THREE ESTIMATION METHODS FOR THE NEGATIVE BINOMIAL PARAMETER k

By

ROBERT L. STEINER, JR.

Bachelor of Science
University of Oklahoma
Norman, Oklahoma
1978

Bachelor of Science
East Central University
Ada, Oklahoma
1984

Master of Arts
University of Oklahoma
Norman, Oklahoma
1987

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 1993

THREE ESTIMATION METHODS FOR THE

NEGATIVE BINOMIAL PARAMETER k

Thesis Approved:

R. Leroy Folks
Thesis Advisor

P. Larry Claypool

John C. Nolan

W. M. Warde

Thomas C. Collins
Dean of the Graduate College

ACKNOWLEDGEMENTS

The Statistics department is composed of people who truly want the student to learn and to succeed in the program. This is especially true of the members of my committee-- Dr. Folks, Dr. Claypool, and Dr. Warde. Each in his own way has made an extra effort to help me during my stay at Oklahoma State University--and these efforts are genuinely appreciated.

I would also like to thank the Statistics department and Dr. Folks for giving me the opportunity to teach and for the financial support that I received during the last three and one-half years. The teaching process helped me to continue learning not only about statistics but also about conveying the subject more effectively. This experience was invaluable.

My parents have always encouraged and supported me in my endeavors. As a child they provided a stable and loving home life that nurtured my growth. In retrospect I realize how lucky I was to have them as parents.

Esther, my fiancé and best friend, has given me love and support throughout the trials of graduate school. She is also a graduate student, and so she understands the pressures of overcoming the obstacles that we encounter in graduate school. I appreciate her love and assistance

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
II.	SOME PROPERTIES OF THE NEGATIVE BINOMIAL AND PRIOR RESEARCH IN THE ESTIMATION OF k	4
	Properties of the Negative Binomial Distribution	4
	Problems in the Estimation of k	8
	Previous Work in the Estimation of k . . .	15
III.	LARGE LIKELIHOOD ESTIMATION OF k	19
	An Intuitive Explanation of the Large Likelihood Estimator	19
	Simulation Methods	21
	Simulation Results	32
	Some Examples	36
	Existence of a Finite Positive Large Likelihood Estimator	37
	Asymptotic Properties of the Large Likelihood Extimator	40
	Consistency of the Estimator	40
	Asymptotic Normality for the Estimator	45
	Asymptotic Efficiency for the Estimator	46
	Discussion of Assumptions 1,2, and 3.	47
IV.	CHOICE OF OPTIMAL C AND THE ADJUSTED LARGE LIKELIHOOD ESTIMATOR	55
	Optimal Choice of C	55
	Adjusted Large Likelihood Estimator	57
	Table Summary of Simulation Results	59
	Interpretation of Simulation Results . . .	63
	Problems with the Adjusted Large Likelihood Estimator	64

Chapter Page

V. RE-WEIGHTING: A PRELIMINARY APPROACH TO SOLVE THE PROBLEM OF UNDER-DISPERSION 66

Simulation Results for Re-weighting 69
Conclusions Regarding the Re-weighting Technique 73

IV. DISCUSSION AND SUMMARY 75

Success of the Three Estimation Methods . . 75
Areas for Further Research 77

REFERENCES 82

APPENDIX A - APPROXIMATE PROBABILITY OF TYPE I AND TYPE II ERRORS 83

APPENDIX B - BOX PLOT COMPARISON OF LARGE LIKELIHOOD, MAXIMUM LIKELIHOOD, AND METHOD OF MOMENTS ESTIMATION 85

APPENDIX C - SAS-IML SOURCE PROGRAM FOR THE SIMULATION 99

LIST OF TABLES

Table Page

1. Percent of Under-Dispersed Samples in 7000 Samples of Size 50 11

2. Comparison of Estimated Bias for the LLE, MLE, and MME For 7000 Samples of Size 50 . . 23

3. Comparison of $S_{\hat{k}}$ for the LLE, MLE, and MME for 7000 Samples of Size 50 23

4. Comparison of the Estimated MSE for the LLE, MLE, and MME for 7000 Samples of Size 50 . . 24

5. Comparison of the Average $\hat{k}$ for the LLE, MLE, and MME for 2000 Samples of Size 50 25

6. Comparison of the Median $\hat{k}$ for the LLE, MLE, and MME for 2000 Samples of Size 50 25

7. Comparison of Q_1 for the LLE, MLE, and MME for 2000 Samples of Size 50 26

8. Comparison of Q_3 for the LLE, MLE, and MME for 2000 Samples of Size 50 26

9. Comparison of the Estimated MSE for the LLE, MLE, and MME for 2000 Samples of Size 50 . . 27

10. Comparison of $S_{\hat{k}}$ for the LLE, MLE, and MME for 2000 Samples of Size 50 27

11. Comparison of the 99^{th} Percentiles for the LLE, MLE, and MME for 2000 Samples of Size 50 28

12. Comparison of the Average $\hat{k}$ for the LLE, MLE, and MME for 2000 Samples of Size 250 28

13. Comparison of the Median $\hat{k}$ for the LLE, MLE, and MME for 2000 Samples of Size 250 29

Table | Page

14. Comparison of Q_1 for the LLE, MLE, and MME for 2000 Samples of Size 250 29

15. Comparison of Q_3 for the LLE, MLE, and MME for 2000 Samples of Size 250 30

16. Comparison of the Estimated MSE for the LLE, MLE, and MME for 2000 Samples of Size 250 . . 30

17. Comparison of $S_{\hat{k}}$ for the LLE, MLE, and MME for 2000 Samples of Size 250 31

18. Selected Examples to Illustrate the LLE 32

19. Optimal Values of C for the Nine Parameter Combinations 56

20. Median Second Derivatives for the Nine Parameter Combinations 57

21. Comparison of the Average Adjusted LLE With the Other Three Averages, 3000 Samples of Size 50 61

22. Comparison of the $S_{\hat{k}}$ for the AdjLLE, LLE, MLE, and MME for 3000 Samples of Size 50 61

23. Comparison of the Estimated MSE for the AdjLLE, MLE, LLE, and MME for 3000 Samples of Size 50 62

24. Comparison of the 75th Percentiles for the AdjLLE, LLE, MLE, and MME for 3000 samples of Size 50 62

25. An Example to Illustrate Re-Weighting 68

26. Comparison of Re-weighting Versus Conventional Methods for μ = 1 and k = 1, 2400 Samples of Size 35 70

27. Comparison of Re-weighting Versus Conventional Methods for μ = 1 and k = 3, 2400 Samples of Size 35 70

28. Comparison of Re-weighting Versus Conventional Methods for μ = 1 and k = 5, 2400 Samples of Size 35 71

29. Comparison of Re-weighting Versus Conventional Methods for μ = 3 and k = 1, 2400 Samples of Size 35 71

Table		Page

30. Comparison of Re-weighting Versus Conventional Methods for μ = 3 and k = 3, 2400 Samples of Size 35 72

31. Comparison of Re-weighting Versus Conventional Methods for μ = 3 and k = 5, 2400 Samples of Size 35 72

32. Comparison of Re-weighting Versus Conventional Methods for μ = 5 and k = 5, 2400 Samples of Size 35 73

LIST OF FIGURES

Figure		Page
1.	Probability Mass Functions for $\mu = 1$	5
2.	Probability Mass Functions for $\mu = 3$	6
3.	Probability Mass Functions for $\mu = 5$	7
4.	Sample Variance Plotted Versus Sample Mean to Illustrate Under-Dispersion	10
5.	Plot of the Log Likelihood as a Function of μ and k	13
6.	Plot of Log Likelihood as a Function of k for μ = Sample Mean When the Sample is Under-Dispersed	14
7.	Plot of the Log Likelihood as a Function of k for μ = Sample Mean for an Under-Dispersed Sample	19
8.	Plot of the Log Likelihood Function with a Pointed Maximum	20
9.	Plot of the Log Likelihood Function with a Sharp Maximum	56
10.	Maximization of the Log Likelihood in terms of μ for a Fixed Value of k	78
11.	Log Likelihood Function Following by its Partial Derivative Function	80
12.	Partial Derivative of the Log Likelihood Function	80
13.	Log Likelihood Function with Multiple Maxima . . .	81
14.	Box Plot Comparison of the Methods $\mu = 1$, $k = 1$.	86
15.	Box Plot Comparison of the Methods $\mu = 1$, $k = 3$.	87
16.	Box Plot Comparison of the Methods $\mu = 1$, $k = 5$.	88
17.	Box Plot Comparison of the Methods $\mu = 3$, $k = 1$.	89

Figure Page

18. Box Plot Comparison of the Methods $\mu = 3, k = 3$. 90

19. Box Plot Comparison of the Methods $\mu = 3, k = 5$. 91

20. Box Plot Comparison of the Methods $\mu = 5, k = 1$. 92

21. Box Plot Comparison of the Methods $\mu = 5, k = 3$. 93

22. Box Plot Comparison of the Methods $\mu = 5, k = 5$. 94

23. Box Plot Comparison of the Methods $\mu = 1, k = 1$. 95

24. Box Plot Comparison of the Methods $\mu = 1, k = 3$. 96

25. Box Plot Comparison of the Methods $\mu = 5, k = 3$. 97

26. Box Plot Comparison of the Methods $\mu = 5, k = 5$. 98

CHAPTER I

INTRODUCTION

The negative binomial distribution (NB) is used in entomology, forestry, and accident statistics to name a few fields. The probability mass function is given by

$$P(X = x) = \binom{k + x - 1}{k - 1} \left(\frac{k}{\mu + k} \right)^k \left(\frac{\mu}{\mu + k} \right)^x$$

μ, $k > 0$, $x = 0,1,2,\ldots$, and zero elsewhere. This uses the familiar parameterization with μ and k, having μ as a location parameter. For a fixed μ, k is a shape parameter. μ is adequately estimated by the sample mean. k is traditionally estimated by maximum likelihood or method of moments approaches. These methods give fairly good estimates for k when the sample variance substantially exceeds the sample mean, but problems do occur when the sample mean exceeds the sample variance (under-dispersion) or when the mean is only slightly smaller than the variance. In the former case (under-dispersed) the method of moments (MME) estimate is negative and the maximum likelihood estimate (MLE) does not exist (Levin and Reeds (1977)). Obviously this is very frustrating to the researcher who encounters a sample he believes is from a NB distribution, but is unable to estimate reasonably one of the parameters.

In the second case, in which the sample variance is only slightly larger than the sample mean, estimates for k are often volatile, exhibiting a large degree of variability. In simulations it is common for k estimates to exceed 500 or 1000 when in fact the true value is 5. Again such variability of estimates limits the usefulness of the methods.

The focus of this work is to offer a new method which provides a finite positive estimate for k, for any NB sample, regardless of the relative magnitude of the sample mean and variance. The large likelihood estimator (LLE), which is a variation of the MLE , not only provides a reasonable estimate for k in all NB samples , but also exhibits smaller variability and bias, as compared with the MLE and MME approaches. The intuitive explanation for the LLE, as well as simulation results and asymptotic properties are discussed in chapter III.

Chapter IV examines a refinement of the LLE, called the adjusted LLE. Simulation results are included to illustrate the comparative behavior of the LLE, adjusted LLE, MLE, and MME methods.

Chapter II describes the properties of the NB distribution, the problems encountered in the conventional estimation of k, and a review of previous work in the estimation of k.

An alternative method of estimation, developed in the early stages of research, called re-weighting, is examined in chapter V. The goal of this method is to increase the dispersion-to-mean ratio of the sample in order to decrease the volatility of the estimates. Simulation results comparing re-weighting versus MLE and MME are included to illustrate the reasonable degree of success enjoyed by the re-weighting method.

Finally in chapter VI, concluding remarks are offered concerning the LLE method, with specific emphasis on the benefits of its use and further areas of research that can be explored. In addition, other applications of the method beyond the NB distribution are discussed.

CHAPTER II

SOME PROPERTIES OF THE NEGATIVE BINOMIAL DISTRIBUTION AND PRIOR RESEARCH IN THE ESTIMATION OF k

In this chapter, selected properties of the negative binomial distribution are discussed. In addition, the difficulties in the estimation of k are outlined, as well as the previous work in the estimation of k.

Properties of the Negative Binomial Distribution

The negative binomial probability mass function is given by

$$P(X = x) = \binom{k + x - 1}{k - 1} \left(\frac{k}{\mu + k} \right)^k \left(\frac{\mu}{\mu + k} \right)^x$$

μ, $k > 0$, $x = 0,1,2,\ldots$, and zero elsewhere. The parameter μ is a location parameter; and for a fixed μ, k is a shape parameter. Plots of the function for a variety of values of μ and k illustrate the trend when k is increased and μ is held constant. Specifically, when k is small in relation to μ, the probability mass function is wedge-like with most of the mass being concentrated at the small x values. As k is increased the p.m.f.

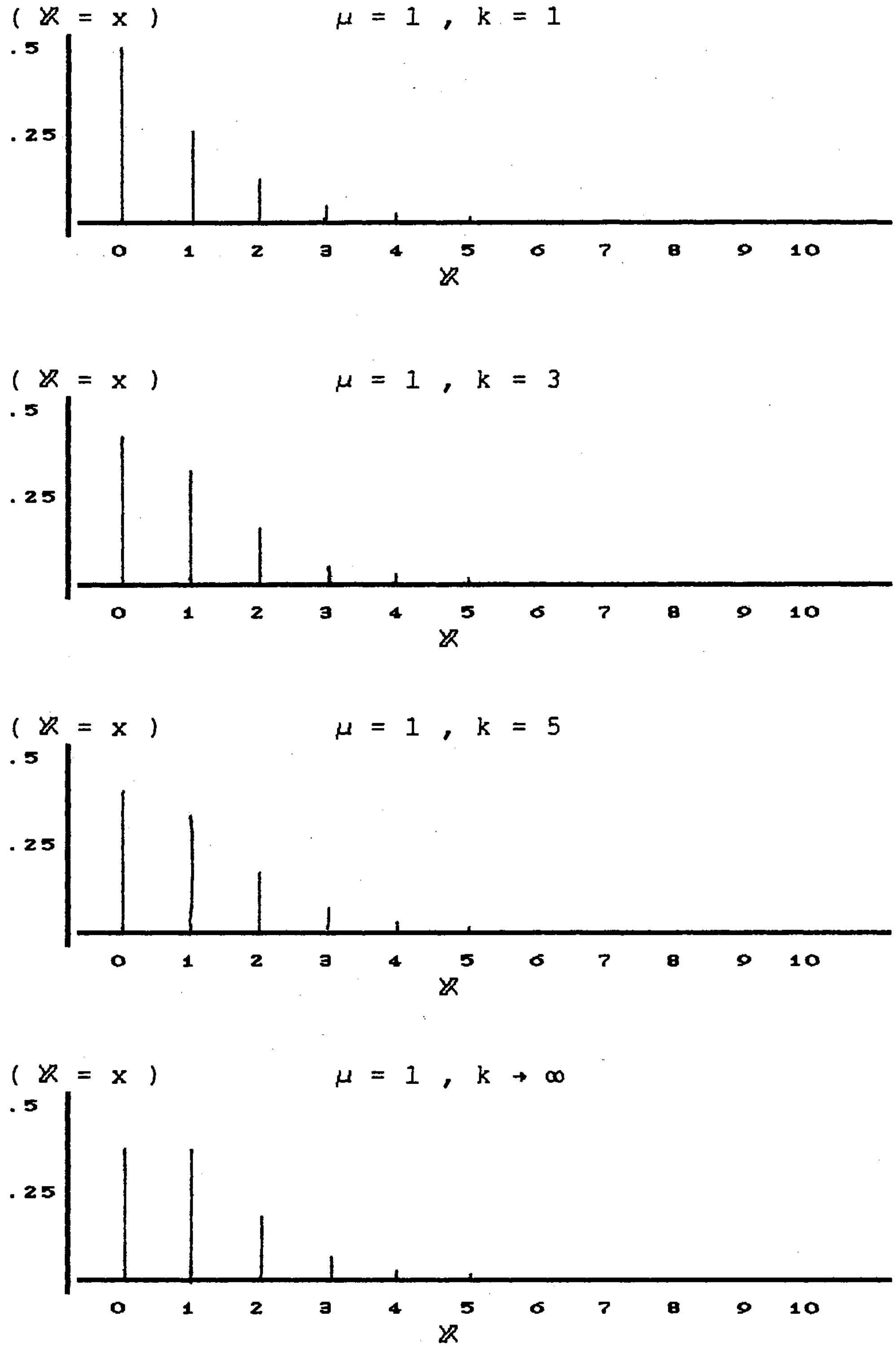

Figure 1. Probability Mass Functions for $\mu = 1$ and Increasing k.

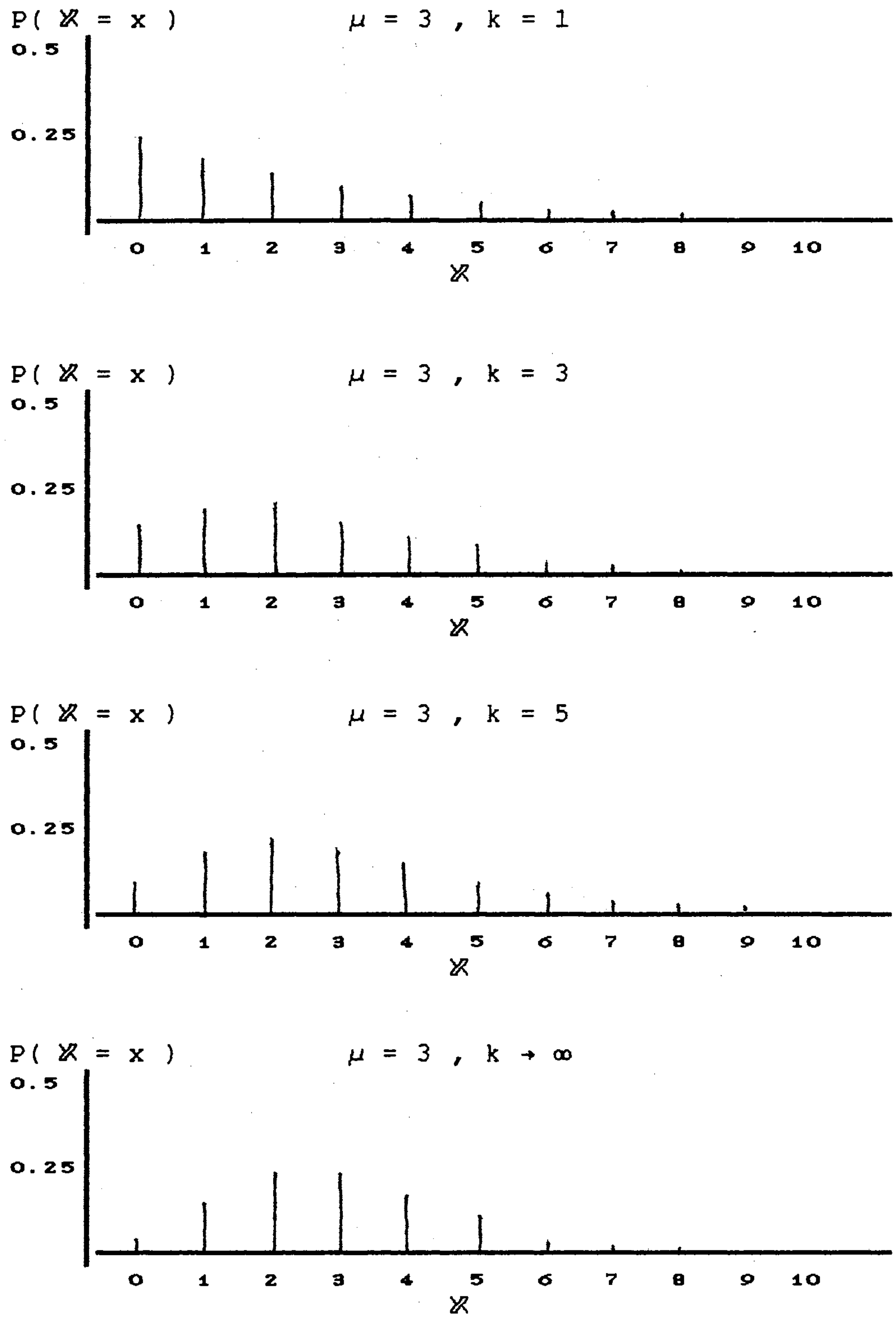

Figure 2. Probability Mass Functions for $\mu = 3$ and Increasing k.

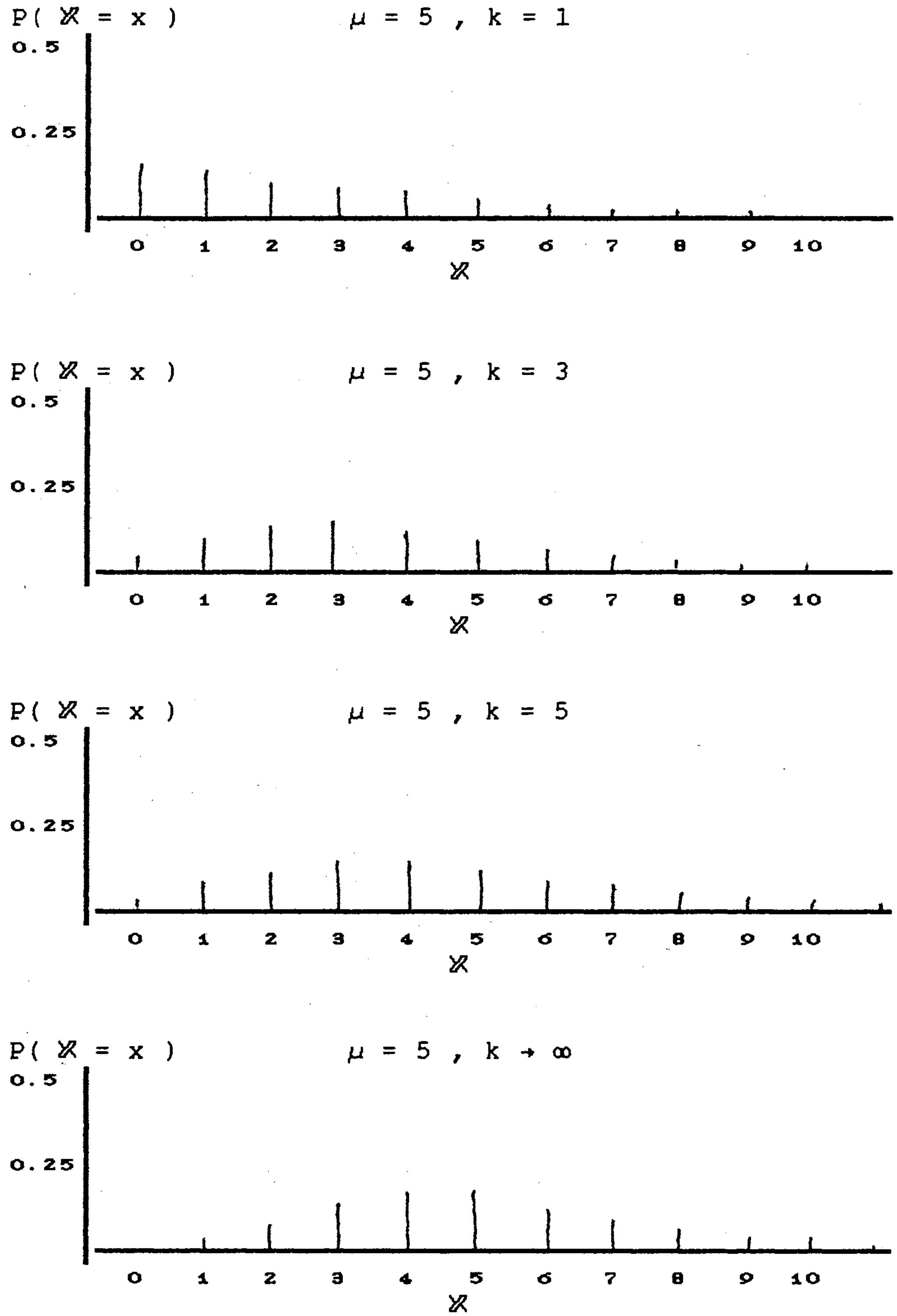

Figure 3. Probability Mass Functions for $\mu = 5$ and Increasing k.

becomes more mound shaped with the larger masses located near the value for μ. The graphs in figures 1,2, and 3 illustrate the trends in the appearance of the p.m.f. as k is increased and μ is fixed. The fourth graph on each page is the p.m.f. for the Poisson distribution, which is the resulting distribution if the NB parameter k is allowed to become infinitely large.

The variance is given by $\sigma^2 = \mu + \mu^2/k$, giving the over-dispersion (variance greater than the mean) which is characteristic of the distribution. The moment generating function is given by

$M(t) = (k/(\mu+k-\mu e^t))^k$ for t, $-h < t < h$ and $h > 0$.

When k is known the negative binomial distribution with parameter μ is a member of the exponential family. Hence $\bar{X}$ is a complete, sufficient statistic and is a mimimum variance unbiased estimator of μ.

Problems in the Estimation of k

Traditionally, k has been estimated by maximum likelihood or method of moments approaches. The method of moments estimator is given by

$$\hat{k} = \bar{X}^2 / (S^2 - \bar{X}) \quad . \qquad (2.1)$$

The maximum likelihood estimator is the solution in k to the following equation: (2.2)

$$n \ln(1+\bar{x}/k) = n_1 1/k + n_2(1/k + 1/(k+1)) + n_3(1/k + 1/(k+1) + 1/(k+2)) + \ldots$$

where n is the sample size, n_1 is the number of ones in the sample, n_2 is the number of twos in the sample, and so on. The next paragraphs specifically discuss the problems with the method of moments approach.

Upon viewing equation (2.1), several problems in estimating k become apparent. First, if the sample variance equals the sample mean then $\hat{k}$ does not exist. In addition if the sample variance is less than the sample mean then $\hat{k}$ is negative. These problems are illustrated in figure 4. In this graph each sample forms an ordered pair with the first coordinate being the sample mean and the second coordinate being the sample variance. The points originate from a NB distribution with $\mu = 1$ and $k = 3$. Note that the points falling below the line $S^2 = \overline{X}$ represent under-dispersed samples, which have a negative method of moments estimate and a non-existent maximum likelihood estimate. Points falling on the line have no method of moments estimate. Ordered pairs falling above the line originate from over-dispersed samples and hence their estimates for k can be found using MME or MLE approaches. It should be noted that the points falling just above the line yield very volatile estimates for k when the MME or MLE methods are used. In this context volatility describes estimates that are very large or very dispersed.

From the above discussion it is evident that under-dispersion presents a significant problem in the

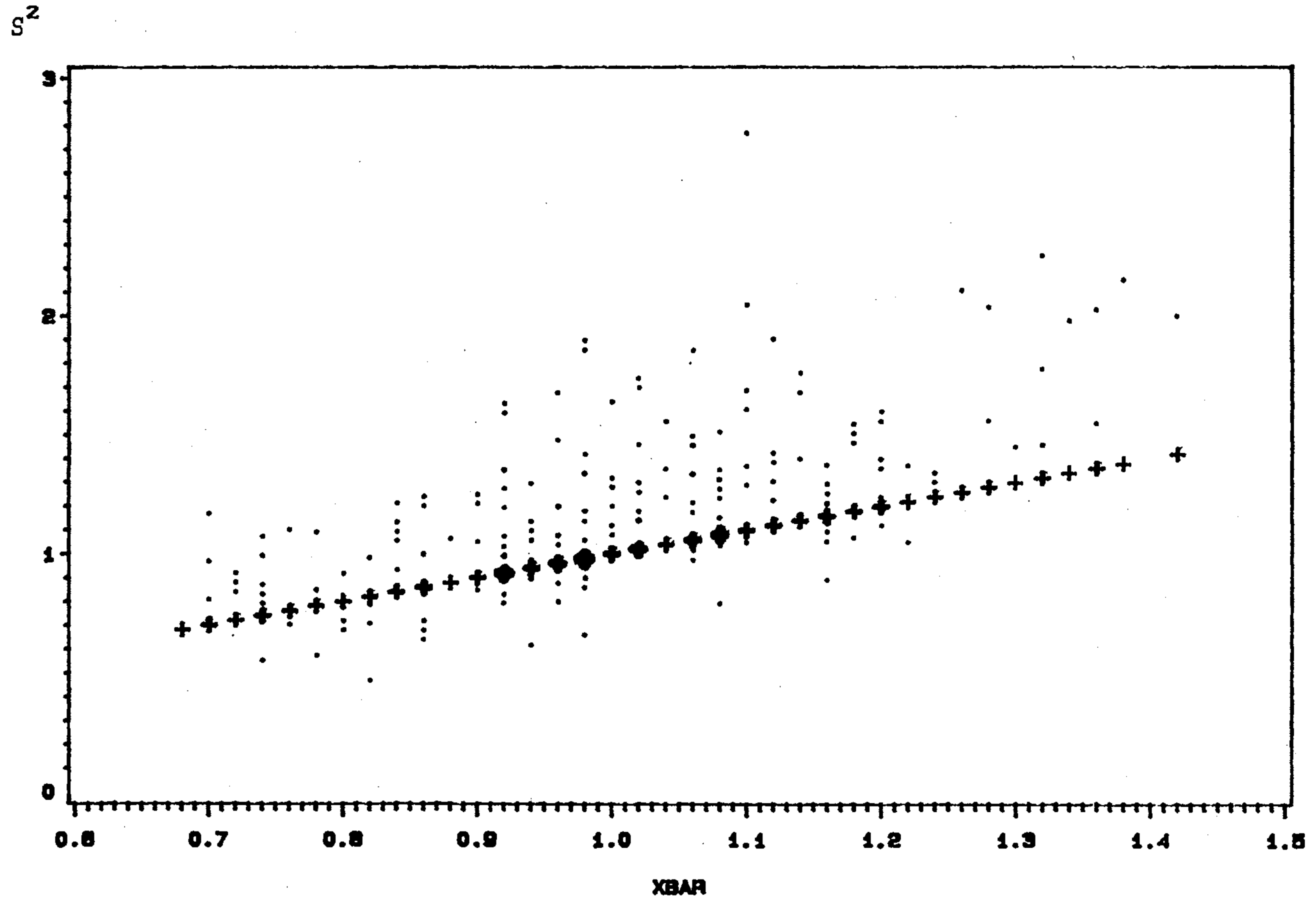

Figure 4. Sample Variance Plotted Versus Sample Mean

estimation process. This is true especially for certain choices of parameters. For example when $\mu = 1$, $k = 5$, and the sample size is 50 then about twenty-eight percent of the samples generated in simulations have a sample mean exceeding the sample variance. Table 1 summarizes the percent of under-dispersed samples for other parameter combinations.

TABLE 1

PERCENT OF UNDER-DISPERSED SAMPLES IN 7000 TRIALS FOR SAMPLE SIZE 50

μ	k	% UNDER-DISP
1	1	1.0
1	3	17.0
1	5	28.0
3	1	0.0
3	3	0.3
3	5	3.0
5	1	0.0
5	3	0.1
5	5	0.4

The table indicates a pronounced problem with under-dispersion when $\mu = 1$. If μ is held constant, then an increase in k is associated with larger percentages of bad samples. Another trend not described in the table is the increase in the percent of under-dispersed samples as the sample size decreases. Next turn to the maximum likelihood approach.

A second common approach to estimating k is maximum likelihood. This estimator is the solution in k for equation (2.2). A graph of the log likelihood function (as a function of μ and k) is presented in figure 5. In estimating k, $\hat{\mu}$ is set to $\bar{x}$ and equation (2.2) is solved for k.

In this method, problems occur when the biased sample variance is less than the sample mean, $\bar{X} > S^2(n-1)/n$. Specifically, if the variance is less than the mean the log likelihood function behaves asymptotically, reaching no maximum (see figure 6)-- hence no finite MLE is attainable (Levin and Reeds (1977)). As a result, maximization of the log likelihood with numerical algorithms yields a 'solution' (often in the millions) that is meaningless. The algorithm simply stops when the log-likelihood function is within the pre-determined tolerance of its supremum, even though no true solution for the partial derivative is attained.

In previous research these under-dispersed samples were discarded, and additional samples were generated until the desired simulation size was reached (Pieters *et al.* (1977)). Discarding ten to thirty percent (for some parameter combinations) of the original samples would certainly have an effect on the overall results of the simulation.

In the work of Anraku and Yanagimoto (1990), the distribution was parameterized with $\alpha = 1/k$. $\hat{\alpha}$ was

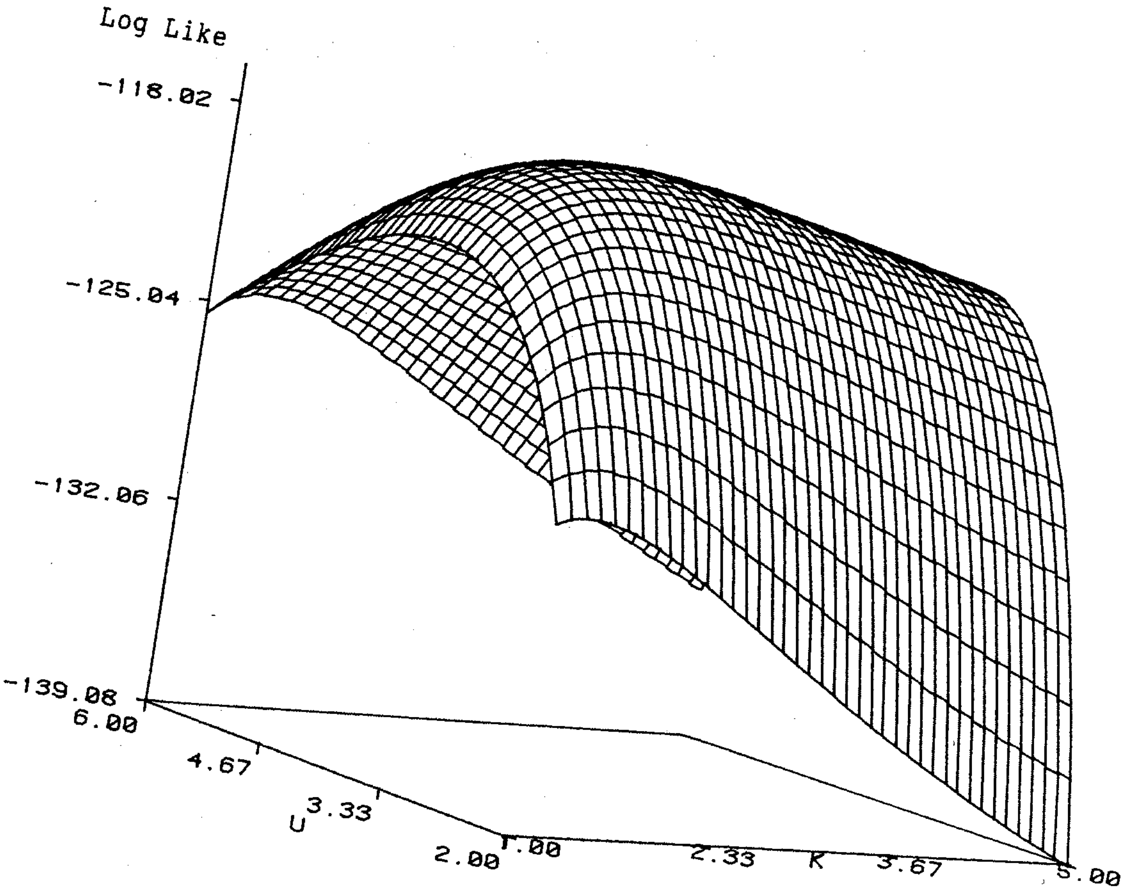

Figure 5. Log Likelihood as a Function of μ and k

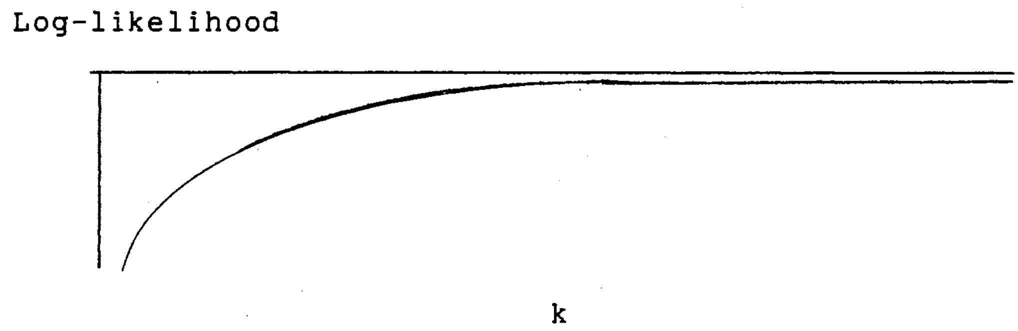

Figure 6. Log-likelihood Function of an Under-dispersed Sample

found by maximizing a conditional likelihood function, and $\hat{\alpha}$ was defined as zero for under-dispersed samples in the simulations.

It can be seen from the above examples that researchers often believed that under-dispersed samples did not belong in negative binomial simulations, and hence discarded such samples or considered them from a Poisson distribution.

Even when the sample variance exceeds the sample mean there are other problems encountered in estimating k. Results from simulations show large upward biases and large mean square errors for both the MME and MLE approaches. These problems are especially pronounced with the MLE. The proposed method, which will be described in subsequent paragraphs, is successful in reducing the bias, reducing the MSE, and eliminating the problem of under-dispersion.

Previous Work in the Estimation of k

In 1977 Pieters *et al.* used the method of moments and maximum likelihood approaches (and two other methods) to estimate k for samples of size 50, 100, and 200. Four hundred samples were generated via computer simulation for each parameter combination with under-dispersed samples being **discarded.** The average $\hat{k}$'s calculated from these simulations displayed the usual upward estimated bias encountered in small sample studies. The variability of the estimates was not specifically summarized in the text of the paper. The authors used t-tests to compare the average $\hat{k}$'s with the known values of the parameters. The MLE and MME approaches produced better results (in general) as compared with the other methods, but still the two preferred methods suffered from the aforementioned upward bias. Again, the variance of the $\hat{k}$'s was not specifically discussed in the paper.

In 1984, Willson *et al.*, used a multistage process to estimate k. In this method the MME was calculated for an initial sample. Five observations were then added to the initial sample and the MME was re-calculated. This process was repeated until sequentially adjacent estimates differed by less than a predetermined tolerance. This multi-stage approach produced lower estimated bias and lower estimated MSE as compared with MME and MLE methods. In some parameter combinations the improvement was dramatic. However these good results are often obtained by reaching very large sample sizes in the multi-stage sampling. For example the average sample size for $\mu = 1$ and $k = 5$ is 150, and the average sample size for $\mu = 1$ and $k = 3$ is 110. It should be noted that even with the multi-stage procedure, a few samples were discarded because of under-dispersion.

The next three groups of researchers parameterized the distribution by setting $\alpha = 1/k$, and focused their efforts on estimating α.

The first group, Clark and Perry (1989), maximized an extended quasi-likelihood function (MQLE) to estimate α. Negative $\hat{\alpha}$'s were allowed in the simulation results and no samples were discarded. Large simulations with 10,000 samples for each parameter combination were run for samples of size 50. The results were tabulated to include average $\hat{\alpha}$, standard deviation of the $\hat{\alpha}$'s, number of negative estimates, and the 50^{th}, 75^{th} and 25^{th} percentiles for the set of $\hat{\alpha}$'s generated for each parameter combination. The authors noted that the results for MQLE were slightly better than MME when the sample size is large. Smaller simulations were run (1,000 samples per simulation) for sample sizes 10, 20, 30 and 50 in order to study the behavior of the MQLE for small samples.

In 1990 Piegorsch followed up the work of Clark and Perry by estimating α with the maximum likelihood approach. This method was absent from the Clark and Perry paper. Piegorsch allowed negative $\hat{\alpha}$'s in his simulation results as long as the inequality was satisfied $\hat{\alpha} > -1/y_n$, where

y_n is the largest observed value in the random sample. The author did not specify how many samples were discarded because of this restriction. Sample sizes and size of simulations were exactly the same as used by Clark and Perry. Piegorsch summarizes his results by recommending the MQLE for small sample estimation and voices a slight preference for the MLE when samples are large, due to its established asymptotic properties.

A third group of researchers, Anraku and Yanagimoto, 1990, maximized a conditional likelihood function to estimate α. They treated under-dispersed samples as originating from a Poisson distribution. In these samples, $\hat{k}$ was **defined** as infinity, and $\hat{\alpha}$ was **defined** to be zero (it can be shown by simulations that the chance of misclassifying a negative binomial sample as Poisson is large, see appendix A for the results of such a simulation study). The tabulated results for simulation studies compared the Conditional Maximum Likelihood Estimator (CMLE), MLE, and MME. The authors concluded that the CMLE and MLE are comparable in estimating α for a single population (they voice a preference in favor of CMLE in terms of overall performance). Again note that in these studies, $\hat{\alpha}$ was **DEFINED** to be zero for under-dispersed samples.

Now turn from the comparison of different simulation studies to the examination of existence and uniqueness for the maximum likelihood estimator.

A common technique used throughout most of the previous papers is maximum likelihood estimation. A question that arises when using this method involves the existence and uniqueness of a maximum for the log-likelihood function. Anscombe (1950) posed such a question and offered a sketch of a proof for the existence of at least one positive finite solution in k for $\partial LN\ L(k,\hat{\mu} = \bar{x})/\ \partial k = 0$ when $(n-1)s^2/n > \bar{x}$. He also stated his belief that such a root is unique and that there is no root when $(n-1)s^2/n \leq \bar{x}$. In 1977 Levin and Reeds confirmed Anscombe's conjecture by proving that the likelihood function has at most one local maximum at $\hat{k}$. This maximum occurs for finite k iff $(n-1)s^2/n > \bar{x}$. So we know that there is a finite MLE iff the sample is over-dispersed and no finite MLE for k otherwise. This concludes the review of previous work. The next chapter is a description of the work completed by this author.

CHAPTER III

LARGE LIKELIHOOD ESTIMATION OF k

An Intuitive Explanation of the Large Likelihood Estimator

Suppose that a researcher wishes to maximize the log likelihood function with respect to k for an under-dispersed sample. A graph of this function, letting $\hat{\mu} = \bar{x}$, illustrates the problem of maximization (see figure 7). The usual method of setting the partial derivative equal to zero and solving for k gives no finite solution. Suppose instead that the partial derivative is set to a small positive constant, say c = 0.13 , and the equation is solved for k. This process yields a reasonable estimate for k in any NB sample, regardless of the relative magnitude of the sample mean and variance. This $\hat{k}$ is the large likelihood estimator.

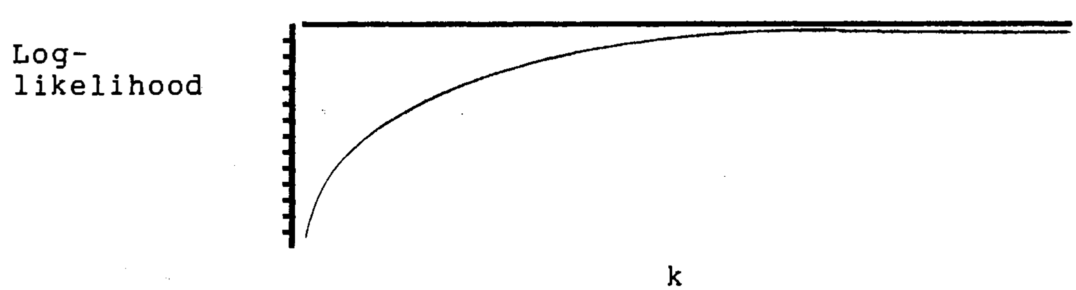

Figure 7. Log-likelihood Function for an Under-dispersed Sample

When this same process is applied to a log likelihood function that can be maximized, the LLE and MLE are close, especially when the concavity around the maximum is extreme. Specifically, samples from distributions with small k (in relation to μ) tend to generate log likelihood functions that display extreme negative concavity at the maximum (the function has a sharp point at the maximum), hence the LLE and MLE results for these parameter combinations are very similar (see figure 8). Samples that originate from distributions with a large k (in relation to μ) generate log-likelihood functions that are moderately concave at the maximum, and hence the differences between the MLE and LLE are more pronounced.

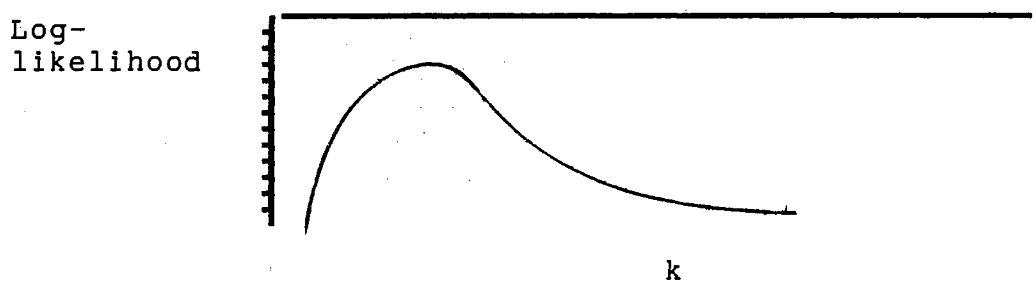

Figure 8. Log-likelihood Function With a Pointed Maximum

It should also be noted that choice of a positive constant c ensures that the LLE will be smaller than the MLE. In simulations this has the effect of producing an average LLE that is closer to the true k as compared with the average results from the MLE and MME approaches.

The large likelihood method also reduces the frequency of extremely large estimates encountered in the MLE and the MME approaches. The volatility of the MLE and MME estimates seen in the borderline NB-Poisson samples seems to be less pronounced with the LLE.

A further advantage of the large likelihood approach is that a positive $\hat{k}$ is always possible (even if the sample is under-dispersed). This is in contrast to Clark and Perry (1989) and Piegorsch (1990) . These researchers parameterized the distribution with $\alpha = 1/k$. Clark and Perry maximized the extended quasi-likelihood to estimate α. No samples were discarded but negative $\hat{\alpha}$'s were encountered in the simulations. Piegorsch investigated the MLE under the $\alpha = 1/k$ parameterization. Again, negative $\hat{\alpha}$'s were allowed as long as $\hat{\alpha} > -1/y_n$, where y_n is the largest observation in the sample.

The previous paragraphs described the intuitive justification for large likelihood estimation, as well as some of its benefits. The results of computer simulations that were run in order to compare the performance of the LLE with the MM and ML methods are included next in this paper.

Simulation Methods

In order to study the behavior of the LLE, as compared with the MLE and MME, three sets of simulations were run. The first set consisted of nine simulations of 7000 samples each (sample size equals 50). These simulations were run for the LLE only, and the results were compared with the results for the MME and MLE published by Willson *et al.* (1984). The nine cases mentioned above are achieved by pairing the values of 1,3 and 5 for μ with the values 1,3, and 5 for k. The results of these nine simulations are summarized in tables 2, 3, and 4.

Tables 2,3 and 4 compare Bias, $S_{\hat{k}}$, and MSE, respectively for large likelihood estimation versus method of moments and maximum likelihood estimation. The results for the MME and MLE originate from 10,000 samples of size 50 (Willson *et al.* (1984)). Under-dispersed samples were discarded. The simulations were run until 10,000 good samples were generrated. A large percent of the samples was thrown out when small μ values (1 and 3) were combined with large k (k = 5). The results for the LLE column are based on 7000 samples of size 50. No samples were discarded from the simulation because of under-dispersion. The estimate for k is the solution in k for

$$\frac{\partial L(k, \bar{x} = \hat{\mu})}{\partial k} = 0.13$$

the partial derivative of the log-likelihood function. The constant 0.13 was chosen because it yielded favorable results in preliminary simulations.

In the second set of simulations, the three methods of estimation were directly compared. Specifically, a sample was generated, and the LLE, MLE, and MME were found

TABLE 2

COMPARISON OF THE BIAS IN THE LLE, MLE, AND MME, N = 50

μ	K	MME	MLE	LLE
1	1	0.57	0.66	0.25
1	3	1.40	2.60	0.03
1	5	0.84	2.90	-1.30
3	1	0.17	0.11	0.09
3	3	1.1	1.2	0.31
3	5	2.9	4.7	-0.05
5	1	0.14	0.09	0.07
5	3	0.54	0.55	0.24
5	5	1.7	2.3	0.24

TABLE 3

COMPARISON OF $S_{\hat{k}}$ FOR THE LLE, MLE, AND THE MME, N = 50

μ	k	MME	MLE	LLE
1	1	1.8	4.2	0.75
1	3	4.5	12.0	1.55
1	5	5.5	15.0	1.66
3	1	0.43	0.38	0.36
3	3	3.9	7.4	1.40
3	5	9.4	27.0	2.00
5	1	0.36	0.3	0.29
5	3	1.7	1.7	1.17
5	5	7.0	18.0	1.97

TABLE 4

COMPARISON OF THE MSE FOR THE LLE, MLE, AND MME WITH N = 50

μ	k	MME	MLE	LLE
1	1	3.6	18.0	0.63
1	3	23.0	150.0	2.39
1	5	31.0	240.0	4.44
3	1	0.22	0.15	0.14
3	3	16.0	56.0	2.05
3	5	97.0	740.0	4.00
5	1	0.15	0.1	0.09
5	3	3.2	3.3	1.44
5	5	52.0	330.0	3.95

for that sample. Under-dispersed samples produced a finite LLE but the MLE and MME were designated as missing values. This plan was carried out for 2,000 samples of size 50. Again the same nine parameter combinations were studied. The results of this second set of simulations are summarized in tables 5 through 11. An accompanying set of box plots illustrates quartile information for these simulations (see appendix B).

In tables 5 through 11 the percent under-dispersion column represents the percent of samples where no MLE was possible or where the MME was negative. Again, the LLE was found for all samples. After 2000 samples were estimated the average $\hat{k}$, median $\hat{k}$, estimated MSE, $S_{\hat{k}}$, Q_1, Q_3, and 99^{th} percentile were found for the three methods.

TABLE 5

AVERAGE $\hat{k}$ FOR THE LLE, MLE, AND MLE
FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	1.263	1.765	1.566	0.95
1	3	2.982	5.771	4.570	15.00
1	5	3.745	8.670	6.050	28.15
3	1	1.104	1.123	1.171	0.00
3	3	3.258	4.218	3.994	0.25
3	5	4.970	8.827	7.827	2.60
5	1	1.074	1.086	1.137	0.00
5	3	3.202	3.451	3.438	0.00
5	5	5.127	6.894	6.520	0.15

TABLE 6

MEDIAN $\hat{k}$ FOR THE LLE, MLE, AND MME
FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	1.041	1.069	1.125	0.95
1	3	2.629	2.877	2.903	15.00
1	5	3.545	4.062	3.912	28.15
3	1	1.046	1.060	1.096	0.00
3	3	2.930	3.138	3.132	0.25
3	5	4.502	5.323	5.222	2.60
5	1	1.032	1.040	1.086	0.00
5	3	3.012	3.151	3.118	0.00
5	5	4.707	5.245	5.168	0.15

TABLE 7

25^{th} PERCENTILE FOR THE LLE, MLE, AND MME
FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	0.768	0.782	0.803	0.95
1	3	1.825	1.872	1.892	15.00
1	5	2.412	2.442	2.391	28.15
3	1	0.852	0.860	0.865	0.00
3	3	2.309	2.415	2.376	0.25
3	5	3.488	3.856	3.769	2.60
5	1	0.870	0.877	0.879	0.00
5	3	2.433	2.510	2.474	0.00
5	5	3.790	4.066	3.980	0.15

TABLE 8

75^{th} PERCENTILE FOR THE LLE, MLE, AND MME
FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	1.511	1.603	1.670	0.95
1	3	3.889	5.357	5.060	15.00
1	5	4.902	7.729	6.894	28.15
3	1	1.274	1.296	1.381	0.00
3	3	3.887	4.433	4.459	0.25
3	5	6.027	8.304	7.962	2.60
5	1	1.231	1.245	1.340	0.00
5	3	3.707	3.955	3.985	0.00
5	5	6.052	7.278	7.092	0.15

Table 9

ESTIMATED MSE FOR THE LLE, MLE, AND MME FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	0.660	60.713	3.974	0.95
1	3	2.305	101.635	25.339	15.00
1	5	4.421	359.921	36.331	28.15
3	1	0.141	0.160	0.224	0.00
3	3	1.918	67.221	16.989	0.25
3	5	4.015	170.011	79.056	2.60
5	1	0.089	0.096	0.152	0.00
5	3	1.213	2.345	2.363	0.00
5	5	3.600	91.954	40.307	0.15

TABLE 10

$S_{\hat{k}}$ FOR THE LLE, MLE, AND MME FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	0.769	7.756	1.912	0.95
1	3	1.518	9.695	4.783	15.00
1	5	1.688	18.619	5.937	28.15
3	1	0.361	0.381	0.442	0.00
3	3	1.361	8.109	4.001	0.25
3	5	2.004	12.467	8.431	2.60
5	1	0.289	0.298	0.365	0.00
5	3	1.082	1.463	1.473	0.00
5	5	1.893	9.402	6.165	0.15

TABLE 11

99^{th} PERCENTILE FOR THE LLE, MLE, AND MME FROM 2000 SAMPLES OF SIZE 50

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	4.255	11.316	9.800	0.95
1	3	7.197	46.738	23.529	15.00
1	5	8.025	60.984	29.481	28.15
3	1	2.322	2.434	2.704	0.00
3	3	7.874	18.836	16.268	0.25
3	5	10.735	67.623	48.580	2.60
5	1	1.987	2.036	*	0.00
5	3	6.707	8.571	8.808	0.00
5	5	11.214	30.742	30.637	0.15

TABLE 12

AVERAGE $\hat{k}$ FOR THE LLE, MLE, AND MME FROM 1500 SAMPLES OF SIZE 250

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	1.054	1.063	1.081	0.00
1	3	3.171	3.871	3.823	0.00
1	5	4.822	8.529	7.893	3.60
3	1	1.018	1.021	1.039	0.00
3	3	3.055	3.108	3.118	0.00
3	5	5.121	5.447	5.421	0.00
5	1	1.013	1.015	1.026	0.00
5	3	3.061	3.091	3.091	0.00
5	5	5.067	5.203	5.204	0.00

TABLE 13

MEDIAN $\hat{k}$ FOR THE LLE, MLE, AND MME FROM 1500 SAMPLES OF SIZE 250

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	1.014	1.021	1.040	0.00
1	3	2.867	3.032	3.055	0.00
1	5	4.507	5.292	5.130	3.60
3	1	1.006	1.008	1.028	0.00
3	3	2.968	3.011	3.008	0.00
3	5	4.955	5.166	5.133	0.00
5	1	1.004	1.006	1.018	0.00
5	3	3.031	3.059	3.062	0.00
5	5	4.929	5.037	5.017	0.00

TABLE 14

25^{th} PERCENTILE FOR THE LLE, MLE, AND MME FROM 1500 SAMPLES OF SIZE 250

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	0.884	0.889	0.887	0.00
1	3	2.281	2.353	2.364	0.00
1	5	3.426	3.712	3.667	3.60
3	1	0.921	0.923	0.919	0.00
3	3	2.628	2.659	2.631	0.00
3	5	4.246	4.371	4.347	0.00
5	1	0.934	0.936	0.924	0.00
5	3	2.730	2.751	2.724	0.00
5	5	4.392	4.471	4.457	0.00

TABLE 15

75^{th} PERCENTILE FOR THE LLE, MLE, AND MME FROM 1500 SAMPLES OF SIZE 250

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	1.177	1.187	1.221	0.00
1	3	3.687	4.113	4.091	0.00
1	5	5.882	8.150	8.090	3.60
3	1	1.100	1.103	1.142	0.00
3	3	3.383	3.446	3.473	0.00
3	5	5.782	6.131	6.113	0.00
5	1	1.082	1.084	1.122	0.00
5	3	3.343	3.380	3.409	0.00
5	5	5.588	5.746	5.802	0.00

TABLE 16

ESTIMATED MSE FOR THE LLE, MLE, AND MME FROM 1500 SAMPLES OF SIZE 250

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	0.062	0.067	0.085	0.00
1	3	1.666	14.192	11.610	0.00
1	5	3.478	197.5	105.8	3.60
3	1	0.019	0.019	0.030	0.00
3	3	0.367	0.421	0.492	0.00
3	5	1.577	4.150	3.693	0.00
5	1	0.013	0.013	0.022	0.00
5	3	0.215	0.232	0.274	0.00
5	5	0.891	1.101	1.191	0.00

TABLE 17

$S_{\hat{k}}$ FOR THE LLE, MLE, AND MME
FROM 1500 SAMPLES OF SIZE 250

μ	k	LLE	MLE	MME	% UNDER-DISP MLE & MME
1	1	0.244	0.251	0.281	0.00
1	3	1.279	3.666	3.307	0.00
1	5	1.857	13.608	9.875	3.60
3	1	0.138	0.139	0.171	0.00
3	3	0.603	0.640	0.691	0.00
3	5	1.250	1.988	1.875	0.00
5	1	0.114	0.114	0.147	0.00
5	3	0.460	0.473	0.516	0.00
5	5	0.942	1.030	1.072	0.00

Finally, a third set of simulations was run to compare the behavior of LLE, MLE,and MME for large samples (n = 250). These simulations were carried out exactly as the second set but with a larger sample size. The results are summarized in tables 12 through 17. The constant C=0.13 was used in both the second and third sets of simulations because it produced good results in preliminary trials.

All of the simulations were run on PC-SAS using the IML procedure. Newton's method was used to find the MLE and LLE.

Simulation Results

To summarize the results from the first set of simulations (tables 2, 3, and 4), note that the LLE exhibited a smaller estimated bias in all parameter combinations except $\mu = 1$ and $k = 5$. In this case the MME yielded a smaller estimated bias. For certain parameter combinations ($\mu = 1$ and $k = 3$; $\mu = 3$ and $k = 3$; and $\mu = 5$ and $k = 5$) the bias from the LLE is dramatically smaller than the results from the other two methods.

The LLE produces a marked decrease in MSE in comparison with the MLE and MME approaches. This reduction occurred for all parameter combinations. Similar results hold for $S_{\hat{k}}$. The previous paragraphs discussed the results for the first set of simulations, where the large likelihood estimator (run in 1992) was compared to the maximum likelihood and method of moments estimators published in 1984 by Willson *et al.* Now turn to the discussion of the second set of simulations, in which a sample was subjected to all three methods of estimation (LLE, MLE, and MME) within the same simulation.

In table 5 the average $\hat{k}$'s for $n = 50$ are presented. Note that the average LLE is closer to the true value of k for all parameter combinations except $\mu = 1$ and $k = 5$. In this case the average MME is closer to 5 than the other two methods. It should also be noted that the differences between the three methods are more pronounced (with the LLE having the advantage in most cases) for combinations $\mu = 5$ and $k = 5$, $\mu = 3$ and $k = 5$, $\mu = 3$ and $k = 3$, and $\mu = 1$ and $k = 3$. In contrast, the three methods behave similarly when k is small ($k = 1$). This similarity should be expected, especially between the LLE and MLE, since the samples from distributions with a small k 's tend to give rise to log likelihood functions which exhibit extreme concavities around the maximum. Hence the LLE and MLE are close.

Next turn to the variability of the $\hat{k}$'s. Table 9 summarizes the estimated MSE's for samples of size 50. In all nine simulations the MSE's for the LLE's are smaller than the estimates for the other two methods, with most of the simulations showing a dramatic difference. Obviously the LLE's do not exhibit the extreme variability seen in the MLE's. As with the results for the average $\hat{k}$'s, the three methods behave similarly when k is small in relation to μ.

The results for $S_{\hat{k}}$ (table 10) correspond to those of the estimated MSE 's. Again in most cases the standard errors for the LLE 's are much smaller than the estimates for the other two methods. Only when k is small in relation to μ do the three methods behave similarly.

For the 75^{th} percentiles (Table 8), the Q_3LLE is smaller than the other two Q_3's. This would indicate that the benfits seen in the LLE are gained not only by eliminating the extreme values seen in the MLE and MME, but also by producestimates that are in general less variable than the other two methods. The 99^{th} percentiles (table 11) for the LLE are in all cases smaller, and in most cases dramatically smaller than the other 99^{th} percentiles.

Before moving to the discussion of the next set of simulations it should be noted that a decrease in sample size is associated with an increase in the percent of under-dispersed samples. This trend is reflected in these sets of simulations (compare the percent under-dispersed for n = 50 with n = 250). Another trend that appears is that the smaller sample size accentuates the difference between the LLE and the other two methods, favoring the LLE. As you will see in the next set of simulations, larger sample sizes produce results that are more alike for the three methods. So the larger sample sizes produce similar results for the LLE, MLE, and MME, but the benefits of the LLE become more evident for the smaller sample sizes (n = 50). The next discussion will focus on the comparative behavior of the three methods when sample size is 250.

Table 12 summarizes the average $\hat{k}$'s for n = 250. A familiar trend continues in the sense that the results for the three methods are very similar when k is small. On the other hand the largest differences between the methods occur when k = 5, with the LLE giving the better results in every case. It should be noted that the difficulty with the average LLE for $\mu = 1$, $k = 5$, and $n = 50$ is no longer a problem in the simulations for $n = 250$. Next turn to the variability of the estimates.

The estimated MSE 's are given in table 16. Again the results for the LLE, MLE, and MME are very similar in the simulations with $k = 1$. Differences in performance occur when $k = 5$, with the LLE being less variable than the other two methods. The same can be said for the combination $\mu = 1$ and $k = 3$. Corresponding results hold for the estimated standard errors (see table 17).

The **medians** (table 13) for all the methods are reasonably close to the true k, with the median LLE being the smallest in all cases. The worst results are obtained (for any method) when $\mu = 1$, or 3 , and $k = 5$. The median LLE is especially low when $\mu = 1$ and $k = 5$. Recall however that this value is based on the full 1500 samples, whereas the medians for the other two methods are based on 96.4 % of 1500. The 75^{th} percentiles (table 15) are comparable for the three methods except when $\mu = 1$ and $k = 5$. In this case Q_3LLE is 5.882 as compared with Q_3MLE = 8.15 and Q_3MME = 8.09.

It is clear that increasing the sample size decreases the differences between the three methods. Taking the larger n improves the performance of all the methods by decreasing Bias and decreasing the variability of the estimates.

Some Examples

The samples included in this section were selected from simulations in order to illustrate the properties of the LLE, with special emphasis on its behavior for under-dispersed or borderline NB-Poisson samples. The values for the parameters μ and k represent the numbers used in the simulations.

TABLE 18

SPECIFIC EXAMPLES FOR THE LLE, MLE, AND MME FOR N = 50

μ	k	X	S_1^2	MME	MLE	LLE
3	1	3.16	11.614	1.149	1.254	1.236

X:	0	1	2	3	4	5	6	7	8	14	16
FREQ:	9	13	5	7	3	2	4	2	3	1	1

μ	k	X	S_1^2	MME	MLE	LLE
1	5	0.94	0.896	-34.916	.	5.155

X:	0	1	2	3	4
FREQ:	19	19	9	2	1

μ	k	X	S_1^2	MME	MLE	LLE
1	5	1.02	1.059	16.993	20.885	4.236

X:	0	1	2	3	4
FREQ:	20	14	12	3	1

Existence of a Finite Positive Large Likelihood Estimator

At present, every sample generated in the course of the simulations yielded a finite positive LLE for k. However it is important to prove that a finite positive LLE exists for any negative binomial sample. This amounts to showing that

$$\frac{\partial L(k, \hat{\mu} = \bar{x})}{\partial k} - C = 0$$

has a finite positive solution in k, where C is a small positive constant. For convenience of notation let

$$\frac{\partial L(k, \hat{\mu} = \bar{x})}{\partial k} \equiv \frac{\partial L}{\partial k}$$

The first part of the proof will show that the derivative of the log-likelihood is positive for some small positive k, *i.e.* show that

$$\frac{\partial L}{\partial k} > 0 \qquad \text{for some small positive k.}$$

This requires equations (2.2) and (2.3) from Theorem 2.1, published in Willson *et al.* (1986). I am expanding these two equations for clarity.

$$\frac{\partial L}{\partial k} = n_1(1/k) + n_2(1/k + 1/(k+1)) +$$

$$n_3(1/k + 1/(k+1) + 1/(k+2)) - n \ln(1 + \bar{x}/k) \qquad (3.1)$$

For small positive k we have $\ln(1 + k/\bar{x}) < 1$.

$\Rightarrow \quad \ln(\bar{x} + k) - \ln(\bar{x}) < 1.$

$\Rightarrow \quad \ln(\bar{x} + k) - \ln(k) < 1 + \ln(\bar{x}) - \ln(k).$

$\Rightarrow \quad \ln((\bar{x} + k) / k) < 1 + \ln(\bar{x} / k).$

$\Rightarrow \quad \ln(1 + \bar{x}/k) < 1 + \ln(\bar{x}) + \ln(1/k).$

Let $b = \ln(1/k)$, hence the right hand side of the inequality becomes $1 + \ln(\bar{x}) + b$, which is linear in b. Recall that a linear function can be bounded above (for large enough b) by an exponential function, *i.e.*

$$1 + \ln(\bar{x}) + b < (n_1 + n_2 + \ldots\ldots)\, e^b / n$$

$$= (n_1 + n_2 + \ldots) / (nk).$$

But b is large when k is a small positive, therefore

$$\ln(1 + \bar{x}/k) < \ln(\bar{x}/k) + 1 = \ln(\bar{x}) + b + 1$$

$$< (n_1 + \ldots\ldots)/(nk)$$

$$< (n_1(1/k) + n_2(1/k + 1/(k+1)) +$$

$$n_3(1/k + 1/(k+1) + 1/(k+2)) + \ldots)/n$$

the above inequality is true for some small postitive k.

For the second part of the proof we want to show that the derivative gets epsilon close to 0 as k approaches positive infinity. Begin by taking the following limit.

$$\lim_{k \to \infty} \frac{\partial L}{\partial k} = \lim_{k \to \infty} (n_1(1/k) + n_2(1/k + 1/(k+1)) + \ldots)$$

$$- \lim_{k \to \infty} n \ln(1 + \bar{x} / k) = 0$$

Denote $n_1(1/k) + n_2(1/k + 1/(k+1)) + \ldots$ as $\Sigma 1$ and $n \ln(1 + \bar{x}/k)$ as LN. If $\Sigma 1 >$ LN for large k then the proof is completed. On the other hand, if LN $> \Sigma 1$ for large k then the proof can be completed by arguing that the partial derivative is continuous in k for $k > 0$, the partial derivative is positive for some small positive k, and the derivative is negative for some large k, therefore the partial derivative is equal to some small positive b for some $k > 0$. Hence the LLE exists for all negative binomial samples. So a finite, positive $\hat{k}$ can be found using the large likelihood method for over or under-dispersed negative binomial samples, giving the large likelihood method an advantage over the method of moments and maximum likelihood approaches. ∎

In the previous sections the properties of the large likelihood estimator were considered for finite sample sizes. The next step is to examine the large sample properties of the estimator. Specifically, the consistency, asymptotic normality, and asymptotic efficiency will be studied for the LLE. Simulation results suggest consistency when the average $\hat{k}$'s are compared for n = 50 versus n = 250. It is the goal of the next section to establish the above large sample properties with mathematical rigor. The proof, which is a variation of Cramer's proof (1946), starts with the question of consistency, progresses to asymptotic normality, and finishes with asymptotic efficiency.

Asymptotic Properties of the Large Likelihood Estimator

The results of simulations suggest that the average LLE moves closer to the true value for k as the sample size n increases. This trend can be seen when comparing the results for simulations of n = 250 versus n = 50 (tables 12 and 5 respectively). In response to this evidence of consistency this section will offer a proof of consistency for the LLE, as well as arguments to establish the asymptotic normality and asymptotic efficiency of the estimator. The proof is based on Cramer's argument (1946) to show consistency, asymptotic normality, and asymptotic efficiency for the maximum likelihood estimator. The appropriate changes have been incorporated into Cramer's proof in order to demonstrate the results for the LLE. Also some steps in the proof have been expanded for the sake of clarity.

<u>Consistency of the Estimator</u>

In this section it will be shown that the solution for

$$\frac{\partial \text{ LN L}}{\partial k} - C = 0$$

converges in probability to the true value of k, as $n \to \infty$, where L is the likelihood function for a sample of size n, and C a small positive constant. f will denote the negative binomial probability mass function where $f = f(x ; \mu , k)$ for $x = 0,1,2,3,\ldots.$, $\mu > 0$ and $k > 0$. The following three assumptions are made for the proof.

1) For all $x \in \{0, 1, 2, 3, \ldots\}$

$$\frac{\partial \, LN\, f}{\partial k}, \quad \frac{\partial^2 LN\, f}{\partial k^2}, \quad \frac{\partial^3 LN\, f}{\partial k^3} \quad \text{exist for all } k \in A$$

where $A = (a,b)$ and $b > a > 0$.

2) for all $k \in A$,

$$\left| \frac{\partial f}{\partial k} \right| < G_1(x), \quad \left| \frac{\partial^2 f}{\partial k^2} \right| < G_2(x), \text{ and } \left| \frac{\partial^3 LNf}{\partial k^3} \right| < H(x)$$

where $G_1(x)$ and $G_2(x)$ are integrable functions over $(-\infty, \infty)$, or in the case of discrete random variables

$$\sum_{x=0}^{\infty} G_1(x) < \infty \text{ and } \sum_{x=0}^{\infty} G_2(x) < \infty. \text{ Also } \sum_{x=0}^{\infty} H(x)\, f < M$$

where M is independent of k and $M < \infty$.

3) For all $k \in A$, $\sum_{x=0}^{\infty} \left(\frac{\partial \, LN\, f}{\partial k} \right)^2 f$ is finite and positive.

Next use Taylor's theorem to express the derivative of the log(f) as a sum of three terms, where the subscript o on ( ) denotes that a term is to be evaluated at the true value k_o of the parameter k. It is assumed that $k_o \in A$.

$$\frac{\partial \, LN\, f}{\partial k} = \left(\frac{\partial \, LN\, f}{\partial k} \right)_o + (k - k_o) \left(\frac{\partial^2 LN\, f}{\partial k^2} \right)_o + \frac{1}{2}\, \theta\, (k - k_o)^2\, H(x),$$

where $|\theta| < 1$. From the above algebraic expression, the likelihood equation (after division by n) may be written in the form

$$\frac{\partial \, LN \, L}{n \, \partial k} = B_o + (k - k_o) B_1 + \frac{1}{2} \theta (k - k_o)^2 B_2 = 0 ,$$

$$\text{where} \quad B_o = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial \, LN \, f_i}{\partial k} \right]_o , \quad B_1 = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial^2 LN \, f_i}{\partial k^2} \right]_o$$

and $B_2 = \frac{1}{n} \sum_{i=1}^{n} H(x_i)$. f_i denotes $f(x_i ; k , \mu)$.

Subtracting the constant C from both sides of the equation gives the large likelihood equation shown below (eq.3.1).

$$\frac{1}{n} \left[\frac{\partial \, LN \, L}{\partial k} - C \right] = \frac{1}{n} \left[\sum_{i=1}^{n} \left[\frac{\partial \, LN \, f_i}{\partial k} \right]_o - C \right] +$$

$$(k - k_o) B_1 + \frac{1}{2} \theta (k - k_o)^2 B_2$$

Next it is important to examine the B_i 's as n approaches infinity. This is an intermediate step towards the goal of showing that the root of eq.3.1 converges in probability to k_o.

As before, f_i denotes $f(x_i ; \mu , k)$. Hence

$$\sum_{x=0}^{\infty} f_i = 1, \text{ and so } \frac{\partial}{\partial k} \sum_{x=0}^{\infty} f_i = 0$$

Using assumptions 1) and 2), the derivative and the sum can be interchanged (see lemma 4) giving

$$\frac{\partial}{\partial k} \sum_{x=0}^{\infty} f_i = \sum_{x=0}^{\infty} \frac{\partial f_i}{\partial k} = 0 \quad . \text{ Similarly}$$

$$\frac{\partial^2}{\partial k^2} \sum_{x=0}^{\infty} f_i = \sum_{x=0}^{\infty} \frac{\partial^2 f_i}{\partial k^2} = 0 \quad . \text{ Then for all } k \in A,$$

$$E\left[\frac{\partial \, LN \, f}{\partial k} \right]_0 = \sum_{x=0}^{\infty} \left[\frac{1}{f_i} \frac{\partial f_i}{\partial k} \right]_0 f_i = 0 \quad \text{and}$$

$$E\left[\frac{\partial^2 \, LNf}{\partial k^2} \right]_0 = \sum_{x=0}^{\infty} \left[\frac{1}{f_i} \frac{\partial^2 f_i}{\partial k^2} - \left[\frac{1}{f_i} \frac{\partial f_i}{\partial k} \right]^2 \right]_0 f_i =$$

$$= - E\left[\frac{\partial \, LN \, f}{\partial k} \right]_0^2 = -\gamma^2 .$$

And by assumption 3), $\gamma > 0$. So B_0 is the mean of n i.i.d. random variables with mean zero. From Khintchine's theorem it follows that B_0 converges in probability to zero. Similarly $B_1 \xrightarrow{P} -\gamma^2$, and $B_2 \xrightarrow{P} E(H(x)) < M$. Letting C be a constant random variable gives $C/n \xrightarrow{P} 0$.

Let δ and ε be arbitrarily small positive numbers. Let P(.) denote the joint probability function of the random variables $x_1, x_2, x_3, \ldots, x_n$. For $n > n_0 = n_0(\delta, \varepsilon)$, the following four inequalities hold:

$P_1 = P(|B_o| \geq \delta^2) < \varepsilon/ 4$,

$P_2 = P(B_1 \geq -\gamma^2/ 2) < \varepsilon/ 4$,

$P_3 = P(|B_2| \geq 2M) < \varepsilon/ 4$,

$P_4 = P(- C/n \leq -\delta^2) < \varepsilon/ 4$.

Further let S be the set of points $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_n)$ such that all four inequalites $|B_o| < \delta^2$, $B_1 < -\gamma^2/ 2$, $|B_2| < 2 M$, and $C/n < \delta^2$ are satisfied. The complement of S, denoted S', consists of points $\mathbf{x}$ such that at least one of the four inequalities is not satisfied. So $P(S') \leq P_1 + P_2 + P_3 + P_4 < \varepsilon$, and thus $P(S) > 1 - \varepsilon$. This implies that the probability that $\mathbf{x} \in S$ is $1 - \varepsilon$, whenever $n > n_o$.

In the concluding step of the proof, it will be shown that $(\partial \text{ LN } L)/ (\partial k) - C$, which is a continuous function of $k \in A$, is both positive and negative for k values in a neighborhood of the true value k_o, and hence $(\partial \text{ LN } L)/ (\partial k) - C$ has a solution within $k_o \pm \delta$ for $n > n_o$. To accomplish this let $k = k_o \pm \delta$. The right hand side of (3.1) becomes $B_o - C/n \pm B_1 \delta + (\theta/ 2)B_2\delta^2$. If $\mathbf{x} \in S$ then $B_o - C/n + (\theta/ 2)B_2\delta^2 < (1 + M)\delta^2$ and $B_1\delta < - \delta\gamma^2/ 2$. If $\delta < \gamma^2/(2(1 + M))$ then

$- B_1\delta > B_o - C/n + (\theta/2)B_2\delta^2$, and so for $k = k_o \pm \delta$ the sign of $(\partial \text{ LN } L)/(\partial k) - C$ will be determined by $\pm B_1\delta$. For $k = k_o - \delta$, $(\partial \text{ LN } L)/(\partial k) - C > 0$, and $(\partial \text{ LN } L)/(\partial k) - C < 0$ for $k = k_o + \delta$. So by continuity of $(\partial \text{ LN } L)/(\partial k) - C$ in k, and for arbitrarily small δ and ε, the large likelihood expression will, with probability exceeding $1-\varepsilon$, have a root between $k_o \pm \delta$, for $n > n_o(\delta,\varepsilon)$. Thus, consistency of the large likelihood estimator is proved. The next section establishes asymptotic normality for the estimator.

Asymptotic Normality for the Estimator

Let k_* be the solution to equation 3.1, the consistency of which was proved in the previous section. If k_* is substituted into 3.1, the equation becomes

$$0 = B_o - C/n + B_1(k_*-k_o) + \frac{\theta}{2} B_2(k_*-k_o)^2 .$$

Algebraic manipulation of the equation yields the expression

$$\gamma\sqrt{n}\ (k_*-k_o) = \frac{\frac{1}{\gamma}\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\frac{\partial \text{ LN } f_i}{\partial k}\right]_o - \frac{1}{\gamma}\frac{1}{\sqrt{n}}C}{(-B_1/\gamma^2) - (\theta/2)B_2(k_*-k_o)/\gamma^2} .$$

The denominator of the right hand side converges in probability to 1. The subtracted expression in the numerator converges in probability to 0. The expression $\left[\frac{\partial \, LN \, f}{\partial \, k} \right]_o$ is a variable with mean zero and variance γ^2, hence by the Central Limit Theorem (Lindeburg-Lévy) the sum $\sum_{i=1}^{n} \left[\frac{\partial \, LN \, f_i}{\partial \, k} \right]_o$ is asymptotically normal with mean zero and variance $\gamma^2 n$. From this it can be concluded that the first term in the numerator is asymptotically normal(0,1). Hence the left hand side of eq. 3.1 is asymptotically normal(0,1), and so k_* is asymptotically normal with mean k_o and variance $1/(\gamma^2 n)$. Thus the second result is shown. The next argument will show that large likelihood estimator is asymptotically efficient.

<u>Asymptotic Efficiency of the Estimator</u>

In the case of unbiased estimators that obey regularity conditions, the asymptotic efficiency is defined as

$$e_o(k_*) = \lim_{n \to \infty} e(k_*) = 1/(w^2 E\{(\partial LN \, f)/ \partial k \}^2)$$

where the variance of k_* is of order w^2/n where w is a constant. However when the estimator is not unbiased, k_* is approximately normal (for large n) with mean k and variance $1/(n \, E\{(\partial \, LN \, f)/ \partial k \}_o^2)$, i.e. w^2/ n .

So the asymptotic efficiency of k_* is given by $e_O(k_*)$, and substitution of $w^2 = 1/(E\{(\partial \text{ LN } f)/\partial k\}_o^2)$ into the equation for $e_o(k_*)$ gives a ratio of 1. Hence k_* is asymptotically efficient. ■

<u>Discussion of Assumptions 1),2), and 3)</u>

In this section the validity of the three assumptions will be examined, as well as a discussion of interchanging the order of differentiation and summation (infinite).

Lemma 1. Begin with the first assumption, which involves the existence of the first, second, and third partial derivatives of LN f (with respect to k).

$$\frac{\partial \text{ LN } f}{\partial k} = \frac{1}{k+x-1} + \ldots + \frac{1}{k} + \text{LN}\left(\frac{k}{\mu+k}\right) + \frac{\mu - x}{\mu + k}$$

Let $k \in A$. When $x = 0$, the sum $\frac{1}{k+x-1} + \ldots + \frac{1}{k}$ is zero. Hence for $x = 0, 1, 2, \ldots$, and $k \in A$, the first partial derivative exists.

Next $$\frac{\partial^2 \text{LN } f}{\partial k^2} = \frac{-1}{(k+x-1)^2} + \ldots + \frac{-1}{k^2} + \frac{1}{k} - \frac{1}{\mu+k} - \frac{\mu - x}{(\mu + k)^2}$$

Again if $x = 0$ then the sum $\frac{-1}{(k+x-1)^2} + \ldots + \frac{-1}{k^2}$ equals zero, and so for $k \in A$ and $x = 0, 1, 2, \ldots$, the second partial derivative exists. Taking the third partial derivative gives

$$\frac{\partial^3 \text{LN } f}{\partial k^3} = \frac{2}{(k+x-1)^3} + \ldots + \frac{2}{k^3} - \frac{1}{k^2} + \frac{1}{(\mu+k)^2} + \frac{2(\mu - x)}{(\mu + k)^3} .$$

Again the third partial exists for $x = 0, 1, 2, \ldots$ and $k \in A$. So the first assumption of the proof holds.

Lemma 2. In this section it must be shown that the first and second partial derivatives of f (with respect to k) are bounded by integrable (summable) functions. Also it will be demonstrated that the third partial derivative of LNf is bounded by a function with finite, positive expected value.

Let $k \in A$, $C_1 = \left(\frac{k}{\mu+k} \right)^k$, and $C_2 = \left(\frac{\mu}{\mu+k} \right)^x$.

$$\frac{\partial f}{\partial k} = \frac{1}{(x!)} C_1 C_2 \Big((k+x-2)\ldots k + (k+x-1)(k+x-3)\ldots k + \ldots$$

$$\ldots + (k+x-1)\ldots(k+1) + \mu/(\mu+k) + LN(k/(\mu+k)) -$$

$$x (k+x-1) \ldots k /(\mu + k) \Big) .$$

$$\therefore \left| \frac{\partial f}{\partial k} \right| \le \frac{C_1 C_2}{(x!)} \Big((k+x-2)\ldots k + (k+x-1)(k+x-3)\ldots k +$$

$$\ldots + (k+x-1)\ldots(k+1) + \mu/(\mu+k) + LN((\mu+k)/k) +$$

$$x (k+x-1) \ldots k /(\mu + k) \Big) .$$

$$\le \frac{1}{x\,!} C_1 C_2 \Big(x (k+x-1)\ldots(k + 1) k/k +$$

$x\ (k+x-1)\ \ldots\ k\ /(\mu + k) + LN((\mu+k)/k)C_3\ (k+x-1)\ldots\ldots k$

$+\ \mu/(\mu+k)\ (k+x-1)\ldots k\ C_3\ )$

$= \frac{x\ f}{k} + \frac{x\ f}{\mu+k} + C_3\ LN((\mu+k)/k)\ f + C_3\ \mu\ /\ (\mu+k)\ f =$

$= G_1(x)$. Where C_3 is a real number such that

$(k+x-1)\ \ldots\ k\ C_3 \geq 1$. Hence $\sum_{x=0}^{\infty} G_1(x) =$

$= \mu/k + \mu\ /\ (\mu + k) + C_3\ LN((\mu+k)/\mu) + C_3\mu\ /(\mu+k) < \infty,$

showing that the absolute value of the partial derivative is bounded by a summable function G_1 for $k \in A$.

Next, it will be shown that the absolute value of the second partial derivative is bounded by a function $G_2(x)$, where $G_2(x)$ is summable over $x = 0, 1, 2, \ldots$.
Let $k \in A$; and C_1, and C_2 be defined as before.

$$\left| \frac{\partial^2 f}{\partial k^2} \right| = \frac{1}{x\,!} \left| C_1\ (LN(k/(\mu+k)) + \mu/(\mu+k))\ C_2 \cdot \right.$$

$\{(k+x-2)\ldots k + (k+x-1)(k+x-3)\ldots k + \ldots + (k+x-1)\ldots(k+1) +$

$LN(k/(k+\mu)) + \mu/(\mu+k) - x\ (k+x-1)..k/(\mu+k)\ \} +$

$C_1\ C_2\ (-\ x/(\mu+k)$ { as before } +

$C_1C_2 \left[(k+x-3)\ldots k + (k+x-2)(k+x-4)\ldots k + \ldots + (k+x-2)..\right.$

$(k+x-3)\ldots k+(k+x-1)(k+x-4)\ldots k + \ldots + (k+x-1)(k+x-3)\ldots$

$+ \ldots +$

$(k+x-2)\ldots(k+1)+\ldots+ (k+x-1)\ldots(k+2) \,] + 1/k - 1/(\mu+k)$

$- \mu/(\mu+k)^2 - x\,(- (k+x-1)\ldots k\,/(\mu+k)^2 +$

$((k+x-2)\ldots k + (k+x-1)(k+x-3)\ldots k + \ldots + (k+x-1)\ldots(k+1)\,)$

$\div (\mu+k)\ \ ) \ \Big|$

$\leq \dfrac{C_1C_2}{(x!)}\,(\,(\,LN((\mu+k)/\mu) + \mu/(\mu+k))\{\, x(k+x-1)\ldots(k+1)k/k$

$+(LN((\mu+k)/k)+\mu/(\mu+k))(k+x-1)\ldots k\ C_3+x(k+x-1)\ldots k/(\mu+k)\ \}$

$+(x/(\mu+k))\{\text{as before}\}+x(x-1)(k+x-1)\ldots(k+2)(k+1)k/(k^2+k)$

$+ (\ 1/k + 1/(\mu+k) + \mu/(\mu+k)^2\ )\ (k+x-1)\ldots k\ C_3 +$

$x\ (k+x-1)\ldots k/(\mu+k)^2 + x^2/(\mu+k)\ \ (k+x-1)\ldots(k+1)k/k\ )$

$= x/k\ (\ LN((\mu+k)/k) + \mu/(\mu+k)\ )f + C_3(\ LN((\mu+k)/k)$

$+ \mu/(\mu+k)\ )\ f + x\ f\ /\ (\mu+k) + x^2\ f\ /(\mu k + k^2) +$

$+ C_3\ x/(\mu+k)\ (\ LN((\mu+k)/k) + \mu/(\mu+k)\ )\ f +$

$x^2 f\ /(\mu+k)^2 + (x^2-x)\ f\ /(k^2+k) + C_3\ f\ (\ 1/k + 1/(\mu+k)+$

$\mu/(\mu+k)^2\ ) + x\ f\ /(\mu+k)^2 + x^2\ f\ /\ (\mu k+k^2) = G_2(x)\ .$

And so $\sum_{x=0}^{\infty} G_2(x) = \mu/k \; (LN((\mu+k)/k) + \mu/(\mu+k))$

$+ C_3 \; (LN((\mu+k)/k) + \mu/(\mu+k)) + \mu/(\mu+k) + (\sigma^2 + \mu^2)\div$

$(\mu k + k^2) + C_3 \; \mu/(\mu+k) \; (LN((\mu+k)/k) + \mu/(\mu+k))$

$+ (\sigma^2 + \mu^2)/(\mu + k)^2 + (\sigma^2 + \mu^2 - \mu)/(k^2+k) \; +$

$C_3 \; (1/k + 1/(\mu+k) + \mu/(\mu+k)^2) + \mu/(\mu+k)^2 \; +$

$(\sigma^2 + \mu^2)/(\mu k + k^2) \; < \infty$

Therefore for $k \in A$, and $\mu > 0$, the absolute value of the second partial derivative of f is bounded by a summable function, $G_2(x)$.

And finally for assumption 2) it will be shown that the absolute value of the third partial derivative of f is bounded by a function H(x), whose expected value is less than a finite number M (with M independent of k). Let $k \in A$.

$$\left| \frac{\partial^3 LNf}{\partial k^3} \right| = \left| \frac{2}{(k+x-1)^3} + \ldots + \frac{2}{k^3} - \frac{1}{k^2} + \frac{1}{(\mu+k)^2} + \frac{2(\mu-x)}{(\mu + k)^3} \right|$$

$$\leq \frac{2x}{k^3} + \frac{1}{k^2} + \frac{1}{(\mu+k)^2} + \frac{2\mu + 2x}{(\mu + k)^3} = H(x)$$

The expected value is $\sum_{x=0}^{\infty} H(x)\ f(x\ ;\ k) =$

$$\frac{2\mu}{k^3} + \frac{1}{k^2} + \frac{1}{(\mu+k)^2} + \frac{2\mu}{(\mu+k)^3} + \frac{2\mu}{(\mu+k)^3} \leq M < \infty .$$

Recall that $A = (a,b)$ where $b > a > 0$. If $a \geq 1$ then $M = 6\mu + 2$. If $0 < a < 1$, then M is equal to the following

$$\frac{2\mu}{a^3} + \frac{1}{a^2} + \frac{1}{(\mu+a)^2} + \frac{4\mu}{(\mu+a)^3},$$ which is finite and independent of k. Hence all of the conditions of assumption 2) have been verified.

Lemma 3. For the final assumption 3) it is necessary to show that for $k \in A$, the expected value of the square of the first partial of f is finite and positive. Let the sum $1/(k+x-1) + 1/(k+x-2) + \ldots + 1/k$ be denoted by ψ for ease of notation. The square of $(\partial \text{LN } f / \partial k)$ equals

$\psi^2 + 2\psi \text{ LN}(k/(\mu+k)) + (\text{LN}(k/(\mu+k)))^2 +$

$((\mu-x)/(\mu+k))^2 + 2\psi(\mu-x)/(\mu+k) + 2 \text{ LN}(k/(\mu+k))(\mu-x)/(\mu+k)$.

The above expression is bounded from above by

$(x/k)^2 + 2x/k \text{ LN}((\mu+k)/k) + (\text{LN}(k/(\mu+k)))^2 +$

$((\mu-x)/(\mu+k))^2 + 2x/k\ (\mu+x)/(\mu+k) + 2 \text{ LN}((\mu+k)/k)(\mu+x)/(\mu+k)$.

Hence, the expected value is bounded from above as the next inequality demonstrates. Let $k \in A$.

$$\sum_{x=0}^{\infty} \left(\frac{\partial \text{ LN } f}{\partial k} \right)^2 f(x\ ;\ k) \leq (\sigma^2+\mu^2) / k^2 + 2\mu/k \text{ LN}((\mu+k)/k) +$$

$(LN(k/(\mu+k)))^2 + \sigma^2/(\mu+k) + (\sigma^2 + 3 \mu^2)/((\mu+k)k) +$ $4 \mu/ (\mu+k) LN((\mu+k)/k) < \infty$. So the expected value is finite. Next it will be shown that the expected value is positive. Consider $\sum_{x=0}^{\infty} \left[\frac{\partial LN f}{\partial k} \right]^2 f(x ; k)$. The square term is non-negative and f is positive. If the expected value equals zero, then this means that $(\partial LN f)/ \partial k$ is equal to zero for all x = 0, 1, 2, . . . and $k \in A$. This conclusion about the derivative implies that the log likelihood function is constant for x = 0, 1, 2, ... and $k \in A$, which it is not. Therefore the expected value is positive as required. The final lemma addresses the question of exchanging the order of the derivative and the infinite sum.

Lemma 4. This lemma is from Folland (1984). Let $f : X \times [a, b] \longrightarrow \mathbb{R}$, and $\sum_{x=0}^{\infty} f(x; k_1) < \infty$ for fixed $k_1 \in [a, b]$, where $- \infty < a < b < \infty$.

Let $F(k) = \sum_{x=0}^{\infty} f(x ; k)$ for $k \in [a, b]$. Suppose that $\partial f/\partial k$ exists and there is a g(x) such that

$\sum_{x=0}^{\infty} g(x) < \infty$, and $|\partial f(x;k)/\partial k| \leq g(x)$ for all $k \in [a,b]$ and x = 0, 1, 2, ... , then F(k) is differentiable and, $\sum_{x=0}^{\infty} \frac{\partial f}{\partial k} (x,k) = \frac{\partial}{\partial k} \sum_{x=0}^{\infty} f(x,k)$.

Proof: Let k_n be any sequence converging to k_o. Therefore

$$\partial f(x,k_o)/\partial k = \lim_{k_n \to k_o} h_n(x) \quad \text{where}$$

$$h_n(x) = (f(x, k_n) - f(x,k_o)) / (k_n - k_o) .$$

By the Mean Value Theorem

$$| h_n(x) | \leq \sup_{k \in [a,b]} | \partial f(x,k)/\partial k | \leq g(x).$$

Applying the Dominated Convergence Theorem gives

$$\frac{\partial}{\partial k} \sum_{x=o}^{\infty} f(x, k_o) = \lim_{k_n \to k_o} \left[\frac{\sum_{x=o}^{\infty} f(x,k_n) - \sum_{x=o}^{\infty} f(x,k_o)}{k_n - k_o} \right]$$

$$= \lim_{k_n \to k_o} \sum_{x=o}^{\infty} \left[\frac{f(x,k_n) - f(x,k_o)}{k_n - k_o} \right] = \sum_{x=o}^{\infty} \frac{\partial f(x,k)}{\partial k}.$$

■

This concludes the proof of the asymptotic properties of the large likelihood estimator. The next chapter will be devoted to a refinement of the LLE, namely the adjusted large likelihood estimator. This new method sets the derivative equal to a small positive C, but the value of C changes in response to the concavity of the log likelihood function.

CHAPTER VI

CHOICE OF OPTIMAL C AND ADJUSTED LARGE LIKELIHOOD ESTIMATOR

Optimal Choice of C

The choice of C = 0.13 for the LLE provides an overall improvement for estimation in terms of bias and variability, as compared with the method of moments and maximum likelihood approaches. However, fixing C causes the average LLE to over-estimate or under-estimate the true value of k by (in some cases) a considerable degree--still the LLE does provide the researcher with an improved method of estimation. The question arises " what value of C will produce an average LLE that is essentially unbiased? ". In an effort to reach an answer, many simulations were run in order to find the optimal (minimizes estimated bias) C for each parameter combination. These C values are summarized in table 19.

It should be noted that the resulting bias is quite robust to the choice of C when k is small in relation to μ (k = 1 and μ = 1 ,3 , or 5). Specifically, C can be varied from 0.5 to 0.8 with no great change in the average LLE. This is intuitively reasonable because the concavity of the log likelihood function (at its maximum) is in general much more extreme when k = 1. With extreme concavities (the log likelihood has a sharp point at its maximum), C can be varied greatly without changing the average LLE of the simulation (see figure 9).

TABLE 19

OPTIMAL VALUES FOR C

μ		1	3	5
	5	0.05	0.13	0.18
k	3	0.13	0.25	0.31
	1	0.69	0.89	0.97

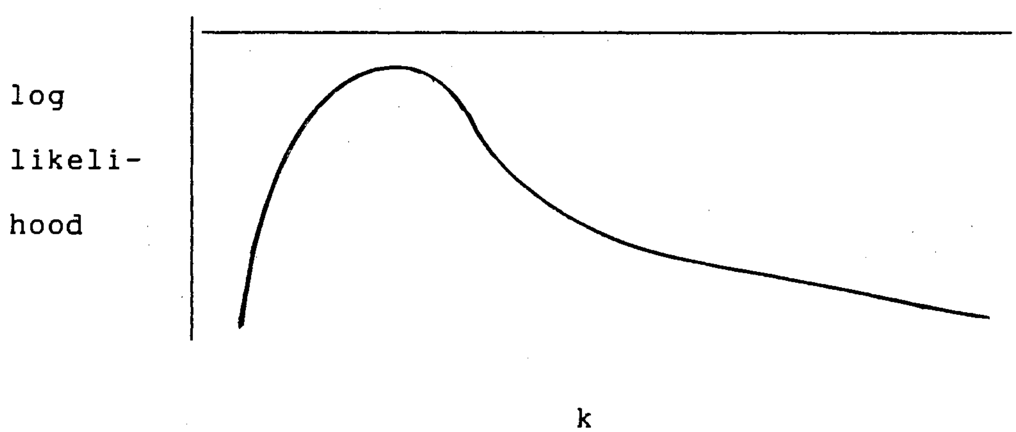

Figure 9. Log Likelihood Function With a Sharp Point

On the other hand, the choice of C when k is large in relation to μ (k = 5) is crucial. Again an appeal to the concavity will explain the reason. The log likelihood function arising from a population with k = 5 and μ = 1 (for example) exhibits moderate to mild concavity at its maximum (if the maximum exists). Hence a small change in C produces large changes in the average LLE.

Adjusted Large Likelihood Estimator

A conclusion that can be drawn from the above discussion is that for a fixed value of μ , the concavities of the log likelihood functions (at the maximum) follow a marked trend as k is increased from 1 to 3 and again to 5. To demonstrate this, nine simulations of 1500 samples of size 50 were run, and the median second derivative (at the maximum) was found. The differences between the median second derivatives can be quite dramatic as k is increased from 1 to 3 and then to 5 (see table 20). This being the case, it is natural to think that the concavity of a log likelihood function (at the maximum) could give information to the researcher in regard to the choice of C. Such an idea was implemented in the adjusted large likelihood estimator.

TABLE 20

MEDIAN SECOND PARTIAL DERIVATIVE
FROM 1500 SAMPLES OF SIZE 50

μ		1	3	5
	5	0.047	0.102	0.204
k	3	0.132	0.491	0.898
	1	3.492	10.413	14.158

The adjustment to the LLE comes in the form of changing the constant C in response to the concavity of the log likelihood function. If the sample is under-dispersed, the value of C is set to 0.09, because there is no maximum to evaluate the second derivative. The algorithm for the Adjusted LLE (adjLLE) uses the sample mean and the absolute value of the second partial derivative of the log likelihood function (denoted 'sec') to find an adjusted C. The goal of the process is to tailor the constant so the the average adjLLE will be closer to the true value of k than the average LLE, MLE, or MME. A function $C = c(\bar{x}, \text{sec})$ will be developed that chooses a C in response to the sample mean and concavity. The derivation of this function follows.

Using the data from table 19, a non-linear model is fit with C as the response variable and μ and k as explanatory variables. The choice of the model

$C = c_o + (c_1 \mu^{c2}) / (k^{c3})$ was suggested by the data.

In fact during the course of the research, many other models were fitted, but the end results are not as promising as the final model described above. The SAS procedure NLIN yielded (4.1)

$$C = -0.0589 + (0.7049 \mu^{0.2791}) / (k^{1.0277}),$$

with absolute residuals being bounded from above by 0.0438. The same process was carried out for the data in table 20, giving equation (4.2)

$$|\text{sec}| = -0.1715 + 4.6066 (\mu^{0.6847}) / (k^{2.3556})$$

with absolute residuals bounded from above by 0.94.

It is helpful to visualize the two grids of data stacked vertically. When the sample is encountered and k is to be estimated, the true value of k in the parent population is the same regardless of which grid that is examined. Hence equations (4.1) and (4.2) are each solved for k, giving $k = ((c_1 \mu^{c2}) / (C - c_0))^{1/c3}$ and $k = ((b_1 \mu^{b2}) / (-sec - b_0))^{1/b3}$. The right hand side of both expressions are equated, and then the equality is solved for C giving

$$C = c_0 + (c_1 \mu^{c2}) / ((b_1 \mu^{b2}) / (-sec - b_0))^{c3/b3} .$$

Substitution of $\bar{x}$ for μ gives a value for C that depends on the sample mean and concavity. This C is then used in the usual LLE algorithm instead of the fixed 0.13.

Table Summary of the Simulation Results

Nine simulations, consisting of 3000 samples of size 50, were run using the same parameter combinations as in previous chapters. The program generates a sample and the MME and MLE are calculated if possible. If there is a valid maximum for the log likelihood function then this information is used in the form of 'sec' to find an adjusted C in the adjusted LLE. The LLE is found for all samples whether they be over or under-dispersed. So for each sample four estimates are found-- namely the MME, MLE, LLE, and adjLLE. If the sample is under-dispersed the MME and MLE will be missing values; the constant for the LLE will be 0.13; and the C for the adjLLE will be 0.09. 0.09 was chosen because it allows moderate improvement in the bias for the μ = 1 and k = 5 combination without increasing the bias too much in the μ = 1 and k = 3 case. To elaborate, if the C for the adjLLE is set to 0.05 (for under dispersed samples) then the bias for the μ = 1 and k = 5 will be greatly improved, but this will be at the expense of the μ = 1 and k = 3 case. Using 0.05 in the latter parameter combination produces an average adjLLE that is 3.8 to 4.0. And so an improvement for one parameter combination causes a deterioration in the results for an adjacent parameter combination. The results for the nine simulations are summarized in tables 21 through 24.

TABLE 21

AVERAGE $\hat{k}$ FOR 3000 SAMPLES OF SIZE 50

μ	k	MME	MLE	LLE	adjLLE	% UNDER-DISP
1	1	1.575	1.660	1.278	1.210	1.30
1	3	4.797	6.967	2.988	3.111	13.97
1	5	6.136	9.028	3.715	4.006	27.07
3	1	1.165	1.114	1.095	1.024	0.00
3	3	4.127	4.507	3.349	3.267	0.33
3	5	7.720	10.155	4.908	4.981	3.13
5	1	1.137	1.089	1.077	1.016	0.00
5	3	3.518	3.513	3.248	3.121	0.03
5	5	6.589	6.877	5.223	5.191	0.30

TABLE 22

$S_{\hat{k}}$ FOR 3000 SAMPLES OF SIZE 50

μ	k	MME	MLE	LLE	adjLLE	% UNDER-DISP
1	1	1.826	3.983	0.787	0.856	1.30
1	3	5.472	20.376	1.521	1.868	13.97
1	5	6.099	25.693	1.649	2.096	27.07
3	1	0.439	0.384	0.363	0.431	0.00
3	3	4.126	15.316	1.430	1.552	0.33
3	5	9.634	44.254	2.028	2.353	3.13
5	1	0.373	0.309	0.300	0.282	0.00
5	3	1.593	1.561	1.148	1.184	0.03
5	5	5.451	7.031	1.936	2.135	0.30

TABLE 23

MSE FOR 3000 SAMPLES OF SIZE 50

μ	k	MME	MLE	LLE	adjLLE	% UNDER-DISP
1	1	3.664	16.291	0.696	0.776	1.30
1	3	33.158	430.763	2.313	3.502	13.97
1	5	38.477	676.050	4.370	5.377	27.07
3	1	0.220	0.163	0.141	0.117	0.00
3	3	18.292	236.776	2.166	2.478	0.33
3	5	100.177	1984.29	4.119	5.535	3.13
5	1	0.158	0.103	0.096	0.080	0.00
5	3	2.806	2.698	1.380	1.415	0.03
5	5	32.225	52.945	3.798	4.593	0.30

TABLE 24

75^{th} PERCENTILE FOR $\hat{k}$ FROM 3000 SAMPLES OF SIZE 50

μ	k	MME	MLE	LLE	adjLLE	% UNDER-DISP
1	1	1.680	1.619	1.522	1.402	1.30
1	3	5.312	5.741	3.882	4.056	13.97
1	5	7.087	7.852	4.824	5.438	27.07
3	1	1.387	1.279	1.258	1.174	0.00
3	3	4.551	4.603	3.999	3.899	0.33
3	5	7.812	8.253	6.052	6.208	3.13
5	1	1.332	1.247	1.234	1.162	0.00
5	3	4.069	4.051	3.778	3.624	0.03
5	5	7.291	7.450	6.146	6.136	0.30

Interpretation of the Simulation Results

For 8 of the 9 parameter combinations the average adjLLE is closer to the true k than the average LLE, MLE, or MME. In the exceptional case of $\mu = 1$ and $k = 3$, the average LLE is $\cong 2.93$ and the average adjLLE is $\cong 3.11$. The overall improvement in the bias that is provided by the adjLLE is paid for at the price of a slightly larger variability (in some cases). Only for parameter combinations $\mu = 3$ and $k = 1$, and $\mu = 5$ and $k = 1$ is estimated standard error for the $\hat{k}$'s smaller in the adjLLE as compared with the LLE. Still it should be remembered that the variation of the LLE based methods are very similar and in most cases much smaller than the variation of the MME and MLE approaches. So in general the adjLLE provides a slightly smaller bias and a slightly larger variability as compared with the LLE, but the two new methods give much better results than the MLE and MME.

The adjLLE somewhat improves the problem encountered with the LLE for $\mu = 1$ and $k = 5$. Recall for this parameter combination that the average LLE under-estimated the true value of k by about 1.3, whereas the average MME over-estimated the true k by about 1.14. Using the adjLLE decreases the under-estimation to about 1 (the average adjLLE is 4.008). Hence the absolute bias for the adjLLE is slightly smaller than the bias seen with the LLE and MME approaches (and certainly the MLE).

The upper quartile (Q_3) for the adjLLE is in 6 of the 9 cases smaller than the quartile for the LLE (and hence much smaller than the quartiles for the two conventional methods). In the case of $\mu = 1$ and $k = 5$, Q_3 for the adjLLE is bigger than the quartile for the LLE, but this is desirable since the LLE under-estimates k for this parameter combination. Also for $\mu = 1$ and $k = 3$, and $\mu = 3$ and $k = 5$ the upper quartiles for the LLE are slightly smaller than those for the adjLLE.

In general the results for the LLE and adjLLE are very similar, with a slight improvement in the bias offered by the adjLLE. Both LLE based methods show a great improvement over the MME and MLE approaches in terms of bias, variability, and upper quartiles.

Problems With the Adjusted Large Likelihood Estimator

The adjLLE is based on simulation results for samples of size 50. It is a matter for further research to determine if the grids (of values) and the equations for sample size 50 can be used with reasonable success for sample sizes that are close to 50. This idea is offered since the fixed C = 0.13 worked well for n = 50 and for n = 250.

The original goal of the adjLLE was to reduce the bias encountered in the LLE approach. In almost all cases the adjLLE did reduce the bias, but the results were somewhat less dramatic than hoped for. This marginal improvement could be due to the fact that the non-linear equation that was used to fit the median second derivatives did not model the grid of values well for $\mu = 1$. The residuals were quite large with $| \text{resid}_i | < 0.9$. Hence there was a systematic error in the equations that related second derivative and choice of C. Another explanation that can be offered is that the grid of optimal C's was found by fixing C, running a simulation, and finding the average LLE. C was adjusted until the process yielded an average LLE that was within 0.05 of the true value for k. But in the adjLLE algorithm, the C's varied due to variation in sample mean and sample concavity (the usual sample variation). It was hoped that these variable C's would (over the course of the simulation) produce an average LLE that was very close to the true value of the parameter. This success was enjoyed to a certain degree.

The next chapter examines a preliminary solution to the problem of under-dispersion. The technique of re-weighting a sample produces favorable results in most negative binomial samples, but some samples are not responsive to the approach. The details are discussed in chapter V.

CHAPTER V

REWEIGHTING: A PRELIMINARY APPROACH TO SOLVE THE PROBLEM OF UNDER-DISPERSION

One of the main goals of this research was to find a method of estimating k that worked for all negative binomial samples. Attention was focused on enlarging the sample variance by increasing the frequency (or weight) of the smallest and largest observations in the sample. If an under-dispersed sample is encountered, extra weight is given to the observation(s) that produce(s) the largest increase in the $S_1^2 / \bar{X}$ ratio, where $S_1^2 = S^2 (n-1)/n$. The phrase "extra weight" refers to increasing the frequency (by one) of the largest and smallest observations in the sample. This re-weighting increases the above ratio, and repeated application of the technique will usually render the sample over-dispersed, *i.e.* $s_1^2 / \bar{x} > 1$. This inequality can be attained in most cases, but not in all possible samples. There are some samples that re-weighting cannot over-disperse.

Figure 4 on page 10 illustrates the problem. For For this graph S_1^2 is plotted versus $\bar{X}$ for 200 samples of size 35. The samples are from a negative binomial distribution with $\mu = 1$ and $k = 3$. Each dot represents a sample, with the ordered pair being ($\bar{x}$, s_1^2). The plus symbols form the line $S_1^2 = \bar{X}$. Dots falling below this line represent under-dispersed samples, and hence a finite positive $\hat{k}$ is not available. The goal of re-weighting is to increase the variance-to-mean ratio of the sample, thus moving the dot above the line so that reasonable k estimates can be found by MLE and MME approaches.

It should be noted that dots falling just above the line represent samples that exhibit very volatile estimates for k. Finite positive $\hat{k}$'s are possible for the MLE and MME approaches, but these estimates can often be very large, *i.e.* $\hat{k}$ for mle = 1256.67 when the actual value is 5. This volatility can be explained in the MME by examining the denominator of the estimator, which is $S^2 - \bar{X}$. When the sample mean and variance are very close together, then the denominator will be nearly zero, forcing the fraction to increase dramatically.

The following example illustrates the re-weighting procedure. A sample of size 35 was generated using $\mu = 3$ and $k = 3$. The frequencies are listed in the following table.

TABLE 25

AN EXAMPLE TO ILLUSTRATE RE-WEIGHTING

observation	0	1	2	3	4	5	6	7	8
frequency	4	4	11	6	4	3	1	1	1

$\bar{x} = 2.7428$ $s_1^2 = 3.6195$ MME = 7.6518

Increase the frequency of observation 8 to 2.

new $\bar{x} = 2.8888$ $s_1^2 = 4.2654$ MME = 5.5696

Enlarging the frequency of 8 from 1 to 2 increases the variance-to-mean ratio and decreases the MME estimate for k from 7.6518 to 5.5696. This re-weighting will also decrease the MLE for k.

If an under-dispersed sample is obtained the re-weighting process will be repeated until $s_1^2 > \bar{x}$.

In general the algorithm can be outlined as follows.

Calculate s_1^2 and $\bar{X}$ for the sample.

Do while $s_1^2 <= \bar{X}$

re-weight (obtain the observation that yields the maximum variance-to-mean ratio and increase its frequency)

Calculate s_1^2 and $\bar{X}$.

End of do loop.

Re-weight once more.

Calculate MM and MLE for k.

Simulation Results for Re-weighting

Simulation results for various parameter choices are given in tables 26 through 32. Each table summarizes the results from 2400 samples of size 35. Note that the values in the column denoted Conventional are obtained without re-weighting. For example, when $\mu = 1$ and $k = 1$, 3.1 % of the 2400 samples were discarded because $S_1^2 < \bar{X}$. So the aveMLE 1.987 is based on 2325 samples. On the other hand, the values in the column denoted by Re-weight are based on the full 2400 samples (there were 0 % under-dispersed samples after re-weighting). Note that this method also dramatically reduces the estimated MSE. Max $\hat{k}$ refers to the largest $\hat{k}$ calculated in the 2400 samples. This value is also reduced by re-weighting.

TABLE 26

COMPARISON OF RE-WEIGHTING TO THE
MLE AND MME, μ = 1 and k = 1

	MLE	RE-WEIGHT
% UNDER-DISPERSED	3.1	0
AVE $\hat{k}$	1.987	1.398
MSE	28.066	1.308
MAX $\hat{k}$	120.58	8.47
AVE N	35	36.26
	MME	**RE-WEIGHT**
AVE $\hat{k}$	1.744	1.445
MSE est	5.311	1.656
MAX $\hat{k}$	31.08	7.87

TABLE 27

COMPARISON OF RE-WEIGHTING TO THE
MLE AND MME, μ = 1 and k = 3

	MLE	RE-WEIGHT
% UNDER-DISPERSED	20.25	0
AVE $\hat{k}$	7.422	3.050
MSE	451.238	3.385
MAX $\hat{k}$	475.02	9.41
AVE N	35	38.05
	MME	**RE-WEIGHT**
AVE $\hat{k}$	4.515	3.218
MSE	27.441	3.765
MAX $\hat{k}$	41.44	8.97

TABLE 28

COMPARISON OF RE-WEIGHTING TO THE
MLE AND MME, μ = 1 and k = 5

	MLE	RE-WEIGHT
% UNDER-DISPERSED	33.1	0
AVE $\hat{k}$	9.977	3.668
MSE	618.52	5.166
MAX $\hat{k}$	447.15	10.331
AVE N	35	39.21
	MME	**RE-WEIGHT**
AVE $\hat{k}$	5.686	3.842
MSE	32.846	5.161
MAX $\hat{k}$	41.45	8.97

TABLE 29

COMPARISON OF RE-WEIGHTING TO THE
MLE AND MME, μ = 3 and k = 1

	MLE	RE-WEIGHT
% UNDER-DISPERSED	0	0
AVE $\hat{k}$	1.218	1.083
MSE	1.081	0.295
MAX $\hat{k}$	34.09	12.25
AVE N	35	36.0
	MME	**RE-WEIGHT**
AVE $\hat{k}$	1.281	1.164
MSE	0.769	0.375
MAX $\hat{k}$	20.10	12.08

TABLE 30

COMPARISON OF RE-WEIGHTING TO THE MLE AND MME, μ = 3 and k = 3

	MLE	RE-WEIGHT
% UNDER-DISPERSED	0.8	0
AVE $\hat{k}$	5.145	3.457
MSE	110.782	5.091
MAX $\hat{k}$	338.61	19.48
AVE N	35	36.03
	MME	RE-WEIGHT
AVE $\hat{k}$	4.575	3.437
MSE	27.502	5.220
MAX $\hat{k}$	63.73	17.25

TABLE 31

COMPARISON OF RE-WEIGHTING TO THE MLE AND MME, μ = 3 and k = 5

	MLE	RE-WEIGHT
% UNDER-DISPERSED	6.8	0
AVE $\hat{k}$	14.212	5.599
MSE	5959.9	11.126
MAX $\hat{k}$	2305.6	25.18
AVE N	35	36.28
	MME	RE-WEIGHT
AVE $\hat{k}$	8.314	5.459
MSE	107.459	10.739
MAX $\hat{k}$	107.05	20.97

TABLE 32

COMPARISON OF RE-WEIGHTING TO THE MLE AND MME, μ = 5 and k = 5

	MLE	RE-WEIGHT
% UNDER-DISPERSED	1.4	0
AVE $\hat{k}$	10.051	5.695
MSE	1744.41	15.715
MAX $\hat{k}$	1319.98	37.21
AVE N	35	36.05
	MME	RE-WEIGHT
AVE $\hat{k}$	7.649	5.517
MSE	106.14	13.36
MAX $\hat{k}$	150.87	31.73

Conclusions Regarding the Re-weighting Technique

Re-weighting provided successful results in simulation studies of sample size 35. The technique decreases the percent of under-dispersed samples, produces a smaller estimated bias as compared with the conventional MLE and MME, and produces a smaller estimated MSE as compared with the conventional methods. However, several problems arose with the new technique. The first being that re-weighting could not fix all samples. There are certain samples that cannot be over-dispersed by increasing the frequency of extreme observations. Another problem involved knowing when to stop the re-weighting process. Trial and error demonstrated that re-weighting bad samples until the aforementioned ratio exceeds one, and then re-weighting once more, produced fairly good simulation results (good samples were also re-weighted one time). But some parameter combinations were more closely estimated when samples were re-weighted twice. Hence if a researcher used the technique, it is not obvious when to stop the process. These pitfalls in the re-weighting process led this author to look for another method of estimation which truly gives a reasonable $\hat{k}$ for all NB samples--this method being the large likelihood estimator which was discussed at length in previous chapters.

CHAPTER VI

DISCUSSION AND SUMMARY

Success of the Three Estimation Methods

The goal of this research was to develope methods to overcome the problems of under-dispersion and volatility which are encountered in MME and MLE approaches. All three of the techniques (re-weighting, LLE, and adjLLE) accomplished these goals with varying degrees of success.

The technique of re-weighting reduced the fraction of samples that were under-dispersed, but it did not eliminate the problem entirely. As a result the method made partial gains against under-dispersion. On the positive side, re-weighting does produce estimates that display smaller variability and smaller absolute bias as compared with the convention MME and MLE methods.

An intuitive drawback to the method of re-weighting involves the excessive emphasis (or weight) given to the largest and smallest observations in the sample. If the frequency of each extreme observation is only boosted by one, and the original sample size is 50, then re-weighting probability causes very little change in the nature of the original sample. However, if it is necessary to re-weight each extreme value 30 times, then the heavy-ended new sample is far removed from its original state. With excessive re-weighting some would argue that the data tampering produces a new sample which bears little resemblance to the original. Still the overall results for the re-weighting technique are favorable.

The large likelihood estimator goes much further in overcoming the obstacles of under-dispersion and volatility. Since a finite positive LLE estimate is attainable for all NB samples, the problem of under-dispersion is eliminated. All NB samples give rise to a reasonable LLE estimate of k, regardless of the relative magnitude of the sample mean and variance. An additional bonus for the LLE is its smaller variability. Simulations demonstrate the reductions (sometimes dramatic) in $S_{\hat{k}}$ and $\widehat{MSE}$ for the LLE as compared with the MLE and MME approaches. Also the extremely large estimates that were seen with the two traditional methods simply are not present with the LLE. It is surprising that 'almost maximizing' a function gives better results than actually setting the derivative equal to zero and solving for k.

Fixing C = 0.13 does render the LLE somewhat unresponsive when estimating samples from populations having large k (in relation to $\mu = 1$). Simulations show that the LLE tends to under-estimate the true value for k more than the average MME over-estimates the true k. The adjusted LLE was developed in order to overcome this problem. It was thought that the concavity of the log likelihood function could serve as a guide in choosing the small positive C-- in other words, adjusting C in response to the concavity could produce estimates whose average was closer to the true value than the LLE, MLE, or MME. Simulation results demonstrate that this goal is for the most part realized, but the improvement is not as dramatic as was hoped. Still when comparing the results of the adjLLE (and also the LLE) to the results of the MLE or MME, the results for the new methods are very encouraging. It is possible that additional work on the adjLLE will improve its performance even more in regard to reducing the bias. Perhaps better models can be found that fit the nine grid points exactly (cubic splines were used with disappointing results). In the next section, ideas for further research are explored.

Areas for Further Research

In order to discuss applications of the LLE to other problems it helps to examine why the method works for the NB distribution. When viewing the log likelihood function of a NB sample of size n, the maximum in terms of μ (for a fixed k) is the sample mean. This is true for any positive k (see figure 10).

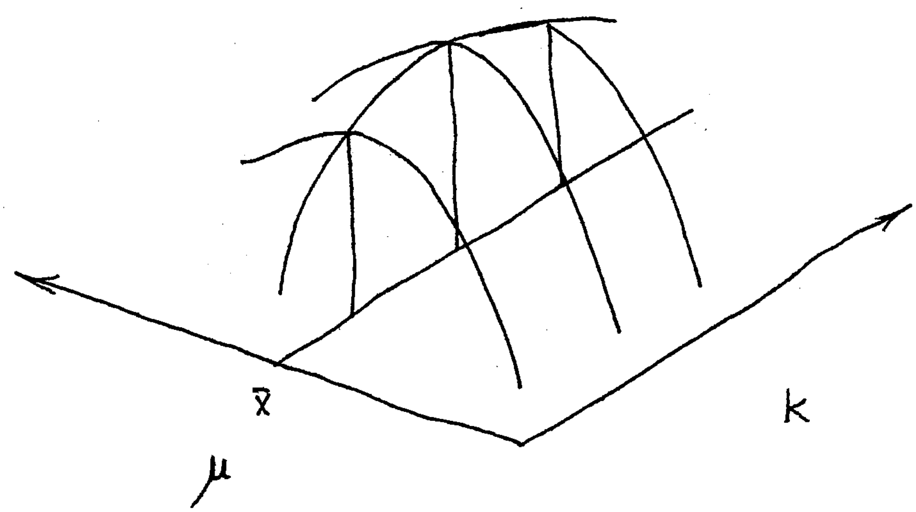

Figure 10. Maximum of Log Likelihood Surface in terms of μ

So estimation of k is reduced to a one variable problem. This is evident from the partial derivative

$\partial L / \partial \mu = -nk / (\mu + k) + \sum x_i \, (1/\mu - 1/(\mu + k))$.

Setting the derivative equal to zero and solving for $\hat{\mu}$ yields $\bar{x}$, which is independent of k. As a result, the estimation of k is essentially a problem of finding the root (in k) of $\partial L / \partial k = 0$. Recall figures 7 and 8.

Another useful characteristic possessed by the NB is the unimodality of its likelihood function (Levin and Reeds). Hence the problem is well suited to solution by Newton's method--the iterative process is started with a preliminary $\hat{k}$ near zero, and the estimate is increased until if falls within the pre-determined tolerance. Examination of a log likelihood function and its corresponding derivative function illustrates the usefulness of unimodality (see figures 11 and 12). If on the other hand the log likelihood function has more than one local maximum (see figure 13), then the problems of distinguishing the global max from the local maximums come into play. The LLE could still be used, but the algorithm would have to be modified so that the global max is found first, and then the large likelihood process would begin.

So here is a summary of the characteristics of the NB log likelihood that make it suitable for large likelihood estimation.

1. It is a univariate estimation problem.
2. The log likelihood function for the NB is unimodal.
3. The partial derivative of the log likelihood function does not yield a closed form solution for the parameter of interest.

This list is not meant to preclude the use of the LLE for problems that violate one or more of the above conditions. It would just be easier to apply the technique in its present state if these conditions hold.

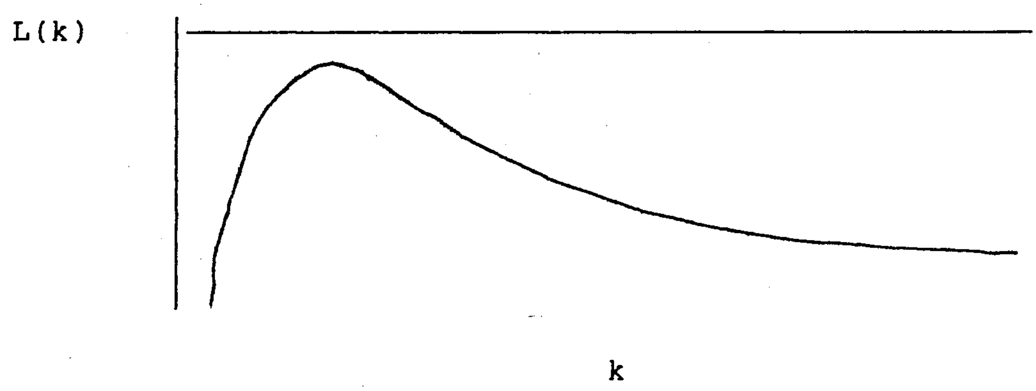

Figure 11. Log Likelihood Function

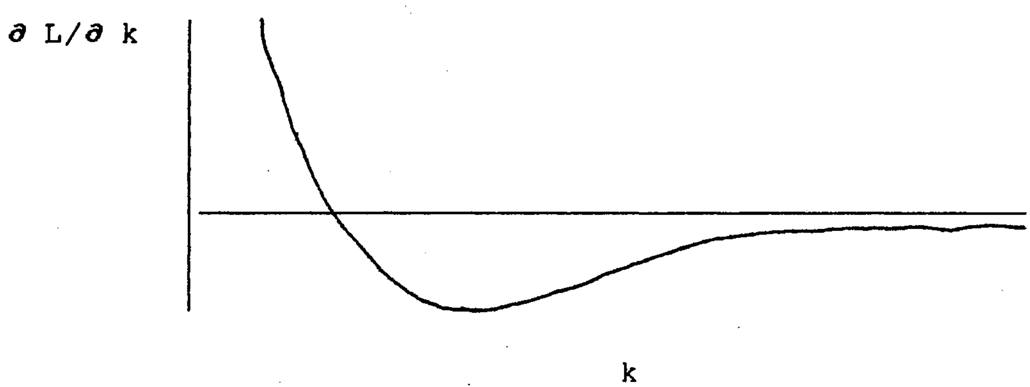

Figure 12. The Derivative Function for the Log Likelihood Function in Figure 11.

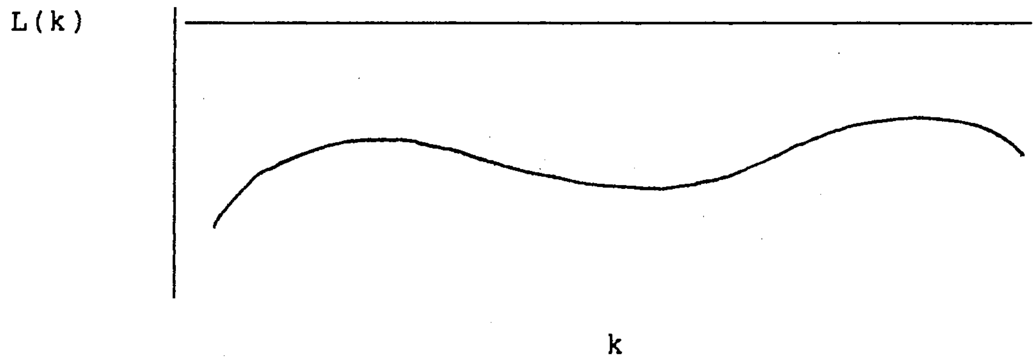

Figure 13. Log Likelihood Function With Multiple Maxima

REFERENCES

Anraku, K., and Yanagimoto, T. (1990). Estimation for the Negative Binomial Distribution Based on the Conditional Likelihood. *Communications in Statistics-Simulation* 19(3), 771-786.

Anscombe, F. J. (1950). Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. *Biometrika* 37, 358-382.

Clark, S. J. and Perry, J. N. (1989). Estimation of the Negative Binomial Parameter k by Maximum Quasi-Likelihood. *Biometrics* 45, 309-316.

Cramer, Harald (1946). Mathematical Methods of Statistics. Princeton University Press.

Folland, Gerald B. (1984). Real Analysis: Modern Techniques and Their Applications. Wiley Pure and Applied Mathematical Series.

Levin, B. and Reeds, J. (1977). Compound Multinomial Likelihood Functions are Unimodal: Proof of a Conjecture of I. J. Good. *The Annals of Statistics* 5, 79-87.

Piegorsch, W. W. (1990). Maximum Likelihood Estimation for Negative Binomial Dispersion Parameter. *Biometrics* 46, 863-867.

Pieters, E. G., Gates, C. E., Matis, J. H., and Sterling, W. L. (1977). Small Sample Comparisons of Different Estimators of Negative Binomial Parameters. *Biometrics* 33, 718-723.

Willson, L. J., Folks, J. L., and Young, J. H., (1984). Multistage Estimation Compared with Fixed-Sample-Size Estimation of the Negative Binomial Parameter K. *Biometrics* 40, 109-117.

Willson, L. J., Folks, J. L., and Young, J. H., (1986). Complete Sufficiency and Maximum Likelihood Estimation of the Two-Parameter Negative Binomial Distribution. *Metrika* 33, 349-362.

APPENDIX A

APPROXIMATE PROBABILITY OF TYPE I AND TYPE II ERRORS

These numbers represent the approximate Type I and Type II probabilities encountered when under-dispersion is used to distinguish between negative binomial and Poisson distributions. For the first row of the table, 2000 samples of size 35 from a negative binomial distribution were classified (reject NB if the sample was under-dispersed and accept NB if the sample was over-dispersed). The negative binomial parameters were μ = 1 and k = 3. The second row of the table summarizes the same decision process when 2000 samples of size 35 were generated from a Poisson distribution with μ = 1.

	Reject NB	Accept NB
Ho: NB	0.2025	0.7975
Ha: Poi	0.6050	0.3950

APPENDIX B

BOX PLOT COMPARISON OF LARGE LIKELIHOOD, MAXIMUM LIKELIHOOD, AND METHOD OF MOMENTS ESTIMATION

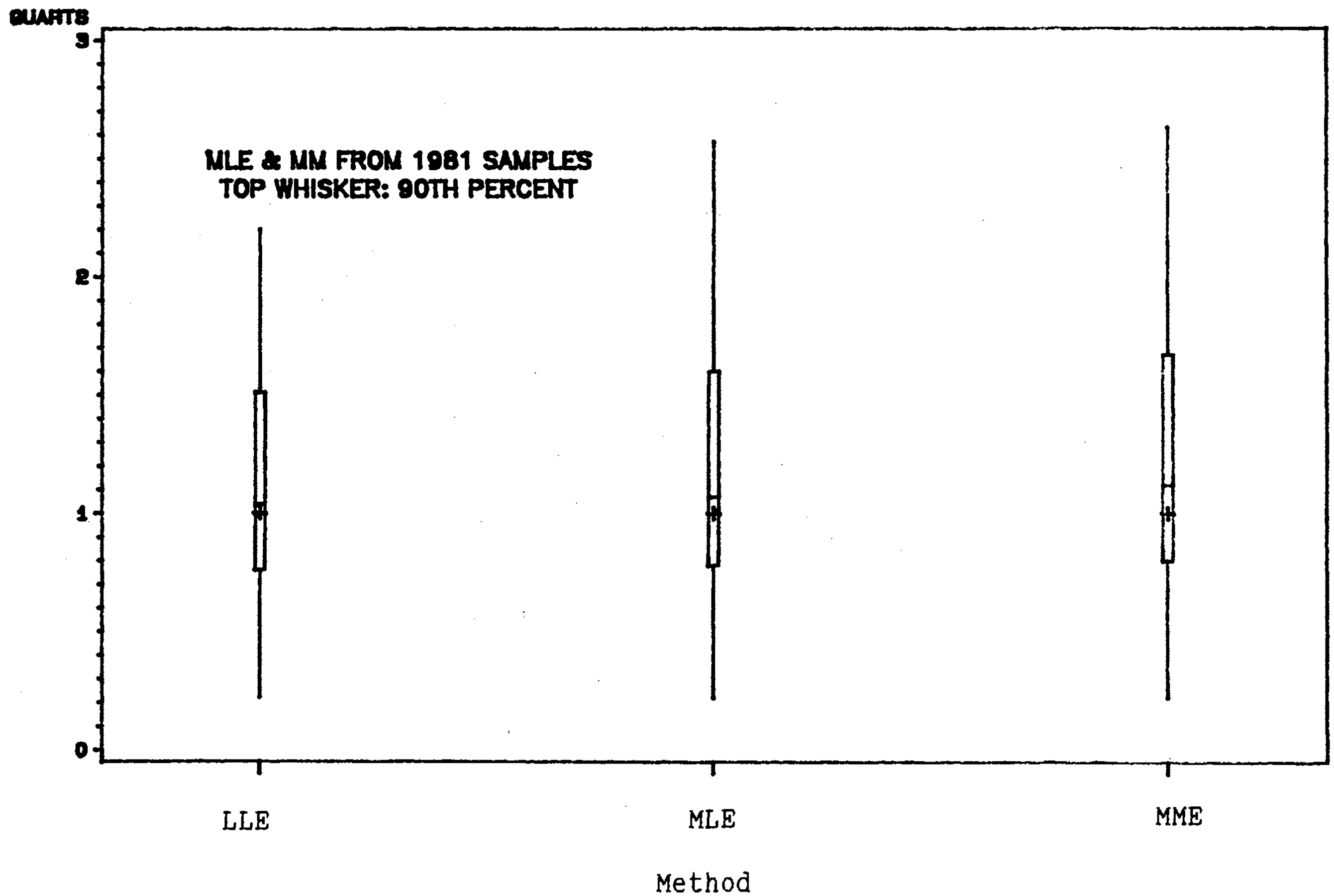

Figure 14. Comparison of the Three Estimation Methods, $\mu = 1$ and $k = 1$

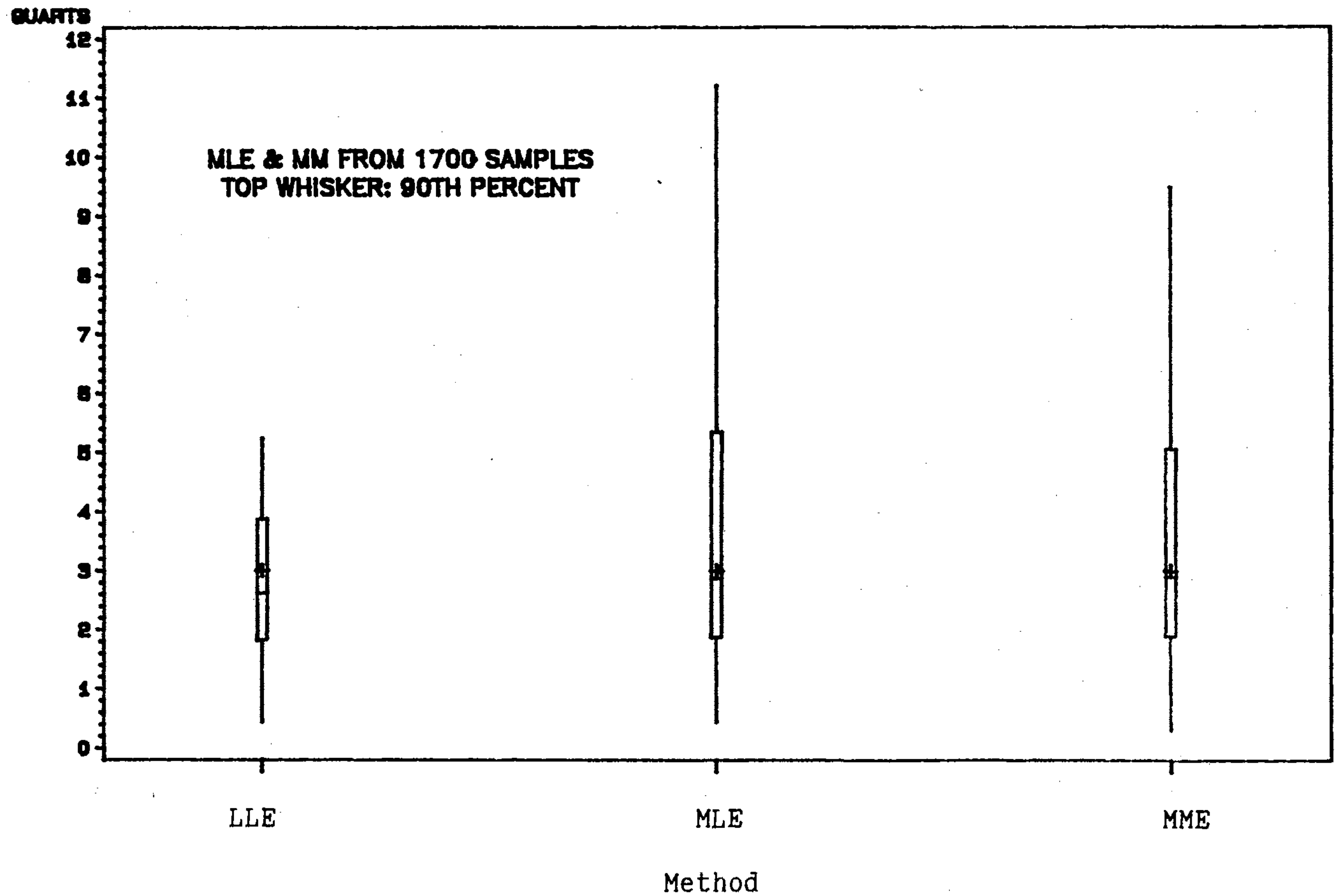

Figure 15. Comparison of the Three Estimation Methods, $\mu = 1$ and $k = 3$

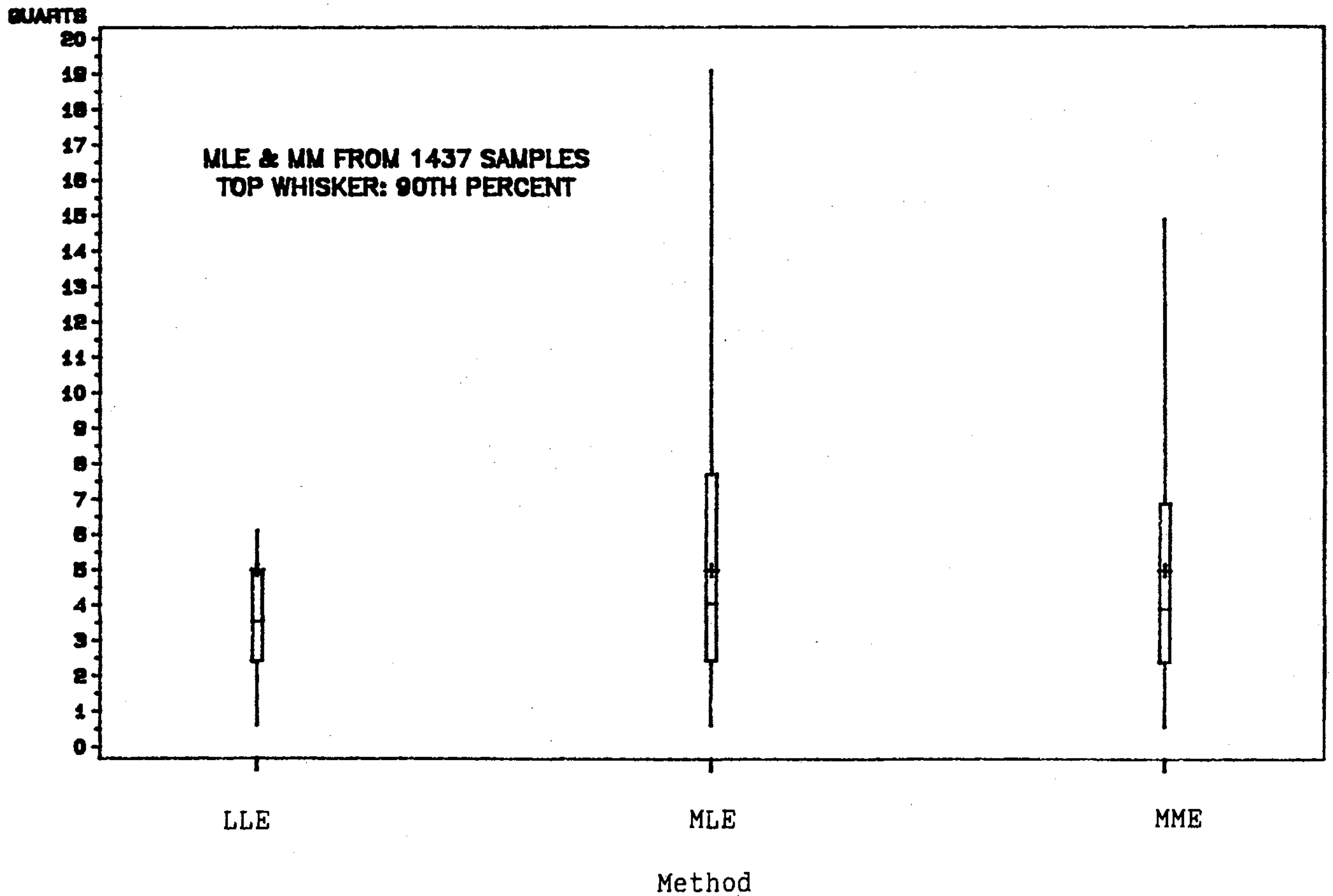

Figure 16. Comparison of the Three Estimation Methods, $\mu = 1$ and $k = 5$

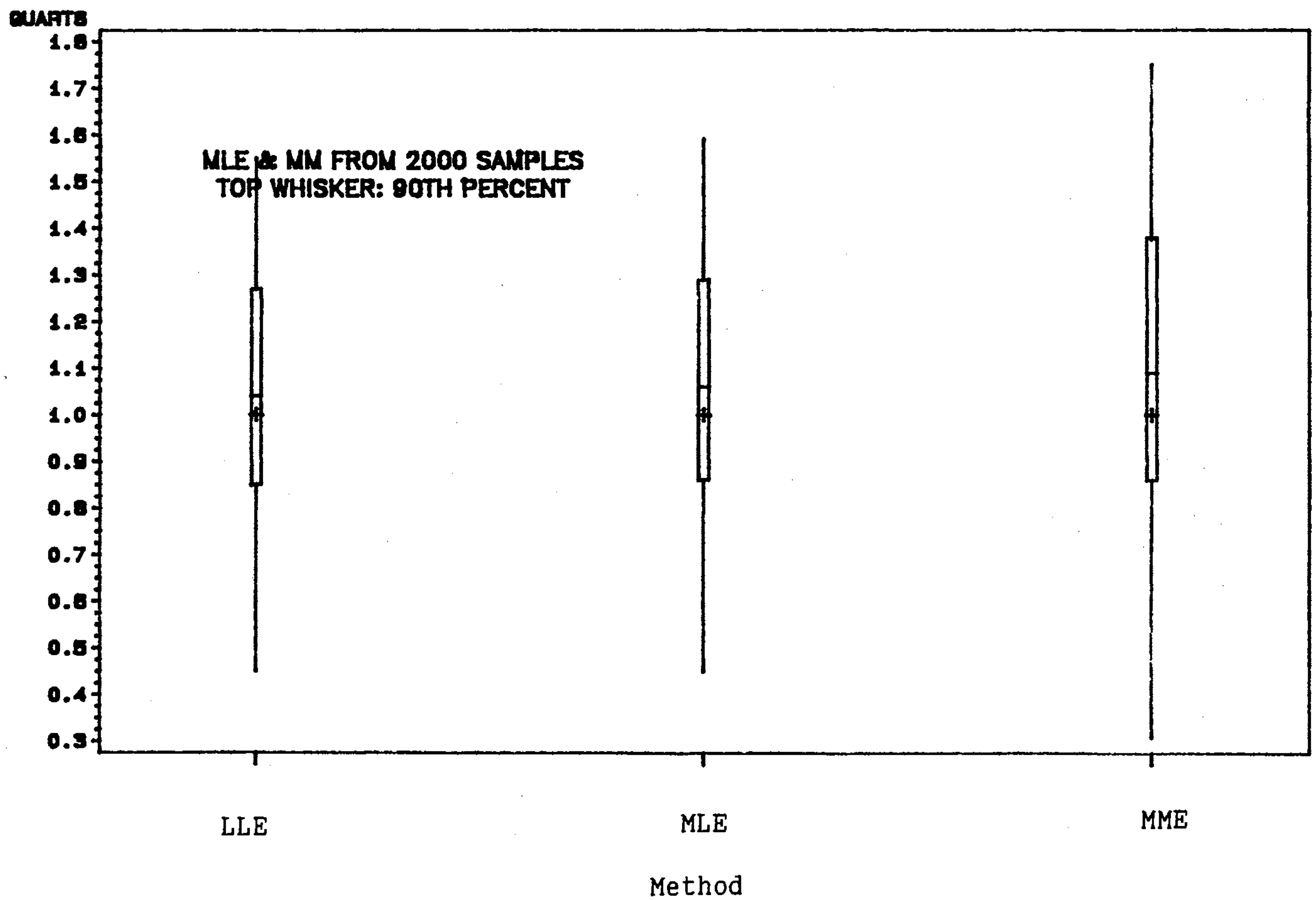

Figure 17. Comparison of the Three Estimation Methods, μ = 3 and k = 1

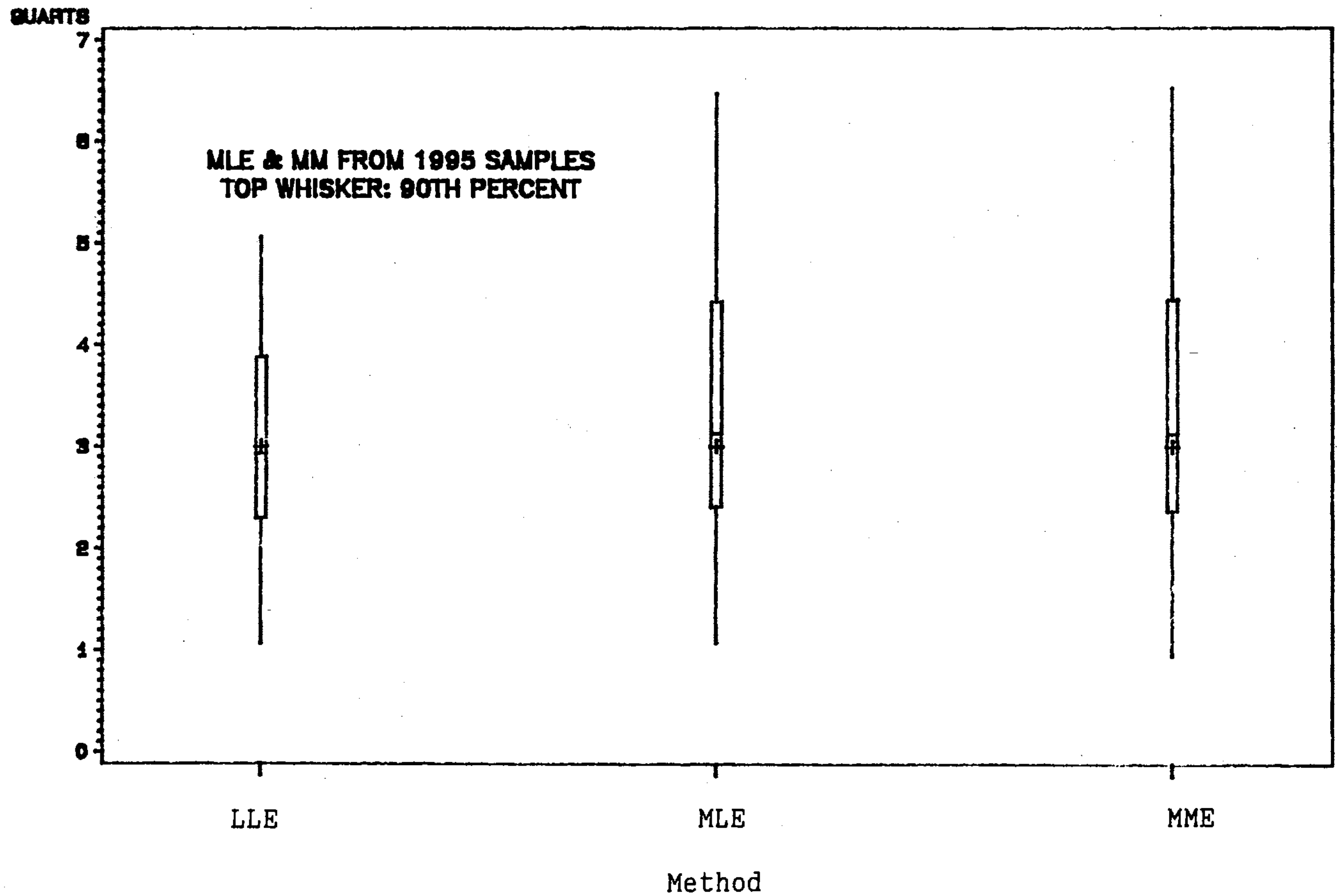

Figure 18. Comparison of the Three Estimation Methods, $\mu = 3$ and $k = 3$

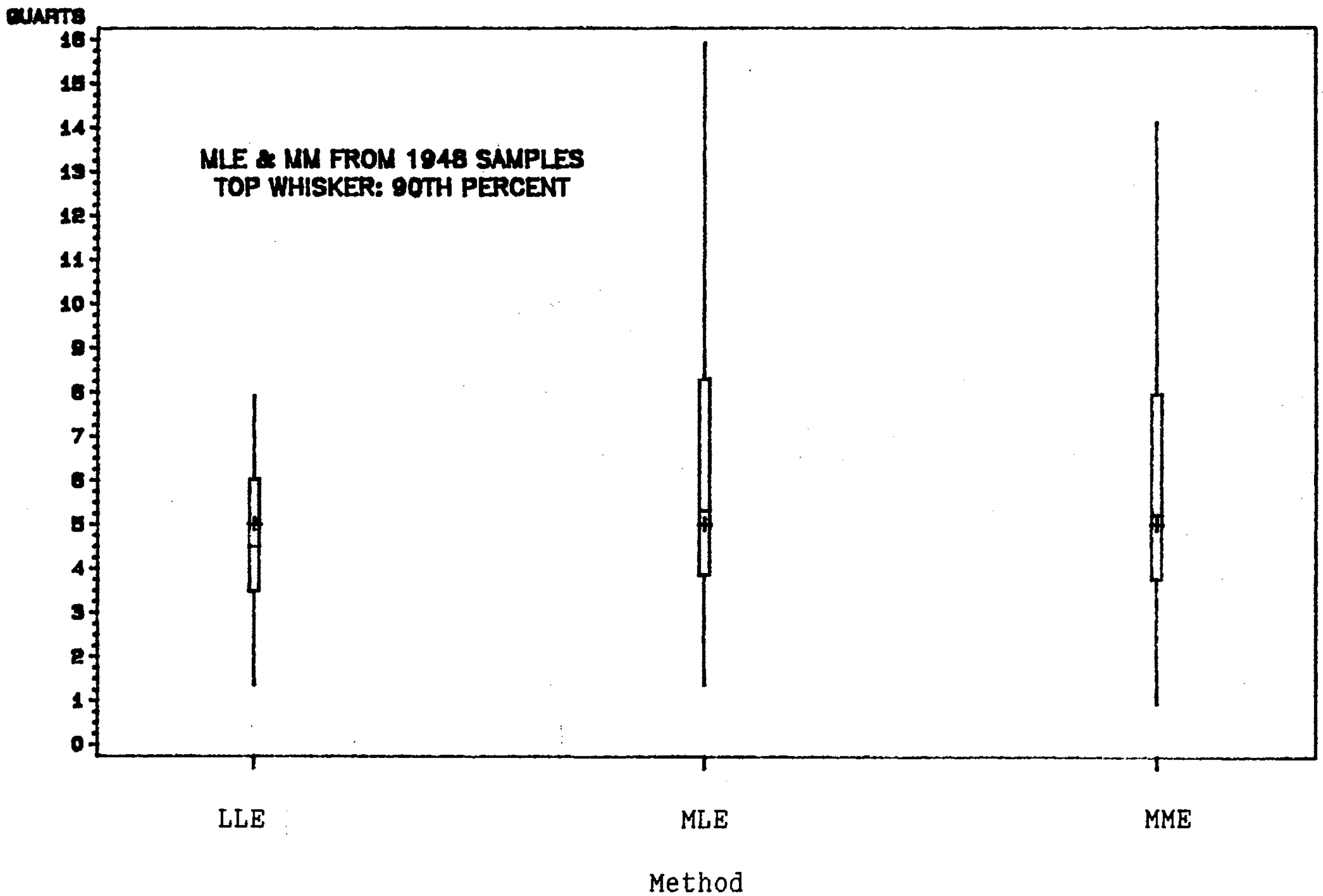

Figure 19. Comparison of the Three Estimation Methods, μ = 3 and k = 5

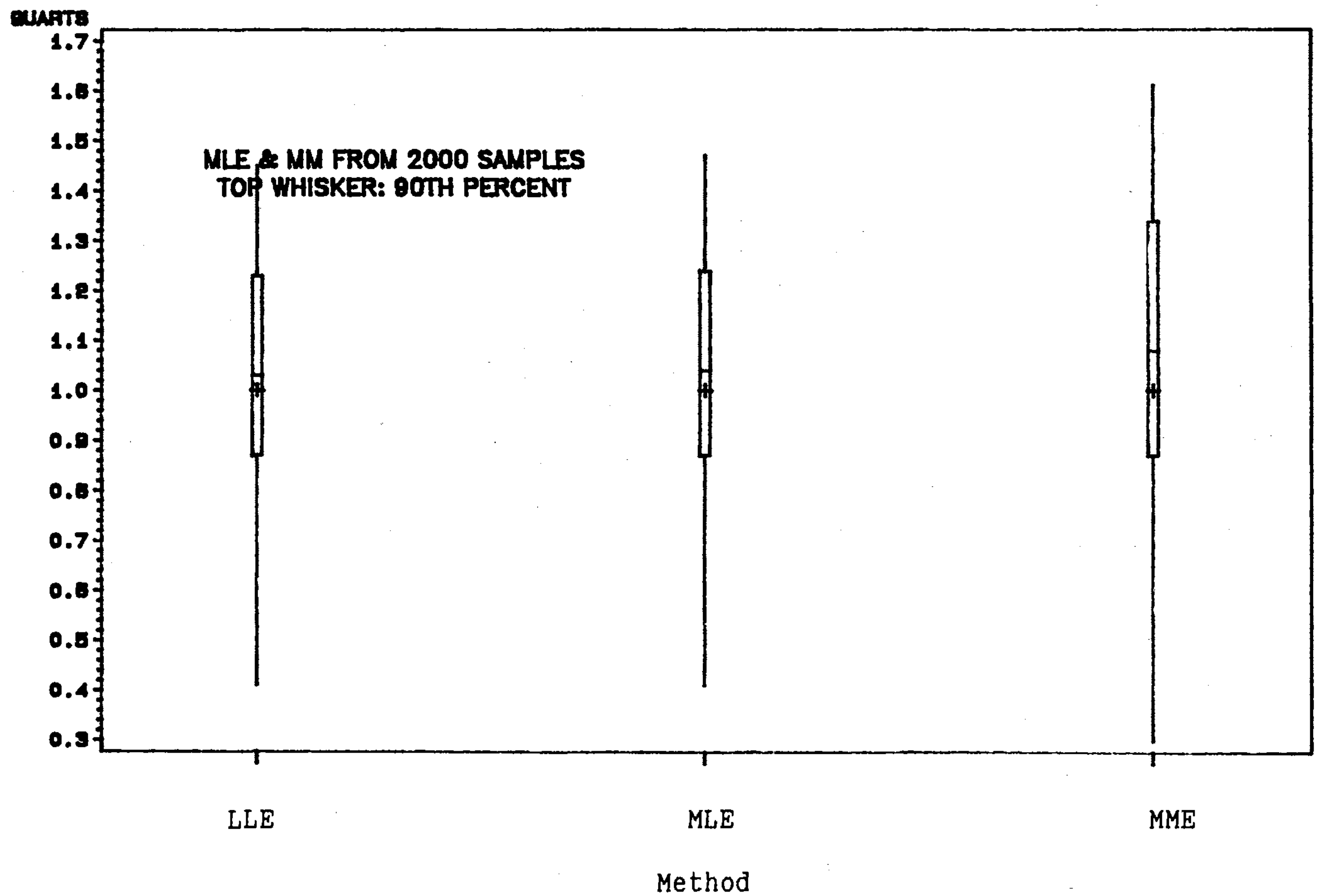

Figure 20. Comparison of the Three Estimation Methods, $\mu = 5$ and $k = 1$

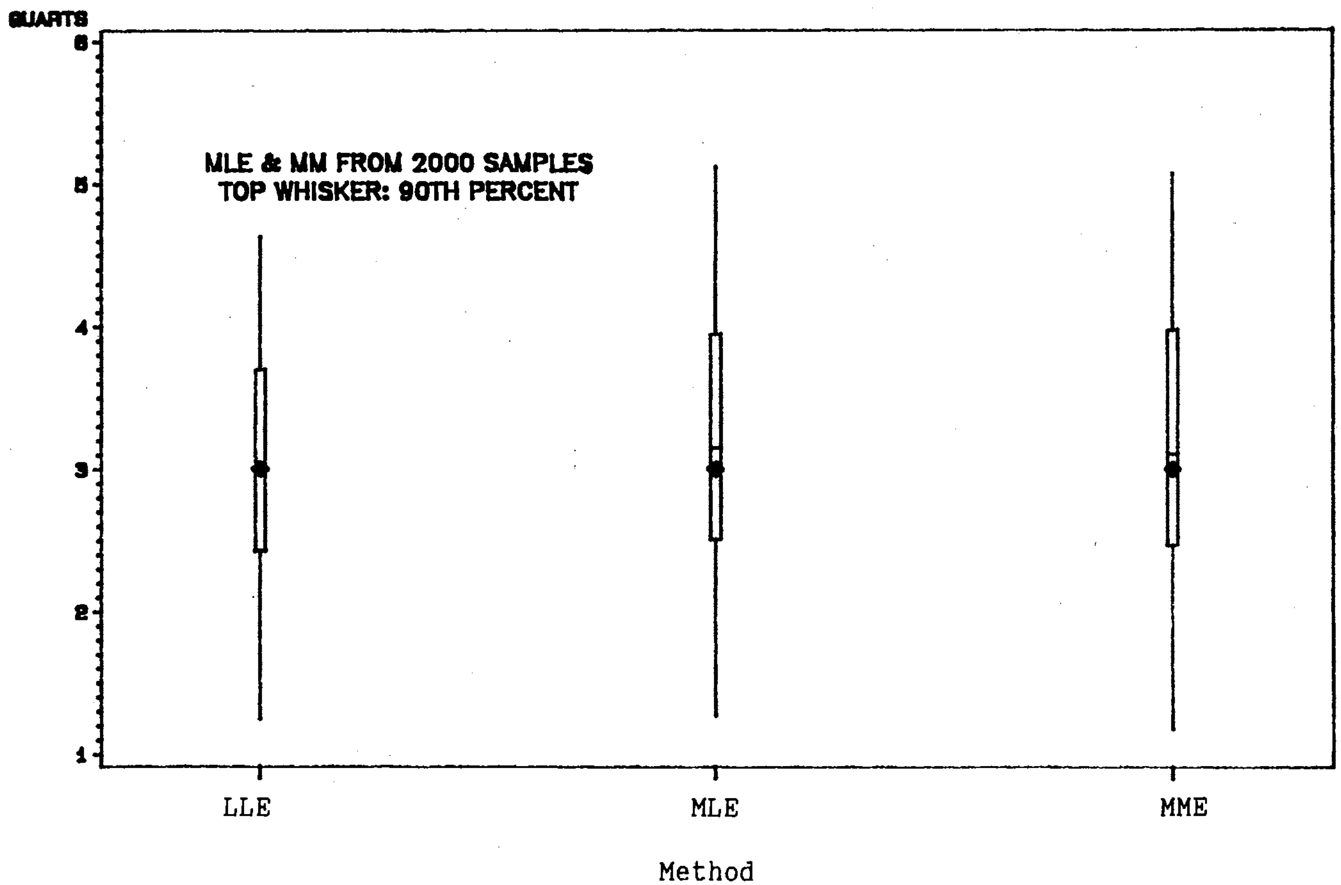

Figure 21. Comparison of the Three Estimation Methods, μ = 5 and k = 3

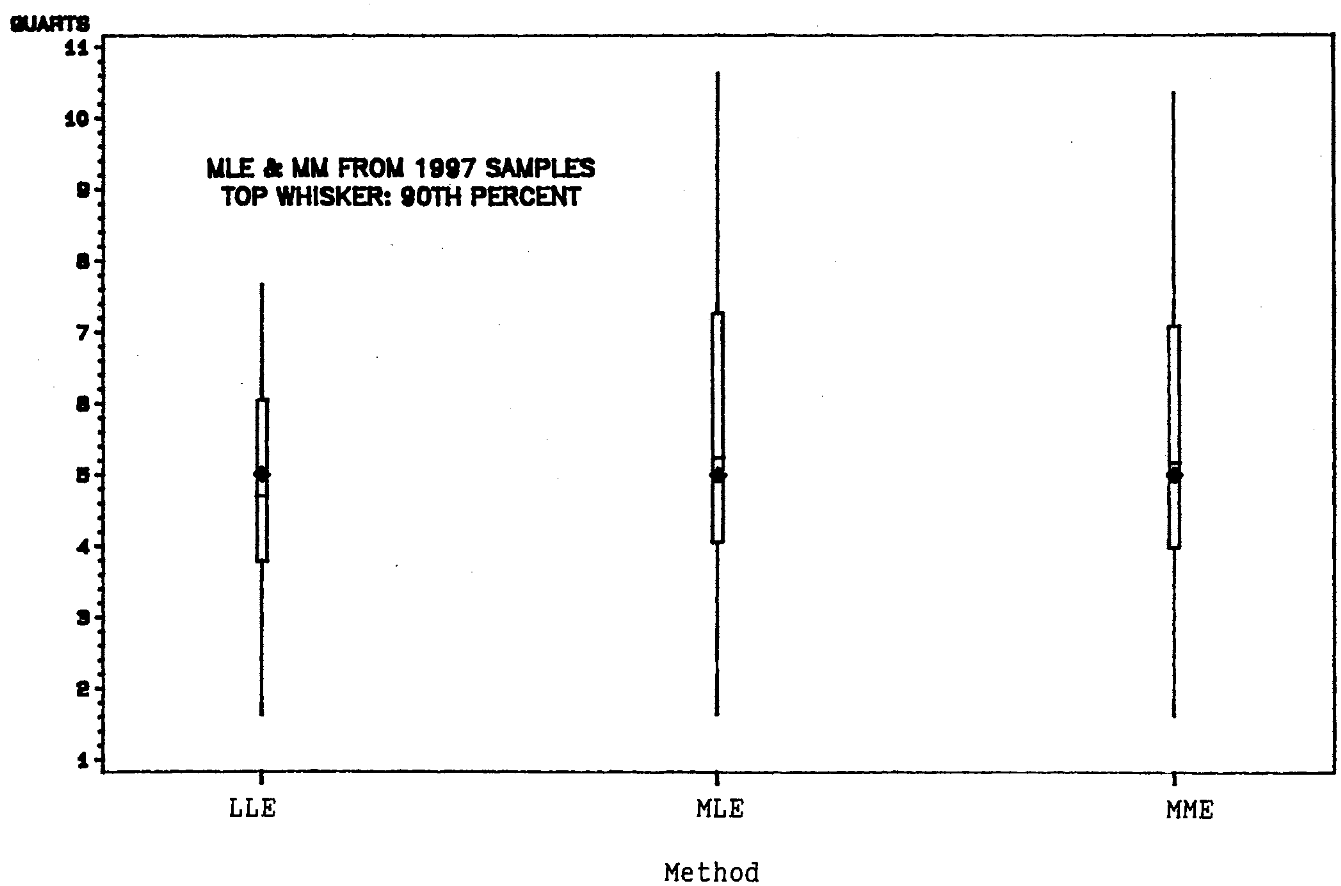

Figure 22. Comparison of the Three Estimation Methods, μ = 5 and k = 5

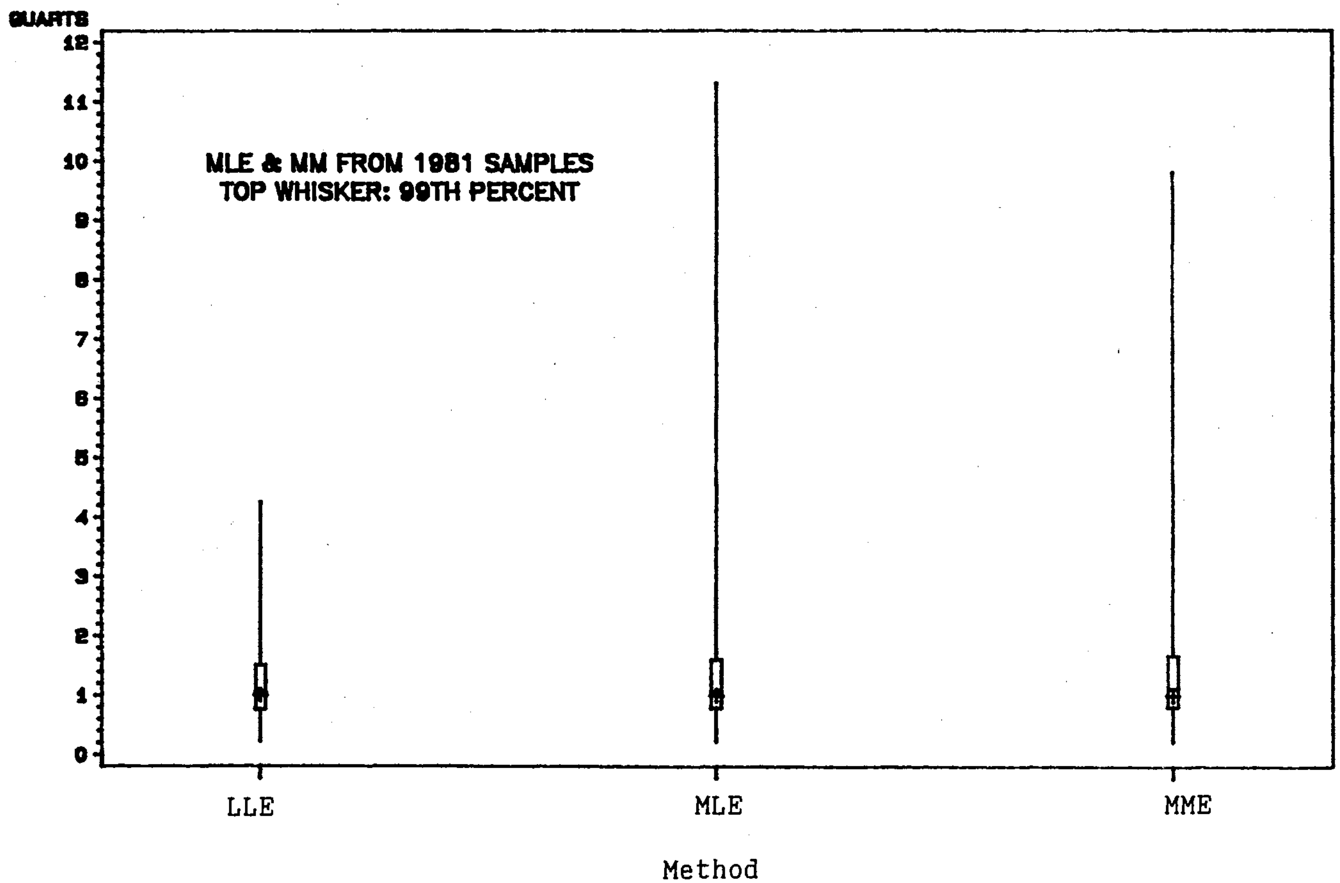

Figure 23. Comparison of the Three Estimation Methods, $\mu = 1$ and $k = 1$

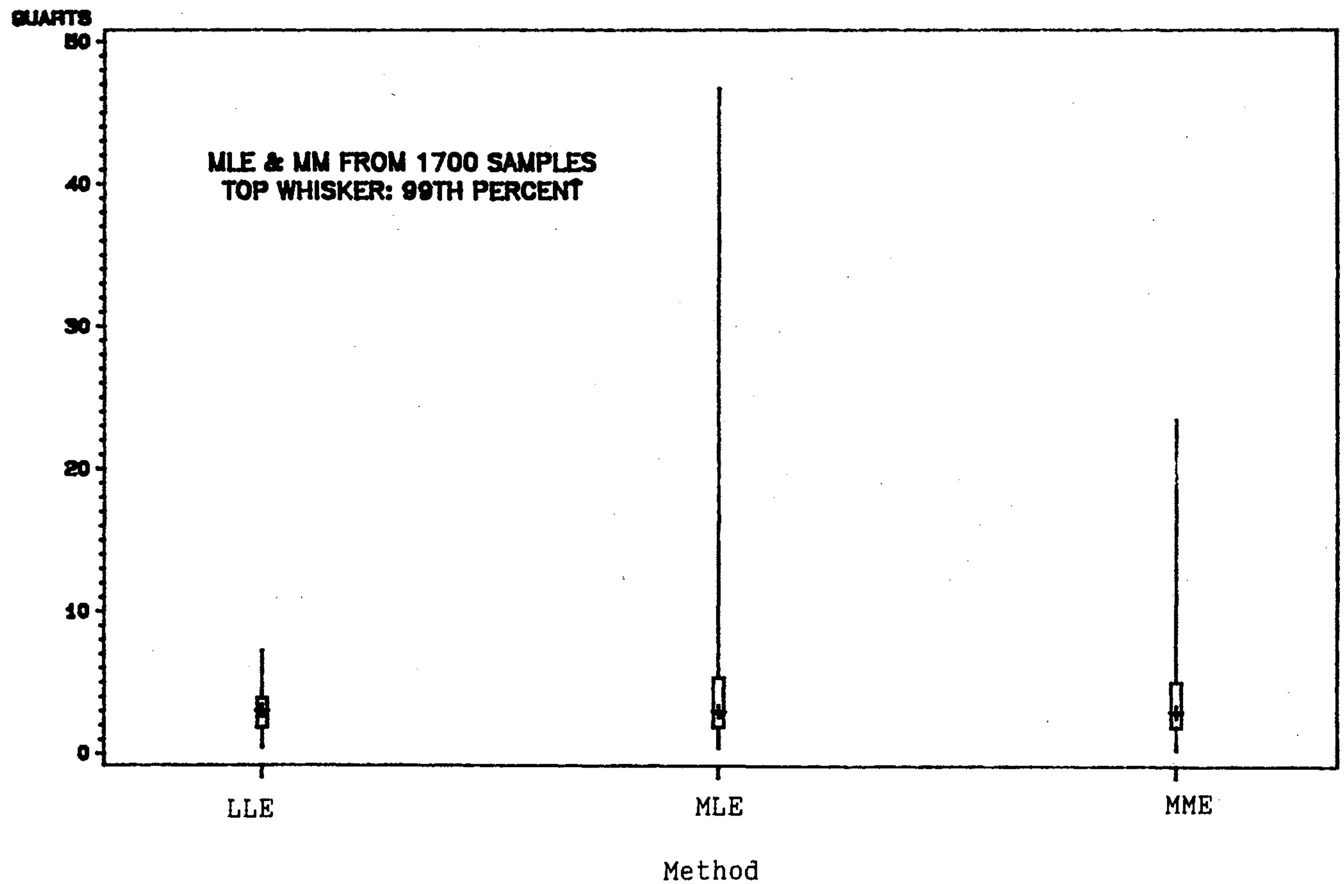

Figure 24. Comparison of the Three Estimation Methods, $\mu = 1$ and $k = 3$

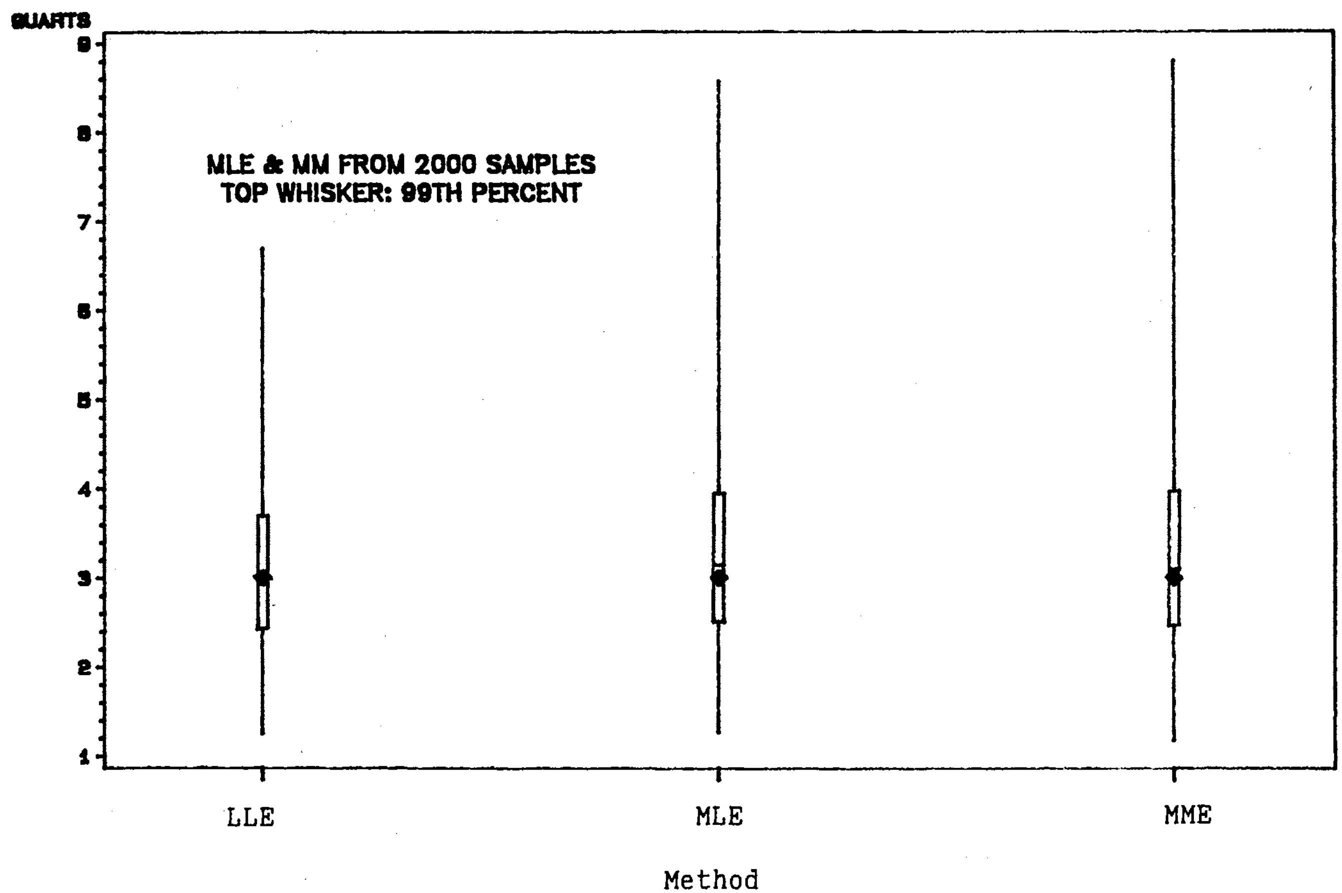

Figure 25. Comparison of the Three Estimation Methods, μ = 5 and k = 3

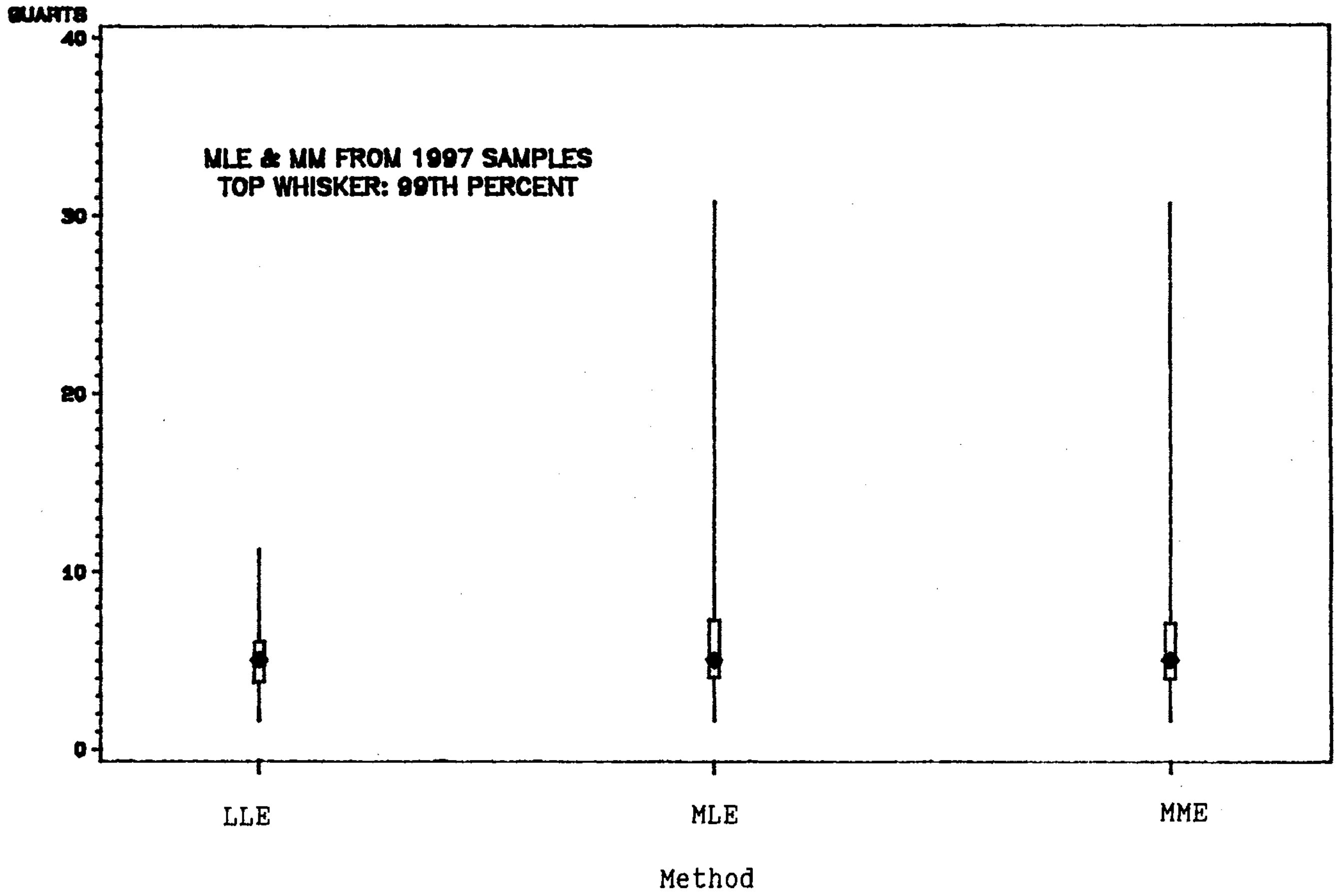

Figure 26. Comparison of the Three Estimation Methods, μ = 5 and k = 5

APPENDIX C

SAS-IML SOURCE PROGRAM FOR THE SIMULATIONS

```
proc iml;reset noname;  start main;
 ss= 50; big=500 ;  seed=0;
    filename uuuu 'c:\rls\u1k3.dat';
    file uuuu;

    valu = j( 4,1,0);
    valu (|1|) =  0   ;
    valu (|2|) =  0.13;

    do iiii = 1 to big;
    x = j(ss,1,0);
    u=1; k=3;
    p= k/(u+k);
    do i = 1 to ss;
    xx = 0;    zzz=0;
    pp= uniform(seed);
   do while ( zzz=0 );
   if pp < probnegb(p,k ,xx) then do;
     zzz=1;
     if xx=0 then hold =0;
     else hold =xx;
   end;
    xx=xx+1;
    end;
   x(| i |) =hold;
   end;

   xbar = sum(x) / ss ;
   square =(ssq(x)-sum(x)*sum(x)/ss)/(ss-1);
   square1=(ssq(x)-sum(x)*sum(x)/ss)/(ss)  ;
   rat = square1/ xbar;
   maxx = max(x);
   fre = j( maxx, 1,0);

   do j = 1 to ss;
    if x(|j|) > 0 then do;
    fre(| x(|j|) |) = fre(|  x(|j|) |) +1;
    end;  *Of do;
   end; *of j;

   if square1 > xbar then do;
   ggo = 1;  stopp= 4;
   mme = (xbar*xbar)/(square - xbar ); end;
   else do; ggo = 2; stopp = 3; valu(|3|) = 0.09;
   mme = .; mle = .; end;
    do yyy = ggo to stopp;
    tol=0.0001;top=35;newk=0.08;oldk=newk+1;lim=0;
    do while (abs(newk-oldk)>tol & lim <top);
    lim   = lim   + 1;
    hold = 0;
    do j = 1 to maxx;
    sum=0;
```

```
      do i = j to maxx;
      sum = sum + fre(|i|);
      end;
     hold = hold + sum /(  ( newk+ j - 1 ));
     end;

     numer=hold-ss*log(1+xbar/newk)-valu(|yyy|);
     hold = 0;
     do j = 1 to maxx;
      sum=0;
      do i = j to maxx;
    sum = sum + fre(|i|);
      end;
     hold = hold - sum /((newk+ j - 1 )**2 );
     end;
     denom = hold +ss* xbar/((newk+xbar)*newk);
      oldk=newk;
     newk = oldk - numer/denom;
     end;  if yyy =1 then do; mle = newk;
    newx = x`;
    do tt = 1 to 2;
    macx = max( newx );
    minx = min( newx );
    do t = 1 to ncol( newx);
    if newx(|t|) = macx then maci = t;
    if newx(|t|) = minx then mini = t;
    end;
    newx = remove( newx , maci||mini );
    end;
    mu = sum( newx )/ ncol(newx) ;
valu(|3|)=(0.7050*(mu##0.2792)/(0.8639*(mu##1.0284)
/(rat - 0.9752 ) )##1.0518 ) - 0.0589 ;
valu(|4|)=(0.7050*(mu##0.2792)/(4.6379*(mu##0.6814)/
( - denom  + 0.1863 ) )##0.4391 ) - 0.0589 ;
if valu(|3|) < < 0.025 then valu(|3|) = 0.025;
   if valu(|3|) > 5.50 then valu(|3|) = 5.50;
   if valu(|4|) < 0.025 then valu(|4|) = 0.025;
   if valu(|4|) > 5.50 then valu(|4|) = 5.50;
  end; * of do;
  else if yyy =2 then lle  = newk;
  else if yyy =3 then llerat  = newk;
  else if yyy =4 then llesec = newk;

    end ;  * of yyy;
    if xbar >= square1 then  llesec=llerat;
    put  @1  xbar 5.2 +1 mu   7.3 +1  mme 8.4 +1
  mle 8.4 +1 lle 8.4 +1 llesec 8.4 +1 llerat 8.4
  ;
   end;  * o f   b i g   l o o p ;
    closefile uuuu;
   finish; run main;
```

VITA

Robert L. Steiner, Jr.

Candidate for the Degree of

Doctor of Philosophy

Thesis: THREE ESTIMATION METHODS FOR THE NEGATIVE BINOMIAL PARAMETER k

Major Field: Statistics

Biographical:

Personal Data: Born in Ada, Oklahoma, December 12, 1953, the son of Bob and Mary Steiner.

Education: Graduated from Ada High School, Ada, Oklahoma, May 1972; Received Bachelor of Science in Pharmacy from the University of Oklahoma, Norman, May 1978; Received Bachelor of Science in Mathematics from East Central University, Ada, Oklahoma, May 1984; Received Master of Arts in Mathematics from the University of Oklahoma in May 1987; Completed the requirements for the Doctor of Philosophy in May, 1993, Oklahoma State University, Stillwater, Oklahoma.

Professional Experience: Registered Pharmacist, Valley View Hospital, Ada, Oklahoma from May 1978 to April 1982. Graduate Teaching Assistant University of Oklahoma, August 1984 to May 1989; Graduate Teaching Associate, Oklahoma State University, August 1989 to May 1993.