

University of Windsor

Scholarship at UWindor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2007

Shrinkage estimation of variance components with applications to microarray data

Lihua An

University of Windsor

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>

Recommended Citation

An, Lihua, "Shrinkage estimation of variance components with applications to microarray data" (2007). *Electronic Theses and Dissertations*. 4688.

<https://scholar.uwindsor.ca/etd/4688>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

**Shrinkage Estimation of Variance Components
with Applications to Microarray Data**

by

Lihua An

**A Dissertation
Submitted to the Faculty of Graduate Studies
through Mathematics and Statistics
in Partial Fulfillment of the Requirements of
the Degree of Doctor of Philosophy at the
University of Windsor**

Windsor, Ontario, Canada

2007



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-35080-5
Our file *Notre référence*
ISBN: 978-0-494-35080-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

All rights reserved, 2007

© Lihua An

Abstract

The difficulty is to know what problems to combine together, ... Why should not all our estimation problems be lumped together on one grand melee.

George Barnard (1962)

Microarray technology has become an important tool for simultaneously obtaining quantitative measurements for the expression of thousands of genes presented in a biological sample. A major concern with microarray experiments is that they have little replication. The variance estimates obtained from an individual gene may be very imprecise, and therefore the inferences reached may not be trustworthy due to the weak variance estimate. An appealing idea for improving inferences from microarray experiments would be combining information across genes.

In this dissertation, the estimation problem of variance components in various contexts is investigated. In particular, we consider the following problems:

- Multi-sample analysis of population variances;
- Simultaneous estimation of variances when some uncertain prior information (UPI) about the parameter is available;
- The estimation of variance components in meta-analysis with random effects.

In chapter two, we consider the large-sample inference of population variance and investigate its statistical properties in a multi-sample set up when random samples are

drawn from arbitrary populations. Asymptotic statistical procedures are developed for testing the homogeneity of these variances and an interval estimation based on combined data from various sources is proposed. An extensive sampling experiment is conducted to investigate the performance of the suggested methods over a wide range of models including skewed distributions. The problem of kurtosis estimation arising from inferences about variance parameter is also discussed extensively; improved estimators of population kurtosis are proposed.

In chapters three and four, we develop some alternative estimation strategies based on pretest (Bancroft (1944)) and James-Stein (James and Stein (1961)) principles when the information regarding the homogeneity of all the variances may not be precise. Asymptotic properties of the pre-test and shrinkage estimators are discussed and compared to the maximum likelihood estimator and the pooled estimator. It is demonstrated that the positive part James-Stein estimator utilizes the sample and non-sample information in a superior way relative to the ordinary shrinkage estimator. Two shrinkage-type optimal weight combination estimators are derived from finite samples under the quadratic and the entropy loss functions. Simulation study shows that our estimators are superior to other combination estimators.

In chapter five, we deal with the estimation of the variance components in meta-analysis with random effects. James-Stein type estimators are proposed and their risk is simulated and compared to some existing estimators. The simulation study shows that our shrinkage estimators are minimax and admissible with respect to the base estimator.

Acknowledgements

First of all, I wish to express my most sincere appreciation to my supervisor, Dr. Ejaz Ahmed, for his invaluable guidance, encouragement and patience throughout the whole process of this dissertation. His support and knowledge sparked my interest in both theoretical and applied statistics.

Special thanks go to Dr. Nancy Reid, my dissertation external examiner, for taking her precious time to read through my thesis and for providing constructive comments.

I am very grateful to all of the members of my dissertation advisory committee, Dr. Fazle Baki, Dr. Karen Fung, and Dr. Abdul Hussein. I appreciate their time and their constructive feedback, which helped further improve the quality of my thesis.

I would also like to thank my friendly and knowledgeable group members, Shakhawat Hossain and Razi Ghor, for all of the nice discussions on our research.

Another sincere gratitude is to all of the faculty, staff members and the fellow graduate students in the mathematics and statistics department. Everybody has been so friendly and helpful. Thanks to them, I have had a very enjoyable and memorable experience during my four years of study here.

Last, but never least, I wish to express my deepest gratitude to my family. I owe more than I can express to my parents, my husband, and my two sons for their unconditional love, constant encouragement, understanding, and patience. Nothing would have been possible without their support and faith.

This research is supported in part by NSERC, OGS and OGSST scholarships.

Abbreviations

<i>ADB</i>	asymptotic distributional bias
<i>ADF</i>	asymptotic distribution function
<i>ADQR</i>	asymptotic distributional quadratic risk
$ch_{min}(\mathbf{A})$	smallest eigenvalue of \mathbf{A}
$ch_{max}(\mathbf{A})$	largest eigenvalue of \mathbf{A}
<i>EBE</i>	empirical Bayes estimator
<i>II</i>	imprecise information
<i>IPTE</i>	improved preliminary test estimator
<i>JSE</i>	James-Stein type estimator
<i>LSE</i>	linear shrinkage estimator
<i>LCE</i>	linear combination estimator
<i>MLE</i>	maximum likelihood estimator
<i>MME</i>	method of moment estimator
<i>MSE</i>	mean squared error
<i>NSI</i>	non-sample information
<i>OWCE</i>	optimal weight combination estimator
<i>PJSE</i>	positive part James-Stein type estimator
<i>PTE</i>	preliminary test estimator
<i>QB</i>	quadratic bias
<i>RE</i>	restricted estimator

<i>RE</i>	relative efficiency
<i>REMLE</i>	restricted maximum likelihood estimator
<i>SB</i>	simulated bias
<i>SE</i>	standard error
<i>SHE</i>	simple heterogeneity variance estimator
<i>SMSE</i>	simulated mean squared error
<i>UE</i>	unrestricted estimator
<i>UPI</i>	uncertain prior information
<i>VCE</i>	variance component type estimator

List of Symbols

\mathbf{D}	$= \text{Diag}(\omega_{1,n}, \dots, \omega_{k,n})$
H_0	null hypothesis
\mathbf{I}	identity matrix
k	the number of independent samples
$K_{(n)}$	local alternative hypothesis
$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$	a loss function when estimating $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$
n	sample size in a balanced design or $\sum n_i$
n_i	sample size for the i th independent sample
Q	test statistic for homogeneity of variance in meta analysis
\mathbf{Q}	a positive semi-definite weight matrix in the quadratic loss function
$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$	a risk function when estimating $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$
T_l	test statistics, where $l = 1, 2, 3, 4$
v_i	the i th within group (study) variance
W	Levene test statistic
α	level of significance
β_0, β_1	coefficients in $\hat{\gamma}^{(N3)}$
$\boldsymbol{\delta}$	a fixed real valued vector contained in $K_{(n)}$
δ_i	random effect of the i th group(study) in the meta analysis

γ^*	population kurtosis
γ	$= \gamma^* - 3$, standardized population kurtosis
γ_i	the i th population kurtosis
$\hat{\gamma}_{(R)}^*$	a pooled estimator of the common kurtosis parameter
$\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M$	kurtosis estimators currently adapted by different software packages
$\hat{\gamma}^{N1}, \hat{\gamma}^{N2}, \hat{\gamma}^{N3}$	kurtosis estimators proposed in this thesis
Γ	covariance matrix of $\hat{\sigma}^2$
Γ_1^{-1}	$= \Omega_1$
$\hat{\Gamma}_2^{-1}$	$= \frac{\Omega_2}{(\hat{\gamma}_{(R)}^* - 1)(\hat{\sigma}_{(R)}^2)^2}$
$\hat{\Gamma}_3^{-1}$	$= \hat{\Omega}_3$
$\hat{\Gamma}_4^{-1}$	$= \frac{\Omega_2}{(\hat{\gamma}_{(R)}^* - 1)}$
Λ	test statistic
$\lambda_{n,\alpha}$	critical value of Λ
$\omega_{i,n}$	$= n_i / \sum n_i$ weight of the i th group
Ω	dispersion matrix of $\hat{\sigma}^2$
Ω_1	$= \text{Diag} \left(\frac{\omega_{1,n}}{(\gamma_1^* - 1)\sigma_{1o}^2}, \dots, \frac{\omega_{k,n}}{(\gamma_k^* - 1)\sigma_{ko}^2} \right)$
Ω_2	$= \text{Diag} (\omega_{1,n}, \dots, \omega_{k,n})$
$\hat{\Omega}_3$	$= \text{Diag} \left(\frac{\omega_{1,n}}{(\hat{\gamma}_1^* - 1)}, \dots, \frac{\omega_{k,n}}{(\hat{\gamma}_k^* - 1)} \right)$
$\Psi_k(x; \Theta)$	noncentral χ^2 distribution function with non-centrality parameter Θ and degrees of freedom k

σ^2	population variance
σ_i^2	the i th population variance
$\hat{\sigma}_{(R)}^2$	a pooled estimator of the common variance
σ^2	$= (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)'$
$\hat{\sigma}^2$	$= (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_k^2)'$
$\hat{\sigma}_{(JS)}^2$	the James-Stein estimator of σ^2
$\hat{\sigma}_{(PP)}^2$	the positive part James-Stein estimator of σ^2
$\hat{\sigma}_{(PT)}^2$	the preliminary test estimator of σ^2
$\hat{\sigma}_{(R)}^2$	the restricted estimator of σ^2
$\hat{\sigma}_{(SP)}^2$	the improved preliminary test estimator of σ^2
$\hat{\sigma}_{(SR)}^2$	the linear shrinkage estimator of σ^2
τ^2	the among group(study) variance in the meta analysis
θ	any parameter vector to be estimated
θ_0	a guessed value of θ based on a prior non-sample information
$\hat{\theta}_{(IPT)}$	the improved preliminary test estimator of θ
$\hat{\theta}_{(JS)}$	the James-Stein estimator of θ
$\hat{\theta}_{(PP)}$	the positive part James-Stein estimator of θ
$\hat{\theta}_{(PT)}$	the preliminary test estimator of θ
$\hat{\theta}_{(R)}$	the restricted estimator of θ under a null hypothesis
$\hat{\theta}_{(SR)}$	the linear shrinkage estimator of θ
$\mathbf{1}_k$	a vector of 1's

Contents

Abstract	i
Acknowledgements	iii
Abbreviations	iv
List of Symbols	vi
1 Introduction and Literature Review	1
1.1 Estimation of Variance in Microarray Analysis	1
1.2 The History of Shrinkage Estimation	3
1.3 Meta-analysis in Microarray Studies	10
1.4 Outline of the Thesis	13
2 Multi-Sample Analysis of Population Variances and Kurtosis Esti- mation	16
2.1 Introduction	16
2.2 Preliminaries and Large Sample Results	17
2.3 Hypothesis Testing	18

<i>CONTENTS</i>	x
2.4 Variance Stabilizing Transformation	21
2.5 A Power Study	23
2.6 Interval Estimation	24
2.7 Estimation of Kurtosis	26
2.7.1 Preliminaries and Background	27
2.7.2 Proposed Estimation Strategies	29
2.7.3 A Simulation Study of Kurtosis	31
2.7.4 An Application in the Estimation of Coefficient of Variation	42
2.8 Simulation Studies for Variances	47
2.8.1 Simulation Study 1: Test Statistics	48
2.8.2 Simulation Study 2: Interval Estimation	56
2.9 Application to the Iris Data	64
2.10 Concluding Remarks and Further Research	67
3 Simultaneous Estimation of Variance under Imprecise Information	69
3.1 Introduction	69
3.2 Estimation Strategies Under Imprecise Information	72
3.2.1 Restricted Estimator	73
3.2.2 Linear Combination Estimator	78
3.2.3 Preliminary Test Estimator	79
3.2.4 Improved Preliminary Test Estimator	81
3.2.5 James-Stein Type Estimator	85
3.2.6 Positive Rule James-Stein Estimator	86
3.3 Numerical ADQR Analysis	90

<i>CONTENTS</i>	xi
3.4 Simulation Results	96
3.4.1 Under Quadratic Loss Function	97
3.4.2 Under Entropy Loss Function	103
3.5 Concluding Notes	108
4 Small-Sample Risk Analysis of Variance Estimation with Applica- tion in Differential Gene Expression	111
4.1 Introduction	111
4.2 Two Optimal Weight Combination Estimators (OWCE)	114
4.2.1 The OWC Estimator Based on the Original Data	114
4.2.2 An OWC Estimator Based on Log-transformed Data	118
4.3 Risk Study of the Estimators	123
4.3.1 Under the Quadratic Loss Function	123
4.3.2 Under Entropy Loss Function	124
4.4 Simulation Results	126
4.5 Concluding Notes	133
5 Estimation of the Variance Components for Meta-Analysis with Ran- dom Effects	134
5.1 Introduction	134
5.2 Risk Study of the Existing Estimators	136
5.2.1 Simulation Study of the Risks	138
5.3 James-Stein Type Estimators	140
5.3.1 Risk Simulation	140

<i>CONTENTS</i>	xii
5.4 Conclusion and Discussion	141
6 Conclusions and Future Research	147
Bibliography	152
Vita Auctoris	164

List of Tables

2.1	The relative efficiency of the kurtosis estimators	31
2.2	The coefficients in $\hat{\gamma}^{N3}$	42
2.3	Performance of $\hat{\tau}$ based on different $\hat{\gamma}$	45
2.4	Power simulation of the tests based on T_2 and T_4 for normal samples ($\alpha = 0.05$)	53
2.5	Power simulation of the tests based on T_2 and T_4 for Student-t samples ($\alpha = 0.05$)	54
2.6	Power simulation of the tests based on T_2 and T_4 for χ^2 samples ($\alpha = 0.05$)	54
2.7	Power simulation of the tests based on T_2 and T_4 for samples from a mixture of two normal populations ($\alpha = 0.05$)	55
2.8	Coverage probability and mean width for uncombined normal sample data	57
2.9	Coverage probability and mean width for combined normal samples data	58
2.10	Coverage probability and mean width based on uncombined sample data from t_6 distribution	58

2.11 Coverage probability and mean width based on combined samples data from t_6 distribution	59
2.12 Coverage probability and mean width based on single sample data from $\chi^2_{(4)}$ distribution	60
2.13 Coverage probability and mean width based on combined samples data from $\chi^2_{(4)}$ distribution	61
2.14 Coverage probability and mean width based on uncombined sample data from a mixture of normal distribution	62
2.15 Coverage probability and mean width based on combined data from a mixture of normal distribution	62
2.16 Four measurements on three species of Iris (in centimeters)	64
3.1 SRE of the estimators under quadratic loss with $k = 4$, $\alpha = 0.05$, $n_i =$ 50 when population is normal.	100
3.2 SRE of the estimators under quadratic loss with $k = 4$, $\alpha = 0.05$, $n_i =$ 20 when population is normal.	100
3.3 SRE of the estimators under quadratic loss with $k = 10$, $\alpha = 0.05$, $n_i =$ 50 when population is normal.	101
3.4 SRE of the estimators under quadratic loss with $k = 10$, $\alpha = 0.05$, $n_i =$ 20 when population is normal.	101
3.5 SRE of the estimators under quadratic loss with $k = 10$, $\alpha = 0.20$, $n_i =$ 50 when population is normal.	101
3.6 SRE of the estimators under entropy loss with $k = 4$, $\alpha = 0.05$, $n_i = 20$ when population is normal.	106

3.7	SRE of the estimators under entropy loss with $k = 4$, $\alpha = 0.05$, $n_i = 50$ for normal data.	106
3.8	SRE of the estimators under entropy loss with $k = 10$, $\alpha = 0.05$, $n_i =$ 50 for normal data.	107
3.9	SRE of the estimators under entropy loss with $k = 10$, $\alpha = 0.05$. $n_i =$ 20 for normal data.	107
3.10	SRE of the estimators under entropy loss with $k = 10$, $\alpha = 0.20$. $n_i =$ 50 when population is normal.	107
3.11	SRE under entropy loss for mixture of normal data with $k = 4$. $\alpha =$ 0.05, $n_i = 20$	108
3.12	SRE under entropy loss for mixture of normal data with $k = 4$. $\alpha =$ 0.05, $n_i = 50$	108
3.13	SRE under entropy loss for mixture of normal samples with $k = 10$. $\alpha = 0.05$, $n_i = 50$	109
3.14	SRE under entropy loss for mixture of normal samples with $k = 10$. $\alpha = 0.05$, $n_i = 20$	109
4.1	The values of m and V for selected degrees of freedom	121
5.1	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 10$	141
5.2	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 20$	142

5.3	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 30$	142
5.4	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 50$	142
5.5	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 10$	143
5.6	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 20$	143
5.7	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 30$	143
5.8	Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 50$	144

List of Figures

2.1	Bias and <i>MSE</i> of the kurtosis estimators for normal distributions. $E_1, E_2, E_3, E_4,$ and E_5 denote the estimators $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2},$ respectively.	30
2.2	Bias comparison of the kurtosis estimators for a mixture of two normal populations. $E_1, E_2, E_3, E_4, E_5,$ and <i>theory</i> represent the expected values of the estimators $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1}, \hat{\gamma}^{N2},$ and the theoretical value $\gamma,$ respectively. The bias is measured by the distance between the mean of the estimator and the theoretical value.	34
2.3	<i>MSE</i> of the kurtosis estimators for mixtures of two normal populations. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2},$ respectively.	35
2.4	Bias of the kurtosis estimators for Student-t distributions. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2},$ respectively.	36
2.5	<i>MSE</i> of the kurtosis estimators for Student-t distributions. ν is the degrees of freedom which ranges from 5 to 30. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2},$ respectively.	36

2.6 Bias of the kurtosis estimators for χ^2 distributions. The degrees of freedom ν ranges from 1 to 40. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively. 38

2.7 *MSE* of the kurtosis estimators for χ^2 distributions. The bottom two plots use enlarged vertical scale in order to have a better view when the degrees of freedom ν is large. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively. 39

2.8 $\hat{\gamma}^{N2}$ vs γ scatter plot 41

2.9 Distribution of the test statistic T_2 (original data). The test statistics with “BC” were calculated based on bias corrected estimate of kurtosis in table 2.2. and the rest were based on $\hat{\gamma}^U$ 49

2.10 Distribution of the test statistic T_4 (transformed data). The test statistics with “BC” were calculated based on bias corrected estimate of kurtosis in table 2.2, and the rest were based on $\hat{\gamma}^U$ 50

3.1 ADQR/k of the estimators at different α level when $k=4$ 92

3.2 ADQR/k of the estimators at different α level when $k = 10$ 93

3.3 ARE of the estimators for $k = 4$ 94

3.4 ARE of the estimators for $k = 10$ 95

3.5 *SRE* performance comparison for $k = 4, \alpha = .05$ and various n when population is normal. $SRE_1, SRE_2, SRE_3,$ and SRE_4 represent the simulated relative efficiency of $\hat{\sigma}_{(SR)}^2, \hat{\sigma}_{(SP)}^2, \hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$, respectively. 98

3.6 *SRE* comparison for $k = 10$, $\alpha = .05$ and various n when population is normal. SRE_1 , SRE_2 , SRE_3 , and SRE_4 represent the simulated relative efficiency of $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(SP)}^2$, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$, respectively. 98

3.7 *SRE* comparison for mixture of normal distributions 102

3.8 *SRE* comparison for mixture of normal distributions when α differs 103

3.9 *SRE* comparison under entropy loss for normal population 105

3.10 *SRE* under entropy loss for normal population under different α 105

4.1 Relative efficiency of the proposed estimators under quadratic loss for normal data when $k = 4$ and 10. 127

4.2 Relative efficiency of the proposed estimators under quadratic loss for normal data when $k = 100$. The plots on the right are a closer look of the plots on the left. 128

4.3 Relative efficiency of the proposed estimators under entropy loss (normal data) for $k = 4$ and 10, and $n = 4$ and 10. 129

4.4 RE comparison under entropy and quadratic losses (normal data). The top two plots are under entropy loss, and the bottom two plots are under quadratic loss. 130

4.5 Relative efficiency of the estimators under quadratic loss for a mixture of normal data. 131

5.1 RE comparison under quadratic loss 144

5.2 RE comparison under entropy loss 145

Chapter 1

Introduction and Literature

Review

1.1 Estimation of Variance in Microarray Analysis

The development of microarray technology has revolutionized the study of molecular biology and become a standard tool in genomics research. Instead of working on a gene-by-gene basis, microarray technology allows scientists to simultaneously view the expression of thousands of genes from an experimental sample. Due to the cost, it is common that thousands of genes are measured with a small number of replications (Lonnstedt and Speed (2002), Kendziorski *et al.* (2003)). As a consequence, we are faced with a “large k , small n ” paradigm, where k is the total number of genes and n is the number of replications for each gene.

The variance estimates obtained from individual genes are unreliable due to the relatively small number of replications. Consequently, the commonly used statistical methods such as t-test or F-test for detecting differentially expressed genes based on the gene specific variance estimates have low power (Cui and Churchill (2003) and Baldi and Long (2001)).

However, measures of thousands of genes are readily available in a microarray experiment. Borrowing strength and combining data across genes seems to be an appealing idea. Many approaches to improving the variance estimation initiated from this thought have been developed in the past a few years.

Assuming the dependence between the average intensity and the variance of the intensity differences, Kamb and Ramaswami (2001) suggested a simple difference-averaging method by averaging a neighborhood of genes whose expression levels are close to one another. Their method enables determination of variances as a function of signal intensities by using information over the entire data set. However, in their simple average method, all the genes within a window are treated equally. Huang and Pan (2002) improved upon this idea and proposed a weighted average method, where the weights depend on the Euclidean distance between the observations. Jain *et al.* (2003) and Comander *et al.* (2004) proposed local-pooled estimation procedures that pooled the variances of genes with similar intensities. Their methods effectively identify significant differential expression patterns with a small number of replicated arrays. Baldi and Long (2001) initiated a regularized t-test replacing the usual variance estimate with a hierarchical Bayes estimator. Lonnstedt and Speed (2002) brought forth an empirical Bayes approach that combines information across

genes. Kendzioriski *et al.* (2003) extended the empirical Bayes method using hierarchical gamma-gamma and lognormal-normal models. The SAM t-test developed by Tusher *et al.* (2001) added a small constant to the gene-specific variance estimate in order to stabilize the small variance estimates for the purpose of eliminating some false positives associated with low values of variance.

Cui *et al.* (2005) proposed a James-Stein type shrinkage estimator of variance and used it to construct an F-like test statistic. It is favorable compared to other F-like statistics based on the gene-based estimator or the simple pooled estimator across all k genes in terms of power, false positive rate, and robustness. This shrinkage type estimation of variance has drawn increasing attention in the application of microarray analysis (Fan and Ren (2006), Allison *et al.* (2006) and Leek *et al.* (2006)). However, so far, the research has concentrated on the applications, but little is known about the theoretical properties of various shrinkage variance estimators.

1.2 The History of Shrinkage Estimation

The history of shrinkage estimation spanned from the famous Stein's paradox discovered by Charles Stein in 1956. Professor Stephen Stigler stated in his 1988 Neyman Memorial Lecture, "*One of the most provocative results in mathematical statistics of the past 35 years is the phenomenon known variously as Stein's paradox, shrinkage estimation, or the James-Stein estimator,*" (Stigler (1990)). Stein (1956) and James and Stein (1961) discovered that in three or more dimensions, the ordinary maximum likelihood estimator (MLE) of the vector of means of a multivariate normal

distribution is inadmissible. In its simplest form, the situation is this: a collection of independent measurements X_1, X_2, \dots, X_k is available, each normally distributed with mean θ_i and variance 1. The θ_i 's are fixed unknown parameters which are not necessarily related to one another, and it is desired to estimate all the θ_i 's with a composite loss function

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^k (\theta_i - \hat{\theta}_i)^2, \quad (1.1)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)'$. The performance of the joint estimator $\hat{\boldsymbol{\theta}}$ is to be judged by the risk function,

$$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta}}[L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})]. \quad (1.2)$$

The startling discovery of Stein was that the obvious or "ordinary" estimator $\hat{\theta}_i^0 = X_i$ is inadmissible if $k \geq 3$; in fact, for $k \geq 3$ any estimator of the form

$$\hat{\theta}_i^{JS} = \left(1 - \frac{c}{S^2}\right) X_i \quad (1.3)$$

has a uniformly smaller risk for all θ_i , where $S^2 = \sum_{i=1}^k X_i^2$ and c is any constant with $0 < c < 2(k-2)$. (The best choice of c is $k-2$.)

A consequence of the above finding is the following counterintuitive result: when three or more unrelated parameters are estimated, the total risk can be reduced by using a combined estimator such as the James-Stein estimator; whereas when each parameter is estimated separately, the ordinary maximum likelihood estimator is admissible. This paradox has caused some to sarcastically ask whether, in order to estimate the speed of light, one should jointly estimate tea consumption in Taiwan and hog weight in Montana. The response is that the James-Stein estimator always improves upon the *total* risk, i.e., the sum of the expected losses of each component.

Therefore, the total risk in measuring light speed, tea consumption and hog weight would improve by using the James-Stein estimator. However, any particular component (such as the speed of light) would improve for some parameter values, and deteriorate for others.

Because $\hat{\theta}_i^{JS}$ may be considered as a weighted average of 0 and X_i , it has been described as “shrinking” the ordinary estimator $\hat{\theta}_i^0 = X_i$ toward 0 despite the fact that if $S^2 < c$, it “shrinks past” 0.

In a practical application of the James-Stein estimator, all of the savings will disappear if any of the $|\theta_i|$ are very large. In that case, S^2 will be large and $\hat{\theta}^{JS} = (1 - \frac{k-2}{S^2}) \mathbf{X}$ will be close to the MLE $\hat{\theta}^0 = \mathbf{X}$. $\hat{\theta}^{JS}$ shrinks \mathbf{X} toward the origin, reaping the savings depending on the origin being well chosen for the problem at hand. As a result, in the estimation of the speed of light, borrowing information from tea consumption and hog weight will probably not give us much of an advantage.

Of course we can choose any origin we want by subtracting an arbitrary constant from the data; or we can let the data choose the origin for us by shrinking toward a central value for the k observed values x_i . Efron and Morris (1973) used an Empirical Bayes approach and devised a variant of the original James-Stein estimator, which shrinks X_i toward \bar{X} . These estimators are of the form

$$\hat{\theta}_i^{EM} = \bar{X} + \left(1 - \frac{c}{S'^2}\right) (X_i - \bar{X}), \quad (1.4)$$

where $S'^2 = \sum_{i=1}^k (X_i - \bar{X})^2$, and c is any constant with $0 < c < 2(k-3)$. The optimal value of c is $k-3$. These estimators dominate the ordinary estimator as long as $k \geq 4$. The dimension of the parameter space is reduced from k to $k-1$, which is reflected in the use of the constant $k-3$ instead of $k-2$.

Since the discovery of the Stein's phenomenon, much research has been done and numerous papers focusing on the shrinkage estimation of location parameters have been published (Casella and Hwang (1986), Ahmed and Saleh (1993) and Ahmed and Saleh (1999)). On the contrary, little can be found on the shrinkage estimation of a scale parameter. Stein (1964) proved the inadmissibility of the usual variance estimator with unknown mean even though he stated, "I find it hard to take the problem of estimating σ^2 with quadratic loss function very seriously," in the same paper. However, Brown (1968) and then Brewster and Zidek (1974) took the problem seriously and improved upon Stein's result for point estimation. Kubokawa (1994) took a unified approach to improving the point and interval estimation of variance. Yet all of this research is concerned with a single variance, which is not applicable to microarray data analysis. Some research has been devoted to the shrinkage estimation of a covariance matrix (Sinha and Ghosh (1987), Kubokawa and Srivastava (2003)). However, all of these methods required $n > k$ to ensure non-singularity of the sample covariance matrix.

In this thesis, we propose new optimal shrinkage estimators of variances that borrow strength across genes. Interestingly, our Stein-type shrinkage estimators naturally arise from a pure shrinkage estimation via the preliminary test estimation principle.

In statistical literature, preliminary test estimation was introduced by Bancroft (1944) to estimate the parameters of a model when it is suspected that some "uncertain prior information" on the parameter of interest is available. The method involves a statistical test of the "uncertain prior information" based on an appropriate statistic and a decision on whether the model based sample estimate or the prior information

based estimate of the model parameters should be taken.

To this end, let us consider the estimation of a vector of any unknown parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$ when a priori, for example, $\boldsymbol{\theta} = \boldsymbol{\theta}^0 = (\theta_1^0, \dots, \theta_k^0)'$ is suspected. This natural origin $\boldsymbol{\theta}^0$ could be any sort of a priori information about $\boldsymbol{\theta}$. In many applied problems, usually the experimenter has some idea about the value of parameter $\boldsymbol{\theta}$ based on past experiences or acquaintance with the problem under consideration. It is reasonable, then, to move the “ordinary”, “classical” or “benchmark” estimator, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}^0$. Let us define a linear shrinkage estimator (*LSE*) as

$$\hat{\boldsymbol{\theta}}_{(SR)} = \hat{\boldsymbol{\theta}} - \pi(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0), \quad \pi \in (0, 1), \quad (1.5)$$

where π is the degree of trust in the prior information. The value of $\pi \in [0, 1]$ may be assigned by the experimenter according to her/his belief in the prior values θ_i^0 . Ahmed (1992), Ahmed and Krzanowski (2004), Bickel and Doksum (2001) and others pointed out that such an estimator yields smaller mean squared error (*MSE*) when a priori information is correct or nearly correct, however, at the expense of poorer performance in the rest of the parameter space. We will demonstrate that $\hat{\boldsymbol{\theta}}_{(SR)}$ has a smaller *MSE* than $\hat{\boldsymbol{\theta}}$ near the restriction, that is, $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. However, $\hat{\boldsymbol{\theta}}_{(SR)}$ becomes considerably biased and inefficient when the restriction may not be judiciously justified. Thus, the performance of this shrinkage procedure depends upon the correctness of the uncertain prior information (*UPI*). The above insight leads to the pretest estimation when the hypothesis information is rather suspicious and it is useful to construct a compromised estimator by performing a preliminary test on the H_0 . Therefore, one may obtain an improved preliminary test estimator as a convex combinations of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{(SR)}$ via a test-statistic for testing H_0 to achieve a bounded

quadratic risk. As such, when the prior information is rather suspicious, it may be desirable to construct an improved preliminary test estimator (*IPTE*) denoted by $\hat{\theta}_{(IPT)}$, which incorporates a preliminary test on $\theta = \theta^0$. Thus, the estimator $\hat{\theta}$ or $\hat{\theta}_{(SR)}$ is selected depending upon the outcome of the preliminary test. If the prior information is tenable, one may use $\hat{\theta}_{(SR)}$; otherwise $\hat{\theta}$ will be chosen.

Naturally, the improved preliminary test estimator is defined as

$$\hat{\theta}_{(IPT)} = \hat{\theta} - \pi(\hat{\theta} - \theta^0)I(T < c_{n,\alpha}), \quad (1.6)$$

where T is a suitable test statistic for the null hypothesis $H_0 : \theta = \theta^0$. $c_{n,\alpha}$ is the critical value, i.e., the $(1 - \alpha) \times 100\%$ percentile, of T under the null hypothesis, and $I(A)$ is the indicator function of a set A . If we substitute $\pi = 1$ in the above relation, then we obtain

$$\hat{\theta}_{(PT)} = \hat{\theta} - (\hat{\theta} - \theta^0)I(T < c_{n,\alpha}). \quad (1.7)$$

The estimator $\hat{\theta}_{(PT)}$ is known as the classical preliminary test estimator (*PTE*), due to Bancroft (1944). Essentially, we have replaced π , a fixed constant, in equation (1.5) by a dichotomous random quantity, $I(T < c_{n,\alpha})$, to obtain a preliminary test estimator. In return, we achieve an estimator with a bounded risk.

In recent literatures, a useful discussion about preliminary testing can be found in Giles and Giles (1993), Magnus (1999), Ohanti (1999), Reif and Vlcek (2002), and Khan and Ahmed (2003). Nevertheless, it is important to remark that $\hat{\theta}_{(PT)}$ performs better than $\hat{\theta}$ in some part of the parameter space. The use of $\hat{\theta}_{(PT)}$ may, however, be limited due to the large size of the preliminary test. The estimators based on the pretest method are sensitive to the departure from H_0 and may not be

useful for all θ . The performance of the preliminary test estimator depends upon the correctness of the hypothesized information. To overcome this shortcoming, we propose James-Stein type estimators (Stein (1956)), which combine the sample and non-sample information in a more superior way than the preceding estimators.

Following Ahmed and Saleh (1993), the *James-Stein type shrinkage estimator (JSE)* is defined by

$$\hat{\theta}_{(JS)} = \hat{\theta} - \{(k-3)T^{-1}\}(\hat{\theta} - \theta^0), \quad k \geq 4. \quad (1.8)$$

Interestingly, the above suggested estimator is simply obtained by replacing the dichotomous quantity (indicator function) in (1.7) by a continuous quantity $((k-3)T^{-1})$. Hence, the suggested shrinkage estimator arises in a natural way. The proposed James-Stein type estimator provides a uniform improvement over $\hat{\theta}$. It is, however, not a convex combination of θ^0 and $\hat{\theta}$. Also, the proposed estimator may not remain non-negative. To avoid this odd behavior of $\hat{\theta}_{(JS)}$, we truncate $\hat{\theta}_{(JS)}$, leading to a convex combination of $\hat{\theta}$ and θ^0 , which is called *positive part James-Stein estimator (PJSE)*. The PJSE is defined as follows:

$$\begin{aligned} \hat{\theta}_{(PP)} &= \hat{\theta} - (k-3)T^{-1}(\hat{\theta} - \theta^0) - \\ &\quad \{1 - (k-3)T^{-1}\}I(T < k-3)(\hat{\theta} - \theta^0), \quad k \geq 4. \end{aligned} \quad (1.9)$$

We shall examine the asymptotic properties of the proposed estimators using the following weighted quadratic loss function:

$$L(\hat{\theta}_*, \theta) = n(\hat{\theta}_* - \theta)'Q(\hat{\theta}_* - \theta), \quad (1.10)$$

where $\hat{\theta}_*$ is an appropriate estimator of θ and Q is a given positive semi-definite matrix. Assume that $G(y) = \lim_{n \rightarrow \infty} P\{\sqrt{n}(\hat{\theta}_* - \theta) \leq y\}$. Then we define the

asymptotic distributional quadratic risk (ADQR) by

$$ADQR(\hat{\theta}_*, \theta) = \int \int \cdots \int \mathbf{y}' \mathbf{Q} \mathbf{y} dG(\mathbf{y}) = \text{trace}(\mathbf{Q} \mathbf{Q}^*), \quad (1.11)$$

where $\mathbf{Q}^* = \int \int \cdots \int \mathbf{y} \mathbf{y}' dG(\mathbf{y})$.

Further, we consider the following contiguous sequence of alternatives to establish the needed asymptotic results:

$$K_{(n)} : \theta = \theta_n, \quad \text{where } \theta_n = \theta_o + \frac{\Delta}{\sqrt{n}}, \quad \Delta \text{ is a real fixed vector.} \quad (1.12)$$

Note that $\Delta = \mathbf{0}$ implies $\theta_n = \theta_o$.

Furthermore, we also compare the risks of suggested estimators under entropy loss functions. Simulation shows that our estimators perform well when sample sizes are small.

1.3 Meta-analysis in Microarray Studies

On the one hand, most of the microarray studies have small to moderate sample sizes, and thus have low statistical power to detect significant relationships between gene expression levels and outcomes of interest; on the other hand, the increasing availability and maturity of DNA microarray technology has led to an explosion of cancer profiling studies. One will not be surprised that a number of research groups perform similar studies on the same subject under alike conditions. In this case, another type of combination of data, i.e., pooling data across multiple studies often improves the power for detecting new relationships. However, this type of pooling is complicated by the fact that the technologies used to isolate, purify, label and

measure RNA are so diverse that the gene expression measurements from different microarray platforms are not directly comparable. Nevertheless, the literature of microarray and other medical researches has become so massive, and the requirements of many scientific journals have made all expression data related to the publication of research available to general public. Due to the ease of access to these data, it is considered to be profitable to overcome the many hurdles when incorporating gene expression data from multiple sources. To extract maximum value from accumulating mass of publicly available gene expression data, methods are needed to evaluate, integrate, and inter validate multiple data sets. Some successful methods for mining information from disparate data sets prove that it is worth the effort. Troyanskaya *et al.* (2003) introduced a MAGIC (Multisource Association of Genes by Integration of Clusters) system, which applied a Bayesian framework for combining heterogeneous data sources for gene function prediction. Johnson *et al.* (2007) proposed parametric and nonparametric empirical Bayes approaches for adjusting the so-called batch effects from different studies. In some studies, instead of combining separate data sets and performing analysis on the unified data, one may skip the raw data and combine the analytic results from multiple studies. In general, these techniques for the aggregation and synthesis of prior research are called “Research synthesis” or “Meta-analysis”.

Meta-analysis aims to compare and combine estimates of effect across related studies. Suppose that a parameter θ of interest can be estimated from k independent sets of data. θ might be a mean, a difference of means, a factorial contrast, a log odds ratio, etc. which is often termed as an overall effect or effect size in meta-analysis.

The k sets might correspond to k independent studies by different investigators. Let t_1, \dots, t_k be the estimates, assumed for the moment normally distributed around θ with known variances v_1, \dots, v_k , calculated from the internal variability within the separate sets of data. If these assumptions are a reasonable basis for the analysis, the “best” combined estimate of θ is

$$\hat{t} = (\sum t_j/v_j)/(\sum 1/v_j), \quad (1.13)$$

which is called a weighted mean of the t_j 's.

A number of assumptions are made in (1.13). The principal one is that the parameter θ is indeed the same for all sets of data. This is a fixed effect model. It is an important general principle that before merging information from different sources, mutual consistency should be checked. A χ^2 test with the following test statistic will serve this purpose:

$$Q = \sum (t_j - \hat{t})^2. \quad (1.14)$$

Suppose that (1.14) shows that the separate estimates t_j differ by more than they should under the assumption of homogeneity, and a specific explanation of the variation in θ is absent, it may be taken as random, i.e., we may take a representation

$$\theta_j = \theta + \delta_j, \quad (1.15)$$

where δ_j are independent random variables normally distributed with zero mean and variance τ^2 . This is equivalent to assuming t_1, \dots, t_k to be independently normally distributed around θ with variances $v_1 + \tau^2, \dots, v_k + \tau^2$. It is unrealistic to assume that v_i and τ^2 are known. The within study variances v_i may be estimated by the

methods mentioned in sections 1.1 and 1.2. Regarding the estimation of the among-study variance τ^2 , many approaches have been proposed over the years. DerSimonian and Laird (1986) proposed a method of moments estimator (MM); Hedges (1983) suggested a variance-component type estimator (VC); Sidik and Jonkman (2005) proposed a simple heterogeneity variance estimator (SH). All of these estimators are simple to compute. In contrast, the maximum likelihood estimator (ML) by Hardy and Thompson (1996), the approximate restricted maximum likelihood estimator (REML) by Morris (1983), and the empirical Bayes estimator (EB) by Morris (1983) are more computational intensive and require iterative solutions.

We construct shrinkage estimators based on the existing estimators. Our simulation study shows that our shrinkage estimators improve upon the base estimators in terms of risk under quadratic and entropy loss functions.

1.4 Outline of the Thesis

This thesis is composed of six chapters.

Chapter 1 provides an overview of the problems that are considered for research in this investigation as well as the literature survey.

In chapter 2, we consider the large-sample inference of population variance and investigate its statistical properties in a multi-sample set up when random samples are drawn from arbitrary populations. Further, asymptotic statistical procedures are developed for testing the homogeneity of these variances. The interval estimation based on combined data from various sources is also proposed. An extensive sampling

experiment is conducted to investigate the performance of the suggested methods over a wide range of data sampling designs such as symmetric and skewed distributions. We tackle the inference problem for variance in a multi-sample situation without the stringent normality assumption. Our contribution is to study point estimation, interval estimation and testing procedures of the population variance parameters when samples are drawn from $k > 1$ arbitrary populations. We provide a total inferential package for the problem at hand. The problem of kurtosis estimation arisen from inferences about variance parameter is also discussed extensively in this chapter.

Chapter 3 develops some alternative estimation strategies when the information regarding the homogeneity of all the variances may not be precise. Assuming that homogeneity holds, it is advantageous to combine the data to estimate the common parameter. However, the combined estimator becomes inconsistent when the equality of the hypothesis does not hold. In this situation, estimators based on pretest (Bancroft (1944)) and the James-Stein (James and Stein (1961)) principles are proposed. Asymptotic properties of the shrinkage estimator, positive-part and pretest estimators are discussed and compared with the standard and combined estimators. It is demonstrated that the positive part estimator utilizes the sample and non-sample information in a superior way relative to the ordinary shrinkage estimator. A simulation study is performed for finite samples, and the result shows that a moderate sample size may be sufficient to utilize the same dominance picture in practical situations.

In chapter 4, we study the risks of a class of linear weighted combination estimators of variance based on finite samples. Two optimal weight combination estimators under

quadratic loss function are proposed. An application of the estimation strategies in the microarray study is included in this chapter.

In chapter 5, we deal with the estimation of the variance components in meta-analysis with random effects. Some existing methods are reviewed and the risks are compared. Shrinkage estimators are proposed based on the risk comparison. A simulation study shows that our shrinkage estimators outperform the base estimator. Further, they behave more robustly with respect to other estimators under study in the sense that they dominate other estimators in most of the parameter space.

Finally, in chapter 6, some general conclusions and directions for the future research are offered.

Chapter 2

Multi-Sample Analysis of Population Variances and Kurtosis Estimation

2.1 Introduction

In this chapter, we consider the large-sample inference of population variance and investigate its statistical properties in a multi-sample set up when random samples are drawn from arbitrary populations. Further, asymptotic statistical procedures are developed for testing the homogeneity of these variances. The interval estimation based on combined data from various sources is also proposed. An extensive sampling experiment is conducted to investigate the performance of the suggested methods over a wide range of data sampling designs such as symmetric and skewed distributions. We tackle the inference problem for variance in a multi-sample situation without

the stringent normality assumption. Our contribution is to study point estimation, interval estimation and testing procedures of the population variance parameters when samples are drawn from $k > 1$ arbitrary populations.

The structure of this chapter is as follows. In Section 2, we present the preliminary notations and basic assumptions. Section 3 deals with the testing problem for the equality of several variances. A variance stabilizing transformation is introduced and the testing problems are revisited in Section 4. In Section 5, we investigate the properties of power functions of all the proposed tests. The interval estimation problem is tackled in section 6. In Section 7 the problem of kurtosis estimation arising from hypothesis testing is discussed extensively. The results of a simulation study on the proposed asymptotic procedures are reported in section 8. The purpose of the simulation study is to evaluate the properties of the proposed methods for small, moderate and large samples from normal and non-normal populations. To illustrate our method, an example is given in section 9. The concluding remarks are presented in the last section.

2.2 Preliminaries and Large Sample Results

In this section, we establish large sample results for inference purposes.

As usual, our starting point is a basic random experiment with an underlying sample space Ω and a probability measure P . Let Y be a real-valued random variable for the experiment with mean μ and variance σ^2 . However, we are interested in a multi-sample situation. Suppose that k (greater than 1) similar and indepen-

dent experiments have been completed. Let $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$, $i = 1, 2, \dots, k$, be k independent samples from underlying populations having a distribution with mean parameter μ_i , and variance parameter σ_i^2 . Let $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ be the estimators of μ_i and σ_i^2 respectively, where

$$\hat{\mu}_i = \frac{1}{n} \sum_{j=1}^{n_i} Y_{ij}, \quad \hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2.$$

Assume that the fourth moment, μ_{4i} , exists, then

$$n_i^{1/2}(\hat{\mu}_i - \mu_i) \xrightarrow{D} \mathcal{N}(0, \sigma_i^2), \quad n_i^{1/2}(\hat{\sigma}_i^2 - \sigma_i^2) \xrightarrow{D} \mathcal{N}(0, \mu_{4i} - \sigma_i^4), \quad (2.1)$$

where the notation \xrightarrow{D} means *convergence in distribution*. Further, the asymptotic variance can be written in the following canonical form,

$$\mu_{4i} - \sigma_i^4 = (\gamma_i^* - 1)\sigma_i^4 \quad (2.2)$$

where $\gamma_i^* = \mu_{4i}/\sigma_i^4$ and is called the *kurtosis* of the distribution.

Further, if we assume that the populations are normal then

$$n_i^{1/2}(\hat{\mu}_i - \mu_i) \xrightarrow{D} \mathcal{N}(0, \sigma_i^2), \quad n_i^{1/2}(\hat{\sigma}_i^2 - \sigma_i^2) \xrightarrow{D} \mathcal{N}(0, 2\sigma_i^4)$$

Now we turn to the main objective of this investigation. Let us define the parameter vector $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)'$ which is estimated by $\hat{\boldsymbol{\sigma}}^2 = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_k^2)'$. In the following section, we consider a large-sample testing problem for the variances.

2.3 Hypothesis Testing

In this section, we conduct some statistical hypothesis testing problems for the population variances. For the simple null versus global alternative, let's suppose it is

desired to test the following hypothesis

$$H_0 : \sigma^2 = \sigma_o^2, \quad \text{where } \sigma_o^2 = (\sigma_{1o}^2, \sigma_{2o}^2, \dots, \sigma_{ko}^2)'$$

against the global alternative

$$H_a : \sigma^2 \neq \sigma_o^2.$$

It is natural to propose the following test statistic for the null hypothesis, which is defined by normalized distance of $\hat{\sigma}^2$ from σ_o^2 :

$$T_1 = n(\hat{\sigma}^2 - \sigma_o^2)' \hat{\Gamma}_1^{-1} (\hat{\sigma}^2 - \sigma_o^2),$$

where $n = \sum_{i=1}^k n_i$, and

$$\hat{\Gamma}_1^{-1} = \hat{\Omega}_1, \quad \hat{\Omega}_1 = \text{Diag} \left(\frac{\omega_{1,n}}{(\hat{\gamma}_1^* - 1)\sigma_{1o}^4}, \dots, \frac{\omega_{k,n}}{(\hat{\gamma}_k^* - 1)\sigma_{ko}^4} \right),$$

where $\omega_{i,n} = \frac{n_i}{n}$ and $\hat{\gamma}_i^*$, $i = 1, \dots, k$, is a consistent estimator of γ^* . Estimation of γ^* will be discussed in detail in section 2.7. Here, we assume that $\lim_{n \rightarrow \infty} (\omega_{i,n}) = \omega_i$ ($0 < \omega_i < 1$) is fixed for $i = 1, \dots, k$. When the null hypothesis is true, the large sample distribution of T_1 converges to a central χ^2 distribution with k degrees of freedom. Hence, the upper α -level critical value of T_1 may be approximated by the $(1 - \alpha)$ th percentile of the central χ^2 distribution with k degrees of freedom. Finally, from a practitioner's point of view we present the above test statistic in the following scalar form which is easily computable:

$$T_1 = \sum_{i=1}^k \left[\frac{n_i}{\hat{\gamma}_i^* - 1} \left(\frac{\hat{\sigma}_i^2 - \sigma_{io}^2}{\sigma_{io}^2} \right)^2 \right]. \quad (2.3)$$

A more interesting hypothesis would be whether the variation of underlying populations, as measured by the σ_i^2 , can be regarded as homogeneous across studies, i.e.,

to test $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$, where σ_i^2 denotes the population variance in study i .

Thus, the hypotheses are

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2(\text{unknown}) \quad \text{vs} \quad H_a : \sigma_i^2 \neq \sigma_h^2, \\ \text{for at least one pair of } (i, h) \quad (2.4)$$

where $(i, h) \in (1, 2, \dots, k)$, and $i \neq h$.

In sequel, for the sake of brevity we assume homogeneity of the kurtosis, i.e., $\gamma_1^* = \gamma_2^* = \dots = \gamma_k^* = \gamma^*$ (unknown) and the remaining discussion follows. We define the estimate of the common variance

$$\hat{\sigma}_{(R)}^2 = \frac{1}{n-k} (n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2 + \dots + n_k \hat{\sigma}_k^2). \quad (2.5)$$

We propose the following test statistic for the null hypothesis (2.4), which is defined by the normalized distance of $\hat{\sigma}^2$ from $\hat{\sigma}_{(R)}^2$ and is given by

$$T_2 = n(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2 \mathbf{1}_k)' \hat{\Gamma}_2^{-1} (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2 \mathbf{1}_k), \quad (2.6)$$

where $\mathbf{1}_k$ is a unit vector, $\hat{\Gamma}_2^{-1} = \frac{\mathbf{\Omega}_2}{(\hat{\gamma}_{(R)}^* - 1)(\hat{\sigma}_{(R)}^2)^2}$, $\mathbf{\Omega}_2 = \text{Diag}(\omega_{1,n}, \dots, \omega_{k,n})$, and $\hat{\gamma}_{(R)}^*$ is the pooled estimator of the common kurtosis parameter, γ^* , and is given by

$$\hat{\gamma}_{(R)}^* = \frac{1}{n} (n_1 \hat{\gamma}_1^* + n_2 \hat{\gamma}_2^* + \dots + n_k \hat{\gamma}_k^*). \quad (2.7)$$

Since, $\hat{\gamma}_{(R)}^*$ and $\hat{\sigma}_{(R)}^2$ are consistent estimators of γ^* and σ^2 , then for large n , it is appropriate to approximate the distribution of T_2 by a χ^2 distribution with degrees of freedom $(k-1)$. Thus, an approximate test of H_0 is obtained by referring T_2 to the tables of the chi-square distribution with $(k-1)$ degrees of freedom. We may rewrite the proposed test statistic in the following computationally attractive form:

$$T_2 = \sum_{i=1}^k \left[\frac{n_i}{\hat{\gamma}_{(R)}^* - 1} \left(\frac{\hat{\sigma}_i^2 - \hat{\sigma}_{(R)}^2}{\hat{\sigma}_{(R)}^2} \right)^2 \right] \quad (2.8)$$

Recall that the asymptotic distribution of $n_i^{1/2}\{\hat{\sigma}_i^2 - \sigma_i^2\}$ depends on two unknown parameters σ_i^2 and the kurtosis γ_i^* . We present a variance stabilizing transformation so that asymptotic variance will not depend on the parameter σ_i^2 itself.

2.4 Variance Stabilizing Transformation

The asymptotic distribution of the quantity $n_i^{1/2}\{\hat{\sigma}_i^2 - \sigma_i^2\}$ follows a normal distribution with mean 0 and variance $(\gamma_i^* - 1)(\sigma_i^2)^2$. Since the asymptotic variance depends on unknown parameters σ_i^2 , the question is whether we could find a suitable transformation of $\hat{\sigma}_i^2$ for which the asymptotic variance does not depend on the unknown parameter σ_i^2 . To obtain a variance-stabilizing transformation we must define a function $g[\hat{\sigma}_i^2]$ such that $n_i^{1/2}\{g[\hat{\sigma}_i^2] - g[\sigma_i^2]\} \xrightarrow{D} N(0, a_i^2)$, where a_i is independent of σ_i^2 . In the following, we show that this may be achieved by taking a natural *log* transformation.

Theorem 2.4.1. *For large n_i , $n_i^{1/2}\{\log[\hat{\sigma}_i^2] - \log[\sigma_i^2]\} \xrightarrow{D} N(0, (\gamma_i^* - 1))$.*

The following lemma known as the Delta method is used for the proof of the theorem.

Lemma 2.4.2. *(Lehmann (1999)) Let $\hat{\theta}_i$ be an estimator of the parameter θ_i . Suppose that $n_i^{1/2}\{\hat{\theta}_i - \theta_i\} \xrightarrow{D} N(0, \sigma_i^2)$ and let g be a continuous function such that $g'(\theta_i)$ exists and $g'(\theta_i) \neq 0$. Then it follows that $n_i^{1/2}\{g(\hat{\theta}_i) - g(\theta_i)\}/\sigma_i g'(\theta_i) \xrightarrow{D} N(0, 1)$.*

Proof of the Theorem: Letting $g(x) = \log(x)$, we have $g'(x) = x^{-1}$ and $g'(\sigma_i^2) = \frac{1}{\sigma_i^2} \neq 0$. Thus, by the above lemma the result of the theorem follows.

Again, we consider two classes of testing problems for arbitrary populations:

- (i) simple null versus global alternative, and
- (ii) test for homogeneity.

First, consider

$$H_0 : \sigma^2 = \sigma_o^2 \quad \text{vs.} \quad H_a : \sigma^2 \neq \sigma_o^2.$$

Based on the log transformation, we have

$$T_3 = n (\log[\hat{\sigma}^2] - \log[\sigma_o^2])' \hat{\Gamma}_3^{-1} (\log[\hat{\sigma}^2] - \log[\sigma_o^2]), \quad (2.9)$$

where

$$\hat{\Gamma}_3^{-1} = \hat{\Omega}_3, \quad \hat{\Omega}_3 = \text{Diag} \left(\frac{\omega_{1,n}}{(\hat{\gamma}_1^* - 1)}, \dots, \frac{\omega_{k,n}}{(\hat{\gamma}_k^* - 1)} \right). \quad (2.10)$$

Alternatively, T_3 can be written as

$$T_3 = \sum_{i=1}^k \left[\frac{n_i}{\hat{\gamma}_i^* - 1} (\log[\hat{\sigma}_i^2] - \log[\sigma_{io}^2])^2 \right] \quad (2.11)$$

When the null hypothesis is true, the large sample distribution of T_3 converges to a central χ^2 distribution with k degrees of freedom. Hence, the upper α -level critical value of T_3 may be approximated by central χ^2 distribution with k degrees of freedom.

Next, let

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \quad \text{and} \quad H_a : \sigma_i^2 \neq \sigma_h^2 \quad \text{for at least one pair of } i \neq h.$$

Then a reasonable test statistic is

$$T_4 = n (\log[\hat{\sigma}^2] - \log[\hat{\sigma}_{(R)}^2] \mathbf{1}_k)' \hat{\Gamma}_4^{-1} (\log[\hat{\sigma}^2] - \log[\hat{\sigma}_{(R)}^2] \mathbf{1}_k), \quad (2.12)$$

where

$$\hat{\Gamma}_4^{-1} = \frac{\Omega_2}{(\hat{\gamma}_{(R)}^* - 1)}, \quad \Omega_2 = \text{Diag} (\omega_{1,n}, \dots, \omega_{k,n}).$$

The above test statistic may be written as follows:

$$T_4 = \sum_{i=1}^k \left[\frac{n_i}{\hat{\gamma}_{(R)}^* - 1} (\log[\hat{\sigma}_i^2] - \log[\hat{\sigma}_{(R)}^2])^2 \right]. \quad (2.13)$$

Since $\hat{\gamma}_{(R)}^*$ is a consistent estimator of the common parameter γ^* , for large n_i , it is appropriate to approximate the distribution of T_4 by the chi-squared distribution with degrees of freedom $k - 1$.

In the following section, we study the power of our proposed tests.

2.5 A Power Study

It is important to note that for a fixed alternative hypothesis, the power of all four tests statistics proposed earlier will converge to 1 as $n \rightarrow \infty$. Thus, to study the asymptotic power properties of T_l ($l = 1, 2, 3, 4$), we confine ourselves to a sequence of local alternatives $\{K_n\}$. When σ^2 is the parameter of interest, such a sequence may be specified by

$$K_n : \sigma^2 = \sigma_o^2 + \frac{\delta}{n^{1/2}}, \quad (2.14)$$

where δ is a vector of fixed real numbers. Evidently, σ^2 approaches σ_o^2 at a rate proportional to $n^{-1/2}$. Stochastic convergence of $\hat{\sigma}^2$ to σ^2 ensures convergence under local alternatives as well.

Theorem 2.5.1. *Under the local alternatives in (2.14), if $n_i/n \rightarrow \omega_i$ ($0 < \omega_i < 1$) as $n_i \rightarrow \infty$, then*

$$\begin{aligned} n^{1/2}\{\hat{\sigma}^2 - \sigma_o^2\} &\xrightarrow{D} \mathcal{N}_k(\delta, \Gamma_1), \\ n^{1/2}\{\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2 \mathbf{1}_k\} &\xrightarrow{D} \mathcal{N}_k(\mathbf{J}\delta, \Gamma_2), \end{aligned}$$

$$n^{1/2}\{\log[\hat{\sigma}^2] - \log[\sigma_0^2]\} \xrightarrow{D} \mathcal{N}_k(\boldsymbol{\delta}, \boldsymbol{\Gamma}_3),$$

$$n^{1/2}\{\log[\hat{\sigma}^2] - \log[\hat{\sigma}_{(R)}^2]\mathbf{1}_k\} \xrightarrow{D} \mathcal{N}_k(\mathbf{J}\boldsymbol{\delta}, \boldsymbol{\Gamma}_4),$$

where

$$\begin{aligned} \boldsymbol{\Gamma}_1 &= \text{Diag}\left(\frac{(\gamma_1^* - 1)(\sigma_{1o}^2)^2}{\omega_1}, \dots, \frac{(\gamma_k^* - 1)(\sigma_{ko}^2)^2}{\omega_k}\right), \\ \boldsymbol{\Gamma}_2 &= \frac{1}{4}(\gamma^* - 1)(\sigma^2)^2 \boldsymbol{\Omega}_2^{-1} \mathbf{J}', \quad \mathbf{J} = \mathbf{I} + \mathbf{1}\mathbf{1}'\boldsymbol{\Omega}_2, \\ \boldsymbol{\Gamma}_3 &= \text{Diag}\left(\frac{(\gamma_1^* - 1)}{\omega_1}, \dots, \frac{(\gamma_k^* - 1)}{\omega_k}\right), \\ \boldsymbol{\Gamma}_4 &= (\gamma^* - 1)\boldsymbol{\Omega}_2^{-1} \mathbf{J}'. \end{aligned}$$

Therefore, each $T_j, j = 1, 2, 3, 4$ has asymptotically a noncentral chi-square distribution with non-centrality parameters Δ_l , where

$$\begin{aligned} \Delta_1 &= \boldsymbol{\delta}'\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\delta}, \\ \Delta_2 &= (\mathbf{J}\boldsymbol{\delta})'\boldsymbol{\Gamma}_2^{-1}(\mathbf{J}\boldsymbol{\delta}), \\ \Delta_3 &= \boldsymbol{\delta}'\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\delta}, \\ \Delta_4 &= (\mathbf{J}\boldsymbol{\delta})'\boldsymbol{\Gamma}_4^{-1}(\mathbf{J}\boldsymbol{\delta}) \end{aligned}$$

Hence, calculation for the power of the test statistic can be done by using noncentral chi-square distribution.

2.6 Interval Estimation

In this section, we propose interval estimation procedures for σ_i^2 . Note that,

$$Pr \left\{ \log[\hat{\sigma}_i^2] - z_{\alpha/2} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2} \leq \log[\sigma_i^2] \leq \log[\hat{\sigma}_i^2] + z_{\alpha/2} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2} \right\}$$

converges to $1 - \alpha$ as $n \rightarrow \infty$. This, in turn, implies that

$$Pr \left\{ e^{\log[\hat{\sigma}_i^2] - z_{\alpha/2} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2}} \leq \sigma_i^2 \leq e^{\log[\hat{\sigma}_i^2] + z_{\alpha/2} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2}} \right\}$$

converges to $1 - \alpha$. So, the probability that the random interval

$$\left(e^{\log[\hat{\sigma}_i^2] - z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2}}, e^{\log[\hat{\sigma}_i^2] + z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2}} \right) \quad (2.15)$$

or

$$\left(\hat{\sigma}_i^2 e^{-z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2}}, \hat{\sigma}_i^2 e^{z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2}} \right). \quad (2.16)$$

includes the unknown population variance σ_i^2 is $1 - \alpha$, and serves as an asymptotic $100(1 - \alpha)\%$ confidence interval for σ_i^2 .

On the other hand,

$$Pr \left\{ \hat{\sigma}_i^2 - z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2} \hat{\sigma}_i^2 \leq \sigma_i^2 \leq \hat{\sigma}_i^2 + z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_i^* - 1}{n_i} \right)^{1/2} \hat{\sigma}_i^2 \right\} \quad (2.17)$$

converges to $1 - \alpha$ as $n_i \rightarrow \infty$. Thus,

$$\left(1 \pm z_{\frac{\alpha}{2}} \left\{ \frac{\hat{\gamma}_i^* - 1}{n_i} \right\}^{1/2} \right) \hat{\sigma}_i^2 \quad (2.18)$$

represents an asymptotic $100(1 - \alpha)\%$ confidence interval for σ_i^2 based on the original data.

If the null hypothesis of homogeneity of the variances is not rejected, then it may be of interest to obtain a $100(1 - \alpha)\%$ confidence interval about the common value of σ^2 . A $100(1 - \alpha)\%$ confidence interval about σ^2 may be obtained by using the combined data. Noting that

$$Pr \left\{ \log[\hat{\sigma}_{(R)}^2] - z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \leq \log[\sigma^2] \leq \log[\hat{\sigma}_{(R)}^2] + z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \right\}$$

converges to $1 - \alpha$ as $n \rightarrow \infty$, we have

$$Pr \left\{ e^{\log[\hat{\sigma}_{(R)}^2] - z_{\frac{\alpha}{2}}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \leq \sigma^2 \leq e^{\log[\hat{\sigma}_{(R)}^2] + z_{\frac{\alpha}{2}}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \right\}$$

converges to $1 - \alpha$. This means that

$$\left(e^{\log[\hat{\sigma}_{(R)}^2] - z_{\frac{\alpha}{2}}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2}, e^{\log[\hat{\sigma}_{(R)}^2] + z_{\frac{\alpha}{2}}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \right) \quad (2.19)$$

serves as an asymptotic $100(1 - \alpha)\%$ confidence interval for the common variance parameter σ^2 .

For original data, as $n \rightarrow \infty$,

$$Pr \left\{ \hat{\sigma}_{(R)}^2 - z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \hat{\sigma}_{(R)}^2 \leq \sigma^2 \leq \hat{\sigma}_{(R)}^2 + z_{\frac{\alpha}{2}} \left(\frac{\hat{\gamma}_{(R)}^* - 1}{n} \right)^{1/2} \hat{\sigma}_{(R)}^2 \right\} \quad (2.20)$$

converges to $1 - \alpha$. The endpoints of an asymptotic $100(1 - \alpha)\%$ confidence interval for the common variance σ^2 are given by

$$\hat{\sigma}_{(R)}^2 \left(1 \pm z_{\frac{\alpha}{2}} \left\{ \frac{\hat{\gamma}_{(R)}^* - 1}{n} \right\}^{1/2} \right). \quad (2.21)$$

We realize that the asymptotic variance of the limiting distribution is a function of the kurtosis of the distribution. Thus, a better estimation of the kurtosis would become an important task. In the following section, several estimators are studied and a new bias corrected estimator of kurtosis is proposed.

2.7 Estimation of Kurtosis

Over the years, many estimators of the population kurtosis have been proposed in the reviewed literature (Fisher (1929), Hogg (1972), Balanda and MacGillivray (1988), Moors (1988), An and Ahmed (2007), among others).

2.7.1 Preliminaries and Background

The population kurtosis parameter is traditionally defined as the standardized fourth population moment about the mean (provided it exists),

$$\gamma^* = \frac{E(X - \mu)^4}{(E(X - \mu)^2)^2} = \frac{\mu_4}{\sigma^4}. \quad (2.22)$$

Alternatively,

$$\gamma = \gamma^* - 3 \quad (2.23)$$

is often used so that the normal distribution has a kurtosis of zero.

Let X_1, X_2, \dots, X_n be a random sample of size n from an arbitrary population. then a commonly used consistent estimator of γ is given by

$$\hat{\gamma} = \frac{n \sum (x_i - \bar{x})^4}{[\sum (x_i - \bar{x})^2]^2} - 3, \quad \text{where } \bar{x} = \frac{\sum x_i}{n}. \quad (2.24)$$

Note that the above estimator is not unbiased. Indeed, the bias and variance of the estimator depend on the underlying population distribution. For example, Cramer (1946) gave the following results for normal distributions:

$$\text{Bias}(\hat{\gamma}) = -6/(n + 1), \quad (2.25)$$

$$\text{Var}(\hat{\gamma}) = \frac{24n(n - 2)(n - 3)}{(n + 1)^2(n + 3)(n + 5)}. \quad (2.26)$$

Another frequently used estimator of γ is defined as

$$\hat{\gamma}^U = \frac{(n - 1)}{(n - 2)(n - 3)} \{(n + 1)\hat{\gamma} + 6\}. \quad (2.27)$$

It has been proved that $\hat{\gamma}^U$ is unbiased for normal distributions. We refer to Fisher (1929), Kendall *et al.* (1987), Joanes and Gill (1998) and others.

Note that $\hat{\gamma}^U$ has been adapted by various computing packages, such as SAS, SPSS, and S-Plus. For example, SAS PROC MEAN gives $\hat{\gamma}^U$. On the other hand, using BIASKUR option in CALIS procedure yields $\hat{\gamma}$.

In contrast, the kurtosis measure adapted by MINITAB is defined by

$$\hat{\gamma}^M = \left(\frac{n-1}{n} \right)^2 (\hat{\gamma} + 3) - 3. \quad (2.28)$$

Joanes and Gill (1998) showed that for normal distributions,

$$\text{Bias}(\hat{\gamma}^M) = 3 \left[\frac{(n-1)^3}{n^2(n+1)} \right] - 3 \simeq -\frac{12}{n+1}. \quad (2.29)$$

It is seen from expressions (2.27) and (2.28) that:

$$\text{Var}(\hat{\gamma}^U) = \left\{ \frac{(n-1)(n+1)}{(n-2)(n-3)} \right\}^2 \text{Var}(\hat{\gamma}) \simeq \left(1 + \frac{10}{n} \right) \text{Var}(\hat{\gamma}), \quad (2.30)$$

$$\text{Var}(\hat{\gamma}^M) = \left(\frac{n-1}{n} \right)^4 \text{Var}(\hat{\gamma}) \simeq \left(1 - \frac{4}{n} \right) \text{Var}(\hat{\gamma}). \quad (2.31)$$

Hence, in general,

$$\text{Var}(\hat{\gamma}^M) < \text{Var}(\hat{\gamma}) < \text{Var}(\hat{\gamma}^U). \quad (2.32)$$

However, in order to provide a fair comparison we use the *MSE* criterion. For a normal population,

$$MSE(\hat{\gamma}) < MSE(\hat{\gamma}^M) < MSE(\hat{\gamma}^U).$$

Thus, $\hat{\gamma}$, available in SAS and S-Plus, has the smallest *MSE*, more importantly, when sample size is small. In this case the performance of the unbiased estimator $\hat{\gamma}^U$ is the worst due to the large magnitude of the variance. Interestingly, this estimator is also adapted by SAS, SPSS and S-Plus.

This comparison gives some insight and forms a basis for introducing some new estimators, particularly when sample size is small or moderate. We propose two new improved estimators in the following Section.

2.7.2 Proposed Estimation Strategies

Since both $\hat{\gamma}$ and $\hat{\gamma}^M$ are biased estimators with lower variance compared to $\hat{\gamma}^U$, correcting the bias given in (2.25) and (2.29) yields two new estimators as follows:

$$\hat{\gamma}^{N1} = \hat{\gamma} + 6/(n+1) \quad (2.33)$$

and

$$\hat{\gamma}^{N2} = \hat{\gamma}^M + 3 \left[1 - \frac{(n-1)^3}{n^2(n+1)} \right]. \quad (2.34)$$

Consequently, for normal data, $\hat{\gamma}^{N1}$ and $\hat{\gamma}^{N2}$ are both unbiased estimators of γ .

Now let $E_1, E_2, E_3, E_4,$ and E_5 denote the estimators $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively. It is a straightforward matter to derive that when underlying population distribution is normal,

$$MSE(E_1) = \frac{24n^3 - 84n^2 + 432n + 540}{(n+1)^2(n+3)(n+5)}, \quad (2.35)$$

$$MSE(E_2) = \frac{24n(n-1)^2}{(n-2)(n-3)(n+3)(n+5)}, \quad (2.36)$$

$$MSE(E_3) = \frac{24(n-1)^4(n-2)(n-3)}{n^3(n+1)^2(n+3)(n+5)} + \frac{9[(n-1)^3 - n^2(n+1)]^2}{n^4(n+1)^2}, \quad (2.37)$$

$$MSE(E_4) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}, \quad (2.38)$$

$$MSE(E_5) = \frac{24(n-1)^4(n-2)(n-3)}{n^3(n+1)^2(n+3)(n+5)}. \quad (2.39)$$

It can be shown that for all $n > 3$,

$$MSE(E_5) < MSE(E_4) < MSE(E_1) < MSE(E_3) < MSE(E_2), \quad (2.40)$$

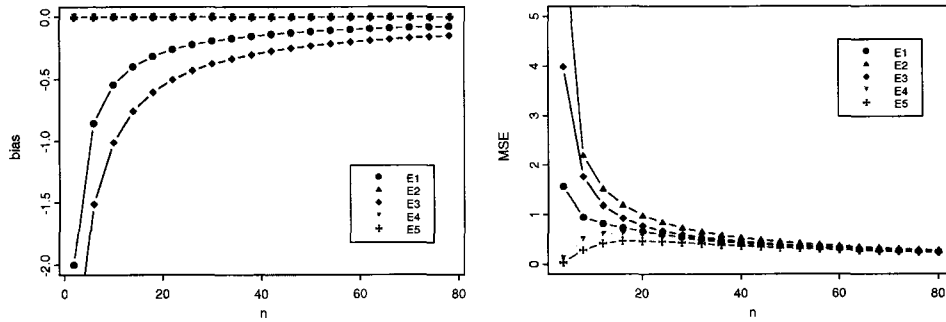


Figure 2.1: Bias and MSE of the kurtosis estimators for normal distributions. E_1 , E_2 , E_3 , E_4 , and E_5 denote the estimators $\hat{\gamma}$, $\hat{\gamma}^U$, $\hat{\gamma}^M$, $\hat{\gamma}^{N1}$, and $\hat{\gamma}^{N2}$, respectively.

keeping in mind that E_4 and E_5 are the proposed estimators. Hence, the two proposed estimators are superior to the existing estimators in terms of MSE for normal data.

Figure 2.1 is a visualization of the bias and MSE comparison of the five estimators when sample size n varies. The graphical analysis reveals that both E_1 and E_3 underestimate the population kurtosis and the negative bias is substantially large when n is small. The magnitude of bias decreases as sample size n increases and eventually becomes negligible. The estimators E_2 , E_4 , and E_5 are all unbiased. Furthermore, the graph clearly indicates that our estimators E_4 and E_5 have significantly smaller MSE , especially when n is small, and hence are superior to the existing estimators.

In order to quantify the relative efficiencies of the estimators, we define the efficiency of the estimators relative to E_2 by

$$RE_j = \frac{MSE(E_2)}{MSE(E_j)}, \quad \text{where } j = 1, 3, 4, 5. \quad (2.41)$$

The REs are calculated for selected sample sizes and listed in Table 2.1. It can

Table 2.1: The relative efficiency of the kurtosis estimators

n	RE_1	RE_3	RE_4	RE_5
5	3.200	1.307	16.00	39.01
10	2.053	1.274	3.125	4.763
20	1.490	1.271	1.700	2.087
30	1.314	1.215	1.414	1.619
40	1.230	1.173	1.293	1.431
50	1.181	1.144	1.227	1.330
100	1.088	1.076	1.106	1.152

be observed that when n is small, our estimators E_4 and E_5 are highly efficient compared to the others. For example, when $n = 5$, RE_4 and RE_5 are 16.00 and 39.01, respectively. In passing, we would like to remark here that as the sample size increases, the MSE difference among all the estimators becomes smaller and eventually negligible.

In summary, all indicators show that our estimators are more advantageous compared to the existing ones adapted by various software packages when data are from normal population. However, data can arise from non-normal populations. In the following section, we will look at their relative performance when data come from an arbitrary population through simulation studies.

2.7.3 A Simulation Study of Kurtosis

In this section we showcase a Monte Carlo simulation study to investigate the performance of various estimators for some sampling designs. Moreover, an empirical bias-corrected estimator for skewed and heavy-tailed data is proposed.

The bias and MSE of the kurtosis estimators are now simulated for contaminated

normal, heavy-tailed, and skewed data, each based on 5000 Monte-Carlo simulations.

The results are presented in Figures 2.2 to 2.7.

Contaminated Normal Data The probability density function (pdf) of a mixture of k normal random variable X is defined by

$$f(x) = \sum_{j=1}^k p_j \phi_j(x; \mu_j, \sigma_j^2). \quad (2.42)$$

where, for $j = 1, \dots, k$,

$$\phi_j(x; \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x - \mu_j)^2}{2\sigma_j^2}\right)$$

is the pdf of a normal random variable with mean μ_j and variance σ_j^2 , and

$$0 \leq p_j \leq 1, \quad \sum_{j=1}^k p_j = 1.$$

The population kurtosis of this distribution is given in Wang (2000) as follows:

$$\gamma^* = \frac{1}{\sigma^4} \sum_{j=1}^k p_j [3\sigma_j^4 + 6(\mu_j - \mu)^2 \sigma_j^2 + (\mu_j - \mu)^4],$$

where $\mu = \sum_{j=1}^k p_j \mu_j$ and $\sigma^2 = \sum_{j=1}^k p_j (\sigma_j^2 + \mu_j^2) - \mu^2$ are the mean and variance of the mixture distribution respectively.

Result 2.7.1. For a mixture of two normal random variables, if $\mu_1 = \mu_2$, then

$$\gamma^* = \frac{3(p_1 \sigma_1^4 + p_2 \sigma_2^4)}{(p_1 \sigma_1^2 + p_2 \sigma_2^2)^2}.$$

Note that minimum occurs when either $p_1 = 0$ or 1 , with the minimum value equal to 3. Further, it is maximized when $p_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$ and $p_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$, and the maximum value is $\frac{3}{4}(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} + 2)$.

Thus, in our simulation study, we considered mixtures of two normal distributions with the following parameter configurations:

Configuration 1. $\mu_1 = \mu_2, \sigma_1/\sigma_2 = 1/2, n = 20, 0 \leq p_1 \leq 1;$

Configuration 2. $\mu_1 = \mu_2, \sigma_1/\sigma_2 = 1/3, n = 20, 0 \leq p_1 \leq 1;$

Configuration 3. $\mu_1 = \mu_2, \sigma_1/\sigma_2 = 1/2, n = 50, 0 \leq p_1 \leq 1;$

Configuration 4. $\mu_1 = \mu_2, \sigma_1/\sigma_2 = 1/3, n = 50, 0 \leq p_1 \leq 1.$

Figure 2.2 (a)-(d) exhibits the bias comparison of the five estimators. The values of the population kurtosis together with the Monte-Carlo means of the five estimators for above four parameter configurations are presented in the figure. The graphs show that when σ_1/σ_2 is fixed, the population kurtosis γ^* starts from 3 as $p_1 = 0$, increases with the increasing of the weight p_1 , reaches its maximum value as $p_1 = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$, and decreases back to 3 when $p_1 = 1$. For example, kurtosis takes its maximum value at $p_1 = 0.8$ when $\sigma_1/\sigma_2 = 0.5$. This agrees with Result 2.7.1. It is observed as well that all estimators underestimate the population kurtosis in the entire parameter space. Regarding the magnitude of the simulated bias (SB), the graph reveals that

$$SB(\hat{\gamma}^U) < SB(\hat{\gamma}^{N1}) < SB(\hat{\gamma}^{N2}) < SB(\hat{\gamma}) < SB(\hat{\gamma}^M). \quad (2.43)$$

The bias decreases as the sample size increases. When $\sigma_1/\sigma_2 \rightarrow 1$, the distribution becomes normal. Conversely, when σ_1/σ_2 moves away from 1, the performance of the estimators become very poor around the maximum value of the kurtosis.

The simulated mean squared error ($SMSE$) of the estimators is given in Figure 2.3. The $SMSE$ increases dramatically as σ_1/σ_2 moves away from 1. Increasing sample size n improves the estimation, yet not good enough to get an acceptable

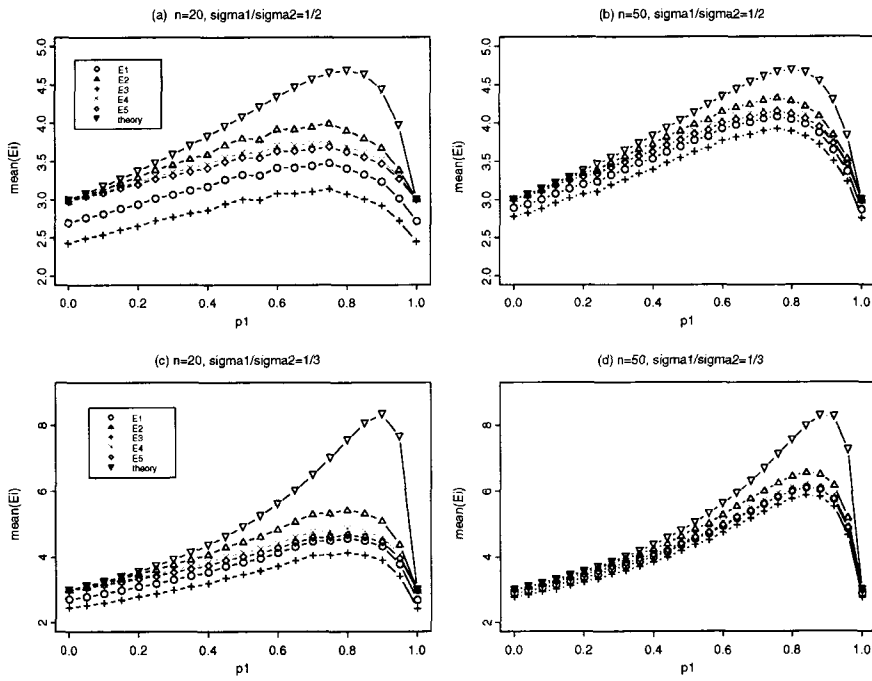


Figure 2.2: Bias comparison of the kurtosis estimators for a mixture of two normal populations. E_1, E_2, E_3, E_4, E_5 , and *theory* represent the expected values of the estimators $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1}, \hat{\gamma}^{N2}$, and the theoretical value γ , respectively. The bias is measured by the distance between the mean of the estimator and the theoretical value.

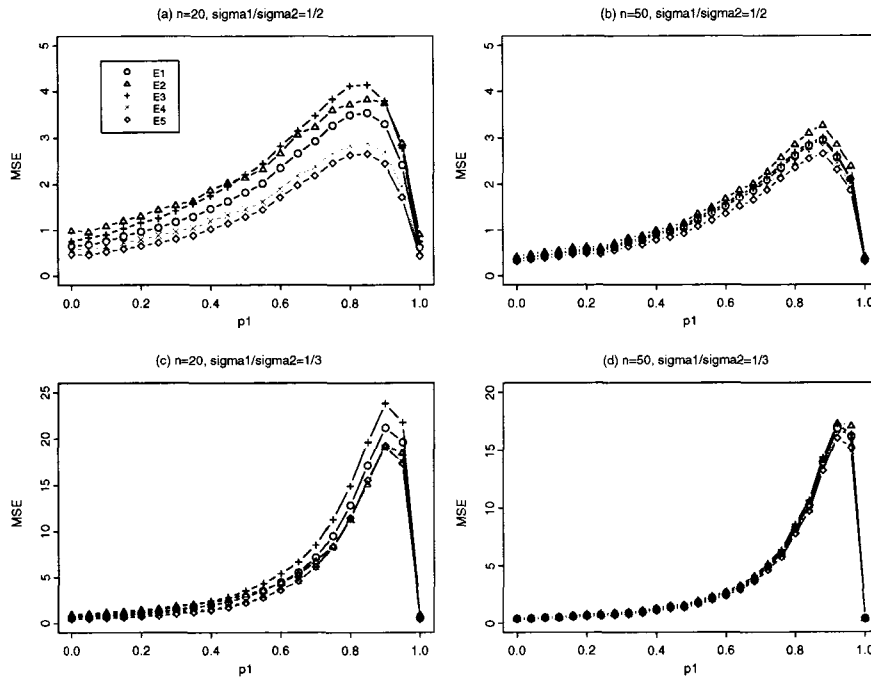


Figure 2.3: *MSE* of the kurtosis estimators for mixtures of two normal populations. E_1 , E_2 , E_3 , E_4 , and E_5 represent $\hat{\gamma}$, $\hat{\gamma}^U$, $\hat{\gamma}^M$, $\hat{\gamma}^{N1}$, and $\hat{\gamma}^{N2}$, respectively.

performance. However, comparing the *SMSE* of the five estimators, we notice that $\hat{\gamma}_2^{N2}$ and $\hat{\gamma}_2^{N1}$ perform relatively better.

Heavy-tailed Data Samples of sizes 20 and 50 are generated from Student- t distributions with degrees of freedom v ranging from 5 to 25. The *SB* of the five estimators for the t distribution is presented in Figure 2.4. Part (a) is based on samples of size 20 and part (b) is based on samples of size 50. It is observed that for fixed sample size, the bias of all five estimators is considerably large when the degrees of freedom is small (say ≤ 6). The bias decreases and approaches 0 as $v \rightarrow \infty$. When sample size is small, substantial bias difference is observed among the five estimators.

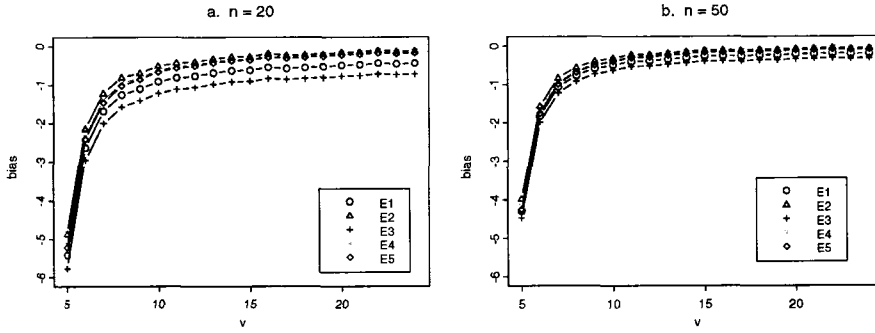


Figure 2.4: Bias of the kurtosis estimators for Student-t distributions. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively.

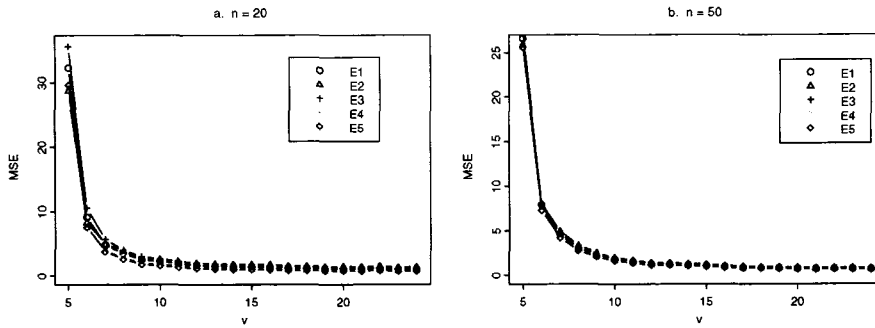


Figure 2.5: MSE of the kurtosis estimators for Student-t distributions. ν is the degrees of freedom which ranges from 5 to 30. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively.

Figures indicate that $\hat{\gamma}^U, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$ have the smallest bias, whereas $\hat{\gamma}$ and $\hat{\gamma}^M$ have the largest bias. Interestingly, this pattern is the same as that for normal data. Comparing figure 2.4 (a) and (b), we observe that for large degrees of freedom the amount of bias is negligible for both sample sizes.

Figure 2.5 presents a comparison on the $SMSE$ of the five estimators. It demonstrates that for a fixed sample size n , the $SMSE$ of all estimators are very large (around 30 compared to less than 2 in normal case) when the degrees of freedom is small, i.e., 5.

This is due to the large deviation of the student-t from normal distribution when v is small, which causes huge negative bias (around -5). It is not surprising that $\hat{\gamma}^U$ has the smallest MSE followed by $\hat{\gamma}^{N2}$ and $\hat{\gamma}^{N1}$ when v is small because of their relatively small bias compared to $\hat{\gamma}^M$ and $\hat{\gamma}$. The $SMSE$ of all estimators drops dramatically when v increases; interestingly, the order of the five $MSEs$ changes in favor of our estimators. For example, the following relation is observed for both sample sizes 20 and 50 when $v > 6$:

$$SMSE(\hat{\gamma}^{N2}) < SMSE(\hat{\gamma}^{N1}) < SMSE(\hat{\gamma}^U) < SMSE(\hat{\gamma}) < SMSE(\hat{\gamma}^M). \quad (2.44)$$

Hence, $\hat{\gamma}^{N2}$ and $\hat{\gamma}^{N1}$ dominate the existing estimators for the said condition. i.e., when v is not too small. Generally speaking, there is no clear cut winner in this case. Incidentally, for a large sample, say 50 and so the performance of all the estimators is similar.

Skewed Data To simulate skewed data, 5000 samples of sizes 20 and 50 are randomly taken from χ^2 distributions with degrees of freedom 1 to 40 in each case. Figures 2.6 and 2.7 compare the SB and $SMSE$ of the five estimators based on sample sizes 20 and 50, respectively. The graphical analysis is similar to that of Student-t distribution. Figure 2.6 shows that all five estimators underestimate the population kurtosis in the entire parameter space. The following relation in terms of bias is observed.

$$SB(\hat{\gamma}^U) < SB(\hat{\gamma}^{N1}) < SB(\hat{\gamma}^{N2}) < SB(\hat{\gamma}) < SB(\hat{\gamma}^M). \quad (2.45)$$

Seemingly, the magnitude of bias is tremendous when the data is highly skewed. However, it reduces rapidly as the degree of freedom v increases. The graph shows

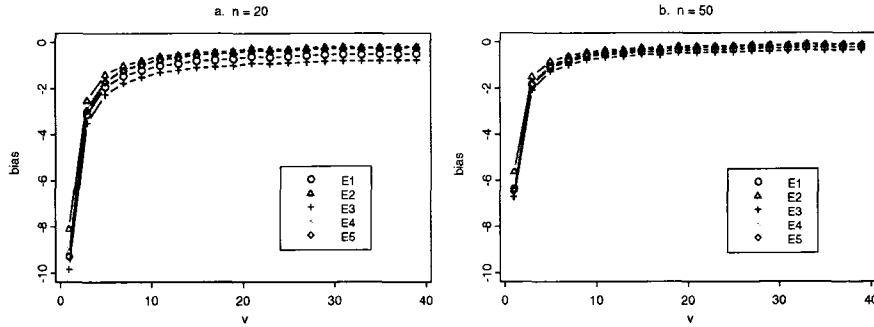


Figure 2.6: Bias of the kurtosis estimators for χ^2 distributions. The degrees of freedom ν ranges from 1 to 40. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively.

that larger sample size improves the estimation, yet, huge bias is observed when v is small.

Figure 2.7 shows the behavior of the $SMSE$ of the estimators. For small v ,

$$SMSE(\hat{\gamma}^U) < SMSE(\hat{\gamma}^{N1}) < SMSE(\hat{\gamma}^{N2}) < SMSE(\hat{\gamma}) < SMSE(\hat{\gamma}^M). \quad (2.46)$$

On the other hand, as v increases, the domination of bias diminishes, resulting in the change of the order in the above relation. For example, when $n = 20$, the following relation is observed for $v > 10$:

$$SMSE(\hat{\gamma}^{N2}) < SMSE(\hat{\gamma}^{N1}) < SMSE(\hat{\gamma}^M) < SMSE(\hat{\gamma}) < SMSE(\hat{\gamma}^U) \quad (2.47)$$

A similar but slower change for larger sample size is detected as well. Again, the proposed estimators $\hat{\gamma}^{N2}$ and $\hat{\gamma}^{N1}$ outperform the existing estimators under reasonable conditions. However, no one can claim to be the winner for a general case.

In summary, based on the bias and MSE behavior of the estimators, the proposed estimators $\hat{\gamma}^{N1}$ and $\hat{\gamma}^{N2}$ outshine the other estimators for normal as well as non-normal populations in many situations. The simulation study reveals that in general,

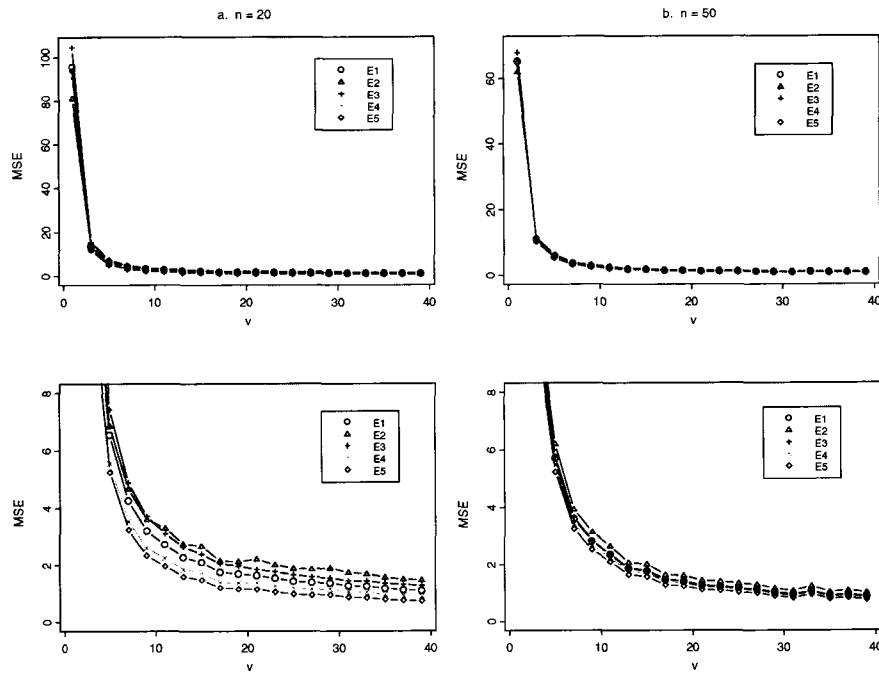


Figure 2.7: MSE of the kurtosis estimators for χ^2 distributions. The bottom two plots use enlarged vertical scale in order to have a better view when the degrees of freedom ν is large. $E_1, E_2, E_3, E_4,$ and E_5 represent $\hat{\gamma}, \hat{\gamma}^U, \hat{\gamma}^M, \hat{\gamma}^{N1},$ and $\hat{\gamma}^{N2}$, respectively.

our estimators perform better than the existing ones adapted by various software packages such as SAS, S-Plus, SPSS, and MINITAB in terms of bias and *MSE*. Noting that all estimators are biased for non-normal populations and the bias and *MSE* are substantially high when the distribution is far away from normal, a bias correction for these estimators is necessary in order to achieve a better performance and hence better inferences.

An Empirical Bias-corrected Estimator for Non-normal Distributions

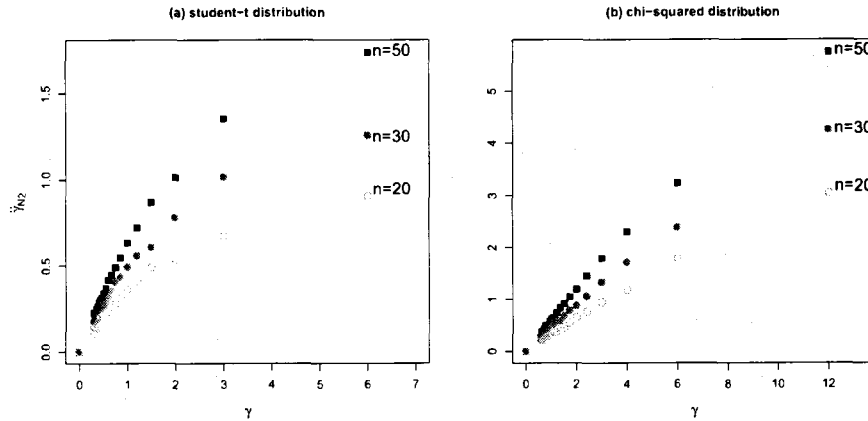
Noticing that all five estimators are biased for non-normal populations, and bias is inflated in a range of the parameter space, it seems to be an appealing idea to construct a bias-corrected estimator. Since the variance is negligible compared to *Bias*², we suggest to employ a bias-reduction technique based on our proposed best performing estimator $\hat{\gamma}^{N^2}$.

Recall that, for *t* and chi-squared distributions, the bias depends on the sample size as well as the degrees of freedom. Examining the scatter plots in Figure 2.8(a) and (b), with γ as *X* axis and $mean(\hat{\gamma}^{N^2})$ as *Y* axis respectively, we see smooth quadratic curves for both distributions. Hence, a new bias-corrected estimator may be constructed in the following form:

$$\hat{\gamma}^{N^3} = \hat{\beta}_1 \hat{\gamma}^{N^2} + \hat{\beta}_2 (\hat{\gamma}^{N^2})^2. \quad (2.48)$$

The values of the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ for $n = 20, 30,$ and 50 are computed and listed in Table 2.2.

Based on the bias-correction in equation (2.48) and Table 2.2, a simulation experiment is conducted to inspect the bias and *MSE* of the estimators. The result

Figure 2.8: $\hat{\gamma}^{N^2}$ vs γ scatter plot

shows that these estimators effectively reduced the bias to a negligible level; however, extremely large variance was introduced due to the quadratic form, resulting in inflated MSE . A similar correction was applied to the other four estimators, and the performance of these bias-corrected estimators were similar. This form of correction is not ideal.

Since the uncorrected estimators perform reasonably well when the degrees of freedom is not very small (say ≥ 10 for t distribution and ≥ 5 for χ^2 distribution), it might be more practical to consider a correction only when the degrees of freedom is very small (i.e., < 10 for t distribution and < 5 for χ^2 distribution). Note that in Figure 2.8, when the degrees of freedom is small (corresponding to large kurtosis), the relationship between $\hat{\gamma}$ and γ can be well approximated by a linear equation; thus, a simple linear regression model without independent variable $(\hat{\gamma}^{N^2})^2$ may be fitted. Our simulation found that the variance of this fitted estimator is greater than the original biased estimator, however it is not inflated too much. The result is not shown

Table 2.2: The coefficients in $\hat{\gamma}^{N3}$

n	t-distribution		χ^2 -distribution	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
20	0.6453	6.2886	2.8691	0.3435
30	0.6290	2.9618	2.0371	0.1852
50	0.6430	1.4809	1.5310	0.0968

here.

For the mixture of two normal distributions, the bias of the estimation depends on the shift $(\mu_1 - \mu_2)$, the scale σ_1/σ_2 , the weights p_1 and p_2 , as well as the sample size n . It is very difficult to obtain a general equation to correct the bias. Luckily, such a correction is not needed in many situations, which will be discussed in the next section.

2.7.4 An Application in the Estimation of Coefficient of Variation

The coefficient of variation $\theta = \frac{\sigma}{\mu}$ is a descriptive measure of relative dispersion that can be found in virtually all introductory statistics text books. It is commonly used in medical and biological sciences as a very useful measure of relative variability of the data. A commonly used consistent estimator of θ is defined by $\hat{\theta} = \frac{s}{\bar{x}}$, where s is the usual sample standard deviation and \bar{x} is the sample mean. Ahmed (2002) showed that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0, \tau^2), \quad \text{where} \quad \tau^2 = \theta^4 - \theta^3 \frac{\mu_3}{\sigma^3} + \frac{1}{4} \theta^2 (\gamma + 2). \quad (2.49)$$

It is seen that the variance parameter τ^2 is not *stable* in the sense that it is a function of θ and γ . A variance stabilizing transformation can be obtained to get rid of θ ;

however, it will be still a function of γ . To this end, a consistent estimator of τ^2 is given by

$$\hat{\tau}^2 = \hat{\theta}^4 - \hat{\theta}^3 \frac{\hat{\mu}_3}{\hat{\sigma}^3} + \frac{1}{4} \hat{\theta}^2 (\hat{\gamma} + 2),$$

where $\hat{\mu}_3$, $\hat{\sigma}$, and $\hat{\gamma}$ are suitable estimators of μ_3 , σ , and γ respectively. Hence, an asymptotic $(1 - \alpha)100\%$ confidence interval for θ can be calculated by

$$(\hat{\theta} - z_{\alpha/2} \sqrt{\hat{\tau}^2/n}, \hat{\theta} + z_{\alpha/2} \sqrt{\hat{\tau}^2/n}), \quad (2.50)$$

where, $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution.

Illustrative Examples

Example 1: Point Estimation (large sample) The Hong Kong Medical Technology Association has conducted the Quality Assurance Programme for medical laboratories in Hong Kong since 1989. See Fung and Tsing (1998). The aim of the programme is to promote the quality and standards of medical laboratory technology. Coefficient of variation is commonly presented in the Annual Report of the programme. In the specialty of haematology and serology, two whole blood samples (one normal and the other abnormal) were sent to participants for measurement of Hb, RBC, MCV, Hct, WBC and Platelet in each survey. The data collected from the third survey of 1996 are

14.0, 14.0, 14.2, 14.2, 14.3, 14.3, 14.3, 14.4, 14.5, 14.5, 14.5, 14.5, 14.5, 14.6, 14.6,
 14.6, 14.6, 14.6, 14.6, 14.6, 14.7, 14.7, 14.7, 14.7, 14.7, 14.7, 14.7, 14.7, 14.7,
 14.7, 14.7, 14.7, 14.7, 14.7, 14.8, 14.8, 14.8, 14.8, 14.8, 14.8, 14.8, 14.8, 14.8,
 14.8, 14.9, 14.9, 14.9, 14.9, 14.9, 14.9, 14.9, 14.9, 14.9, 14.9, 14.9, 15.0, 15.0, 15.0,
 15.0, 15.0, 15.0, 15.0, 15.0, 15.0, 15.0, 15.0, 15.1, 15.1, 15.3, 17.3

For this sample data, $\hat{\theta} = 0.0265$, and

$$\hat{\gamma} = 22.29, \quad \hat{\gamma}^U = 23.99, \quad \hat{\gamma}^M = 21.61, \quad \hat{\gamma}^{N1} = 22.37, \quad \hat{\gamma}^{N2} = 21.77.$$

Consequentially, the corresponding asymptotic variance estimates are:

$$(\hat{\tau}_1^2, \hat{\tau}_2^2, \hat{\tau}_3^2, \hat{\tau}_4^2, \hat{\tau}_5^2) = (0.004200, 0.004497, 0.004079, 0.004214, 0.004107).$$

Seemingly, $\hat{\gamma}^M$ gives the smallest variance estimation. However, the actual τ^2 is unknown; hence, a conclusion cannot be drawn as to which estimate is better. Nevertheless, we observed in our simulation study in section 2.7.3 that all five estimators seem to underestimate the true kurtosis parameter.

Clearly, the data came from a non-normal population, perhaps due to an outlier. Fung and Tsing (1998) pointed out that 17.3 is an outlier in this data set. Recalculating the above statistics after removing this outlier, the following results are obtained:

$$\hat{\theta} = 0.01717 \quad \text{and}$$

$$\hat{\gamma} = 0.8131, \quad \hat{\gamma}^U = 0.9607, \quad \hat{\gamma}^M = 0.7079, \quad \hat{\gamma}^{N1} = 0.8953, \quad \hat{\gamma}^{N2} = 0.8706.$$

The small values of the kurtosis estimates indicate that the data resemble normal distribution fairly well after the removal of the outlier. Further, the variance estimates are:

$$(\hat{\tau}_1^2, \hat{\tau}_2^2, \hat{\tau}_3^2, \hat{\tau}_4^2, \hat{\tau}_5^2) = (.0002116, .0002225, .0002038, .0002176, .0002158).$$

In an effort to get a clear picture for the relative performance of various estimators in this example, the bias and the standard error (SE) of $\hat{\tau}_1^2$ to $\hat{\tau}_5^2$ are computed using

Table 2.3: Performance of $\hat{\tau}$ based on different $\hat{\gamma}$

estimator	full data set			outlier removed		
	bias($\times 10^{-3}$)	SE($\times 10^{-3}$)	RE	bias($\times 10^{-6}$)	SE($\times 10^{-5}$)	RE
$\hat{\tau}_1^2$	-1.635	1.779	0.986	-9.583	4.773	1.107
$\hat{\tau}_2^2$	-1.455	1.907	1	-0.598	5.122	1
$\hat{\tau}_3^2$	-1.711	1.7317	0.9727	-17.077	4.642	1.073
$\hat{\tau}_4^2$	-1.620	1.779	0.995	-3.525	4.773	1.145
$\hat{\tau}_5^2$	-1.682	1.731	0.988	-5.081	4.642	1.200

a bootstrap technique. Furthermore, the efficiency of the estimators relative to $\hat{\tau}_2^2$ is estimated accordingly. The result is listed in Table 2.3.

Table 2.3 reveals that the bias and SE are relatively small without the outlier as compared to that of the full data set. This makes sense since the data approximately follows a normal distribution after removing the outlier. However, the data is highly skewed when the outlier exists. Thus, the bias of the estimates is inflated (from 10^{-6} to 10^{-3}) as a result of the deviation from normality. The standard error of the variance estimates is exploded as well due to the outlier. Inspecting the relative efficiencies, we find that the variance estimates $\hat{\tau}_4^2$ and $\hat{\tau}_5^2$ based on our kurtosis estimators $\hat{\gamma}^{N1}$ and $\hat{\gamma}^{N2}$ perform the best when the outlier is removed. In contrast, in the presence of an outlier, the performance of all the estimators is comparable with $\hat{\tau}_2^2$, $\hat{\tau}_4^2$ and $\hat{\tau}_5^2$ slightly better than the other two. Keeping in mind that $n = 73$ is a reasonably large sample size, as a result, the difference among the performance of the estimators is not very significant. More importantly, these observations are consistent with our theoretical and simulation studies.

Example 2: Interval Estimation (small sample) Ott and Longnecker describe a

study where percent potency reduction is measured in a random sample of 20 pesticide containers that have been stored at room temperature for six months, see Douglas (2006). The sample data are 0.2, 0.5, 1.1, 1.4, 1.8, 2.3, 2.5, 2.7, 3.5, 4.4, 4.6, 5.4, 5.4, 5.7, 5.8, 5.9, 6.0, 6.6, 7.1, 7.9. Here

$$\hat{\theta} = s/\bar{x} = 0.5777, \quad \hat{\gamma}^{N1} = -0.9511, \quad \hat{\gamma}^{N2} = -0.8584.$$

Consequently, $\hat{\tau}^2(N1) = 0.2279$ and $\hat{\tau}^2(N2) = 0.2356$. Thus, 95% confidence intervals for θ based on $\hat{\gamma}^{N1}$ and $\hat{\gamma}^{N2}$ are (0.3685, 0.7869) and (0.3650, 0.7904) respectively. A 95% bootstrap- t interval and a bootstrap percentile interval were also computed based on 500×500 re-samples and 5000 re-samples, respectively. The bootstrap confidence intervals are (0.3205, 0.7560) and (0.3823, 0.7842), respectively. These intervals are comparable and all shorter than the one given in Douglas (2006). However, Douglas (2006) is estimating the coefficient of quartile variation by removing 50% of the observations from the original data set.

We summarize this section as follows:

We have compared the performance of several kurtosis measures adapted by SAS, SPSS, S-Plus, Minitab, and other statistical packages. We have proposed several new measures of kurtosis. It has been demonstrated both analytically and numerically that our proposed estimators outperform the existing estimators for normal population based on the MSE criterion. Moreover, an extensive simulation study has been conducted for non-normal populations. Results indicate that the proposed estimators are superior to the existing ones in many practical situations. Bearing this in mind, all the estimators substantially underestimate kurtosis parameter when under-

lying population distribution is highly skewed or heavy tailed. In order to correct the bias, empirical formulas are provided for Student-t and Chi-squared distributions. However, empirical estimates are subject to extra variation introduced and result in inflated *MSE*. Perhaps, some re-sampling methods such as bootstrap and jackknife may be considered to reduce the bias as well as keeping a relatively lower variance. Example 2 showed that all estimators are very sensitive to outliers. Some non-parametric estimators that are more robust may be developed.

2.8 Simulation Studies for Variances

Now we return to the main problem of this chapter and provide some simulation results for variances.

The main purpose of this simulation is to examine the quality of statistical inferences based on large-sample methodology in moderate sample situations under various scenarios. The developed theoretical aspects of the test statistic and interval estimation in previous sections are now examined through the Monte Carlo simulation. In the simulation study, we consider some of the frequently encountered models in practice. The simulation study is performed using normal (symmetric), Student's *t* (heavy tailed), chi squared (skewed), and mixture of normal (contaminated) distributions. Representatives of skewed and heavy tailed distributions were chosen because they frequently occur in practice and are particularly troublesome. A mixture of normal distributions is studied since it has provided an extremely flexible method of modelling a wide variety of random phenomena in economics, sociology, and medical science, and has received increasing attention in many fields. Further, these specific

distributions were considered for their ease of generation and mathematical manipulation.

All simulations were run in S-plus.

2.8.1 Simulation Study 1: Test Statistics

Distribution of the test statistics

First, a simulation study was carried out to assess the distribution of the test statistics for testing the homogeneity of several variances. We consider the test statistics T_2 and T_4 given in relations (2.8) and (2.13), respectively. For the given sample sizes and fixed k , random samples were generated from a normal, a Student's t , a chi-squared and a mixture of two normal distributions. In each case, 5000 samples were randomly generated. Normal random samples were generated with mean 0 and standard deviation 1, Student t random samples were drawn with degrees of freedom 6, the chi-squared distributed samples were obtained for 4 degrees of freedom, and the mixture of normal distributions was composed of 90% $N(0, 1^2)$ and 10% $N(0, 2^2)$.

In each case, both test statistics T_2 and T_4 were computed based on 5000 replications. The results are presented in Figures 2.9 and 2.10. For the Student- t and chi-squared samples, the test statistics were computed based on two different kurtosis estimators: one is $\hat{\gamma}^U$ as defined in equation (2.27), and the other is based on the bias corrected estimator given in Table 2.2 (labeled with "BC" in the graphs). Figure 2.9 compares the distribution of the test statistic T_2 which is computed from original data with the theoretical asymptotic distribution $\chi_{(k-1)}^2$, and figure 2.10 compares that of T_4 which is computed from transformed data with $\chi_{(k-1)}^2$. These figures are

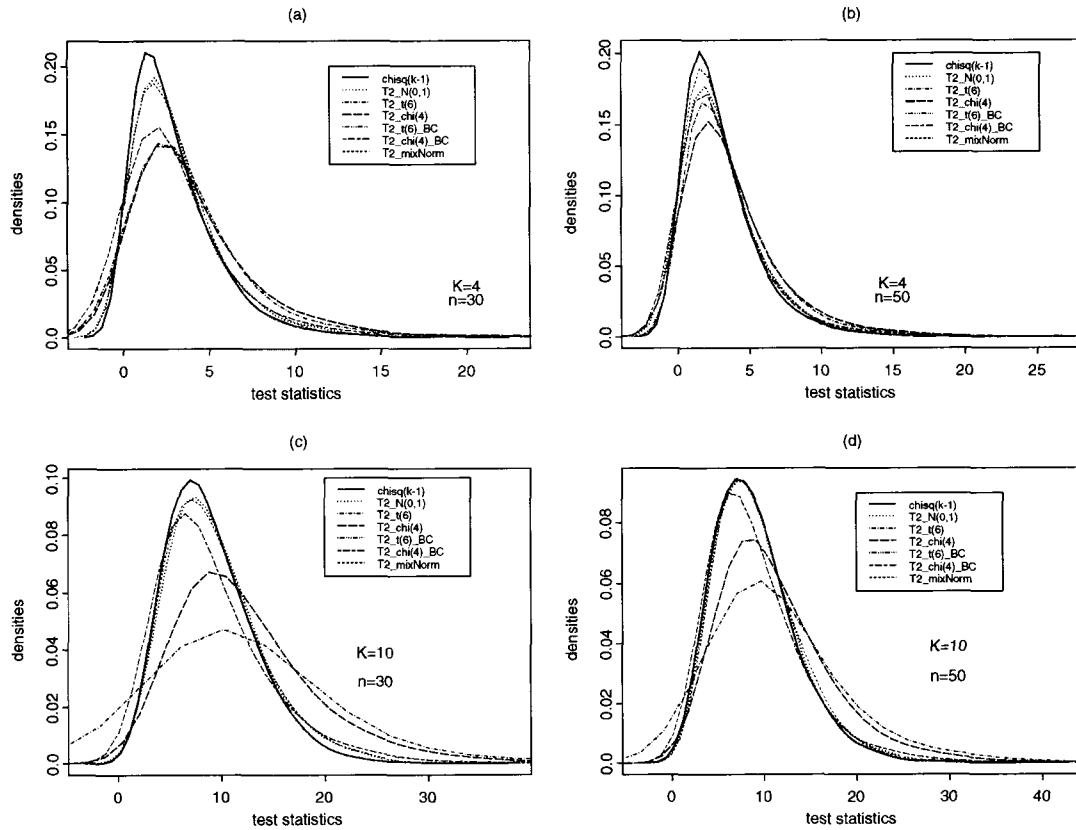


Figure 2.9: Distribution of the test statistic T_2 (original data). The test statistics with “BC” were calculated based on bias corrected estimate of kurtosis in table 2.2, and the rest were based on $\hat{\gamma}^U$.

based on four different population distributions, two different estimators of kurtosis and different combinations of k and n . In the figures, a solid line represents the theoretical asymptotic distribution of the test statistic, i.e., $\chi^2_{(k-1)}$; dotted and dashed lines portray the simulated sampling distributions of the test statistic T_2 or T_4 based on the samples of size 30 or 50, as marked in the graphs, from $N(0, 1)$, $t(6)$, χ^2_4 and the mixture of 90% $N(0, 1^2)$ and 10% $N(0, 2^2)$ data, respectively.

The graphical analysis reveals that the distributions of the test statistics T_2 and T_4

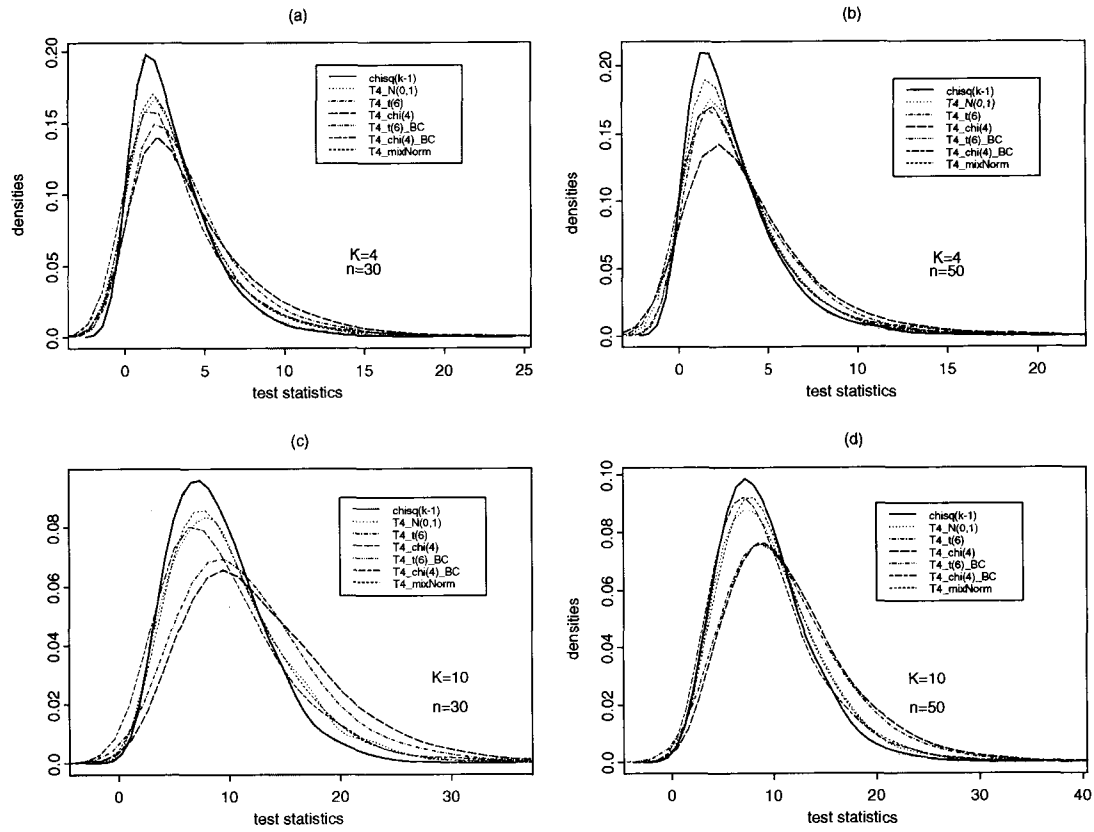


Figure 2.10: Distribution of the test statistic T_4 (transformed data). The test statistics with “BC” were calculated based on bias corrected estimate of kurtosis in table 2.2, and the rest were based on $\hat{\gamma}^U$.

are reasonably approximated by a chi-squared distribution with a respective degree of freedom for a moderate sample size n and not too large number k , of samples for all four population distributions. For example, when $k = 4$, the distribution of the right tails, which is our major concern, of the T_2 and T_4 resemble $\chi^2_{(k-1)}$ very well for n as small as 30. When k increases, T_2 and T_4 deviate from $\chi^2_{(k-1)}$ significantly, especially for $t(6)$ and $\chi^2(4)$. A larger sample size is required in order to get a better approximation. The simulation shows that in general, T_2 and T_4 resemble χ^2 very well for normal and mixture of normal populations even for k as large as 10 and n as small as 30. The bias correction to the kurtosis estimator significantly improves the χ^2 approximation for t and χ^2 population, especially on the right tails, for both original data and transformed data and for all combinations of k and n . Interestingly, T_2 and T_4 resemble $\chi^2_{(k-1)}$ surprisingly well for the mixture of normal distributions even though $\hat{\gamma}^U$ is used to estimate γ without any bias correction. It leads us to conclude that a bias correction is not necessary for the mixture of normal case in our procedure. The figures also demonstrate that the effect of the variance stabilizing transformation is not very significant in this case.

Power Simulation

A simulation to study the power of the above tests was carried out as well. We considered the case when $k=4$ for each of normal, Student-t, chi-squared and mixture of two normal populations. The null hypothesis of interest is

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

In order to simulate the power, let

$$\Delta = \frac{\sum(\sigma_i^2 - \sigma_{(R)}^2)^2/k}{\sigma_{(R)}^2}$$

where $\sigma_{(R)}^2$ is the mean of σ_i^2 . Δ serves as a measure of deviation of the true status of the parameter from the null hypothesis. In the simulation, random samples were taken from populations with different variances to vary the value of Δ . Powers were calculated for different Δ and selected sample sizes 20 and 50. In order to make a comparison of our tests with the established methods, we included Levene's test, the most popularly used test for homogeneity of variance when the populations are not normal, in our simulation as well.

Levene's test (Levene (1960)) is used to test the homogeneity of variance when normality is violated. Let y_{ij} be the j^{th} measurement in the i^{th} group, where $i = 1, \dots, k$ and $j = 1, \dots, n_i$. The Levene test statistic is defined as:

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^k n_i (\bar{z}_i - \bar{z}_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2}$$

where N is the total sample size, z_{ij} can have one of the following three definitions:

1. $z_{ij} = |y_{ij} - \bar{y}_i|$, where \bar{y}_i is the mean of the i th group.
2. $z_{ij} = |y_{ij} - \tilde{y}_i|$, where \tilde{y}_i is the median of the i th group.
3. $z_{ij} = |y_{ij} - \bar{y}'_i|$, where \bar{y}'_i is the 10% trimmed mean of the i th group.

\bar{z}_i are the group means of the z_{ij} and $\bar{z}_{..}$ is the overall mean of the z_{ij} .

The three choices for defining z_{ij} determine the robustness and power of Levene's test. Levene's original paper only proposed using the mean. Brown and Forsythe (1974) extended the Levene's test to use either the median or the trimmed mean in addition to the mean. Simulation studies indicated that using trimmed mean

Table 2.4: Power simulation of the tests based on T_2 and T_4 for normal samples ($\alpha = 0.05$)

Δ	T_2		T_4		Levene test	
	n=20	n=50	n=20	n=50	n=20	n=50
0	0.055	0.063	0.072	0.069	0.053	0.054
0.333	0.402	0.822	0.478	0.845	0.332	0.767
0.500	0.788	0.995	0.848	0.996	0.719	0.988
0.600	0.940	1	0.962	1	0.907	1
0.778	0.999	1	1	1	0.998	1

performed best when the underlying data followed a Cauchy distribution and the median performed the best when the underlying data followed χ_4^2 distribution. Using the mean provided the best power for symmetric, moderate-tailed distributions.

The Levene's test rejects the hypothesis that the variances are equal if

$$W > F_{(\alpha, k-1, N-k)}$$

where $F_{(\alpha, k-1, N-k)}$ is the upper critical value of the F distribution with $k - 1$ and $N - k$ degrees of freedom at a significant level of α .

The results of the power simulation are presented in Tables 2.4 to 2.7. The kurtosis estimate is based on $\hat{\gamma}^{(U)}$ unless otherwise stated.

Table 2.4 shows the power of the three tests when the population distribution is normal. Keeping the level of significance at 0.05, we calculated the powers of the three tests, T_2 , T_4 , and W , for sample sizes $n=20$ and 50 when $\Delta = 0, 0.333, 0.5, 0.6$, and 0.778 respectively. The simulation shows that the test T_4 based on transformed data provides the highest power among the three when the alternative hypothesis is true, however, with a cost of higher probability of type one error (0.072 and 0.069)

Table 2.5: Power simulation of the tests based on T_2 and T_4 for Student-t samples ($\alpha = 0.05$)

Δ	T_2		T_4		Levene test	
	n=20	n=50	n=20	n=50	n=20	n=50
0	.072(.055)	.063(.047)	.083(.066)	.074(.058)	.055	.059
0.143	.148(.072)	.171(.089)	.180(.092)	.182(.103)	.073	.085
0.333	.304(.138)	0.523(.315)	.344(.158)	.546(.324)	.127	.262
0.500	.459(.178)	.786(.390)	.508(.205)	.812(.424)	.195	.442

The numbers in parentheses are based on bias corrected kurtosis estimates in Table 2.2.

Table 2.6: Power simulation of the tests based on T_2 and T_4 for χ^2 samples ($\alpha = 0.05$)

Δ	T_2		T_4		Levene test	
	n=20	n=50	n=20	n=50	n=20	n=50
0	.098(.061)	.100(.066)	.121(.076)	.121(.077)	.041	.045
0.333	.308(.118)	.459(.326)	.476(.235)	.544(.414)	.248	.624
0.500	.450(.196)	.808(.672)	.662(.393)	.875(.777)	.559	.971
0.600	.592(.316)	.934(.851)	.787(.627)	.966(.920)	.776	.999
0.778	.863(.500)	.984(.904)	.963(.838)	.998(.982)	.986	1

The numbers in parentheses are based on bias corrected kurtosis estimates in Table 2.2.

compared to the nominal value 0.05. Therefore, we might have to correct the α level and recalculate the corrected power. On the other hand, T_2 and Levene's tests behave reasonably well; T_2 has a slightly higher power than Levene's when the alternative is true. It is also observed that the power of the tests increases as sample size n and the deviation, Δ , from the null hypothesis increases.

The case of t-distribution is shown in table 2.5. It can be seen that bias correction improves the behavior of the test statistics T_2 and T_4 when the null hypothesis is true. Both of our tests have a higher power than Levene's when the population distribution is t.

Table 2.6 shows the result when the population distribution is chi-squared (skewed).

As for t distribution, we can see that the bias correction improves the behavior of

Table 2.7: Power simulation of the tests based on T_2 and T_4 for samples from a mixture of two normal populations ($\alpha = 0.05$)

Δ	T_2		T_4		Levene test	
	n=20	n=50	n=20	n=50	n=20	n=50
0	0.101(0.055)	0.081(0.046)	0.121(0.070)	0.088(0.051)	0.059	0.051
0.333	0.346(0.258)	0.634(0.531)	0.435(0.346)	0.684(0.600)	0.256	0.602
0.500	0.614(0.517)	0.937(0.893)	0.712(0.631)	0.956(0.926)	0.480	0.911
0.600	0.767(0.667)	0.989(0.977)	0.856(0.788)	0.994(0.988)	0.638	0.981
0.778	0.949(0.908)	1(1)	0.984(0.971)	1(1)	0.891	1

The numbers in parentheses are based on bias corrected kurtosis estimates in Table 2.2.

the test statistics T_2 and T_4 when the null hypothesis is true, but with lower power when the alternative is true. For this skewed distribution, Levene's test outperforms both of T_2 and T_4 .

The power comparison of the three test statistics for the mixture of two normal populations is displayed in Table 2.7. Previous simulation shows that the right tail distribution of T_4 and T_2 resembles $\chi_{(k)}^2$ better than $\chi_{(k-1)}^2$ under the null hypothesis. Therefore, a power simulation using both $\chi_{(k,\alpha)}^2$ and $\chi_{(k-1,\alpha)}^2$ as critical values was conducted. The numbers in parentheses are based on $\chi_{(k,\alpha)}^2$. Simulation shows that the power of T_4 increases the fastest, followed by T_2 in the second place, and Levene's test with the lowest power for mixture of normal population. It can be observed that our tests are much more powerful than Levene's test when sample size n is small. When $\chi_{(k)}^2$ is applied, the size of the tests T_2 and T_4 are closer to the nominal α level 0.05 than that when $\chi_{(k-1)}^2$ is used. The power of T_2 and T_4 are slightly lowered when using k degrees of freedom, still significantly higher than Levene's test. As sample size n increases, the power of all three tests increases, and the difference among the three tests becomes smaller. The kurtosis estimation here is based on $\hat{\gamma}^U$. The power based

on the theoretical kurtosis was also simulated; no difference was observed between using $\hat{\gamma}^U$ and γ , which confirms our previous conclusion that no bias correction is needed for our purpose.

2.8.2 Simulation Study 2: Interval Estimation

In this section, we present the result of our simulation study conducted to assess the performance of our interval estimation procedures proposed in Section 2.6. The coverage probability and the average mean width of the interval are reported using the single sample and combined k-sample data. The simulated intervals were computed in two situations. In one case, the true value of kurtosis was used and in the other situation, the kurtosis was estimated from the simulated data. All the kurtosis estimation is based on $\hat{\gamma}^U$ defined in 2.27 unless otherwise stated. The percentage of simulated confidence intervals that contained the true index value was determined. In addition, an average length of simulated intervals was calculated based on 5000 replications. The simulation study was performed on four different distributions: normal, student t , chi-squared and the mixture of two normal distributions.

For the intervals based on combined data we considered only the case for $k = 4$ with equal sample sizes. The observed coverage probability and the mean width of the generated confidence intervals using relations (2.16), (2.18), (2.19), and (2.21) are presented in Tables 2.8 to 2.13.

Tables 2.8 and 2.9 record the coverage probability and average width for the single and combined data, respectively, sampled from a normal population.

By examining the values in Table 2.8, it can be stated that the empirical coverage

Table 2.8: Coverage probability and mean width for uncombined normal sample data

		Without Transformation		With Log Transformation	
Normal Distribution	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
when true value of γ^* is used	100	0.94	1.11	0.94	1.12
	70	0.93	1.32	0.94	1.34
	50	0.93	1.56	0.94	1.60
	30	0.92	2.04	0.93	2.12
	20	0.89	2.49	0.93	2.63
when γ^* is estimated	100	0.93	1.10	0.93	1.12
	70	0.92	1.31	0.93	1.34
	50	0.92	1.55	0.93	1.59
	30	0.90	1.99	0.92	2.08
	20	0.88	2.43	0.90	2.56

probability is not very far from the nominal value (0.95) for a moderate sample size (≥ 70). We observe that the variance stabilizing transformation dramatically improves the performance of the confidence intervals, especially when the samples are small. As expected from the theory, the average width of the simulated confidence intervals decreases as n increases for all intervals (tending to validate the simulation).

The combined data provides a better result, as was expected. As seen in Table 2.9, the combined data produces shorter confidence intervals and more accurate coverage probability than the uncombined one does. The average mean width is reduced by 50% as compared to that of uncombined data. The effect of the transformation is visible, even though not as significant as for uncombined data.

For t distribution with 6 degrees of freedom, the simulation was also conducted at various sample sizes and the results are presented in Tables 2.10 and 2.11. The numbers in parentheses are based on bias corrected estimate of kurtosis in Table 2.2.

Table 2.9: Coverage probability and mean width for combined normal samples data

		Without Transformation		With Log Transformation	
Normal Distribution	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
When true value of γ^* is used	4 × 100	0.95	0.55	0.95	0.55
	4 × 70	0.94	0.66	0.95	0.66
	4 × 50	0.94	0.78	0.95	0.78
	4 × 30	0.95	1.02	0.95	1.02
	4 × 20	0.93	1.24	0.95	1.26
when γ^* is estimated	4 × 100	0.95	0.55	0.95	0.55
	4 × 70	0.95	0.66	0.95	0.66
	4 × 50	0.94	0.78	0.95	0.78
	4 × 30	0.93	1.01	0.94	1.01
	4 × 20	0.92	1.22	0.94	1.25

Table 2.10: Coverage probability and mean width based on uncombined sample data from t_6 distribution

		Without Transformation		With Log Transformation	
student-t (df=6)	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
when true value of γ^* is used	100	0.95	1.32	0.96	1.35
	70	0.95	1.59	0.97	1.65
	50	0.93	1.87	0.96	1.97
	30	0.93	2.40	0.96	2.67
	20	0.91	2.94	0.96	3.47
when γ^* is estimated	100	0.88(.90)	1.15(1.24)	.90(.91)	1.19(1.28)
	70	0.88(.89)	1.33(1.49)	.90(.91)	1.41(1.57)
	50	0.86(.88)	1.53(1.60)	.89(.91)	1.65(1.88)
	30	0.84(.88)	1.93(2.54)	.86(.91)	2.12(2.63)
	20	0.82(.87)	2.28(3.27)	.86(.93)	2.57(3.56)

The higher probability "0.93" in the last row is probably due to sampling error.

Table 2.11: Coverage probability and mean width based on combined samples data from t_6 distribution

Student-t (df=6)	Sample Size	Without Transformation		With Log Transformation	
		Coverage Probability	Mean Width	Coverage Probability	Mean Width
when true value of γ^* is used	4×100	0.95	0.65	0.96	0.66
	4×70	0.95	0.78	0.96	0.79
	4×50	0.95	0.93	0.96	0.94
	4×30	0.95	1.20	0.97	1.23
	4×20	0.94	1.47	0.96	1.52
when γ^* is estimated	4×100	.92(.93)	0.57(0.65)	.92(.94)	0.58(0.66)
	4×70	.91(.92)	0.67(0.78)	.92(.94)	0.67(0.79)
	4×50	.90(.92)	0.77(0.93)	.91(.93)	0.78(0.95)
	4×30	.88(.91)	0.95(1.19)	.90(.93)	0.97(1.24)
	4×20	.87(.90)	1.13(1.48)	.87(.92)	1.16(1.53)

The numbers in parentheses are based on bias corrected kurtosis estimates in Table 2.2.

It is seen that the empirical coverage probability is reasonably close to the nominal value (0.95) for a sample size as small as 20 when the true values of the kurtosis are used. However, when the kurtosis is estimated from sample data, properly applying the proposed interval estimation methodologies requires a much bigger sample size, especially for uncombined data. The result shows that combining the data, applying variance stabilizing transformation, and correcting the bias all significantly improve the estimation. Further, confidence intervals based on the combined data produce shorter confidence intervals than that of unconstrained intervals. The average mean width is reduced by 50%.

Now, we consider a skewed distribution: a chi-squared distribution for simulation purposes. Tables 2.12 and 2.13 record the proportion of the coverage and average width of confidence intervals, respectively, based on single sample and combined data for a chi squared process with degrees of freedom 4. The numbers in parenthesis

Table 2.12: Coverage probability and mean width based on single sample data from $\chi^2_{(4)}$ distribution

		Without Transformation		With Log Transformation	
Chi Squared (df=4)	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
when true value of γ^* is used	100	0.94	7.02	0.95	7.19
	70	0.93	8.38	0.95	8.67
	50	0.92	9.81	0.96	10.5
	30	0.90	12.7	0.96	14.2
	20	0.88	15.6	0.96	18.2
when γ^* is estimated	100	.89(.90)	6.47(7.08)	.90(.92)	6.76(7.04)
	70	.87(.89)	7.79(8.14)	.90(.92)	7.97(9.63)
	50	.84(.85)	8.66(9.27)	.87(.89)	9.39(10.1)
	30	-	-	-	-
	20	-	-	-	-

The numbers in parentheses are based on bias corrected kurtosis estimates. The coverage probabilities in the last two rows are very low and omitted from the table.

were computed based on the bias-corrected estimate of the kurtosis. The effect of combining, transformation and bias correction is clearly shown in the tables. Not surprisingly, the coverage probability is closer to the nominal value when using the bias-corrected estimate of kurtosis than using the uncorrected estimate. Combining and transforming the data dramatically improve the performance of the confidence intervals. Note that desired nominal coverage level can be achieved even for samples of size 20 when the exact value of the kurtosis is used combined with transformation.

Finally, the mixture of two normal distributions is considered. The confidence intervals for the population variance were simulated using our large sample approximation in Section 2.6 for a mixture of 90% $N(0, 1^2)$ and 10% $N(0, 2^2)$ population and the results are shown in Tables 2.14 and 2.15. It is observed that when the true values of the kurtosis and variance stabilization transformation are used, the

Table 2.13: Coverage probability and mean width based on combined samples data from $\chi^2_{(4)}$ distribution

		Without Transformation		With Log Transformation	
Chi Squared (df=4)	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
when true value of γ^* is used	4×100	0.94	3.49	0.95	3.53
	4×70	0.94	4.16	0.95	4.24
	4×50	0.95	4.94	0.95	5.03
	4×30	0.94	6.40	0.95	6.61
	4×20	0.92	7.82	0.95	8.19
when γ^* is estimated	4×100	.93(.95)	3.24(3.70)	.93(.94)	3.28(3.51)
	4×70	.91(.93)	3.77(4.18)	.91(.94)	3.88(4.23)
	4×50	.90(.92)	4.40(4.91)	.91(.93)	4.53(5.08)
	4×30	.89(.91)	5.43(6.31)	.89(.91)	5.54 (6.46)
	4×20	.85(.88)	6.38(7.66)	.86(.89)	6.54 (7.98)

coverage probabilities are nearly identical to the nominal value 0.95 for both single sample and combined data for any sample sizes (as low as 20 in our study range). A bigger sample size is required to achieve such a good approximation without the log transformation. When the kurtosis is estimated, the approximation performs more poorly, especially when the data is not combined. Increasing sample size, combining samples, and transforming the data all considerably improve the performance of the approximation. However, a better estimation for population kurtosis seems more important here.

Note that for t and chi-squared distributions, the bias-correction notably improves the coverage probability, but at the expense of a wider confidence interval. This is due to the extra variation introduced by estimating the bias. The quadratic form in the bias correction is disadvantageous since it inflates the standard error. It is also

Table 2.14: Coverage probability and mean width based on uncombined sample data from a mixture of normal distribution

		Without Transformation		With Log Transformation	
Normal Mixture	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
when true value of γ^* is used	100	0.94	0.95	0.95	0.97
	70	0.94	1.134	0.95	1.16
	50	0.93	1.33	0.95	1.40
	30	0.92	1.72	0.95	1.86
	20	0.90	2.10	0.95	2.32
when γ^* is estimated	100	0.90	0.90	0.93	0.93
	70	0.90	1.07	0.91	1.09
	50	0.87	1.22	0.91	1.30
	30	0.85	1.54	0.89	1.67
	20	0.83	1.84	0.88	2.02

Table 2.15: Coverage probability and mean width based on combined data from a mixture of normal distribution

		Without Transformation		With Log Transformation	
Normal Mixture	Sample Size	Coverage Probability	Mean Width	Coverage Probability	Mean Width
When true value of γ^* is used	4×100	0.95	0.47	0.95	0.47
	4×70	0.94	0.56	0.95	0.57
	4×50	0.95	0.67	0.95	0.67
	4×30	0.94	0.86	0.95	0.88
	4×20	0.94	1.05	0.95	1.08
when γ^* is estimated	4×100	0.92	0.45	0.94	0.45
	4×70	0.92	0.53	0.92	0.53
	4×50	0.91	0.61	0.92	0.62
	4×30	0.90	0.77	0.91	0.78
	4×20	0.88	0.91	0.90	0.93

observed that the width of the intervals based on the estimate is consistently shorter than that based on the true kurtosis for the same sample size. This is due to the fact that the estimator has a negative bias which means the estimate of kurtosis is smaller than the true value in general. Meanwhile the width of the confidence interval depends on the standard error of the estimator, which is an increasing function of kurtosis (see Equation (2.2)). Thus, an underestimated kurtosis will result in a shorter confidence interval. The width of the intervals across four distributions is not comparable here since the variance parameters are different (2, 1.5, 8, and 1.3 respectively). The kurtosis exists for t distribution with at least 5 degrees of freedom, which produces the maximum variance $5/3$. In order to have a variance as low as this, the chi-squared distribution is required to have a maximum degree of freedom $5/6$, which is not very reasonable.

In this subsection, the large sample approach of interval estimation for a single population variance proposed in Section 2.6 is simulated for normal, Student t, chi-squared, and a mixture of two normal distributions. The study reveals that a good estimation of kurtosis is very important to improve the convergence of the probabilities to $(1 - \alpha)$, i.e., to achieve the nominal coverage probability. A variance stabilizing transformation significantly speeds up the convergence as well. Combining sample data when the null hypothesis is true improves the estimation due to the rising of the sample size. If a good estimation of kurtosis is achieved, our approach of interval estimation for population variance is very appealing when the population is not normal.

Finally, the simulation study provides some guidelines on the use of the proposed

Table 2.16: Four measurements on three species of Iris (in centimeters)

Iris setosa				Iris versicolor				Iris virginica			
Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
		⋮				⋮				⋮	

asymptotic method when parent populations are not normal. A numerical example based on the published Iris data set will be given to illustrate how the proposed procedures are applied in the next section.

2.9 Application to the Iris Data

The Iris data set (Anderson (2003)), partially shown in table 2.16, involves four measurements (Sepal length, sepal width, petal length, and petal width) on three species (Iris setosa, Iris versicolor, and Iris virginica) of iris. The sample size is 3×50 .

Testing Problem

Suppose we are interested in testing the homogeneity of the variance of sepal width for the three species. The hypotheses would be

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \quad \text{against} \quad H_a : \sigma_i^2 \neq \sigma_h^2 \quad \text{for at least one } i \neq h. \quad (2.51)$$

To test this null hypothesis, we may apply either T_2 in (2.8) or T_4 in (2.13) as the test statistic.

Based on the sample data, $\hat{\sigma}_1^2 = 0.1437$, $\hat{\sigma}_2^2 = 0.0985$, and $\hat{\sigma}_3^2 = 0.1040$. The pooled estimator of the common variance $\hat{\sigma}_{(R)}^2 = (\hat{\sigma}^2_1 + \hat{\sigma}^2_2 + \hat{\sigma}^2_3)/3 = 0.1154$. To calculate $\hat{\gamma}^*$, we will apply the formula given in (2.20). $\hat{\gamma}^*_{1} = 3.9547$, $\hat{\gamma}^*_{2} = 2.6338$, and $\hat{\gamma}^*_{3} = 3.7061$. The pooled estimator of the common kurtosis is $\hat{\gamma}^*_{(R)} = 3.4315$.

Now substituting the numbers into (2.8) and (2.13) we have $T_2 = 1.8799$ and $T_4 = 1.729$. T_2 and T_4 have asymptotically $\chi^2_{(2)}$ distribution. The p-values for these two statistics are 0.3906 and 0.4213 respectively. We fail to reject the null hypothesis according to both test statistics at the significant level $\alpha = 0.05$.

Applying Shapiro-Wilk's test for normality to the three samples results in p-values 0.2715, 0.338 and 0.1809 respectively. Therefore these samples can be considered drawn from normal populations. Bartlett's test is applied and the corresponding p-value is 0.3517. Levene's test statistic is also calculated, resulting a p-value 0.5527. The fact that the p-values of our test statistics are in between Bartlett's and Levene's test confirms that Bartlett's test is the most sensitive when populations are normal, our tests follow closely in the second and third place, while Levene's is most conservative.

Interval Estimation

Using the confidence interval formula based on transformed data in relation (2.16), the 95% asymptotic confidence intervals for σ_1^2 , σ_2^2 and σ_3^2 are computed, respectively as follows:

$$(0.089, 0.231), (0.0691, 0.1403), (0.0659, 0.1641).$$

Applying (2.18) based on original data, the 95% confidence intervals for σ_1^2, σ_2^2

and σ_3^2 are respectively calculated as follows:

$$(0.0752, 0.2122), (0.0636, 0.1334), (0.0566, 0.1514),$$

From the calculation, it can be seen that the width of the intervals based on transformed data and based on original data are comparable, no significant difference is observed. However, it can be observed that the intervals based on transformed data are shifted rightward comparing to those based on original data.

Since the test showed that the samples are from normal populations, a $(1 - \alpha)100\%$ confidence interval for σ_i^2 can be calculated by $((n - 1)s^2/\chi_{\alpha/2}^2, (n - 1)s^2/\chi_{1-\alpha/2}^2)$, which produces the following 95% confidence intervals for σ_1^2 , σ_2^2 , and σ_3^2 respectively:

$$(0.100, 0.223), (0.0687, 0.1529), (0.0726, 0.1615)$$

These intervals are very close to our intervals based on transformed data, which confirms that our procedure is very reasonable.

Assuming that the null hypothesis in (2.51) is true, applying (2.19) and (2.21), the 95% confidence intervals based on transformed and based on original data for the common variance are (0.0899, 0.1481) and (0.0866, 0.1442) respectively. Again, the width of the two intervals are very close. Comparing the width of the confidence intervals for the common variance and the individual variances, we find that the width from combined data is shorter than that from uncombined data, by approximately $1/\sqrt{k}$.

In this section we applied the proposed methods (estimation and test of hypothesis) to the Iris data. The suggested techniques can be easily implemented and interpreted in a classical probability setting for a wide class of process populations. Hence,

the current study further strengthened the usefulness and applications of $\hat{\sigma}^2$ beyond the normal data.

2.10 Concluding Remarks and Further Research

In this chapter a contribution is made regarding the inference about variance in a multi-sample setting for arbitrary populations. We developed asymptotic tests and asymptotic interval estimation procedures and hence provided a total inferential package. The statistical properties of the proposed inference procedures were investigated analytically and numerically. Our estimation methods are easy to understand, simple to implement, and have good behavior in finite-sample examples. A numerical example based on a real data set demonstrates how to implement and use the proposed methodologies. The simulation study supports our theoretical findings. It is reinforced that a much larger sample should be taken if the parent population is skewed. In many situations the appropriate values of n far exceeded the folklore values of 25 or 30 suitable in the case of asymptotic results associated with t distribution. In this case, instead of using χ_α^2 as the critical value of the test, we could employ a bootstrap resampling method, and use the empirical percentile. This could possibly improve the inference. Research on the statistical implications of these and other estimators is ongoing.

The combined estimator outperforms the estimator based on single data. However, the performance of combined estimator is superior only when the hypothesis of homogeneity of the parameters is true (nearly true). If the opposite holds, then the $\hat{\sigma}_{(R)}^2$ becomes very poor and the analysis based on using $\hat{\sigma}_{(R)}^2$ will be misleading and

inconsistent. In the next chapter, we deal with this critical issue and propose various estimation strategies for estimating σ^2 when the information regarding the equality of variances is rather imprecise.

The kurtosis parameter estimation is embedded in many statistical estimation problems and applications. The estimation of kurtosis parameter is studied extensively in this chapter. We have compared the performance of several kurtosis measures adapted by SAS, SPSS, S-Plus, Minitab, and other statistical packages. We have proposed several new measures of kurtosis. It has been demonstrated both analytically and numerically that our proposed estimators outperform the existing estimators for normal population based on the *MSE* criterion. Moreover, an extensive simulation study has been conducted for non-normal populations. The result indicates that the proposed estimators are superior to the existing ones in many practical situations. Bearing this in mind, all the estimators substantially underestimate kurtosis parameter when underlying population distribution is highly skewed or heavy tailed. In order to correct the bias, empirical formulas are provided for student-t and chi-squared distributions. However, empirical estimates are subject to extra variation introduced and result in inflated *MSE*. Perhaps, some re-sampling methods such as bootstrap and jackknife may be considered to reduce the bias as well as keeping a relatively lower variance.

Chapter 3

Simultaneous Estimation of Variance under Imprecise Information

3.1 Introduction

In Chapter 2, we mentioned that the superior performance of the pooled estimator depends on the equality assumption of all the variances. In this chapter, we develop some alternative estimation strategies when the information regarding the homogeneity of all the variances may not be precise. Again, assuming that homogeneity holds, it is advantageous to combine the data to estimate the common parameter. However, the combined estimator becomes inconsistent when the equality of the hypothesis does not hold. In this situation, estimators based on pretest (Bancroft (1944)) and the James-Stein (James and Stein (1961)) principles are proposed. Asymptotic properties of the shrinkage, positive-part and pretest estimators are discussed and compared with the standard and combined estimators. It is demonstrated that the positive part estimator utilizes the sample and non-sample information in a superior way relative to the ordinary shrinkage estimator.

The use of *imprecise information (II)* is well established and documented in the inference based on the conditional specification arena of statistical procedure. Such information is usually incorporated into classical models. More generally, this methodology abounds in a wide range of statistical applications as evidenced by the research publications and applications of such procedures. In this chapter, we consider the estimation of the the population variances in the presence of *II*.

Professor Efron in *RSS News of January, 1995* wrote:

The empirical Bayes/James-Stein category was the entry in my list least affected by computer developments. It is ripe for a computer-intensive treatment that brings the substantial benefits of James-Stein estimation to bear on complicated, realistic problems. A side benefit may be at least a partial reconciliation between frequentist and Bayesian perspectives as they apply to statistical practice.

It may be worth mentioning that this is one of the two areas Professor Efron predicted for continuing research for the early 21st century. Shrinkage and likelihood-based methods continue to play vital roles in statistical inference. These methods provide extremely useful techniques for combining data from various sources, and recent asymptotic theory has advanced our understanding of the fundamental role of the likelihood function for much the same purpose. Scientific inference is the process of reasoning from observed data back to its underlying mechanism. The two great schools of statistical inference, Bayesian and frequentist, have competed over the past two centuries, often bitterly, for scientific supremacy. Shrinkage/Empirical Bayes, a

novel hybrid, appeared in the 1950s, showing promise of immense possible gains in inferential accuracy. Nevertheless, it has languished in the statistics literature, with its gains viewed as suspicious and even paradoxical by Bayesians and frequentists alike. New scientific technology, exemplified by gene microarrays, has suddenly revived interest in these methods (Cui *et al.* (2005), Ibrahim *et al.* (2002) and references therein).

As in chapter 2, the simultaneous estimation of population variances is considered in a multi-sample situation. We assume here that $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$ ($i = 1, 2, \dots, k$) is a random sample of size n_i taken from the i -th arbitrary population. Let the mean parameter vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)'$ and the covariance structure $\boldsymbol{\Sigma} = \sigma_i^2 \mathbf{I}_{(k \times k)}$, where \mathbf{I} is an identity matrix. If μ_i and σ_i^2 are unknown, then the *unrestricted estimator* of (μ_i, σ_i^2) is $(\hat{\mu}_i, \hat{\sigma}_i^2)$, where

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2. \quad (3.1)$$

Further, there is the possibility that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$. The assumption of the equality of variation is common in many experiments. In this chapter, we consider the problem of simultaneous estimation of σ^2 under this situation. Pre-test and Stein-type estimators are considered to be best suited to the situation. Both methods combine the sample and non-sample information via a test statistic for testing the homogeneity hypothesis.

In the present investigation, we propose several estimators by combining the sample and imprecise information which is given by

$$H_o : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2 \text{ (unknown)}. \quad (3.2)$$

The estimator of σ^2 given in (3.1) is usually used in the case when there is no information available on the vector parameter of interest σ^2 . However, in the situations where we have some information on the space of the parameter of interest, it is advantageous to use this additional information together with the sample information in the hope of obtaining improved estimators, and we intend to explore the same.

A plan of this chapter is as follows. In Section 3.2, several improved estimation strategies are proposed; the expressions for bias and risk (under quadratic loss function) of the estimators and discussion on the risk behavior of the proposed estimators are contained in this section. The numerical risk analysis under a special quadratic loss function is performed in Section 3.3. Furthermore, a simulation study of risk under quadratic and entropy loss functions for finite samples is presented in Section 3.4. Throughout this chapter, the boldface symbols represent vectors/matrices.

3.2 Estimation Strategies Under Imprecise Information

We will propose several estimation strategies of σ^2 which incorporate both the non-sample and sample information.

For the full model, the *separate or unrestricted estimator (UE)* of $\sigma^2 = (\sigma_1^2, \dots, \sigma_k^2)'$, is defined as

$$\hat{\sigma}^2 = (\hat{\sigma}_1^2, \dots, \hat{\sigma}_k^2)', \quad (3.3)$$

where $\hat{\sigma}_i^2$, $i = 1, \dots, k$, is the maximum likelihood estimator of σ_i^2 .

Lemma 3.2.1. *Let $\omega_{n,i} = n_i/n$ and assume the fourth moment of Y_i exists. If $\omega_{n,i} \rightarrow \omega_i$ as $n_i \rightarrow \infty$, then*

$$n^{1/2}\{\hat{\sigma}^2 - \sigma^2\} \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \Gamma), \quad (3.4)$$

where Γ is the covariance matrix of $\hat{\sigma}^2$, and

$$\Gamma = \text{Diag} \left(\frac{(\gamma_1^* - 1)(\sigma_1^2)^2}{\omega_1}, \dots, \frac{(\gamma_k^* - 1)(\sigma_k^2)^2}{\omega_k} \right), \quad (3.5)$$

where ω_i is a constant between 0 and 1, and γ_i^* is the kurtosis of the i^{th} group, as defined in chapter 2, $i = 1, \dots, k$.

However, it may be fruitful to use the information at hand to obtain improved estimates of σ^2 . The information given in (3.2) may be explicitly incorporated into the estimation process by modifying the parameter space. In this situation, the new (restricted) parameter space is a subspace of the original one (reduced in dimension). Statistically speaking, the reduction in dimensionality provides efficient parameter estimates. However, in the case of an incorrect restriction, opposite conclusions will hold. We plan to investigate such characteristics for the estimation problem at hand.

3.2.1 Restricted Estimator

Under the restriction in (3.2), we propose *pooled/combined/restricted estimator (RE)* of σ^2 , defined by

$$\hat{\sigma}_{(R)}^2 = (\hat{\sigma}_{(R)}^2 \cdots, \hat{\sigma}_{(R)}^2)' = \hat{\sigma}_{(R)}^2 \mathbf{1}_k \quad \text{where} \quad \hat{\sigma}_{(R)}^2 = \sum_{i=1}^k n_i \hat{\sigma}_i^2 / (n - k), \quad n = \sum_{i=1}^k n_i. \quad (3.6)$$

Lemma 3.2.2. *Under the usual regularity conditions and the restriction in (3.2), if $n_i/n \rightarrow \omega_i$ ($0 < \omega_{i,n} < 1$) as $n_i \rightarrow \infty$, then*

$$n^{1/2} \{ \hat{\sigma}^2 - \hat{\sigma}_{(R)}^2 \} \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \Gamma_{(R)}), \quad (3.7)$$

where

$$\Gamma_{(R)} = (\gamma^* - 1)(\sigma^2)^2 \mathbf{D}^{-1} \mathbf{J}', \quad \mathbf{J} = \mathbf{I} + \mathbf{1} \mathbf{1}' \mathbf{D}, \quad \mathbf{D} = \text{Diag}(\omega_{1,n}, \dots, \omega_{k,n}). \quad (3.8)$$

The proof that $\hat{\sigma}_{(R)}^2$ is more efficient than $\hat{\sigma}^2$ under the restriction in (3.2) is relatively simple, since the difference between the covariance matrices of $\hat{\sigma}^2$ and

$\hat{\sigma}_{(R)}^2$ is a positive semi-definite matrix. Thus, the pooled estimator $\hat{\sigma}_{(R)}^2$ is more asymptotically efficient (or, at least, no less efficient) than the usual estimator $\hat{\sigma}^2$ when the restriction is true (or the information is precise). However, an interesting and more challenging question is what happens when the restriction is not correct (or the information is imprecise). Intuitively, the restricted estimators will, in general, be biased.

Having said that $\hat{\sigma}_{(R)}^2$ is biased when the constraint in (3.2) fails to hold, it is still of interest to find how well it performs in the entire parameter space induced by the restriction. We shall define our loss function and use the risk measure in order to have a fair comparison.

We are mainly interested in estimating the unknown parameter vector σ^2 by means of an estimator σ_*^2 . A loss function $L(\sigma_*^2, \sigma^2)$ will show the loss incurred by making a wrong decision about σ^2 using the estimator σ_*^2 . For a given positive semi-definite matrix \mathbf{Q} , consider the quadratic loss

$$L(\sigma_*^2, \sigma^2) = n(\sigma_*^2 - \sigma^2)' \mathbf{Q} (\sigma_*^2 - \sigma^2). \quad (3.9)$$

Then the risk of σ_*^2 is defined by the expected loss:

$$R(\sigma_*^2, \sigma^2) = E[L(\sigma_*^2, \sigma^2)] = n \text{trace}[\mathbf{Q}\{E(\sigma_*^2 - \sigma^2)(\sigma_*^2 - \sigma^2)'\}] = \text{trace}(\mathbf{Q}\Omega), \quad (3.10)$$

where $\Omega = n\{E(\sigma_*^2 - \sigma^2)(\sigma_*^2 - \sigma^2)'\} =$ dispersion matrix of $n^{1/2}(\sigma_*^2 - \sigma^2)$.

Further, σ_*^2 is an *inadmissible* estimator of σ^2 if there exists an alternative estimator σ_\diamond^2 such that

$$R(\sigma_\diamond^2, \sigma^2) \leq R(\sigma_*^2, \sigma^2) \quad (3.11)$$

for all σ^2 , and with strict inequality for some σ^2 . We also say that σ_\diamond^2 dominates σ_*^2 .

If instead of (3.11) holding for every n , we have

$$\lim_{n \rightarrow \infty} R(\sigma_\diamond^2, \sigma^2) \leq \lim_{n \rightarrow \infty} R(\sigma_*^2, \sigma^2) \quad (3.12)$$

for all σ^2 with strict inequality holding for some σ^2 , then σ_*^2 is termed an *asymptotically inadmissible* estimator of σ^2 .

In a large sample set-up, for fixed alternative, $\hat{\sigma}_{(R)}^2$ will have an unbounded risk. which is stated more formally in the following theorem.

Theorem 3.2.3. *For large n , if $\sigma^2 \notin H_0$, then $\hat{\sigma}_{(R)}^2$ will have unbounded asymptotic risk.*

We outline the proof as follows:

Considering $\hat{\sigma}_{(R)}^2$, for any $\sigma^2 \notin H_0$,

$$\hat{\sigma}_{(R)}^2 - \sigma^2 \neq 0, \text{ and}$$

$$n(\hat{\sigma}_{(R)}^2 - \sigma^2)'(\hat{\sigma}_{(R)}^2 - \sigma^2) \rightarrow \infty, \text{ as } n \rightarrow \infty.$$

Hence, the risk of $\hat{\sigma}_{(R)}^2$, for any $\sigma^2 \notin H_0$, approaches $+\infty$.

The finding of the theorem tells us that in a large-sample situation and under fixed alternative there is not much to do. In an effort to obtain some interesting and meaningful results, we shall therefore restrict ourselves to contiguous alternatives. More specifically, we consider a sequence $K_{(n)}$ of contiguous (or local) alternatives defined by

$$K_{(n)} : \sigma_{(n)}^2 = \sigma_0^2 + \frac{\theta}{\sqrt{n}}, \text{ where } \sigma_0^2 = \sigma_0^2 \mathbf{1}_k, \theta \text{ is a real fixed vector.} \quad (3.13)$$

Note that $\theta = 0$ implies $\sigma_n^2 = \sigma_0^2 \mathbf{1}_k$, so (3.2) is a particular case of $\{K_{(n)}\}$.

First, we introduce the *asymptotic distribution function (ADF)* of $\sqrt{n}(\sigma_*^2 - \sigma_n^2)$ under $K_{(n)}$ by:

$$G(\mathbf{y}) = \lim_{n \rightarrow \infty} P\{\sqrt{n}(\sigma_*^2 - \sigma_n^2) \leq \mathbf{y} | K_{(n)}\}. \quad (3.14)$$

Then we define the *asymptotic distributional quadratic risk (ADQR)* by

$$ADQR(\sigma_*^2, \sigma^2) = \int \int \cdots \int \mathbf{y}' \mathbf{Q} \mathbf{y} dG(\mathbf{y}) = \text{trace}(\mathbf{Q} \mathbf{Q}^*), \quad (3.15)$$

where $\mathbf{Q}^* = \int \int \cdots \int \mathbf{y} \mathbf{y}' dG(\mathbf{y})$ is the dispersion matrix obtained from G .

We may be able to compute the asymptotic risk by replacing \mathbf{Q}^* with the limit of the actual dispersion matrix of $n^{1/2}(\sigma_*^2 - \sigma^2)$, i.e., $\mathbf{\Omega}$. However, this may require extra regularity conditions to suit the problem in hand. This point has been explained in various other contexts by Sen (1984) and others.

In the same vein, we define the *asymptotic distributional bias (ADB)* of an estimator σ_*^2 of σ^2 as

$$\text{ADB}(\sigma_*^2, \sigma^2) = \int \int \cdots \int \mathbf{y} dG(\mathbf{y}) = \lim_{n \rightarrow \infty} E[n^{1/2}(\sigma_*^2 - \sigma^2)]. \quad (3.16)$$

In order to present a clear-cut picture of various bias functions, we transform the bias functions in scalar form by defining the *Quadratic Bias (QB)* of σ_*^2 as

$$QB(\sigma_*^2) = [\mathbf{B}(\sigma_*^2)]' \mathbf{\Gamma}^{-1} [\mathbf{B}(\sigma_*^2)], \quad (3.17)$$

where $\mathbf{B}(\sigma_*^2) = \text{ADB}(\sigma_*^2)$.

Now, we consider the computation of biases and risk of the estimator under the local alternatives $K_{(n)}$.

Theorem 3.2.4. *Under $K_{(n)}$ and the usual regularity conditions,*

$$\text{ADB}(\hat{\sigma}^2, \sigma^2) = \mathbf{0},$$

$$\begin{aligned} QB(\hat{\sigma}^2, \sigma^2) &= 0, \\ ADQR(\hat{\sigma}^2, \sigma^2) &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}). \end{aligned}$$

Note that $\hat{\sigma}^2$ is asymptotically an unbiased estimator and its ADQR is a constant, with the value of $\text{trace}(\mathbf{Q}\mathbf{\Gamma})$. Obviously, $\hat{\sigma}^2$ does not use the information given in (3.2).

Theorem 3.2.5. *Under $K_{(n)}$ and the usual regularity conditions,*

$$\begin{aligned} \text{ADB}(\hat{\sigma}_{(R)}^2, \sigma^2) &= -\boldsymbol{\theta}^*, \\ QB(\hat{\sigma}_{(R)}^2, \sigma^2) &= \Theta, \\ ADQR(\hat{\sigma}_{(R)}^2, \sigma^2) &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}) - \text{trace}(\mathbf{Q}\mathbf{C}) + \Theta_k, \end{aligned}$$

where $\boldsymbol{\theta}^* = \mathbf{J}\boldsymbol{\theta}$, $\Theta_k = \boldsymbol{\theta}^{*\prime}\mathbf{Q}\boldsymbol{\theta}^*$, $\Theta = \boldsymbol{\theta}^{*\prime}\mathbf{\Gamma}^{-1}\boldsymbol{\theta}^*$, $\mathbf{C} = \mathbf{\Gamma} - \tau_o^2\mathbf{1}\mathbf{1}'$, $\tau_o^2 = (1 - \gamma^*)\sigma_o^4$.

It is easy to see that the magnitude of the quadratic bias of $\hat{\sigma}_{(R)}^2$ increases without a bound and tends to ∞ as $\Theta \rightarrow \infty$. Moreover, the bias does not vanish as the sample gets larger. The ADQR of $\hat{\sigma}_{(R)}^2$ has similar characteristics. The ADQR of $\hat{\sigma}^2$ is constant (independent of Θ) with the value $\text{trace}(\mathbf{Q}\mathbf{\Gamma})$, while the risk of $\hat{\sigma}_{(R)}^2$ becomes unbounded as the hypothesis error grows. Furthermore, we note that

$$ADQR(\hat{\sigma}_{(R)}^2; \sigma^2) \leq ADQR(\hat{\sigma}^2; \sigma^2) \quad \text{if } \Theta_k \leq \text{trace}(\mathbf{Q}\mathbf{C}).$$

Thus, $\hat{\sigma}_{(R)}^2$ dominates $\hat{\sigma}^2$ in the interval $[0, \text{trace}(\mathbf{Q}\mathbf{C})]$. Clearly, when Θ_k moves away from the origin and beyond the value of $\text{trace}(\mathbf{Q}\mathbf{C})$, the ADQR of $\hat{\sigma}_{(R)}^2$ increases without a bound. The ADQR of the linear combination estimator has similar characteristics.

3.2.2 Linear Combination Estimator

The *linear combination estimator (LCE)* or *linear shrinkage estimator (LSE)* of σ^2 may be defined as

$$\hat{\sigma}_{(SR)}^2 = \hat{\sigma}^2 - \pi(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2), \quad \pi \in (0, 1), \quad (3.18)$$

where π is a constant, and may be regarded as the degree of trust in the null hypothesis. If $\pi = 1$ then we obtain the *RE*. Clearly, $\hat{\sigma}_{(SR)}^2$ is a convex combination of $\hat{\sigma}^2$ and $\hat{\sigma}_{(R)}^2$ through a given value of $\pi \in (0, 1)$. Like $\hat{\sigma}_{(R)}^2$, the LCE $\hat{\sigma}_{(SR)}^2$ has a smaller quadratic risk than $\hat{\sigma}^2$ in an interval near the null hypothesis at the expense of poorer performance in the rest of the parameter space. Indeed, its risk becomes unbounded as the hypothesis error grows. The following theorem provides some asymptotic results.

Theorem 3.2.6. *Under $K_{(n)}$ and the usual regularity conditions, the ADB, QB, and ADQR of the LSE are*

$$\begin{aligned} \text{ADB}(\hat{\sigma}_{(SR)}^2, \sigma^2) &= -\pi\theta^*, \\ \text{QB}(\hat{\sigma}_{(SR)}^2, \sigma^2) &= \pi^2\Theta, \\ \text{ADQR}(\hat{\sigma}_{(SR)}^2, \sigma^2) &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}) - \pi(2 - \pi)\text{trace}(\mathbf{Q}\mathbf{C}) + \pi^2\Theta_k. \end{aligned}$$

Thus, $\hat{\sigma}_{(SR)}^2$ also has no control on its quadratic bias since $\Theta \in [0, \infty)$. However, the bias of $\hat{\sigma}_{(SR)}^2$ approaches to infinity slower than that of $\hat{\sigma}_{(R)}^2$, depending on the value of π . The ADQR behavior of $\hat{\sigma}_{(SR)}^2$ has similar characteristics as that of $\hat{\sigma}_{(R)}^2$. The value of π controls the magnitude of ADQR of $\hat{\sigma}_{(SR)}^2$, in this sense $\hat{\sigma}_{(SR)}^2$ is a better choice than $\hat{\sigma}_{(R)}^2$ alone. Further, it is seen that $\hat{\sigma}_{(SR)}^2$ dominates $\hat{\sigma}^2$ in a wider interval than $\hat{\sigma}_{(R)}^2$ does, so $\hat{\sigma}_{(SR)}^2$ has an edge over $\hat{\sigma}_{(R)}^2$ in this sense.

In summary, if the information is *precise* in the sense that the hypothesis error is negligible, then the pooled estimator is superior to $\hat{\sigma}^2$ and $\hat{\sigma}_{(R)}^2$ and $\hat{\sigma}_{(SR)}^2$ offer

a substantial gain over $\hat{\sigma}^2$. Alternatively, if the information is *imprecise*, i.e., the hypothesis error is substantial, then as expected, $\hat{\sigma}^2$ performs better than the pooled estimators. More importantly, one seldom knows whether the information is imprecise or not.

The above insight leads to preliminary test estimation when the hypothesis information is rather uncertain and it is useful to construct a compromised estimator by performing a preliminary test on the given information in the form of the null hypothesis.

3.2.3 Preliminary Test Estimator

The *preliminary test estimator (PTE)* can be readily obtained by replacing π , a fixed number in (3.18), by a binary random quantity. Therefore, PTE of σ^2 is defined by

$$\hat{\sigma}_{(PT)}^2 = \hat{\sigma}^2 - (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)I(\Lambda < \lambda_{n,\alpha}), \quad (3.19)$$

where $I(A)$ is an indicator function of a set A and $\lambda_{n,\alpha}$ is the the upper α -level critical value of Λ . Further, from chapter 2,

$$\Lambda = n(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)' \hat{\Gamma}_n^{-1} (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2), \quad (3.20)$$

where $\hat{\Gamma}_n^{-1} = \frac{\Omega_2}{\hat{\tau}^2}$, $\Omega_2 = \text{Diag}(\omega_{1,n}, \dots, \omega_{k,n})$, $\omega_{i,n} = \frac{n_i}{n}$, $\hat{\tau}^2 = (1 - \hat{\gamma}^*)(\hat{\sigma}_{(R)}^2)^2$. Thus, when the null hypothesis is true, the large sample distribution of Λ converges in distribution to a central χ^2 distribution with $(k - 1)$ degrees of freedom. Hence, the upper α -level critical value of Λ defined by $\lambda_{n,\alpha}$ may be approximated by $(1 - \alpha)100^{\text{th}}$ percentile of the central χ^2 distribution with $(k - 1)$ degrees of freedom.

In turn, an estimator with a good control on its ADQR function is achieved. Indeed, the preliminary test estimators are obtained as convex combinations of *unrestricted* and *combined* estimators via a test-statistic, Λ , for testing H_0 in (3.2). Meanwhile, a bounded quadratic risk is achieved.

The biases and risk of $\hat{\sigma}_{(PT)}^2$ under the local alternatives are given in the following theorem:

Theorem 3.2.7. *Let $\Psi_k(x; \Theta)$ be the noncentral chi-squared distribution function with non-centrality parameter Θ and degrees of freedom k , then under $K_{(n)}$ and the regularity conditions,*

$$\begin{aligned} \text{ADB}(\hat{\sigma}_{(PT)}^2) &= -\theta^* \Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta), \\ \text{QB}(\hat{\sigma}_{(PT)}^2) &= \Theta [\Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta)]^2, \\ \text{ADQR}(\hat{\sigma}_{(PT)}^2, \sigma^2) &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}) - \text{trace}(\mathbf{Q}\mathbf{C}) \Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta) + \\ &\quad \Theta_k \{2\Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta) - \Psi_{k+3}(\chi_{k-1, \alpha}^2; \Theta)\}. \end{aligned}$$

Proof. By definition,

$$\begin{aligned} \text{ADB}(\hat{\sigma}_{(PT)}^2) &= \lim_{n \rightarrow \infty} \sqrt{n} E[\hat{\sigma}_{(PT)}^2 - \sigma_n^2] \\ &= \lim_{n \rightarrow \infty} \sqrt{n} E[\hat{\sigma}^2 - (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2) I(\Lambda < \chi_{n, \alpha}^2) - \sigma_n^2] \\ &= \lim_{n \rightarrow \infty} \sqrt{n} E[(\hat{\sigma}^2 - \sigma_n^2) I(\Lambda > \chi_{n, \alpha}^2) + (\hat{\sigma}_{(R)}^2 - \sigma_n^2) I(\Lambda < \chi_{n, \alpha}^2)] \\ &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\sigma}_{(R)}^2 - \sigma_n^2) I(\Lambda < \chi_{n, \alpha}^2)] \\ &= -\theta^* \Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta). \end{aligned}$$

The last step is by Stein's identities given by Judge and Bock (1978).

The proof of $\text{QB}(\hat{\sigma}_{(PT)}^2) = \Theta [\Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta)]^2$ follows immediately from the result of $\text{ADB}(\hat{\sigma}_{(PT)}^2)$ above.

Similarly,

$$\begin{aligned} \text{ADQR}(\hat{\sigma}_{(PT)}^2, \sigma^2) &= \lim_{n \rightarrow \infty} n E[(\hat{\sigma}_{(PT)}^2 - \sigma_n^2)' \mathbf{Q} (\hat{\sigma}_{(PT)}^2 - \sigma_n^2)] \\ &= \lim_{n \rightarrow \infty} n E[(\hat{\sigma}_{(R)}^2 - \sigma_n^2)' \mathbf{Q} (\hat{\sigma}_{(R)}^2 - \sigma_n^2)] I(\Lambda < \chi_{n, \alpha}^2) \\ &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}) - \text{trace}(\mathbf{Q}\mathbf{C}) \Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta) \\ &\quad + \Theta_k \{2\Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta) - \Psi_{k+3}(\chi_{k-1, \alpha}^2; \Theta)\}. \end{aligned}$$

□

Useful discussions on some of the implications of the pretest and shrinkage estimators in parametric theory are given in Bancroft (1944), Efron and Morris (1975), Judge and Bock (1978), Stigler (1990), Ahmed (2002), and Ahmed *et al.* (2006) among others. For some asymptotic results on the subject we refer to Ahmed *et al.* (2001), Ahmed (2002), Ahmed (2005), and An *et al.* (2006). It is important to remark that $\hat{\sigma}_{(PT)}^2$ performs better than $\hat{\sigma}^2$ in some part of the parameter space. Further, the use of $\hat{\sigma}_{(PT)}^2$ may, however, be limited due to the large size of the preliminary test.

Recall that $\hat{\sigma}_{(SR)}^2$ provides a wider range than $\hat{\sigma}_{(R)}^2$ does in which it dominates $\hat{\sigma}^2$. Thus, it is logical to replace $\hat{\sigma}_{(R)}^2$ by $\hat{\sigma}_{(SR)}^2$ in (3.19) to obtain an improved preliminary test estimator.

3.2.4 Improved Preliminary Test Estimator

We obtain the *improved preliminary test estimator (IPTE)* by multiplying the random quantity $I(\Lambda < \lambda_{n,\alpha})$ by a fixed (π) number to improve upon both $\hat{\sigma}_{(SR)}^2$ and $\hat{\sigma}_{(PT)}^2$. Indeed, the process is natural.

$$\hat{\sigma}_{(SP)}^2 = \hat{\sigma}^2 - \pi(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)I(\Lambda < \lambda_{n,\alpha}). \quad (3.21)$$

Ahmed (1992) demonstrated that $\hat{\sigma}_{(SP)}^2$ significantly improves upon $\hat{\sigma}_{(PT)}^2$ in the size of the test, and dominates $\hat{\sigma}^2$ for a large portion of the parameter space. However, $\hat{\sigma}_{(PT)}^2$ is more efficient than $\hat{\sigma}_{(SP)}^2$ near the restriction. There is no clear-cut winner in the entire parameter space. Meanwhile, $\hat{\sigma}_{(SP)}^2$ and $\hat{\sigma}_{(PT)}^2$ involve the test statistic Λ which adjusts the estimator for any empirical departure from the null hypothesis. For large values of Λ , both estimators yield $\hat{\sigma}^2$, while for small values of Λ their performance is different. Indeed, $\hat{\sigma}^2$ and $\hat{\sigma}_{(PT)}^2$ may be considered as special cases of

$\hat{\sigma}_{(SP)}^2$: for $\pi = 0$, $\hat{\sigma}_{(SP)}^2 = \hat{\sigma}^2$ and for $\pi = 1$, $\hat{\sigma}_{(SP)}^2 = \hat{\sigma}_{(PT)}^2$.

Theorem 3.2.8. *Let $\Psi_k(x; \Theta)$ be the noncentral chi-squared distribution function with non-centrality parameter Θ and degrees of freedom k , then under $K_{(n)}$ and the regularity conditions,*

$$\begin{aligned} \text{ADB}(\hat{\sigma}_{(SP)}^2) &= -\pi\theta^* \Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta), \\ \text{QB}(\hat{\sigma}_{(SP)}^2) &= \pi^2 \Theta [\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta)]^2, \\ \text{ADQR}(\hat{\sigma}_{(SP)}^2; \sigma^2) &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}) - \pi(2 - \pi) \text{trace}(\mathbf{Q}\mathbf{C}) \Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) + \\ &\quad \Theta_k \{2\pi \Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) - \pi(2 - \pi) \Psi_{k+3}(\chi_{k-1,\alpha}^2; \Theta)\}. \end{aligned}$$

Proof. The proof follows the same arguments as in Theorem 3.2.7. □

For $\pi = 1$, we obtain the biases and ADQR of $\hat{\sigma}_{(PT)}^2$ given in Theorem 3.2.7.

Clearly $B(\hat{\sigma}_{(SP)}^2) = \pi B(\hat{\sigma}_{(PT)}^2) < B(\hat{\sigma}_{(PT)}^2)$ for $\pi \in (0, 1)$. Hence, $\hat{\sigma}_{(SP)}^2$ has asymptotically less bias than that of $\hat{\sigma}_{(PT)}^2$, depending upon the value of π . One may also view π as a bias reduction factor in the preliminary test estimation.

It is important to note that the ADQR of $\hat{\sigma}_{(SP)}^2$ is bounded in Θ and it begins with an initial value of $[\text{trace}(\mathbf{Q}\mathbf{\Gamma}) - \text{trace}(\mathbf{Q}\mathbf{C})\pi(2 - \pi) \Psi_{k+1}(\chi_{k-1,\alpha}^2; 0)]$. Hence the ADQR of $\hat{\sigma}_{(PT)}^2$ is also a bounded function of the non-centrality parameter Θ . Further, as the value of Θ deviates from the null hypothesis the ADQR of $\hat{\sigma}_{(SP)}^2$ increases, exceeds the ADQR of $\hat{\sigma}^2$ and achieves a maximum value; then the ADQR of $\hat{\sigma}_{(SP)}^2$ decreases and approaches the ADQR of $\hat{\sigma}^2$. The risk function of $\hat{\sigma}_{(PT)}^2$ also follows a similar pattern.

Comparing the risk of $\hat{\sigma}_{(SP)}^2$ with the risk of $\hat{\sigma}^2$ we note that

$$\begin{aligned} \text{ADQR}(\hat{\sigma}_{(SP)}^2; \sigma^2) &\leq \text{ADQR}(\hat{\sigma}^2; \sigma^2) \quad \text{if} \\ \Theta_k &\leq \frac{\text{trace}(\mathbf{Q}\mathbf{C})(2 - \pi) \Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta)}{2\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) - (2 - \pi) \Psi_{k+3}(\chi_{k-1,\alpha}^2; \Theta)}. \end{aligned}$$

Thus, $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}^2$ for some values of Θ . On the other hand,

$$ADQR(\hat{\sigma}_{(PT)}^2; \sigma^2) \leq ADQR(\hat{\sigma}^2; \sigma^2) \quad \text{if}$$

$$\Theta_k \leq \frac{\text{trace}(\mathbf{QC})\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta)}{2\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) - \Psi_{k+3}(\chi_{k-1,\alpha}^2; \Theta)}.$$

Hence, the performance of $\hat{\sigma}_{(SP)}^2$ is better than $\hat{\sigma}^2$ in a larger portion of parameter space compared to $\hat{\sigma}_{(PT)}^2$ for $\pi \in (0, 1)$. The fact that $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}^2$ in the interval $[0, \text{trace}(\mathbf{QC})(2 - \pi)\pi^{-1})$ and $\hat{\sigma}_{(PT)}^2$ performs better than $\hat{\sigma}^2$ when $\Theta \in [0, \text{trace}(\mathbf{QC}))$ shows the superiority of $\hat{\sigma}_{(SP)}^2$ over $\hat{\sigma}_{(PT)}^2$.

Next we will compare $\hat{\sigma}_{(SP)}^2$ with $\hat{\sigma}_{(R)}^2$.

$$\frac{ADQR(\hat{\sigma}_{(SP)}^2; \sigma^2)}{ADQR(\hat{\sigma}_{(R)}^2; \sigma^2)} \leq 1 \quad \text{if}$$

$$\Theta_k \geq \frac{\text{trace}(\mathbf{QC})\{1 - \pi(2 - \pi)\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta)\}}{1 - 2\pi\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) + \pi(2 - \pi)\Psi_{k+3}(\chi_{k-1,\alpha}^2; \Theta)}.$$

Hence, $\hat{\sigma}_{(SP)}^2$ is superior to $\hat{\sigma}_{(R)}^2$ if

$$\Theta_k \in \left(\frac{\text{trace}(\mathbf{QC})\{1 - \pi(2 - \pi)\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta)\}}{1 - 2\pi\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) + \pi(2 - \pi)\Psi_{k+3}(\chi_{k-1,\alpha}^2; \Theta)}, \infty \right),$$

and $\hat{\sigma}_{(R)}^2$ will perform better than $\hat{\sigma}_{(SP)}^2$ outside this interval. Consequently, we conclude that none of the estimators is asymptotically superior, since the risk functions of the estimators are crossing each other in the parameter space.

Finally, we compare $\hat{\sigma}_{(SP)}^2$ and $\hat{\sigma}_{(PT)}^2$ and determine the conditions under which $\hat{\sigma}_{(SP)}^2$ outperforms $\hat{\sigma}_{(PT)}^2$. First, we consider the case when H_0 is true. In this situation,

$$ADQR(\hat{\sigma}_{(SP)}^2; \sigma^2) - ADQR(\hat{\sigma}_{(PT)}^2; \sigma^2) = \text{trace}(\mathbf{QC})(1 - \pi)^2\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta).$$

Since the right hand side of the above expression is a positive quantity, $\hat{\sigma}_{(PT)}^2$ has a smaller risk than $\hat{\sigma}_{(SP)}^2$ does under H_0 . However, the picture is somewhat different when the hypothesis deviates from the true value of the parameters. Since

$$ADQR(\hat{\sigma}_{(SP)}^2; \mathbf{Q}) - ADQR(\hat{\sigma}_{(PT)}^2; \sigma^2) = \text{trace}(\mathbf{QC})(1 - \pi)^2 \Psi_{k+3}(\chi_{k-1, \alpha}^2; \Theta) - \Theta_k \{2(1 - \pi) \Psi_{k+1}(\chi_{k-1, \alpha}^2; \Theta) - (1 - \pi)^2 \Psi_{k+3}(\chi_{k-1, \alpha}^2; \Theta)\}; \quad (3.22)$$

the risk of $\hat{\sigma}_{(SP)}^2$ will be larger than that of $\hat{\sigma}_{(PT)}^2$ in the neighborhood of the null hypothesis, which may be negligible for larger values of π . On the other hand, as null hypothesis deviates from the true value of the parameters, the difference in (3.22) becomes negative and $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}_{(PT)}^2$ in the remaining parameter space. For a given π , let Θ_{k_π} be a point in the parameter space at which the risk of $\hat{\sigma}_{(SP)}^2$ and $\hat{\sigma}_{(PT)}^2$ intersects. Then, for $\Theta \in (0, \Theta_{k_\pi}]$, $\hat{\sigma}_{(PT)}^2$ performs better than $\hat{\sigma}_{(SP)}^2$; while for $\Theta_k \in (\Theta_{k_\pi}, \infty)$, $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}_{(PT)}^2$. Further, for large values of π (close to 1), the difference between $\hat{\sigma}_{(PT)}^2$ and $\hat{\sigma}_{(SP)}^2$ may not be significant on the interval $(0, \Theta_{k_\pi}]$. Nonetheless, $\hat{\sigma}_{(PT)}^2$ and $\hat{\sigma}_{(SP)}^2$ share a common asymptotic property that as hypothesis error grows and tends to ∞ , their risks converge to a common limit, i.e., to the ADQR of $\hat{\sigma}^2$.

Thus, in light of the above discussion, none of the four estimators $\hat{\sigma}_{(PT)}^2$, $\hat{\sigma}^2$, $\hat{\sigma}_{(R)}^2$ and $\hat{\sigma}_{(SP)}^2$ is asymptotically superior than the others.

The estimators based on the preliminary test method are sensitive to departure from H_0 and may not be useful for all σ^2 , and neither $\hat{\sigma}_{(SP)}^2$ nor $\hat{\sigma}_{(PT)}^2$ perform uniformly better than $\hat{\sigma}^2$. The performance of the estimators depends on the correctness of the information regarding the homogeneity of the parameters. This is somewhat

not helpful from practical point of view due to the fact that the information regarding parameters is generally imprecise in real life situations.

To overcome this shortcoming, we propose an estimator using the optimal weight for combining the data and non-data information. The proposed estimator combines the sample and non-sample information in a superior way than the preceding estimators. Further, it resembles the Stein-type estimator (Stein, 1956).

3.2.5 James-Stein Type Estimator

For $\pi = \frac{c}{\Lambda}$ in (3.18) we obtain a James-Stein type estimator, where c is so-called the shrinkage constant and $0 < c < 2(k - 3)$, and Λ is as defined in (3.20). Hence, a *James-Stein estimator (JSE)* of σ^2 is defined by

$$\hat{\sigma}_{(JS)}^2 = \hat{\sigma}^2 - \frac{c}{\Lambda}(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2), \quad 0 < c < 2(k - 3). \quad (3.23)$$

The proposed Stein type estimator dominates $\hat{\sigma}^2$. It is, however, not a convex combination of $\hat{\sigma}_{(R)}^2$ and $\hat{\sigma}^2$. Hence, the proposed estimator $\hat{\sigma}_{(JS)}^2$ may not remain non-negative. This feature can be seen easily by rewriting the above relation as

$$\hat{\sigma}_{(JS)}^2 = \hat{\sigma}_{(R)}^2 + \left\{ 1 - \frac{k - 3}{\Lambda} \right\} (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2). \quad (3.24)$$

Here we used the optimal value of c , which is $k - 3$, for the current estimation problem. To avoid the over-shrinking inherent in $\hat{\sigma}_{(JS)}^2$, a truncation is made leading to a convex combination of $\hat{\sigma}^2$ and $\hat{\sigma}_{(R)}^2$, which is called *positive rule James-Stein estimator (PJSE)*.

3.2.6 Positive Rule James-Stein Estimator

The positive-rule James-Stein estimator may be defined as follows:

$$\hat{\sigma}_{(PP)}^2 = \hat{\sigma}_{(R)}^2 + \left\{ 1 - \frac{k-3}{\Lambda} \right\}^+ (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2), \quad (3.25)$$

where $U^+ = \max(0, U)$. Further, $\hat{\sigma}_{(PP)}^2$ may also be written as:

$$\hat{\sigma}_{(PP)}^2 = \hat{\sigma}_{(JS)}^2 - \left\{ 1 - \frac{k-3}{\Lambda} \right\} I(\Lambda \leq k-3) (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2). \quad (3.26)$$

Theorem 3.2.9. *If $\sigma^2 \notin H_0$ then $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ have the same finite risk as that of $\hat{\sigma}^2$.*

Proof. We outline the proof as follows. By investigating the behavior of $\hat{\sigma}_{(JS)}^2$, we note that

$$\begin{aligned} \sqrt{n}(\hat{\sigma}_{(JS)}^2 - \hat{\sigma}^2)' \mathbf{Q} \sqrt{n}(\hat{\sigma}_{(JS)}^2 - \hat{\sigma}^2) &= (k-3)^2 \Lambda^{-2} \{ \sqrt{n}(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)' \mathbf{Q} \sqrt{n}(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2) \} \\ &\leq (k-3)^2 \{ n(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)' \mathbf{Q} (\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2) \}^{-1} \\ &\quad \{ ch_{\max}(\mathbf{Q}\Omega^{-1}) \}^2. \end{aligned}$$

where $ch_{\max}(\cdot)$ is the largest eigenvalues of (\cdot) . Also, under H_0 , we have $\hat{\sigma}_{(JS)}^2 = \hat{\sigma}_{(R)}^2$. Further, for $\sigma^2 \notin H_0$,

$$E\{\Lambda^{-1} I(\Lambda > 0)\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

In other words, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}^2$ become asymptotically risk equivalent for every σ^2 not in H_0 . A similar analysis holds for $\hat{\sigma}_{(PP)}^2$. \square

Now we will investigate the ADQR under the local alternatives defined in (3.13) and compare the respective performance of the proposed estimators.

We present the expressions for the biases and ADQRs of the estimators in the following theorem.

Theorem 3.2.10.

$$\begin{aligned} \text{ADB}(\hat{\sigma}_{(PP)}^2) &= -\theta^* [\Psi_{k+1}(k-3; \Theta) + E\{\chi_{k+1}^{-2}(\Theta) I[\chi_{k+1}^2(\Theta) > (k-3)]\}] \\ \text{QB}(\hat{\sigma}_{(JS)}^2) &= (k-3)^2 \Theta [E(\chi_{k+1}^{-2}(\Theta))]^2, \end{aligned}$$

$$\begin{aligned}
 QB(\hat{\sigma}_{(PP)}^2) &= \Theta [\Psi_{k+1}(k-3; \Theta) + E\{\chi_{k+1}^{-2}(\Theta)I[\chi_{k+1}^2(\Theta) > (k-3)]\}]^2 \\
 ADQR(\hat{\sigma}_{(JS)}^2; \sigma^2) &= \text{trace}(\mathbf{Q}\mathbf{\Gamma}) + \Theta_k(k-3)(k+1)E(\chi_{k+3}^{-4}(\Theta)) - \\
 &\quad (k-3)\text{trace}(\mathbf{Q}\mathbf{C})\{2E(\chi_{k+1}^{-2}(\Theta)) - (k-3)E(\chi_{k+1}^{-4}(\Theta))\}. \\
 ADQR(\hat{\sigma}_{(PP)}^2; \sigma^2) &= R(\hat{\sigma}_{(JS)}^2; \sigma^2) - \text{trace}(\mathbf{Q}\mathbf{C}) \\
 &\quad E[\{1 - (k-3)\chi_{k+1}^{-2}(\Theta)\}^2 I[\chi_{k+1}^2(\Theta) \leq (k-3)]] + \\
 &\quad \Theta_k [E\{2\{1 - (k-3)\chi_{k+1}^{-2}(\Theta)\} I[\chi_{k+1}^2(\Theta) \leq (k-3)]] - \\
 &\quad E[\{1 - (k-3)\chi_{k+3}^{-2}(\Theta)\}^2 I[\chi_{k+3}^2(\Theta) \leq (k-3)]]].
 \end{aligned}$$

Proof. The proof follows the same arguments as in Theorem 3.2.7. \square

The quadratic bias of $\hat{\sigma}_{(JS)}^2$ starts from 0 at $\Theta = 0$, increases to a maximum value, then decreases towards 0, since $E(\chi_{\nu}^{-2}(\Theta))$ is a decreasing log-convex function of Θ . The behavior of $\hat{\sigma}_{(PP)}^2$ is similar to $\hat{\sigma}_{(JS)}^2$, however, the bias curve of $\hat{\sigma}_{(PP)}^2$ remains below the curve of $\hat{\sigma}_{(JS)}^2$ for all values of Θ .

We now turn to investigate the comparative statistical properties of the shrinkage-type estimators. First we compare it with $\hat{\sigma}^2$ when the null hypothesis is true.

$$ADQR(\hat{\sigma}^2; \sigma^2) - ADQR(\hat{\sigma}_{(JS)}^2; \sigma^2) = \text{trace}(\mathbf{Q}\mathbf{C})(k-3)E\{2\chi_{k+1}^{-2}(0) - (k-3)\chi_{k+1}^{-4}(0)\}$$

is a positive quantity. Hence, we conclude that the Stein-type estimator dominates $\hat{\sigma}^2$ when the null hypothesis is true. Also, the maximum risk gain of $\hat{\sigma}_{(JS)}^2$ over $\hat{\sigma}^2$ is achieved at the null hypothesis.

Now, we characterize a class of positive semi-definite matrices by

$$\mathbf{Q}^D = \left\{ \mathbf{Q} \mid \frac{\text{trace}(\mathbf{Q}\mathbf{\Gamma})}{ch_{max}(\mathbf{Q}\mathbf{\Gamma})} \geq \frac{k+1}{2} \right\} \quad (3.27)$$

In order to provide a meaningful comparison of the various estimators, we state the following theorem .

Courant Theorem If \mathbf{C} and \mathbf{D} are two positive semi-definite matrices with \mathbf{D} nonsingular, both of order $(q \times q)$, and \mathbf{x} is a column vector of order $(q \times 1)$, then

$$ch_{\min}(\mathbf{CD}^{-1}) \leq \frac{\mathbf{x}'\mathbf{C}\mathbf{x}}{\mathbf{x}'\mathbf{D}\mathbf{x}} \leq ch_{\max}(\mathbf{CD}^{-1})$$

where $ch_{\min}(\cdot)$ and $ch_{\max}(\cdot)$ are the smallest and largest eigenvalues of (\cdot) respectively.

We note that the above lower and upper bounds are equal to the infimum and supremum, respectively, of the ratio $\frac{\mathbf{x}'\mathbf{C}\mathbf{x}}{\mathbf{x}'\mathbf{D}\mathbf{x}}$ for $\mathbf{x} \neq \mathbf{0}$. Also, for $\mathbf{D} = \mathbf{I}$, the ratio is known as Rayleigh quotient for matrix \mathbf{C} .

As a consequence of the above Courant theorem, we have

$$ch_{\min}(\mathbf{Q}\mathbf{\Gamma}) \leq \frac{\boldsymbol{\theta}^{*\prime}\mathbf{Q}\boldsymbol{\theta}^*}{\boldsymbol{\theta}^{*\prime}\mathbf{\Gamma}^{-1}\boldsymbol{\theta}^*} \leq ch_{\max}(\mathbf{Q}\mathbf{\Gamma}), \quad \text{for } \boldsymbol{\theta}^* \neq \mathbf{0} \text{ and } \mathbf{Q} \in \mathbf{Q}^D.$$

Thus, under the class of matrices defined in relation (3.27) we conclude that $ADQR(\hat{\boldsymbol{\sigma}}_{(JS)}^2; \boldsymbol{\sigma}^2) \leq ADQR(\hat{\boldsymbol{\sigma}}^2; \boldsymbol{\sigma}^2)$ for all $\boldsymbol{\theta}^*$, where strict inequality holds for some $\boldsymbol{\theta}^*$. It clearly indicates the asymptotic inadmissibility of $\hat{\boldsymbol{\sigma}}^2$ relative to $\hat{\boldsymbol{\sigma}}_{(JS)}^2$ under local alternatives. The risk of $\hat{\boldsymbol{\sigma}}_{(JS)}^2$ begins with an initial value of 3 and increases monotonically towards $\text{trace}(\mathbf{Q}\mathbf{\Gamma})$ as hypothesis error moves away from 0. The risk of $\hat{\boldsymbol{\sigma}}_{(JS)}^2$ is uniformly smaller than $\hat{\boldsymbol{\sigma}}^2$, where the upper limit is attained when $\|\boldsymbol{\theta}^*\| \rightarrow \infty$. The result is valid as long as the expectation exists, and the expectation in the expression of $ADQR(\hat{\boldsymbol{\sigma}}_{(JS)}^2; \boldsymbol{\sigma}^2)$ in theorem (3.2.10) exists whenever $k \geq 4$.

We now wish to compare $\hat{\boldsymbol{\sigma}}_{(JS)}^2$ and $\hat{\boldsymbol{\sigma}}_{(R)}^2$ under H_0 . We have

$$ADQR(\hat{\boldsymbol{\sigma}}_{(JS)}^2; \boldsymbol{\sigma}^2) - ADQR(\hat{\boldsymbol{\sigma}}_{(R)}^2; \boldsymbol{\sigma}^2) = \text{trace}(\mathbf{Q}\mathbf{C}) - \frac{k-3}{k-1}\text{trace}(\mathbf{Q}\mathbf{\Gamma}) > 0. \quad (3.28)$$

Therefore, the ADQR of $\hat{\boldsymbol{\sigma}}_{(R)}^2$ is smaller than the ADQR of $\hat{\boldsymbol{\sigma}}_{(JS)}^2$ when the null hypothesis is true. Alternatively, as $\boldsymbol{\theta}^*$ departs from the null vector, Θ_k increases and

$E(\chi_{k+1}^{-4}(\Theta))$ decreases, so $\hat{\sigma}_{(JS)}^2$ has smaller ADQR than $\hat{\sigma}_{(R)}^2$. Generally speaking $\hat{\sigma}_{(JS)}^2$ does not perform better than $\hat{\sigma}_{(R)}^2$ in a small neighborhood of the null hypothesis and $\hat{\sigma}_{(JS)}^2$ dominates $\hat{\sigma}_{(R)}^2$ in the rest of the parameter space. Hence, under local alternatives neither of $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(R)}^2$ is asymptotically better than the other.

Next, we compare $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(SP)}^2$ under H_0 .

$$\begin{aligned} & ADQR(\hat{\sigma}_{(JS)}^2, \sigma^2) - ADQR(\hat{\sigma}_{(SP)}^2, \sigma^2) \\ &= \text{trace}(\mathbf{QC}) \left\{ \pi(2 - \pi)\Psi_{k+1}(\chi_{k-1,\alpha}^2; 0) - \frac{(k-3)}{(k-1)} \right\} > 0, \end{aligned}$$

whenever

$$\pi < 1 - \sqrt{1 - \frac{(k-3)}{(k-1)\Psi_{k+1}(\chi_{k-1,\alpha}^2; 0)}}, \quad \text{and} \quad \Psi_{k+1}(\chi_{k-1,\alpha}^2; 0) < \frac{(k-3)}{(k-1)}. \quad (3.29)$$

otherwise $\hat{\sigma}_{(JS)}^2$ has a smaller risk. For $\pi = 1$ we get comparison for $\hat{\sigma}_{(PT)}^2$ versus $\hat{\sigma}_{(JS)}^2$. In general, Under H_0 , when (3.29) holds, we may order the dominance of the estimators as: $\hat{\sigma}_{(R)}^2 \succ \hat{\sigma}_{(PT)}^2 \succ \hat{\sigma}_{(SP)}^2 \succ \hat{\sigma}_{(JS)}^2 \succ \hat{\sigma}^2$, where the notation \succ stands for dominance. However, the dominance picture changes to $\hat{\sigma}_{(R)}^2 \succ \hat{\sigma}_{(JS)}^2 \succ \hat{\sigma}_{(PT)}^2 \succ \hat{\sigma}_{(SP)}^2 \succ \hat{\sigma}^2$, whenever (3.29) fails to hold. Thus, (3.29) specifies a range of values of π and α for which $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}_{(JS)}^2$. The picture changes as θ^* moves away from the null vector. The ADQR of $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(SP)}^2$ intersect at $\Theta_k = \Theta_{k,\pi,\alpha}$ if the condition (3.29) is satisfied, otherwise there is no intersecting point in the parameter space. If $\Theta_k \in [0, \Theta_{k,\pi,\alpha})$, then $\hat{\sigma}_{(SP)}^2 \succ \hat{\sigma}_{(JS)}^2$ while for $\Theta_k \in (\Theta_{k,\pi,\alpha}, \infty)$, $\hat{\sigma}_{(JS)}^2 \succ \hat{\sigma}_{(SP)}^2$. If (3.29) is not satisfied then $\hat{\sigma}_{(JS)}^2 \succ \hat{\sigma}_{(SP)}^2$ for $\Theta_k \in [0, \infty)$. In fact, the ADQR of $\hat{\sigma}_{(JS)}^2$ always lies below this asymptotic value. Also, we note that the application of $\hat{\sigma}_{(JS)}^2$ is constrained by the requirement that $k \geq 4$. If $k < 4$, then $\hat{\sigma}_{(SP)}^2$ may be the sensible choice for estimating the parameter vector σ^2 .

Finally, we wish to compare the risk performance of $\hat{\sigma}_{(PP)}^2$ and $\hat{\sigma}_{(JS)}^2$. We may conclude from theorem (3.2.10) that

$$\frac{ADQR(\hat{\sigma}_{(PP)}^2; \sigma^2)}{ADQR(\hat{\sigma}_{(JS)}^2; \sigma^2)} \leq 1, \quad \text{for all } \Theta,$$

with strict inequality for some Θ . Therefore, $\hat{\sigma}_{(PP)}^2$ asymptotically dominates $\hat{\sigma}_{(JS)}^2$ under local alternatives. Consequently, $\hat{\sigma}_{(PP)}^2 \succ \hat{\sigma}_{(JS)}^2 \succ \hat{\sigma}^2$. The dominance pattern of $\hat{\sigma}_{(PP)}^2$ relative to $\hat{\sigma}_{(SP)}^2$ is similar to that of $\hat{\sigma}_{(JS)}^2$ to $\hat{\sigma}_{(SP)}^2$ and hence is not discussed here to save space.

3.3 Numerical ADQR Analysis

It is noted that the risk of all the estimators depend on the matrices \mathbf{Q} and $\mathbf{\Gamma}$. In order to facilitate numerical computation of the ADQR functions, we consider the particular case $\mathbf{Q} = \mathbf{\Gamma}^{-1}$ and obtain the value of risk expressions on a computer. With this substitution in theorems (3.2.4), (3.2.6), (3.2.8), and (3.2.10), the risks, denoted by $ADQR_l$, $l = 1, \dots, 5$ for simplicity, are given in the following corollary:

Corollary 3.3.1. *Let $\mathbf{Q} = \mathbf{\Gamma}^{-1}$, then the ADQR of $\hat{\sigma}^2$, $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(SP)}^2$, $\hat{\sigma}_{(JS)}^2$, and $\hat{\sigma}_{(PP)}^2$ are*

$$\begin{aligned} ADQR_1 &= k \\ ADQR_2 &= k - \pi(2 - \pi)(k - 1) + \pi^2\Theta \\ ADQR_3 &= k - \pi(2 - \pi)(k - 1)\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) + \\ &\quad \Theta\{2\pi\Psi_{k+1}(\chi_{k-1,\alpha}^2; \Theta) - \pi(2 - \pi)\Psi_{k+3}(\chi_{k-1,\alpha}^2; \Theta)\}, \\ ADQR_4 &= k + \Theta(k - 3)(k + 1)E(\chi_{k+3}^{-4}(\Theta)) - \\ &\quad (k - 1)(k - 3)\{2E(\chi_{k+1}^{-2}(\Theta)) - (k - 3)E(\chi_{k+1}^{-4}(\Theta))\}, \\ ADQR_5 &= R_4 - (k - 1)E[\{1 - (k - 3)\chi_{k+1}^{-2}(\Theta)\}^2 I[\chi_{k+1}^2(\Theta) \leq (k - 3)]] + \\ &\quad \Theta[2E[\{1 - (k - 3)\chi_{k+1}^{-2}(\Theta)\} I[\chi_{k+1}^2(\Theta) \leq (k - 3)]] - \\ &\quad E[\{1 - (k - 3)\chi_{k+3}^{-2}(\Theta)\}^2 I[\chi_{k+3}^2(\Theta) \leq (k - 3)]]], \end{aligned}$$

respectively.

We have plotted $ADQR_1$ to $ADQR_5$ versus Θ for $\pi = 0.5$ under different values of α for $k = 4$ and 10 in figures 3.1 to 3.2. The expectations in the ADQRs was calculated using Maple9.5, and the rest of the computation was carried out in S-plus. In order to have a fair comparison for different k values, a correction is done by dividing the risks by k . Delta in the figure corresponds to Θ/k . The graphs exhibit that all the estimators dominate $\hat{\sigma}^2$ at and near the origin (or for small values of Θ) only. More importantly, $\hat{\sigma}_{(PP)}^2$ and $\hat{\sigma}_{(JS)}^2$ outshine $\hat{\sigma}^2$ for all the values of Θ . Interestingly, $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}_{(PP)}^2$ and $\hat{\sigma}_{(JS)}^2$ for a range of α and for small values of Θ when k is small, say $k = 4$. However, when k increases, say $k = 10$, $\hat{\sigma}_{(SP)}^2$ loses all the advantages, and is dominated by $\hat{\sigma}_{(PP)}^2$ and $\hat{\sigma}_{(JS)}^2$ in the whole parameter space. Similar trends are observed for the relative performance of $\hat{\sigma}_{(PP)}^2$ and $\hat{\sigma}_{(JS)}^2$ compared to $\hat{\sigma}_{(SR)}^2$. The graphical analysis shows that increasing k enhances the dominance of $\hat{\sigma}_{(PP)}^2$ and $\hat{\sigma}_{(JS)}^2$ to the other estimators. Further, $\hat{\sigma}_{(PP)}^2$ dominates $\hat{\sigma}_{(JS)}^2$ in the whole parameter space, and the improvement by using the positive part of the $\hat{\sigma}_{(JS)}^2$ is very significant when Delta is small. More importantly, the components of $\hat{\sigma}_{(PP)}^2$ has the same sign as that of $\hat{\sigma}^2$. We notice that all the estimators have maximum risk reduction as compared to $\hat{\sigma}^2$ at $\Theta = 0$.

Using the numerical risks, we define the notion of *Asymptotic Relative Efficiency* (*ARE*) of an estimator, σ_*^2 , compared to another estimator σ_\diamond^2 by

$$ARE(\sigma_*^2 : \sigma_\diamond^2) = \frac{ADQR(\sigma_\diamond^2)}{ADQR(\sigma_*^2)}. \quad (3.30)$$

A value of *ARE* greater than 1 indicates the degree of superiority of σ_*^2 over σ_\diamond^2 . Thus, the asymptotic efficiency of the various proposed estimators relative to $\hat{\sigma}^2$, are

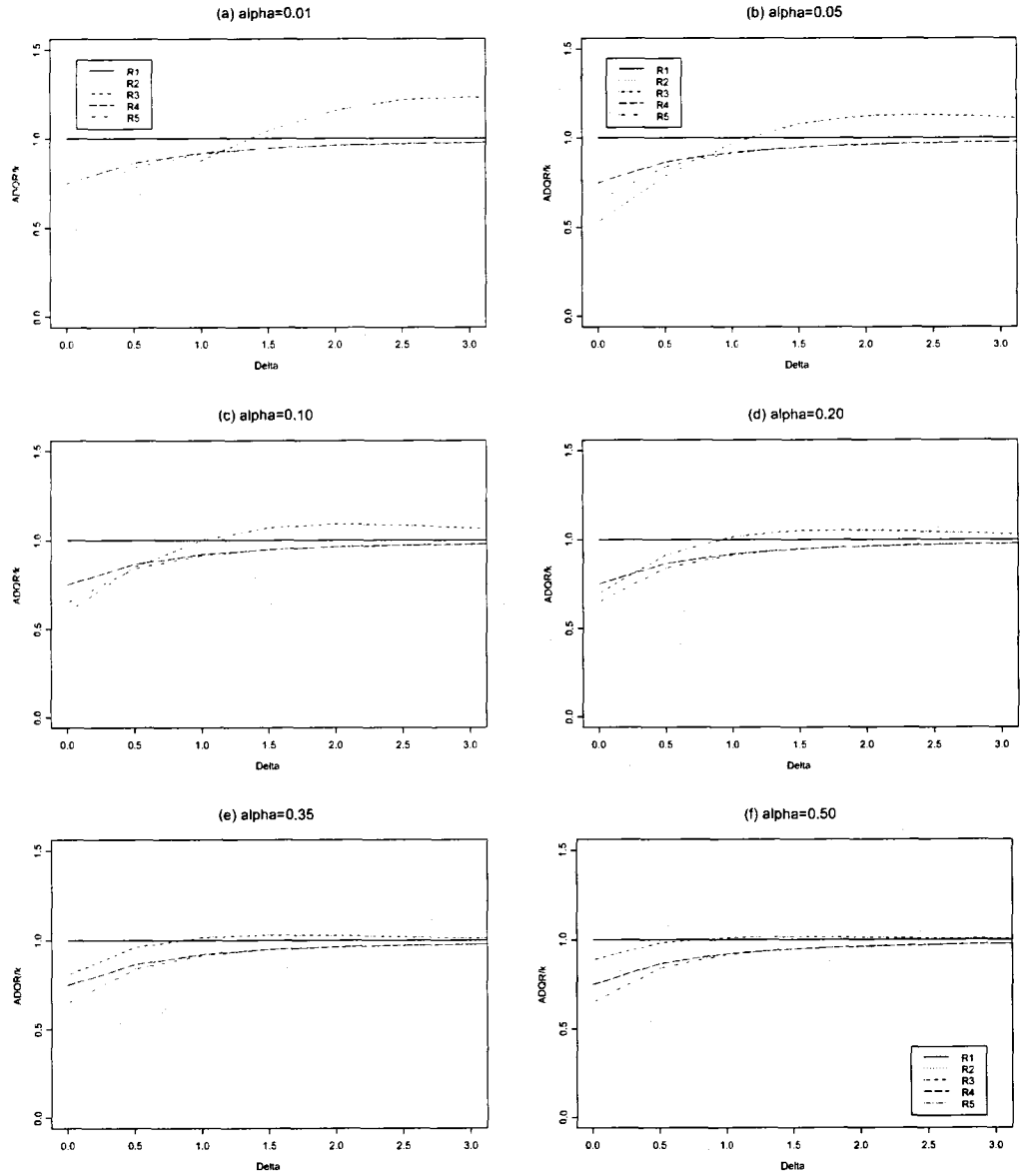


Figure 3.1: ADQR/k of the estimators at different α level when $k=4$

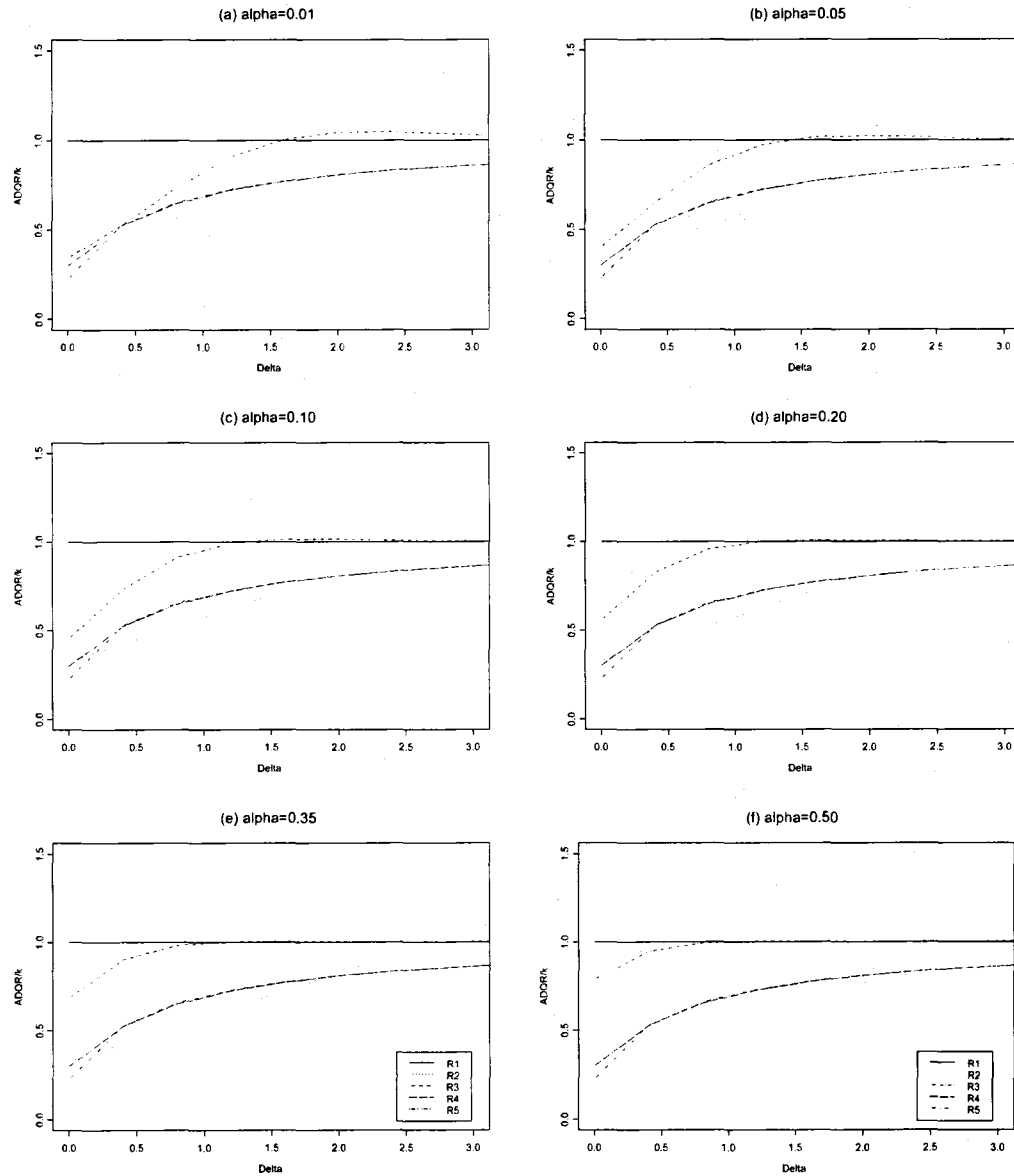


Figure 3.2: ADQR/k of the estimators at different α level when $k = 10$

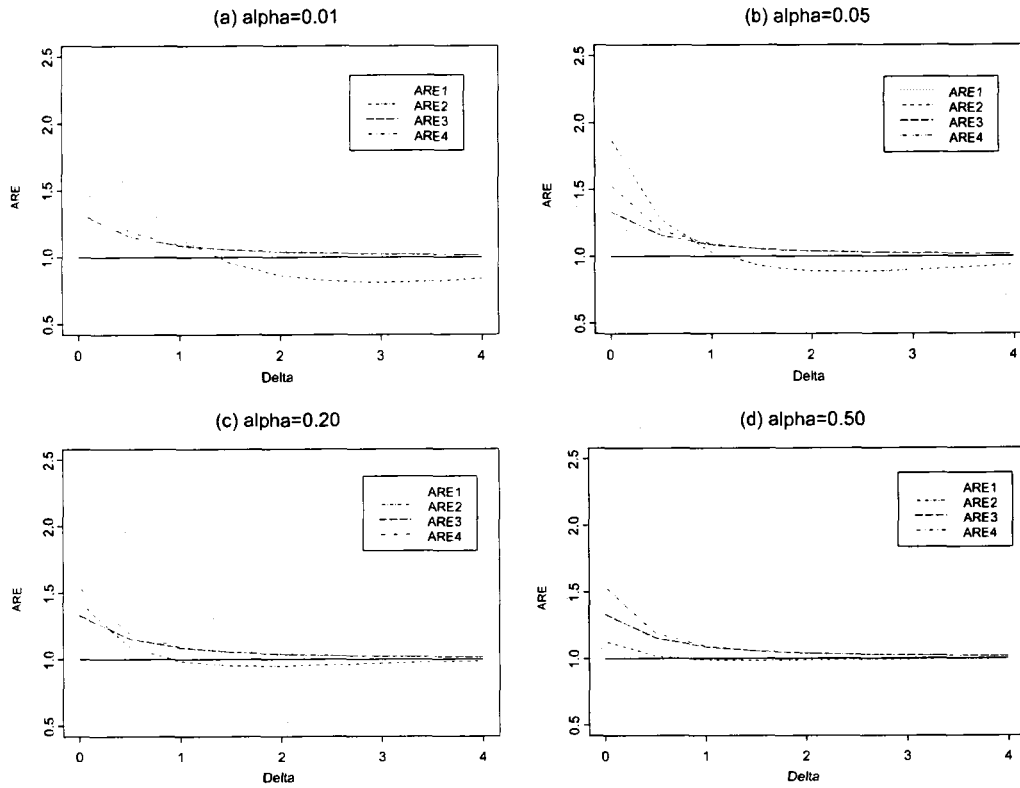


Figure 3.3: ARE of the estimators for $k = 4$

given by

$$ARE_{v-1} = \frac{ADQR_1}{ADQR_v}; \quad v = 2, 3, 4, 5.$$

Thus, ARE_1, ARE_2, ARE_3 , and ARE_4 are the efficiency of $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(SP)}^2$, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ respectively, relative to $\hat{\sigma}^2$. The AREs are plotted in figures 3.3 and 3.4.

From a practitioner perspective, the most important question is:

What is the adequate value of sample size n to apply the proposed methodology?

From a statistical perspective, the most important question is:

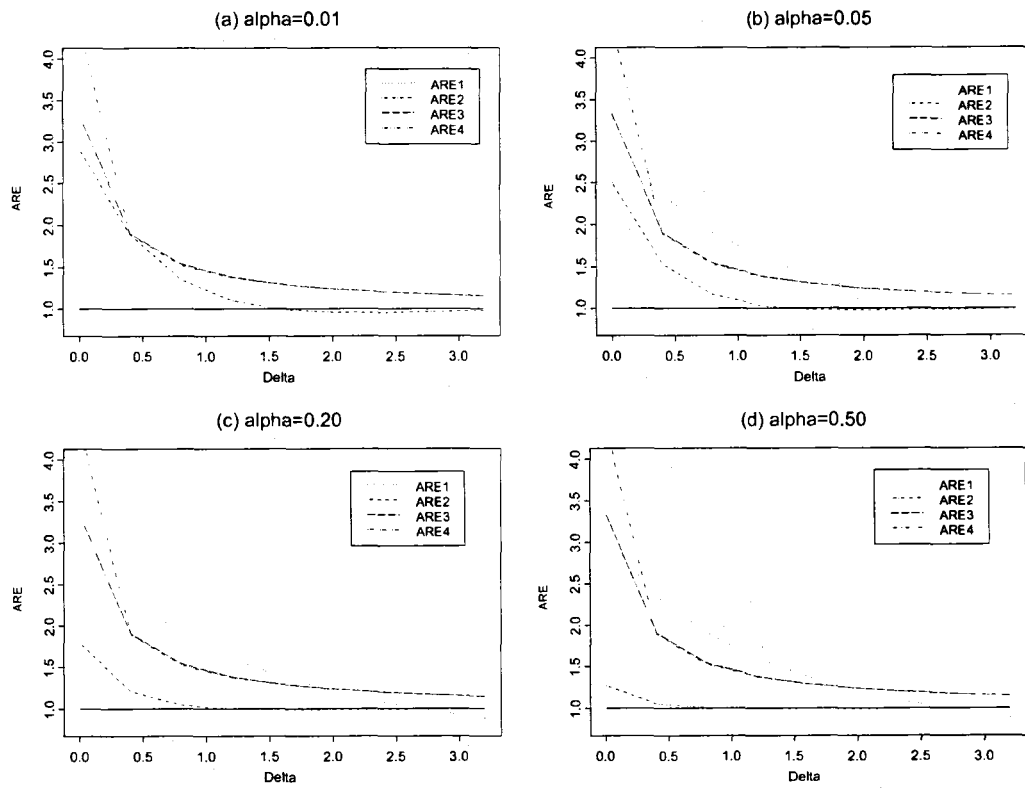


Figure 3.4: ARE of the estimators for $k = 10$

Does the loss function (squared) used to draw the inference make sense or folly?

In the following section by means of simulation we answer these questions. We conduct two simulation experiments using quadratic and entropy loss functions, respectively.

3.4 Simulation Results

The developed theoretical aspects of parameter estimation is examined in this section through Monte Carlo simulation under two loss functions. The main purpose of this simulation is to examine the quality of statistical inference based on large-sample methodology in moderate sample situations and to justify the use of quadratic loss function. The simulation study was conducted for two different population distributions - a normal distribution and a mixture of two normal distributions. All the simulation was carried out using S-plus.

First, we introduce the notation of the simulated relative efficiency. The *simulated relative efficiency (SRE)* of an estimator σ_*^2 to another estimator σ_\diamond^2 is defined by

$$SRE(\sigma_*^2 : \sigma_\diamond^2) = \frac{R(\sigma_\diamond^2)}{R(\sigma_*^2)},$$

where $R(\sigma_\diamond^2)$ and $R(\sigma_*^2)$ are the simulated risks of the estimator σ_*^2 and σ_\diamond^2 respectively. Again, keep in mind that a value of *SRE* greater than 1 indicates the degree of superiority of σ_*^2 over σ_\diamond^2 . Thus, the simulated efficiency of various proposed estimators relative to $\hat{\sigma}^2$ is given by:

$$SRE_{v-1} = \frac{R_1}{R_v}, \quad v = 2, 3, 4, 5,$$

where R_1, R_2, R_3, R_4 and R_5 are the simulated risks of $\hat{\sigma}^2, \hat{\sigma}_{(SR)}^2, \hat{\sigma}_{(SP)}^2, \hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$, respectively.

3.4.1 Under Quadratic Loss Function

In a k -sample setup, the risk of an estimator $\hat{\sigma}_*^2$ under the quadratic loss function, based on fixed sample sizes $n_i, i = 1, \dots, k$, is defined by

$$MSE(\hat{\sigma}_*^2) = E(\hat{\sigma}_*^2 - \sigma^2)'(\hat{\sigma}_*^2 - \sigma^2)$$

The expectation is estimated by the mean of 5000 Monte Carlo simulations.

Normal population The samples of sizes 20 and 50 are taken from normal populations with equal mean 0 and various variance σ_i^2 . For simplicity, only the case with equal sample sizes is considered in our study. Based on simulated values of $\hat{\sigma}^2$ and $\hat{\sigma}_{(R)}^2$, the test statistic Λ is computed first using (3.20); then $\hat{\sigma}_{(SP)}^2, \hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ are calculated accordingly. The distribution of the test statistic Λ has been simulated under the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ in chapter 2, and the simulation shows that Λ can be reasonably approximated by $\chi_{(k-1)}^2$ distribution for a moderate n . Therefore, the cut off points of distribution of the test-statistic is obtained. Further, we define the parameter $\Delta = \sum_{i=1}^k (\sigma_i^2 - \sigma_{(R)}^2)^2 / k$, which is a measure of the distance between the null hypothesis and the true value of the parameters. It depends on the values of σ_i^2 .

The efficiency of the various estimators is calculated based on 5,000 simulations for different choices of k and α . Tables 3.1 to 3.5 and Figures 3.5 to 3.6 provide the estimated relative efficiency for the various estimates over $\hat{\sigma}^2$ when $n = 20$ and 50, $k = 4$ and 10, and $\alpha = .05$. Simulation studies show that maximum efficiency of

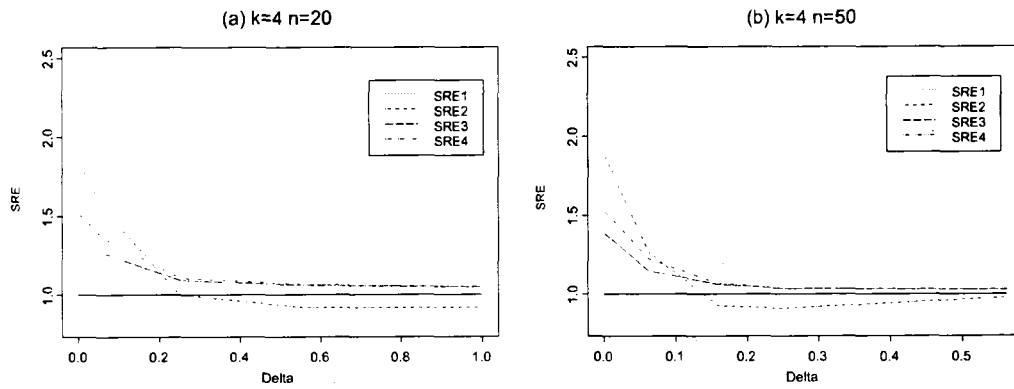


Figure 3.5: *SRE* performance comparison for $k = 4$, $\alpha = .05$ and various n when population is normal. SRE_1 , SRE_2 , SRE_3 , and SRE_4 represent the simulated relative efficiency of $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(SP)}^2$, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$, respectively.

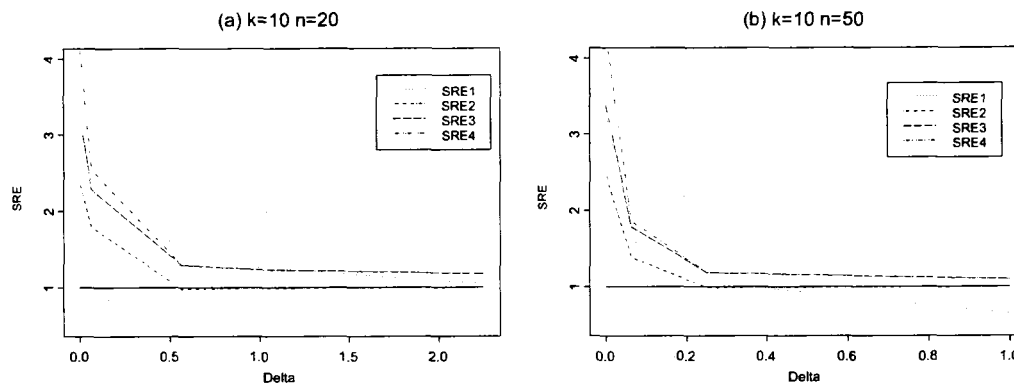


Figure 3.6: *SRE* comparison for $k = 10$, $\alpha = .05$ and various n when population is normal. SRE_1 , SRE_2 , SRE_3 , and SRE_4 represent the simulated relative efficiency of $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(SP)}^2$, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$, respectively.

all the estimators relative to $\hat{\sigma}^2$ occurred at $\Delta = 0$. It is apparent from these tables that $\hat{\sigma}_{(SR)}^2$ dominates the other three estimators near the null hypothesis. However, as Δ increases, the performance of $\hat{\sigma}_{(SR)}^2$ becomes the worst and hence it is not a desirable strategy. On the other hand, the performance of $\hat{\sigma}_{(SP)}^2$ is less sensitive for such departures, for instance, the relative efficiency achieves its maximum value at $\Delta = 0$, drops as Δ increases, to below 1 and then tends to 1. Further, when k is small, the relative efficiency of $\hat{\sigma}_{(SP)}^2$ is higher than that of $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ near the null hypothesis. However, when k is large, $\hat{\sigma}_{(SP)}^2$ does not show any advantage over $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ on the entire parameter space. It is obvious that for larger value of Δ , $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ are superior to all other estimators for all k values. More importantly, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ are superior to $\hat{\sigma}^2$ for all the values of Δ , and $\hat{\sigma}_{(PP)}^2$ dominates $\hat{\sigma}_{(JS)}^2$. In short, Tables 3.1 to 3.5 reveal that for Δ close to 0, all the proposed estimators are highly efficient relative to $\hat{\sigma}^2$; for larger values of Δ , the performance of the estimators is similar to the analysis of asymptotic provided in section 3.3. The advantage of $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ over other estimators is enlarged when k is large.

The simulation study also shows the effect of the sample size n on the *SRE*. It is observed that for a fixed value of k , decreasing n will slow down the change of the *SRE* with respect to Δ . For example, for $k = 4$, when $n_i = 50$, RE_1 decreases from 2.273 to 0.740 as Δ increases from 0 to 0.5625; but when $n_i = 20$, SRE_1 drops from about the same level (2.298) only to 1.262. Similar analysis applies to other estimators too. This observation allows us to conclude that our proposed pre-test and shrinkage estimators have more advantages in a wider range over the unrestricted estimator in

Table 3.1: SRE of the estimators under quadratic loss with $k = 4$, $\alpha = 0.05$, $n_i = 50$ when population is normal.

Δ	SRE_1	SRE_2	SRE_3	SRE_4
0.0000	2.273	1.891	1.392	1.532
0.1600	1.252	0.931	1.060	1.069
0.2500	0.956	0.912	1.039	1.039
0.5625	0.740	0.978	1.030	1.030

Table 3.2: SRE of the estimators under quadratic loss with $k = 4$, $\alpha = 0.05$, $n_i = 20$ when population is normal.

Δ	SRE_1	SRE_2	SRE_3	SRE_4
0.0000	2.298	1.846	1.359	1.522
0.2500	1.481	1.002	1.094	1.108
0.5625	1.262	0.918	1.063	1.067
1.0000	1.092	0.920	1.051	1.053

terms of efficiency for small samples.

We have also assessed the performance of $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ relative to $\hat{\sigma}_{(SP)}^2$ for larger size of the test, α . Tables 3.3 and 3.5 give us a rough idea how α affects the performance of the estimators. It is observed that only RE_2 changes with the change of α which is sensible since α is only involved in the construction of $\hat{\sigma}_{(SP)}^2$. It is seen that as α increases, $\hat{\sigma}_{(SP)}^2$ tends to $\hat{\sigma}^2$.

Finally, based on our simulation study we find that the trend of the relative dominance of the various estimators for a fixed sample size n is similar to the case of asymptotic, except that the curves are stretched out to the right when n is small.

Mixture of normal populations The samples of sizes 20 and 50 are taken from a mixture population of two k -variate normal distributions, i.e., 90% $N(\mathbf{0}, \mathbf{I}\sigma_1^2)$ and

Table 3.3: SRE of the estimators under quadratic loss with $k = 10$, $\alpha = 0.05$, $n_i = 50$ when population is normal.

Δ	SRE_1	SRE_2	SRE_3	SRE_4
0.0000	3.096	2.474	3.393	4.340
0.0625	2.238	1.385	1.787	1.865
0.2500	1.065	0.990	1.188	1.188
1.0000	0.652	1.000	1.104	1.104

Table 3.4: SRE of the estimators under quadratic loss with $k = 10$, $\alpha = 0.05$, $n_i = 20$ when population is normal.

Δ	SRE_1	SRE_2	SRE_3	SRE_4
0.0000	3.065	2.376	3.268	4.137
0.5625	1.451	0.981	1.294	1.294
1	1.264	0.993	1.235	1.235
4	0.939	1	1.150	1.150

Table 3.5: SRE of the estimators under quadratic loss with $k = 10$, $\alpha = 0.20$, $n_i = 50$ when population is normal.

Δ	SRE_1	SRE_2	SRE_3	SRE_4
0.0000	3.082	1.767	3.323	4.359
0.0625	2.240	1.136	1.771	1.852
0.2500	1.063	0.999	1.192	1.192
1.000	0.654	1.000	1.104	1.104

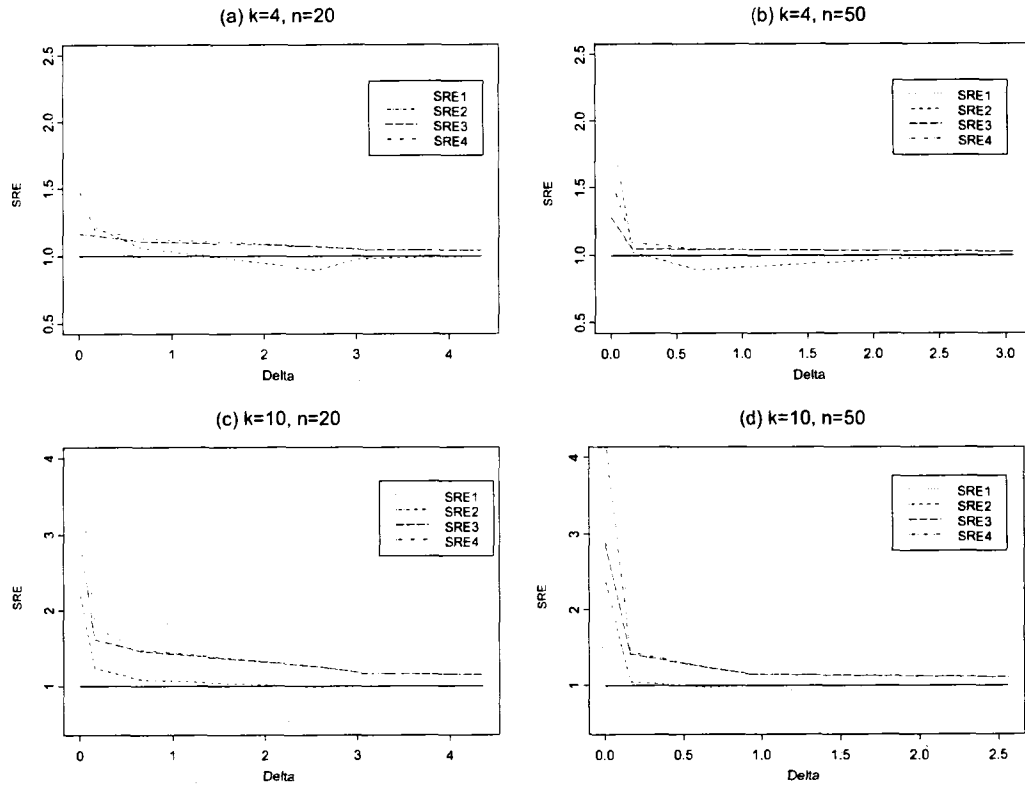


Figure 3.7: *SRE* comparison for mixture of normal distributions

10% $N(\mathbf{0}, \mathbf{I}\sigma_2^2)$, where $\sigma_i^2 = (\sigma_{i1}^2, \dots, \sigma_{ik}^2)$, $i = 1, 2$. Changing the values of σ_1^2 and σ_2^2 will change the covariance matrix of the mixture k -variate distribution and eventually change the values of Δ . The simulation procedure is analogous to that for normal population. The simulated relative efficiencies are plotted in Figures 3.7 to 3.8.

The plots of simulated relative risk from mixture samples presented in Figures 3.7 and 3.8 are very similar to those in Figures 3.5 and 3.6 from normal samples. The maximum efficiency of all estimators relative to $\hat{\sigma}^2$ is achieved at $\Delta = 0$; it decreases as Δ increases. $\hat{\sigma}_{(SR)}^2$ and $\hat{\sigma}_{(SP)}^2$ perform well when the null hypothesis is true or nearly true; however, it gets worse as the true status departs from H_0 and the

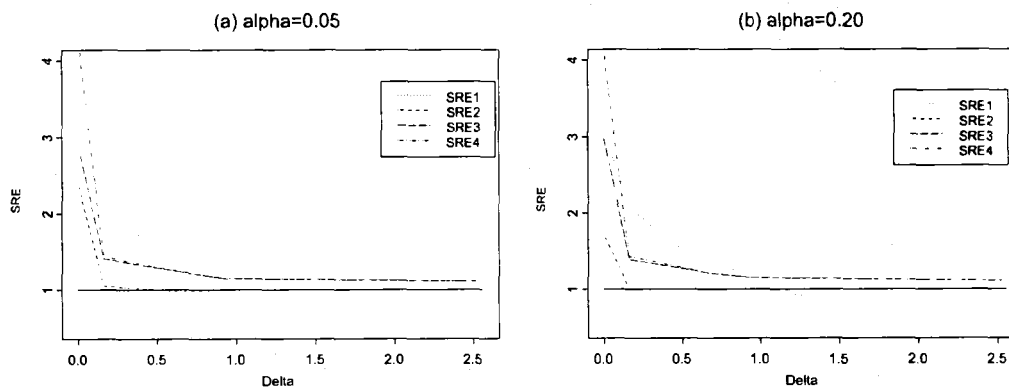


Figure 3.8: *SRE* comparison for mixture of normal distributions when α differs

SRE becomes lower than 1. $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$, on the other hand, dominate $\hat{\sigma}^2$ in the entire parameter space. The advantages of $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ over other estimators are especially extended when k is large.

3.4.2 Under Entropy Loss Function

Quadratic risk, or mean squared error (MSE) is a reasonable criterion when comparing estimators of a location parameter, but may not be a good idea for scale parameters. One problem is that MSE penalizes equally for overestimation and underestimation which is fine in the location case. However, in the scale case, "0" is a natural lower bound, the estimation problem is not symmetric. Use of MSE tends to be forgiving of underestimation. Maatta and Casella (1990) used a scaled version of squared error loss - $L(\tilde{\sigma}^2, \sigma^2) = \frac{1}{\sigma^2}(\tilde{\sigma}^2 - \sigma^2)^2$. However, it does not solve the mentioned problem.

Stein (1964) found a loss function for which the usual unbiased estimator of variance $\hat{\sigma}^2$ is best invariant. This loss function is called entropy loss function and defined

as:

$$L_{En}(\hat{\theta}, \theta) = \frac{\hat{\theta}}{\theta} - 1 - \ln \frac{\hat{\theta}}{\theta} \quad (3.31)$$

Note that $L_{En}(\hat{\theta}, \theta) \geq 0$ and attains the value 0 uniquely at $\hat{\theta} = \theta$. Also, $L_{En}(\hat{\theta}, \theta)$ is strictly convex in $\hat{\theta}$ and $\lim_{\hat{\theta} \rightarrow \infty} L_{En}(\hat{\theta}, \theta) = \lim_{\hat{\theta} \rightarrow 0} L_{En}(\hat{\theta}, \theta) = \infty$. Gross underestimation is penalized just as heavily as gross overestimation. Brown (1968) and Brown (1990) discussed a number of nice properties of $L_{En}(\hat{\theta}, \theta)$ when estimating a single scale parameter. Brown (1990) stated, “My own feeling is that the loss L_{En} is the most appropriate for general studies of estimation of scale parameters.”

In a k-sample situation, let Σ be the covariance matrix and $\hat{\Sigma}$ be an estimator of Σ . Then the entropy loss when estimating Σ by $\hat{\Sigma}$ is defined as:

$$L_{En}(\hat{\Sigma}, \Sigma) = \text{tr}(\hat{\Sigma}\Sigma^{-1}) - \log \det(\hat{\Sigma}\Sigma^{-1}) - p. \quad (3.32)$$

In our application, the entropy risk is then calculated by

$$R_{En} = E(L_{En}) = E \left[\sum_{i=1}^k \left(\frac{\hat{\sigma}_{(*)i}^2}{\sigma_i^2} - 1 - \ln \frac{\hat{\sigma}_{(*)i}^2}{\sigma_i^2} \right) \right] \quad (3.33)$$

The relative efficiency of the estimators based on entropy risk is defined in the same way as in previous subsection, and denoted by REE_1 to REE_4 for $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(SP)}^2$, $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ respectively. The simulation is carried out analogous to section 3.4.1. The sampling procedures from two different distributions (normal and mixture of two normals) are as described previously.

The simulation result from normal populations is shown in tables 3.6 to 3.10 and figures 3.9 to 3.10.

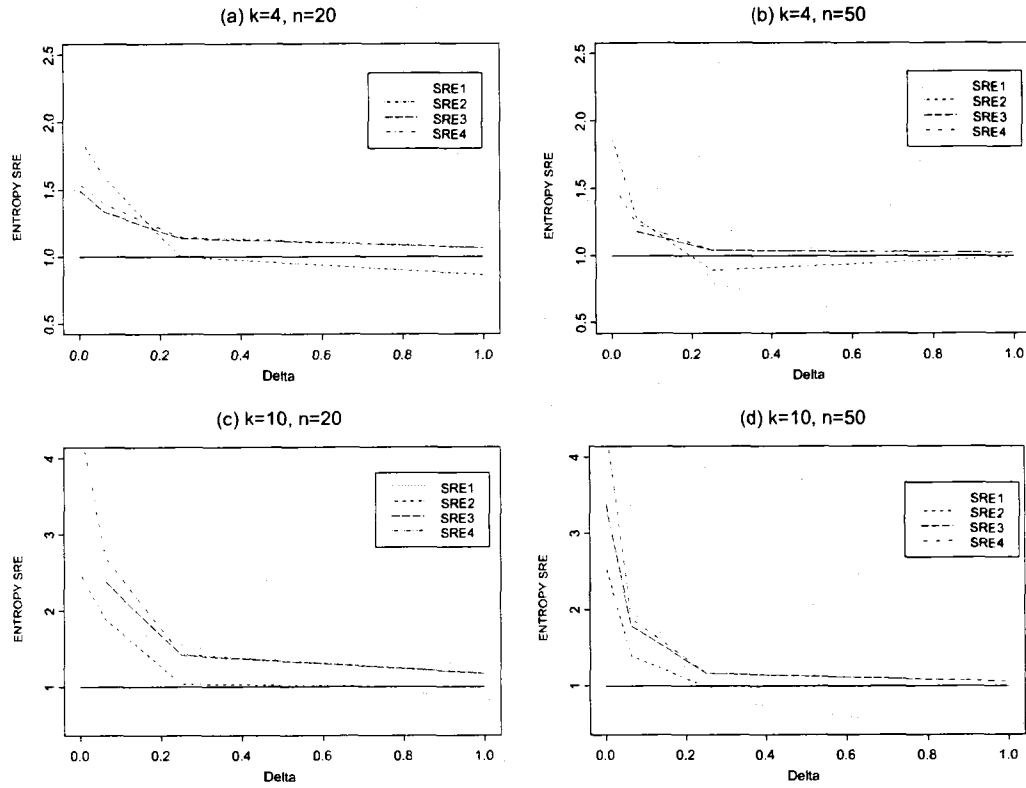


Figure 3.9: SRE comparison under entropy loss for normal population

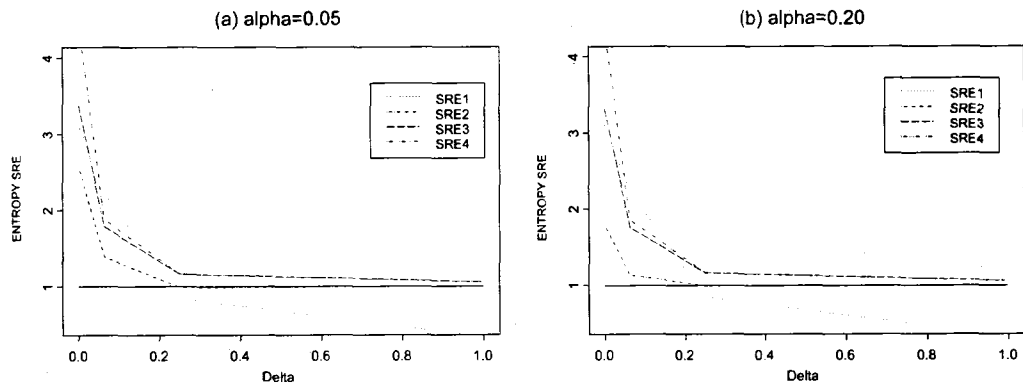


Figure 3.10: SRE under entropy loss for normal population under different α

Table 3.6: SRE of the estimators under entropy loss with $k = 4$, $\alpha = 0.05$, $n_i = 20$ when population is normal.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	2.384	1.886	-	1.542
0.0625	2.073	1.589	1.335	1.393
0.2500	1.350	1.000	1.137	1.148
1.000	0.690	0.863	1.066	1.067

“-” stands for a negative variance estimate due to overshrinking.

Table 3.7: SRE of the estimators under entropy loss with $k = 4$, $\alpha = 0.05$, $n_i = 50$ for normal data.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	2.316	1.890	-	1.546
0.0625	1.765	1.274	1.182	1.237
0.2500	0.792	0.891	1.042	1.043
1.000	0.325	0.992	1.023	1.023

“-” stands for a negative variance estimate due to overshrinking.

The simulation shows that the relative efficiencies of the estimators are very similar under two different loss functions. The dominance relation among the estimators follow the same trend in both cases even though the curves of the risks versus Δ for individual estimators (not shown here) are very different under two loss functions. For instance, $MSE(\hat{\sigma}_{(SR)}^2)$ is linearly correlated to Δ , whereas $Risk(\hat{\sigma}_{(SR)}^2)$ vs. Δ is not a straight line under entropy loss function. However, these details seem not affect the efficiency of the estimators relative to $\hat{\sigma}^2$. Note that there are some missing values for the efficiency of $\hat{\sigma}_{(JS)}^2$ when $\Delta = 0$, this is because negative values of $\hat{\sigma}_{(JS)i}^2$ were calculated due to over shrinking in the simulation, causing undefined values of logarithmic. This, in general, cannot be avoided when Δ is small.

Tables 3.11 to 3.14 present the simulated relative efficiencies of the estimators under entropy loss function for mixtures of normal distributions. The relative perfor-

Table 3.8: SRE of the estimators under entropy loss with $k = 10$, $\alpha = 0.05$, $n_i = 50$ for normal data.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	3.122	2.559	3.387	4.435
0.0625	2.235	1.394	1.791	1.886
0.2500	0.843	0.990	1.164	1.164
1.000	0.333	1.000	1.059	1.059

Table 3.9: SRE of the estimators under entropy loss with $k = 10$, $\alpha = 0.05$, $n_i = 20$ for normal data.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	3.192	2.482	-	4.413
0.0625	2.756	1.889	2.380	2.688
0.2500	1.557	1.037	1.417	1.430
1.000	0.749	0.988	1.173	1.173

“-” stands for a negative variance estimate due to overshrinking.

Table 3.10: SRE of the estimators under entropy loss with $k = 10$, $\alpha = 0.20$, $n_i = 50$ when population is normal.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	3.109	1.787	3.319	4.360
0.0625	2.223	1.139	1.761	1.862
0.2500	0.853	0.998	1.169	1.169
1.000	0.332	1.000	1.064	1.064

Table 3.11: SRE under entropy loss for mixture of normal data with $k = 4$, $\alpha = 0.05$, $n_i = 20$.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	2.274	1.911	-	1.610
0.0625	1.984	1.688	-	1.519
0.3600	1.598	1.380	-	1.346
1.1025	1.291	1.155	-	1.215
1.4400	0.955	0.926	-	1.068

“-” stands for a negative variance estimate due to overshrinking.

Table 3.12: SRE under entropy loss for mixture of normal data with $k = 4$, $\alpha = 0.05$, $n_i = 50$.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	2.253	1.898	-	1.560
0.0625	1.864	1.473	-	1.355
0.3600	1.346	1.097	-	1.150
1.1025	0.995	0.933	-	1.070
1.4400	0.645	0.840	1.000	1.010

“-” stands for a negative variance estimate due to overshrinking.

mance of the estimators is very similar to that from normal population. The trend remains the same even when the sample size is as low as 20. An unusual phenomenon we observed in this simulation is that the observations in the column of REE_3 are mostly missing. This is due to the “over-shrinking” that caused some negative estimates of the variance. This shows the importance of the improved positive-rule Stein-type estimator from a different aspect.

3.5 Concluding Notes

Several estimation strategies for pooling data are presented for estimating the population variances and their risks are studied under different loss functions. It is concluded that the positive rule Stein-type estimator dominates the usual shrinkage

Table 3.13: SRE under entropy loss for mixture of normal samples with $k = 10$. $\alpha = 0.05$, $n_i = 50$.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	3.052	2.385	-	4.201
0.0625	2.419	1.626	-	2.220
0.3600	1.677	1.138	-	1.423
1.1025	1.196	0.999	-	1.187
1.4400	0.751	0.965	1.021	1.022

“-” stands for a negative variance estimate due to overshrinking.

Table 3.14: SRE under entropy loss for mixture of normal samples with $k = 10$. $\alpha = 0.05$, $n_i = 20$.

Δ	REE_1	REE_2	REE_3	REE_4
0.0000	3.114	2.452	-	4.419
0.0625	2.667	1.922	-	2.886
0.3600	2.117	1.463	-	1.909
1.1025	1.669	1.228	-	1.495
1.4400	1.191	1.026	-	1.167

“-” stands for a negative variance estimate due to overshrinking.

type estimator and they both dominate the unrestricted estimator $\hat{\sigma}^2$ in terms of the asymptotic distributional quadratic risk we defined. On the other hand, the performance of restricted estimator heavily depends on the quality of non-data information. The $\hat{\sigma}_{(JS)}^2$ has a disquieting feature that it may shrink beyond the hypothesis vector. We have improved $\hat{\sigma}_{(JS)}^2$ by $\hat{\sigma}_{(PP)}^2$. The positive-rule estimator is particularly important to control the over-shrinking inherent in the shrinkage estimator. The performance of $\hat{\sigma}_{(R)}^2$, $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(PT)}^2$ and $\hat{\sigma}_{(SP)}^2$ depends on the value of Θ , i.e., the correctness of the null hypothesis. It is shown that the range in which $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}^2$ is wider than the range in which PTE dominates UE .

A risk simulation is conducted for finite samples under the usual quadratic loss and the entropy loss functions. The simulation result indicates that moderate sample

sizes are sufficient enough for the proposed estimators to follow the same dominance pattern (based on asymptotic risk) in practical situations. Though the estimators are proposed based on the quadratic risk, they perform similarly under the entropy loss function in terms of the relative efficiency. Finally, it should be noted that $\hat{\sigma}_{(JS)}^2$ and $\hat{\sigma}_{(PP)}^2$ can only be used for $k > 3$.

Chapter 4

Small-Sample Risk Analysis of Variance Estimation with Application in Differential Gene Expression

4.1 Introduction

Microarray technology has become an important tool for obtaining quantitative measurements for the expression of thousands of genes present in a biological sample simultaneously. DNA microarrays have been used to monitor changes in gene expression during important biological processes (e.g., cellular replication and the response to changes in the environment), and to study variation in gene expression across collections of related samples (e.g., tumor samples from patients with cancer). In order to compare two types of cells (e.g., a cancer cell versus a normal cell), the DNA materials are extracted from both cell types; one is labeled with fluorescence cy5 (red) and the other with cy3 (green). The microarray which provides a large-scale medium for matching known and unknown DNA segments based on base-pairing rules is then exposed to the mixture of the two DNA samples for hybridization. When mRNA for

a gene is more abundant in the cancer cell than in the normal cell, for example, the array spot corresponding to that gene will show a red color. Numerically, a vector of length G is reported, where G is the number of spots (genes) on the array, and each entry of the vector records the ratios (or log ratios) of the fluorescence intensities ($cy5/cy3$). When more than two types of cells are considered, the microarray data often takes the form of a $G \times p$ matrix, where each column corresponds to a cell type (e.g., lymphoma cell, leukemia cell, and normal cell) or a treatment, and each row corresponds to a gene. Thus, through the use of DNA microarrays, one can monitor simultaneously the expression levels of thousands of genes in different types of cells. We refer to Tusher *et al.* (2001), Tibshirani *et al.* (2002), Cherepinsky *et al.* (2003) among others for a detailed discussion of the micro-array experiments. A major statistical task is to understand the structure of the data from such studies, and the first level of analysis requires determining whether observed differences in expression are significant or not. Numerous statistical methods have been developed to deal with this task in recent years. We refer to Efron (2004), Efron *et al.* (2001), Ji *et al.* (2006), Datta and Datta (2005), Klebanov *et al.* (2006) and Cui *et al.* (2005) among others.

One commonly used approach in the current literature is the fold change approach, in which a gene is declared to have significantly changed if its average expression level varies by more than a constant factor, typically 2, between the treatment and control conditions. This simple 2-fold rule is unlikely to yield optimal results because it neglects the random sampling variation.

Another approach is to use a t-test which is more precise in a statistical sense (Broberg (2003)). However, a problem with microarray experiments is that the repli-

cation numbers are usually very small, hence the variance estimates obtained from an individual gene may be very imprecise, and therefore the inferences reached may not be trustworthy due to the weak variance estimates. Baldi and Long (2001) proposed a regularized t-test replacing the usual variance estimate with a hierarchical Bayes estimator which compares favorably with simple t-test or fold methods in terms of consistency. The SAM t-test proposed by Tusher *et al.* (2001) adds a small constant to the gene-specific variance estimate in order to stabilize the small variances. The idea is to eliminate some false positives associated with low values of variance. The technique is described in detail by Chu *et al.* (n.d.). Cui *et al.* (2005) proposed a James-Stein type estimator and used it to construct an F-like test statistic. It is favorable compared to other F-like statistics which are based on a gene-based estimator or a pooled estimator across all G genes in terms of power, false positive rate, and robustness.

In chapter 3, we have constructed several estimators of population variance vector σ^2 based on pretest and shrinkage rules. These estimators, basically, are all weighted combinations of two estimators - the unrestricted estimator, $\hat{\sigma}^2$, and the pooled estimator $\hat{\sigma}_{(R)}^2$. The asymptotic distributional quadratic risk (ADQR) of the proposed estimators were obtained and the risks of the estimators were simulated for finite samples under quadratic and entropy loss functions. The simulation study showed that the James-Stein type estimators dominate others in a wide range of parameter space; it especially showed substantial advantages when the number of samples k was large and n was small. In this chapter, we study the risks of these combination estimators for finite samples. Two optimal weight estimators are proposed under the

quadratic loss function in section 4.2; some theoretical results are obtained in section 4.3 and the simulation study is included in section 4.4. An application of the estimation strategies in the microarray study is included in the last section.

4.2 Two Optimal Weight Combination Estimators (OWCE)

In chapter 3, we have defined the unrestricted estimator, $\hat{\sigma}^2$, and the restricted estimator $\hat{\sigma}_{(R)}^2$ of $\sigma^2 = (\sigma_1^2 \cdots, \sigma_k^2)'$, as

$$\hat{\sigma}^2 = (\hat{\sigma}_1^2 \cdots, \hat{\sigma}_k^2)', \text{ where } \hat{\sigma}_i^2 = s_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1},$$

and

$$\hat{\sigma}_{(R)}^2 = (\hat{\sigma}_{(R)}^2 \cdots, \hat{\sigma}_{(R)}^2)' = \hat{\sigma}_{(R)}^2 \mathbf{1}_k, \quad \hat{\sigma}_{(R)}^2 = \sum_{i=1}^k n_i \hat{\sigma}_i^2 / n, \quad n = \sum_{i=1}^k n_i,$$

respectively. Some weighted combinations of these two estimators have been studied. The optimal weights for the two components will be estimated under some assumptions in this section.

4.2.1 The OWC Estimator Based on the Original Data

Without any distributional assumption, we have that

$$\hat{\sigma}^2 \sim (\sigma^2, \Psi), \tag{4.1}$$

and

$$\hat{\sigma}_{(R)}^2 \sim (\sigma^2 + \gamma, \Phi), \tag{4.2}$$

where γ is a $(k \times 1)$ bias vector and Ψ and Φ are positive-semi-definite covariance matrices. We also define

$$\text{cov} \begin{bmatrix} \hat{\sigma}^2 \\ \hat{\sigma}_{(R)}^2 \end{bmatrix} = \begin{bmatrix} \Psi & \Sigma \\ \Sigma' & \Phi \end{bmatrix}. \quad (4.3)$$

Our objective is to find the optimal weight for the combination of the two estimators with smaller quadratic risk than the estimator $\hat{\sigma}^2$. Following the idea of Judge and Mittelhammer (2004), we define a new estimator

$$\hat{\sigma}^2(\alpha) = \alpha \hat{\sigma}^2 + (1 - \alpha) \hat{\sigma}_{(R)}^2. \quad (4.4)$$

The quadratic risk of $\hat{\sigma}^2(\alpha)$ is given by

$$\begin{aligned} \text{MSE}(\hat{\sigma}^2(\alpha)) &= E[(\alpha(\hat{\sigma}^2 - \sigma^2) + (1 - \alpha)(\hat{\sigma}_{(R)}^2 - \sigma^2))' \\ &\quad (\alpha(\hat{\sigma}^2 - \sigma^2) + (1 - \alpha)(\hat{\sigma}_{(R)}^2 - \sigma^2))] \\ &= \alpha^2 \text{tr}(\Psi) + (1 - \alpha)^2 [\text{tr}(\Phi) + \gamma' \gamma] + 2\alpha(1 - \alpha) \text{tr}(\Sigma). \end{aligned} \quad (4.5)$$

To minimize $\text{MSE}(\hat{\sigma}^2(\alpha))$, let

$$\frac{\partial \text{MSE}(\hat{\sigma}^2(\alpha))}{\partial \alpha} = 0. \quad (4.6)$$

Solving (4.6) yields the first-order necessary condition for α :

$$\begin{aligned} \alpha^* &= 1 - \frac{\text{tr}(\Psi) - \text{tr}(\Sigma)}{\gamma' \gamma + \text{tr}(\Psi) + \text{tr}(\Phi) - 2\text{tr}(\Sigma)} \\ &= \frac{\gamma' \gamma + \text{tr}(\Phi) - \text{tr}(\Sigma)}{\gamma' \gamma + \text{tr}(\Psi) + \text{tr}(\Phi) - 2\text{tr}(\Sigma)}. \end{aligned} \quad (4.7)$$

Since $\partial^2 \text{MSE}(\hat{\sigma}^2(\alpha)) / \partial \alpha^2 > 0$, the optimal weight combination estimator $\hat{\sigma}^2(\alpha^*) = \alpha^* \hat{\sigma}^2 + (1 - \alpha^*) \hat{\sigma}_{(R)}^2$ is superior to the unrestricted estimator under quadratic loss.

When the null hypothesis in (3.2) is true, $\gamma = \mathbf{0}$ and $\alpha^* = 0$; the optimal estimator would be $\hat{\sigma}_{(R)}^2$.

Assuming normality and independence, we have

$$\Psi = \begin{pmatrix} \frac{2\sigma_1^4}{n_1-1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{2\sigma_2^4}{n_2-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \frac{2\sigma_k^4}{n_k-1} \end{pmatrix}, \quad (4.8)$$

and

$$\Phi = \frac{2}{k^2} \times \sum_{i=1}^k \frac{\sigma_i^4}{n_i-1} \mathbf{1}_{k \times k}, \quad (4.9)$$

where $\mathbf{1}_{k \times k}$ is a $k \times k$ matrix with all elements 1.

$$\Sigma = \frac{2}{k} \begin{pmatrix} \sigma_1^4/n_1 - 1 & \cdots & \sigma_1^4/n_1 - 1 \\ \sigma_2^4/n_2 - 1 & \cdots & \sigma_2^4/n_2 - 1 \\ \vdots & & \vdots \\ \sigma_k^4/n_k - 1 & \cdots & \sigma_k^4/n_k - 1 \end{pmatrix}_{k \times k}. \quad (4.10)$$

When the sample sizes are all equal, the above expressions are simplified to

$$\Psi = \frac{2}{n-1} \begin{pmatrix} \sigma_1^4 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^4 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_k^4 \end{pmatrix}, \quad (4.11)$$

and

$$\Phi = \frac{2 \sum_{i=1}^k \sigma_i^4}{(n-1)k^2} \mathbf{1}_{k \times k}, \quad (4.12)$$

and

$$\Sigma = \frac{2}{k(n-1)} \begin{pmatrix} \sigma_1^4 & \cdots & \sigma_1^4 \\ \sigma_2^4 & \cdots & \sigma_2^4 \\ \vdots & & \vdots \\ \sigma_k^4 & \cdots & \sigma_k^4 \end{pmatrix}_{k \times k} \quad (4.13)$$

Note that

$$\begin{aligned} E[(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)'(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)] &= E[(\hat{\sigma}^2 - \sigma^2)'(\hat{\sigma}^2 - \sigma^2)] \\ &\quad + E[(\hat{\sigma}_{(R)}^2 - \sigma^2)'(\hat{\sigma}_{(R)}^2 - \sigma^2)] - 2E[(\hat{\sigma}_{(R)}^2 - \sigma^2)'(\hat{\sigma}^2 - \sigma^2)] \\ &= tr(\Psi) + [tr(\Phi) + \gamma'\gamma] - 2tr(\Sigma). \end{aligned} \quad (4.14)$$

Thus, $(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)'(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)$ is an unbiased estimator of the denominator term in the α^* expression (4.7). Regarding the numerator term,

$$tr(\Psi) - tr(\Sigma) = \frac{2(k-1)}{(n-1)k} ((\sigma^2)'\sigma^2).$$

Since $\hat{\sigma}^2$ is an unbiased estimator of σ^2 , we define an estimator of the optimal weight α^* in the form

$$\hat{\alpha}^* = 1 - \frac{\frac{2(k-1)}{(n-1)k} ((\hat{\sigma}^2)'\hat{\sigma}^2)}{(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)'(\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2)}; \quad (4.15)$$

yielding the corresponding optimal weight combination estimator

$$\begin{aligned} \hat{\sigma}^2(\hat{\alpha}^*) &= \hat{\sigma}^2 + \frac{2(k-1)(\hat{\sigma}^2)'\hat{\sigma}^2}{(n-1)k\|\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2\|^2} (\hat{\sigma}_{(R)}^2 - \hat{\sigma}^2) \\ &= \hat{\sigma}_{(R)}^2 - \left(1 - \frac{2(k-1)(\hat{\sigma}^2)'\hat{\sigma}^2}{(n-1)k\|\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2\|^2}\right) (\hat{\sigma}_{(R)}^2 - \hat{\sigma}^2). \end{aligned} \quad (4.16)$$

This is a James-Stein type estimator; however, it is not a convex combination of the two estimators. In order to keep the weights in $[0, 1]$ and avoid overshinking,

it is necessary to take the positive part of the weight. Thus, we obtain the positive James-Stein type estimator,

$$\hat{\sigma}^2(\hat{\alpha}^*)^+ = \hat{\sigma}_{(R)}^2 - \left(1 - \frac{2(k-1)(\hat{\sigma}^2)' \hat{\sigma}^2}{(n-1)k \|\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2\|^2} \right)^+ (\hat{\sigma}_{(R)}^2 - \hat{\sigma}^2), \quad (4.17)$$

where for any number a , a^+ denotes $\max(a, 0)$.

4.2.2 An OWC Estimator Based on Log-transformed Data

In section 4.2.1 we have defined an OWC estimator of the variance vector. Since the covariance matrix of the estimator depends on the the parameter itself, we have to replace the parameter σ^2 by its unbiased estimator $\hat{\sigma}^2$ in the estimation of the optimal weight $\alpha(\ast)$. However, a log-transformation to the parameters will overcome this problem.

Let

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_k \end{pmatrix} = \begin{pmatrix} \ln \sigma_1^2 \\ \ln \sigma_2^2 \\ \vdots \\ \ln \sigma_k^2 \end{pmatrix}.$$

We assume normality and independence, and consider equal sample size case only for mathematical simplicity. Then $s_i^2 \sim \sigma_i^2 \cdot \chi_\nu^2 / \nu$, where χ_ν^2 is a chi-squared random variable with ν degrees of freedom. Thus, $\ln s_i^2 \sim \ln \sigma_i^2 + \ln \frac{\chi_\nu^2}{\nu}$. Let

$$E \left(\ln \frac{\chi_\nu^2}{\nu} \right) = m,$$

and

$$\text{Var} \left(\ln \frac{\chi_\nu^2}{\nu} \right) = V.$$

Then

$$E(\ln s_i^2) = \ln \sigma_i^2 + m,$$

and

$$\text{Var}(\ln s_i^2) = V.$$

m and V depend only on the degrees of freedom $\nu = n - 1$, and can be evaluated numerically through integration. Using a second order Taylor Expansion.

$$\begin{aligned} \ln \frac{\chi^2}{\nu} &\approx \frac{1}{\nu}(\chi_\nu^2 - \nu) - \frac{1}{2\nu^2}(\chi_\nu^2 - \nu)^2 \\ m &\approx E\left(\frac{1}{\nu}(\chi_\nu^2 - \nu) - \frac{1}{2\nu^2}(\chi_\nu^2 - \nu)^2\right) = -\frac{1}{\nu} \\ V &\approx \text{Var}\left(\frac{1}{\nu}(\chi_\nu^2 - \nu) - \frac{1}{2\nu^2}(\chi_\nu^2 - \nu)^2\right) \\ &= \frac{1}{\nu^2}\text{Var}(\chi_\nu^2) + \frac{1}{4\nu^4}\text{Var}[(\chi_\nu^2 - \nu)^2] \\ &= \frac{2}{\nu} + \frac{1}{4\nu^4}(\mu_4 - \mu_2^2) = \frac{2}{\nu} + \frac{1}{4\nu^4}(48\nu + 12\nu^2 - 4\nu^2) \\ &= \frac{2}{\nu} + \frac{2}{\nu^2} + \frac{12}{\nu^3}. \end{aligned}$$

The values of m and V and their second order approximations for some selected ν are listed in Table 4.1. Thus, we define an unbiased estimator of the parameter τ in the following form:

$$\hat{\tau} = \begin{pmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \\ \vdots \\ \hat{\tau}_k \end{pmatrix} = \begin{pmatrix} \ln s_1^2 - m \\ \ln s_2^2 - m \\ \vdots \\ \ln s_k^2 - m \end{pmatrix} \sim (\tau, \Psi), \tag{4.18}$$

and a pooled estimator

$$\hat{\tau}_{(R)} = \begin{pmatrix} \hat{\tau}_{(R)} \\ \hat{\tau}_{(R)} \\ \vdots \\ \hat{\tau}_{(R)} \end{pmatrix} = \begin{pmatrix} \overline{lns^2} - m \\ \overline{lns^2} - m \\ \vdots \\ \overline{lns^2} - m \end{pmatrix} \sim (\tau + \gamma, \Phi), \quad (4.19)$$

where $\overline{lns^2} = \sum n_i lns_i^2/n$.

Define

$$cov \begin{bmatrix} \hat{\tau} \\ \hat{\tau}_{(R)} \end{bmatrix} = \begin{bmatrix} \Psi & \Sigma \\ \Sigma' & \Phi \end{bmatrix}. \quad (4.20)$$

Then

$$\Psi = Cov(\hat{\tau}) = V\mathbf{I}_k, \quad (4.21)$$

$$\Phi = Cov(\hat{\tau}_{(R)}) = \frac{V}{k^2} \mathbf{1}_k \mathbf{1}'_k, \quad (4.22)$$

$$\Sigma = \frac{V}{k} \mathbf{1}_k \mathbf{1}'_k, \quad (4.23)$$

$$tr(\Psi) = kV,$$

$$tr(\Phi) = V/k,$$

$$tr(\Sigma) = V.$$

Analogous to section 4.2.1, we define the weighted combination estimator for τ

$$\hat{\tau}(\alpha) = \alpha \hat{\tau} + (1 - \alpha) \hat{\tau}_{(R)}. \quad (4.24)$$

The optimal weight, under the quadratic loss, is given by

$$\hat{\alpha}^* = 1 - \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2}. \quad (4.25)$$

Table 4.1: The values of m and V for selected degrees of freedom

ν	$-m$	$\frac{1}{\nu}$	V	$\frac{2}{\nu} + \frac{2}{\nu^2} + \frac{12}{\nu^3}$
1	1.270	1.000	4.935	16.00
2	0.577	0.500	1.645	3.000
3	0.369	0.333	0.935	1.333
4	0.270	0.250	0.645	0.812
5	0.213	0.200	0.490	0.576
6	0.176	0.167	0.395	0.444
7	0.150	0.143	0.330	0.362
8	0.130	0.125	0.284	0.305
9	0.115	0.111	0.249	0.263
10	0.103	0.100	0.221	0.232
15	0.068	0.067	0.143	0.146
20	0.051	0.050	0.105	0.106
30	0.034	0.033	0.069	0.069
40	0.025	0.025	0.051	0.051
50	0.020	0.020	0.041	0.041

The corresponding OWC estimator is

$$\begin{aligned} \hat{\tau}(\hat{\alpha}^*) &= \hat{\tau} + \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2} (\hat{\tau}_{(R)} - \hat{\tau}) \\ &= \hat{\tau}_{(R)} - \left(1 - \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2}\right) (\hat{\tau}_{(R)} - \hat{\tau}). \end{aligned} \tag{4.26}$$

Again, to avoid over-shrinking, we take the positive part of the weight to obtain a convex combination, defined by

$$\hat{\tau}(\hat{\alpha}^*)^+ = \hat{\tau}_{(R)} - \left(1 - \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2}\right)^+ (\hat{\tau}_{(R)} - \hat{\tau}). \tag{4.27}$$

Transforming back to the original scale and correcting the bias result in the regular and positive-part James-Stein estimators for σ^2 based on log-transformed

$$\hat{\sigma}^2(\hat{\alpha}^*)_{Tr} = e^{-m} \cdot \exp \left[\hat{\tau}_{(R)} - \left(1 - \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2}\right) (\hat{\tau}_{(R)} - \hat{\tau}) \right], \tag{4.28}$$

$$\hat{\sigma}^2(\alpha^*)_{Tr}^+ = \epsilon^{-m} \cdot \exp \left[\hat{\tau}_{(R)} - \left(1 - \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2} \right)^+ (\hat{\tau}_{(R)} - \hat{\tau}) \right], \quad (4.29)$$

where ϵ^{-m} is a bias corrector, and for any vector \mathbf{V} , $\exp[\mathbf{V}]$ means taking the exponential of each element. If we substitute (4.18) and (4.19) into (4.28), and let

$$a = \frac{(k-1)V}{\|\hat{\tau} - \hat{\tau}_{(R)}\|^2},$$

then each component of $\hat{\sigma}^2(\alpha^*)_{Tr}$ in (4.28) can be expressed in the following form:

$$\hat{\sigma}_i^2(\alpha^*)_{Tr} = \left[\left(\prod_{j=1}^k s_j^2 \right)^{1/k} \right]^{(1-a)} \cdot (s_i^2)^a, \quad (4.30)$$

which is a weighted geometric mean of the unrestricted estimator and the “geometrically pooled” estimator. (4.29) is very similar to the estimator given by Cui et al (2005),

$$\hat{\sigma}_i^2(JS+) = \left(\prod (s_i^2)^{1/G} \right) \epsilon^{-m} \exp \left[\left(1 - \frac{(k-3)V}{\sum (\ln(\nu s_i^2) - \overline{\ln(\nu s_i^2)})^2} \right)_+ \left(\ln(\nu s_i^2) - \overline{\ln(\nu s_i^2)} \right) \right], \quad (4.31)$$

with one major difference. We have $(k-1)$, while they use $(k-3)$ in the weight component. The approaches by which the two estimators are constructed are totally different. Their $(k-3)$ is associated with the degrees of freedom of the test statistic, while our approach is a pure optimization process unrelated to any test statistic. We did estimate the denominator term in (4.7) with its unbiased estimator; however, it does not make much sense to subtract the number of parameters estimated from $(k-1)$ in the context that degrees of freedom is not present. Further, the simulation in section 4.4 shows that our estimator in (4.29) has lower risk than that in (4.31).

4.3 Risk Study of the Estimators

The objective of this section is to assess the performance of various estimators using squared error loss and entropy loss functions. All the results in this section are subject to normality and independence assumptions.

4.3.1 Under the Quadratic Loss Function

When estimating σ^2 using some estimator $\hat{\sigma}^2_{(*)}$, the quadratic loss is

$$L(\hat{\sigma}^2_{(*)}) = (\hat{\sigma}^2_{(*)} - \sigma^2)'(\hat{\sigma}^2_{(*)} - \sigma^2). \tag{4.32}$$

The risk of $\hat{\sigma}^2_{(*)}$ is

$$R(\hat{\sigma}^2_{(*)}) = E((\hat{\sigma}^2_{(*)} - \sigma^2)'(\hat{\sigma}^2_{(*)} - \sigma^2)).$$

Since $\hat{\sigma}^2 \sim \mathbf{F}(\sigma^2, \text{Diag}(2\sigma_i^4/(n-1)))$, where \mathbf{F} is some kind of k-variate probability distribution. It is straight forward to derive that

$$R(\hat{\sigma}^2) = \frac{2}{(n-1)}(\sigma^2)' \sigma^2 = \frac{2 \sum \sigma_i^4}{(n-1)}. \tag{4.33}$$

We should note that the quadratic risk of this unrestricted estimator is not a constant, i.e., it depends on the parameter σ^2 .

It is also easy to see that

$$\begin{aligned} R(\hat{\sigma}^2_{(R)}) &= (\sigma^2 - \sigma^2_{(R)})'(\sigma^2 - \sigma^2_{(R)}) + \frac{2}{(n-1)k} \sigma^{2'} \sigma^2 \\ &= \sum_{i=1}^k (\sigma_i^2 - \sigma_{(R)}^2)^2 + \frac{2 \sum \sigma_i^4}{(n-1)k}. \end{aligned} \tag{4.34}$$

It is obvious that when the null hypothesis in (3.2) is true, the first term on the right hand side of equation (4.34) is zero. Thus, $R(\hat{\sigma}^2_{(R)})$ is much smaller than $R(\hat{\sigma}^2)$

in the microarray data analysis where k is usually thousands. However, when the null hypothesis is not true, $R(\hat{\sigma}_{(R)}^2)$ is potentially larger than $R(\theta, \hat{\sigma}^2)$, and it is unbounded.

Comparing the right hand side of (4.33) and (4.34), it is straight forward to prove the following theorem.

Theorem 4.3.1. *when $\frac{1}{k}(\sigma^2 - \sigma_{(R)}^2)'(\sigma^2 - \sigma_{(R)}^2) < \frac{2(k-1)}{(n-3)k+2}\sigma_{(R)}^2' \sigma_{(R)}^2$, $\hat{\sigma}_{(R)}^2$ has smaller quadratic risk than $\hat{\sigma}^2$.*

When $n \leq 3$ and $k \geq 2$, the inequality in theorem 4.3.1 always holds. Consequently, we obtain the following corollary:

Corollary 4.3.2. *When $n \leq 3$ and $k \geq 2$, $\hat{\sigma}_{(R)}^2$ has smaller quadratic risk than $\hat{\sigma}^2$.*

The risk of the OWLC estimator $\hat{\sigma}^2(\alpha^*)$ is

$$\begin{aligned} R(\hat{\sigma}^2(\alpha^*)) &= E[(\hat{\sigma}^2(\alpha^*) - \sigma^2)'(\hat{\sigma}^2(\alpha^*) - \sigma^2)] \\ &= E[(\hat{\sigma}_{(R)}^2 - (1-a)(\hat{\sigma}_{(R)}^2 - \hat{\sigma}^2) - \sigma^2)'(\hat{\sigma}_{(R)}^2 - (1-a)(\hat{\sigma}_{(R)}^2 - \hat{\sigma}^2) - \sigma^2)], \end{aligned}$$

where

$$a = \frac{2(k-1)(\hat{\sigma}^2)' \hat{\sigma}^2}{(n-1)k \|\hat{\sigma}^2 - \hat{\sigma}_{(R)}^2\|^2}.$$

The risk is a highly nonlinear function of the data; it can be estimated by a bootstrap technique or the first-order asymptotic technique suggested by Judge and Mittelhammer (2004).

4.3.2 Under Entropy Loss Function

As discussed in chapter 3, entropy risk is considered as a more reasonable measure when comparing estimators of a scale parameter. When estimating σ^2 by $\hat{\sigma}^2$, the

entropy loss is defined as

$$L(\hat{\sigma}^2) = \frac{\hat{\sigma}^2}{\sigma^2} - 1 - \ln \frac{\hat{\sigma}^2}{\sigma^2}. \quad (4.35)$$

It can be shown that under the entropy loss function, the risk

$$\begin{aligned} R(\hat{\sigma}^2) &= \sum E \left(\frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 - \ln \frac{\hat{\sigma}_i^2}{\sigma_i^2} \right) \\ &= - \sum E \left(\ln \frac{\hat{\sigma}_i^2}{\sigma_i^2} \right) \\ &= -kE \left(\ln \frac{\chi_v^2}{v} \right) \\ &= -k[\psi(v/2) - \ln(v/2)], \end{aligned} \quad (4.36)$$

where χ_v^2 is a chi-squared random variable with v degrees of freedom, and $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function. Here, $v = n - 1$. Therefore, the entropy risk of the unrestricted estimator is a constant independent of the parameter σ^2 . Selected values of the expectation in (4.36) is given in table 4.1. By using the second order Taylor expansion,

$$R(\hat{\sigma}^2) \approx \frac{k}{n-1}.$$

Under the entropy loss defined in (4.35), the risk of $\hat{\sigma}_{(R)}^2$ is

$$R(\hat{\sigma}_{(R)}^2) = \sum_{i=1}^k \left[E \left(\frac{\hat{\sigma}_{(R)}^2}{\sigma_i^2} - 1 + \ln \frac{\sigma_{(R)}^2}{\sigma_i^2} \right) \right]. \quad (4.37)$$

Under the null hypothesis,

$$\begin{aligned} R(\hat{\sigma}_{(R)}^2) &= -kE \left(\ln \frac{\sigma_{(R)}^2}{\sigma^2} \right) \\ &= -kE \left(\ln \frac{\chi_{\nu_2}^2}{\nu_2} \right), \text{ where } \nu_2 = nk - 1 \\ &\approx \frac{k}{nk-1}. \end{aligned} \quad (4.38)$$

It is easy to see that $R(\hat{\sigma}_{(R)}^2)$ is much smaller than $R(\hat{\sigma}^2)$ when the null hypothesis is true.

The entropy risks of the proposed shrinkage type estimators are quite complicated. Therefore, we conduct a simulation study and present it in section 4.4.

4.4 Simulation Results

In order to compare the performance of the various estimators, an extensive simulation study was conducted for k ranging from 4 to 1000 and n ranging from 2 to 20 under quadratic and entropy loss functions based on the E.Coli data provided by Arfin *et al.* (2000). Eight estimators were studied in our simulation. However, the issue of risk comparison of the pretest estimators and the Stein-type estimators was addressed in chapter 3 and the Stein-type estimators are preferred over the pretest estimators. Thus, we only exhibit the comparison among the pooled, the two OWC estimators proposed in section 4.2, and the Stein-type estimator proposed by Cui *et al.* (2005). Although the OWC estimators were derived based on the normality assumption, we still would like to know how they perform when the data are from a non-normal population. Therefore, the data were sampled from non-normal as well as normal distributions. The non-normal distributions are represented by mixtures of two normal populations in this study.

In the simulation, the efficiencies of the RE, OWC1 (from original data), OWC2 (from log-transformed data) and Cui's estimators relative to the unrestricted estimator are computed for values of Δ ranging from 0 to 2. The simulation result for selected combinations of k and n is presented in Figures 4.1 to 4.5.

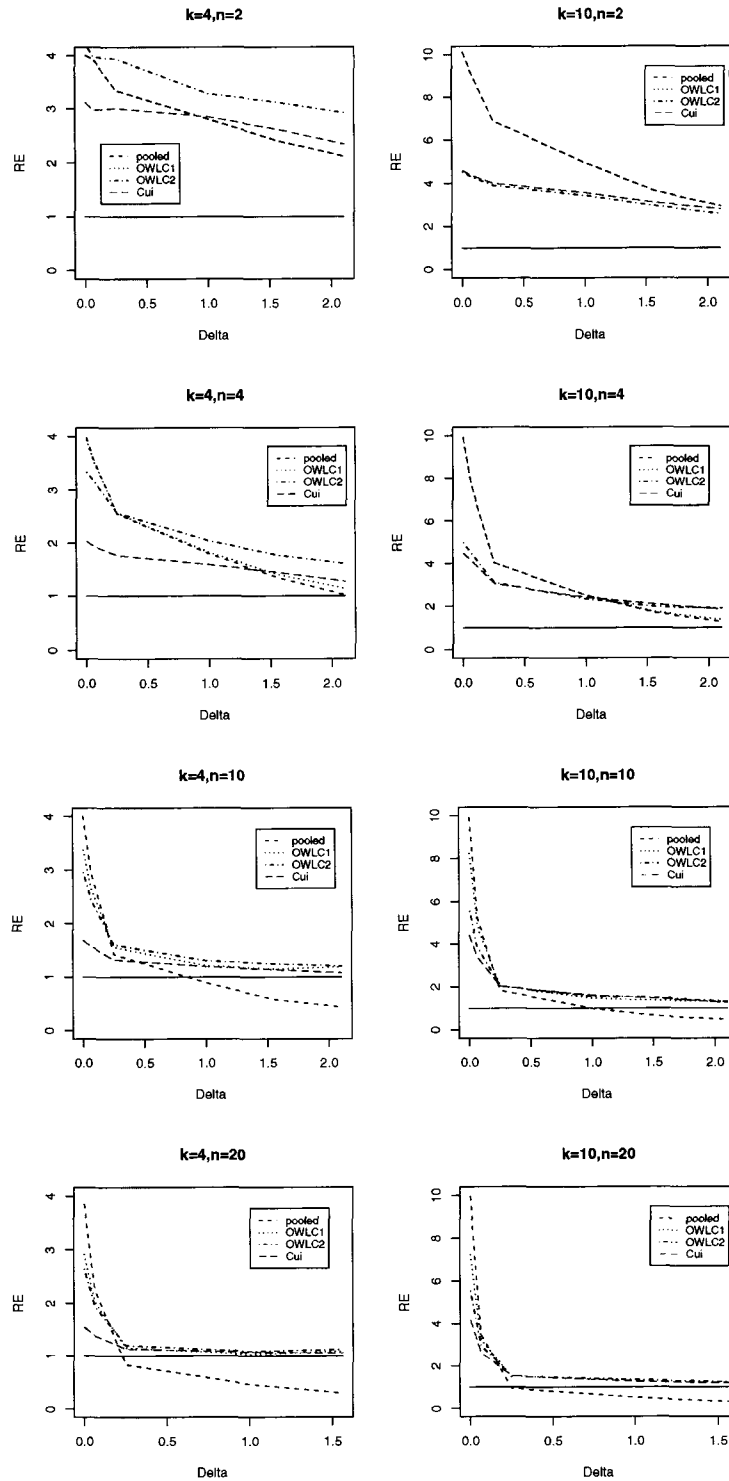


Figure 4.1: Relative efficiency of the proposed estimators under quadratic loss for normal data when $k = 4$ and 10.

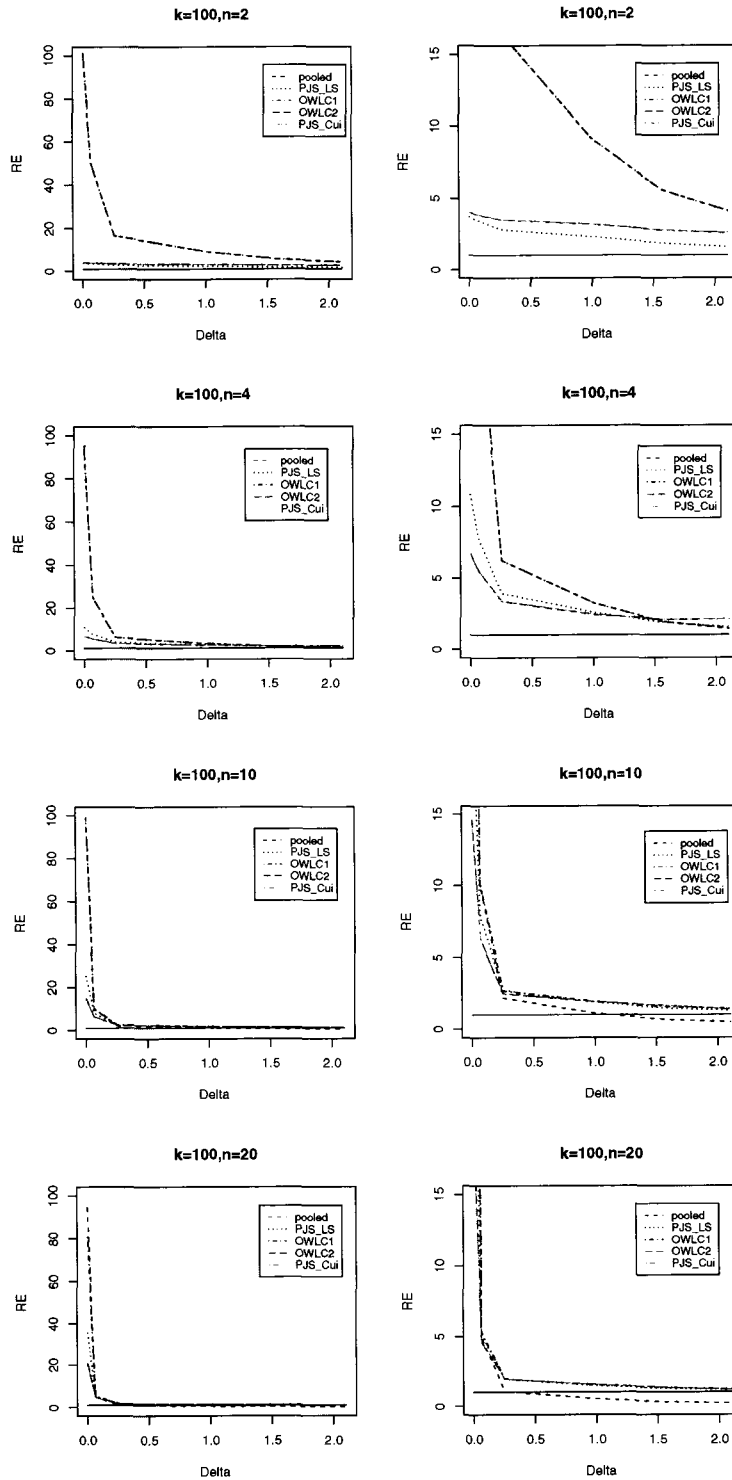


Figure 4.2: Relative efficiency of the proposed estimators under quadratic loss for normal data when $k = 100$. The plots on the right are a closer look of the plots on the left.

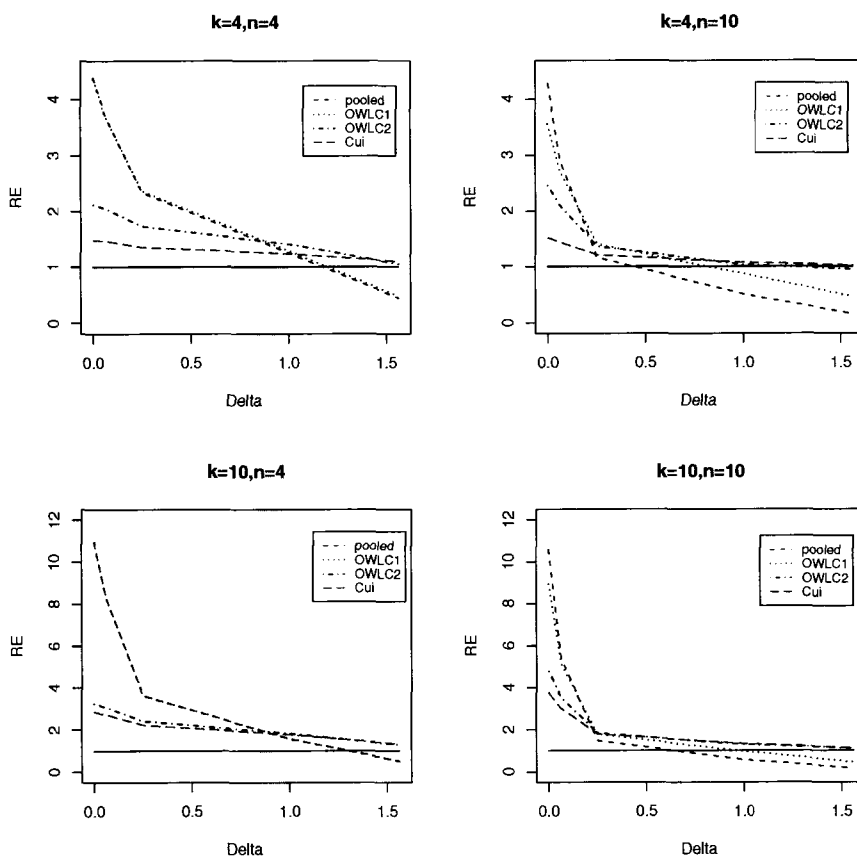


Figure 4.3: Relative efficiency of the proposed estimators under entropy loss (normal data) for $k = 4$ and 10 , and $n = 4$ and 10 .

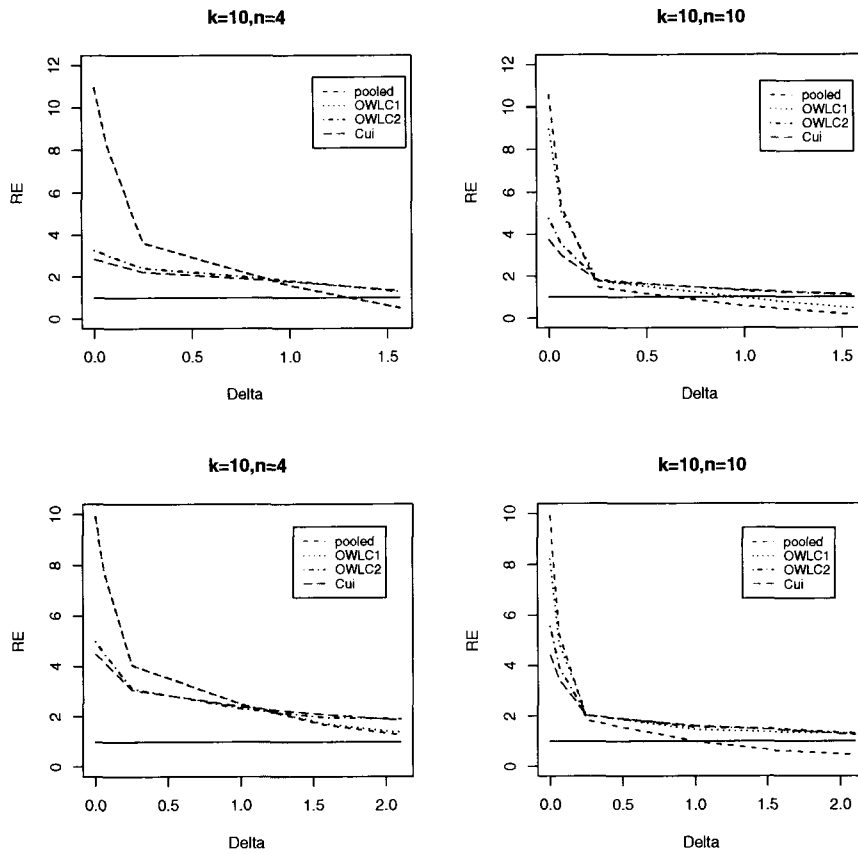


Figure 4.4: RE comparison under entropy and quadratic losses (normal data). The top two plots are under entropy loss, and the bottom two plots are under quadratic loss.

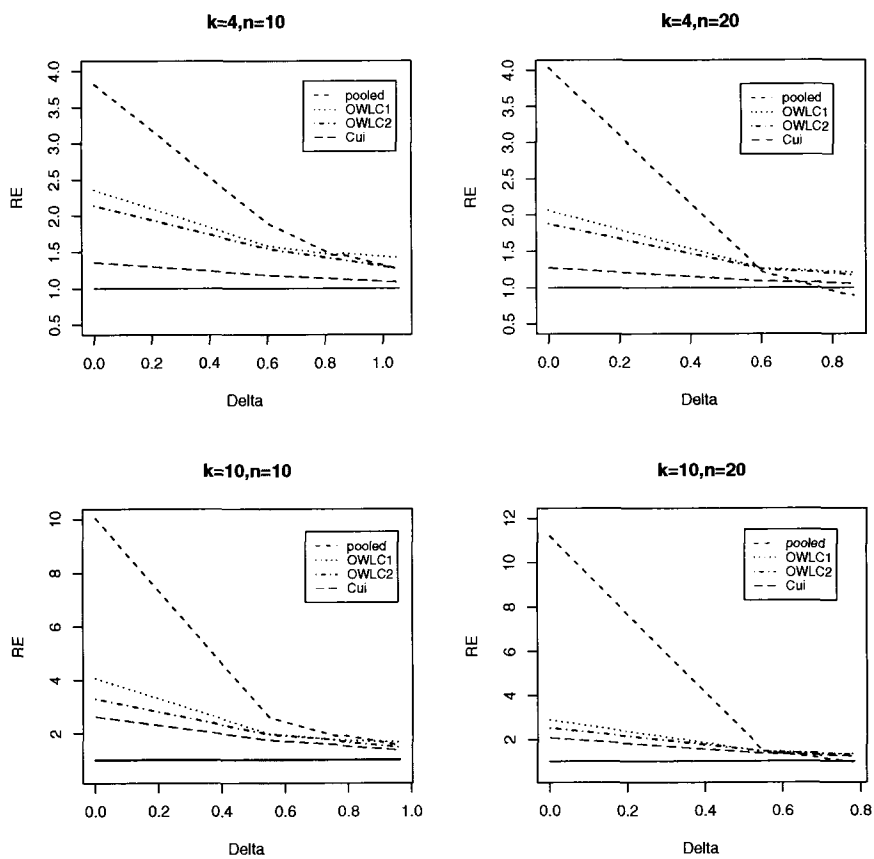


Figure 4.5: Relative efficiency of the estimators under quadratic loss for a mixture of normal data.

Figures 4.1 to 4.2 show the relative efficiencies of the estimators under quadratic loss when data are sampled from normal populations. It is observed that when $n=2$ $\hat{\sigma}_{(R)}^2$ dominates the unrestricted estimator (with $RE > 1$) within the entire studied range of the parameter. This result agrees with Theorem 4.3.1. When n is moderate, the RE of $\hat{\theta}^{RE}$ decreases rapidly and turns lower than 1 as Δ increases. on the other hand, the RE of the other three estimators all approach, but stay well above 1. The two OWC estimators perform better than Cui's estimator in general. The advantage is substantial when Δ is small. As k increases, the difference between Cui's estimator and the OWLC2 disappears.

Figure 4.3 demonstrates the relative efficiencies when using entropy loss function, while Figure 4.4 presents a side by side comparison under two different loss functions for $k = 10$ and $n = 4$ and 10 . The top plots are the cases using entropy loss, whereas the bottom ones represent quadratic loss. An important observation is that the RE of OWC1 estimator becomes lower than 1 as Δ increases. Hence, it does not retain the property of a Stein-type estimator under entropy loss function. The RE of OWC2 and Cui's estimator both maintain above 1 within the range of our study.

Figure 4.5 shows the relative efficiencies under quadratic loss when the data are sampled from mixtures of two normal populations for selected k and n . Simulation results show that all the estimators are highly efficient relative to $\hat{\theta}^U$ within the experimental range of the E.Coli data. Further, the OWC estimators are more efficient than Cui's estimator.

4.5 Concluding Notes

Two optimal weighted combination estimators are proposed based on assumptions of normality and independence under the quadratic loss function in the context of finite samples. The risks of these estimators are studied and compared to the estimator given by Cui *et al.* (2005). The simulation shows that all three estimators resemble the Stein-type estimator and keep the nice properties, e.g., dominating the unrestricted estimator, under quadratic loss functions. Some theoretical results are obtained. When the entropy loss function is used, these estimators have lower risks compared to the unrestricted estimator in a wide range of parameter space, but do not dominate in the entire parameter space. The performance of the OWC estimators and Cui's estimator are similar when k is large. However, when k is small, OWCE1 and OWCE2 have smaller risks compared to Cui's estimator near the null hypothesis; Cui's estimator becomes better when Δ increases.

Keeping in mind that normality and independence may not hold in microarray data, we simulated the risks and relative efficiencies when data were sampled from non-normal populations. The simulation shows that the relative efficiency is robust to the violation of normality assumption. The risk performance of the estimators in the dependent case was not studied. However, our approach can be extended to correlated data; we relegate this pursue for future work.

Chapter 5

Estimation of the Variance Components for Meta-Analysis with Random Effects

5.1 Introduction

Meta-analysis is defined as the statistical analysis of a collection of analytic results for the purpose of integrating the finding. Such analysis are becoming increasingly popular in medical research where information on efficacy of a treatment is available from a number of clinical studies with similar treatment protocols and in microarray data analysis where many research groups study the same subject under very similar conditions. If considered separately, any one study may be either too small or too limited in scope to come to unequivocal or generalizable conclusions about the effect of treatment. Combining the findings across such studies represents an attractive alternative to strengthen the evidence about the treatment efficacy.

Meta-analysis provides an objective way of combining information from separate studies looking at the same question and has been applied most often to treatment effects in randomized clinical trials. For example, consider k randomized trials com-

paring a new medication with placebo, in which each trial's treatment effect is estimated in terms of a difference in means of a quantitative variable, or, for a binary outcome, in terms of the log odds ratio. Standard meta-analysis methods for providing an overall estimate of the treatment effect rely on certain assumptions (Hardy and Thompson (1996)). The fixed effect model is based on homogeneity of treatment effects across all k studies included in the meta-analysis. In other words it must be assumed that the estimated treatment effect $\hat{\theta}_i$ has a common mean θ and individual variance σ_i^2 for each study $i = 1, \dots, k$. The treatment effect θ can then simply be estimated as a weighted average of the individual study estimates, that is

$$\hat{\theta} = \frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i}, \tag{5.1}$$

where w_i is the weight given to study i and generally taken to be the reciprocal of the variance σ_i^2 for study i . In practice, σ_i^2 must be estimated. The estimation of σ_i^2 in a fixed effect model is considered in the previous chapters.

However, in practical medical researches or microarray experiments, homogeneity is rare due to the nature of the studies and the many variables involved. Hence a degree of statistical heterogeneity might be anticipated. As a result, the assumption underlying the fixed effect model does not hold. A formal test of statistical homogeneity can be performed using the test statistic

$$Q = \sum_{i=1}^k w_i (\hat{\theta}_i - \hat{\theta})^2, \tag{5.2}$$

which has approximately a χ_{k-1}^2 distribution under the null hypothesis

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_k = \theta. \tag{5.3}$$

The random effects model provides a way of incorporating heterogeneity into the estimate by including a between-study component of variance τ^2 .

Consider k independent studies, with effect measure variables Y_i and true study specific effects θ_i for $i = 1, \dots, k$. The random effects model is given by

$$Y_i = \theta_i + \varepsilon_i \quad \text{and} \quad \theta_i = \theta + \delta_i, \quad i = 1, \dots, k,$$

where ε_i and δ_i are assumed to be independent and normally distributed with $\varepsilon_i \sim N(0, \sigma_i^2)$ and $\delta_i \sim N(0, \tau^2)$. Hence the marginal distribution of Y_i is $N(\theta, \sigma_i^2 + \tau^2)$. Here σ_i^2 is the within-study variance and τ^2 is the across-study variance. Both σ_i^2 and τ^2 are unknown fixed values. The generally accepted practice in meta-analysis is to use the estimated values of these variances.

In the previous chapters, we proposed several shrinkage type estimators for the within variance σ_i^2 . In this chapter we will focus on the estimation of the between variance τ^2 . The statistical properties of the proposed estimators will be investigated via an extensive Monte Carlo simulation.

5.2 Risk Study of the Existing Estimators

Over the years many estimators of τ^2 have been proposed in the reviewed literature. The method of moments estimator (MM) by DerSimonian and Laird (1986), the variance-component type estimator (VC) by Hedges (1983), the simple heterogeneity variance estimator (SH) by Sidik and Jonkman (2005), the maximum likelihood estimator (ML) by Hardy and Thompson (1996), the approximate restricted maximum likelihood estimator (REML) by Morris (1983), and the empirical Bayes estimator (EB) by Morris (1983) are the most popular ones among others.

The MM estimator is given by

$$\hat{\tau}_{MM}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k \hat{w}_i (y_i - \hat{\theta})^2 - (k-1)}{\sum_{i=1}^k \hat{w}_i - \sum_{i=1}^k \hat{w}_i^2 / \sum_{i=1}^k \hat{w}_i} \right\}, \quad (5.4)$$

where $\hat{\theta} = \sum_{i=1}^k \hat{w}_i y_i / \sum_{i=1}^k \hat{w}_i$ and $\hat{w}_i = 1/\hat{\sigma}_i^2$.

The VC estimator is given by

$$\hat{\tau}_{VC}^2 = \max \left\{ 0, \frac{1}{k-1} \sum_{i=1}^k (y_i - \bar{y})^2 - \frac{1}{k} \sum_{i=1}^k \hat{\sigma}_i^2 \right\}, \quad (5.5)$$

where $\bar{y} = \sum_{i=1}^k y_i/k$. Note that $\hat{\tau}_{VC}^2$ is also a method of moments estimator.

The SH estimator is given by

$$\hat{\tau}_{SH}^2 = \frac{1}{k-1} \sum_{i=1}^k \hat{v}_i^{-1} (y_i - \hat{\theta}_i)^2, \quad (5.6)$$

where $\hat{v}_i = \hat{r}_i + 1$, $\hat{r}_i = \hat{\sigma}_i^2 / \left(\sum_{i=1}^k (y_i - \bar{y})^2 / k \right)$, and $\hat{\theta}_i = \sum_{i=1}^k \hat{v}_i^{-1} y_i / \sum_{i=1}^k \hat{v}_i^{-1}$.

The ML estimator can be calculated by iterating the equation

$$\hat{\tau}_{ML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k \hat{w}_i^2 \left\{ (y_i - \hat{\theta})^2 - \hat{\sigma}_i^2 \right\}}{\sum_{i=1}^k \hat{w}_i^2} \right\}, \quad (5.7)$$

where $\hat{\theta} = \sum_{i=1}^k \hat{w}_i y_i / \sum_{i=1}^k \hat{w}_i$ and $\hat{w}_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{ML}^2)$.

Similarly, the REML estimator is computed using the iterative equation

$$\hat{\tau}_{REML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k \hat{w}_i^2 \left\{ (k/(k-1))(y_i - \hat{\theta})^2 - \hat{\sigma}_i^2 \right\}}{\sum_{i=1}^k \hat{w}_i^2} \right\}, \quad (5.8)$$

with $\hat{w}_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{REML}^2)$.

An EB estimator $\hat{\tau}_{EB}^2$ can be computed using the same iterative equation given for the REML estimator, except replacing \hat{w}_i^2 with \hat{w}_i .

Among the above six estimators, the MM, VC and SH estimators are simple to compute. On the other hand, the ML, REML, and EB estimators are more computational intensive, and require iterative solutions.

5.2.1 Simulation Study of the Risks

We performed simulation studies to compare the risks of the MM, VC, SH, and ML estimators under quadratic and entropy loss functions. Since the REML and EB estimators are similar to the ML estimator, we did not include these two estimators in our simulation.

The simulation study is based on the meta-analysis data from nine studies on preventing pre-eclampsia using diuretics (Collins *et al.* (1985), Thompson and Pocock (1991), Sidik and Jonkman (2006)). The design of the simulation is similar to that discussed in Sidik and Jonkman (2006). The measure of the effect size is log odds ratio. The true overall effect θ was set at -0.5 because the estimated overall log odds ratio for the diuretics trials in the studies is approximately -0.5. We chose four different meta-analysis sample sizes, $k = 10, 20, 30,$ and 50 . For each sample size k , we considered six values of τ^2 ranging from 0 to 3. For each combination for k and τ^2 , we generated 10,000 2x2 tables. Specifically, we first generated θ_i from $N(\theta, \tau^2)$ for $i = 1, \dots, k$. For a given k , equal sample sizes n_i for the control and treatment groups were randomly chosen from the integers 20 - 1000 since in most of the diuretics trials the sample sizes for the control and treatment groups are similar and roughly within the above range. Next, we generated the responses x_{iC} for the control groups based on a binomial (n_i, p_{iC}) distribution for $i = 1, \dots, k$, where the true binomial probability p_{iC} was randomly selected from a uniform distribution on the interval from 0.02 to 0.5 (following the range of the observed proportions for the control groups in the studies). The responses x_{iT} for the treatment groups were generated from a Binomial (n_i, p_{iT}) distribution with $p_{iT} = p_{iC} \exp\{\theta_i\} / (1 - p_{iC} + p_{iC} \exp\{\theta_i\})$ for $i = 1, \dots, k$.

Using this method, which maintains the true within-study log odds ratio for the i th study at θ_i , we generated k 2x2 tables with the cell counts $(x_{iC}, n_i - x_{iC}, x_{iT}, n_i - x_{iT})$ for the i th table. This was replicated for 10,000 times for each combination of k and τ^2 . For each replicate, the sample log odds ratios y_i , $i = 1, \dots, k$, were computed. The estimated within-study variances $\hat{\sigma}_i^2$ were calculated by the asymptotic formula:

$$\hat{\sigma}_i^2 = 1/(n_i p_{iT}(1 - p_{iT})) + 1/(n_i p_{iC}(1 - p_{iC})). \quad (5.9)$$

Finally, the quadratic and entropy risks of the MM, VC, SH, and ML estimators were computed in the same manner described in chapters 3 and 4 for each combination of k and τ^2 . In the simulation of the entropy risk, the minimum value of τ^2 was 0.1, since a value of zero for the parameter would make the entropy loss undefined. In addition, the efficiency of each estimator relative to the VC estimator was calculated and presented in columns 2 to 4 of Tables 5.1 to 5.8.

Tables 5.1 to 5.4 exhibit the relative efficiencies under the quadratic loss, and Tables 5.5 to 5.8 show the relative efficiencies of the variance estimators under the entropy loss function. The simulation showed that when the random effect variance τ^2 is small, the efficiencies of the MM and ML estimators both are substantially higher than that of the VC estimator. As τ^2 increases, the relative efficiencies of MM and ML estimators decrease rapidly and become lower than 1 at some value of τ^2 . This observation allows us to conclude that when τ^2 is small, the $\hat{\tau}_{MM}^2$ and $\hat{\tau}_{ML}^2$ have lower risks, and therefore are better choices to estimate τ^2 ; when τ^2 is large, $\hat{\tau}_{VC}^2$ is better in terms of risk. The change of the relative efficiency of the SH estimator is less significant and does not strictly follow this pattern under quadratic loss for $k = 20, 30, 50$. The RE of $\hat{\tau}_{SH}^2$ is lower than 1 when $\tau^2 = 0$, jumps up above 1 when

τ^2 increases to 0.1, and decreases to near 1 as τ^2 continues to increase. When the entropy loss function is employed in the simulation, we did not observe any RE of $\hat{\tau}_{SH}^2$ lower than 1 since the value of τ^2 started from 0.1 instead of 0.

5.3 James-Stein Type Estimators

Noticing the change of the efficiencies of $\hat{\tau}_{MM}^2$ and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ with respect to τ^2 motivates us to propose James-Stein type estimators that shrink the VC estimator toward either the MM or the ML estimators:

$$\hat{\tau}_{JS1}^2 = \hat{\tau}_{MM}^2 + \left(1 - \frac{k-3}{Q}\right)_+ (\hat{\tau}_{VC}^2 - \hat{\tau}_{MM}^2), \tag{5.10}$$

$$\hat{\tau}_{JS2}^2 = \hat{\tau}_{ML}^2 + \left(1 - \frac{k-3}{Q}\right)_+ (\hat{\tau}_{VC}^2 - \hat{\tau}_{ML}^2), \tag{5.11}$$

where $Q = \sum_{i=1}^k w_i(\hat{\theta}_i - \hat{\theta})^2$, as given in (5.2), $\hat{\theta}_i = y_i$, $\hat{\theta} = \sum_i w_i y_i / \sum_i w_i$, and $w_i = 1/\sigma_i^2$. Q is the sum of squares of the treatment effects about the mean where the i th square is weighted by the reciprocal of the estimated variance. Under the null hypothesis:

$$H_0 : \tau^2 = 0, \tag{5.12}$$

Q follows asymptotically a χ^2 distribution with $k - 1$ degrees of freedom. Thus, when each study has a large sample size, Q may be used to test the null hypothesis in (5.12). Ideally, $\hat{\tau}_{JS1}^2$ and $\hat{\tau}_{JS2}^2$ both dominate $\hat{\tau}_{VC}^2$ in the entire parameter space.

5.3.1 Risk Simulation

Simulation study was conducted to compare the risks of these two James-Stein estimators with other estimators. The efficiencies of these estimators relative to the

Table 5.1: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 10$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.0	34.08	1.94	69.89	3.69	4.22
0.1	6.38	1.85	4.85	1.39	1.35
0.5	2.11	1.27	1.76	1.07	1.04
1.0	0.97	1.14	1.26	1.04	1.02
2.0	0.24	1.06	0.94	1.04	1.01
3.0	0.13	1.06	0.91	1.04	1.01

VC estimator were computed and listed in the last two columns of Tables 5.1 to 5.8.

The comparison can also be visualized by Figures 5.1 and 5.2.

The simulation results show that $\hat{\tau}_{SR1}^2$ and $\hat{\tau}_{SR2}^2$ perform similarly in terms of risk. Under the quadratic loss function, when the null hypothesis is true, both $\hat{\tau}_{SR1}^2$ and $\hat{\tau}_{SR2}^2$ are highly efficient compared to $\hat{\tau}_{VC}^2$. For instance, when $k = 20$, $RE_{SR1} = 3.37$, $RE_{SR2} = 4.43$. $\hat{\tau}_{SR2}^2$ is more efficient than $\hat{\tau}_{SR1}^2$ since the efficiency of $\hat{\tau}_{ML}^2$ is much higher than that of $\hat{\tau}_{MM}^2$ at $\tau^2 = 0$. As τ^2 increases, the efficiencies of the James-Stein type estimators decrease and approach 1 as expected. The relative performance of the estimators is similar when the entropy loss function is employed except that the risk is not defined when $\tau^2 = 0$ and thus not observed.

5.4 Conclusion and Discussion

The risks of MM, VC, SH, and ML estimators for the random effect variance component τ^2 in meta-analysis were studied. Simulation showed that the efficiencies of these estimators change differently when the true value of τ^2 changes. In general, the SH estimator is slightly more efficient than the VC estimator when H_0 is not true.

Table 5.2: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 20$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.0	14.52	0.64	30.77	3.37	4.43
0.1	3.79	1.33	2.63	1.41	1.29
0.5	1.07	1.18	1.61	1.11	1.04
1.0	0.44	1.14	1.48	1.08	1.02
2.0	0.19	1.11	1.42	1.07	1.01
3.0	0.10	1.09	1.36	1.07	1.01

Table 5.3: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 30$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.0	9.09	0.42	17.45	3.73	5.42
0.1	3.64	1.34	3.03	1.33	1.26
0.5	2.09	1.26	1.73	1.10	1.04
1.0	0.53	1.20	1.53	1.08	1.02
2.0	0.09	1.12	1.29	1.08	1.01
3.0	0.04	1.06	0.96	1.05	1.01

Table 5.4: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under quadratic loss for $k = 50$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.0	29.57	0.54	61.64	8.36	11.83
0.1	4.94	1.25	4.08	1.42	1.36
0.5	3.08	1.19	1.70	1.08	1.03
1.0	0.79	1.17	1.40	1.05	1.02
2.0	0.10	1.08	1.04	1.02	1.01
3.0	0.04	1.03	0.77	1.00	1.00

Table 5.5: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 10$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.1	3.11	2.61	2.57	1.32	1.26
0.5	1.61	1.29	1.41	1.05	1.03
1.0	0.73	1.13	1.03	1.02	1.01
2.0	0.17	1.05	0.81	1.02	1.01
3.0	0.08	1.04	0.76	1.02	1.01

Table 5.6: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 20$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.1	2.76	2.04	2.11	1.42	1.31
0.5	0.77	1.20	1.40	1.08	1.03
1.0	0.32	1.13	1.30	1.06	1.02
2.0	0.13	1.10	1.25	1.05	1.01
3.0	0.07	1.09	1.22	1.06	1.01

Table 5.7: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 30$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.1	3.39	1.72	2.63	1.38	1.27
0.5	1.66	1.25	1.50	1.08	1.03
1.0	0.41	1.18	1.31	1.06	1.02
2.0	0.08	1.12	1.20	1.06	1.01
3.0	0.03	1.07	0.97	1.05	1.01

Table 5.8: Simulated RE of $\hat{\tau}_{MM}^2$, $\hat{\tau}_{SH}^2$, and $\hat{\tau}_{ML}^2$ relative to $\hat{\tau}_{VC}^2$ under entropy loss for $k = 50$

τ^2	$\hat{\tau}_{MM}^2$	$\hat{\tau}_{SH}^2$	$\hat{\tau}_{ML}^2$	$\hat{\tau}_{SR1}^2$	$\hat{\tau}_{SR2}^2$
0.1	5.31	1.75	4.20	1.52	1.43
0.5	2.80	1.24	1.62	1.08	1.03
1.0	0.69	1.18	1.32	1.04	1.01
2.0	0.09	1.08	0.99	1.01	1.00
3.0	0.03	1.03	0.75	1.00	1.00

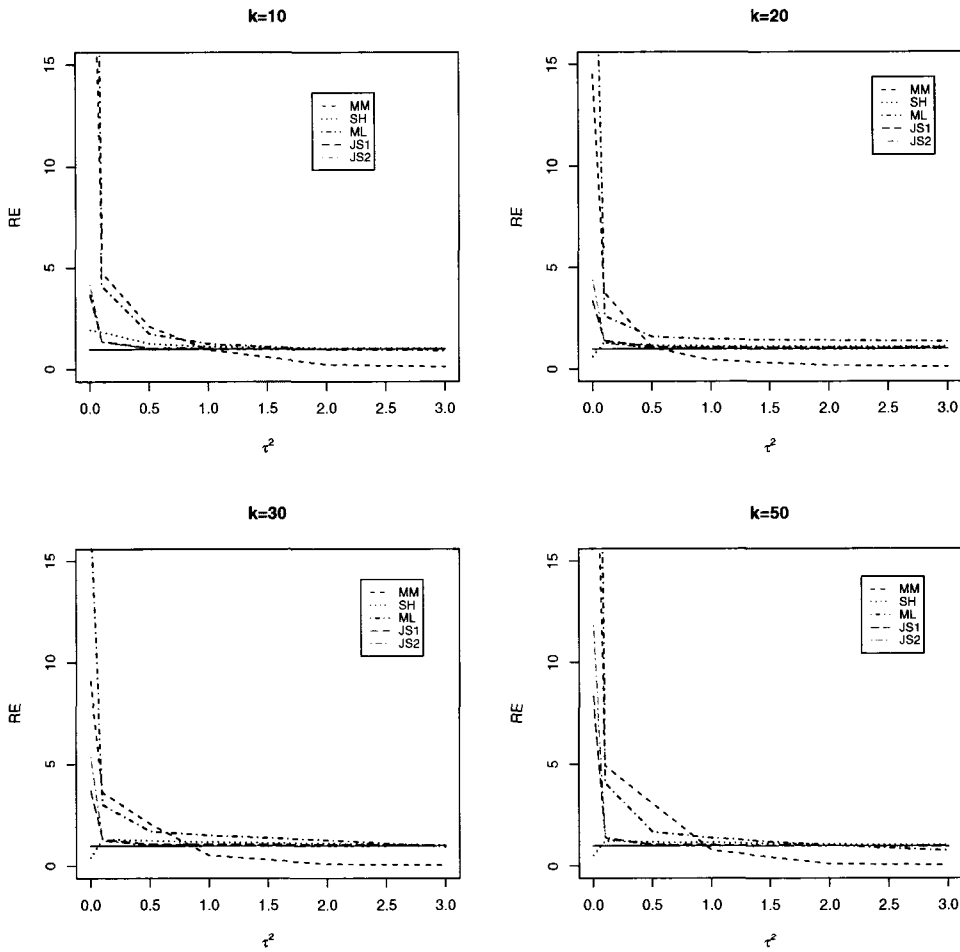


Figure 5.1: RE comparison under quadratic loss

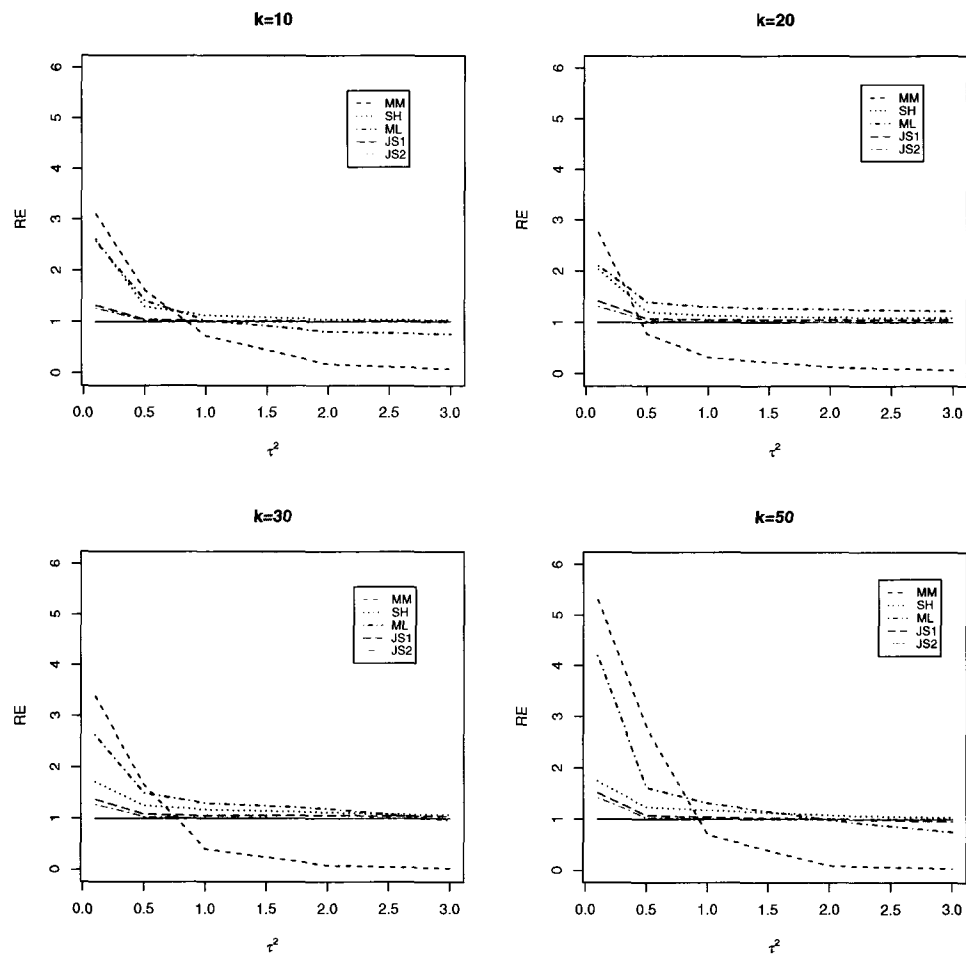


Figure 5.2: RE comparison under entropy loss

On the contrary, the MM and ML estimators have substantially higher efficiencies when H_0 is true or nearly true. As τ^2 increases, the efficiencies of $\hat{\tau}_{MM}^2$ and $\hat{\tau}_{ML}^2$ decrease and eventually become lower than that of $\hat{\tau}_{VC}^2$. Based on this observation, we constructed two James-Stein type estimators $\hat{\tau}_{JS1}^2$ and $\hat{\tau}_{JS2}^2$ which dominate $\hat{\tau}_{VC}^2$ in the range of our simulation study. The simulation showed that our estimators outperform $\hat{\tau}_{VC}^2$. Further, they behave more robustly with respect to other estimators under study in the sense that they dominate other estimators in most of the parameter space. Therefore we highly recommend to apply our proposed James-Stein type estimators in statistical inferences.

Chapter 6

Conclusions and Future Research

This thesis concentrates on problems related to variance estimation in various contexts. First of all, asymptotic tests and asymptotic interval estimation procedures about variance are developed for arbitrary populations. The kurtosis estimation problem embedded in the variance estimation procedures is studied extensively and two novel estimators of kurtosis are proposed. These estimators are improved upon the existing estimators in terms of risk. In the case of simultaneous estimation of k population variances, several estimation strategies based on pretest and James-Stein principles are developed and their risks are studied. In the finite-sample context, two optimal weight combination estimators (OWCE) are derived based on original data and log transformed data, respectively. The OWCEs resemble the James-Stein type estimator and dominate the unrestricted estimator. Further, two shrinkage type estimators of the variance components in meta-analysis with random effects are also suggested. Simulation shows that our shrinkage estimators perform better than the base estimators.

In chapter two, we developed asymptotic tests and asymptotic interval estimation procedures for population variances and hence provided a wide inferential package.

The statistical properties of the proposed inference procedures were investigated analytically and numerically. The simulation study corroborates our theoretical findings. It is reinforced that a much larger sample should be taken if the parent population is skewed. In many situations, the minimum values of n , in order to use the asymptotic results, far exceeded the folklore values of 25 or 30. In this case, employing a bootstrap or permutation test may be a better idea. Research on the statistical implications of these and other estimators is ongoing.

The kurtosis parameter estimation is embedded in many statistical estimation problems and applications. The estimation of kurtosis parameter is studied extensively in this chapter. We have compared the performance of several kurtosis measures adapted by SAS, SPSS, S-Plus, Minitab, and other statistical packages. We have proposed several new measures of kurtosis. It has been both analytically and numerically demonstrated that our proposed estimators outperform the existing estimators for normal population based on the *MSE* criterion. Moreover, an extensive simulation study has been conducted for non-normal populations. The result indicates that the proposed estimators are superior to the existing ones in many practical situations. Bearing this in mind, all the estimators substantially underestimate kurtosis parameter when underlying population distribution is highly skewed or heavy tailed. In order to correct the bias, empirical formulas are provided for student-t and chi-squared distributions. However, empirical estimates are subject to extra variation introduced and result in inflated *MSE*. Perhaps, some re-sampling methods such as bootstrap and Jackknife may be considered to reduce the bias as well as keeping a relatively lower variance. Non-parametric techniques can also be considered in order

to develop more robust estimators.

In chapter three, several estimation strategies for pooling data are presented for estimating the population variances and their risks are studied under different loss functions. It is concluded that positive rule Stein-type estimator dominates the usual shrinkage type estimator and they both dominate the unrestricted estimator $\hat{\sigma}^2$ in terms of the asymptotic distributional quadratic risk we defined. On the other hand, the performance of the restricted estimator heavily depends on the quality of non-data information. The James-Stein type estimator $\hat{\sigma}_{(JS)}^2$ has a disquieting feature that it may shrink beyond the hypothesis vector. We have improved $\hat{\sigma}_{(JS)}^2$ by the positive-rule estimator $\hat{\sigma}_{(PP)}^2$. The positive-rule estimator is particularly important to control the over-shrinking inherent in the shrinkage estimator. The performance of $\hat{\sigma}_{(R)}^2$, $\hat{\sigma}_{(SR)}^2$, $\hat{\sigma}_{(PT)}^2$ and $\hat{\sigma}_{(SP)}^2$ depends upon the correctness of the null hypothesis. It is shown that the range in which $\hat{\sigma}_{(SP)}^2$ dominates $\hat{\sigma}^2$ is wider than the range in which *PTE* dominates *UE*.

A risk simulation is conducted for finite samples under the usual quadratic loss and the entropy loss functions. The simulation result indicates that moderate sample sizes are sufficient for the estimators to follow the same dominance pattern (based on asymptotic risk) in practical situations. Though the estimators are proposed based on the quadratic risk, they perform similarly under the entropy loss function in terms of the relative efficiency.

In chapter four, two novel optimal weight combination estimators are derived based on assumptions of normality and independence under quadratic loss function in the finite-sample context. The risks of these estimators are studied and compared

to shrinkage estimator proposed by Cui *et al.* (2005). All three estimators resemble the Stein-type estimator and keep the nice properties under quadratic loss functions. When the entropy loss function is used, these estimators have lower risks compared to the unrestricted estimator in a wide range of the parameter space, but do not dominate in the entire parameter space. The performance of the OWC estimators and Cui's estimator are similar when the number of samples k is large. However, when k is small, the OWCEs have smaller risks compared to Cui's estimator near the null hypothesis; Cui's estimator becomes better when the parameter vector moves away from the null hypothesis. Keeping in mind that normality and independence may not hold in microarray data, we simulated the risks and relative efficiencies when data were sampled from a non-normal population. Simulation shows that the relative efficiency is robust to the violation of normality assumption. The risk performance of the estimators in the dependent case was not studied. However, our approach is extendable to correlated data, and we relegate this pursue for future work.

In chapter five, the risks of some existing estimators for the variance component τ^2 in meta-analysis with random effects were studied. Simulation showed that the efficiencies of these estimators change differently when the true value of the random effect changes. Two James-Stein type estimators are proposed and their risk is simulated and compared to some existing estimators. The simulation study shows that our shrinkage estimators outperform the base estimators. Further, they behave more robustly with respect to other estimators under study in the sense that they dominate other estimators in most of the parameter space. However, the base estimators, and hence the shrinkage type estimators that we discuss here require relatively large sam-

ple sizes, whereas many microarray experiments are conducted using small sample size. Therefore methods that perform better in small sample cases, such as empirical Bayes estimation, may be considered in future studies.

Throughout this dissertation, we combine information across the samples only via variances. However, more strength can be borrowed by pooling the means as well. This may provide a further improved estimation strategy. Our pooled estimator is the simple arithmetic mean of the k individual estimators. Alternatively, we may consider using geometric mean or weighted mean. In the simultaneous estimation of k independent variances, our shrinkage type estimators shrink the classical estimators toward a unique common variance estimator. In the application to microarray data where thousands of parameters are to be estimated simultaneously, we may consider shrinking the gene specific variance estimator toward a locally pooled estimator within a neighborhood of genes.

The least absolute shrinkage and selection operator (LASSO) (Tibshirani (1996)) estimation procedure, originally proposed for parameter selection in linear regression models, has become popular model selection procedures and found its application in the microarray data analysis recently (Gui and Li (2005), Segal *et al.* (2003)). It is considered to be an efficient way to solve the difficulty associated with the estimation in high dimension and low sample size setting. It has been used in the genes selection procedure. Ahmed *et al.* (2007) proposed an absolute penalty estimation (APE) method by extending the LASSO for partially linear models. They compared the APE estimator with the shrinkage and pretest estimators and found each had better performance on different part of the parameter space. Applying LASSO in the

microarray analysis will be a continuation of our research.

Bibliography

- Ahmed, S.E. (1992). Shrinkage preliminary test estimation in multivariate normal distribution. *Journal of Statistical Computation and Simulation* **43**, 177–195.
- Ahmed, S.E. (2002). Simultaneous estimation of coefficient of variation. *Journal of Statistical Planning and Inference* **104**, 31–51.
- Ahmed, S.E. (2005). Assessing the process capability index for non-normal process. *Journal of Statistical Planning and Inference* **129**, 195–206.
- Ahmed, S.E., A.K. Gupta, S.M. Khan and C.J. Nicol (2001). Simultaneous estimation of several intraclass correlation coefficients. *The Institute of Statistical Mathematics* **53**, 354–369.
- Ahmed, S.E. and A.K.Md.E. Saleh (1993). Improved estimation for the component mean vector. *Journal of Japanese Statistical Society* **43**, 177–195.
- Ahmed, S.E. and A.K.Md.E. Saleh (1999). Improved nonparametric estimation of location vectors in multivariate regression models. *Journal of Nonparametric Statistics* **11**, 51–78.
- Ahmed, S.E. and W.J. Krzanowski (2004). Biased estimation in a simple multivariate regression model. *Computational Statistics and Data Analysis* **45**, 689–696.

- Ahmed, S.E., K. Doksum, S. Hossain and J. You (2007). Shrinkage, pretest and LASSO estimators in partially linear models. submitted for publication.
- Ahmed, S.E., L. An and S. Nkurunziza (2006). Improving the estimation of eigenvectors under quadratic loss. *Calcutta Statistical Association Bulletin*. to appear.
- Allison, D.B., X.Q. Cui, G.P. Page and M. Sabripour (2006). Microarray data analysis: From disarray to consolidation and consensus. *Nature Reviews Genetics* **7**(1), 55-65.
- An, L. and S.E. Ahmed (2007). Improving the performance of kurtosis estimator. submitted for publication.
- An, L., S.E. Ahmed and A. Ali (2006). Tumor growth rate approximation-assisted estimation. *Cancer Informatics* **2**, 214-221.
- Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons.
- Arfin, S.M., A.D. Long, E.T. Ito, L. Toller, M.M. Riehle, E.S. Paegle and G.W. Hatfield (2000). Global gene expression profiling in Escherichia Coli K12. *The Journal of Biological Chemistry* **275**, 29672-29684.
- Balanda, K.P. and H.L. MacGillivray (1988). Kurtosis: A critical review. *The American Statistician* **42**, 111-119.
- Baldi, P. and A.D. Long (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* **17**, 509-519.

- Bancroft, T.A. (1944). On biases in estimation due to use of preliminary test of significance. *The Annals of Mathematical Statistics* **15**, 190–204.
- Bickel, P.J. and K.A. Doksum (2001). *Mathematical Statistics*. Prentice Hall.
- Brewster, J.F. and J.V. Zidek (1974). Improving on equivariant estimators. *Annals of Statistics* **2**, 21–38.
- Broberg, P. (2003). Statistical methods for ranking differentially expressed genes. *Genome Biology* **4**, R41. (open access).
- Brown, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Annals of Mathematical Statistics* **39**, 29–48.
- Brown, L. (1990). Comment: Developments in decision-theoretic variance estimation. *Statistical Science* **5**, 90–120.
- Brown, M.B. and A.B. Forsythe (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association* **69**, 364–367.
- Casella, G. and J.T. Hwang (1986). Confidence sets and the Stein-effect. *Communications in Statistics: theory and methods* **15**, 2043–2063.
- Chereminsky, V., J. Feng, M. Rejali and B. Mishra (2003). Shrinkage-based similarity metric for cluster analysis of microarray data. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9668–9673.

- Chu, G., B. Narasimhan, R. Tibshirani and V.G. Tusher (n.d.). *SAM user's guide and technical document*. <http://www-stat.stanford.edu/tibs/SAM/>.
- Collins, R., S. Yusuf and R. Peto (1985). Overview of randomized trials of diuretics in pregnancy. *British Medical Journal* **290**, 17–23.
- Comander, J., S. Natarajan, M.A. Gimbrone and G. Garcia-Cardena (2004). Improving the statistical detection of regulated genes from microarray data using intensity-based variance estimation. *BMC Genomics* **5**, 1–23. Article No.17, published on Feb 27, 2004.
- Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- Cui, X. and G.A. Churchill (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology* **4**, 210. Epub 2003 Mar 17.
- Cui, X., J.T. Hwang, J. Qiu, N.J. Blades and G.A. Churchill (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* **6**, 59–75.
- Datta, S. and S. Datta (2005). Empirical Bayes screening of many p-values with applications to microarray studies. **21**, 1987–1994.
- DerSimonian, R. and N. Laird (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.
- Douglas, G.B. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics and Data Analysis* **50**, 2953–2957.

- Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of American Statistical Association* **99**, 96–104.
- Efron, B. and C.N. Morris (1973). Stein's estimation rule and its competitors - An empirical Bayes approach. *Journal of the American Statistical Association* **68**, 117–130.
- Efron, B. and C.N. Morris (1975). Data analysis using Stein's estimator and its generalization. *Journal of the American Statistical Association* **70**, 311–319.
- Efron, B., R. Tibshirani, J. Storey and V. Tusher (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* **96**, 1151–1160.
- Fan, J.Q. and Y. Ren (2006). Statistical analysis of DNA microarray data in cancer research. *Clinical Cancer Research* **12**(15), 4469–4473.
- Fisher, R.A. (1929). Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society, Series 2* **30**, 199–238.
- Fung, W.K. and T.S. Tsing (1998). Simulation study comparing tests for the equality of coefficients of variation. *Statistics in Medicine* **17**, 2003–2014.
- Giles, J.A. and D.E.A. Giles (1993). Pre-test estimation and testing in econometrics: Recent developments. *Journal of Economic Review Surveys* **7**, 145–197.
- Gui, J. and H. Li (2005). Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics* **21**(13), 3001–3008.

- Hardy, R. and S.G. Thompson (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* **15**, 619–629.
- Hedges, L.V. (1983). A random effects model for effect sizes. *Psychological Bulletin* **93**, 388–395.
- Hogg, R.V. (1972). More light on the kurtosis and related statistics. *Journal of the American Statistical Association* **69**, 909–923.
- Huang, X. and W. Pan (2002). Comparing three methods for variance estimation with duplicated high density oligonucleotide arrays. *Functional & Integrative Genomics* **2**(3), 126–133.
- Ibrahim, J.G., M.H. Chen and R.J. Gray (2002). Bayesian models for gene expression with DNA microarray data. *Journal of the American Statistical Association* **97**, 88–99.
- Jain, N., J. Thatte, T. Braciale, K. Ley, M. O’Connell and J. Lee (2003). Local-pooled error test for identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics* **19**, 1945–1951.
- James, W. and C. Stein (1961). Estimation with quadratic loss. In: *Proceeding of the Fourth Berkeley Symposium On Mathematical Statistics and Probability*. University of California Press, Berkeley, CA.
- Ji, Y., K.W. Tsui and K.M. Kim (2006). A two-stage empirical Bayes method for identifying differentially expressed genes. **50**, 3592–3604.

- Joanes, D.N. and C.A. Gill (1998). Comparing measures of sample skewness and kurtosis. *The Statistician* **47**, 183–189.
- Johnson, W. E., C. Li and A. Rabinovic (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**(1), 118–127.
- Judge, G.G. and M.E. Bock (1978). *The statistical implication of pre-test and Stein-rule estimators in econometrics*. North-Holland, Amsterdam.
- Judge, G.G. and R.C. Mittelhammer (2004). A semiparametric basis for combining estimation problems under quadratic loss. *Journal of the American Statistical Association* **99**, 479–487.
- Kamb, A. and A. Ramaswami (2001). A simple method for statistical analysis of intensity differences in microarray-derived gene expression data. *BMC Biotechnology* pp. 1–8. Epub 2001 Oct.2.
- Kendall, M., A. Stuart and J. K. Ord (1987). *Kendall's Advanced Theory of Statistics*. Oxford University Press, New York.
- Kendziorski, C.M., M.A. Newton, H. Lan and M.N. Could (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine* **22**, 3899–3914.
- Khan, B.U. and S.E. Ahmed (2003). Improved estimation of coefficient vector in a regression model. *Communications in Statistics - Simulation and Computation* **32**, 747–769.

- Klebanov, L., A. Gordon, Y. Xiao, H. Land and A. Yakovlev (2006). A permutation test motivated by microarray data analysis. *Computational Statistics and Data Analysis* **50**, 3619–3628.
- Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics* **22**, 290–299.
- Kubokawa, T. and M.S. Srivastava (2003). Estimating the covariance matrix: A new approach. *Journal of Multivariate Analysis* **86**, 28–47.
- Leek, J.T., E. Mosen, A.R. Dabney and J.D. Storey (2006). EDGE: Extraction and analysis of differential gene expression. *Bioinformatics* **22**(4), 507–508.
- Lehmann, E.L. (1999). *Elements of Large-Sample Theory*. p. 86. Springer.
- Levene, H. (1960). Robust test for equality of variance. In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow and H.B. Mann, Eds.). pp. 278–292. Stanford University Press.
- Lonnstedt, I. and T. Speed (2002). Replicated microarray data. *Statistica Sinica* **12**, 31–46.
- Maatta, J.M. and G. Casella (1990). Developments in decision-theoretic variance estimation. *Statistical Science* **5**, 90–120.
- Magnus, J.R. (1999). The traditional pretest estimator. *Theory of Probability and its Applications* **44**(2), 293–308.

- Moors, J.J.A. (1988). A quantile alternative for kurtosis. *The Statistician* **37**, 25–32.
- Morris, C.N. (1983). Parametric empirical Bayes inference: theory and applications. *The Journal of American Statistical Association* **78**, 47–55.
- Ohanti, K. (1999). MSE performance of a heterogeneous pre-test estimator. *Statistics & Probability Letters* **41**, 65–71.
- Reif, J. and K. Vlcek (2002). Optimal pre-test estimators in regression. *Journal of Econometrics* **110**, 91–102.
- Segal, M.R., K.D. Dahlquist and B.R. Conklin (2003). Regression approaches for microarray data analysis. *Journal of Computational Biology* **10**, 961–980.
- Sen, P.K. (1984). A James-Stein detour of U-statistics. *Communications in Statistics A* **13**, 2725–2747.
- Sidik, K. and J.N. Jonkman (2005). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society Serie C* **54**, 367–384.
- Sidik, K. and J.N. Jonkman (2006). Robust variance estimation for random effects meta-analysis. *Computational Statistics & Data Analysis* **50**, 3681–3701.
- Sinha, B.K. and M. Ghosh (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix and the generalized variance under entropy loss. *Statistical Decisions* **5**, 201–227.
- Stein, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. In: *Proceeding of the Fourth Berkeley Symposium On*

Mathematical Statistics and Probability. University of California Press, Berkeley, CA.

- Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Annals of the Institute of Mathematical Statistics* **16**, 155–160.
- Stigler, S.M. (1990). The 1988 Neyman memorial lecture: A galtonian perspective on shrinkage estimators. *Statistics in Science* **5**, 147–155.
- Thompson, S.G. and S.J. Pocock (1991). Can meta-analysis be trusted?. *Lancet* **338**, 1127–1130.
- Tibshirani, R (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288.
- Tibshirani, R., T. Hastie, B. Narasimhan and G. Chu (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6567–6572.
- Troyanskaya, O.G., K. Dolinski, A.B. Owen, R.B. Altman and D. Botstein (2003). A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *saccharomyces cerevisiae*). *Proceedings of the National Academy of Sciences of the United States of America* **100**(14), 8348–8353.
- Tusher, V.G., R. Tibshirani and G. Chu (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116–5121.

- Wang, J. (2000). Modeling and generating daily changes in market variables using a multivariate mixture of normal distributions. Working Paper.

Vita Auctoris

Lilua An was born in 1969 in LiaoNing, China. She obtained her B.Eng. in Chemical Engineering in 1991 from Tsinghua University, China and her M.Eng. in Chemical Engineering in 1994 from Tianjin University, China. She completed her M.Sc. in Statistics at West Virginia University, U.S.A. in 2000 and recently completed her Ph.D. in Statistics from the University of Windsor.