)	5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01		JTB (1.25, 0.5)	(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)		4.86E-01	JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01		(0.13)	5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01		6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01		7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01		7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																									
JTB (4.0, 1.0)	(0.00)		5.15E-01	6.97E-01	7.64E-01	6.21E-01	6.80E-01	6.18E-01																																																																																																																																																																																																																																	
	(0.78)		4.02E-01						JTB (1.0, 0.5)	(0.00)		5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01	(0.60)	4.30E-01	JTB (1.25, 0.5)		(0.00)	5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01		JTB (1.35, 0.5)	(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)		5.03E-01	JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01		(0.00)	5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01		6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01		7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																				
JTB (1.0, 0.5)	(0.00)		5.38E-01	7.06E-01	7.67E-01	6.29E-01	6.88E-01	6.31E-01																																																																																																																																																																																																																																	
	(0.60)		4.30E-01						JTB (1.25, 0.5)	(0.00)		5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01	(0.24)	4.86E-01	JTB (1.35, 0.5)		(0.00)	6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01		JTB (1.5, 0.5)	(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)		5.30E-01	JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01		(-0.30)	5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01		7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																															
JTB (1.25, 0.5)	(0.00)		5.92E-01	7.23E-01	7.83E-01	6.45E-01	7.06E-01	6.46E-01																																																																																																																																																																																																																																	
	(0.24)		4.86E-01						JTB (1.35, 0.5)	(0.00)		6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01	(0.13)	5.03E-01	JTB (1.5, 0.5)		(0.00)	6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01		JTB (2.0, 0.5)	(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)		5.93E-01	JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01		(-0.75)	7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																																										
JTB (1.35, 0.5)	(0.00)		6.09E-01	7.28E-01	7.88E-01	6.50E-01	7.12E-01	6.51E-01																																																																																																																																																																																																																																	
	(0.13)		5.03E-01						JTB (1.5, 0.5)	(0.00)		6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01	(0.00)	5.30E-01	JTB (2.0, 0.5)		(0.00)	6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01		JTB (4.0, 0.5)	(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)		7.07E-01	JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																																																					
JTB (1.5, 0.5)	(0.00)		6.35E-01	7.38E-01	7.97E-01	6.61E-01	7.23E-01	6.60E-01																																																																																																																																																																																																																																	
	(0.00)		5.30E-01						JTB (2.0, 0.5)	(0.00)		6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01	(-0.30)	5.93E-01	JTB (4.0, 0.5)		(0.00)	7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01		JTB (9.0, 0.5)	(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																																																																
JTB (2.0, 0.5)	(0.00)		6.91E-01	7.60E-01	8.15E-01	6.80E-01	7.43E-01	6.80E-01																																																																																																																																																																																																																																	
	(-0.30)		5.93E-01						JTB (4.0, 0.5)	(0.00)		7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01	(-0.75)	7.07E-01	JTB (9.0, 0.5)		(0.00)	8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																																																																											
JTB (4.0, 0.5)	(0.00)		7.86E-01	8.07E-01	8.47E-01	7.19E-01	7.92E-01	7.27E-01																																																																																																																																																																																																																																	
	(-0.75)		7.07E-01						JTB (9.0, 0.5)	(0.00)		8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01	(-1.00)	7.80E-01																																																																																																																																																																																																																						
JTB (9.0, 0.5)	(0.00)		8.41E-01	8.46E-01	8.67E-01	7.46E-01	8.31E-01	7.70E-01																																																																																																																																																																																																																																	
	(-1.00)		7.80E-01																																																																																																																																																																																																																																						

Table 6 (continued): Power Comparison Study for Symmetric Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart	
Student (5)	(0.00)	k=4	3.29E-01	6.40E-01	7.13E-01	5.51E-01	6.22E-01	5.44E-01	
	(6.00)		2.28E-01						
Student (6)	(0.00)		3.82E-01	6.57E-01	7.29E-01	5.73E-01	6.39E-01	5.63E-01	
	(3.00)		2.74E-01						
Student (8)	(0.00)		4.53E-01	6.81E-01	7.51E-01	6.02E-01	6.63E-01	5.95E-01	
	(1.50)		3.40E-01						
Student (16)	(0.00)		5.54E-01	7.11E-01	7.76E-01	6.35E-01	6.94E-01	6.31E-01	
	(0.50)		4.41E-01						
Student (25)	(0.00)		5.87E-01	7.22E-01	7.85E-01	6.45E-01	7.06E-01	6.43E-01	
	(0.29)		4.76E-01						
Student (32)	(0.00)		5.95E-01	7.23E-01	7.86E-01	6.47E-01	7.07E-01	6.44E-01	
	(0.21)		4.86E-01						
Student (40)	(0.00)		6.06E-01	7.27E-01	7.89E-01	6.51E-01	7.12E-01	6.49E-01	
	(0.17)		4.98E-01						
Student (5)	(0.00)		k=5	4.54E-01	7.50E-01	8.17E-01	6.83E-01	7.34E-01	6.62E-01
	(6.00)			3.37E-01					
Student (6)	(0.00)	5.15E-01		7.68E-01	8.31E-01	7.07E-01	7.53E-01	6.85E-01	
	(3.00)	3.96E-01							
Student (8)	(0.00)	5.90E-01		7.91E-01	8.50E-01	7.37E-01	7.77E-01	7.19E-01	
	(1.50)	4.74E-01							
Student (16)	(0.00)	6.89E-01		8.20E-01	8.71E-01	7.70E-01	8.08E-01	7.58E-01	
	(0.50)	5.83E-01							
Student (25)	(0.00)	7.18E-01		8.30E-01	8.78E-01	7.80E-01	8.18E-01	7.70E-01	
	(0.29)	6.19E-01							
Student (32)	(0.00)	7.26E-01		8.32E-01	8.79E-01	7.82E-01	8.20E-01	7.73E-01	
	(0.21)	6.28E-01							
Student (40)	(0.00)	7.36E-01		8.35E-01	8.81E-01	7.86E-01	8.24E-01	7.77E-01	
	(0.17)	6.40E-01							
Student (5)	(0.00)	k=6		5.57E-01	8.24E-01	8.80E-01	7.75E-01	8.11E-01	7.48E-01
	(6.00)			4.35E-01					
Student (6)	(0.00)		6.19E-01	8.40E-01	8.92E-01	7.97E-01	8.28E-01	7.71E-01	
	(3.00)		5.00E-01						
Student (8)	(0.00)		6.90E-01	8.60E-01	9.07E-01	8.23E-01	8.50E-01	8.04E-01	
	(1.50)		5.80E-01						
Student (16)	(0.00)		7.79E-01	8.85E-01	9.23E-01	8.52E-01	8.76E-01	8.39E-01	
	(0.50)		6.86E-01						
Student (25)	(0.00)		8.04E-01	8.93E-01	9.27E-01	8.61E-01	8.84E-01	8.49E-01	
	(0.29)		7.18E-01						
Student (32)	(0.00)		8.11E-01	8.94E-01	9.28E-01	8.63E-01	8.86E-01	8.52E-01	
	(0.21)		7.28E-01						
Student (40)	(0.00)		8.18E-01	8.97E-01	9.30E-01	8.65E-01	8.89E-01	8.55E-01	
	(0.17)		7.38E-01						

A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

Table 6 (continued): Power Comparison Study for Symmetric Distributions ( $n = 10$ )

Distribution	(skewness) (kurtosis)	k	Combined Sample	R-chart	S-chart	S2-chart	WV-chart	SC-chart																																																																																																																																																																																																																																	
Barnes 3	(3.00)	k=5	1.53E-01	5.89E-01	7.79E-01	4.59E-01	5.53E-01	3.69E-01																																																																																																																																																																																																																																	
	(1049)		7.54E-02						Barnes 1	(0.00)	7.61E-01	8.44E-01	8.88E-01	7.95E-01	8.33E-01	7.88E-01	(6.89)	6.71E-01	JTB (2.0, 1.0)	(0.00)	4.34E-01	7.36E-01	7.93E-01	6.69E-01	7.14E-01	6.65E-01	(3.00)	3.24E-01	JTB (0.75, 0.5)	(0.00)	5.85E-01	7.84E-01	8.34E-01	7.27E-01	7.65E-01	7.24E-01	(1.20)	4.76E-01	JTB (4.0, 1.0)	(0.00)	6.52E-01	8.06E-01	8.60E-01	7.55E-01	7.93E-01	7.43E-01	(0.78)	5.41E-01	JTB (1.0, 0.5)	(0.00)	6.68E-01	8.11E-01	8.60E-01	7.58E-01	7.97E-01	7.52E-01	(0.60)	5.65E-01	JTB (1.25, 0.5)	(0.00)	7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01	(0.24)	6.25E-01	JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1	(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01	JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01
Barnes 1	(0.00)		7.61E-01	8.44E-01	8.88E-01	7.95E-01	8.33E-01	7.88E-01																																																																																																																																																																																																																																	
	(6.89)		6.71E-01						JTB (2.0, 1.0)	(0.00)	4.34E-01	7.36E-01	7.93E-01	6.69E-01	7.14E-01	6.65E-01	(3.00)	3.24E-01	JTB (0.75, 0.5)	(0.00)	5.85E-01	7.84E-01	8.34E-01	7.27E-01	7.65E-01	7.24E-01	(1.20)	4.76E-01	JTB (4.0, 1.0)	(0.00)	6.52E-01	8.06E-01	8.60E-01	7.55E-01	7.93E-01	7.43E-01	(0.78)	5.41E-01	JTB (1.0, 0.5)	(0.00)	6.68E-01	8.11E-01	8.60E-01	7.58E-01	7.97E-01	7.52E-01	(0.60)	5.65E-01	JTB (1.25, 0.5)	(0.00)	7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01	(0.24)	6.25E-01	JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01	JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01			
JTB (2.0, 1.0)	(0.00)		4.34E-01	7.36E-01	7.93E-01	6.69E-01	7.14E-01	6.65E-01																																																																																																																																																																																																																																	
	(3.00)		3.24E-01						JTB (0.75, 0.5)	(0.00)	5.85E-01	7.84E-01	8.34E-01	7.27E-01	7.65E-01	7.24E-01	(1.20)	4.76E-01	JTB (4.0, 1.0)	(0.00)	6.52E-01	8.06E-01	8.60E-01	7.55E-01	7.93E-01	7.43E-01	(0.78)	5.41E-01	JTB (1.0, 0.5)	(0.00)	6.68E-01	8.11E-01	8.60E-01	7.58E-01	7.97E-01	7.52E-01	(0.60)	5.65E-01	JTB (1.25, 0.5)	(0.00)	7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01	(0.24)	6.25E-01	JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01												
JTB (0.75, 0.5)	(0.00)		5.85E-01	7.84E-01	8.34E-01	7.27E-01	7.65E-01	7.24E-01																																																																																																																																																																																																																																	
	(1.20)		4.76E-01						JTB (4.0, 1.0)	(0.00)	6.52E-01	8.06E-01	8.60E-01	7.55E-01	7.93E-01	7.43E-01	(0.78)	5.41E-01	JTB (1.0, 0.5)	(0.00)	6.68E-01	8.11E-01	8.60E-01	7.58E-01	7.97E-01	7.52E-01	(0.60)	5.65E-01	JTB (1.25, 0.5)	(0.00)	7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01	(0.24)	6.25E-01	JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																					
JTB (4.0, 1.0)	(0.00)		6.52E-01	8.06E-01	8.60E-01	7.55E-01	7.93E-01	7.43E-01																																																																																																																																																																																																																																	
	(0.78)		5.41E-01						JTB (1.0, 0.5)	(0.00)	6.68E-01	8.11E-01	8.60E-01	7.58E-01	7.97E-01	7.52E-01	(0.60)	5.65E-01	JTB (1.25, 0.5)	(0.00)	7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01	(0.24)	6.25E-01	JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01		(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																														
JTB (1.0, 0.5)	(0.00)		6.68E-01	8.11E-01	8.60E-01	7.58E-01	7.97E-01	7.52E-01																																																																																																																																																																																																																																	
	(0.60)		5.65E-01						JTB (1.25, 0.5)	(0.00)	7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01	(0.24)	6.25E-01	JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01		(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01		8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																							
JTB (1.25, 0.5)	(0.00)		7.21E-01	8.29E-01	8.75E-01	7.79E-01	8.17E-01	7.71E-01																																																																																																																																																																																																																																	
	(0.24)		6.25E-01						JTB (1.35, 0.5)	(0.00)	7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01	(0.13)	6.44E-01	JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01		(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01		8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01		8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																
JTB (1.35, 0.5)	(0.00)		7.37E-01	8.35E-01	8.80E-01	7.84E-01	8.23E-01	7.78E-01																																																																																																																																																																																																																																	
	(0.13)		6.44E-01						JTB (1.5, 0.5)	(0.00)	7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01	(0.00)	6.71E-01	JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01		(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01		8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01		8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01		8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																									
JTB (1.5, 0.5)	(0.00)		7.60E-01	8.45E-01	8.88E-01	7.95E-01	8.34E-01	7.88E-01																																																																																																																																																																																																																																	
	(0.00)		6.71E-01						JTB (2.0, 0.5)	(0.00)	8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01	(-0.30)	7.30E-01	JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01		(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01		8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01		8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01		8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01		9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																		
JTB (2.0, 0.5)	(0.00)		8.09E-01	8.66E-01	9.03E-01	8.16E-01	8.55E-01	8.11E-01																																																																																																																																																																																																																																	
	(-0.30)		7.30E-01						JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01	(-0.75)	8.25E-01	JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)		4.12E-01	JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01		(1.20)	5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01		8.24E-01	(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01		8.64E-01	8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01		8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01		9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01		9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																											
JTB (4.0, 0.5)	(0.00)	8.80E-01	9.05E-01	9.27E-01	8.53E-01	8.96E-01	8.58E-01																																																																																																																																																																																																																																		
	(-0.75)	8.25E-01						JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01	(-1.00)	8.76E-01	Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1	(0.00)		8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01	JTB (2.0, 1.0)		(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01		JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)		5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01		(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01		8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01		8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01		8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01		9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																						
JTB (9.0, 0.5)	(0.00)	9.15E-01	9.31E-01	9.40E-01	8.76E-01	9.23E-01	8.92E-01																																																																																																																																																																																																																																		
	(-1.00)	8.76E-01						Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01	4.87E-01	(1049)	1.35E-01	Barnes 1		(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01	JTB (2.0, 1.0)		(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01		JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)		5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01		(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01		8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01		8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01		8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01		9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																
Barnes 3	(3.00)	k=6	2.46E-01	7.49E-01	8.76E-01	5.85E-01	7.16E-01		4.87E-01																																																																																																																																																																																																																																
	(1049)		1.35E-01					Barnes 1			(0.00)	8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01	8.65E-01	(6.89)	7.64E-01		JTB (2.0, 1.0)	(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01		JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)		5.74E-01	JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01		(0.78)	6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01		8.30E-01	(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01		8.82E-01	8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01		8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01		9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																										
Barnes 1	(0.00)		8.38E-01	9.04E-01	9.35E-01	8.73E-01	8.96E-01		8.65E-01																																																																																																																																																																																																																																
	(6.89)		7.64E-01					JTB (2.0, 1.0)			(0.00)	5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01	7.42E-01	(3.00)	4.12E-01		JTB (0.75, 0.5)	(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)	5.74E-01		JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)		6.46E-01	JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01		(0.60)	6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01		8.49E-01	(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01		8.88E-01	8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01		8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																					
JTB (2.0, 1.0)	(0.00)		5.29E-01	8.04E-01	8.54E-01	7.53E-01	7.85E-01		7.42E-01																																																																																																																																																																																																																																
	(3.00)		4.12E-01					JTB (0.75, 0.5)			(0.00)	6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01	8.01E-01	(1.20)	5.74E-01		JTB (4.0, 1.0)	(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01		JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)		6.64E-01	JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01		(0.24)	7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01		8.55E-01	(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01		8.97E-01	8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																
JTB (0.75, 0.5)	(0.00)		6.79E-01	8.49E-01	8.90E-01	8.08E-01	8.34E-01		8.01E-01																																																																																																																																																																																																																																
	(1.20)		5.74E-01					JTB (4.0, 1.0)			(0.00)	7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01	8.24E-01	(0.78)	6.46E-01		JTB (1.0, 0.5)	(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01		JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)		7.23E-01	JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01		(0.13)	7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01		8.65E-01	(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																											
JTB (4.0, 1.0)	(0.00)		7.47E-01	8.72E-01	9.14E-01	8.38E-01	8.62E-01		8.24E-01																																																																																																																																																																																																																																
	(0.78)		6.46E-01					JTB (1.0, 0.5)			(0.00)	7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01	8.30E-01	(0.60)	6.64E-01		JTB (1.25, 0.5)	(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01		JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)		7.40E-01	JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01		(0.00)	7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																						
JTB (1.0, 0.5)	(0.00)		7.58E-01	8.74E-01	9.12E-01	8.39E-01	8.64E-01		8.30E-01																																																																																																																																																																																																																																
	(0.60)		6.64E-01					JTB (1.25, 0.5)			(0.00)	8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01	8.49E-01	(0.24)	7.23E-01		JTB (1.35, 0.5)	(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01		JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)		7.64E-01	JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																																	
JTB (1.25, 0.5)	(0.00)		8.05E-01	8.91E-01	9.25E-01	8.58E-01	8.82E-01		8.49E-01																																																																																																																																																																																																																																
	(0.24)		7.23E-01					JTB (1.35, 0.5)			(0.00)	8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01	8.55E-01	(0.13)	7.40E-01		JTB (1.5, 0.5)	(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01		JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																																												
JTB (1.35, 0.5)	(0.00)		8.19E-01	8.96E-01	9.29E-01	8.64E-01	8.88E-01		8.55E-01																																																																																																																																																																																																																																
	(0.13)		7.40E-01					JTB (1.5, 0.5)			(0.00)	8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01	8.65E-01	(0.00)	7.64E-01		JTB (2.0, 0.5)	(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																																																							
JTB (1.5, 0.5)	(0.00)		8.38E-01	9.04E-01	9.34E-01	8.73E-01	8.97E-01		8.65E-01																																																																																																																																																																																																																																
	(0.00)		7.64E-01					JTB (2.0, 0.5)			(0.00)	8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01	8.86E-01	(-0.30)	8.16E-01	JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																																																																		
JTB (2.0, 0.5)	(0.00)		8.76E-01	9.22E-01	9.46E-01	8.91E-01	9.15E-01		8.86E-01																																																																																																																																																																																																																																
	(-0.30)		8.16E-01					JTB (4.0, 0.5)		(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01	(-0.75)	8.89E-01	JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																																																																													
JTB (4.0, 0.5)	(0.00)	9.28E-01	9.50E-01	9.62E-01	9.20E-01	9.45E-01	9.23E-01																																																																																																																																																																																																																																		
	(-0.75)	8.89E-01						JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01	(-1.00)	9.24E-01																																																																																																																																																																																																																								
JTB (9.0, 0.5)	(0.00)	9.51E-01	9.65E-01	9.70E-01	9.35E-01	9.61E-01	9.45E-01																																																																																																																																																																																																																																		
	(-1.00)	9.24E-01																																																																																																																																																																																																																																							

Because the method involves higher sample cumulants  $\kappa_6$  it is recommended to use a sample size of at least 10; the simulation study shows that critical point  $\frac{z_\alpha + t_{n-1,\alpha}}{2}$  is preferable when sample size 10 is used.

An Example

Suppose a chemical manufacturer wants to monitor the viscosity of a particular chemical from the production line and that it is important to detect disturbances which could result in increasing the variability of the process. The random measurements of the viscosity are selected until subgroups are obtained, and corresponding sample variances  $S_{(i)}^2$ ,  $\kappa_{4,(i)}$  and Z6 are calculated and presented in Table 7.

Necessary process parameters are estimated from the preliminary run stage which contains 30 samples sized 10 each. The estimated process variance  $\tilde{S}^2 = 7.398$ , process skewness  $k_3^* = 1.74$  (positively skewed distribution), and process cumulants are  $\kappa_3 = 33.654$ ,  $\kappa_4 = 232.667$  and  $\kappa_6 = 9598.75$ . Because the proposed method recommends using a combined sample, all quantities are obtained from one large sample,  $n = 300$ , by merging the 30 size 10 samples together. Equation (4) is then used to obtain the upper limit of the control chart with critical point  $z_\alpha$

$$UCL = z_\alpha + n^{\frac{1}{2}} \left( \hat{B}_1 + \hat{B}_2 \frac{z_\alpha^2 - 1}{6} \right) = 6.049.$$

The new method is used to construct the control chart for the variability of this positively skewed distribution. Each sample point is the test statistic Z6 of the sample where

$$Z6 = \frac{S_{(i)}^2 - \tilde{S}^2}{\sqrt{\frac{\kappa_{4,(i)} \tilde{S}^2}{n S_{(i)}^2} + \frac{2 \tilde{S}^4}{n-1}}};$$

and  $S_{(i)}^2$  and  $\kappa_{4,(i)}$  are variance and the fourth cumulant of the  $i^{th}$  sample;  $\tilde{S}^2$  is estimated process variance calculated from the preliminary stage process. (See Figure 1). It can be observed that the process is under statistical control during the period of time when the 40 samples were collected, that is, all points are under the Upper Control Limit of 6.049.

If the traditional one-sided R-chart, S-chart,  $S^2$ -chart as well as WV and SC charts are also constructed, then one can observe that - in all charts except the SC-chart - at least one sample point, point 18, is above the Upper Control Limit, which gives a false out-of-control signal.

Conclusion

This study proposed a new charting scheme for the variability of a process. This technique is an adaptation of Long and Sa's (2005) testing procedure and is designed to control the variability of a process without any assumption regarding the form of the underlying distribution.

The Monte Carlo simulation study of type I error rates indicates that the proposed method is robust for all distributions studied. It can achieve significant improvement over the Shewhart R-chart, the S-chart and the  $S^2$ -chart, as well as the WV R-chart and the SC R-chart when the distribution is highly skewed and/or has large kurtosis. It can maintain the type I error rates close to the nominal level  $\alpha = 0.0027$  and shows reasonably good power.

In a real life situation, control charts are constructed even when there is no information about the form of the distribution of the quality characteristic. The method presented herein works well for all distributions studied, which includes the normal distribution.

If sample size is small, then the average of  $z_\alpha$  and  $t_{\alpha,n-1}$  as the critical point is recommended to produce a small number of false alarms and detect shifts reasonably well. Because the proposed method involves higher moments, a sample size of at least 10 is recommended.

## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

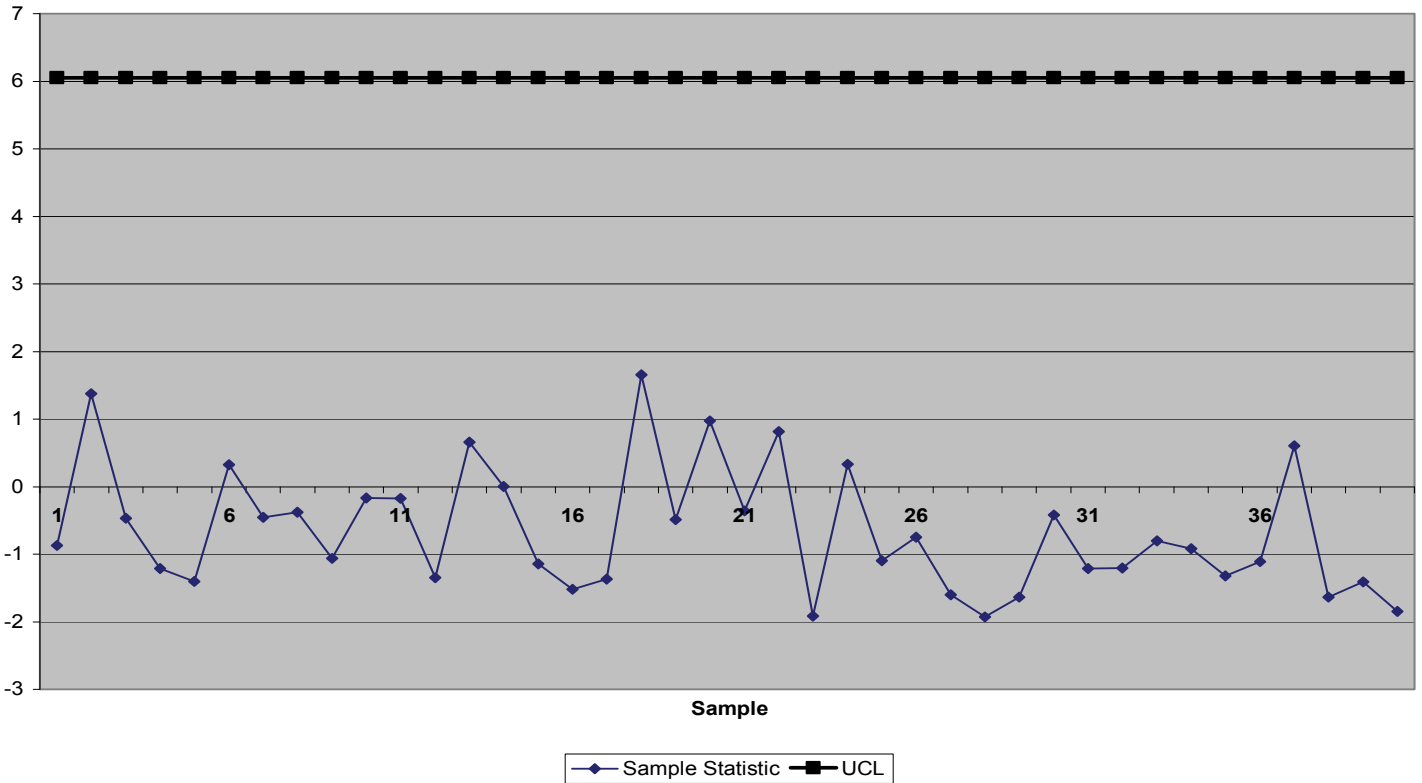
Table 7: Example Data ( $n = 10$ )

Sample	Data										$S^2$	$\kappa_4$	Z6
1	0.1753	0.1993	2.8990	2.6626	2.4803	3.9687	6.6066	1.7451	5.7174	2.3713	4.3754	0.0000	-0.8679
2	1.8987	9.5414	0.4717	3.7043	2.3586	9.9265	5.6399	0.8266	2.2084	1.0139	12.1981	0.0000	1.3730
3	6.7533	0.2629	4.2325	4.3115	3.1053	1.9042	4.7606	1.2661	8.1081	4.5336	5.7830	0.0000	-0.4647
4	1.3454	2.3209	2.8454	0.7297	2.0158	1.9803	1.0899	5.5295	0.2362	1.0541	2.2310	18.2422	-1.2115
5	1.2494	3.7607	0.7006	5.5135	1.6933	1.4810	3.3468	1.9125	2.0569	1.8984	2.0655	6.2677	-1.4055
6	4.9890	0.9675	11.7036	2.4818	4.8241	1.2914	1.6978	3.6436	4.6315	4.3729	9.5299	392.0086	0.3254
7	6.1313	6.3210	1.1931	1.6164	1.2826	6.4058	2.6711	0.4085	5.4223	4.6664	5.8097	0.0000	-0.4570
8	0.2906	1.7878	0.7905	1.1382	7.7531	0.2521	0.3012	2.1627	2.2058	0.6501	5.0435	179.6856	-0.3803
9	6.2531	6.0388	1.6128	2.8829	3.6935	0.7531	2.0489	4.8375	1.7980	2.1992	3.6892	0.0000	-1.0645
10	3.6747	2.2271	2.6493	4.3548	9.9896	2.9346	0.6846	2.2333	4.4169	2.1832	6.4037	217.7216	-0.1638
11	0.0888	1.1911	5.3541	2.9182	1.2550	8.7974	2.1795	2.0307	1.4277	2.0003	6.4750	138.9135	-0.1755
12	3.4825	2.3265	5.2447	0.7256	5.1806	1.6558	3.5688	2.8752	2.0482	0.5028	2.7078	0.0000	-1.3456
13	5.4607	0.7457	3.3457	9.8583	3.3429	2.9802	2.7684	0.1488	8.5222	5.5548	9.7128	0.0000	0.6611
14	0.6061	1.9028	4.7777	1.7799	0.6848	9.3894	1.4367	3.1679	5.5664	4.1000	7.4303	87.0290	0.0055
15	0.6069	0.2926	1.5002	0.9540	2.3739	3.5945	1.8038	1.7982	5.6449	0.2206	2.8074	15.3712	-1.1409
16	4.9356	2.4603	3.6291	4.4440	1.2951	2.3350	1.7887	1.5324	0.7752	1.0778	2.0949	0.0000	-1.5212
17	1.6166	0.6310	4.9091	3.8633	0.5929	2.9032	0.6788	0.8536	3.9993	1.5527	2.6392	0.0000	-1.3653
18	1.8073	4.3716	0.7548	4.0216	1.9322	17.5866	7.5443	1.2864	1.6462	1.7298	26.0279	4037.2395	1.6522
19	0.7649	3.4767	0.3363	3.7036	8.2392	3.9891	0.8094	4.0438	3.6515	2.9723	5.2400	51.3835	-0.4910
20	0.2054	14.2214	5.3427	2.4460	4.6481	5.7322	2.2885	4.3794	4.2059	0.8869	15.2693	1103.0417	0.9704
21	7.6503	1.8991	1.9254	0.8912	1.6838	2.3336	0.6366	6.2804	2.7257	0.2412	5.9172	38.7200	-0.3605
22	9.8933	7.5309	3.2807	1.5821	0.4120	3.3600	0.6842	0.2074	0.1167	1.4909	11.1675	135.4547	0.8177
23	0.9872	2.0582	0.3003	1.1105	1.9264	1.9017	0.1944	0.3366	1.5401	2.6845	0.7320	0.0000	-1.9116
24	2.8916	11.2486	0.7761	0.6566	1.2512	3.0792	4.1080	1.8494	0.6488	2.3294	9.9941	652.4974	0.3328
25	0.5455	2.9088	2.0852	2.5598	6.5651	0.8456	3.5727	2.9086	5.2798	2.2881	3.3818	5.6749	-1.0979
26	1.9177	2.7938	1.3091	7.1956	4.6303	1.6167	1.2484	0.5885	1.9009	1.5335	3.9685	48.8875	-0.7445
27	0.3927	2.2825	1.5281	1.3541	1.9966	0.3233	1.1659	4.0624	0.6790	0.5122	1.3077	4.0065	-1.6038
28	2.1452	2.2046	2.3803	3.4120	1.8255	3.1891	3.2583	1.5925	2.4079	0.7436	0.6824	0.0000	-1.9258
29	0.3452	1.1206	1.5241	2.0653	1.5242	1.7115	0.8049	3.7092	4.2871	1.6163	1.5126	1.5460	-1.6379
30	0.4881	7.1682	4.6524	0.4335	2.2388	0.8668	0.8453	4.9536	0.7459	0.3920	5.9185	1.4050	-0.4228

Table 7 (continued): Example Data ( $n = 10$ )

Sample	Data										$S^2$	$\kappa_4$	Z6
31	1.1352	3.7820	6.0729	2.0385	0.6038	4.9635	3.7010	2.3934	2.7716	1.0346	3.1780	0.0000	-1.2109
32	5.3424	1.4407	2.3453	1.5825	5.1191	3.9414	5.9906	2.5694	1.4657	1.6535	3.2021	0.0000	-1.2040
33	5.5480	2.2404	2.3234	3.5577	0.1689	6.1042	1.2915	3.8142	6.3927	1.6449	4.6073	0.0000	-0.8015
34	3.0710	0.6454	1.1600	2.5251	0.7999	3.8018	2.1754	0.6235	6.6516	2.0064	3.4462	29.9413	-0.9175
35	5.7266	2.5197	0.5746	4.0094	1.8757	3.1076	1.5696	2.2175	1.1754	1.2167	2.3779	7.1928	-1.3236
36	0.4678	3.0494	5.3371	2.9850	0.5926	0.8862	0.7395	4.7288	0.4933	0.7294	3.5218	0.0000	-1.1124
37	1.6734	12.9557	0.7857	0.8905	1.7642	0.7138	0.9902	2.1479	2.7973	2.4034	13.4796	1625.8027	0.6029
38	1.8289	2.9504	3.4233	0.5749	1.2637	4.2572	0.2638	1.0693	2.8841	1.8386	1.7003	0.0000	-1.6343
39	1.9636	4.6574	0.7779	1.2897	1.1516	2.1388	1.4726	4.9557	2.1783	4.4767	2.4909	0.0000	-1.4078
40	3.4110	2.8170	1.3172	3.2577	3.4979	1.7006	1.5000	1.0208	1.3976	1.4541	0.9690	0.0000	-1.8437

Figure 1 Example of the Proposed One-Sided Control Chart Method



## A ROBUST ONE-SIDED VARIABILITY CONTROL CHART

### References

- Bai, D. S., & Choi, I. S. (1995).  $\bar{X}$  and R control charts for skewed populations. *Journal of Quality Technology*, 27(2), 120-131.
- Borysov, P., & Sa, P. (2010). The simulation results for the robust one-sided variability control chart. *Technical Report, Department of Mathematics and Statistics, University of North Florida*, <http://www.unf.edu/coas/math-stat/CRCs/Reports.htm>.
- Burr, I. W. (1967). The effect of non-normality on constants for  $\bar{X}$  and R charts. *Industrial Quality Control*, 23, 563-569.
- Chan, L. K., Hapuarachchi, K. P., & Macpherson, B. D. (1988). Robustness of  $\bar{X}$  and R charts. *IEEE Transactions on Reliability*, 37(1), 117-123.
- Chan, L. K., & Cui, H. J. (2003). Skewness correction  $\bar{X}$  and R charts for skewed distributions. *Naval Research Logistics*, 50, 555-573.
- Fleishman, A. I. (1978). A method of simulating non-normal distributions, *Psychometrika*, 43(4), 521-531.
- Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York, NY: Springer-Verlag, Inc.
- Johnson, M. E., Tietjen, G. L., & Beckman, R. J. (1980). A new family of probability distributions with applications to Monte Carlo studies. *Journal of the American Statistical Association*, 75(370), 276-279.
- Kendall, S. M. (1994). *Distribution theory*. New York, NY: Oxford University Press Inc.
- Long, M. C., & Sa, P. (2005). Right-tailed testing of variance for non-normal distributions. *Journal of Modern Applied Statistical Methods*, 4(1), 187-213.
- Montgomery, D. C. (1996). *Introduction to statistical quality control (3<sup>rd</sup> Ed.)*. New York, NY: John Wiley & Sons, Inc.
- Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Princeton, New Jersey: Van Nostrand Reinhold Company.

### Instructions For Authors

Follow these guidelines when submitting a manuscript:

1. *JMASM* uses a modified American Psychological Association style guideline.
2. Submissions are accepted via e-mail only. Send them to the Editorial Assistant at [ea@jmasm.com](mailto:ea@jmasm.com). Provide name, affiliation, address, e-mail address, and 30 word biographical statements for all authors in the body of the email message.
3. There should be no material identifying authorship except on the title page. A statement should be included in the body of the e-mail that, where applicable, indicating proper human subjects protocols were followed, including informed consent. A statement should be included in the body of the e-mail indicating the manuscript is not under consideration at another journal.
4. Provide the manuscript as an external e-mail attachment in MS Word for the PC format only. (Wordperfect and .rtf formats may be acceptable - please inquire.) Please note that Tex (in its various versions), Exp, and Adobe .pdf formats are designed to produce the final presentation of text. They are not amenable to the editing process, and are **NOT** acceptable for manuscript submission.
5. The text maximum is 20 pages double spaced, not including tables, figures, graphs, and references. Use 11 point Times New Roman font.
6. Create tables without boxes or vertical lines. Place tables, figures, and graphs “in-line”, not at the end of the manuscript. Figures may be in .jpg, .tif, .png, and other formats readable by Adobe Photoshop.
7. The manuscript should contain an Abstract with a 50 word maximum, following by a list of key words or phrases. Major headings are Introduction, Methodology, Results, Conclusion, and References. Center headings. Subheadings are left justified; capitalize only the first letter of each word. Sub-subheadings are left-justified, indent optional. Do not number headings or subheadings.
8. Number all formulas, tables, figures, and graphs, but do not use italics, bold, or underline.
9. Do not use underlining in the manuscript. Do not use bold, except for (a) matrices, or (b) emphasis within a table, figure, or graph. Do not number references. Do not use footnotes or endnotes.
10. In the References section, do not put quotation marks around titles of articles or books. Capitalize only the first letter of books. Italicize journal or book titles, and volume numbers. Use “&” instead of “and” in multiple author listings.
11. *Suggestions for style*: Instead of “I drew a sample of 40” write “A sample of 40 was selected”. Use “although” instead of “while”, unless the meaning is “at the same time”. Use “because” instead of “since”, unless the meaning is “after”. Instead of “Smith (1990) notes” write “Smith (1990) noted”. Do not strike spacebar twice after a period.

### Print Subscriptions

Print subscriptions including postage for professionals are US \$95 per year; for graduate students are US \$47.50 per year; and for libraries, universities and corporations are US \$195 per year. Subscribers outside of the US and Canada pay a US \$10 surcharge for additional postage. Online access is currently free at <http://www.jmasm.com/>. Mail subscription requests with remittances to JMASM, P. O. Box 48023, Oak Park, MI, 48237. Email journal correspondence, other than manuscript submissions, to [ea@jmasm.com](mailto:ea@jmasm.com).

### Notice To Advertisers

Send requests for advertising information to [ea@jmasm.com](mailto:ea@jmasm.com).