Four Essays in Finite-Sample Econometrics

by

Qian Chen

B.A., LaiYang Agricultural University, P.R. China, 2001

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of Economics

Four Essays in Finite-Sample Econometrics

by

Qian Chen

B.A., LaiYang Agricultural University, P.R. China, 2001

Supervisory Committee

Dr. David E.A. Giles (Department of Economics)
Supervisor

Dr. Nilanjana Roy (Department of Economics)
Departmental Member

Dr. Graham M. Voss (Department of Economics)
Departmental Member

Dr. Min Tsao (Department of Mathematics and Statistics)
Outside Member

Supervisory Committee

Dr. David E. A. Giles
Supervisor

Dr. Nilanjana Roy
Departmental Member

Dr. Graham M. Voss
Departmental Member

Dr. Min Tsao
Outside Member

ABSTRACT

In this dissertation, we explore the use of three different analytical techniques for approximating the finite-sample properties of estimators and test statistics. These techniques are the saddlepoint approximation, the large-n approximation and the small-disturbance approximation. The first of these enables us to approximate the complete density or distribution function for a statistic of interest, while the other two approximations provide analytical results for the first few moments of the finite-sample distribution. We consider a range of interesting estimation and testing problems that arise in econometrics and empirical economics. Saddlepoint approximations are used to determine the distribution of the half-life estimator that arises in the empirical purchasing power parity literature, and to show that its moments are undefined. They are also applied to the problem of obtaining accurate critical points for the Anderson-Darling goodness-of-fit test. The large-n approximation is used to study the first two moments of the MLE in the binary Logit model. Finally, we use small-disturbance approximations to examine the bias and mean squared error of some commonly used price index numbers, when the latter are viewed as point estimators.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

Comparison of the Normal-Based and Chi-Squared-Based Saddlepoint Approximation

Comparison of the Lower-Order and Higher-Order Saddlepoint Approximations

CHAPTER 4

ACKNOWLEDGMENTS

I have been aided immeasurably by my supervisor Professor David Giles in the process of initializing the idea for this dissertation and also in the process of research and writing up the final paper. From the bottom of my heart, I want to express my sincerest thanks to Professor David Giles for all of his guidance, encouragement and patience during the course of this work. Also, I would like to thank Dr. Nilanjana Roy, Dr. Graham Voss and Dr. Min Tsao for their valuable comments and critical review. I would also like to extend my appreciation to Professor Andy Wood for very helpful correspondence relating to his own work on the saddlepoint approximation.

I am in debt to the Economics Department at the University of Victoria, for generous financial aid, high quality teaching, encouraging faculty members, and warm and friendly support from Alma, Judy and Lori. I greatly appreciate all of this help.

I would also like to thank my husband for his persistent love and encouragement. I thank all my friends in Victoria for their great help in all ways.

DEDICATION

To my dearest husband!

To my mother, father and brother!

To my parents in-law!

CHAPTER 1:
INTRODUCTION

1. Introduction

Empirical work in economics relies on the use of a large variety of econometric "tools". Many different estimators and statistical tests are used to help economists draw inferences about the parameters of their models, and ultimately about the economy itself. However, in many situations, the exact characteristics of the sampling distributions of the estimators and the test statistics that we use are not known when the sample size is finite. The main justification for many, if not most, of the estimators and test statistics that are used in this field are their asymptotic properties, based on the hypothetical assumption that the sample size is infinitely large. For example, under certain "regularity conditions" on the underlying data density, the Maximum Likelihood estimators (MLE's) that form the basis for much of our inferences are "Best Asymptotically Normal". Similarly, other widely used families such as the class of Instrumental Variables estimators, and Generalized Method of Moments estimators, are also chosen for their large-sample features. Notably, they are at least weakly consistent (if constructed appropriately), and can achieve asymptotic efficiency under some circumstances. However, there are no guarantees when it comes to the finite-sample properties of any of these estimators – they may be unbiased or biased, for example.

Similarly, the general testing strategies that we rely on, such as the Wald test, the Likelihood Ratio test and the Lagrange Multiplier test, also have good asymptotic properties. Moreover, in many cases, and under appropriate regularity conditions, the forms of the limiting distributions of the associated test statistics are known. So, appropriate critical regions can be constructed if the sample size is very large, but their use in the case of finite samples may involve an unknown degree of distortion in the sizes of the tests, with corresponding implications for their power.

Of course, there are exceptions to all of this. For example, in the case of the standard linear multiple regression model with errors that are independent, homoskedastic, and normally distributed, the MLE for the coefficient vector is the minimum variance unbiased estimator. Similarly, when testing the validity of exact linear restrictions on this coefficient vector, the Likelihood Ratio, Wald and Lagrange Multiplier tests are equivalent to the usual uniformly most powerful F-test that we usually use. However, these familiar examples are the exception rather than the rule, and it will be recalled that even in this same model, the MLE for the variance of the error term is biased.

Lack of knowledge about the sampling distributions of our estimators and test statistics in finite samples places severe limitations on our ability to draw appropriate inferences in many practical situations. Not only do we often have to work with samples of only modest size, but we also can't tell in practice if a particular sample is "large enough" for us to be able to appeal to the asymptotic properties with any degree of confidence. What is a "large" sample in some situations is actually very "small" in others. So, it is important to learn more about the finite sample properties of the estimators and test statistics that we use. In particular, we are interested in the finite-sample distribution, or at least the first few moments of this distribution, in the case of estimators. For the tests, we are interested in the full finite-sample distribution of the test statistic, or at least the tails of this distribution, under the null hypothesis if we are to be able to construct valid critical regions. The same information, under the alternative hypothesis, is needed if we want to determine the powers of the tests in finite samples.

There are several well developed methods that can be used to explore the finite sample properties of the estimator and test statistics. For example, Monte Carlo and Bootstrap methods are two which are widely applied throughout the literature. Both of these methods are based on the use of simulation techniques to generate some specific numerical approximations to the sampling distribution, in selected cases. However, in this dissertation, we will focus on applying methods which can generate more general, and *analytic*, results relating to the finite sample properties of estimators and tests. These include the saddlepoint approximations, the large-n approximations and the small-σ approximations. These methods are well developed, but the first of them has not been widely used in econometrics. In this dissertation, we provide some new analytic results relating to the saddlepoint approximation, and all three types of approximations are used to determine the performances of various estimators and an important test. Of course, there are various other methods which can provide the finite sample analytic results of the type that we are concerned with here, such as

Laplace approximations and Edgeworth approximations. The latter, for example, has been used widely in analyzing the distributions of many econometric estimators for simultaneous equations models. We hope that our results will encourage further use of these techniques by econometricians. Now we turn to a brief introduction of the finite sample methods that we will be using. In the case of the saddlepoint approximation, the discussion is more detailed as this technique may be less familiar than the other techniques to readers with an econometrics background.

2. Saddlepoint Approximations

Saddlepoint approximations can be obtained for any statistic which admits a cumulant generating function. We know that once we have the cumulant generating function we can obtain the characteristic function and then use the inversion theorem to obtain the density and the distribution functions. However in many situations, the complexity of the associated integration makes it impossible to get an exact analytic result. Or even if we can, the integration is quite time consuming. Saddlepoint approximations are based on the Fourier inversion formulae for the density, and apply the steepest descent method to the integration to derive an approximation for the density function. Of course, alternative approximations are possible. For example, the Edgeworh expansion provides a very good approximation in the center of the density. However, in some cases, it generates a poor approximation in the tail area of the density, even giving rise to negative "probabilities". Compared to this, the improvements of the saddlepoint approximation over the Edgeworth expansion include: first, the saddlepoint approximation always generates positive probabilities; second, the saddlepoint approximation is very accurate in the tail area where most of the interest lies in the case of testing problems; and third, the performance of the saddlepoint approximation is accurate for small samples, sometime even with only one observation.

The logic behind the saddlepoint approximation for a density function is the following. Let X be a random variable with a density function $f(x)$. Then the associated characteristic function is defined as:

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx,$$

Where "i" is the imaginary number satisfying $i^2 = -1$. The cumulant generating function is $K_X(t) = \log(\phi_X(t))$. By applying the Fourier inversion formula, we can recover the density function, $f(x)$:

$$f(x)=\frac{n}{2\pi}\int_{-\infty}^{+\infty}\phi_X(it)e^{-itx}dt\,, \tag{1}$$

If we let $it=t$, then we can write (1) as

$$f(x)=\frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty}\exp(K_X(t)-tx)dt\,. \tag{2}$$

We define the point $\hat{t}$ that satisfies the equaiton

$$K'_X(t)=x\,.$$

Expanding the argument of the exponential function in (2) around $\hat{t}(x)$, we obtain

$$K_X(t)-tx\approx K_X(\hat{t}(x))-\hat{t}(x)x+\frac{(t-\hat{t}(x))^2}{2}K''_X(\hat{t}(x))\,. \tag{3}$$

Then, we substitute (3) into (2) and integrate with respect to t along the path parallel to the imaginary axis through the point $\hat{t}(x)$. This yields the saddlepoint approximation of the density function $f(x)$ as

$$f(x)\approx\left(\frac{1}{2\pi K''_X(\hat{t}(x))}\right)^{1/2}\exp\{K(\hat{t}(x))-\hat{t}(x)x\}\,.$$

Saddlepoint approximations have been proved useful for a range of distributional problems. Therefore, such approximations have attracted a lot of interest, both in theory and in practice, as is discussed below. The saddlepoint approximation was studied first by Daniels (1954). In this seminal paper, Daniels derived a saddlepoint approximation for the mean of n independent identically distributed random variables. Daniels also discussed the application of the saddlepoint approximation to the case of discrete variables and the distribution of ratios of sums of random variables. With an appropriate renormalization, to ensure that the saddlepoint density function integrates to unity over its support, the accuracy of the approximation can sometimes be improved, and in certain cases it even reproduces the exact density. Daniels (1980) studied the conditions under which the renormalized saddlepoint approximation can reproduce the exact density of an estimator. He found that the normal, gamma and inverse normal are the only three cases for which the renormalized saddlepoint approximation reproduces the exact density of the estimator in the scalar case. In fact, in the cases of normal and inverse normal, the saddlepoint approximation can reproduce the exact density even without renormalization. Goutis and Casella (1999) is a good reference on the subject of the saddlepoint approximation.

Barndorff-Nielsen & Cox (1979) generalize the saddlepoint approximation of the univariate case to the bivariate and the multivariate distributions. They also provide two other important extensions - the single and the double saddlepoint expansions, for the conditional distribution. Barndorff-Nielsen & Cox illustrate the application of the theoretic results to the conditional likelihoods and maximum likelihood ratio test statistics. Skovgaard (1987) also derived the saddlepoint expansions for conditional distributions in bivariate case. Then he generalized this expansion to the conditioning on a p-1-dimensional linear function of a p-dimensional variable. Butler and Huzurbazar (1992) applied the conditional saddlepoint approximation to the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. Jing and Robinson (1994) and Gatto (1996) examined the saddlepoint technique in approximating the marginal densities of some variables. Butler and Sutton (1998) applied the saddlepoint approximation to four multivariate distributions: the multinomial, multivariate hypergeometric, Dirichlet, and multivariate Pólya distributions; and to two tests: outlier discordancy tests and slippage tests.

In many cases, such as hypothesis testing, our primary interest is with one or both tails of the statistic's sampling distribution. So it is natural to focus on the application of the saddlepoint approximation to the tail areas. Lugannini and Rice (1980) produced a very accurate saddlepoint approximation for the cumulative distributive function for the tail areas, called the Lugannini-Rice formula or L-R formula. Some other approaches for the tail areas can be found in Barndorff-Nielsen (1991). Barndorff-Nielsen discussed the relation of the tail area approximation determined by a modified signed log likelihood ratio to a generalization of the L-R approximation. Daniels (1987) provided a full discussion of different saddlepoint approximations for the tail areas. All of the approximations discussed in these works are normal-based approximations, and these seem to provide good accuracy when the statistics have normal limiting distributions. It is obvious to ask if a generalization to the non-normal-based formula may improve the accuracy when the limiting distribution of the variable is non-normal. Wood, *et al.* (1993) proposed the generalization of the non-normal-based saddlepoint approximation, which we will call the WBB formula. They compared the performance of the normal-based and non-normal-based approximations for two examples: a linear combination of chi-squared variables, and the first passage time distribution of a positive persistent random walk. Their results showed that the non-normal-based approximation can be more accurate than the normal-based approximation.

In some situations, the statistic of interest cannot be expressed in a closed form, but is defined by a system of equations. It is often interesting to apply the techniques of the saddlepoint approximation to this situation. Daniels (1983) extended the saddlepoint method to the probability density of an estimator defined by an estimating equation. He derived two formulae: one to approximate the density function, the other to approximate the cumulative distribution function. Spady (1991) extended the saddlepoint approximation from the scalar estimator to a vector of estimators defined by a system of estimating equations. Spady's derivation can be applied to the case where the observations are independently but not identically distributed. The approximation was applied to the least absolute deviations regression, and the results showed that the normal-based saddleppoint approximation provides apparent improvement when the asymptotic normal approximation is very misleading.

The main requirement of the saddlepoint approximation is the knowledge of either the cumulant generating function or the moment generating function. However, sometimes, the cumulant generating function is intractable. Therefore, some statisticians have examined cases when the true cumulant generating function is unavailable. Feuerverger (1989) derived the saddlepoint approximation for the case when the cumulant generating function can be obtained only empirically. Feuerverger gave a full discussion about the properties of the empirical moment generating funciton and the empirical cumulant generating function and the derivatives of these functions which are needed for the saddlepoint approximation. Monti and Ronchetti (1993) investigate the relationship between empirical likelihood and empirical saddlepoint approximations for multivariate M-estimators by comparing the expansions of the empirical log-likelihood and the empirical cumulant generating function. This leads to a nonparametric approximation of the density and the confidence regions of a multivariate M-estimator. Easton and Ronchetti (1986) derived a general saddlepoint approximation. They use the first four cumulants to approximate the cumulant generating function, and then construct the saddlepoint approximation for any statistic. Gatto and Ronchetti (1996) extended the general saddlepoint approximations to approximate the marginal densities and tail probabilities of a general nonlinear statistic.

Many other applications of the saddlepoint approximations have been studied. Robinson (1982) used the saddlpoint approximation to obtain the significance levels and confidence intervals of the permutation tests in the one and two sample problems. Jensen (1988) considered the saddlepoint approximation in the case when the limit of the point evaluated goes to ∞. Fraser *et al.* (1991) described a simple numerical procedure to construct the saddlepoint approximation for a real parameter in an exponential linear model. Jing *et al.* (1994) compared the saddlepoint approximation and the bootstrap approximation. Routledge & Tsao (1997) prove that the derivation of Lugannani and Rice's asymptotic expansion for the cumulative distribution function is the same as Daniels' asymptotic approximation for the corresponding density function. A comprehensive review of the application of the saddlepoint approximations can be found in Reid (1988, 1991) and Field and Ronchetti (1990).

3. Large-n Approximations

The large-n approximation, which is defined below, is one of the methods that are used to approximate the finite sample moments of the econometric estimators and test statistics for moderately large samples. The results derived from the large-n approximation, usually lie between the exact sample result and the large sample asymptotic result. One thing that needs to be clarified is the difference between the large-n approximation and the large sample asymptotic approximation. Inferences from the large sample asymptotic approximation are simply those based on the limiting distribution of the statistic in question when n goes to infinity. On the other hand, the large-n approximation uses an (asymptotic) expansion to approximate the exact distribution or moments of the statistic. Then it provides inferences based on some leading terms of the expansion. The accuracy of this approximation increases as the sample size increase. There are two broad ways of deriving a large-n approximation that have found popularity in econometrics. These are the Edgeworth approximation and what is often termed Nagar's approximation. An Edgeworth approximation is based on an expansion that gives us an approximation to the exact distribution function. On the other hand, Nagar's approximation is based on the expansion that gives us an approximation to the moments of the exact distribution. The large-n approximation we have used in this dissertation is Nagar's approximation, which was introduced first by Nagar (1959). The approximation is based on a Taylor series expansion to approximate the sampling error (the difference between the statistic and the parameter), so that the successive terms are in the descending order of the sample size n, in probability. Suppose that we have a random vector Y whose distribution involves an unknown parameter vector, θ, and let $\hat{\theta}$ be a consistent estimator of θ. Then the sampling error, $\hat{\theta}-\theta$, can be extended in the Taylor Series expansion:

$$\hat{\theta}-\theta = \varepsilon_{-1/2} + \varepsilon_{-1} + \cdots + \varepsilon_{-q/2} + \varepsilon^{*}$$

where $\varepsilon_{-j/2} = O_p(n^{-j/2})$. Then the bias to $O(n^{-1})$ is defined as:

$$Bias(\hat{\theta}) = E(\varepsilon_{-1/2}) + E(\varepsilon_{-1})$$

and the mean square error (MSE) to $O(n^{-2})$ is

$$MSE(\hat{\theta}) = E(\varepsilon_{-1/2}\varepsilon'_{-1/2} + \varepsilon_{-1/2}\varepsilon'_{-3/2} + \varepsilon_{-3/2}\varepsilon'_{-1/2} + \varepsilon_{-1}\varepsilon'_{-1})$$

There have been many applications of the Nagar large-n approximation in econometrics. For example, Srivastava and Giles (1987) applied the large-n approximation to determine the mean, bias and mean squared errors of various estimators for the seemingly unrelated regression model. Grubb and Symons (1987) used the large-n approximation for the OLS estimator of the lagged dependent variable coefficient in the context of a first order stable autoregressive model with exogenous variables (ARX(1)). Also, Kiviet and Phillips provided a series of extensions to Grubb and Symons's works. Kiviet and Phillips (1993, 1994) extended the analysis to the estimator of the full coefficient vector and the higher order dynamic regression models, ARX(p). All the works above are in the context of a stable dynamic model. Kiviet and Phillips (2005) extended the large-n approximation to examine the bias, variance and mean-squared error of the OLS estimator for the whole coefficient vector in a linear dynamic regression model with a unit root. Kiviet *et al.* (1995) also generalized this large-n approximation analysis to the dynamic seemingly unrelated regression model.

Other work has been done on the finite-sample properties of various nonlinear estimators, such as certain MLE's, which do not have an analytic expression in terms of the known variables. Rilstone *et al.* (1996) derived the second-order bias and MSE results for the nonlinear estimator with i.i.d observation data, when the estimator can be written as the solution to a set of moment equations. The application of this method is quite wide, including MLE, OLS, other extremum estimators and method of moment estimators, and of course some GMM estimators. Some other work involving the nonlinear estimator has been undertaken for certain nonlinear models, such as Knight and Satchell (1992) examined the duration models with exponential densities. Bao and Ullah (2006) generalized Rilstone *et* *al.*'s results for time-series models. These results are applicable in both the linear and nonlinear models and are valid for both the normal and nonnormal distribution. However, the form of the exact result depends on the model. Bao and Ullah derived the results for four models: AR(1), ARX(1), VAR(1) and MA(1). Among them, the results for ARX(1) and MA(1) models are also generalized to the non-normal observations.

4. Small-σ Approximations

Another method that is widely used in the finite-sample theory, especially in econometric analysis, is the so-called small-σ (or small disturbance) approximation. This method uses a Taylor series expansion to expand the expression for the sampling error, so that the successive terms are in descending order of σ in probability, in contrast to a large-n asymptotic expansion which orders these terms in descending order of the sample size, n, in probability. The small-σ asymptotic approach is a complement to the large-sample asymptotic approximation. The strength of the small-σ approximation is that it does not require a large sample size or the finiteness of the limiting values of the sample moments. Of course, the small-disturbance approximation also has some limitations. It requires that the disturbance be very small and approach zero in the limit, which implies that the econometric model can explain the variation of the dependent variable very well. In addition, the small-disturbance approximation usually requires knowledge of the underlying distribution, and the results are often quite complicated, and this can make it difficult to draw clear conclusions from the results.

Again, suppose that we have a random vector Y whose distribution involves an unknown parameter vector, θ, and let $\hat{\theta}$ be a consistent estimator of θ. For the small-σ approximation, the expansion of the sampling error in terms of the Taylor series expansion is

$$\hat{\theta}-\theta=\sigma\zeta_1+\sigma^2\zeta_2+\cdots+\sigma^q\zeta_q+\zeta^*$$

where σ is the standard deviation of the sample, y, and $\zeta^*=O(\sigma^{q+s}),\ s>0$.

Then the bias to $O(\sigma^2)$ is defined as:

$$Bias(\hat{\theta})=\sigma E(\zeta_1)+\sigma^2 E(\zeta_2)$$

and MSE to $O(\sigma^4)$ is

$$MSE(\hat{\theta})=\sigma^2 E\zeta_1\zeta_1'+\sigma^3[E\zeta_1\zeta_2'+E\zeta_2\zeta_1']+\sigma^4[E\zeta_2\zeta_2'+E\zeta_3\zeta_1'+E\zeta_1\zeta_3']$$

The small-σ approximation was first proposed by Kadane (1971) in the context of the so-called k-class estimators of the parameters of a single equation in a system of linear simultaneous stochastic equations with normal disturbances. Another early application of this approximation was by Raj and Ullah (1981, pp.134-140). They derived a small-σ approximation to the moment matrix of the Hildreth-Houck estimator for the purely random coefficient model. Ullah, Srivastava, and Roy (1995) generalized the small-σ approximation method to the non-normal but i.i.d case, and provided more explicit expressions for the bias to $O(\sigma^2)$ and MSE to $O(\sigma^4)$. Kiviet and Phillips (1993, 1994) applied the small-σ approximation in the context of ARX models. Other applications of the small-σ approximation have focused on approximating the exact moments of the ratio of quadratic forms in normal random vectors, *e.g.*, Magnus (1986) and Smith (1989). Also, Inder (1986) used the small-σ asymptotics to approximate the null distribution of the Durbin-Watson statistic when the model includes a lagged value of the dependent variable. Ullah and Srivastava (1994) provided more general results for the higher-order moments of the ratio of quadratic forms. Their results apply to the case of i.i.d non-normal distributions, with the normal distribution as a special case of this result. They also illustrate the application of the results to derive the bias of the R^2 for the multiple regression model, and the bias of the estimator in the ARX(1) model. More detailed discussions on the use of finite sample approximations, and particularly small-disturbance approximations, in econometrics can be found in Ullah (2004).

5. Dissertation Overview

In Chapter 2, we aim to solve the Purchasing Power Parity (PPP) puzzle that the persistence of the real exchange rate is too high to be reconciled with the PPP theory. The half-life is the measure which is widely used in this literature to measure the persistence of the real exchange rate. As is common in this literature, we use the OLS estimator of the coefficient in the autoregression (AR) model to estimate the half-life. In this chapter, we apply the saddlepoint approximations to the half-life estimator. We approximate the density and CDF function of the half-life estimator, based on Lieberman's (1994b) saddlepoint approximation for the coefficient in AR(1) models. Based on the density function, we generate numerical results on the point estimator and confidence interval for the half-life estimator. The numerical results show that our point estimator of the half-life estimator is very high and the confidence interval is very wide, which is consistent with most of the empirical literature on the PPP puzzle.

Another contribution of the chapter is that we derive some properties of the half-life estimator based on the density function we derived. We prove that the finite moments of the half-life estimator do not exist if we construct the half-life estimator based on the coefficient in AR model. This property provides an explanation for why most of the empirical works find a wide confidence interval for the half life. This suggests that the PPP puzzle may not really exist. The possible reason for the PPP puzzle is that we use an inappropriate measure for the half-life. Therefore, choosing an appropriate measure of the half-life estimator is the way to solve the PPP puzzle.

In Chapter 3, we apply the different saddlepoint approximations to approximate the CDF for the Anderson-Darling test. The objective of this chapter is to compare the performances of different saddlepoint approximations. The saddlepoint approximations we compare in this chapter include the lower-order and higher-order-normal-based saddlepoint approximations, and the lower-order and higher-order-non-normal-based saddlepoint approximations. The first two approximations derived by Daniels (1954, 1987), provide very accurate approximation when the asymptotic distribution of the statistics is normal. And the third one is from Wood, Booth and Butler (1993) who expect the non-normal-based saddlepoint approximations can provide some improvement over the normal-based saddlepoint approximation when the asymptotic distribution of the statistics is non-normal. The asymptotic CGF of the Anderson-Darling test is a weighted sum of the chi-square distribution, therefore, we expect the non-normal based saddlepoint approximation can provide some improvement over the normal-based saddlepoint approximation. Then we extend the non-normal-based saddlepoint approximations to include the higher-order terms, with the expectation that such inclusion may bring some improvement.

The comparison of the performances of the different saddlepoint approximation is based on the results from Lewis (1961). Lewis simulated some numerical results for the Anderson-Darling test with the sample size ranging from 1 to infinity. The comparison shows that both the non-normal-based and the higher-order saddlepoint approximations can provide some improvement in the central part, but not in the tail areas of the density for the Anderson-Darling test. This means that the lower-order-normal-based saddlepoint approximation is enough if we are only interested in the tail areas.

In Chapter 4, we apply the large-n approximation to the maximum likelihood estimator (MLE) of the Logit model. Based on Rilstone, Srivastava and Ullah's results (1996) for the nonlinear estimator, we derive the bias and MSE of MLE in the Logit model. There is an extensive literature on the asymptotic properties of MLE in the Qualitative Response model. However, the papers on the finite sample properties of MLE in the Qualitative Response models are surprisingly few. So our work makes some contributions to this field. After we derive the analytic expressions for the bias and MSE of MLE in the Logit model, we generate some numerical results based on the analytic results for the Logit model with one regressor. We compare the MLE with its bias-corrected counterparts. We derive two bias-corrected estimators, one based on the true parameter, and another based on an estimator of the parameter. The results show that there is a gain from bias-correction.

In Chapter 5, we apply the small-σ approximation to two price index numbers, which are the Laspeyres Index and the Paasche Index. A price index is used to measure the price change over time or across regions. We derive the bias and MSE of the two index numbers based on Ullah and Srivastava's (1994) small-σ approximation for the bias and MSE of an estimator which can be expressed as a ratio of quadratic forms. We try to decide on the preference between these two index numbers based on their MSE. In order to write the index number as a ratio of quadratic forms, we assume all the price and quantity in the base and current periods are random. Then we derive the analytic function for the bias and MSE of these two index numbers. Some numerical results show that it is hard to choose between these two index numbers. The preference depends on the data set. Different data sets give us different preferences between these two index numbers.

In Chapter 6 of this dissertation, we provide our conclusions from this research, and we also suggest some future directions for further research on each topic that is covered in this dissertation.

CHAPTER 2:

A SADDLEPOINT APPROXIMATION TO THE EXACT DISTRIBUTION OF THE HALF-LIFE ESTIMATOR IN AN AUTOREGRESSIVE MODEL: NEW INSIGHTS INTO THE PPP PUZZLE

1. Introduction

Purchasing power parity (PPP) is a theory about exchange rates between two currencies. Basically, it means that the price for a given basket of services and goods should be the same in two countries, if measured in the same currency. PPP is a building block in international economics. Therefore, the validity of PPP has attracted considerable interest, especially since the advent of flexible exchange rates in the early 1970's. As the real exchange rate is the nominal exchange rate adjusted for the relative price level, the tradition in the literature is to use the real exchange rate to explore PPP theory.

Essentially, there are two empirical puzzles associated with PPP. The first puzzle is the non-stationary behavior of the real exchange rate. PPP theory can be simply re-stated as saying that the real exchange rate is mean reverting. Therefore, the non-stationary behaviour of the real exchange rate implies that any deviation from PPP cannot disappear as time goes. Although few economists view PPP as a short-term phenomenon, non-stationarity implies that the PPP theory does not even hold in the very long run. This poses a problem for international economics because of the important role that PPP plays. The second empirical puzzle arises as follows: The observed degree of persistence in real exchange rates is too high to be reconciled in terms of their short-term volatility. Financial factors, such as monetary or financial shocks, cause the volatility of the exchange rate, and in the presence of price stickiness the effect of such shocks can be exaggerated. However, the high persistence of the deviations from PPP that have been observed in a vast range of empirical studies cannot be explained simply by price stickiness. In the empirical literature relating to PPP, the half-life is a commonly-used measure of the persistence of the deviation from PPP. This is defined as the amount of time it takes for a unit shock to dissipate by 50%. Empirical studies appear to yield a consensus of a half-life of three to five years (*e.g.*, Abuaf and Jorion, 1990; Glen 1992; Cheung and Lai, 1994). This empirical consensus of a half-life of three to five years is well discussed by Rogoff (1996). Rogoff first constructed the phrase "purchasing power parity puzzle" to describe the problem that the high persistence of the real exchange rate is inconsistent with the PPP theory, even considering the stickiness. Only a half-life of no more than one to two years can be explained by price stickiness alone. So, there remains a puzzle that continues to attract considerable attention in the literature, and this provides the motivation for this chapter.

In addition, a number of authors have obtained not just point estimates of the half-life of deviations from PPP, but have also attempted to provide confidence intervals. In this chapter we show that such attempts are meaningless as the usual half-life estimator has no finite moments. This finding also explains why disparate point estimates are obtained in practice. The chapter is constructed as follows. The next section reviews various PPP convergence studies that focus on half-life measures. As the main theoretical results of the this paper are based on the Lieberman's (1994b) saddlepoint approximation for the density and distribution function of the OLS estimator of the AR(1) model, in section 3 we provide, a simple introduction to this approximation. In section 4 we derive the density and distribution functions for the half-life estimator in the AR(1) model, and explore some of the properties of this estimator. These results are extended to the case of the AR(p) model in section 5; and the final section discusses the implications of our results and provides suggestions for future research.

2. Resolutions of the PPP Puzzle

Given the important role of PPP theory it is natural that economists have shown a lot of interest in solving the two puzzles described in the first section. The empirical results relating to the first of these puzzles are mixed. Using standard unit-root tests, most of the studies cannot reject the unit root hypothesis versus the stationary hypothesis for real exchange rates under a floating exchange rate regime (*e.g.*, Meese and Rogoff, 1988; Edison and Fisher, 1991; Grilli and Kaminsky, 1991). In particular, Edison and Fisher could not reject the presence of a unit root in the data even when they allowed for structural breaks. A potential explanation for this is the low power of the test in relatively small samples, so later researches have mainly focused on two directions: the use of long-term historical data (*e.g.*, Diebold *et al.,* 1991; Lothian and Taylor, 1996) and the application of more powerful tests. The studies using long-term data show that if the data set covers both fixed and floating rate periods, then the real exchange rate exhibits mean-reverting behavior. However, the results cannot determine whether or not the mean-reverting behaviour of the real exchange rate still exists in the post-Bretton Woods period. Thus, several more powerful unit root tests have been applied to obtain inferences relating to the post-1973 data set.

Some economists have applied panel unit root tests and have found strong evidence in support of rejecting the unit root (*e.g.*, Lothian, 1997; Wu, 1996; Papell and Theodoridis, 1998; Pedroni, 2004). Taylor and Sarno (1998) constructed two types of powerful unit root tests. One is the multivariate augmented Dickey-Fuller test or MADF test, which is a multivariate analogue of the standard, single-equation augmented Dickey-Fuller (ADF). Like the usual panel unit root test, the null hypothesis of the MADF test is that all of the series in the panel data-set are non-stationary, which may result in an over-rejection problem. Another approach is to apply Johansen's (1988) likelihood ratio (JLR) test statistic to test the null hypothesis, that is, at least one of the series have a unit root. So the JLR test can overcome the over-rejection problem. Taylor, *et al.* (2001) applied the MADF and JLR tests and their results support the stationary property of the real exchange rate, at least at the five percent significance level. However, some economists have argued that the panel studies are sensitive to the choice of sample (*e.g.*, Papell, 1997). Chuang and Lai (1998) uncover the first puzzle based on the post-Bretton Woods data by applying two efficient univariate unit root tests: a modification of the ADF test constructed from generalized least squares (GLS) estimation – *i.e.*, the DF-GLS test suggested by Elliot *et al.* (1996); and a modification of the ADF test constructed from weighted symmetric least squares (WSLS) estimation - the DF-WS test suggested by Park and Fuller (1995). Therefore, by covering longer- term data or exploring more powerful tests, the first puzzle seems to be solved, which means that PPP theory holds, at least in long run.

Accordingly, more recent papers in this field focus on the second PPP puzzle by exploring the possible reasons of over-estimating the persistence of the real exchange rate. From an empirical viewpoint, the second puzzle can be illustrated for the linear AR(1) model as follows.

Let

$$y_t = \alpha y_{t-1} + u_t \; ; \quad t = 0, 1, 2, \ldots, T \tag{1}$$

where y_t is the real exchange rate series, which is conditional on the initial value, y_0, and $u_t \sim i.i.d.N(0, \sigma^2)$. Often, the normality assumption is not explicitly used in the associated empirical literature, as it is not needed for the construction of a half-life measure. We initially retain the normality assumption to establish our main results, but subsequently we show that these results still hold if this assumption is relaxed to some degree.

Based on the model in (1), an estimator of the half-life for the speed of adjustment can be obtained as:

$$\hat{h} = \log(0.5) / \log(\hat{\alpha}) \quad , \tag{2}$$

where $\hat{\alpha}$ is the OLS estimator of α, namely $\hat{\alpha} = (y'_{-1} y_{-1})^{-1} y'_{-1} y$, and we require $\hat{\alpha} \in (0\,,1)$ for the model to be dynamically stable, and for the estimated half-life to be positive.

We know that the second puzzle involves observing an estimated half-life that is "too large", as it is difficult for this to be reconciled with the PPP theory. Only a half-life of less than two years can be explained by the stickiness of prices. Therefore, the second puzzle arises if the estimator $\hat{h}$ yields a value greater than (about) three years, say. In order to provide a more complete picture of the puzzle, most studies also report a confidence interval to supplement the point estimate of the half-life. However, these confidence intervals are usually found to be far too wide in practice to provide any useful information about the PPP puzzle.

The first possible reason for these uninformative confidence intervals that has been put forward is the bias of the OLS estimator in (1). As we know, the OLS estimator in the AR(p) model is biased downward in small samples, and the bias tends to increase with the persistence of the series. Andrews' (1993) median-unbiased estimator for AR(p) models provides a good tool to correct the bias. Unfortunately, though, the studies applying the median-unbiased estimator do not find any support for the PPP theory (*e.g.*, Murray and Papell, 2002; Cashin and McDermott, 2003; Caporale *et al.*, 2005; Lopez, 2005). Actually, the results based on the median-unbiased estimators are that the estimated half-life is higher than with the OLS estimator, and the confidence interval is still so wide that no strong conclusions can be made about the PPP puzzle. Murray and Papell (2005) extended the median-unbiased estimation method to the panel data context, and argued that the shorter half-life of 2-2.5 year based on estimators unadjusted by the median-unbiased estimator from the previous panel data are the results of the implication of inappropriate estimation method. Murray and Papell's results are consistent with Rogoff's PPP puzzle claim. Choi *et al.* (2004) address the bias sources in estimating the half-life of PPP from panel data and found a 5.5 year of half life for 21 OECD countries from 1948-2002. In all, the bias correction seems to drive us away from PPP theory.

Some other economists have tried to find the solution to the puzzle in the actual estimation model. We know that PPP theory is based on the assumption of a complete trade arbitrage world, which means that any deviation from PPP will be eliminated by the trade arbitrage. But in reality, complete trade arbitrage does not exist due to factors such as transaction costs, taxation, or some trade restrictions between two countries. Only when the deviation is large enough to cover the transaction costs will arbitrage occur and drive away the deviation. Therefore, some economists argue that in the presence of transaction costs, a nonlinear model is more reasonable to represent the process of the real exchange rate (*e.g.*, Taylor, Peel and Sarno, 2001; Baum *et al.*, 2001). In the nonlinear models, the mean reversion speed depends on the deviation size from the long-run equilibrium level: the larger are the deviations, the lower are the half-life point estimates and the narrower are the confidence intervals, and *vice versa*. So it is possible to have high persistence as well as low persistence. It seems that the nonlinear models provide a hopeful solution to the PPP puzzle. However, El-Gamal and Ryu (2006) investigate both the autoregression model and nonlinear models such as the Threshold Autoregression (TAR) and Exponential Smooth Threshold Autoregression (ESTAR). They found that the nonlinear models exhibit the same classical decay form of the AR model. As the AR model can capture most of the features of the nonlinear models, their conclusion is that it is unnecessary to introduce complexity by using nonlinear models. Therefore, even if all the efforts provide valuable insights into the puzzle, the puzzle is still there.

A later line of literature starts to question the measure of half-life that is used in this literature. Chortareas and Kapetanios (2004) suggest that the second puzzle may be caused artificially by the measure of half-life that is adopted. They suggest an alternative measure, which can reduce the half-life estimate to a level consistent with the predictions of the sticky price models for the AR(p) model. However, the alternative measure is the same as the traditional one in the special case of the AR(1) model. Therefore, this alternative measure of the half-life can not solve the PPP puzzle for the AR(1) model.

This chapter takes a different position from previous studies that have tried to resolve the PPP puzzle. Essentially, this chapter has two contributions: first, we provide the density and distribution functions for the usual half-life estimator. All the studies in this field up to now construct the density function and confidence intervals on the basis of Monte Carlo or bootstrap simulation methods. No specific density or distribution function of half-life has been given, except that Kilian and Zha (1999) provide the density function for the half-life based on a Bayesian method. However, our method is much easier to calculate and it also overcomes the problem of prior information. Second, based on the density function, this chapter proves that the moments of the half-life estimator do not exist, and we also extend the results to the general AR(p) model. This provides an explanation for the wide confidence intervals in all of the empirical studies, and it also implies that the second puzzle may arise due to the invalid measure of the half-life, as is suggested by Chortareas and Kapetanios (2004).

3. Saddlepoint Approximations for the Distribution

As we can see from (2), the basic definition of the half-life is based on the OLS estimator of the coefficient in an AR(1) model. If we know the density function of $\hat{\alpha}$, then we can easily construct the density function of the half-life. Fortunately, there exist various studies that explore the properties of $\hat{\alpha}$ in (1) (*e.g.*, Philips, 1978; Lieberman, 1994a, 1994b). This section uses Lieberman's results to explore the properties of the half-life estimator. Lieberman (1994b) implemented a saddlepoint approximation to obtain the density and distribution functions for the coefficient estimator in the AR(1) model. Since the seminal paper by Daniels (1954), many applications have shown that the saddlepoint approximation can provide a valuable tool in approximating the density and distribution functions. Lieberman's results strengthen this conclusion. He illustrated that the saddlepoint approximation of the density and distribution of $\hat{\alpha}$ is excellent over the whole interval of $\hat{\alpha}$ even for very small samples.

For equation (1), Lieberman expresses the OLS estimator $\hat{\alpha}$ as:

$$\hat{\alpha} = \frac{v'R'_\alpha C_1 R_\alpha v}{v'R'_\alpha C_2 R_\alpha v}, \qquad v \sim N(0, \sigma^2 I), \tag{3}$$

$$\text{where}\quad C_1 = \begin{bmatrix} 0 & \frac{1}{2} & \dots & 0 & 0 \\ \frac{1}{2} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \frac{1}{2} \\ 0 & 0 & \dots & \frac{1}{2} & 0 \end{bmatrix}_{(T+1)x(T+1)}, \quad C_2 = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}_{(T+1)x(T+1)},$$

$$\text{and}\ R_\alpha = \begin{bmatrix} \mathrm{b} & 0\dots & & & 0 \\ \alpha\mathrm{b} & 1 & 0\dots & & 0 \\ \alpha^2 b & \alpha & 1 & 0\dots & 0 \\ \vdots & \vdots & & & \vdots \\ \alpha^T b & \alpha^{T-1} & \cdots & \alpha & 1 \end{bmatrix}_{(T+1)x(T+1)}, b = \begin{cases} (1-\alpha^2)^{-\frac{1}{2}} & ;\ \text{if}\ \alpha \in (-1,1) \\ 0 & ;\ \text{otherwise,} \end{cases}$$

Then Lieberman derived the saddlepoint approximation for the density of $\hat{\alpha}$ as

$$\hat{f}(\hat{\alpha}) = \frac{\{tr(\hat{A}^{-1}R'_\alpha C_2 R_\alpha)\}\left|\hat{A}\right|^{-\frac{1}{2}}}{\left[4\pi tr\{(\hat{A}^{-1}D)^2\}\right]^{\frac{1}{2}}} \tag{4}$$

where $D = D(\hat{\alpha}) = R'_\alpha(C_1 - \hat{\alpha}C_2)R_\alpha$, $\hat{A} = A(\hat{w}) = I - 2\hat{w}D$ and $\hat{w}$ satisfies

$$tr(\hat{A}^{-1}D) = 0 \quad . \tag{5}$$

Then he generated the distribution function of $\hat{\alpha}$ by integrating the density function and applying the Lugannani-Rice formula, which is:

$$\hat{F}(\hat{\alpha}) = P(\hat{\alpha} < x) = \Phi(\hat{\varepsilon}) + \phi(\hat{\varepsilon})(\frac{1}{\hat{z}} - \frac{1}{\hat{\varepsilon}}) \quad , \tag{6}$$

where $\hat{\varepsilon} = \left(\log\left|\hat{A}\right|\right)^{\frac{1}{2}}\mathrm{sgn}(\hat{w})$, $\hat{z} = \hat{w}\left[2tr\{(\hat{A}^{-1}D)^2\}\right]^{\frac{1}{2}}$, $D = D(x)$; Φ and ϕ are the standard normal distribution and density functions respectively, and $\hat{w}$ is defined by (5).

Lieberman compared the approximation of the distribution with the exact values obtained using Davies' (1973) algorithm for the CDF of a weighted sum of independent chi-square variates, as coded into the DAVIES option of the DISTRIB command in the SHAZAM econometrics package (Whistler *et al.*, 2001) for different sample sizes and values of α . His comparison showed that the saddlepoint approximation is excellent over the whole interval of $\hat{\alpha}$ (Lieberman, 1994a, Table 1).

4. Properties of the Half-Life Estimator in the AR(1) Model

4.1 Density and distribution functions of the half-life estimator

If the real exchange rate follows an AR(1) process, then the half-life estimator is defined by (2). Taking the first puzzle of PPP to be resolved, we can ignore the unit root case and apply Lieberman's result. Then, the saddlepoint approximation for the density of the half-life estimator can be constructed based on the density of $\hat{\alpha}$ using the transformation:

$$f(\hat{h}) = f[\hat{\alpha}(\hat{h})|0 < \hat{\alpha} < 1]J$$

$$= \frac{f(\hat{\alpha}(\hat{h}))J}{\Pr.(0 < \hat{\alpha} < 1)}, \tag{7}$$

where $f(\hat{\alpha}(\hat{h}))$ is the density function obtained by replacing $\hat{\alpha}$ with $(0.5)^{1/\hat{h}}$ in (4); and

$$J = \frac{(0.5)^{1/\hat{h}} \ln 2}{\hat{h}^2},$$

is the Jacobian of the transformation.

$\Pr.(0 < \hat{\alpha} < 1) = \Pr.(\hat{\alpha} < 1) - \Pr.(\hat{\alpha} < 0)$ can be calculated from (6) by letting $x = 0$ and $x = 1$. We let $C = \Pr.(0 < \hat{\alpha} < 1)$, which is a constant number.

So, the approximation to the density function for the half-life estimator is:

$$\hat{f}(\hat{h}) = \frac{\{tr(\widetilde{A}^{-1}R'_\alpha C_2 R_\alpha)\}|\widetilde{A}|^{-1/2}}{[4\pi\, tr\{(\widetilde{A}^{-1}\widetilde{D})^2\}]^{1/2}} \frac{(0.5)^{1/\hat{h}} \ln 2}{C\hat{h}^2} \tag{8}$$

where $\widetilde{D} = D(\hat{h}) = R'_\alpha(C_1 - (0.5)^{1/\hat{h}} C_2)R_\alpha$, $\widetilde{A} = A(\hat{w}) = I - 2\hat{w}\widetilde{D}$ and $\hat{w}$ satisfies

$$tr(\widetilde{A}^{-1}\widetilde{D}) = 0 \quad .$$

Similarly, the approximation to the distribution function of the half-life estimator is:

$$\hat{F}(\hat{h}) = P(\hat{h} < x|0 < \hat{\alpha} < 1)$$

$$= P(\frac{\log(0.5)}{\log(\hat{\alpha})} < x) \Big/ P(0 < \hat{\alpha} < 1)$$

$$= P(\hat{\alpha} < (0.5)^{1/x})/C \quad . \tag{9}$$

Again, the distribution function (9) can be calculated easily, using equation (6).

Based on (8), we generate the numerical values for the density for different choices of α and different sample sizes. We provide some figures to compare the density of $\hat{h}$ for different values of α with the same sample size, and for different sample size with the same value of α. More specifically, Figure 1 shows the density function for α equal to 0.8, 0.9 and 0.97 and sample sizes of 30. Figure 2 shows the density of $\hat{h}$ for sample sizes of 10, 30, and 50 and α equal to 0.95. From the first figure, we can see that the density function is highly skewed to the right, and the density moves to the right and the tails become fatter as α increases. From figure 2, we can see that the location of the density also moves to the right and the tails become fatter as the sample size increases.

Table 1:
Point Estimator and Confidence Intervals of the Half-Life for Different α Values and Sample Sizes

α	T = 10		T = 30	
	Point estimator (Median)	95% Confidence Interval	Point estimator (Median)	95% Confidence Interval
0.6	1.20	[0.27, 7.12]	1.28	[0.47, 3.23]
0.7	1.58	[0.32, 12.15]	1.78	[0.63, 5.02]
0.8	2.19	[0.39, 24.49]	2.72	[0.87, 9.56]
0.9	3.39	[0.49, 60.04]	5.06	[1.27, 33.64]
0.95	4.79	[0.61, 109.52]	8.26	[1.65, 108.48]
0.97	6.07	[0.69, 155.41]	10.95	[1.88, 197.73]

Note: both the median point estimates and the confidence intervals are calculated based on the distribution function (9) using code written for the SHAZAM econometrics package.

Table 1 shows the (median) point estimate and 95% confidence interval of the half-life estimator when the true data process is an AR(1) model. Based on this table, we can see: first, the point estimate increases with the sample size, which is consistent with the figures; second, when α is higher than 0.9, which is almost always the case in the empirical studies in this field, the confidence interval is very wide. Third, when the sample size increases, the confidence interval becomes somewhat tighter, but it is still quite wide, except for the case $\alpha = 0.97$. The actual meaning of the half-life estimate depends on the frequency of the real data. For yearly data, the results are obviously inconsistent with the PPP theory. However, for quarterly data, the sample size is usually over 30 and the autocorrelation coefficient is usually quite high, so we can expect that the PPP puzzle is still there. Therefore our results are consistent with most of the other related empirical work.

4.2 Moments of the half-life estimator

***Theorem 1*:**

Let the data follow a stationary AR(1) process: $y_t = \alpha y_{t-1} + u_t$, with $u_t \sim N(0,\sigma^2)$ and $\alpha \in (-1,1)$. The half-life estimator is defined as $\hat{h} = \log(0.5)/\log(\hat{\alpha})$, where $\hat{\alpha}$ is the least squares estimator of α and suppose $\hat{\alpha} \in (0,1)$. Then the mean of the half-life estimator does not exist.

***Proof*:**

$$M(\hat{h}) = \int_0^\infty \hat{h}\, f(\hat{h})\, d\hat{h},$$

$$= \int_0^1 \frac{\log(0.5)}{\log(\hat{\alpha})} f(\hat{\alpha})\, d\hat{\alpha}$$

Let $u(\hat{\alpha}) = \left\{tr(\hat{A}^{-1}R_\alpha' C_2 R_\alpha)\right\}\left|\hat{A}\right|^{-\frac{1}{2}}$ and $v(\hat{\alpha}) = \left[4\pi\, tr\{(\hat{A}^{-1}D)^2\}\right]^{\frac{1}{2}}$

$$\hat{M}(\hat{h}) = \int_0^1 \frac{\log(0.5)}{\log(\hat{\alpha})} \frac{u(\hat{\alpha})}{v(\hat{\alpha})} d\hat{\alpha}$$

$$= \underset{\varepsilon \to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{\log(0.5)}{\log(\hat{\alpha})} \frac{u(\hat{\alpha})}{v(\hat{\alpha})} d\hat{\alpha}.$$

Since the whole interval of $\hat{\alpha}$ is $(-\infty, \infty)$, $u(\hat{\alpha})$ and $v(\hat{\alpha})$ are continuous functions of $\hat{\alpha}$ on the closed interval $[0, 1]$. According to the extreme value theorem, we can assume:

(i) when $\hat{\alpha} = \overline{\alpha}$, $u(\hat{\alpha})$ gets to its minimum value N and $N \neq 0$.

(ii) when $\hat{\alpha} = \breve{\alpha}$, $v(\hat{\alpha})$ gets to its maximum value M and $M \neq \infty$.

(The justification for assumptions (i) and (ii) is given in the Appendix.)

Given that $f(\hat{\alpha}) \geq \delta$ for some $\delta > 0$ in (0, 1), then:

$$\hat{M}(\hat{h}) > \underset{\varepsilon \to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{\log(0.5)}{\log(\hat{\alpha})} \frac{N}{M} d\hat{\alpha}$$

$$= \log(0.5)\frac{N}{M} \underset{\varepsilon \to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{1}{\log(\hat{\alpha})} d\hat{\alpha}$$

$$= \log(0.5)\frac{N}{M}\left[\underset{\varepsilon \to 0}{Lim}\left(\frac{\hat{\alpha}}{\log(\hat{\alpha})}\Bigg|_\varepsilon^{1-\varepsilon}\right) + \underset{\varepsilon \to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{1}{[\log(\hat{\alpha})]^2} d\hat{\alpha}\right]$$

$$= \log(0.5)\frac{N}{M}\left[\infty - \underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{[\log(\hat{\alpha})]^2}\,d\hat{\alpha}\right].$$

So, the estimated mean of the half-life estimator does not exist.

Based on the inversion formula, we know that

$$f(\hat{\alpha}) = \hat{f}(\hat{\alpha})(1+\ldots) \quad .$$

Therefore if the estimated mean $\hat{M}\ (\hat{h}\ )$ based on the saddlepoint approximation does not exist, then the true mean $M(\hat{h})$ does not exist, either. We know that one or both of $\overline{\alpha}$ and $\breve{\alpha}$ may possibly take values on the boundary of the [0 , 1] interval. In this case, we can set their value(s) to 1-ε or ε (ε⟶0) appropriately. Then we take the limit, and the proof still holds.

So the theorem is proved.

***Corollary 1*:**

Let the data follow a stationary AR(1) process: $y_t = \alpha y_{t-1} + u_t$, with $u_t \sim N(0,\sigma^2)$ and . The half-life estimator is defined as $\hat{h} = \log(0.5)/\log(\hat{\alpha})$, where $\hat{\alpha}$ is the least squares estimator and $\hat{\alpha} \in (0,1)$. Then the integer-order moments of the half-life estimator do not exist.

***Proof*:**

$$\mathrm{M}(\hat{h}^r) = \int_0^{\infty}\hat{h}^r\, f(\hat{h})\,d\hat{h}$$

$$= \int_0^1\left[\frac{\log(0.5)}{\log(\hat{\alpha})}\right]^r f(\hat{\alpha})\,d\hat{\alpha}$$

$$\hat{M}\ (\hat{h}) = \underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\left[\frac{\log(0.5)}{\log(\hat{\alpha})}\right]^r \frac{u(\hat{\alpha})}{v(\hat{\alpha})}\,d\hat{\alpha}$$

$$> \underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\left[\frac{\log(0.5)}{\log(\hat{\alpha})}\right]^r \frac{N}{M}\,d\hat{\alpha}$$

where M and N are defined same as in (i) and (ii) above.

Given that $f(\hat{\alpha}) \geq \delta$ for some $\delta > 0$ in (0, 1), then

$$\hat{M}\ (\hat{h}) > [\log(0.5)]^r\frac{N}{M}\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{[\log(\hat{\alpha})]^r}\,d\hat{\alpha}$$

$$= [\log(0.5)]^r \frac{N}{M}\left[\underset{\varepsilon\to 0}{Lim}\left(\frac{\hat{\alpha}}{[\log(\hat{\alpha})]^r}\Bigg|_{\varepsilon}^{1-\varepsilon}\right) + r\,\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{[\log(\hat{\alpha})]^{r+1}}d\hat{\alpha}\right]$$

$$= [\log(0.5)]^r \frac{N}{M}\left[\infty + r\,\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{[\log(\hat{\alpha})]^{r+1}}d\hat{\alpha}\right]$$

which establishes our result.

5. Properties of the Half-Life Estimator in the AR(p) Model

Often, PPP studies are based on the more general AR(p) model, to take account of more general features of the data. So, it is of interest to see if our results also hold in the AR(p) model, even though this situation is considerably more complicated. In order to make the problem workable, we make some reasonable simplifying assumptions. First, we need to know the formula used to estimate the half-life in the case of the AR(p) model. Essentially there are two ways that are used to estimate the half-life for the AR(p) model in this literature. First, some studies use the impulse response function to estimate the half-life by using some nonparametric method, such as the bootstrap or Monte Carlo method. Second, other studies estimate the half-life based on the formula constructed from the coefficient estimator from an augmented Dickey-Fuller (ADF) regression equation. The empirical work based on both of these methods has found similar results for the PPP puzzle, namely an implausibly large half-life estimate and a very wide confidence interval. In order to derive a specific density and distribution functions for half life, this chapter uses the second method. The basic idea is as follows.

If the real exchange rate y_t follows an autoregressive process of order p, AR(p), then:

$$y_t = \sum_{i=1}^{p}\alpha_i y_{t-i} + u_t \quad . \tag{10}$$

There is no explicit half-life function for the AR(p) model based on the estimator of the coefficients in (10). The formula often used in practice involves approximating the half-life by estimating an ADF equation:

$$\Delta y_t = \beta y_{t-1} + \sum_{i=1}^{p-1}\phi_i \Delta y_{t-i} + u_t, \qquad u_t \sim i.i.d.N(0,\sigma^2) \quad . \tag{11}$$

We suppose that the data are stationary, so that $\beta \in (-1,1)$. Then, based on (11), we estimate the half-life using:

$$\hat{h} = \log(0.5)/\log(1+\hat{\beta}), \qquad 1+\hat{\beta} \in (0,1) \quad . \tag{12}$$

In order to express the OLS estimator $\hat{\beta}$ simply, we first apply some transformations to the data. Let,

$$R_1 = M'\Delta y_t \qquad \text{and} \qquad R_2 = M'y_{t-1}$$

where $M = I - Y(Y'Y)^{-1}Y'$, and $Y = (\Delta y_{t-1}\ \Delta y_{t-2}\ \Delta y_{t-3} \cdots \Delta y_{t-p+1})$, and we are implicitly conditioning on the p initial observations.

Using standard partitioning results,

$$\hat{\beta} = (R_2' R_2)^{-1} R_2' R_1 \ ,$$

which can be rewritten as:

$$\hat{\beta} = \frac{R'FR}{R'GR}, \tag{13}$$

where

$$F = \begin{bmatrix} 0 & 0\ldots & & 0 & 0 & \frac{1}{2} & 0 & 0\cdots & 0 \\ 0 & 0\ldots & & 0 & 0 & 0 & \frac{1}{2} & 0\ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & & & \vdots \\ 0 & 0\ldots & & 0 & 0\ldots & & & & \frac{1}{2} \\ \frac{1}{2} & 0\ldots & & 0 & 0\ldots & & & & 0 \\ 0 & \frac{1}{2} & 0\ldots & 0 & 0\ldots & & & & 0 \\ \vdots & \vdots & & \vdots & \vdots & & & & \vdots \\ 0 & 0\ldots & & 0 & \frac{1}{2} & 0\ldots & & & 0 \end{bmatrix}_{2T\times 2T}$$

$$G = \begin{bmatrix} 1 & 0\ldots & & 0\ldots & & 0 \\ 0 & 1 & 0\ldots & 0\ldots & & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0\cdots & & 1 & 0\ldots & 0 \\ 0 & 0\ldots & & & & 0 \\ 0 & 0\ldots & & & & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0\ldots & & & & 0 \end{bmatrix}_{2T\times 2T} \qquad \text{and} \quad R = \begin{bmatrix} R_2 \\ R_1 \end{bmatrix} .$$

Now, we define the covariance matrix of R to be $\Omega_{2T\times 2T}$ and $\Omega^{-1} = P'P$.

Equation (13) can be written as:

$$\hat{\beta} = \frac{v'P'FPv}{v'P'GPv}, \qquad v \sim N(0, \sigma^2 I). \tag{14}$$

We can see that equations (14) and (3) are quite similar, so based on Lieberman's method, we can derive the density function of $\hat{\beta}$ as,

$$\hat{f}(\hat{\beta}) = \frac{\{tr(\hat{N}^{-1}P'GP)\}|\hat{N}|^{-\frac{1}{2}}}{[4\pi\, tr\{(\hat{N}^{-1}L)^2\}]^{\frac{1}{2}}}, \tag{15}$$

where $L = L(\hat{\beta}) = P'(F - \hat{\beta}G)P$, $\hat{N} = N(\hat{w}) = I - 2\hat{w}L$ and $\hat{w}$ satisfies

$$tr(\hat{N}^{-1}L) = 0 \quad . \tag{16}$$

Then, applying Lugannani-Rice formula, the distribution function of $\hat{\beta}$ is

$$\hat{F}(\hat{\beta}) = P(\hat{\beta} < x) = \Phi(\hat{\varepsilon}) + \phi(\hat{\varepsilon})(\frac{1}{\hat{z}} - \frac{1}{\hat{\varepsilon}}) \quad , \tag{17}$$

where $\hat{\varepsilon} = \left(\log|\hat{N}|\right)^{\frac{1}{2}} \operatorname{sgn}(\hat{w})$, $\hat{z} = \hat{w}[2tr\{(\hat{N}^{-1}L)^2\}]^{\frac{1}{2}}$,

$L = L(x)$;as before, Φ and ϕ are the standard normal distribution and density functions respectively, and $\hat{w}$ is defined by (16).

Let $\tilde{\alpha} = 1 + \hat{\beta}$, and using the fact that the Jacobian is unity, the density function of $\tilde{\alpha}$ is :

$$\hat{f}(\tilde{\alpha}) = \frac{\{tr(\tilde{N}^{-1}P'GP)\}|\tilde{N}|^{-\frac{1}{2}}}{[4\pi\, tr\{(\tilde{N}^{-1}\tilde{L})^2\}]^{\frac{1}{2}}} \quad , \tag{18}$$

where $\tilde{L} = L(\hat{\alpha}) = P'(F - (\tilde{\alpha} - 1)G)P$, $\tilde{N} = N(\hat{w}) = I - 2\hat{w}\tilde{L}$ and $\hat{w}$ satisfies

$$tr(\tilde{N}^{-1}\tilde{L}) = 0 \quad .$$

Based on the definition function (12), and allowing for the Jacobian, the density function of the half-life estimator in the AR(p) model is:

$$f(\hat{h}) = f(\tilde{\alpha}(\hat{h}) | 0 < \tilde{\alpha} < 1)J$$

$$= \frac{f(\tilde{\alpha}(\hat{h}))J}{P(-1 < \hat{\beta} < 0)} \quad , \tag{19}$$

where $f(\tilde{\alpha}(\hat{h}))$ is the density function by replacing $\hat{\alpha}$ with $(0.5)^{1/\hat{h}}$ in (18); and the Jacobian is

$$J = \frac{(0.5)^{1/\hat{h}} \ln 2}{\hat{h}^2}.$$

$\Pr.(-1 < \hat{\beta} < 0) = \Pr.(\hat{\beta} < 0) - P(\hat{\beta} < -1)$ can be calculated from (17) by letting $x = 0$ and $x = 1$. We let $K = \Pr.(-1 < \hat{\beta} < 0)$, which is a constant number.

So, the density function for the half-life estimator is:

$$f(\hat{h}) = \frac{\{tr(\overline{N}^{-1} P'GP)\}\left|\overline{N}\right|^{-\frac{1}{2}}}{\left[4\pi\, tr\{(\overline{N}^{-1}\overline{L})^2\}\right]^{\frac{1}{2}}} \frac{(0.5)^{1/\hat{h}} \ln 2}{K\hat{h}^2} \qquad , \qquad (20)$$

where $\overline{L} = L(\hat{h}) = P'(F - ((0.5)^{1/\hat{h}} - 1)G)P$, $\overline{N} = N(\hat{w}) = I - 2\hat{w}\overline{L}$ and $\hat{w}$ satisfies

$$tr(\overline{N}^{-1}\overline{L}) = 0.$$

Based on the density function (20), we can derive the following theorem.

***Theorem 2*:**

Suppose that the data follow a stationary AR(p) process and satisfy the ADF equation: $\Delta y_t = \beta y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + u_t$, with $u_t \sim N(0, \sigma^2)$ and $\beta \in (-1,1)$, and the half-life is defined as $\hat{h} = \log(0.5)/\log(1+\hat{\beta})$, where $\hat{\beta}$ is the least squares estimator and $\hat{\beta} \in (-1,0)$. Then the mean of the half-life estimator does not exist.

***Proof*:**

$$\mathrm{M}(\hat{h}) = \int_0^\infty \hat{h}\, f(\hat{h})\, d\hat{h}$$

$$= \int_0^1 \frac{\log(0.5)}{\log(\tilde{\alpha})} f(\tilde{\alpha})\, d\tilde{\alpha}$$

Let $u(\tilde{\alpha}) = \{tr(\overline{N}^{-1} P'GP)\}\left|\overline{N}\right|^{-\frac{1}{2}}$ and $v(\tilde{\alpha}) = \left[4\pi\, tr\{(\overline{N}^{-1}\overline{L})^2\}\right]^{\frac{1}{2}}$

$$\hat{M}(\hat{h}) = \int_0^1 \frac{\log(0.5)}{\log(\tilde{\alpha})} \frac{u(\tilde{\alpha})}{v(\tilde{\alpha})} d\tilde{\alpha}$$

$$= \underset{\varepsilon \to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{\log(0.5)}{\log(\tilde{\alpha})} \frac{u(\tilde{\alpha})}{v(\tilde{\alpha})} d\tilde{\alpha} \quad .$$

Since the whole interval of $\tilde{\alpha}$ is $(-\infty, \infty)$, $u(\hat{\alpha})$ and $v(\hat{\alpha})$ are continuous functions of $\tilde{\alpha}$ on the closed interval $[0, 1]$. According to the extreme value theorem, we can assume:

(iii) when $\tilde{\alpha}=\overline{\alpha}$, $u(\tilde{\alpha})$ gets to its minimum value $\tilde{N}$ and $\tilde{N}\neq 0$.

(iv) when $\tilde{\alpha}=\breve{\alpha}$, $v(\tilde{\alpha})$ gets to its maximum value $\tilde{M}$ and $\tilde{M}\neq\infty$.

(The justification for assumptions (iii) and (iv) is similar to that for (i) and (ii) above.)

Given that $f(\hat{\alpha})\geq\delta$ for some $\delta>0$ in (0, 1), then:

$$\hat{M}\,(\hat{h})>\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{\log(0.5)}{\log(\tilde{\alpha})}\frac{\tilde{N}}{\tilde{M}}\,d\tilde{\alpha}$$

$$=\log(0.5)\frac{\tilde{N}}{\tilde{M}}\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{\log(\tilde{\alpha})}\,d\tilde{\alpha}$$

$$=\log(0.5)\frac{\tilde{N}}{\tilde{M}}\left[\underset{\varepsilon\to 0}{Lim}\left(\frac{\tilde{\alpha}}{\log(\tilde{\alpha})}\bigg|_{\varepsilon}^{1-\varepsilon}\right)+\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{[\log(\tilde{\alpha})]^{2}}d\tilde{\alpha}\right]$$

$$=\log(0.5)\frac{\tilde{N}}{\tilde{M}}\left[\infty+\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\frac{1}{[\log(\tilde{\alpha})]^{2}}\,d\tilde{\alpha}\right]\;.$$

So, our result is established.

***Corollary 2*:**

Suppose that the data follow a stationary AR(p) process and satisfy the ADF equation: $\Delta y_t=\beta y_{t-1}+\sum_{i=1}^{p-1}\phi_i\Delta y_{t-i}+u_t$, with $u_t\sim N(0,\sigma^2)$ and $\beta\in(-1,1)$, and the half-life is defined as $\hat{h}=\log(0.5)/\log(1+\hat{\beta})$, where $\hat{\beta}$ is the least squares estimator and $\hat{\beta}\in(-1,0)$. Then none of the integer-order moments of the half-life estimator exist.

***Proof*:**

$$M\,(\hat{h}^{r})=\int_{0}^{\infty}\hat{h}^{r}\,f(\hat{h})\,d\hat{h}$$

$$\hat{M}\,(\hat{h}^{r})=\int_{0}^{1}\left[\frac{\log(0.5)}{\log(\tilde{\alpha})}\right]^{r}f(\tilde{\alpha})\,d\tilde{\alpha}$$

$$=\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\left[\frac{\log(0.5)}{\log(\tilde{\alpha})}\right]^{r}\frac{u(\tilde{\alpha})}{v(\tilde{\alpha})}\,d\tilde{\alpha}$$

$$>\underset{\varepsilon\to 0}{Lim}\int_{\varepsilon}^{1-\varepsilon}\left[\frac{\log(0.5)}{\log(\tilde{\alpha})}\right]^{r}\frac{\tilde{N}}{\tilde{M}}\,d\tilde{\alpha}$$

where $\tilde{M}$ and $\tilde{N}$ are defined same as in (iii) and (iv).

Given that $f(\hat{\alpha}) \geq \delta$ for some $\delta > 0$ in (0, 1), then:

$$M\ (\hat{h}^r) > [\log(0.5)]^r \frac{\widetilde{N}}{\widetilde{M}} \underset{\varepsilon \to 0}{Lim} \int_{\varepsilon}^{1-\varepsilon} \frac{1}{[\log(\widetilde{\alpha})]^r} d\widetilde{\alpha}$$

$$= [\log(0.5)]^r \frac{\widetilde{N}}{\widetilde{M}} \left[\underset{\varepsilon \to 0}{Lim} \left(\frac{\widetilde{\alpha}}{[\log(\widetilde{\alpha})]^r} \Big|_{\varepsilon}^{1-\varepsilon} \right) + r \underset{\varepsilon \to 0}{Lim} \int_{\varepsilon}^{1-\varepsilon} \frac{1}{[\log(\widetilde{\alpha})]^{r+1}} d\widetilde{\alpha} \right]$$

$$= [\log(0.5)]^r \frac{N}{M} \left[\infty + r \underset{\varepsilon \to 0}{Lim} \int_{\varepsilon}^{1-\varepsilon} \frac{1}{[\log(\widetilde{\alpha})]^{r+1}} d\widetilde{\alpha} \right] .$$

So the corollary is proved.

6. Robustness of the Properties of the Half-Life Estimator

In the theorems and corollaries above, we assume that the disturbance term follows a normal distribution. Here, we examine whether our results are robust to a relaxation of this normality assumption. Let H_k represent the regular regression model; and H_k^{-1} represent the first-order autoregressive model. $E_1(n, \Sigma)$ represents the elliptically symmetric family of distributions. Then King (1979; p. 121) proves that "when the disturbance vector of H_k and H_k^{-1} takes an $E_1(n, \Sigma)$ distribution, any linear unbiased or any well-behaved non-linear estimator will have very similar properties to those of the same estimator when the disturbance term is normally distributed." From King's result, we would expect that we will get similar results if we generalize the normality assumption to the assumption that the errors follow an elliptically symmetric distribution. All that is of concern to us here is determining if the non-existence of the moments of the half-life estimator still hold under other distributional assumptions. Here, we apply further results of Lieberman (1997) that relate to the non-normal case, to prove the robustness of the theorems we presented above to the distributional assumption. Lieberman derives the saddlepoint approximation for the density and cumulative distribution function for the estimator $\hat{\alpha}$ in an AR(1) model with some exogeneous variables. Applying his result to (3), we can get the saddlepoint approximation to $\hat{\alpha}$ in (3). First, we let

$$S = v'R'_\alpha C_1 R_\alpha v - \alpha v' R'_\alpha C_2 R_\alpha v$$

$$Z = v'R'_\alpha C_2 R_\alpha v$$

$$B = R'_\alpha C_2 R_\alpha .$$

Then the saddlepoint approximation to the density of $\hat{\alpha}$ is:

$$f(\hat{\alpha}) = \frac{\widetilde{k}_{10} e^{\widetilde{K}_S}}{\sqrt{2\pi \widetilde{k}_2^S}}, \tag{21}$$

with the saddlepoint $\hat{w}$ satisfying

$$K'_S(\hat{w}) = 0 , \tag{22}$$

where $K_S(w)$ is the cumulant generating function of S and

$$\widetilde{K}_S = K_S(\hat{w}) \tag{23}$$

$$\widetilde{k}_2^S = K''_D(\hat{w}) \tag{24}$$

$$k_{10} = E(Z) \tag{25}$$

$$\widetilde{k}_{10} = k_{10}(\hat{w}) . \tag{26}$$

Suppose v has arbitrary cumulants $k^i = 0$, $k^{i,j}$, $k^{i,j,k}, \cdots$, where the cumulants are defined as follows:

$$k^{i,j} = cum(v^i, v^j)$$

$$k^{i,j,k} = cum(v^i, v^j, v^k)$$

Then (24) and (26) can be expressed in terms of v's cumulants $k^{i,j}$, $k^{i,j,k}, \cdots$.

$$\widetilde{k}_2^S = \sum_{ijkl} s_{ij} s_{kl} k^{ij,kl} \tag{27}$$

$$\widetilde{k}_{10} = \sum_{ij} b_{ij} k^{ij} \quad . \tag{28}$$

This specification allows the v's to be correlated. When v is $i.i.d.$, (27) and (28) reduce to

$$\widetilde{k}_2^S = k_4 \sum_{ij} s_{ij}^2 + 2k_2^2 \sum_{ij} s_{ij}^2 \tag{29}$$

$$\widetilde{k}_{10} = k_2 \sum_i b_{ii} \tag{30}$$

where $k_2 = k^{i,i}, k_4 = k^{i,i,i,i}$.

The approximating function in (21) is continuous on a closed interval $\hat{\alpha} \in [0,1]$. We can use the same procedure as for Theorem 1 to prove that the moments of the half-life estimator do not exist.

$$M\,(\hat{h}) = \int_0^\infty \hat{h}\, f(\hat{h})\, d\hat{h}$$

the
$$\hat{M}\,(\hat{h}) = \int_0^1 \frac{\log(0.5)}{\log(\hat{\alpha})} \frac{\widetilde{k}_{10} e^{\widetilde{K}_S}}{\sqrt{2\pi\, \widetilde{k}_2^S}}\, d\hat{\alpha}$$

If ν is $i.i.d.$ and the second cumulant of ν is finite, then $\widetilde{k}_2^S$ and $\widetilde{k}_{10}$ are defined by (29) and (30). And we can also see that both $\widetilde{k}_2^S$ and $\widetilde{k}_{10}$ are continuous functions of $\hat{\alpha}$ on the closed interval $[0,\ 1]$, and they are the sum of a finite number of terms. Therefore, there is a non-zero minimum and maximum for the numerator and denominator of the expression for the density function of $\hat{\alpha}$ in (21). We assume that $\hat{N}$ is the minimum value of the numerator and $\hat{M}$ is the maximum value of the denominator, and $\hat{N} \neq 0$, $\hat{M} \neq 0$.

Then:

$$\begin{aligned}
\hat{M}\,(\hat{h}) &> \underset{\varepsilon\to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{\log(0.5)}{\log(\hat{\alpha})} \frac{\hat{N}}{\hat{M}}\, d\hat{\alpha} \\
&= \log(0.5) \frac{N}{M} \underset{\varepsilon\to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{1}{\log(\hat{\alpha})}\, d\hat{\alpha} \\
&= \log(0.5) \frac{N}{M} \left[\underset{\varepsilon\to 0}{Lim} \left(\frac{\hat{\alpha}}{\log(\hat{\alpha})} \bigg|_\varepsilon^{1-\varepsilon} \right) + \underset{\varepsilon\to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{1}{[\log(\hat{\alpha})]^2}\, d\hat{\alpha} \right] \\
&= \log(0.5) \frac{N}{M} \left[\infty + \underset{\varepsilon\to 0}{Lim} \int_\varepsilon^{1-\varepsilon} \frac{1}{[\log(\hat{\alpha})]^2}\, d\hat{\alpha} \right]
\end{aligned}$$

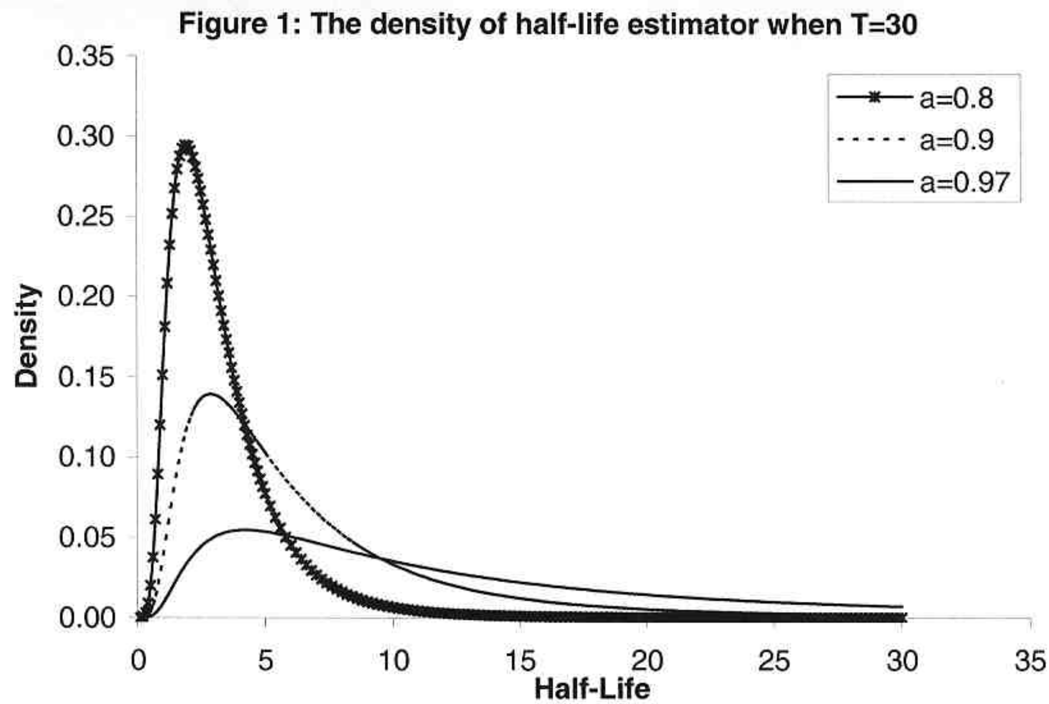
therefore, the mean of $\hat{h}$ does not exist. Similarly, we can prove that none of the integer-order moments of $\hat{h}$ exist either.

Therefore, all of our main results hold as long as ν is $i.i.d.$ and the second cumulant of ν is finite. In addition to the normal distribution, there are many distributions with a finite second cumulant. When we allow the disturbances to be correlated, the situation is more complicated. However, we can still find quite a large class of distributions which will satisfy the conditions of the above proof. For the AR(p) model, we can apply (21) to (18). The situation is almost the same as for the AR(1) model. Therefore, the property that the moments of the half life estimator do not exist is quite robust to the distributional assumption for the errors of the fitted model.

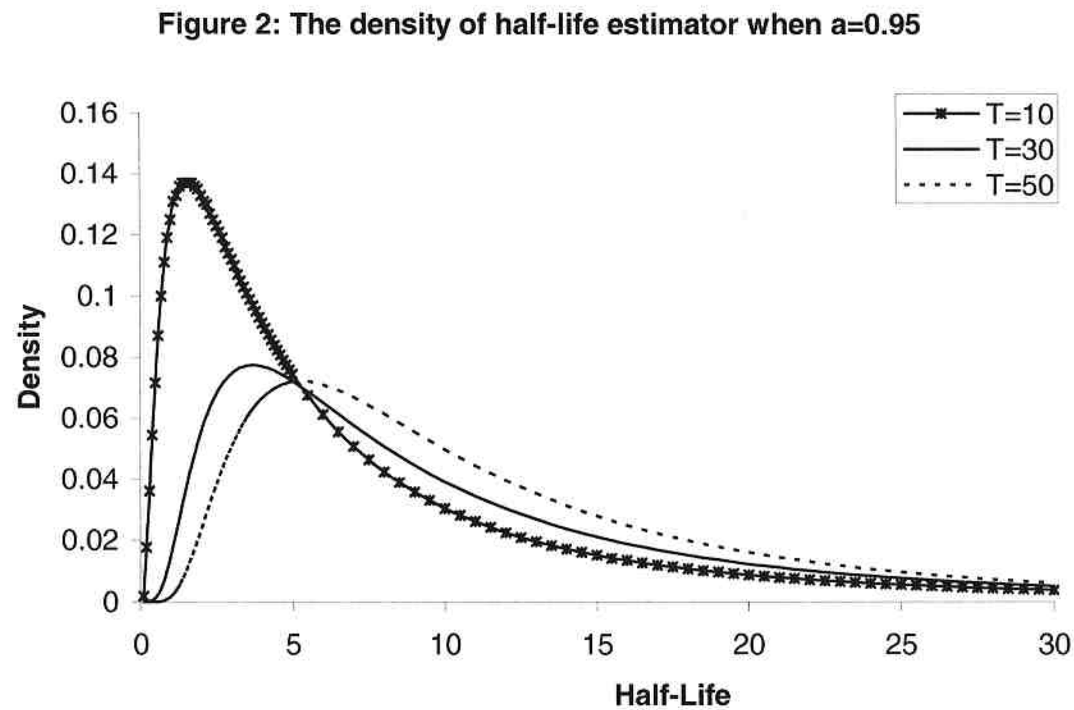
7. Conclusions

Given the important role of PPP theory in economics, it is natural that the "PPP puzzle" has attracted a lot of attention. However, although numerous efforts have been made to try to solve the puzzles by exploring the possible reasons from different perspectives, no one can give a specific answer. This chapter tries to add some distributions to this literature. It provides explicit density and distribution functions for the half life estimator, and proves analytically that the moments of half life estimator actually do not exist, not only for the AR(1) model but also for AR(p) models. Moreover, this result is quite robust to the distributional assumptions. These properties explain clearly the existence of the quite wide confidence intervals in all of the associated empirical studies.

Our results have some implications for future research. First, the poor properties of the half-life estimator may suggest that the measure used in this chapter is not a good one, which supports Chortareas and Kapetanios' (2004) arguments that the puzzle may be caused artificially by the measure we use. So, future work may be better to focus on constructing some appropriate methods to measure the persistence, rather than just explore all possible reasons to improve the accuracy based on the current measure of the half-life. Second, we have not considered the case of nonlinear models. However, the assumption in the nonlinear models employed in this literature is that the arbitrage happens only when the deviation is quite large. So we can imagine that when the deviation is small, the situation for the nonlinear models would be similar with the case we analyzed here. Therefore, nonlinear models can never solve the problem for small deviations. If we think the source of the puzzle lies in the model specification, then we must construct some model which is unrelated to linear models. For example, we know that the PPP theory is based on the complete trade arbitrage world. We can perhaps find an appropriate way to put other variables, which do affect the perfect assumption of the complete trade arbitrage, into the model, rather than just base it on the real exchange rate to study the validity of PPP theory. Third, some studies found that the key engine governing the speed of PPP convergence is the nominal exchange rate, not the price and the slow reversion of PPP is due to the slow reversion of nominal exchange rate (*e.g.*, Engel and Morley, 2001; Cheung, *et al.*, 2004). This can also explain why the empirical consensus is inconsistent with the sticky-price model. Perhaps, we just need to explore more explanations for the long persistence. The puzzle is not a real one.



Note: Figure 1 depicts the saddlepoint density function (8), which is the approximate density of $\hat{h}$ when the sample size is 30 and the true α is 0.8, 0.9, 0.97 respectively.



Note: Figure 2 depicts the saddlepoint density function (8), which is the approximate density of $\hat{h}$ when the sample size is 10, 30, 50 respectively and the true α is 0.95.

Appendix: Proof of Assumptions of $N \neq 0$ and $M \neq \infty$

Here we justify assumptions (i) and (ii) used in the proof of Theorem 1.

First, we prove that $N \neq 0$:

$$N = u(\overline{\alpha}) = \left\{tr(\hat{A}^{-1}R'_{\alpha}C_2R_{\alpha})\right\}\left|\hat{A}\right|^{-\frac{1}{2}} .$$

So, if N =0, then $\left|\hat{A}\right|^{-\frac{1}{2}} = 0$ or $\left\{tr(\hat{A}^{-1}R'_{\alpha}C_2R_{\alpha})\right\} = 0$.

As the density exists, we can rule out the possibility that $\left|\hat{A}\right|^{-\frac{1}{2}} = 0$.

For $$\left\{tr(\hat{A}^{-1}R'_{\alpha}C_2R_{\alpha})\right\} = \sum_{t=0}^{T} \frac{f_t}{1-2\hat{w}d_t},$$

where the d_t are the eigenvalues of matrix D and the f_t are the eigenvalues of $R'_{\alpha}C_2R_{\alpha}$

Since $\left|\hat{A}\right|^{-\frac{1}{2}} = \exp\left\{-\frac{1}{2}\sum_0^T \log(1-2\hat{w}d_t)\right\}$ exists, $\frac{1}{1-2\hat{w}d_t}$ must be positive. Also, $R'_{\alpha}C_2R_{\alpha}$ is a positive definite matrix, so the eigenvalues f_t are all positive. Therefore:

$$\left\{tr(\hat{A}^{-1}R'_{\alpha}C_2R_{\alpha})\right\} > 0 ,$$

and so $N \neq 0$ is proved

Second, we prove that $M \neq \infty$.

$$M = v(\tilde{\alpha}) = \left[4\pi tr\{(\hat{A}^{-1}D)^2\}\right]^{\frac{1}{2}} = \left\{4\pi \sum_0^T \left[d_t^2/(1-2\hat{w}d_t)^2\right]\right\}^{1/2}$$

so, if $M = \infty$, it must be the case that $\frac{1}{(1-2\hat{w}d_t)^2}$ is zero. But from $\left|\hat{A}\right|^{-\frac{1}{2}} \neq 0$, we know this could not happen. Therefore $M \neq \infty$ is proved.

CHAPTER 3:
DIFFERENT SADDLEPOINT APPROXIMATIONS TO THE DISTRIBUTION FUNCTION OF THE ANDERSON-DARLING TEST STATISTIC

1. Introduction

The Anderson-Darling (A-D) test is widely used for testing the hypothesis that a sample of size n has been drawn from a population with a specified continuous cumulative distribution function. Anderson and Darling (1952) derived this test based on the discrepancies between the empirical cumulative distribution function $F_n(x)$ and the specified cumulative distribution function $F(x)$ with different weight function of $\psi(\cdot)$. It is actually a modification of the Kolmogorov-Smirnov (K-S) test. The Anderson-Darling test gives more weight to the tails than the K-S test, which makes the Anderson-Darling test more sensitive at the tail than at the median. Anderson and Darling mainly considered two cases for this test, when the weight function is chosen to be either $\psi(F(x)) = 1/[F(x)(1-F(x))]$ or $\psi(F(x)) = 1$. In this literature, we usually call the statistic based on the former weight function the A-D test.

Anderson and Darling's main purpose was to find the asymptotic distribution of their test statistic and the significance points at different probability levels. However, it turns out that the distribution function is very complicated, even asymptotically. Consequently, Anderson and Darling (1954) used a numerical method previously applied by Birnbaum (1952) to the K-S test, to obtain the asymptotic critical points for significance levels of 1%, 5% and 10% for the A-D test. Lewis (1961) found that the support of the A-D test statistic is essentially from 0 to 8. He computed an extensive table for the cumulative distribution function of this test statistic on the interval from 0.025 to 8, not only for the asymptotic distribution but also for sample sizes from 1 to 8. The results for the asymptotic case are based on the application of Hermite-Gauss numerical quadrature to evaluate the integral associated with the c.d.f. in the asymptotic case, while the results for sample sizes from 2 to 8 are based on Monte Carlo simulation. When n = 1 the expression for the c.d.f. is especially simple and requires no integration. For $n > 8$, the computing time associated with the Monte Carlo method was prohibitive due to the CPU speeds and sampling methods available at that time. Lewis found the convergence speed of this distribution to be very fast as n grows, and the maximum difference between the points on the distribution for $n = 8$ and for $n \to \infty$ are approximately 0.006. Therefore, the practical loss from not considering the larger sample sizes is not great. Recently, by taking advantage of the current CPU speeds and fast sampling methods, Marsaglia and Marsaglia (2004) gave a method to evaluate the distribution function to the fourth digit for any sample size n. Lewis also reported the 90%, 95% and 99% critical points for the sample sizes of $n = 1$ to 8 and $n \to \infty$. The results when $n = 1$ are based on an exact closed-form expression. For the case when $n \to \infty$ the results are "exact" to the extent of the accuracy of the numerical integration involved. For the sample sizes 2 to 8, the critical points were obtained by inverse interpolation from the cumulative distribution table, which in turn is simulation based, so two sources of approximation error are associated with these values.

Although the c.d.f. derived by Anderson and Darling is very complicated, the characteristic function is simple and has some interesting properties. For example, the characteristic function of the A-D test corresponds to an infinite weighted sum of independent chi-squared random variables. Therefore, many of the research contributions related to this statistic are based on the characteristic function. Sinclair and Spurr (1988) used the results on limiting distributions of quadratic forms of Zolotarev (1961) to derive a theoretical function of the upper tail area for the A-D test. The approximations based on this formula are very good above the median, especially in the upper tail of the distribution. However, the approximation is poor in the lower tail, with some areas in excess of unity over a certain range. Sinclair and Spurr derived the first four cumulants of the A-D test using the characteristic function. Based on the skewness and the kurtosis, they concluded that the distribution of the A-D test should lie between the gamma and log-normal distributions. Therefore, Sinclair and Spurr derived an empirical function by fitting a generalization of the logistic distribution to the values given by Lewis. The empirical function corrects the problem of overestimating the upper tail areas and provides a good complement to Lewis's table.

Another important contribution based on the characteristic function is Giles (2001). Giles derived a saddlepoint approximation to the asymptotic distribution function of the A-D statistic by applying the formula proposed by Lugannani and Rice (1980). The numerical results show that the approximations are excellent in both tails, in good agreement with the approximations given by Lewis. Since the characteristic function of the A-D test is an infinite weighted sum of independent chi-squared random variables, Giles suggested that a saddlepoint formula with a chi-square base, not a normal base, might perform better for this test. Following Giles' suggestions, this chapter aims to compare the performances of different types of saddlepoint approximations for the A-D test. These include not only saddlepoint approximations that use different base distributions, but also higher-order saddlepoint approximations so that we can check if the higher-order terms can improve the quality of the approximations. This chapter extends Giles' work in three directions. First, we extend the interval of the test from the tail areas to the whole distribution of the A-D test; and second, we extend the Lugannani-Rice formula to the chi-square base. We then extend it to the higher-order approximation case. Overall, our objective here is to use the Anderson-Darling test as a basis for examining what gains may or may not come from considering such extensions of the usual saddlepoint approximation. Lewis' numerical results provide a useful benchmark for us to use in this regard.

The structure of this chapter is as follows: in section 2, we present the Anderson-Darling test and discuss some of its basic properties. In section 3, we discuss some saddlepoint theory and Giles' saddlepoint approximation of the A-D test statistic. We then derive the saddlepoint approximations of the A-D test statistics with a non-normal base distribution and extend the approximation to higher-order terms. In section 4, we generate some numerical results based on the derivation in section 3 and compare the results with Lewis' numerical results. Some concluding comments follow in section 5.

2. The Anderson-Darling Test Statistic

Anderson and Darling (1952) considered a general test for the hypothesis that n observations have been drawn from a population with distribution function $F(x)$. The associated test statistic is:

$$A_n^2 = n\int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 \psi[F(x)]\, dF(x),$$

where $F_n(x)$ is the empirical distribution function based on n observations; $\psi(t)$ (≥ 0) is some preassigned weight function.

When the weight function is $\psi(t) = 1/[t(1-t)]$, the test is:

$$A_n^2 = n\int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 / [F(x)\{1 - F(x)\}] dF(x).$$

The test can be applied as the following: first, we order the n observations $x_1 \leq x_2 \leq \cdots \leq x_n$ and let $u_i = F(x_i)$. Then we compute:

$$A_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln(u_j) + \ln(1 - u_{n-j+1})\right].$$

The asymptotic test statistic is

$$A^2 = \underset{n\to\infty}{Lim}(A_n^2).$$

As noted already, Lewis (1961) found that the range of A^2 must be between 0 and 8. Anderson and Darling gave an explicit expression for the characteristic function of the limiting distribution A^2, namely:

$$\phi(t) = \prod_{j=1}^{\infty}\left[1 - 2it/(j(j+1))\right]^{-1/2}$$

$$= \prod_{j=1}^{\infty}\left[1 - 2\lambda_j it\right]^{-1/2},$$

where $\lambda_j = 1/(j(j+1))$ and $i^2 = -1$

So, the moment generating function is

$$\theta(t) = \prod_{j=1}^{\infty}\left[1 - 2\lambda_j t\right]^{-1/2}, \tag{1}$$

From equation (1), the cumulant generating function (CGF) of A^2 is:

$$K(t) = \log[\theta(t)] = -0.5\sum_{j=1}^{\infty}\log[1 - 2\lambda_j t]. \tag{2}$$

We know that the essential requirement for a saddlepoint approximation is to know the cumulant generating function. Therefore, we can derive the saddlepoint approximation for the A-D test, based on the characteristic function (1). Giles (2001) derives the saddlepoint approximation for this test by applying the standard Lugannani and Rice saddlepoint formula and provides some numerical approximations to the tail areas of the asymptotic distribution of the A-D test. Giles' approximations compare favorably with Lewis' numerical results. However, Giles suggested that a saddlepoint approximation with non-normal base might provide some improvements to this statistics. There are two reasons for this suggestion. First, as we can see, this cumulant generating function is an infinite sum of independent $\chi^2(1)$ variates, with weights λ_j. Second, Sinclair and Spurr (1988) show that the coefficients of skewness and kurtosis are 5.5865 and 12.036 respectively, which are quite different from those for the normal distribution. Therefore, a saddlepoint approximation with a chi-square base might be expected to provide some improvement in this case. In addition to pursuing this suggestion, we also test if some improvement can be achieved by extending the saddlepoint approximations to higher-order terms. In the next part, we will derive different saddlepoint approximations for the asymptotic distribution of the Anderson-Darling test statistic.

3. Saddlepoint Approximations

3.1 Giles' saddlepoint approximation

Before introducing Giles' results, we first need to introduce the Luganini-Rice (LR) forumulae. The L-R formula for the distribution of the random variable X at point y is:

$$\Pr(X \geq y) \approx 1 - \Phi(\hat{w}) + \phi(\hat{w})\left\{\frac{1}{\hat{u}} - \frac{1}{\hat{w}}\right\}, \tag{3}$$

where Φ and ϕ are the c.d.f. and density function of the standard normal distribution, and

$$\hat{w} = \{2[\hat{t}y - K(\hat{t})]\}^{1/2}\,\mathrm{sgn}(\hat{t}) \tag{4}$$

$$\hat{u} = \hat{t}[K^{(2)}(\hat{t})]^{1/2}. \tag{5}$$

Here, $K(t)$ is the CGF of the variable X, and we let $K^{(r)}(\cdot)$ denote the i^{th} derivative of $K(\cdot)$ with respect to its argument; $\hat{t}$ is the solution to the saddlepoint equation:

$$K^{(1)}(\hat{t}) = y \tag{6}$$

and

$$\mathrm{sgn}(\hat{t}) = \begin{cases} +1, & \text{if } \hat{t} > 0 \\ -1, & \text{if } \hat{t} < 0 \\ 0, & \text{if } \hat{t} = 0 \end{cases}. \tag{7}$$

When $y = E(X) = K^{(1)}(0)$, the root $\hat{t}$ is zero, based on the saddlepoint equation (6). Then $\hat{u}$ and $\hat{w}$ are zero, from (4) and (5). This leads to the collapse of (3). Therefore, at the mean value, we need to take the limit of (3), and by L'Hôpital's rule the formula (3) reduces to:

$$\Pr(X \geq E(x)) \approx \frac{1}{2} - \frac{\zeta_3(0)}{6\sqrt{2\pi}} \tag{8}$$

where $\zeta_r(\hat{t}) = K^{(r)}(\hat{t}) / \{K^{(2)}(\hat{t})\}^{r/2}$.

We now introduce some theorems from Daniels (1954) regarding the existence and properties of the real root of the saddlepoint equation (6). According to Daniels, suppose that the moment generating function of the variable X is:

$$M(T) = e^{K(T)} = \int_{-\infty}^{+\infty} e^{tx} dF(x), \tag{9}$$

Daniels proved that if (9) converges for real $t \in (-c_1,\ c_2)$, where both c_1 and c_2 are nonnegative, and $c_1 + c_2 > 0$, and then the saddlepoint approximation for this variable can be derived. Although in (9), the support of the distribution is the full real line, actually Daniels' arguments are also applicable to the case where the support is limited at either or both ends. The properties of the root of the saddlepoint approximation are provided in the following theorem.

Theorem 1 (Daniels, 1954, p. 638)

Let $F(x) = 0$ for $x < a$, $0 < F(x) < 1$ for $a < x < b$, and $F(x) = 1$ for $b < x$, where $-\infty < a < b < \infty$. Then for every y in $a < y < b$, there is a unique simple root $\hat{t}$ of $K^{(1)}(t) = y$. As t increases from $-\infty$ to ∞, $K^{(1)}(t)$ increases continuously from $y = a$ to $y = b$.

In the case where one or both of a and b are infinite, the conditions

$$\lim_{t \to c_2} K^{(1)}(t) = b,\ \lim_{t \to c_1} K^{(1)}(t) = a$$

are also required for the uniqueness of the root of the saddlepoint equation. Our case is the simple case as Lewis proved that the support of the A-D test statistic is from 0 to 8. Based on the L-R formula (3), Giles first calculated the first two derivatives of the CGF of the Anderson-Darling statistic from (2).

$$K^{(1)}(t) = \sum_{j=1}^{\infty} [\lambda_j / (1 - 2\lambda_j t)] \tag{10}$$

$$K^{(2)}(t) = 2\sum_{j=1}^{\infty}[\lambda_j/(1-2\lambda_j t)]^2 . \tag{11}$$

In addition to the first two derivatives of the CGF, we need further derivatives of the CGF for some of the following results. Therefore, we note them here.

$$K^{(3)}(t) = 8\sum_{j=1}^{\infty}[\lambda_j/(1-2\lambda_j t)]^3 \tag{12}$$

$$K^{(4)}(t) = 48\sum_{j=1}^{\infty}[\lambda_j/(1-2\lambda_j t)]^4 \tag{13}$$

$$K^{(5)}(t) = 384\sum_{j=1}^{\infty}[\lambda_j/(1-2\lambda_j t)]^5 . \tag{14}$$

After calculating the derivatives of the CGF, Giles obtained $\hat{t}$ by solving the saddlepoint equation (6) for any y value of interest and $\hat{u}$ and $\hat{w}$ by substituting $\hat{t}$ into (4), (5) respectively. Therefore, the saddlepoint approximation to the c.d.f. of A^2 is:

$$\Pr(A^2 \geq y) \approx 1 - \Phi(\hat{w}) + \phi(\hat{w})\left\{\frac{1}{\hat{u}} - \frac{1}{\hat{w}}\right\}, \tag{15}$$

where all the variables are defined as in (3).

When $A^2 = E(A^2) = K^{(1)}(0)$, (15) reduces to

$$\Pr(A^2 \geq E(A^2)) \approx \frac{1}{2} - \frac{\zeta_3(0)}{6\sqrt{2\pi}} \quad . \tag{16}$$

Giles also discussed some properties of this saddlepoint equation and its root. For any y, the saddlepoint equation is:

$$\sum_{j=1}^{\infty}[\lambda_j/(1-2\lambda_j t)] - y = 0 \tag{17}$$

where λ_j is defined as in (1).

The solution $\hat{t}$ is well defined, since:

$$\underset{j\to\infty}{Lim}[\lambda_j/(1-2\lambda_j t)] = 0 .$$

From (11), $K^{(2)}(t) > 0$, which means that $K^{(1)}(t)$ is a continuously increasing function in the whole interval. Therefore, from theorem 1, the root of $\hat{t}$ is unique. So, we can say the saddlepoint approximation of the A-D test is well defined. The saddlepoint equation (17) can be easily solved numerically, for example, by the Newton-Raphson algorithm, or by using a line-search. Convergence is very rapid. In this chapter, we use the Newton-Raphson algorithm for our related numerical evaluations.

3.2 Non-normal-based saddlepoint approximation

Wood, Booth and Butler (1993) generalize the LR formula to the non-normal-based saddlepoint approximation, which we call the WBB formula. To obtain the WBB formula, first, Wood, Booth and Butler make the following transformation from a normal base to a non-normal base:

$$G(w_{\hat{\varepsilon}}) - \hat{\varepsilon} w_{\hat{\varepsilon}} = K(\hat{t}) - y\hat{t} \tag{18}$$

where $\hat{t}$ is defined as in equation (3), and $w_{\hat{\varepsilon}}$ is the solution to the saddlepoint equation:

$$G^{(1)}(w_{\hat{\varepsilon}}) = \hat{\varepsilon} \quad . \tag{19}$$

Again, here we let $G^{(r)}(\cdot)$ denote the r[th] derivative of $G(\cdot)$ with respective to its argument.

The WBB formula at point y is:

$$\Pr(X \geq y) \approx 1 - \Gamma(\hat{\varepsilon}) + \gamma(\hat{\varepsilon})\left\{\frac{1}{\hat{u}_{\varepsilon}} - \frac{1}{w_{\hat{\varepsilon}}}\right\}, \tag{20}$$

where Γ and γ are the c.d.f. and density function of the base distribution whose CGF is G.. $\hat{\varepsilon}$ and $w_{\hat{\varepsilon}}$ are defined by (18) and (19). Clearly, we need to solve the root $\hat{t}$ before we calculate $\hat{\varepsilon}$ and $w_{\hat{\varepsilon}}$. After finding $\hat{t}$, we can express $w_{\hat{\varepsilon}}$ as a function of $\hat{\varepsilon}$ based on (19), then we substitute this function into (18) to solve for $\hat{\varepsilon}$ using some numerical method. Now we can substitute the solution of $\hat{\varepsilon}$ into the function of $w_{\hat{\varepsilon}}$ to get the solution of $w_{\hat{\varepsilon}}$. For equation (18), the left-hand side is the Legendre-Frenchel transformation of G., which is a concave function of $\hat{\varepsilon}$, and therefore there are at most two solutions for $\hat{\varepsilon}$, and the choice for the solution of $\hat{\varepsilon}$ is:

$$\hat{\varepsilon} = \begin{cases} \varepsilon_{+}(y), & \text{if } y > K^{(1)}(0) \\ \varepsilon_{-}(y), & \text{if } y < K^{(1)}(0) \\ G'(0), & \text{if } y = K^{(1)}(0) \end{cases} \tag{21}$$

and

$$\hat{u}_{\hat{\varepsilon}} = \hat{u}[G^{(2)}(w_{\hat{\varepsilon}})]^{-1/2} \tag{22}$$

where $\hat{u}$ is defined by (5).

When $y = E(X) = K^{(1)}(0)$, $\hat{\varepsilon} = G^{(1)}(0)$ from (21). Then from (19), $w_{\hat{\varepsilon}} = 0$. Also $\hat{u}_{\hat{\varepsilon}} = 0$, for the same reason as with the LR formula. Therefore, the WBB formula has the same problem as the LR formula. At the mean value, we should calculate the limit of (20), which is:

$$\Pr(X \geq E(x)) \approx 1 - \Gamma(G^{(1)}(0)) + \frac{1}{6}\sqrt{G^{(2)}(0)}\gamma(G^{(1)}(0))\{\zeta_3'(0) - \zeta_3(0)\} \tag{23}$$

where $\zeta_r'(w_{\hat{\varepsilon}}) = G^{(r)}(w_{\hat{\varepsilon}}) / \{G^{(2)}(w_{\hat{\varepsilon}})\}^{r/2}$.

Our formula (23) is different from equation (9) in Wood *et al.* (1993, p.681). We are grateful to Professor Wood for helpful email correspondence through which we were able to confirm that there is a type-setting error in equation (9) of their paper. Our formula (23) is the correct one.

Now based on all these results, we can derive the non-normal-based saddlepoint approximation for the c.d.f. of the Anderson-Darling statistic, which is:

$$\Pr(A^2 \geq y) \approx 1 - \Gamma(\hat{\varepsilon}) + \gamma(\hat{\varepsilon})\left\{\frac{1}{\hat{u}_{\varepsilon}} - \frac{1}{w_{\hat{\varepsilon}}}\right\} \tag{24}$$

where all the variables are defined as above. Γ and γ could represent any distribution and its associated density function. If they represent the normal distribution, then (24) is the LR formula (15). Also when G is the same as the true CGF of the statistic, then the approximation is exact. When A^2 equals its mean, we use the following formula for its c.d.f.:

$$\Pr(A^2 \geq E(A^2)) \approx 1 - \Gamma(G^{(1)}(0)) + \frac{1}{6}\sqrt{G^{(2)}(0)}\gamma(G^{(1)}(0))\{\zeta_3'(0) - \zeta_3(0)\}. \tag{25}$$

The CGF of the Anderson-Darling test is an infinite weighted sum of independent $\chi^2(1)$ variates. Also the skewness and the kurtosis of the Anderson-Darling test suggest that the true distribution should lie between the gamma and the log-normal distributions. Therefore, in this paper, we will show some numerical results when Γ is chosen to be the chi-squared distribution. We expect that (24) could bring some improvements over (15). The CGF of the chi-squared distribution with α degrees of freedom is:

$$G(w_{\hat{\varepsilon}}) = -\frac{\alpha}{2}\log(1-2w_{\hat{\varepsilon}}). \tag{26}$$

As before, we need to calculate some derivatives of (26) for the following results:

$$G^{(1)}(w_{\hat{\varepsilon}}) = \alpha/(1-2w_{\hat{\varepsilon}}) \tag{27}$$

$$G^{(2)}(w_{\hat{\varepsilon}}) = 2\alpha/(1-2w_{\hat{\varepsilon}})^2 \tag{28}$$

$$G^{(3)}(w_{\hat{\varepsilon}}) = 8\alpha/(1-2w_{\hat{\varepsilon}})^3 \tag{29}$$

$$G^{(4)}(w_{\hat{\varepsilon}}) = 48\alpha/(1-2w_{\hat{\varepsilon}})^4 \tag{30}$$

$$G^{(5)}(w_{\hat{\varepsilon}}) = 384\alpha/(1-2w_{\hat{\varepsilon}})^5\,. \tag{31}$$

The process for calculating (25) is as follows: first, solve the root $\hat{t}$ using Giles' approximation. Then, we substitute (27) into (19) to solve $w_{\hat{\varepsilon}}$ in terms of $\hat{\varepsilon}$:

$$\alpha/(1-2w_{\hat{\varepsilon}}) = \hat{\varepsilon}\,,$$

so that

$$w_{\varepsilon} = \frac{1}{2} - \frac{\alpha}{2\varepsilon} \quad . \tag{32}$$

Then we substitute (32) into (18) to solve for the root $\hat{\varepsilon}$:

$$\frac{\alpha}{2}\log(\frac{\varepsilon}{\alpha}) - \frac{\varepsilon}{2} + \frac{\alpha}{2} = -0.5\sum_{j=1}^{\infty}\log[1-2\lambda_j\hat{t}] - y\hat{t} \tag{33}$$

and substitute $\hat{\varepsilon}$ back into (32) to get $w_{\hat{\varepsilon}}$.

After solving for $\hat{t}$, $\hat{\varepsilon}$ and $w_{\hat{\varepsilon}}$, we can get $\hat{u}_{\hat{\varepsilon}}$ based on (22). Now we can calculate (24). There is still another problem we need to solve: the choice of the degrees of freedom for the chi-squared distribution, α. Here we follow Wood *et al.*'s suggestion to choose α. The choice of α made by Wood *et al.* is based on matching the derivatives of $K(\hat{t})$ and $G(w_{\hat{\varepsilon}})$, which means

$$\frac{[K^{(2)}(\hat{t})]^3}{[K^{(3)}(\hat{t})]^2} = \frac{[G^{(2)}(w_{\hat{\varepsilon}})]^3}{[G^{(3)}(w_{\hat{\varepsilon}})]^2} \tag{34}$$

so that

$$\alpha = 8\frac{[K^{(2)}(\hat{t})]^3}{[K^{(3)}(\hat{t})]^2} \quad . \tag{35}$$

Evaluating (34) at $\hat{t} = 0$ and $w_{\hat{\varepsilon}} = 0$, then (35) reduces to

$$\alpha = 2\frac{[K^{(1)}(0)]^2}{[K^{(2)}(0)]} \quad . \tag{36}$$

We know the support of the A-D test statistic is $0 \le A^2 \le 8$. Here we also consider the case of $\alpha = 2$, because the support for the chi-squared distribution with two degrees of freedom is close to that of the A-D test statistic. Of course there are lots of other choices for α, but in this chapter we consider only these three cases.

3.3 Higher-order saddlepoint approximations

In addition to the non-normal-based saddlepoint approximation, another main contribution of this paper is to see if the saddlepoint approximations (15) and (24) can achieve some improvement by including the higher-order terms. In this part, we will extend (15) and (24) to allow for higher order terms.

A. Higher-order Lugannani-Rice formula

Daniels (1987) extended the Lugannani-Rice formula to higher-order terms for the mean of n independent identically distributed random variables in (4.5) (Daniels, 1987, p. 42):

$$P(\overline{X} > \overline{x}) \approx 1 - \Phi(\hat{w}n^{1/2}) + \phi(\hat{w}n^{1/2})\left\{\frac{b_0}{n^{1/2}} + \frac{b_1}{n^{3/2}} + \ldots + \frac{b_k}{n^{(k+1)/2}} + o(n^{-k-3/2})\right\} \tag{37}$$

where

$$b_0 = \frac{1}{\hat{u}} - \frac{1}{\hat{w}}, \quad b_1 = \frac{1}{\hat{u}}(\frac{\zeta_4}{8} - \frac{5}{24}\zeta_3^2) - \frac{\zeta_3}{2\hat{u}^2} - \frac{1}{\hat{u}^3} + \frac{1}{\hat{w}^3} \tag{38}$$

and for brevity, we write $K^{(r)}(\hat{t})$, $G^{(r)}(w_{\hat{\varepsilon}})$, $\zeta_{(r)}(\hat{t})$ and $\zeta'_{(r)}(w_{\hat{\varepsilon}})$ as $K^{(r)}$, $G^{(r)}$, $\zeta_{(r)}$ and $\zeta'_{(r)}$.

Daniels' results can be easily generalized from the mean to a single random variable X:

$$P(X > y) \approx 1 - \Phi(\hat{w}) + \phi(\hat{w})\{b_0 + b_1 + \ldots + b_k + \ldots\} \quad . \tag{39}$$

Then we apply (39) to the Anderson-Darling test statistic to get a higher-order approximation for its CDF by retaining both the b_0 and b_1 terms in (39):

$$\Pr(A^2 \ge y) \approx 1 - \Phi(\hat{w}) + \phi(\hat{w})\left\{(\frac{1}{\hat{u}} - \frac{1}{\hat{w}}) + \frac{1}{\hat{u}}(\frac{\zeta_4}{8} - \frac{5}{24}\zeta_3^2) - \frac{\zeta_3}{2\hat{u}^2} - \frac{1}{\hat{u}^3} + \frac{1}{\hat{w}^3}\right\} \tag{40}$$

When $A^2 = E(A^2) = K^{(1)}(0)$, (40) reduces to

$$\Pr(A^2 \geq E(A^2)) \approx \frac{1}{2} - \frac{1}{2\pi}\left\{\frac{1}{6}\zeta_3 - \frac{1}{40}\zeta_5 + \frac{5}{48}\zeta_3\zeta_4 - \frac{35}{432}\zeta_3^2\right\}, \tag{41}$$

where $\hat{w}$ and $\hat{u}$ are defined as in (15) and ζ_r is defined as in (16).

B. Higher-order-WBB approximation

In this part, we derive the higher-order-WBB formula based on some results from Daniels and Wood, Booth and Butler. The proof is provided in the appendix, and here we just state the results.

The higher-order-WBB formula is:

$$\Pr(A^2 \geq y) \approx 1 - \Gamma(\hat{\varepsilon}) + \frac{\gamma(\hat{\varepsilon})}{\delta_{\hat{\varepsilon}}}\left\{\begin{array}{l} (\frac{1}{\hat{u}_{\hat{\varepsilon}}} - \frac{1}{w_{\hat{\varepsilon}}}) + \frac{1}{8}(\frac{\zeta_4}{\hat{u}_{\hat{\varepsilon}}} - \frac{\zeta_4'}{w_{\hat{\varepsilon}}}) - \frac{5}{24}(\frac{\zeta_3^2}{\hat{u}_{\hat{\varepsilon}}} - \frac{\zeta_3'^2}{w_{\hat{\varepsilon}}}) \\ -\frac{1}{2\sqrt{G^{(2)}}}(\frac{\zeta_3}{\hat{u}_{\hat{\varepsilon}}^2} - \frac{\zeta_3}{w_{\hat{\varepsilon}}^2}) - \frac{1}{G^{(2)}}(\frac{1}{\hat{u}_{\hat{\varepsilon}}^3} - \frac{1}{\hat{w}_{\hat{\varepsilon}}^3}) \end{array}\right\} \tag{42}$$

where

$$\delta_{\hat{\varepsilon}} = 1 + \tfrac{1}{8}\zeta_4' - \tfrac{5}{24}\zeta_3'^2$$

and all of the other variables are defined as in (24) and (25).

When $A^2 = E(A^2) = K^{(1)}(0)$, it can be shown that (41) reduces to

$$\Pr(A^2 \geq E(A^2)) \approx 1 - \Gamma(G^{(1)}(0)) + T_1T_2,$$

where

$$T_1 = \frac{\sqrt{G^{(2)}(0)}\gamma(G^{(1)}(0))}{1 + \frac{1}{8}\zeta_4'(0) - \frac{5}{24}\zeta_3'^2(0)}$$

and

$$\begin{aligned} T_2 &= \tfrac{1}{6}[\zeta_3'(0) - \zeta_3(0)] + \tfrac{1}{40}[\zeta_5'(0) - \zeta_5(0)] \\ &\quad + \tfrac{5}{48}[\zeta_3'(0)\zeta_4'(0) - \zeta_3(0)\zeta_4(0)] + \tfrac{35}{432}[\zeta_3^2(0) - \zeta_3'^2(0)]. \end{aligned} \tag{43}$$

4. Numerical Evaluations

In this section, we provide some numerical results to compare the accuracies of all the saddlepoint approximations discussed in section 3. The programs which generate the numerical results were written in double precision FORTRAN code using the Lahey (1992) F77 compiler. The programs incorporate the "error function" routines from Press *et al.* (1992)

to evaluate the standard normal c.d.f. $\Phi(\cdot)$ and the "gammp" routine to evaluate the gamma function needed for the Chi-Squared distribution, $\Gamma(\cdot)$ above.

Table 1 provides a detailed comparison of all the saddlepoint approximations with Lewis' results for the asymptotic CDF of the Anderson-Darling statistics. This table includes the normal-based distribution and Chi-Squared-based distribution with three different choices of α. Chi-Squared (1) is the chi-squared distribution with α defined by (35); Chi-Squared (2) is the chi-squared distribution with α defined by (36); Chi-Squared (3) is the chi-squared distribution with two degrees of freedom. Giles illustrates the great accuracy of the normal-based-saddlepoint approximation for the tail areas. Here we extend the numerical results to the whole support of the test statistic. From Table 1, we can see Giles' approximation is excellent not only in the tail areas but also over the whole support. The maximum difference between Lewis's and the normal-based approximation is 0.0169, when the value for the A-D test statistic is 0.75 or 0.8. One thing we note here is that the performance of the higher-order formulae for the mean is really poor except for the case of chi-squared (1). The reason for this is not transparent and warrants further investigation.

Next we consider whether the chi-squared-based approximation and the higher-order approximation can improve the accuracy of Giles' normal-based approximation, especially in the middle part of the support. In order to show the comparisons more clearly, we calculate the absolute difference between each saddlepoint approximation and Lewis' results. Then we illustrate these error curves with some graphs. Not surprisingly, the performances of all the approximations except for the case of chi-squared (1) are extremely bad at the mean value, so we delete the point at the mean for each approximation to make the graph more visible.

Figures 1 - 3 provide the comparisons between the normal-based approximation with three types of chi-squared approximations. From the graphs, we can see that the three types of chi-squared-based saddlepoint approximations are strictly better than the normal-based saddlepoint approximation in the middle part of the distribution. However, the normal-based approximation performs better in the tail area, especially in the right tail area.

Figures 4-7 compare each approximation with its corresponding higher-order approximations. The results are similar to those noted already. Each approximation's higher-order correspondence performs strictly better in the middle part of the support. For the first chi-squared distribution, the higher-order approximation improves the approximation almost in the whole interval. However for the other cases, the higher-order approximations perform better in the middle part of the distribution, but they are strictly worse in the right tail areas. Therefore, we conclude that the chi-squared-based saddlepoint approximations do not provide improvement over the whole support compared with the normal-based saddlepoint approximation, even though chi-squared-based saddlepoint approximation seems more close to the true distribution. One possible explanation is that the true distribution is closer to the chi-squared distribution in the center of the support, but not in the tail areas.

More generally, it should be kept in mind that using a higher-order saddlepoint approximation in any context does not necessarily guarantee an improvement for any particular *fixed* finite sample size. Sometime, the corrections from the extra terms worsen the approximation. The improvement from the extra terms can only be guaranteed beyond a certain sample size, but we do not know the exact value of the sample size that is needed for this to occur in practice, as it will vary according to the characteristics of the problem at hand.

5. Concluding Remarks

In this chapter, we examine if the non-normal-based and higher-order saddlepoint approximations can improve Giles' approximation for the asymptotic distribution of the Anderson-Darling distribution. The numerical results show that the chi-squared-based and higher-order approximations perform better than the normal-based saddlepoint approximation in the middle part of the distribution, but not in the tail areas. Therefore, we can conclude that Giles' results are very robust in the tail areas. The accuracy is enough for the interest of the critical point in the tail areas. The approximations we derived are not limited to this case. The distribution function in the non-normal-based formulas could represent any distribution, not just limited to chi-squared distribution. We can apply them to any other interesting problems. For the asymptotic distribution of the Anderson-Darling statistics, the normal-based approximations are very robust. The non-normal-based and higher-order saddlepoint approximations do not bring much improvement. However, this does not mean that they could not achieve better performance in other applications. In future work, we will try to find some interesting cases that benefit from the non-normal-based and higher-order saddlepoint approximations.

Table 1: Values of $F(A^2)$

A^2	Lewis	Normal		Chi-Square (1)		Chi-Square (2)		Chi-Square (3)	
		Low	High	Low	High	Low	High	Low	High
0.100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.125	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
0.150	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014
0.175	0.0042	0.0043	0.0042	0.0043	0.0042	0.0043	0.0042	0.0042	0.0042
0.200	0.0096	0.0099	0.0095	0.0098	0.0095	0.0098	0.0095	0.0096	0.0094
0.225	0.0180	0.0185	0.0178	0.0185	0.0178	0.0184	0.0178	0.0182	0.0177
0.250	0.0296	0.0306	0.0293	0.0305	0.0293	0.0304	0.0292	0.0301	0.0290
0.275	0.0443	0.0459	0.0439	0.0457	0.0438	0.0456	0.0437	0.0453	0.0434
0.300	0.6180	0.6410	0.6110	0.0639	0.0611	0.0638	0.0609	0.0635	0.0605
0.325	0.0817	0.0849	0.0807	0.0845	0.0806	0.0844	0.0805	0.0841	0.0799
0.350	0.1036	0.1077	0.1021	0.1072	0.1020	0.1071	0.1019	0.1068	0.1011
0.375	0.1269	0.1320	0.1250	0.1314	0.1248	0.1314	0.1247	0.1311	0.1239
0.400	0.1513	0.1576	0.1489	0.1568	0.1487	0.1567	0.1486	0.1565	0.1477
0.425	0.1764	0.1838	0.1736	0.1829	0.1733	0.1829	0.1733	0.1826	0.1722
0.450	0.2019	0.2105	0.1987	0.2094	0.1984	0.2094	0.1984	0.2092	0.1972
0.475	0.2276	0.2374	0.2241	0.2361	0.2236	0.2361	0.2237	0.2359	0.2225
0.500	0.2532	0.2641	0.2494	0.2626	0.2489	0.2627	0.2491	0.2625	0.2478
0.525	0.2786	0.2905	0.2747	0.2889	0.2740	0.2889	0.2743	0.2888	0.2730
0.550	0.3036	0.3165	0.2997	0.3147	0.2989	0.3148	0.2993	0.3146	0.2979
0.575	0.3281	0.3419	0.3244	0.3399	0.3234	0.3400	0.3239	0.3399	0.3226
0.600	0.3520	0.3666	0.3486	0.3645	0.3475	0.3646	0.3482	0.3645	0.3468
0.625	0.3753	0.3906	0.3723	0.3884	0.3711	0.3885	0.3719	0.3884	0.3705
0.675	0.4199	0.4362	0.4180	0.4338	0.4164	0.4339	0.4176	0.4338	0.4162
0.700	0.4412	0.4577	0.4400	0.4553	0.4382	0.4554	0.4395	0.4553	0.4162
0.750	0.4815	0.4984	0.4818	0.4959	0.4796	0.4959	0.4814	0.4958	0.4800
0.800	0.5190	0.5359	0.5208	0.5333	0.5182	0.5333	0.5204	0.5332	0.5190
0.850	0.5537	0.5703	0.5569	0.5677	0.5541	0.5676	0.5565	0.5675	0.5552
0.900	0.5858	0.6018	0.5903	0.5993	0.5871	0.5991	0.5899	0.5990	0.5886
0.950	0.6154	0.6307	0.6210	0.6283	0.6176	0.6279	0.6206	0.6279	0.6194
1.000	0.6427	0.6571	0.8952	0.6549	0.6455	0.6544	0.8560	0.6544	0.7651
1.050	0.6680	0.6814	0.6743	0.6794	0.6712	0.6787	0.6746	0.6787	0.6735
1.100	0.6912	0.7037	0.6982	0.7018	0.6948	0.7011	0.6982	0.7011	0.6972
1.150	0.7127	0.7242	0.7202	0.7225	0.7164	0.7216	0.7199	0.7216	0.7189
1.200	0.7324	0.7430	0.7401	0.7415	0.7362	0.7405	0.7398	0.7405	0.7389
1.250	0.7580	0.7603	0.7584	0.7590	0.7545	0.7579	0.7581	0.7580	0.7572
1.300	0.7677	0.7763	0.7751	0.7752	0.7713	0.7740	0.7749	0.7741	0.7740
1.350	0.7833	0.7911	0.7906	0.7901	0.7868	0.7889	0.7903	0.7889	0.7895

Table 1: ***Continued***

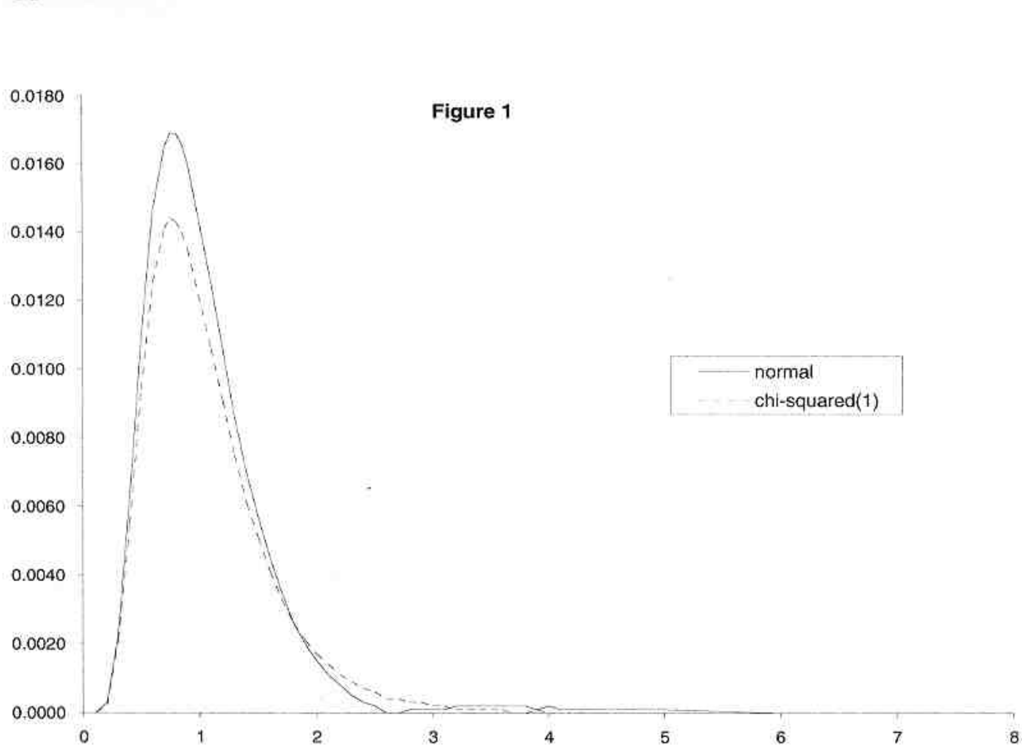
A^2	Lewis	Normal		Chi-Squared (1)		Chi-Squared(2)		Chi-Squared(3)	
		Low	High	Low	High	Low	High	Low	High
1.400	0.7978	0.8048	0.8048	0.8039	0.8011	0.8026	0.8045	0.8027	0.8038
1.450	0.8111	0.8174	0.8179	0.8167	0.8143	0.8153	0.8177	0.8155	0.8169
1.500	0.8235	0.8292	0.8300	0.8286	0.8265	0.8272	0.8298	0.8273	0.8291
1.550	0.8350	0.8401	0.8412	0.8396	0.8378	0.8381	0.8410	0.8383	0.8403
1.600	0.8457	0.8502	0.8515	0.8498	0.8482	0.8483	0.8513	0.8485	0.8507
1.650	0.8556	0.8596	0.8611	0.8593	0.8579	0.8578	0.8609	0.8580	0.8603
1.700	0.8648	0.8683	0.8700	0.8681	0.8670	0.8666	0.8698	0.8668	0.8693
1.750	0.8734	0.8765	0.8783	0.8764	0.8753	0.8749	0.8781	0.8750	0.8776
1.800	0.8814	0.8840	0.8860	0.8840	0.8831	0.8825	0.8858	0.8827	0.8853
1.850	0.8888	0.8911	0.8931	0.8912	0.8904	0.8897	0.8929	0.8899	0.8925
1.900	0.8957	0.8977	0.8997	0.8978	0.8972	0.8964	0.8996	0.8965	0.8992
1.950	0.9021	0.9039	0.9059	0.9041	0.9035	0.9026	0.9058	0.9028	0.9054
2.000	0.9082	0.9097	0.9117	0.9099	0.9094	0.9084	0.9116	0.9086	0.9112
2.050	0.9138	0.9151	0.9171	0.9153	0.9149	0.9139	0.9170	0.9141	0.9166
2.100	0.9190	0.9201	0.9221	0.9204	0.9200	0.9190	0.9220	0.9192	0.9217
2.150	0.9239	0.9249	0.9268	0.9252	0.9248	0.9238	0.9267	0.9240	0.9264
2.200	0.9285	0.9293	0.9312	0.9296	0.9293	0.9283	0.9311	0.9285	0.9308
2.250	0.9328	0.9334	0.9353	0.9338	0.9335	0.9325	0.9352	0.9327	0.9349
2.300	0.9368	0.9373	0.9392	0.9377	0.9374	0.9364	0.9391	0.9366	0.9388
2.350	0.9405	0.9410	0.9428	0.9413	0.9411	0.9401	0.9427	0.9403	0.9424
2.400	0.9441	0.9444	0.9462	0.9448	0.9446	0.9436	0.9461	0.9438	0.9458
2.450	0.9474	0.9476	0.9493	0.9480	0.9478	0.9468	0.9493	0.9470	0.9490
2.500	0.9504	0.9506	0.9523	0.9510	0.9509	0.9499	0.9522	0.9501	0.9520
2.550	0.9534	0.9535	0.9551	0.9539	0.9537	0.9528	0.9550	0.9529	0.9548
2.600	0.9561	0.9561	0.9577	0.9565	0.9564	0.9555	0.9576	0.9556	0.9574
2.650	0.9586	0.9586	0.9601	0.9590	0.9589	0.9580	0.9601	0.9582	0.9599
2.700	0.9610	0.9610	0.9624	0.9614	0.9613	0.9604	0.9624	0.9606	0.9622
2.750	0.9633	0.9632	0.9646	0.9636	0.9635	0.9627	0.9645	0.9628	0.9644
2.800	0.9654	0.9653	0.9666	0.9657	0.9656	0.9648	0.9666	0.9649	0.9664
2.850	0.9674	0.9673	0.9685	0.9676	0.9676	0.9668	0.9685	0.9669	0.9683
2.900	0.9692	0.9691	0.9703	0.9695	0.9694	0.9686	0.9703	0.9688	0.9701
2.950	0.9710	0.9708	0.9720	0.9712	0.9711	0.9704	0.9720	0.9705	0.9718
3.000	0.9726	0.9725	0.9736	0.9728	0.9728	0.9720	0.9736	0.9722	0.9734
3.050	0.9742	0.9740	0.9751	0.9744	0.9743	0.9736	0.9751	0.9737	0.9749
3.100	0.9756	0.9755	0.9765	0.9758	0.9758	0.9751	0.9765	0.9752	0.9764
3.150	0.9770	0.9768	0.9778	0.9772	0.9771	0.9765	0.9778	0.9766	0.9777
3.200	0.9783	0.9781	0.9791	0.9784	0.9784	0.9778	0.9790	0.9779	0.9789
3.250	0.9795	0.9793	0.9803	0.9796	0.9796	0.9790	0.9802	0.9791	0.9801

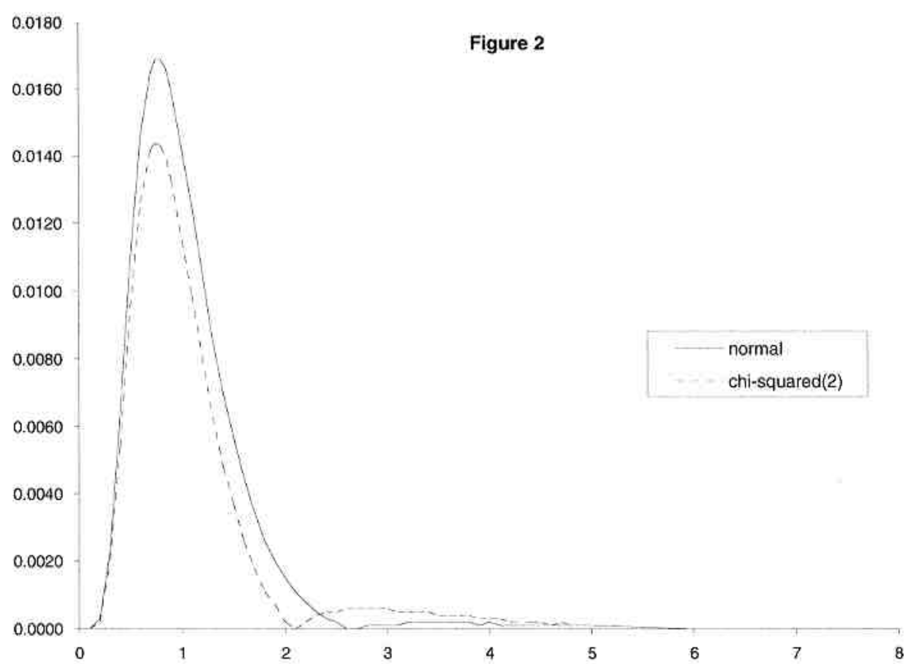
Table 1: *Continued*

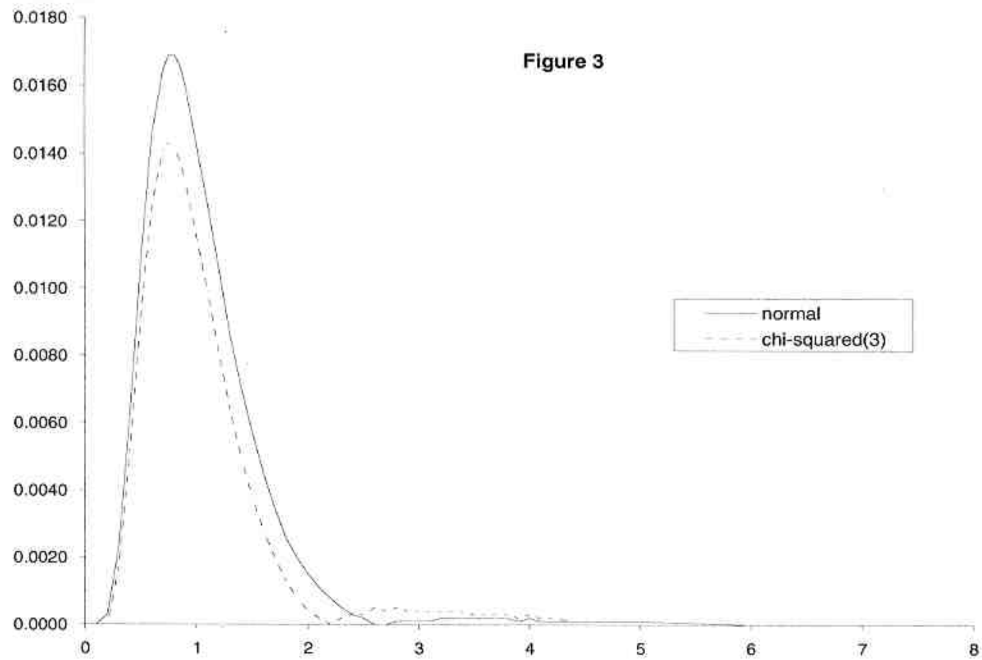
A^2	Lewis	Normal		Chi-Squared (1)		Chi-Squared(2)		Chi-Squared(3)	
		Low	High	Low	High	Low	High	Low	High
3.300	0.9807	0.9805	0.9814	0.9808	0.9808	0.9802	0.9813	0.9803	0.9812
3.350	0.9818	0.9816	0.9824	0.9819	0.9818	0.9813	0.9824	0.9814	0.9823
3.400	0.9828	0.9826	0.9834	0.9829	0.9828	0.9823	0.9834	0.9824	0.9833
3.450	0.9837	0.9835	0.9843	0.9838	0.9838	0.9833	0.9843	0.9834	0.9842
3.500	0.9846	0.9844	0.9852	0.9847	0.9847	0.9842	0.9852	0.9843	0.9851
3.550	0.9855	0.9853	0.9860	0.9856	0.9855	0.9851	0.9860	0.9852	0.9859
3.600	0.9863	0.9861	0.9868	0.9864	0.9863	0.9859	0.9867	0.9860	0.9867
3.650	0.9870	0.9869	0.9875	0.9871	0.9871	0.9867	0.9875	0.9867	0.9874
3.700	0.9878	0.9876	0.9882	0.9878	0.9878	0.9874	0.9882	0.9875	0.9881
3.750	0.9884	0.9883	0.9888	0.9885	0.9885	0.9881	0.9888	0.9881	0.9888
3.800	0.9891	0.9889	0.9895	0.9891	0.9891	0.9887	0.9894	0.9888	0.9894
3.850	0.9897	0.9895	0.9900	0.9897	0.9897	0.9893	0.9900	0.9894	0.9900
3.900	0.9902	0.9901	0.9906	0.9903	0.9903	0.9899	0.9906	0.9900	0.9905
3.950	0.9908	0.9906	0.9911	0.9908	0.9908	0.9905	0.9911	0.9905	0.9910
4.000	0.9913	0.9911	0.9916	0.9913	0.9913	0.9910	0.9916	0.9910	0.9915
4.050	0.9917	0.9916	0.9920	0.9918	0.9918	0.9915	0.9920	0.9915	0.9920
4.100	0.9922	0.9921	0.9925	0.9922	0.9922	0.9919	0.9925	0.9920	0.9924
4.150	0.9926	0.9925	0.9929	0.9926	0.9926	0.9924	0.9929	0.9924	0.9928
4.200	0.9930	0.9929	0.9933	0.9930	0.9930	0.9928	0.9933	0.9928	0.9932
4.250	0.9934	0.9933	0.9936	0.9934	0.9934	0.9932	0.9936	0.9932	0.9936
4.300	0.9938	0.9936	0.9940	0.9938	0.9938	0.9935	0.9940	0.9936	0.9939
4.350	0.9941	0.9940	0.9943	0.9941	0.9941	0.9939	0.9943	0.9939	0.9943
4.400	0.9944	0.9943	0.9946	0.9944	0.9944	0.9942	0.9946	0.9943	0.9946
4.500	0.9950	0.9949	0.9952	0.9950	0.9950	0.9948	0.9952	0.9949	0.9951
4.600	0.9955	0.9954	0.9957	0.9955	0.9955	0.9954	0.9957	0.9954	0.9957
4.700	0.9960	0.9959	0.9961	0.9960	0.9960	0.9958	0.9961	0.9959	0.9961
4.800	0.9964	0.9963	0.9965	0.9964	0.9964	0.9963	0.9965	0.9963	0.9965
4.900	0.9968	0.9967	0.9969	0.9968	0.9968	0.9967	0.9969	0.9967	0.9969
5.000	0.9971	0.9970	0.9972	0.9971	0.9971	0.9970	0.9972	0.9970	0.9972
5.500	0.9983	0.9983	0.9984	0.9983	0.9983	0.9983	0.9984	0.9983	0.9984
6.000	0.9990	0.9990	0.9991	0.9990	0.9990	0.9990	0.9991	0.9990	0.9991
7.000	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
8.000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Note: "Low" represent the lower-order saddlepoint approximation; "High" represent the higher-order saddlepoint approximation.

Comparison of the Normal-Based and Chi-Squared-Based Saddlepoint Approximation

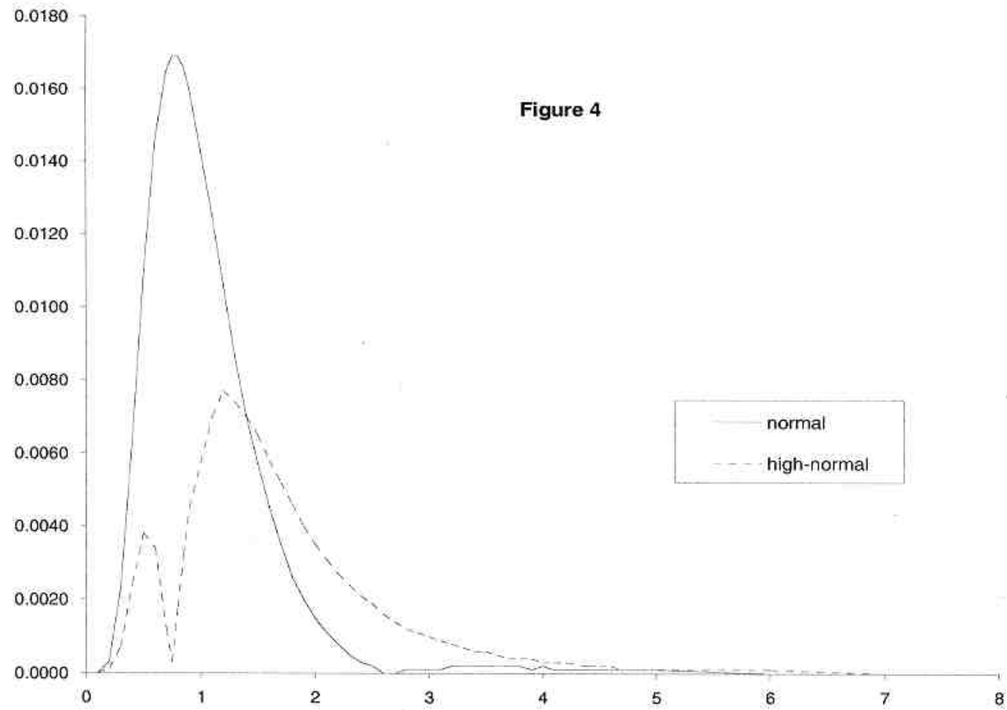


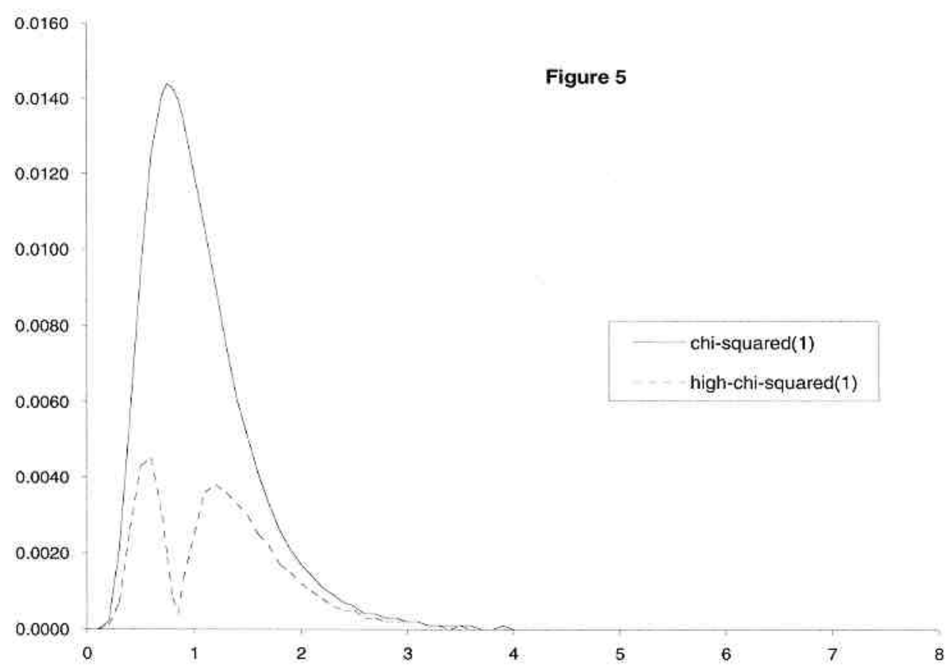


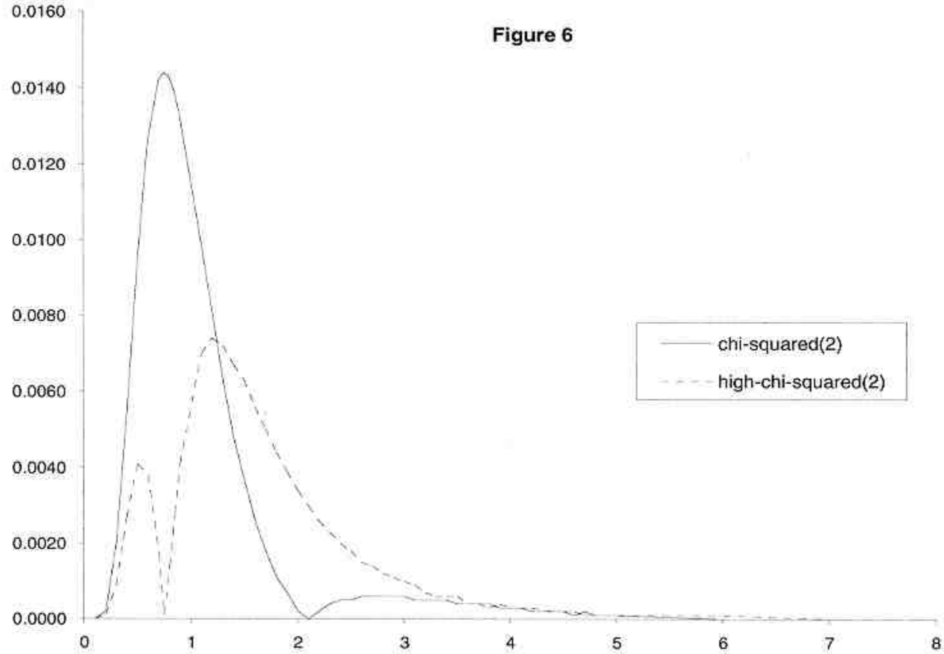


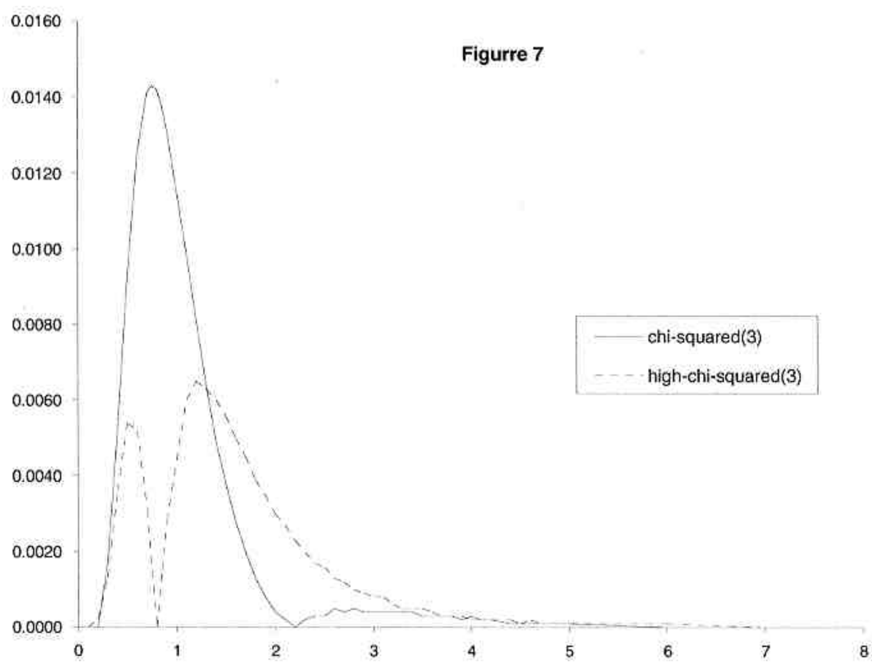
Figures 1-3 compare the normal-based and chi-squared-based distribution. The horizontal axis represents the value of the Anderson-Darling test statistic; the vertical axis represents the absolute error of these two distributions compared with Lewis' results. Chi-squared (1) is the chi-squared distribution with α defined by (34); chi-squared (2) is the chi-squared distribution with α defined by (35); chi-squared (3) is the chi-squared distribution with degrees of freedom of 2.

Comparison of the Lower-Order and Higher-Order Saddlepoint Approximations









Figures 4 -7 compare the lower-order and higher-order saddlepoint approximations. The horizontal axis represents the value of the Anderson-Darling test statistic; the vertical axis represents the absolute error of these two distributions compared with Lewis' results. Chi-squared (1) is the chi-squared distribution with α defined by (34); chi-squared (2) is the chi-squared distribution with α defined by (35); chi-squared (3) is the chi-squared distribution with degrees of freedom of 2.

Appendix: Proof of higher-order WBB saddlepoint approximation

From the Inversion Formula for the tail probability:

$$P(X \geq y) = \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{K(t)-ty}\frac{dt}{t} \tag{44}$$

let

$$G(w_{\varepsilon}) - \hat{\varepsilon}w_{\varepsilon} = K(t) - ty \tag{45}$$

$$P(X \geq y) = \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\hat{\varepsilon}w_{\varepsilon}}\left(\frac{dt}{tdw_{\varepsilon}}\right)dw_{\varepsilon}$$

$$= \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\left(\frac{1}{w_{\varepsilon}} + \frac{dt}{tdw_{\varepsilon}} - \frac{1}{w_{\varepsilon}}\right)dw_{\varepsilon}$$

$$= \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\frac{1}{w_{\varepsilon}}dw_{\varepsilon} + \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\left(\frac{dt}{tdw_{\varepsilon}} - \frac{1}{w_{\varepsilon}}\right)dw_{\varepsilon} \tag{46}$$

Again, from the Inversion Formula, the first part of (46) is:

$$\frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\hat{\varepsilon}w_{\varepsilon}}\frac{1}{w_{\varepsilon}}dw_{\varepsilon} = 1 - \Gamma(\hat{\varepsilon})$$

then the second part of (46) can be written as:

$$\frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\left(\frac{dt}{tdw_{\varepsilon}} - \frac{1}{w_{\varepsilon}}\right)dw_{\varepsilon}$$

$$= \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\left(\frac{dt}{tdw_{\varepsilon}}\right)dw_{\varepsilon} - \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\frac{dw_{\varepsilon}}{w_{\varepsilon}}.$$

Let

$$I_1 = \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\left(\frac{dt}{tdw_{\varepsilon}}\right)dw_{\varepsilon}$$

$$I_2 = \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{G(w_{\varepsilon})-\varepsilon w_{\varepsilon}}\frac{dw_{\varepsilon}}{w_{\varepsilon}}$$

from (45), I_1 can be written as:

$$I_1 = \frac{1}{2\pi i}\int_{\hat{\varepsilon}-i\infty}^{\hat{\varepsilon}+i\infty} e^{K(t)-ty}\frac{dt}{t}.$$

Daniels used a routine application of the saddlepoint method to the inversion formula (44) to get the higher-order saddlepoint approximation. Using Daniels' result (3.3) (1987), we can get:

$$I_1 = \frac{e^{K-\hat{t}y}}{\hat{t}\sqrt{2\pi K^{(2)}}}(1+\frac{1}{8}\zeta_4 - \frac{5}{24}\zeta_3^2 - \frac{\zeta_3}{2\hat{t}(K^{(2)})^{1/2}} - \frac{1}{\hat{t}^2 K^{(2)}})$$

$$I_2 = \frac{e^{G(w_{\hat{\varepsilon}})-\hat{\varepsilon}w_{\hat{\varepsilon}}}}{w_{\hat{\varepsilon}}\sqrt{2\pi G^{(2)}}}(1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2 - \frac{\zeta_3'}{2w_{\hat{\varepsilon}}(G^{(2)})^{1/2}} - \frac{1}{w_{\hat{\varepsilon}}^2 G^{(2)}}).$$

Here, we need to use some results from Daniels. Daniels' saddlepoint approximation to the density function at any specified point y up to b_1 term is:

$$f(y) = g(y)\left\{1+\tfrac{1}{8}\zeta_4(\hat{t}) - \tfrac{5}{24}\zeta_3^2(\hat{t}) + \ldots\right\} \tag{47}$$

we let:

$$\gamma(\hat{\varepsilon}) = \frac{e^{G(w_{\hat{\varepsilon}})-\hat{\varepsilon}w_{\hat{\varepsilon}}}}{\sqrt{2\pi G^{(2)}}}\left\{1+\tfrac{1}{8}\zeta_4'(\hat{t}) - \tfrac{5}{24}\zeta_3'^2(\hat{t}) + \ldots\right\}. \tag{48}$$

Now we can rewrite I_1 and I_2 as:

$$I_1 = \frac{e^{G(w_{\hat{\varepsilon}})-\hat{\varepsilon}w_{\hat{\varepsilon}}}}{\sqrt{2\pi G^{(2)}}}\frac{\{1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2\}}{\{1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2\}}\frac{\sqrt{2\pi G^{(2)}}}{\hat{t}\sqrt{2\pi K^{(2)}}}(1+\frac{1}{8}\zeta_4 - \frac{5}{24}\zeta_3^2 - \frac{\zeta_3}{2\hat{t}(K^{(2)})^{1/2}} - \frac{1}{\hat{t}^2 K^{(2)}})$$

$$= \gamma(\hat{\varepsilon})\frac{\sqrt{G^{(2)}}}{\hat{t}\sqrt{K^{(2)}}\{1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2\}}(1+\frac{1}{8}\zeta_4 - \frac{5}{24}\zeta_3^2 - \frac{\zeta_3}{2\hat{t}(K^{(2)})^{1/2}} - \frac{1}{\hat{t}^2 K^{(2)}})$$

$$I_2 = \frac{e^{G(w_{\hat{\varepsilon}})-\hat{\varepsilon}w_{\hat{\varepsilon}}}}{\sqrt{2\pi G^{(2)}}}\frac{\{1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2\}}{w_{\hat{\varepsilon}}\{1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2\}}(1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2 - \frac{\zeta_3'}{2w_{\hat{\varepsilon}}(G^{(2)})^{1/2}} - \frac{1}{w_{\hat{\varepsilon}}^2 G^{(2)}})$$

$$= \gamma(\hat{\varepsilon})\frac{1}{w_{\hat{\varepsilon}}\{1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2\}}(1+\frac{1}{8}\zeta_4' - \frac{5}{24}\zeta_3'^2 - \frac{\zeta_3'}{2w_{\hat{\varepsilon}}(G^{(2)})^{1/2}} - \frac{1}{w_{\hat{\varepsilon}}^2 G^{(2)}}).$$

Let

$$\hat{u}_{\hat{\varepsilon}} = \hat{t}\sqrt{K^{(2)}}[G^{(2)}]^{-1/2}$$

$$\delta_{\hat{\varepsilon}} = 1+\tfrac{1}{8}\lambda_4' - \tfrac{5}{24}\lambda_3'^2.$$

Then,

$$P(X \geq y) = 1-\Gamma(\hat{\varepsilon}) + I_1 - I_2$$

$$\approx 1-\Gamma(\hat{\varepsilon}) + \gamma(\hat{\varepsilon})\frac{1}{\hat{u}_{\hat{\varepsilon}}\delta_{\hat{\varepsilon}}}(1+\frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2 - \frac{\lambda_3}{2\hat{t}(K^{(2)})^{1/2}} - \frac{1}{\hat{t}^2 K^{(2)}})$$

$$- \gamma(\hat{\varepsilon})\frac{1}{\hat{w}\delta_{\hat{\varepsilon}}}(1+\frac{1}{8}\lambda_4' - \frac{5}{24}\lambda_3'^2 - \frac{\lambda_3'}{2\hat{w}(G^{(2)})^{1/2}} - \frac{1}{\hat{w}^2 G^{(2)}})$$

$$=1-\Gamma(\hat{\varepsilon})+\frac{\gamma(\hat{\varepsilon})}{\delta_{\hat{\varepsilon}}}\left[\begin{array}{l}(\frac{1}{\hat{u}_{\hat{\varepsilon}}}-\frac{1}{w_{\hat{\varepsilon}}})+\frac{1}{8}(\frac{\lambda_4}{\hat{u}_{\hat{\varepsilon}}}-\frac{\lambda_4'}{w_{\hat{\varepsilon}}})-\frac{5}{24}(\frac{\lambda_3^2}{\hat{u}_{\hat{\varepsilon}}}-\frac{\lambda_3'^2}{w_{\hat{\varepsilon}}}) \\ -\frac{1}{2\sqrt{G^{(2)}}}(\frac{\lambda_3}{\hat{u}_{\hat{\varepsilon}}^2}-\frac{\lambda_3'}{w_{\hat{\varepsilon}}^2})-\frac{1}{G^{(2)}}(\frac{1}{\hat{u}_{\hat{\varepsilon}}^3}-\frac{1}{\hat{w}_{\hat{\varepsilon}}^3})\end{array}\right]. \tag{49}$$

When $X=E(X)$, $w_{\hat{\varepsilon}}=0$ and $\hat{u}_{\varepsilon}=0$. Therefore, we need to take the limit of (49). First, (49) can be written as:

$$\Pr(X\geq y)=S_1+\frac{\gamma(\hat{\varepsilon})}{\delta_{\hat{\varepsilon}}}\sqrt{G^{(2)}}\left(S_2-S_3+S_4\right) \tag{50}$$

where

$$S_1=1-\Gamma(\hat{\varepsilon})+\frac{\gamma(\hat{\varepsilon})}{\delta_{\hat{\varepsilon}}}(\frac{1}{\hat{u}_{\hat{\varepsilon}}}-\frac{1}{w_{\hat{\varepsilon}}})$$

$$S_2=\frac{\zeta_4}{8\hat{t}\sqrt{K^{(2)}}}-\frac{5\zeta_3^2}{24\hat{t}\sqrt{K^{(2)}}}-\frac{\zeta_3}{2\hat{t}^2K^{(2)}}-\frac{1}{\hat{t}^3(K^{(2)})^{3/2}}+\frac{1}{[2(\hat{t}K^{(1)}-K(\hat{t})]^{3/2}}$$

$$S_3=\frac{\zeta_4'}{8w_{\hat{\varepsilon}}\sqrt{G^{(2)}}}-\frac{5\zeta_3'^2}{24w_{\hat{\varepsilon}}\sqrt{G^{(2)}}}-\frac{\zeta_3'}{2w_{\hat{\varepsilon}}^2G^{(2)}}-\frac{1}{\hat{w}_{\hat{\varepsilon}}^3(G^{(2)})^{3/2}}+\frac{1}{[2(w_{\hat{\varepsilon}}G^{(1)}-G(w_{\hat{\varepsilon}})]^{3/2}}$$

$$S_4=\frac{1}{[2(w_{\hat{\varepsilon}}G^{(1)}-G(w_{\hat{\varepsilon}})]^{3/2}}-\frac{1}{[2(\hat{t}K^{(1)}-K(\hat{t})]^{3/2}}.$$

We can see S_1 is the same as the WBB forumula. Therefore the limit of S_1 is (24). Then we need to calculate the limit of the remaining parts of (50).

S_2 and S_3 are the same with b_1 in (38). Therefore, taking Daniels' result, we can get the limit of S_2 and S_3 as:

$$\underset{\hat{t}\to 0}{Lim}\,S_2=\tfrac{1}{40}\zeta_5-\tfrac{5}{48}\zeta_3\zeta_4+\tfrac{35}{432}\zeta_3^2$$

$$\underset{\hat{t}\to 0}{Lim}\,S_3=\tfrac{1}{40}\zeta_5'-\tfrac{5}{48}\zeta_3'\zeta_4'+\tfrac{35}{432}\zeta_3'^2.$$

Then we apply L'Hôpital's rule to get the limit of S_4.

$$\underset{\hat{t}\to 0}{Lim}\,S_4=\underset{\hat{t}\to 0}{Lim}\frac{[\hat{t}K^{(1)}-K(\hat{t})]^{3/2}-[w_{\hat{\varepsilon}}G^{(1)}-G(w_{\hat{\varepsilon}})]^{3/2}}{2\sqrt{2}[w_{\hat{\varepsilon}}G^{(1)}-G(w_{\hat{\varepsilon}})]^{3/2}[\hat{t}K^{(1)}-K(\hat{t})]^{3/2}}. \tag{51}$$

Differentiating both denominator and numerator of (50):

$$\underset{\hat{t}\to 0}{Lim}\,S_4$$

$$=\underset{\hat{t}\to 0}{Lim}\frac{1}{2\sqrt{2}}\frac{\frac{3}{2}\hat{t}K^{(2)}\sqrt{\hat{t}K^{(1)}-K}-\frac{3}{2}\hat{t}K^{(2)}\sqrt{w_{\hat{\varepsilon}}G^{(1)}-G}}{\frac{3}{2}\hat{t}K^{(2)}\sqrt{\hat{t}K^{(1)}-K}[w_{\hat{\varepsilon}}G^{(1)}-G]^{3/2}+\frac{3}{2}\hat{t}K^{(2)}\sqrt{w_{\hat{\varepsilon}}G^{(1)}-G}[\hat{t}K^{(1)}-K]^{3/2}}$$

$$= \underset{\hat{t}\to 0}{Lim} \frac{1}{2\sqrt{2}} \frac{\sqrt{\hat{t}K^{(1)} - K} - \sqrt{w_{\hat{\varepsilon}}G^{(1)} - G}}{\sqrt{\hat{t}K^{(1)} - K}[w_{\hat{\varepsilon}}G^{(1)} - G]^{3/2} + \sqrt{w_{\hat{\varepsilon}}G^{(1)} - G}[\hat{t}K^{(1)} - K]^{3/2}}$$

Differentiating again and applying L'Hôpital's rule a second time :

$$\underset{\hat{t}\to 0}{Lim}\, S_4 = \underset{\hat{t}\to 0}{Lim} \frac{1}{2\sqrt{2}} \frac{\sqrt{w_{\hat{\varepsilon}}G^{(1)} - G} - \sqrt{\hat{t}K^{(1)} - K}}{(w_{\hat{\varepsilon}}G^{(1)} - G)^2 + 6(\hat{t}K^{(1)} - K)(w_{\hat{\varepsilon}}G^{(1)} - G) + (\hat{t}K^{(1)} - K)^2}$$

$$\underset{\hat{t}\to 0}{Lim}\, S_4 = \underset{\hat{t}\to 0}{Lim} - \frac{64}{2\sqrt{2}} \left[\frac{1}{(w_{\hat{\varepsilon}}G^{(1)} - G)^{3/2}} - \frac{1}{(\hat{t}K^{(1)} - K)^{3/2}} \right] = \underset{\hat{t}\to 0}{Lim} - 64 S_4$$

so that $\underset{\hat{t}\to 0}{Lim}\, S_4 = 0$.

Thus, when $X = E(X)$, we obtain the formula in (43).

CHAPTER 4:
THE FINITE-SAMPLE MOMENTS OF THE MLE FOR THE BINARY LOGIT MODEL

1. Introduction

Qualitative Response (QR) models are very widely used in empirical economics, and in many other areas of application. There are many different settings for the QR models, but the common feature for all the QR models is that the dependent variable is qualitative, rather than quantitative. Then to make the model estimable, these qualitative attributes are "coded" numerically so as to partition the sample data appropriately. For example, the decision of whether or not to pursue a Ph.D. degree involves a qualitative, "yes – no" choice, as does deciding whether or not to accept a job offer. A further example would be the decision over the mode of transport to use to get to school – perhaps, walking, taking the bus, or using a bicycle. In this case there are three qualitative choices, and we have a multinomial situation. There is a vast and readily accessible literature relating to inference in the context of qualitative response models and it is not our intention to survey this literature here. For example, see Maddala (1983), Wooldridge (2002), Hensher *et al.* (2005) and Cameron and Trivedi (2005).

The binary choice model is the most widely used of the QR models. In the binary choice model, the dependent variable is coded as unity or zero if a certain event occurs or not. In this case, it is well known that the conventional (linear) regression methods are inappropriate. For instance, the predicted probabilities that are generated by the linear probability model can be negative, or exceed unity. In addition, the error must be heteroskedastic, and standard inferences are confounded by the fact that the error term clearly cannot follow a normal distribution. These problems can be overcome by making the probability of the event of interest occurring (*i.e.*, the probability that the dependent variable is assigned the value of unity) a non-linear, rather than a linear function of the covariates. In particular, if this function is taken to be a cumulative distribution function, it will be monotonically non-decreasing, and bounded between zero and unity. This then prevents nonsensical predicted probabilities from arising. Usually, the distribution that is chosen for this purpose is the normal distribution (which gives rise to the so-called Probit model), the logistic distribution (which gives us the

Logit model), the Weibull distribution, or the extreme value distribution. Different distributions lead to different non-linear models with somewhat different features. The Logit and Probit models are the two that are encountered most frequently in practice, and they generally yield similar estimates. The former has a computational advantage over the latter - the distribution function can be written in closed form, rather than having to be expressed as an integral.

For both the Probit and Logit models the likelihood function can be shown to be strictly concave, so it has a unique maximum. The asymptotic properties of the Maximum Likelihood Estimators (MLE) of the parameters in the QR models are standard. The likelihood functions satisfy the usual regularity conditions, so these MLE's are weakly consistent and "Best Asymptotically Normal". However, it is surprising that there have been very few studies of the finite sample properties of the MLE in the QR models. In this chapter, we derive analytic expressions for some of the finite sample properties of the MLE of the parameter vector in the Logit model. The approach that we use could also be used to extend our results to other QR models.

As we stated above, little work has been done in this field. One known work is Amemiya (1980). He derived the n^{-2}-order mean squared error (MSE) of MLE and the minimum chi-square estimators (MCS) of the dichotomous Logit model and provided some numerical results on the relative quality of these two estimators, based on their MSE's. The MCS estimator was first introduced by Berkson (1944) for the dichotomous logit model, which is the model in which there are a number of observations of the dependent variable for each value of the independent variables. Taylor (1953) showed that the MCS estimator and the MLE have the same asymptotic normal distribution when the number of the observations of the dependent variable for each value of the independent variables goes to infinity. Berkson (1955) approximate the finite-sample bias and MSE of the MLE and the MCS estimator for four simple models, and showed that the MCS is preferred to the MLE in terms of MSE in all four of these cases. Following Amemiya's work, several studies made some further advances on Berkson and Amemiya's results. For example, Ghosh and Sinha (1981) provided the theory to give necessary and sufficient conditions for improving the MSE of the MLE, and applied this to Berkson's dichotomous Logit model. They also showed the preference between the MLE and the MCS estimator in terms of MSE actually depends on the models selected.

Davis (1984) found some examples in which the MLE has smaller MSE than the MCS estimator, and Hughes and Savin (1994) provided further results indicating that the choice between these two estimators is not straightforward. Another somewhat related study is that of Mackinnon and Smith (1998). They discussed methods for reducing the bias of consistent estimators that are biased in finite samples, and applied their methods to the parameter estimator in the AR(1) model and the Logit model. Finally, Li (2005) used a Monte Carlo experiment to examine the small sample properties of the MLE for three different models - the Probit model, the Logit model and the binary choice model where the underlying distribution is the Extreme Value distribution. From the results of her Monte Carlo experiment, Li reached two main conclusions regarding the root mean squared percentage error of each estimator. She found that the Probit MLE ranks first, while the Logit MLE is the least preferred one, on this basis. She also found that if the underlying distributional process is mis-specified, this increases the MSE for each of the estimators.

In this chapter, we will apply Rilstone *et al.*'s (1996) results to derive analytic expressions for the bias and MSE functions for the MLE in the Logit model. This approach was also used by Rilstone and Ullah (2002) in the context of Heckman's sample selection estimator. Based on the analytic bias and MSE expressions, we can derive a bias-corrected estimator and the standard error associated with the bias-corrected estimator. We also provide some numerical results based on these analytic results. The numerical results show that the bias correction works very well even based on the estimated results, instead of the exact results. In order to apply Rilstone *et al.*'s results, we need to assume that both the dependent and independent variables are random and the observations are i.i.d., which makes our results incomparable with Amemiya's and Davis's results. Because all of the observations are i.i.d random, the expectations of the random variables or any function of the random variables are the same for all of the observations, which simplifies the derivation of Rilstone *et al*'s results. On the other hand, if the independent variables are fixed, then the expectations of the independent variables or any function involving the independent variables are different for different observations, which complicates the derivations to the extent that they are sample specific. Although in one sense the case of random independent variables includes the case of fixed independent variables as a special case, it also gives rise to some practical issues of implementation. In future work, we plan to consider this issue further.

The next section introduces the Logit model. Then, in section 3 we introduce Rilstone *et al.*'s results and derive the analytic results for the bias and MSE of the MLE in the Logit model. Some numerical results follow in section 4. The final section provides our conclusions and suggestions for further research.

2. The Logit Model and the Maximum Likelihood Estimator

In a binary choice model, if the response to an event is "yes", we assign a value of "1" to the dependent variable; while if the response is "no", we assign "0" to the dependent variable. We use a latent dependent variable to represent the total effect from a series of factors (or covariates) that affect the decision. The latent regression is:

$$y_i^* = X_i'\beta + \varepsilon \ , \tag{1}$$

where y^* is the latent dependent variable, and the row vector, X_i', represents the i^{th} observation on all of the factors (covariates) which affect the person's choice.

Then, the dependent variable can be defined as,

$$\begin{aligned} y_i &= 1; \quad \text{if } X_i'\beta + \varepsilon \geq a \\ y_i &= 0; \quad \text{if } X_i'\beta + \varepsilon < a \end{aligned} \tag{2}$$

where a is the threshold. As is well understood, as long as an intercept is included among the regressors, the threshold value for determining the dependent variable is actually irrelevant, and may be set to zero. Then, (1) and (2) can be simplified to

$$y_i^* = X_i'\beta_i + \varepsilon$$

and

$$\begin{aligned} y_i &= 1; \quad \text{if } y_i^* \geq 0 \\ y_i &= 0; \quad \text{if } y_i^* < 0 . \end{aligned} \tag{3}$$

The values of "1"and "0" that are assigned to y_i have no quantitative meaning – they are simply categorical scores that enable us to partition the sample into two parts. What is of interest is the probability of the person ending up with one choice or the other. So the basic model can be structured as follows:

$$P_i = \Pr(y_i = 1|X) = F(X_i'\beta)$$

$$1 - P_i = \Pr(y_i = 0|X) = 1 - F(X_i'\beta) .$$

The form of the cumulative distribution function, $F(X_i'\beta)$, will determine which particular model we end up with. In this chapter, we focus on the Logit model. Therefore, the model we study is as follows:

$$P_i = \Pr(y_i = 1|X) = F(X_i'\beta) = \Lambda_i \tag{4}$$

where

$$\Lambda_i = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)} \tag{5}$$

is the c.d.f. for the Logistic distribution.

The MLE for the parameter vector in (4) is derived as follows. The (conditional) joint data density function for the sample is:

$$\Pr(Y_1 = y_1, Y_2 = y_2, \cdots, Y_n = y_n|\beta, X) = \prod_{y_i=1} \Lambda_i \prod_{y_i=0} (1-\Lambda_i),$$

so the (conditional) likelihood function is:

$$L(\beta|X, y) = \prod_{i=1}^{N} \Lambda_i^{y_i} (1-\Lambda_i)^{(1-y_i)},$$

and the (conditional) log-likelihood function is:

$$\log L = \sum_{i=1}^{n} y_i \log \Lambda_i + (1-y_i)\log(1-\Lambda_i).$$

The log-likelihood equations are:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} (y_i - \Lambda_i) X_i = 0 \quad . \tag{6}$$

The MLE of β is the solution to (6). Since the log-likelihood function is strictly concave, the MLE is unique, but as (6) is highly non-linear in the parameters, it must be solved numerically. That is, the MLE cannot be written as closed-form expression, and this substantially complicates the task of evaluating the characteristics of its sampling distribution.

3. Analytic Results

Before we derive the analytic results for the Bias and MSE of the MLE in the Logit model, we first introduce Rilstone *et al.*'s results. The class of estimators considered by Rilstone *et al.* (RSU) includes those which can only be expressed implicitly as a function of the data. Suppose we have a regression model of the form

$$y_i = f(X_i;\beta) + \varepsilon_i .$$

The regressor vector, X_i, could include the lagged values of the dependent variable, or any other endogenous and exogenous variables. In order to make the derivation simple, RSU assume that all of the variables are random. Let $Z_i = (y_i, X_i)$ and let $Z_1, Z_2, Z_3, \ldots$ be a sequence of m dimensional i.i.d. random vectors. θ_0 represents the true parameter vector, which could include only β, or any other parameters of interest. The estimator $\hat{\theta}$ could be written in the following form.

$$\psi_n(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta}) = 0 , \tag{7}$$

where $g_i(\theta) = g_i(z_i,\theta)$ is a $k \times 1$ vector involving the known variables and the parameters, and $E[g_i(\theta)] = 0$ only for the true value θ_0. Preceding the derivation of the proposition below, RSU made some assumptions about the function $g_i(\theta)$.

Assumption 1 (Ullah, 2004, p.31)

The *s*th order derivatives of $g_i(\theta)$ exist in a neighborhood of θ and $E\left\|\nabla^s g_i(\theta)\right\|^2 < \infty$, where $\left\|A\right\|$, for a matrix A, denotes the usual norm, trace $[AA']^{1/2}$, and $\nabla^s A(\theta)$ is the matrix of s^{th} - order partial derivations of a matrix $A(\theta)$ with respect to θ and obtained recursively.

Assumption 2 (Ullah, 2004, p.31)

For some neighborhood of θ, $(\nabla \psi_n(\theta))^{-1} = O_p(1)$.

Assumption 3 (Ullah, 2004, p.31)

$\left\|\nabla^s g_i(\theta) - \nabla^s g_i(\theta_0)\right\| \le \left\|\theta - \theta_0\right\| M_i$ for some neighborhood of θ_0, where M_i satisfies the condition $E\left|M_i\right| \le C < \infty$, $i = 1, 2, \ldots$

In the following, we will ignore the argument part for any function of θ when there is no confusion. So, $g_i(\theta)$ will be written as g_i. Then, RSU derived the following Lemma.

Lemma 1 (RSU, 1996; Ullah, 2004, p.32)

Let assumptions 1-3 hold for some $s \geq 2$. Then the bias of $\hat{\theta}$ to order $O(n^{-1})$ is

$$B(\hat{\theta}) = \frac{1}{n} Q\left\{\overline{V_1 d_1} - \frac{1}{2}\overline{H}_2 \overline{[d_1 \otimes d_1]}\right\} \quad , \tag{8}$$

where $\overline{H}_j = \overline{\nabla}^j g_i$, $Q = [\overline{\nabla} g_i]^{-1}$, $V_i = [\nabla g_i - \overline{\nabla} g_i]$, and $d_i = Q g_i$. A bar over a function indicates its expectation, so that $\overline{\nabla} g_i = E[\nabla g_i]$. Further, if Assumptions 1-3 hold for some $s \geq 3$, then the MSE of $\hat{\theta}$ to order $O(n^{-2})$ is

$$MSE(\hat{\theta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^3}(\Pi_3 + \Pi_4 + \Pi_4') \tag{9}$$

where

$$\Pi_1 = \overline{d_1 d_1'}$$

$$\Pi_2 = Q\left\{-\overline{V_1 d_1 d_1'} + \tfrac{1}{2}\overline{H}_2 \overline{[d_1 \otimes d_1] d_1'}\right\}$$

$$\begin{aligned}
\Pi_3 = {} & Q\left\{\overline{V_1 d_1 d_2' V_2'} + \overline{V_1 d_2 d_1' V_2'} + \overline{V_1 d_2 d_2' V_1'}\right\} Q \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{[d_1 \otimes d_1][d_2' \otimes d_2']} + \overline{[d_1 \otimes d_2][d_1' \otimes d_2']} + \overline{[d_1 \otimes d_2][d_2' \otimes d_1']}\right\} \overline{H}_2' Q \\
& - \tfrac{1}{2} Q\left\{\overline{V_1 d_1 d_2' \otimes d_2'} + \overline{V_1 d_2 [d_1' \otimes d_2']} + \overline{V_1 d_2 [d_2' \otimes d_1']}\right\} \overline{H}_2' Q \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{d_1 \otimes d_1 d_2' V_2'} + \overline{[d_1 \otimes d_2] d_1' V_2'} + \overline{[d_1 \otimes d_2] d_2' V_1'}\right\} Q
\end{aligned}$$

$$\begin{aligned}
\Pi_4 = {} & Q\left\{\overline{V_1 Q V_1 d_2 d_2'} + \overline{V_1 Q V_2 d_1 d_2'} + \overline{V_1 Q V_2 d_2 d_1'}\right\} \\
& - \tfrac{1}{2} Q\left\{\overline{V_1 Q\overline{H}_2 [d_1 \otimes d_2] d_2'} + \overline{V_1 Q\overline{H}_2 [d_2 \otimes d_1] d_2'} + \overline{V_1 Q\overline{H}_2 [d_2 \otimes d_2] d_1'}\right\} \\
& + \tfrac{1}{2} Q\left\{\overline{W_1 [d_1 \otimes d_2] d_2'} + \overline{W_1 [d_2 \otimes d_1] d_2'} + \overline{W_1 [d_2 \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{[d_1 \otimes Q V_1 d_2] d_2'} + \overline{[d_1 \otimes Q V_2 d_1] d_2'} + \overline{[d_1 \otimes Q V_2 d_2] d_1'}\right\} \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{d_1 \otimes Q\overline{H}_2 [d_1 \otimes d_2] d_2'} + \overline{d_1 \otimes Q\overline{H}_2 [d_2 \otimes d_1] d_2'} + \overline{d_1 \otimes Q\overline{H}_2 [d_2 \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{[Q V_1 d_1 \otimes d_2] d_2'} + \overline{[Q V_1 d_2 \otimes d_1] d_2'} + \overline{[Q V_1 d_2 \otimes d_2] d_1'}\right\}
\end{aligned}$$

$$+ \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{[Q\overline{H}_2 [d_1 \otimes d_1] \otimes d_2] d_2'} + \overline{[Q\overline{H}_2 [d_1 \otimes d_2] \otimes d_1] d_2'} + \overline{[Q\overline{H}_2 [d_1 \otimes d_2] \otimes d_2] d_1'}\right\}$$

$$- \tfrac{1}{6} Q\overline{H}_3 \left\{\overline{[d_1 \otimes d_1 \otimes d_2] d_2'} + \overline{[d_1 \otimes d_2 \otimes d_1] d_2'} + \overline{[d_1 \otimes d_2 \otimes d_2] d_1'}\right\} \tag{10}$$

where $W_i = [\nabla^2 g_i - \overline{\nabla}^2 g_i]$.

Now we apply the above Lemma to derive the bias and MSE of the MLE in the Logit model. First, we assume that both the dependent and independent variable in the Logit model are random, and the observations are i.i.d. Comparing (6) and (7), we can see that for the Logit model, we should set $g_i = (y_i - \Lambda_i)X_i$, and we know that $E(g_i \mid X_i) = 0$, then according to the law of iterated expectations, $E(g_i) = 0$.

Now we have the following results:

$$\nabla g_i = -\Lambda_i^{(1)} X_i X_i'; \qquad \overline{H}_1 = \overline{\nabla} g_i = -E(\Lambda_i^{(1)} X_i X_i')$$

$$\nabla^2 g_i = -\Lambda_i^{(2)} X_i (X_i' \otimes X_i'); \qquad \overline{H}_2 = \overline{\nabla}^2 g_i = -E[\Lambda_i^{(2)} X_i (X_i' \otimes X_i')]$$

$$\nabla^3 g_i = -\Lambda_i^{(3)} X_i (X_i' \otimes X_i' \otimes X_i'); \qquad \overline{H}_3 = \overline{\nabla}^3 g_i = -E[\Lambda_i^{(3)} X_i (X_i' \otimes X_i' \otimes X_i')]$$

$$Q = (\overline{\nabla} g_i)^{-1} = -[E(\Lambda_i^{(1)} X_i X_i')]^{-1}; \qquad d_i = Q g_i = -[E(\Lambda_i^{(1)} X_i X_i')]^{-1} (y_i - \Lambda_i) X_i$$

$$V_i = \nabla g_i - \overline{\nabla} g_i = -\Lambda_i^{(1)} X_i X_i' + E(\Lambda_i^{(1)} X_i X_i')$$

$$W_i = \nabla^2 g_i - \overline{\nabla}^2 g_i = -\Lambda_i^{(2)} X_i (X_i' \otimes X_i') + E[\Lambda_i^{(2)} X_i (X_i' \otimes X_i')] \quad , \tag{11}$$

where $\Lambda_i^{(s)}$ is the s^{th} order derivative of Λ_i with respect to the argument of $X_i'\beta$ and

$$\Lambda_i^{(1)} = \frac{\exp(X_i'\beta)}{[1+\exp(X_i'\beta)]^2}$$

$$\Lambda_i^{(2)} = \frac{\exp(X_i'\beta)[1-\exp(X_i'\beta)]}{[1+\exp(X_i'\beta)]^3}$$

$$\Lambda_i^{(3)} = \frac{\exp(X_i'\beta)\{1-4\exp(X_i'\beta)+[\exp(X_i'\beta)]^2\}}{[1+\exp(X_i'\beta)]^4} \quad . \tag{12}$$

Then we can derive the following theorem and corollary.

Theorem 1

If assumptions 1-3 hold for some $s \geq 2$. Then the bias of the MLE in the Logit model, to the order of $O(n^{-1})$ is

$$Bias(\hat{\beta}) = \frac{1}{2n} Q\overline{H}_2 vecQ \tag{13}$$

Further if Assumptions 1-3 hold for some $s \geq 3$, then the MSE of MLE in the Logit model to order $O(n^{-2})$ is

$$MSE(\hat{\beta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^3}(\Pi_3 + \Pi_4 + \Pi_4') \tag{14}$$

where

$$\Pi_1 = -Q$$

$$\Pi_2 = -Q\left\{E(\Lambda_1^{(1)} V_1 Q X_1 X_1' Q) - \tfrac{1}{2}\overline{H}_2 E\left\{\Lambda_1^{(2)}[vec(QX_1X_1'Q)]X_1'Q\right\}\right\}$$

$$\Pi_3 = Q\left\{E[V_1 Q(\Lambda_2^{(1)} X_2 X_2')QV_1']\right\}Q \ + \tfrac{1}{4}Q\overline{H}_2\{(vecQ)(vecQ)' + (Q\otimes Q)$$
$$+ (Q\otimes Q)\left\{E[(vec\Lambda_1^{(1)} X_2 X_1')(vec\Lambda_2^{(1)} X_1 X_2')']\right\}(Q\otimes Q)\}\overline{H}_2'Q$$

$$\Pi_4 = -QE(V_1QV_1Q) + \tfrac{1}{4}Q\overline{H}_2(Q\otimes Q)E\left\{\Lambda_1^{(1)}\Lambda_2^{(1)}(X_1\otimes\overline{H}_2)[vec(QX_2X_1'Q)X_2'\right.$$
$$+ vec(QX_1X_2'Q)X_2' + vec(QX_2X_2'Q)X_1']\}Q$$
$$+ \tfrac{1}{4}Q\overline{H}_2 E\left\{\Lambda_1^{(1)}\Lambda_2^{(1)}\left[(Q\overline{H}_2 vec(QX_1X_1'Q)\otimes QX_2)X_2'\right.\right.$$
$$+ (Q\overline{H}_2 vec(QX_2X_1'Q)\otimes QX_1)X_2' + (Q\overline{H}_2 vec(QX_2X_1'Q)\otimes QX_2)X_1']\}Q$$
$$- \tfrac{1}{6}Q\overline{H}_3 E\left\{\Lambda_1^{(1)}\Lambda_2^{(1)}\left[(vec(QX_1X_1'Q)\otimes QX_2)X_2' + (vec(QX_2X_1'Q)\otimes QX_1)X_2'\right.\right.$$
$$+ (vec(QX_2X_1'Q)\otimes QX_2)X_1']\}Q \tag{15}$$

Now we consider a simple case of (4) with only one regressor, which implies that the constant term α in the latent regression model (1) equals the true threshold a in (2). For this simple model, we derive the following corollary.

Corollary 1

If assumptions 1-3 hold for some $s \geq 2$. Then the bias of the MLE of β in the Logit model with only one regressor, to the order of $O(n^{-1})$, is

$$Bias(\hat{\beta}) = -\frac{1}{2n}\frac{E(\Lambda_i^{(2)} x_i^3)}{[E(\Lambda_i^{(1)} x_i^2)]^2} \ . \tag{16}$$

Further if Assumptions 1-3 hold for some $s \geq 3$, then the MSE of the MLE of β in the Logit model with only one regressor, to order $O(n^{-2})$ is

$$MSE(\hat{\beta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^3}(\Pi_3 + \Pi_4 + \Pi_4') \tag{17}$$

where

$$\Pi_1 = \frac{1}{E(\Lambda_1^{(1)} X_1^2)}$$

$$\Pi_2 = -\frac{1}{[E(\Lambda_1^{(1)}X_1^2)]^3}\left\{E(\Lambda_1^{(1)}X_1^2)^2 - [E(\Lambda_1^{(1)}X_1^2)]^2 + \frac{[E(\Lambda_1^{(2)}X_1^3)]^2}{2E(\Lambda_1^{(1)}X_1^2)}\right\}$$

$$\Pi_3 = \frac{1}{[E(\Lambda_1^{(1)}X_1^2)]^3}\left\{E(\Lambda_1^{(1)}X_1^2)^2 - [E(\Lambda_1^{(1)}X_1^2)]^2 + \frac{3[E(\Lambda_1^{(2)}X_1^3)]^2}{4E(\Lambda_1^{(1)}X_1^2)}\right\}$$

$$\Pi_4 = \frac{1}{[E(\Lambda_1' X_1^2)]^3}\left\{E(\Lambda_1^{(1)}X_1^2)^2 - [E(\Lambda_1^{(1)}X_1^2)]^2 + \frac{3[E(\Lambda_1^{(2)}X_1^3)]^2}{2E(\Lambda_1^{(1)}X_1^2)} - \tfrac{1}{2}E(\Lambda_1^{(3)})X_1^4]\right\} \quad (18)$$

The proofs of Theorem 1 and Corollary 1 are given in the Appendix.

4. Numerical Results

In this section, we will derive some numerical results based on Corollary 1. These evaluations are conducted for the one-regressor Logit model with different distributional assumptions for the covariates, and different sample sizes. We choose values for the parameters which ensure a sensible signal/noise ratio for the model. The latter is determined by considering the goodness-of-fit for the model. There are many goodness-of-fit measures suggested for the Logit models, and we have used the one suggested by Efron (1978):

$$R_{Ef}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{P}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad . \quad (19)$$

Here, we replace the predicted probability $\hat{P}_i$ with the P_i calculated from y_i^*. Usually, with the cross-section data, only modest goodness-of-fit values are achieved. Therefore, we choose the parameters which generate values for this goodness-of-fit measure between about 0.7 and 0.8 for each model.

The distributions from which the regressor values are generated include the standard normal distribution, the uniform distribution in the range from -2 to 2, and the Chi-Square distribution with 3 degrees of freedom. So, we include both symmetric and asymmetric distributions for the processes that generate the covariate. In Tables 1 to 3 the first two columns give the true value for the parameter and the sample size. The third column, $\hat{\beta}$, is the average MLE of β based on 2,000 replications of a separate Monte Carlo experiment.

More specifically, the steps associated with the Monte Carlo experiment are as follows:

(i) Set a value for the parameter.

(ii) Generate an (nx1) vector of observations for the random regressor X, with values drawn from the distribution we choose.

(iii) Generate (nx1) vectors of observations for y^* and y based on (3) with a logistic-distributed disturbance term.

(iv) Estimate a logit model based on y and X without a constant term, and record the MLE for the β and the asymptotic standard error of the MLE of β.

(v) Repeat steps (i)-(iv) 2,000 times.

(vi) Calculate the averages of the 2,000 MLE's of β and of its asymptotic standard error to get the values referred to as $\hat{\beta}$ in column (1) and ASE($\hat{\beta}$) in column (6) of the tables.

Two bias-adjusted estimators, $\hat{\beta}_{BC}$ and $\tilde{\beta}_{BC}$, are then defined as follows:

$$\hat{\beta}_{BC} = \hat{\beta} - B(\hat{\beta}),$$

and

$$\tilde{\beta}_{BC} = \hat{\beta} - \hat{B}(\hat{\beta}),$$

where $B(\hat{\beta})$ is the bias based on (16) and the true parameter β, and $\hat{B}(\hat{\beta})$ is the bias based on (16) and the estimator $\hat{\beta}$. In practice, $\hat{\beta}_{BC}$ is an infeasible estimator since it involves the true parameter, whose value we do not know. However, $\tilde{\beta}_{BC}$ is the feasible estimator which can be easily obtained, based on the estimator of the parameter. The sixth column in the tables gives the asymptotic standard error from the maximum likelihood estimation across the 2,000 repetitions. The standard deviation, SD($\hat{\beta}_{BC}$) and the standard error, SE($\hat{\beta}_{BC}$), corresponding to the bias-adjusted estimators $\hat{\beta}_{BC}$, are provided in the following columns. As the bias of $\hat{\beta}_{BC}$ is zero, it follows that SD($\hat{\beta}_{BC}$) and SE($\hat{\beta}_{BC}$) are defined as follows:

$$\text{SD}(\hat{\beta}_{BC}) = \sqrt{MSE(\hat{\beta}) - B(\hat{\beta})B'(\hat{\beta})}\ , \tag{20}$$

and

$$\text{SE}(\hat{\beta}_{BC}) = \sqrt{\hat{M}SE(\hat{\beta}) - \hat{B}(\hat{\beta})\hat{B}'(\hat{\beta})}\ . \tag{21}$$

From (20) and (21), we can also see that SD($\hat{\beta}_{BC}$) and SE($\hat{\beta}_{BC}$) are also the second order approximations to the standard deviation and the standard error of $\hat{\beta}$. In the last two columns, we report the MSE of $\hat{\beta}$ and $\hat{\beta}_{BC}$. MSE($\hat{\beta}$) is based on equation (17) and the true parameter value. MSE($\hat{\beta}_{BC}$) is the square of SD($\hat{\beta}_{BC}$), because $\hat{\beta}_{BC}$ is unbiased.

For example, the first line in Table 1 tells us the following. For the model with the standard normal regressor, the true parameter of 1.4 and sample size of 100, the MLE of $\hat{\beta}$ has an asymptotic standard error of 0.3262, a finite-sample standard error of 0.3053, a finite-sample standard deviation of 0.2974 and a mean squared error of 0.0903. So, although both the asymptotic and finite-sample standard errors over-state the true finite-sample standard deviation, in this case the finite-sample standard error is more accurate than its asymptotic counterpart. Further, the bias-corrected estimator, $\hat{\beta}_{BC}$, has a standard deviation of 0.2974, a standard error of 0.3053 and a mean squared error of 0.0884. So, again, the standard error overstates the true standard deviation. In addition, comparing the MSE for $\hat{\beta}_{BC}$ with that for $\hat{\beta}$ given above, there is a gain in efficiency by correcting for bias.

The Monte Carlo experiment was conducted with code written for the SHAZAM package (Whistler, *et al.*, 2001). In principle, equations (20) and (21) would be inapplicable if the bias term exceeded the MSE term. This situation did not arise, however, for the range of parameter values that we considered. The results of the Monte Carlo experiment were found to be somewhat sensitive to the choice of the seed for the random number generator. A similar situation arose when the experiment was replicated with code written for the EViews package (see Quantitative Micro Software, 2004). However, this sensitivity was not serious enough to affect our main conclusions, or our evaluations based on the analytic results in Theorem 1 and Corollary 1. In practice, we just need to obtain an estimate by applying the MLE to the actual data set, and substitute this estimate into the bias and MSE formulae from Theorem 1 and Corollary 1 to obtain the estimated bias and estimated MSE. As we showed above, the estimated bias is closer to the exact analytic result as the sample size increases. Then we can use the estimated bias to correct the estimator.

We have chosen the same sample sizes and the same range of parameters when the regressor follows a standard normal distribution and a uniform distribution. However, the range of the parameter values and the sample sizes are different for the chi-square regressor. This was a practical necessity. To obtain the MLE for the Logit model by simulation, there is always a chance that problems will be encountered. For example, the MLE is not defined if the dependent variables can be perfectly predicted by the model. Another possibility is that the algorithm being used to solve the non-linear likelihood equations may fail to converge. With artificial data and many, many repetitions of the experiment these issues can be important. The parameter values and sample sizes for the case of the chi-square regressor were chosen to avoid such problems. Tables 1, 2, and 3 report the results for cases where the regressor follows the standard normal distribution, the uniform distribution on (-2, 2) and the chi-square distribution with three degrees of freedom, respectively. In order to show the performance of the bias correction more clearly, Figures 1 to 6 show the absolute difference between the different estimators $\hat{\beta}$, $\hat{\beta}_{BC}$, $\tilde{\beta}_{BC}$ and the true parameter β.

From the information in the tables and the graphs, we can see that the bias-corrected estimators are closer to the true parameter than is the MLE, no matter if it is corrected by the true analytic bias or the estimated analytic bias. For each model, the two bias-corrected estimators generally become closer to the true parameter as the sample size increases. These results are consistent with the argument that the results derived from the large-n approximations lie between the true value and the asymptotic approximations, and the accuracy increases as the sample size increase. The comparison between the true analytic result and estimated analytic results depends on the parameter value, the regressor data, and the sample size. However, we can find that for each choice of data, when the sample size increases, the estimated analytical results become closer to the true analytic results.

Further, from the columns for ASE($\hat{\beta}$), SE($\hat{\beta}$) and SD($\hat{\beta}$), we can see that both the asymptotic standard error and the finite-sample standard error tend to overestimate the finite-sample standard deviation of the MLE, and the asymptotic standard error tends to worse in this respect than the finite-sample standard error of the MLE. From the columns for SD($\hat{\beta}_{BC}$) and SE($\hat{\beta}_{BC}$), we can see that for the bias-corrected estimator, its standard error tends to overestimate its standard deviation. In terms of MSE, the bias-corrected estimator obtains some efficiency gains over the uncorrected MLE.

5. Conclusions

In this chapter, we apply RSU's result to examine the finite sample properties of the MLE in the Logit model. We derive the second order bias and MSE function for the MLE in the Logit model and undertake some numerical evaluations to illustrate the analytic results. From these numerical results, we can reach the following conclusions. First, the bias correction we provide in the Corollary 1 can bring the MLE closer to the true parameter, and the bias-corrected estimator is more efficient than the uncorrected MLE. Second, the accuracy of the corrected estimator generally increases as the sample size increase. All of this is consistent with the argument that the large-n approximation provides a result between the true value and the asymptotic value, and the accuracy of the approximation increase as the sample size increases. Third, the estimated analytical results are closer to the analytical results as the sample size increases. Fourth, the asymptotic standard error overestimates the finite-sample standard deviation of the MLE. In the future, we expect to generalize the results to the case where the independent variables are fixed and make the results comparable with other works.

Table 1: Parameter Estimator and Standard Error Estimates

With Standard Normal Regressor

β	N	$\hat{\beta}$	$\hat{\beta}_{BC}$	$\tilde{\beta}_{BC}$	ASE($\hat{\beta}$)	SD($\hat{\beta}$) (SD($\hat{\beta}_{BC}$))	SE($\hat{\beta}$) (SE($\hat{\beta}_{BC}$))	MSE($\hat{\beta}$)	MSE($\hat{\beta}_{BC}$)
1.4	100	1.4655	1.4222	1.4197	0.3262	0.2974	0.3053	0.0903	0.0884
	200	1.4210	1.3993	1.3989	0.2247	0.2159	0.2178	0.0471	0.0466
1.5	100	1.5770	1.5298	1.5267	0.3425	0.3096	0.3192	0.0981	0.0958
	200	1.5185	1.4948	1.4945	0.2347	0.2252	0.2269	0.0513	0.0507
1.6	100	1.6785	1.6273	1.6240	0.3587	0.3222	0.3323	0.1064	0.1038
	200	1.6208	1.5952	1.5947	0.2455	0.2348	0.2369	0.0558	0.0551
1.7	100	1.7708	1.7154	1.7123	0.3731	0.3352	0.3446	0.1154	0.1123
	200	1.7205	1.6928	1.6923	0.2564	0.2448	0.2469	0.0607	0.0599
1.8	100	1.8633	1.8034	1.8006	0.3888	0.3485	0.3571	0.1250	0.1215
	200	1.8195	1.7895	1.7891	0.2677	0.2551	0.2571	0.0660	0.0651
1.9	100	1.9722	1.9078	1.9044	0.4069	0.3622	0.3722	0.1353	0.1312
	200	1.9304	1.8982	1.8974	0.2807	0.2657	0.2689	0.0716	0.0706
2.0	100	2.0835	2.0143	2.0102	0.4263	0.3761	0.3880	0.1463	0.1415
	200	2.0285	1.9939	1.9932	0.2928	0.2765	0.2797	0.0777	0.0765
2.1	100	2.1928	2.1186	2.1138	0.4457	0.3904	0.4038	0.1579	0.1524
	200	2.1235	2.0864	2.0858	0.3046	0.2877	0.2904	0.0841	0.0828
2.2	100	2.2663	2.1869	2.1834	0.4593	0.4049	0.4147	0.1702	0.1640
	200	2.2282	2.1885	2.1878	0.3179	0.2991	0.3024	0.0910	0.0895
2.3	100	2.3675	2.2828	2.2789	0.4788	0.4197	0.4299	0.1833	0.1761
	200	2.3375	2.2952	2.2941	0.3322	0.3108	0.3152	0.0984	0.0966

Table 2: Parameter Estimator and Standard Error Estimates

With Uniform Distribution (-2, 2) Regressor

β	N	$\hat{\beta}$	$\hat{\beta}_{BC}$	$\tilde{\beta}_{BC}$	ASE($\hat{\beta}$)	SD($\hat{\beta}$) (SD($\hat{\beta}_{BC}$))	SE($\hat{\beta}$) (SE($\hat{\beta}_{BC}$))	MSE($\hat{\beta}$)	MSE($\hat{\beta}_{BC}$)
1.4	100	1.4807	1.4437	1.4401	0.2874	0.2584	0.2678	0.0681	0.0668
	200	1.4126	1.3941	1.3939	0.1946	0.1872	0.1883	0.0354	0.0350
1.5	100	1.5843	1.5428	1.5388	0.3032	0.2701	0.2804	0.0747	0.0730
	200	1.5194	1.4986	1.4982	0.2055	0.1963	0.1981	0.0390	0.0385
1.6	100	1.6582	1.6119	1.6089	0.3145	0.2824	0.2898	0.0819	0.0797
	200	1.6267	1.6035	1.6028	0.2172	0.2059	0.2086	0.0429	0.0424
1.7	100	1.7619	1.7104	1.7070	0.3317	0.2952	0.3033	0.0898	0.0871
	200	1.7286	1.7028	1.7020	0.2288	0.2161	0.2190	0.0473	0.0467
1.8	100	1.8639	1.8069	1.8031	0.3504	0.3084	0.3171	0.0984	0.0951
	200	1.8265	1.7980	1.7973	0.2406	0.2266	0.2295	0.0522	0.0514
1.9	100	1.9751	1.9122	1.9074	0.3716	0.3221	0.3326	0.1077	0.1038
	200	1.9343	1.9028	1.9017	0.2543	0.2377	0.2416	0.0575	0.0565
2.0	100	2.0665	1.9972	1.9927	0.3888	0.3361	0.3457	0.1178	0.1130
	200	2.0313	1.9966	1.9956	0.2672	0.2492	0.2528	0.0633	0.0621
2.1	100	2.1790	2.1028	2.0972	0.4113	0.3505	0.3621	0.1287	0.1229
	200	2.1396	2.1015	2.1001	0.2823	0.2610	0.2659	0.0696	0.0681
2.2	100	2.3119	2.2286	2.2200	0.4393	0.3652	0.3819	0.1403	0.1334
	200	2.2355	2.1938	2.1925	0.2963	0.2733	0.2778	0.0764	0.0747
2.3	100	2.4312	2.3402	2.3295	0.4668	0.3801	0.3999	0.1527	0.1445
	200	2.3412	2.2957	2.2940	0.3122	0.2860	0.2913	0.0839	0.0818

Table 3: Parameter Estimator and Standard Error Estimates

With Chi-Square (3) Regressor

β	N	$\hat{\beta}$	$\hat{\beta}_{BC}$	$\tilde{\beta}_{BC}$	ASE($\hat{\beta}$)	SD($\hat{\beta}$) (SD($\hat{\beta}_{BC}$))	SE($\hat{\beta}$) (SE($\hat{\beta}_{BC}$))	MSE($\hat{\beta}$)	MSE($\hat{\beta}_{BC}$)
1.7	200	1.8024	1.7204	1.7086	0.3651	0.2421	0.2574	0.0653	0.0586
	500	1.7455	1.7127	1.7107	0.2206	0.1772	0.1838	0.0325	0.0314
1.8	200	1.9066	1.8130	1.7996	0.3898	0.2570	0.2720	0.0748	0.0661
	500	1.8502	1.8127	1.8103	0.2361	0.1918	0.1992	0.0382	0.0368
1.9	200	1.9962	1.8901	1.8770	0.4113	0.2711	0.2836	0.0848	0.0735
	500	1.9521	1.9097	1.9069	0.2512	0.2067	0.2146	0.0445	0.0427
2.0	200	2.0917	1.9720	1.9585	0.4346	0.2841	0.2947	0.0950	0.0807
	500	2.0557	2.0078	2.0046	0.2667	0.2219	0.2305	0.0515	0.0492
2.1	200	2.2114	2.0769	2.0591	0.4641	0.2956	0.3063	0.1055	0.0874
	500	2.1622	2.1085	2.1046	0.2832	0.2374	0.2471	0.0592	0.0563
2.2	200	2.3383	2.1879	2.1640	0.4962	0.3052	0.3148	0.1158	0.0932
	500	2.2645	2.2044	2.2000	0.2989	0.2531	0.2633	0.0677	0.0640
2.3	200	2.4645	2.2970	2.2661	0.5281	0.3127	0.3184	0.1258	0.0978
	500	2.3656	2.2987	2.2939	0.3148	0.2689	0.2794	0.0768	0.0723
2.4	200	2.5582	2.3723	2.3406	0.5519	0.3173	0.3169	0.1352	0.1007
	500	2.4628	2.3885	2.3836	0.3302	0.2849	0.2949	0.0867	0.0812
2.5	200	2.6745	2.4690	2.4314	0.5822	0.3183	0.3087	0.1436	0.1013
	500	2.5643	2.4821	2.4767	0.3465	0.3009	0.3112	0.0973	0.0905
2.6	200	2.7732	2.5467	2.5070	0.6082	0.3149	0.2943	0.1504	0.0991
	500	2.6750	2.5844	2.5777	0.3644	0.3169	0.3288	0.1086	0.1004

Figure 1: Bias of MLE for N(0,1) Regressor and N=100

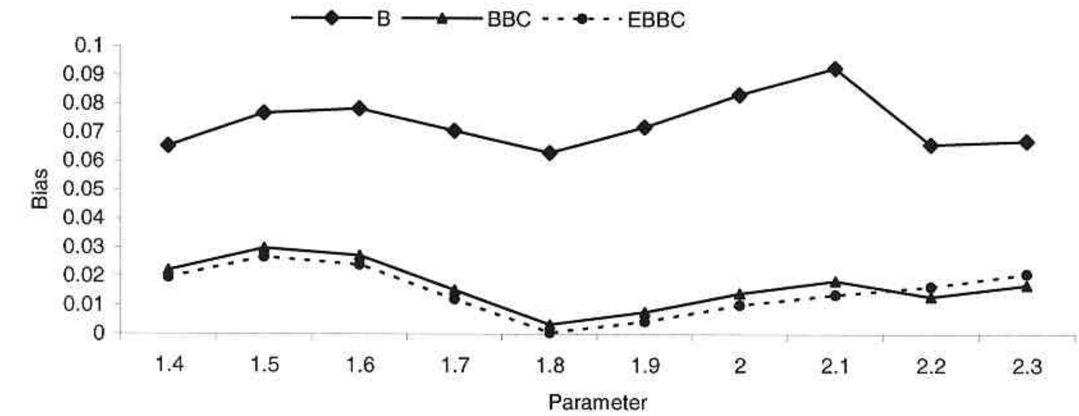


Figure 2: Bias of MLE for N(0,1) Regressor and N=200

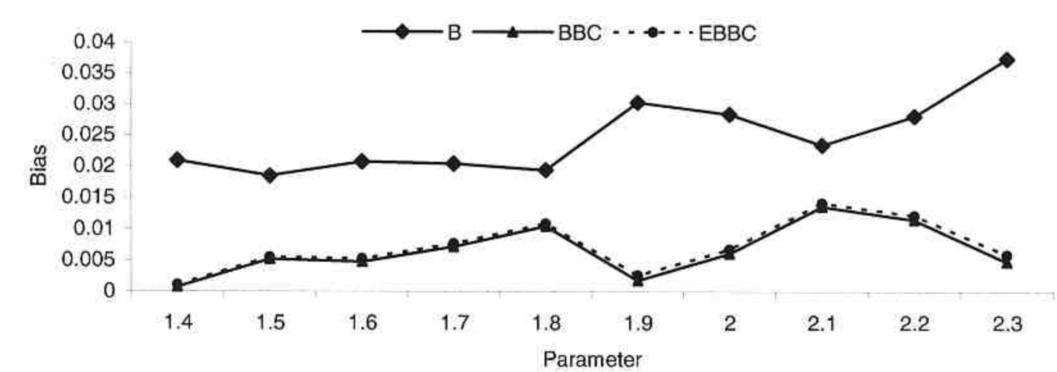


Figure 3: Bias of MLE for Uniform(-2,2) Regressor and N=100

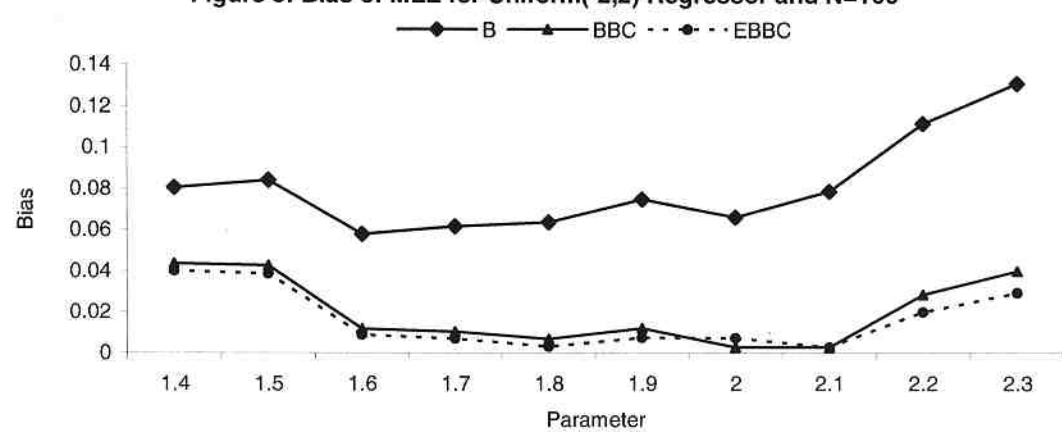


Figure 4: Bias of MLE for Uniform(-2,2) Regressor and N=200

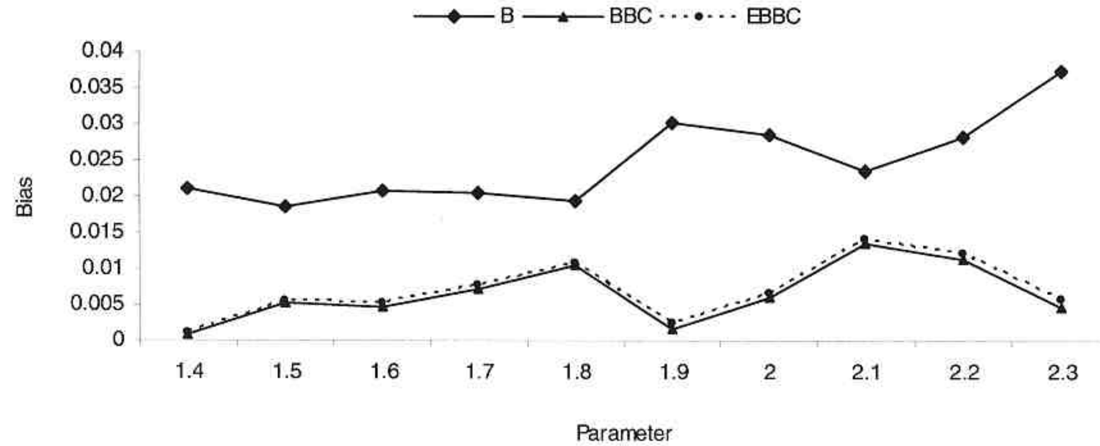


Figure 5: Bias of MLE for Chi-Square(3) Regressor and N=200

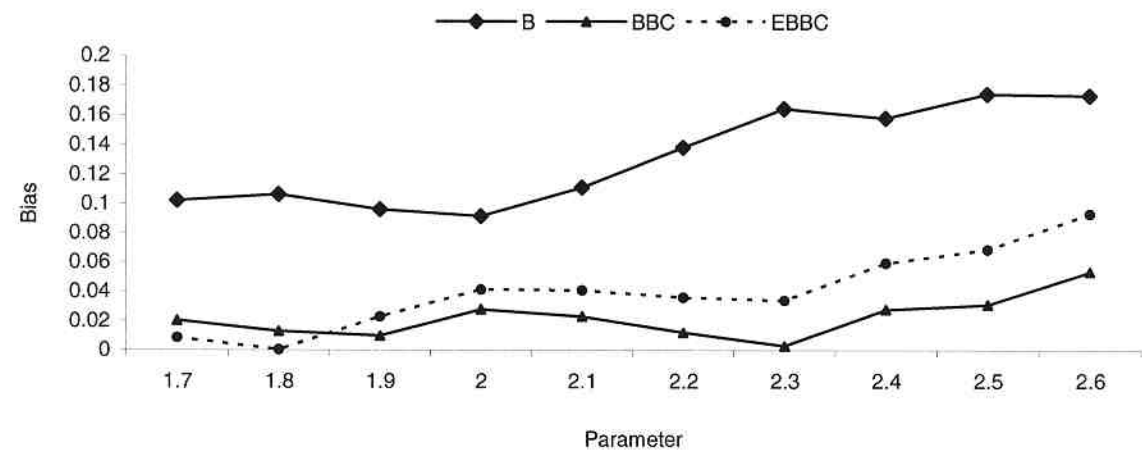
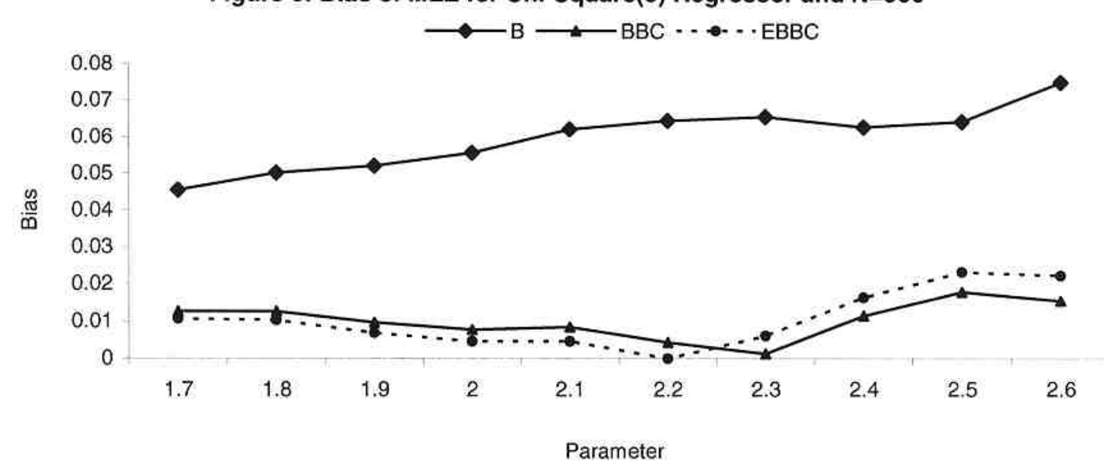


Figure 6: Bias of MLE for Chi-Square(3) Regressor and N=500



Note: in all the graphs above, the series "B" is the difference between $\hat{\beta}$ and β; the series "BBC" is the difference between $\hat{\beta}_{BC}$ and β; the series "EBBC" is the difference between $\tilde{\beta}_{BC}$ and β.

Appendix: Proof of Theorem 1 and Corollary 1

Proof of Theorem 1

First, for the Logit model in (4), we know the following propositions.

$$E(y_i^j|X) = \Lambda_i \quad (22)$$

By applying (11) and the law of iterated expectations, we can derive the following results.

$$\overline{V_1 d_1} = 0$$

$$\overline{d_1 \otimes d_1} = -vecQ$$

$$\overline{V_1 d_1 d_1'} = E(\Lambda_1^{(1)} V_1 Q X_1 X_1' Q)$$

$$\overline{[d_1 \otimes d_1] d_1'} = E\{\Lambda_1^{(2)} [vec(QX_1 X_1' Q)] X_1' Q\}$$

$$\overline{V_1 d_1 d_2' V_2'} = 0$$

$$\overline{V_1 d_2 d_1' V_2'} = 0$$

$$\overline{V_1 d_2 d_2' V_1'} = E[V_1 Q(\Lambda_2^{(1)} X_2 X_2') Q V_1']$$

$$\overline{[d_1 \otimes d_1][d_2' \otimes d_2']} = (vecQ)(vecQ)'$$

$$\overline{[d_1 \otimes d_2][d_1' \otimes d_2']} = Q \otimes Q$$

$$\overline{[d_1 \otimes d_2][d_2' \otimes d_1']} = (Q \otimes Q)\{E[(vec\Lambda_1^{(1)} X_2 X_1')(vec\Lambda_2^{(1)} X_1 X_2')']\}(Q \otimes Q)$$

$$\overline{V_1 d_1 d_2' \otimes d_2'} = 0$$

$$\overline{V_1 d_2 [d_1' \otimes d_2']} = 0$$

$$\overline{V_1 d_2 [d_2' \otimes d_1']} = 0$$

$$\overline{d_1 \otimes d_1 d_2' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_1' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_2' V_1'} = 0$$

$$\overline{V_1 Q V_1 d_2 d_2'} = -E V_1 Q V_1 Q$$

$$\overline{V_1 Q V_2 d_1 d_2'} = 0$$

$$\overline{V_1 Q V_2 d_2 d_1'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_1 \otimes d_2] d_2'} = 0$$

$$\overline{V_1 Q\overline{H}_2[d_2 \otimes d_1]d_2'} = 0$$

$$\overline{V_1 Q\overline{H}_2[d_2 \otimes d_2]d_1'} = 0$$

$$\overline{W_1[d_1 \otimes d_2]d_2'} = 0$$

$$\overline{W_1[d_2 \otimes d_1]d_2'} = 0$$

$$\overline{W_1[d_2 \otimes d_2]d_1'} = 0$$

$$\overline{[d_1 \otimes QV_1 d_2]d_2'} = 0$$

$$\overline{[d_1 \otimes QV_2 d_1]d_2'} = 0$$

$$\overline{[d_1 \otimes QV_2 d_2]d_1'} = 0$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_1 \otimes d_2]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}(Q \otimes Q)(X_1 \otimes \overline{H}_2)[vec(QX_2X_1'Q)]X_2'Q$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_1]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}(Q \otimes Q)(X_1 \otimes \overline{H}_2)[vec(QX_1X_2'Q)]X_2'Q$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_2]d_1'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}(Q \otimes Q)(X_1 \otimes \overline{H}_2)[vec(QX_2X_2'Q)]X_1'Q$$

$$\overline{[QV_1 d_1 \otimes d_2]d_2'} = 0$$

$$\overline{[QV_1 d_2 \otimes d_1]d_2'} = 0$$

$$\overline{[QV_1 d_2 \otimes d_2]d_1'} = 0$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_1] \otimes d_2]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_1X_1'Q) \otimes QX_2]X_2'Q$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_1]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_2X_1'Q) \otimes QX_1]X_2'Q$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_2]d_1'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_2X_1'Q) \otimes QX_2]X_1'Q$$

$$\overline{[d_1 \otimes d_1 \otimes d_2]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_1X_1'Q) \otimes QX_2]X_2'Q$$

$$\overline{[d_1 \otimes d_2 \otimes d_1]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_2X_1'Q) \otimes QX_1]X_2'Q$$

$$\overline{[d_1 \otimes d_2 \otimes d_2]d_1'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_2X_1'Q) \otimes QX_2]X_1'Q \tag{23}$$

Therefore, based on Lemma 1 and (23), Theorem 1 is proved.

Proof of Corollary 1

When the Logit model only include one regressor, (23) reduces to

$$\overline{V_1 d_1} = 0$$

$$\overline{d_1 \otimes d_1} = -\frac{1}{E(\Lambda_1^{(i)} X_1^2)}$$

$$\overline{V_1 d_1 d_1'} = 1 - \frac{E(\Lambda_1^{(i)} X_1^2)^2}{[E(\Lambda_1^{(i)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_1] d_1'} = -\frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(i)} X_1^2)]^2}$$

$$\overline{V_1 d_1 d_2' V_2'} = 0$$

$$\overline{V_1 d_2 d_1' V_2'} = 0$$

$$\overline{V_1 d_2 d_2' V_1'} = \frac{E(\Lambda_1^{(i)} X_1^2)^2 - [E(\Lambda_1^{(i)} X_1^2)]^2}{E(\Lambda_1^{(i)} X_1^2)}$$

$$\overline{[d_1 \otimes d_1][d_2' \otimes d_2']} = \frac{1}{[E(\Lambda_1^{(i)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2][d_1' \otimes d_2']} = \frac{1}{[E(\Lambda_1^{(i)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2][d_2' \otimes d_1']} = \frac{1}{[E(\Lambda_1^{(i)} X_1^2)]^2}$$

$$\overline{V_1 d_1 d_2' \otimes d_2'} = 0$$

$$\overline{V_1 d_2 [d_1' \otimes d_2']} = 0$$

$$\overline{V_1 d_2 [d_2' \otimes d_1']} = 0$$

$$\overline{d_1 \otimes d_1 d_2' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_1' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_2' V_1'} = 0$$

$$\overline{V_1 Q V_1 d_2 d_2'} = 1 - \frac{E(\Lambda_1^{(i)} X_1^2)^2}{[E(\Lambda_1^{(i)} X_1^2)]^2}$$

$$\overline{V_1 Q V_2 d_1 d_2'} = 0$$

$$\overline{V_1 Q V_2 d_2 d_1'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_1 \otimes d_2] d_2'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_2 \otimes d_1] d_2'} = 0$$

$$\overline{V_1 Q\overline{H}_2[d_2 \otimes d_2]d_1'} = 0$$

$$\overline{W_1[d_1 \otimes d_2]d_2'} = 0$$

$$\overline{W_1[d_2 \otimes d_1]d_2'} = 0$$

$$\overline{W_1[d_2 \otimes d_2]d_1'} = 0$$

$$\overline{[d_1 \otimes QV_1 d_2]d_2'} = 0$$

$$\overline{[d_1 \otimes QV_2 d_1]d_2'} = 0$$

$$\overline{[d_1 \otimes QV_2 d_2]d_1'} = 0$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_1 \otimes d_2]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_1]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_2]d_1'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[QV_1 d_1 \otimes d_2]d_2'} = 0$$

$$\overline{[QV_1 d_2 \otimes d_1]d_2'} = 0$$

$$\overline{[QV_1 d_2 \otimes d_2]d_1'} = 0$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_1] \otimes d_2]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_1]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_2]d_1'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[d_1 \otimes d_1 \otimes d_2]d_2'} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2 \otimes d_1]d_2'} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2 \otimes d_2]d_1'} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2} \tag{24}$$

Based on (24) and Lemma 1, Corollary 1 is proved.

CHAPTER 5:
FINITE-SAMPLE MOMENTS FOR STOCHASTIC INDEX NUMBERS

1. Introduction

Index numbers measure the aggregate change in any collection of variables of interest over some time period, or across countries or regions. Therefore, index numbers can be used to assist in making both intertemporal and multilateral spatial comparisons. Some widely used index number series include the consumer price index (CPI), the producer price index (PPI), stock market price indices, the national health index, *etc.* The concept of an index number is widely used in many disciplines, especially economics. Due to their important role, index numbers have attracted the attention of many economists and statisticians over the past hundred years. There are many formulae that have been proposed for calculating index numbers – these include Laspeyres Index, Paasche Index, the Edgeworth-Marshall Index, the Drobisch Index, *etc.* Several other index numbers are based on Laspeyres index and Paasche index, such as Fisher's "ideal" index and the Stuvel index. Different index formulae are motivated by different objectives. For intertemporal comparisons, Laspeyres index, Paasche index and Fisher's index are widely used. For multilateral comparisons, the CCD index derived by Caves *et al.* (1982) and the generalized CCD are well-known. Another well-known index is the Divisia index proposed by Divisia (1925). One property of the Divisia index is that its growth rate is a weighted average of the growth rates of the components.

In this chapter, we focus on two widely used indices: Laspeyres index and the Paasche index. Let p_{i0} and p_{it} represent the price of the i^{th} commodity in periods 0 and t. q_{i0} and q_{it} are the quantities corresponding to p_{i0} and p_{it}. All of the prices and quantities are non-negative and the number of goods is more than one. Then Laspeyres' price index for period *t*, with base value of unity in period 0, is defined as:

$$I_{0,t}^{L} = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \tag{1}$$

and the corresponding Paasche index is defined as:

$$I_{0,t}^{P}=\frac{\sum_{i=1}^{n}p_{it}q_{it}}{\sum_{i=1}^{n}p_{i0}q_{it}} \qquad . \tag{2}$$

Laspeyres' index, proposed by Laspeyres (1871), measures the change in the price level, if we assume that the consumption levels in the current period are the same as those in the base period. Paasche index, proposed by Paasche (1874), measures the change in the price level, if we assume that the base period consumption levels are the same as the current period ones. When the base period is fixed, we call it a fixed base index. However, when the current period is far away from the base period, the chain index approach is strongly recommended. A chain index is defined as:

$$I_{0,t}^{c}=\prod_{s=1}^{t}I_{s-1,s} \tag{3}$$

where I is defined as any index number, such as those in (1) or (2). If we interchange the role of prices and quantities in the formulae above, then the indices measure the aggregate movement in quantities.

Essentially, there are two ways to motivate these indexes. One method, called the functional approach, constructs the index number by maximizing agents' utility or minimizing their cost. The advantages of this method stem from its strong micro-foundations, but its disadvantage is that the form of the utility function must be known or must be assumed. Frisch (1936) initiated the functional method, and Fisher and Shell (1972), Samuelson and Swamy (1974) and Diewert (1976, 1978) made significant contributions to this approach.

Another method, called the axiomatic approach, identifies certain ideal axioms that the index number formulae should obey, and then assesses the index on the basis of these axioms. Fisher (1922) suggested some tests to narrow down the choice among different index number formulae, such as the time reversal test (which means that if we interchange the base period and the current period, then the index number derived should be the reciprocal of the original index number) , and the circularity test (which means that if the periods cover three or more periods, then the chain index from the first to the last periods should be equal to the fixed base index between these two periods). A more comprehensive summary of the tests that are available to assist in narrowing down the index number formulae can be found in Eichorn and Voeller (1983) and Diewert (1992). Different index formulae can satisfy certain of the tests suggested by Fisher. However, none of them satisfy all of the tests. For example, even Fisher's "ideal" index satisfies all of the tests except the circularity test.

Price index numbers are essentially a weighted average of the relative price between two periods that we want to compare. Different indices can be constructed by choosing different weights for the arithmetic or geometric mean of the relative prices of the n component goods. The fact that different formulae generate different numbers for an index leads to a problem, since all the indices supposedly measure the same price change. So which one should we use? Which one is more reliable (in some sense)? Based on the index number formulae themselves, we cannot answer these questions. Fisher's tests can only reduce the range of choices, but it does not provide a definitive answer. In response to this, there is a body of research that studies the properties of these indices, with a view to determining the choice of the index number formula. Some studies focus on the distortions that arise when a price index is used as a measure of the true "cost of living". A "cost of living" index is defined as the ratio of the (minimum) costs of a given level of living in two price situations (Samuelson and Swamy, 1974). However, the true cost of living index cannot be calculated because the quantities of the commodities purchased in at least one situation are not observable – they are "hypothetical". It is well-known that Laspeyres index is distorted upwards, and the Paasche index is distorted downwards, relative to a true (hypothetical) cost-of-living index, when prices are rising, and *vice versa* when they are falling. See Diewert (1981), Lloyd (1975) and Koop (1986). On the other hand, some studies examine the distortion in the index numbers that is induced by the use of only a subset of the commodities, and the randomness of the data. Allen (1975) and Banerjee (1975) first suggested considering the sampling aspects of the problem in order to examine the properties of price relatives. Kott (1984) applied the super-population theory of Godambe (1955) to study the sampling bias of Laspeyres' price index from the sampling and estimation viewpoints. Braithwait (1980) studied the bias of Laspeyres' price index by taking account of the substitution effects among the different commodities from the changes in the relative prices.

In this chapter, we aim to provide an alternative way to narrow down the choice of the index numbers. The method we take is to examine the stochastic properties of the indices from the point of view of the randomness of the variables. Specifically, we derive analytic results for the first two moments for Laspeyres and Paasche indices under suitable assumptions about the randomness of the data, and then derive some specific numerical results for their bias and

Mean Square Error (MSE) using some illustrative data sets. It is hoped that these analytic and numerical results can provide us with some guidance in the choice between these two types of indices, and our methodology can also be applied to other price index number formulae. Essentially, the approach we introduce has some connections with another method that has been used to interpret certain index numbers, namely the "stochastic approach". Therefore, we introduce the stochastic before we go to discuss our own methodology.

Edgeworth (1925) proposed a method of index number construction called the "stochastic approach". The stochastic approach contributes to the index number literature in two ways. First, the stochastic approach introduces a new way to construct the index number. The stochastic method shifts the emphasis from the descriptive construction of an index number formula to the inferential features of the index number formula. Second, the stochastic method provides a way to study the statistical properties of index numbers from the randomness of the underlying variables. This view of index number construction has been discussed widely in the literature since its introduction, as discussed below. For the stochastic method, we construct a suitable regression model, and then the index is the estimator for the coefficient of the regression model. Therefore, the stochastic approach allows us to find the standard error of the index at each point in the sample; construct confidence intervals; and even determine its full distribution. So, it is possible to test some interesting properties of the index by adopting this approach.

These advantages of the stochastic view of index number construction have attracted considerable interest. Clements and Izan (1987) used the stochastic approach to index theory to estimate the inflation rate and its standard error under the assumption that each commodity price change is independent on the underlying rate of inflation. Selvanathan (1988, 1989, 1991, 1993) undertook a series of studies based on the stochastic approach. For example, Selvanathan (1988) applied the stochastic method to the prices of groups of goods and the prices within groups. Selvanathan (1989) applied the within-group results to alcoholic beverage data and his simulation results showed that the estimates are unbiased, but the asymptotic standard errors underestimate the true sampling variability. Selvanathan (1991) used the stochastic approach to derive the standard errors of Laspeyres and Paasche indices, and he showed that the standard errors increase as the variability of relative prices increases. Finally, Selvanathan (1993) extended these results by allowing for systematic changes in relative prices.

The stochastic approach has also been extended to multilateral price comparisons. Diewert undertook a series of studies on the Country Product Dummy (CPD) method for making international comparisons. The CPD method can actually be viewed as an example of stochastic index number construction. Diewert (2004) used the information on the expenditure weights to derive the weighted generalization of the CPD method. Diewert (2005) applied the unweighted CPD and weighted CPD with expenditure weight or quantity weight to two countries or two period data sets. He also showed that with different weights and different regression variables, the estimator of the regression could be a different index, such as the Walsh price index, the Jevons index and the Dutot index. Coondoo, *et al.* (2004) also proposed a method for estimating multilateral regional price index numbers, which is quite similar to the CPD methods. In addition, there are several theoretical developments and applications relating to multilateral price comparison methods, such as those of Balk (1996), Hill (1997) and Prasada Rao (1997).

Giles and McCann (1994) suggested using the Seemingly Unrelated Regression (SUR) model to account for the underlying correlation between the prices of different commodities. They showed that the SUR method can provide more efficient estimation of the index numbers; it improves the precision of the interval estimates of the price index; and it allows us to test various interesting hypotheses, such as whether or not the change in the index from one period to the next is statistically significant. Crompton (2000) proposed a new estimator of the error variance by using the share of the expenditure to weight the relative price movements. This is robust to the unknown forms of heteroskedasticity in the errors of the underlying regression model. Selvanathan and Prasada Rao (1994) and Clements *et al.* (2006) provide a detailed discussion of the stochastic index number methodology.

Let us illustrate how to use the stochastic approach to construct the index numbers. Following Selvanathan and Prasada Rao (1994), if we can make the (rather strong) assumption that the changes in relative prices across commodities have an expected value of zero, are uncorrelated over commodities, and have a common variance, then we can construct the following regression model:

$$p_{it}^{0} = \gamma_t + \varepsilon_{it}, \qquad i = 1, \cdots, \text{n}$$

where

$$p_{it}^{0} = (p_{it} / p_{i0}) \text{ is the } i^{th} \text{ relative price,}$$

and

$$E[\varepsilon_{it}]=0\,;\;\; Var[\varepsilon_{it}]=\sigma_t^2\,;\;\; Cov[\varepsilon_{it},\varepsilon_{jt}]=0\,,\; i\neq j \tag{4}$$

Then the OLS estimator of γ_t is

$$\hat{\gamma}_t=\frac{1}{n}\sum_{i=1}^{n}p_{it}^{0}$$

which is just a simple un-weighted average of all of the price relatives.

Now, if we replace the assumption in (4) with the following:

$$E[\varepsilon_{it}]=0\,;\;\; Var[\varepsilon_{it}]=\sigma_t^2\big/w_{i0}\,;\;\; Cov[\varepsilon_{it},\varepsilon_{jt}]=0\,,\; i\neq j$$

where

$$w_{i0}=p_{i0}q_{i0}\Big/\sum\nolimits_{i=1}^{n}p_{i0}q_{i0}\,,$$

then the GLS estimator of γ_t is

$$\hat{\gamma}_t=\frac{\sum_{i=1}^{n}p_{it}q_{i0}}{\sum_{i=1}^{n}p_{i0}q_{i0}}\,,$$

which is Laspeyres' price index at time t, with a base value of unity in period 0.

Further, if we replace the assumption in (4) with the following:

$$E[\varepsilon_{it}]=0\,;\;\; Var[\varepsilon_{it}]=\sigma_t^2\big/w_{i0}^{t}\,;\;\; Cov[\varepsilon_{it},\varepsilon_{jt}]=0\,,\; i\neq j$$

where

$$w_{i0}^{t}=p_{i0}q_{it}\Big/\sum\nolimits_{i=1}^{n}p_{i0}q_{it}\,,$$

then the GLS estimator of γ_t is

$$\hat{\gamma}_t=\frac{\sum_{i=1}^{n}p_{it}q_{it}}{\sum_{i=1}^{n}p_{i0}q_{it}}\,,$$

which is Paasche price index in period t, with a base value of unity in period 0..

In a similar manner, we can derive certain other common index numbers. In addition, these index numbers can be obtained by alternative regression estimators, such as Instrumental

Variables (IV) estimation. Suppose the regression model is the following:

$$p_{it} = \gamma_t p_{i0} + \varepsilon_{it} \qquad i = 1, \; ...., \mathrm{n} \quad .$$

If both of the p_{i0} and p_{it} are endogenous, then we can suggest q_{i0} as an instrumental variable for p_{i0}, and the IV estimator for γ_t is

$$\hat{\gamma}_t = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \quad .$$

Alternatively, if we suggest q_{it} as the instrumental variable for p_{i0}, then the IV estimator for γ_t is

$$\hat{\gamma}_t = \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{it}} \quad .$$

Therefore, Laspeyres and Paasche price indices can be constructed as IV estimators, period by period. In addition, if the instruments are asymptotically uncorrelated with the errors in the underlying equations, these estimators will be weakly consistent, although this is not of primary interest here. Within this framework it is also easy to derive the standard errors and confidence intervals for the index, period by period. In this case all of the price and quantity data are random in the model, and this is also the position that we take in our own subsequent analysis in this paper. One difference between our framework and the simple IV framework, however, is that all of the price and quantity variables are assumed to be correlated across the goods and across time in our analysis. This introduces a significant generalization. One way of approaching this, in principle is to adopt a GLS/IV framework.

However, in sections 2 and 3 we do not formulate the indices explicitly as regression coefficient estimators because of some technical complications that then arise with the theoretical results that we draw on. More specifically, idempotency requirements for certain of the matrices that arise in the associated quadratic forms are not satisfied for this problem. This precludes making use of certain established results on ratios of quadratic forms as a basis for setting up small-σ or large-n approximations. None the less, the above discussion on obtaining standard price indices from a regression environment portrays the basic idea of the values of the indices being point estimates at each point in time, with the sample being taken across the quantities and prices of the different goods.

Ullah and Srivastava (1994) apply the small-σ approximation method to derive analytic results for the finite sample moments of statistics that can be written as ratios of quadratic forms of quite general random vectors. In this chapter, we apply Ullah and Srivastava's results to derive the analytic expression for the first two moments of Laspeyres and Paasche price indices. Given the assumption that all the price and quantity variables are random, we can express each index as a ratio of quadratic forms. Then we can apply Ullah and Srivastava's results to obtain some of the finite sample properties of these two indices. We illustrate these analytic results by applying them to two real data sets. In this chapter we essentially make two contributions to the index number literature. First, we provide a new way to examine the statistical properties of certain index numbers. Although we take a position that is similar to that associated with the stochastic approach, our method has some advantages over the stochastic approach. One is that we overcome the need to specify a regression model that involves some very strong, and unrealistic, assumptions. Second, we relax the assumption that the data are uncorrelated across the commodities included in the "basket" of goods that is used to construct the index.

The plan of the rest of this chapter is as follows: section 2 focuses on the theory for approximating the exact finite-sample moments of the ratio of quadratic forms provided by Ullah and Srivastava (1994) and presents our theoretical results for the Laspeyres and Paasche indices based on Ullah and Srivastava's results. Section 3 includes some empirical results that illustrate the theoretical results from section 2. Section 4 contains some conclusions and final comments.

2. Moments of the Indices

In this section, we apply Ullah and Srivastava's (1994) theory to approximate the finite-sample moments of Laspeyres and Paasche indices. First, we introduce Ullah and Srivastava's results. Consider a general ratio of quadratic forms:

$$q^{r_1 r_2} = (Y'N_1Y)^{r_1}(Y'N_2Y)^{-r_2}, \tag{5}$$

where

$$Y \sim N(u, \Omega), \tag{6}$$

N_1 and N_2 are non-stochastic symmetric matrices, and r_1 and r_2 are nonnegative real numbers. First, we standardize, by defining

$$y = \Omega^{-\frac{1}{2}}Y,\ m = \Omega^{-\frac{1}{2}}u,\ M_1 = \Omega^{\frac{1}{2}}N_1\Omega^{\frac{1}{2}} \text{ and } M_2 = \Omega^{\frac{1}{2}}N_2\Omega^{\frac{1}{2}}. \tag{7}$$

So, $y \sim N(m, I)$. Before we state the theory, we introduce the following notation:

$$\begin{aligned}
&\theta_1 = m'M_1m, \qquad \theta_2 = m'M_2m, \\
&c_j = [r_1(r_1-1)\ldots(r_1-j+1)]/\theta_1^j, \\
&A_2 = c_1M_1 + 2c_2M_1mm'M_1 + 2c_1\underline{c}_1M_1mm'M_2, \\
&A_4 = \tfrac{1}{2}c_2M_1 + \tfrac{1}{2}c_1\underline{c}_1M_2 + 2c_3M_1mm'M_1 + 2c_1\underline{c}_2M_2mm'M_2 + 4\underline{c}_1c_2M_1mm'M_2, \\
&A_4^* = \tfrac{2}{3}c_4M_1mm'M_1 + 2c_2\underline{c}_2M_2mm'M_2 + \tfrac{8}{3}\underline{c}_1c_3M_1mm'M_2.
\end{aligned} \tag{8}$$

$\underline{c}_j$ is the same as c_j with r_1 replaced by $-r_2$ and θ_1 by θ_2. c_j is defined for $1 \le j \le r_1$, otherwise $c_j = 0$ for $j > r_1$. But $\underline{c}_j$ is defined for all j. Furthermore, $\underline{A}_j$ in the theory below is A_j with c_j interchanged with $\underline{c}_j$ and M_1 interchanged with M_2. Another restriction we need for the following theory is

$$\theta_2 \neq 0. \tag{9}$$

Then we will apply the following result:

Lemma 1 (Ullah and Srivastava, 1994, p.133)

If the random vector follows (6), and (9) holds, the expectation of (5) to $O(\sigma^4)$ is:

$$E(Y'N_1Y)^{r_1}(Y'N_2Y)^{-r_2} = \theta_1^{-r_1}\theta_2^{-r_2}[1 + \lambda_2 + \lambda_4^*], \tag{10}$$

where

$$\begin{aligned}
\lambda_2 &= tr(A_2 + \underline{A}_2), \\
\lambda_4^* &= (trM_1)(trA_4) + (trM_2)(tr\underline{A}_4) + 2tr(M_1A_4 + M_2\underline{A}_4) \\
&\quad + (m'M_1^2m)(trA_4^*) + (m'M_2^2m)(tr\underline{A}_4^*) + 2m'(M_1A_4^*M_1 + M_2\underline{A}_4^*M_2)m.
\end{aligned}$$

We apply Lemma 1 to Laspeyres' and Paasche's price indices, and determine the analytical expressions to approximate the exact first two moments of these two indices. Although we do not follow the stochastic method to construct the index number based on the regression model, we do follow the spirit of the stochastic approach and assume that all of the price and quantity variables are random. Then, the index number can be written as a ratio of quadratic forms and Lemma 1 can be applied. Here, we define the random vector as

$$Y = (P_0' \;\; P_t' \;\; Q_0' \;\; Q_t')' .$$

We assume that the random vector Y follows (6) - that is, $Y \sim N(u, \Omega)$.

Here, $P_0 = (p_{10}, p_{20}, \cdots p_{n0})'$ represents the price vector of n goods in the base period

$P_t = (p_{1t}, p_{2t}, \cdots p_{nt})'$ represents the price vector of n goods in the current period

$Q_0 = (q_{10}, q_{20}, \cdots q_{n0})'$ represents the quantity vector of n goods in the base period

$Q_t = (q_{1t}, q_{2t}, \cdots q_{nt})'$ represents the quantity vector of n goods in the current period.

From the above definition, we know that the dimension of Y is $4n \times 1$. Therefore, u is $4n \times 1$ and Ω is $4n \times 4n$. We also know that Y is time-dependent, so a more appropriate representation would be

$$Y_t \sim N(u_t, \Omega_t).$$

Clearly, any variable which is of a function of Y_t is also time-dependent. However, in this chapter, to simplify the notation, we drop the "t" subscript that is being used to signify the time-dependence of certain variables.

The Paasche index can be written as

$$I^P = \frac{Y' N_1^P Y}{Y' N_2^P Y} \tag{11}$$

where

$$N_1^P = \frac{1}{2}\begin{bmatrix} 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & I_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & I_{n\times n} & 0_{n\times n} & 0_{n\times n} \end{bmatrix}_{4n\times 4n}, \qquad N_2^P = \frac{1}{2}\begin{bmatrix} 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & I_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ I_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \end{bmatrix}_{4n\times 4n}$$

and $I_{n\times n}$ is an ($n \times n$) identity matrix. We apply the same transformation as in (7), and then we can apply Lemma 1 to approximate the exact moments of Paasche index.

For the mean of the Paasche index, $r_1 = r_2 = 1$. Then, according to equation (8):

$$\theta_1 = u'N_1^P u, \qquad \theta_2 = u'N_2^P u,$$

$$c_1 = 1/\theta_1,\ c_j = 0,\ j = 2,\ 3,\ 4,\ \cdots,$$

$$\underline{c}_1 = -1/\theta_2,\ \underline{c}_2 = 2/\theta_2^2,\ \underline{c}_3 = -6/\theta_2^3,\ \underline{c}_4 = 24/\theta_2^4.$$

$$A_2 = \Omega^{\frac{1}{2}}(\frac{N_1^P}{\theta_1} - 2\frac{N_1uu'N_2^P}{\theta_1\theta_2})\Omega^{\frac{1}{2}},$$

$$\underline{A}_2 = -\Omega^{\frac{1}{2}}(\frac{N_2^P}{\theta_2} - 4\frac{N_2^Puu'N_2^P}{\theta_2^2} + 2\frac{N_2^Puu'N_1^P}{\theta_1\theta_2})\Omega^{\frac{1}{2}},$$

$$A_4 = -\Omega^{\frac{1}{2}}(\frac{N_2^P}{2\theta_1\theta_2} - 4\frac{N_2^Puu'N_2^P}{\theta_1\theta_2^2})\Omega^{\frac{1}{2}},$$

$$\underline{A}_4 = \Omega^{\frac{1}{2}}(\frac{N_2^P}{\theta_2^2} - \frac{N_1^P}{2\theta_1\theta_2} - 12\frac{N_2^Puu'N_2^P}{\theta_2^3} + 8\frac{N_2^Puu'N_1^P}{\theta_1\theta_2^2})\Omega^{\frac{1}{2}},$$

$$A_4^* = 0,$$

$$\underline{A}_4^* = 16\Omega^{\frac{1}{2}}(\frac{N_2^Puu'N_2^P}{\theta_2^4} - \frac{N_2^Puu'N_1^P}{\theta_1\theta_2^3})\Omega^{\frac{1}{2}}.$$

Substituting the expressions above into (10), we get

$$E(I^P) = \frac{\theta_1}{\theta_2}(1 + \lambda_2 + \lambda_4^*) \qquad . \tag{12}$$

Similarly, for the second raw moment of the Paasche index, $r_1 = r_2 = 2$, according to (8),

$$\theta_1 = u'N_1^P u, \qquad \theta_2 = u'N_2^P u,$$

$$c_1 = 2/\theta_1,\ c_2 = 2/\theta_1^2,\ c_j = 0,\ j = 3, 4, \cdots,$$

$$\underline{c}_1 = -2/\theta_2,\ \underline{c}_2 = 6/\theta_2^2,\ \underline{c}_3 = -24/\theta_2^3,\ \underline{c}_4 = 120/\theta_2^4.$$

$$A_2 = 2\Omega^{\frac{1}{2}}(\frac{N_1^P}{\theta_1} + 2\frac{N_1^Puu'N_1^P}{\theta_1^2} - 4\frac{N_1^Puu'N_2^P}{\theta_1\theta_2})\Omega^{\frac{1}{2}},$$

$$\underline{A}_2 = -2\Omega^{\frac{1}{2}}(\frac{N_2^P}{\theta_2} - 6\frac{N_2^Puu'N_2^P}{\theta_2^2} + 4\frac{N_2^Puu'N_1^P}{\theta_1\theta_2})\Omega^{\frac{1}{2}},$$

$$A_4 = \Omega^{\frac{1}{2}}(\frac{N_1^P}{\theta_1^2} - 2\frac{N_2^P}{\theta_1\theta_2} + 24\frac{N_2^Puu'N_2^P}{\theta_1\theta_2^2} - 16\frac{N_1^Puu'N_2^P}{\theta_1^2\theta_2})\Omega^{\frac{1}{2}},$$

$$\underline{A}_4 = \Omega^{\frac{1}{2}}(3\frac{N_2^P}{\theta_2^2} - 2\frac{N_1^P}{\theta_1\theta_2} - 48\frac{N_2^Puu'N_2^P}{\theta_2^3} - 8\frac{N_1^Puu'N_1^P}{\theta_1^2\theta_2} + 48\frac{N_2^Puu'N_1^P}{\theta_1\theta_2^2})\Omega^{\frac{1}{2}},$$

$$A_4^* = 24\frac{\Omega^{\frac{1}{2}} N_2^P uu' N_2^P \Omega^{\frac{1}{2}}}{\theta_1^2\theta_2^2},$$

$$\underline{A}_4^* = 4\Omega^{\frac{1}{2}}(20\frac{N_2^P uu' N_2^P}{\theta_2^4} + 6\frac{N_1^P uu' N_1^P}{\theta_1^2\theta_2^2} - 32\frac{N_2^P uu' N_1^P}{\theta_1\theta_2^3})\Omega^{\frac{1}{2}}.$$

Substituting above expressions into (10), we get

$$E[(I^P)^2] = \frac{\theta_1^2}{\theta_1^2}(1 + \lambda_2 + \lambda_4^*) \tag{13}$$

Similarly, we can use the same method to find the mean and variance of Laspeyres' index. To write Laspeyres' index as the ratio of quadratic forms, we define

$$N_1^L = \frac{1}{2}\begin{bmatrix} 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & I_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & I_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \end{bmatrix}_{4n\times 4n}, \qquad N_2^L = \frac{1}{2}\begin{bmatrix} 0_{n\times n} & 0_{n\times n} & I_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ I_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \end{bmatrix}_{4n\times 4n}.$$

Then the mean and variance of Laspeyres' index are obtained from (12) and (13), just replacing N_1^P and N_2^P with N_1^L and N_2^L. Of course, we can also obtain the higher-order moments for the Paasche and Laspeyres indices in the same manner, if these are of interest.

The analytic results above depend on the normality assumption. The normality assumption is quite restrictive here in the sense that both the prices and quantities are non-negative. Therefore, strictly, the random vector Y we defined cannot be drawn from a normal distribution. Relaxing the normality assumption could make the results more attractive and applicable. Allowing the random vector to be both non-normal and non-i.i.d will make the problem quite complicated. Since the random vector Y includes both prices and quantities, and given the interaction of prices and quantities in the market, we keep the non-i.i.d assumption, but we have retained the normality assumption.

3. Some Numerical Evaluations

The theoretical results that have been derived above are now illustrated by applying them to some real data sets. The first set of data that we consider relates to Australian prices and expenditure for alcoholic beverages, and is reported by Clements and Daryal (1999). The second data set is the well-known food consumption data for Sweden, from Wold and Jureen (1953, pp.278-279). Obviously, in practice we will not know the true mean and covariance matrix of the stochastic data, therefore, we will just use the sample mean and covariance matrix to estimate the moments of the index based on (12) and (13). Of course, based on the first and second moments, we can then calculate the bias and mean squared error. We need to note here is that the notion of "bias" that we have in mind here is different from the standard definition of bias. Since we don't know the true value of the index, the bias is defined here as the difference between the mean value of the index and the value of the index calculated from the sample. Therefore, we calculated the bias as the mean in (12) minus the index calculated from (1) or (2) depending on which index we are using. We also calculate the percentage bias by dividing the bias by the estimator from (1) or (2). Similar comments apply to the so-called "mean squared error" values that we report.

The main practical difficulty that is encountered in providing numerical illustrations is that of estimating the mean and covariance matrix of the stochastic vector with only one observation for each year. We have considered different methods to estimate the mean and covariance matrix for the stochastic vector by using different samples. First, we just use the information from the base period and the current period to construct an estimator. With this method, the estimated mean vector $\hat{u}$ is:

$$\hat{u} = (\hat{u}_1, \hat{u}_2, \cdots \hat{u}_{4n})' \tag{14}$$

where

$$\hat{u}_i = \frac{p_{i0} + p_{it}}{2}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

$$\hat{u}_{i+n} = \frac{p_{i0} + p_{it}}{2}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

$$\hat{u}_{i+2n} = \frac{q_{i0} + q_{it}}{2}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

$$\hat{u}_{i+3n} = \frac{q_{i0} + q_{it}}{2}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

Then, the estimated covariance matrix $\hat{\Omega}$ is constructed as follows:

$$\hat{\Omega}_{ij} = (y_i - \hat{u}_i)(y_j - \hat{u}_j), \qquad \text{for } \mathrm{i},\ \mathrm{j} = 1, 2, \cdots, 4\mathrm{n}. \tag{15}$$

where $\hat{\Omega}_{ij}$ is the element in the i^{th} row and the j^{th} column of $\hat{\Omega}$.

We also consider a "rolling method", which means that as time goes by, we include all available information to undertake the estimation. With this method, the mean vector $\hat{u}$ is constructed as follows:

$$\hat{u}_i = \sum_{j=0}^{t} p_{ij}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

$$\hat{u}_{i+n} = \sum_{j=0}^{t} p_{ij}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

$$\hat{u}_{i+2n} = \sum_{j=0}^{t} q_{ij}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n};$$

$$\hat{u}_{i+3n} = \sum_{j=0}^{t} q_{ij}, \qquad \text{for } \mathrm{i} = 1, 2, \cdots, \mathrm{n}; \tag{16}$$

and $\hat{\Omega}$ is constructed the same way as (15), just replacing the mean vector (14) with (16).

Of course, other similar methods can also be considered. For example, we can use three-period information to estimate the mean vector. However, in practice, for all of these similar methods, when the current period is far away from the base period, the estimated covariance matrix can be negative-definite. This seems to suggest that it is not advisable to base the estimator on information which covers a long time-span. One possible explanation for this problem is that the prices tend to rise over time. Therefore, more stable results can be obtained by limiting the time-span over which the information is drawn. This point is consistent with the argument that when the current period is far from the base period, we should use a chain index number as in (3), instead of a fixed-base index number. Another possibility when it comes to estimating the mean vector is to use a quantity-weighted average across time periods, rather than a simple arithmetic average. Hopefully, further research into these matters will help to determine a preferred way of dealing with this empirical issue.

Therefore, we also consider another strategy to estimate the mean vector and the covariance matrix. We use the base period and just the following period to represent the information for the base period. Similarly, we use the current period and the previous one period to represent the information for the current period. This approach means that we have two observations for the base and current period, not only one. Then we can use this information to estimate the mean and the covariance matrix. We call this method the "two-matrix method". Based on this method, the random vector Y is defined as follows:

$$Y_j = (P'_{0j} \quad P'_{tj} \quad Q'_{0j} \quad Q'_{tj})', \qquad \text{for } j = 1,\ 2 \tag{17}$$

where

$$P_{0j} = (p_{1(j-1)},\ p_{2(j-1)},\ \cdots,\ p_{n(j-1)})';$$

$$P_{tj} = (p_{1(j-2+t)},\ p_{2(j-2+t)},\ \cdots,\ p_{n(j-2+t)})';$$

$$Q_{0j} = (q_{1j-1},\ q_{2j-1},\ \cdots,\ q_{nj-1})';$$

$$Q_{tj} = (q_{1(j-2+t)},\ q_{2(j-2+t)},\ \cdots,\ q_{n(j-2+t)})'.$$

Therefore, we can easily estimate the mean vector and covariance matrix based on these two observations for the random vector Y. The estimated mean vector $\hat{u}$ is:

$$\hat{u}_i = \sum_{j=1}^{2} p_{i(j-1)}, \qquad \text{for } i = 1, 2, \cdots, n;$$

$$\hat{u}_{i+n} = \sum_{j=1}^{2} p_{i(j-2+t)}, \qquad \text{for } i = 1, 2, \cdots, n;$$

$$\hat{u}_{i+2n} = \sum_{j=1}^{2} q_{i(j-1)}, \qquad \text{for } i = 1, 2, \cdots, n;$$

$$\hat{u}_{i+3n} = \sum_{j=1}^{2} q_{i(j-2+t)}, \qquad \text{for } i = 1, 2, \cdots, n. \tag{18}$$

The estimated covariance matrix, $\hat{\Omega}$, is constructed as follows:

$$\hat{\Omega}_{ij} = \frac{\sum_{t=1}^{2} (y_{it} - \hat{u}_i)(y_{jt} - \hat{u}_j)}{2}, \qquad \text{for } i,\ j = 1, 2, \cdots, 4n. \tag{19}$$

For this method, if we set the first period as the base period, then the mean for the index in the second period will be unity. Therefore, it is better to ignore the inference derived for the second year. We also try another way to estimate the covariance matrix for the base period and the current period. After we get the mean in (18), we combine the estimated mean vector (18) and the construction of the covariance matrix in (15) to estimate the covariance matrix, which means we ignore the information from another period for the covariance matrix. And we also expand these two methods to include three or more periods for the base and the current period. Among all of the methods, the two-matrix method turns out to be the best one, in the sense that all of the other methods give some unstable results for certain periods or certain data sets.

Table 1 provides the empirical results for the mean, bias, and MSE and the percentage bias of the Paasche index and Laspeyres index for the Australian alcoholic beverages data set. Table 2 gives the corresponding results for the food consumption data set for Sweden. For the alcoholic beverages data set, we set 1988 as the base year. The results show that for both the Paasche and Laspeyres indices, the bias is negative and the percentage bias ranges from 3% to 6%. The Laspeyres index has smaller MSE than the Paasche index, but the difference is not very significant. All of this suggests that there is very little difference between the Paasche and Laspeyres indices for the alcoholic beverage data set. For the food consumption data set, 1921 is set as the base year. The results show that both of the indices exhibit a positive bias. The Paasche index has a percentage bias between 10% and 30%, and Laspeyres index has a percentage bias between 15% and 30%. A comparison of MSE shows that the Paasche index is preferred to Laspeyres index, which is the opposite result from that for the alcoholic beverage data set. The economic implications of the numerical results are also different for these two data sets. For the first data set, the bias is relatively small, so the effects of ignoring the bias in practice are not likely to be big. In particular, if the usual calculated values of the price indices are used to measure inflation over time, and to influence economic policy, this should not be problematic. However, for the second data set, ignoring the bias may have some important economic implications. In this case more care would need to be taken if the usual calculated values of the indices were used as the basis for policy analysis.

Comparing the results from these two tables, we reach the following conclusions. First, the signs of the biases of the Paasche and Laspeyres indices are the same, and the magnitudes are very similar, for a fixed data set. Second, these signs and magnitudes depend on the data set. For example, for the alcoholic beverages data set, both of the indices have a negative bias, and the percentage biases are very similar. However, for the food consumption data set, the biases are positive and the magnitude of the percentage biases suggest that we definitely should pay attention to them when we construct these two indices from these data. Third, any comparison of the MSE's of these two indices also depends on the data set. For the first data set, Laspeyres' index has a smaller MSE than the Paasche index, but for the second data set, the converse is true. Basically, no definitive conclusions can be drawn. The biases and MSEs depend on the data set in question and so do the economic implication of these numerical results. In addition, of course, the quality of our numerical results depends on how well we can estimate the mean vector and the covariance matrix.

It is important to recognize that our analytical results are based on quite strong assumptions about the underlying data. In particular, the normality assumption and the methods that we have used to estimate the mean vector and the covariance matrix assume that the data are stationary. From the tables, we can see the indices for the alcoholic beverages data set increase over time, so we might expect that the stationary assumption may cause more serious problem for this data set than for the Swedish consumption data set. Some basic unit root testing suggests that all of the price *indices* are non-stationary. Of course, in these examples the sample sizes are rather small and the usual unit root tests will have very low power. None the less, this rather tenuous evidence might lead us to question the related assumption that the *individual* price and quantity data are drawn from a stationary distribution. This is an issue that warrants further study.

4. Conclusions

In this chapter we have derived analytic results to approximate the exact moments of the Paasche and Laspeyres index number series. However, the analytical results are too complicated to enable us to draw any general conclusions. We illustrate the results by means of some numerical applications based on two real sets of data. These numerical calculations enable us to get a feel for the signs and orders of magnitudes of the biases associated with these index numbers, when they are viewed as statistics. In the same way, we are able to explore the relative efficiencies of the Paasche and Laspeyres indices in specific contexts.

There are three aspects of our analytical and numerical results that deserve further consideration in future research. The first is the assumed normality of the price and quantity data. This rather strong assumption deserves to be relaxed, as does the assumption of i.i.d. data. In this regard the general results provided by Lieberman (1997) and by Bao and Ullah (2007) are likely to be helpful. The second feature of our work that needs further development is the methodology for evaluating the biases and mean squared errors of the indices in practice. Alternative methods for estimating the mean and the covariance matrix of the data need to be considered in order to make our results more useful in practical situations. Finally, as the numerical results that we have provided here are clearly sensitive in various ways to the choice of data, some Monte Carlo analysis would be warranted in order to explore which particular characteristics of the data are important to the nature of the results.

Table 1: Price Indices for Alcohol Expenditure in Australia

	Paasche Index					Laspeyres Index				
Year	Index	Mean	"Bias"	"MSE"	Percentage "Bias" (%)	Index	Mean	"Bias"	"MSE"	Percentage "Bias" (%)
1989	1.0544	1.0000	-0.0544	0.0030	-5.1616	1.0545	1.0000	-0.0545	0.0030	-5.1709
1990	1.1205	1.0588	-0.0617	0.0039	-5.5070	1.1202	1.0590	-0.0612	0.0039	-5.4647
1991	1.1699	1.1155	-0.0544	0.0033	-4.6539	1.1700	1.1183	-0.0517	0.0030	-4.4218
1992	1.2039	1.1594	-0.0444	0.0053	-3.6904	1.2039	1.1628	-0.0411	0.0048	-3.4149
1993	1.2467	1.1927	-0.0540	0.0038	-4.3281	1.2441	1.1946	-0.0494	0.0034	-3.9743
1994	1.2891	1.2344	-0.0548	0.0036	-4.2480	1.2853	1.2328	-0.0524	0.0033	-4.0795
1995	1.3399	1.2798	-0.0602	0.0040	-4.4913	1.3363	1.2773	-0.0590	0.0038	-4.4125
1996	1.3948	1.3309	-0.0639	0.0044	-4.5789	1.3917	1.3292	-0.0625	0.0041	-4.4907
1997	1.4291	1.3749	-0.0541	0.0044	-3.7888	1.4260	1.3736	-0.0524	0.0040	-3.6722
1998	1.4476	1.4007	-0.0469	0.0046	-3.2418	1.4437	1.3992	-0.0446	0.0042	-3.0868

Table 2: Price Indices for Food Consumption in Sweden

	Paasche Index					Laspeyres Index				
Year	Index	Mean	"Bias"	"MSE"	Percentage "Bias" (%)	Index	"Mean"	"Bias"	"MSE"	Percentage "Bias" (%)
1922	0.7731	1.0000	0.2269	0.0515	29.3430	0.7733	1.0000	0.2267	0.0514	29.3177
1923	0.6667	0.8446	0.1779	0.0403	26.6871	0.6696	0.8441	0.1745	0.0439	26.0591
1924	0.6448	0.7679	0.1231	0.0327	19.0924	0.6501	0.7764	0.1263	0.0406	19.4299
1925	0.6732	0.8160	0.1427	0.1106	21.2019	0.6878	0.8681	0.1803	0.1648	26.2196
1926	0.6356	0.7475	0.1119	0.0270	17.6100	0.6488	0.7792	0.1304	0.0346	20.0930
1927	0.5995	0.7065	0.1071	0.0288	17.8601	0.6151	0.7600	0.1449	0.0526	23.5624
1928	0.5921	0.6922	0.1001	0.0443	16.9063	0.6127	0.7434	0.1307	0.0597	21.3306
1929	0.5760	0.6783	0.1023	0.0402	17.7632	0.6028	0.7323	0.1295	0.0539	21.4903
1930	0.5352	0.6470	0.1118	0.0351	20.8960	0.5590	0.7144	0.1554	0.0574	27.7964
1931	0.4716	0.5592	0.0876	0.0119	18.5630	0.4936	0.6251	0.1315	0.0284	26.6348
1932	0.4530	0.5180	0.0650	0.0137	14.3552	0.4683	0.5656	0.0972	0.0210	20.7644
1933	0.4424	0.5183	0.0760	0.0256	17.1793	0.4607	0.5751	0.1144	0.0437	24.8210
1934	0.4517	0.5040	0.0523	0.0160	11.5757	0.4703	0.5502	0.0799	0.0229	16.9793
1935	0.4912	0.5602	0.0691	0.0466	14.0582	0.5170	0.6213	0.1044	0.0712	20.1899
1936	0.5125	0.5705	0.0580	0.0214	11.3217	0.5424	0.6267	0.0843	0.0318	15.5349
1937	0.5368	0.6108	0.0740	0.0371	13.7898	0.5661	0.6725	0.1065	0.0526	18.8059
1938	0.5656	0.6352	0.0696	0.0345	12.3126	0.5968	0.6979	0.1010	0.0500	16.9313

CHAPTER 6:

SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

1. Overview

Many of the principal results that we draw on in econometrics rely on the use of the asymptotic distributions of our estimators and test statistics. However, in many situations we have only limited knowledge of their finite sample properties. Moreover, very often what knowledge we do have is based on rather limited Monte Carlo evidence that may lack generality. Inferences that are based on the asymptotic properties of an estimator or a test statistic may be very unreliable for small or even moderately large samples. Therefore, it is important to broaden our knowledge of the finite-sample properties of such econometric estimators and tests, and it is especially important to do this by deriving further analytical results that are general and will help us to get a more complete understanding of the quality of the inferences that we draw in practice. So, the objective of this dissertation is to develop several new analytical finite-sample results that will be of direct benefit to econometricians.

Specifically, we have explored the use of several finite-sample methods, including the saddlepoint approximation, the large-n approximation, and the small-σ approximation. We have developed some new analytic results relating to the saddlepoint approximation, and we have used these to study the finite-sample properties of an interesting estimator that arises in the empirical macroeconomics literature, and a goodness-of-fit test that is very widely used throughout the statistics literature. Our use of the large - n (Nagar) approximation allows us to extend the surprisingly limited literature relating to the small-sample properties of the Maximum Likelihood estimator for binary choice models. We are also able to provide new insights into the properties of some standard price indices by viewing them as estimators of the true underlying prices, and applying small-σ approximations to assess their bias and mean squared error.

The objective of Chapter 2 is to provide a resolution to the so-called Purchasing Power Parity (PPP) puzzle. Essentially, we make two contributions to the assessment of the PPP puzzle. First, we apply Lieberman's (1994a, 1994b) results in order to derive saddlepoint approximations to the density and distribution functions of the convergent half-life estimator.

As far as is known, these are the first such analytic results that have been derived, although Kilian and Zha (1999) provide some related Bayesian results. Second, we prove that none of the integer-order moments of this estimator exist. This suggests that the popular use of the half-life based on an AR model may not be an appropriate way to measure convergence to PPP. It also explains why, in practice, various authors have reported unrealistically wide "confidence intervals" for their half-life estimates.

In Chapter 3, we compare the accuracy of four different saddlepoint approximations when they are applied to the distribution of the Anderson-Darling (A-D) test statistic. These include the normal-based and non-normal-based saddlepoint approximations to order $O(n^{-1})$ and $O(n^{-2})$ respectively. The contributions of this chapter include both theoretical and empirical results. Our derivation of the non-normal based saddllepoint approximation to order $O(n^{-2})$ is apparently quite novel, and this can be used in a range of other applications.

The numerical results in this chapter show the following. Improvements in the approximation to the middle part of the distribution can be obtained in two directions: by moving from the normal-based saddlepoint approximation to the non-normal-based one; and by moving from the lower-order saddlepoint approximation to the higher-order one. The preference of the non-normal-based saddlepoint approximation over the normal-based one in the middle of the density is due to the fact that the A-D test statistic is non-normally distributed asymptotically. We might expect that a normal-based saddlepoint approximation will perform well when the asymptotic distribution of the statistic is normal; and the non-normal-based saddlepoint approximation might perform well when the asymptotic distribution of the statistic is non-normal. However, these improvements that come from the non-normal-based and higher-order saddlepoint approximation are limited to the middle of the density, and do not apply to the tail areas, at least in the case that we studied. Therefore, given that we are interested primarily in the tail areas, for construction of accurate critical regions for the test, the normal-based saddlepoint approximation to order $O(n^{-1})$ appears to be adequate.

In chapter 4, we apply the large-n approximation to derive an analytic expression for the mean and MSE of the MLE of the coefficients in the binary Logit model. We also illustrate the analytic results by applying them to a simple model with only one regressor. Since there are very few results on the finite-sample properties of estimators for qualitative response models, we have very little basis for comparing our results with other previous work. Our results are more general than any others relating to this topic as we allow the covariates in the model to be random, rather than fixed in repeated samples.

Based on the analytic results, we report some numerical results for the MLE, its bias-corrected counterpart, and the associated actual and estimated standard deviations. We also report the *estimated* bias-corrected estimator. We find that incorporating the (estimated) bias correction can improve the MLE, not only in terms of bias reduction, but also in terms of reduced variability.

In Chapter 5, we use a small-σ approximation for the ratio of quadratic forms in a random vector to derive analytic expressions for the biases and MSE of two price index numbers, namely the Laspeyres Index and the Paasche Index. Then we generate some illustrative numerical results based on our analytic results. We report the mean, bias and MSE and the percentage bias of these two index numbers for two data sets. One is an Australian alcoholic beverages data set, and the other one is a food consumption data set for Sweden.

Our numerical results show that for a particular data set, the difference between these two index numbers is very small. It is difficult to make a choice between these two index numbers based on a comparison of their mean squared errors. The signs and magnitudes of the biases of these two index numbers depend on the data set. For one of the data sets that we consider, the biases of both of the prices indices are small. For the second data set, the biases are much larger, which would be of some consequence in practice. Certainly, one can conclude that care should be taken when constructing these price indices and then using the results as point estimates of the underlying population price changes.

2. Future Research

In this dissertation we have derived some encouraging analytic and numerical results based on finite-sample theory. However, there is still a lot of potential for further research into each of the topics that we have examined.

In chapter 2, our results suggest the convergence half-life is not the appropriate measure to use when considering PPP convergence. Therefore, finding a more appropriate measure, and in particular one whose estimator has finite moments, is an important problem that needs to be resolved in future research into the PPP puzzle.

In chapter 3, we derived a new higher-order, non-normal-based saddlepoint approximation, which can be applied to quite a large number of distributional problems. In the future, we expect to apply the higher-order saddlepoint approximation to some other statistics to illustrate its performance.

In chapter 4, we only illustrate our new analytic results by applying them to the Logit model with only a single regressor. In future work, we expect to extend the numerical results to the binary Logit model with multiple regressors. In addition, the analytic results could potentially be extended to the multinomial Logit model. Taking the same approach in the case of the widely used Probit model could be problematic, given the integrals involved, but this is certainly worthy of further consideration.

In relation to chapter 5, we expect to extend the results in at least three directions. First, we assumed that the random vector which includes the prices and quantities in the base and current periods is normal distributed, and it would be useful to relax this rather strong assumption. Second, as there is only one observation for each time period, in the future, we expect to consider alternative ways of constructing the sample moments when putting our analytic results into practice. Finally, the small-σ approximation that we used to evaluate the Laspeyres and the Paasche price indices in this chapter could also be applied to other related indices, such as Fisher's ideal index and the Divisia index. There is considerable potential for further research into this stochastic approach to index number construction.

Overall, we feel that this research has provided some interesting results, which in turn have important implications for economists undertaking empirical work. In order to make our conclusions more comprehensive and robust, there is scope for further research into these topics.

REFERENCES

Abuaf, N. and Jorion, P. (1990), "Purchasing Power Parity in the Long Run", *Journal of Finance*, 45, 157-174.

Allen, R. G. D. (1975), *Index Numbers in Theory and Practice*, New York, Macmillan.

Amemiya, T. (1980), "The n^{-2}-order Mean Squared Errors of the Maximum Likelihood and the Minimum Logit Chi-square Estimator", *Annals of Statistics,* 8, 488-505.

Andrews, D. W. K. (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models", *Econometrica*, 61, 139-165.

Anderson, T. W. and Darling, D. A. (1952), "Asymptotic Theory of Certain "Goodness of Fit", Criteria Based on Stochastic Processes", *Annals of Mathematical Statistics*, 23, 193-212.

Andrews, T. W. and Darling, D. A. (1954), "A Test of Goodness of Fit", *Journal of the American Statistical Association*, 49, 765-769.

Balk, B. M. (1996), "A Comparison of Ten Methods for Multilateral International Price and Volume Comparison", *Journal of Official Statistics*, 12, 199-222.

Banerjee, K. S. (1975), *Cost of Living Index Numbers-Practice, Decision and Theory*, Marcel Dekker, New York

Bao, Y. and Ullah, A. (2006), "The Second-Order Bias and Mean Squared Error of Estimators in Time-Series Model," *Journal of Econometrics*, forthcoming.

Barndorff-Nielsen (1991), "Modified Signed Log-Likelihood Ratio", *Biometrika*, 78, 557-563.

Barndorff-Nielsen, O. and Cox, D. R. (1979),"Edgeworth and Saddle-Point Approximations with Statistical Applications", *Journal of the Royal Statistical Society*, B, 41, 279-312.

Baum, C. F., Barkoulas, J. T. and Caglayan, M. (2001), "Nonlinear Adjustment to Purchasing Power Parity in the Post-Bretton Woods Era", *Journal of International Money and Finance,* 20, 379-399.

Berkson, J. (1944), "Application of the Logistic Function to Bio-Assay", *Journal of the American Statistical Association*, 39, 357-365.

Berkson, J. (1955), "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function", *Journal of the American Statistical Association*, 50, 130-162.

Braithwait, S. D. (1980), "The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes", *American Economic Review*, 70, 64-77.

Birnbaum, Z. W. (1952), "Numerical Tabulation of the Distribution of Kolmogorov's Statistic for Finite Sample Size", *Journal of the American Statistical Association*, 47, 425-441.

Butler, R. W. and Huzurbazar, S. (1992), "Saddlepoint Approximations for the Bartlett-Nanda- Pillai Trace Statistic in Multivariate Analysis", *Biometrika*, 79, 705-715.

Butler, R. W. and Sutton, R. W. (1998), "Saddlepoint Approximation for Multivariate Cumulative Distribution Functions and Probability Computations in Sampling Theory and Outlier Testing", *Journal of the American Statistical Association*, 93,596-604.

Cameron, A. C. and P. K. Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge.

Caporale, G. M., Cerrato, M. and Spagnolo, N. (2005), "Measuring Half-Lives Using a Non-Parametric Bootstrap Approach", *Applied Financial Economics Letters*, 1, 1-4.

Cashin, P. and McDermott, C. J. (2003), "An Unbiased Appraisal of Purchasing Power Parity", *International Monetary Fund Staff Papers*, 50, 321-351.

Caves, D. W., Christensen, L. R. and Diewert, W. E. (1982), "Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers", *Economic Journal,* 92, 73-86.

Cheung, Y. W. and Lai, K. S. (1994), "Mean Reversion in Real Exchange Rates", *Economics Letters,* 46, 251-256.

Cheung, Y. W. and Lai, K. S. (1998), "Parity Reversion in Real Exchange Rates during the Post-Bretton Woods period", *Journal of International Money and Finance*, 17, 597-614.

Cheuing, Y. W., Lai, K. S. and Bergman, M. (2004), "Dissecting the PPP Puzzle: the Unconventional Roles of Nominal Exchange Rate and Price Adjustments", *Journal of International Economics*, 64, 135-150

Choi, C. Y., Mark, N. C. and Sul. D. (2004), "Unbiased Estimation of the Half-Life to PPP Convergence in Panel Data", National Bureau of Economic Research, Working Paper No. W10641.

Chortareas, G. and Kapetanios, G. (2004), "How Puzzling is the PPP Puzzle? an Alternative Half-Life Measure of Convergence to PPP", University of London, Working Paper No. 522.

Clements, K. W. and Daryal, M. (1999), "The Economic of Marijuana Consumption", mimeo., Economic Research Centre, Department of Economics, University of Western Australia.

Clements, K. W. and Izan, H. Y. (1987), "The Measurement of Inflation: A Stochastic Approach", *Journal of Business and Economic Statistics*, 5, 339-350.

Clements, K. W., Izan, H.Y. and Selvanathan, E. A. (2006), "Stochastic Index Numbers: A Review", *International Statistical Review*, 74, 235-270.

Coondoo, D., Majumder, A. and Ray, R. (2004), "A Method of Calculating Regional Consumer Price Differentials with Illustrative Evidence from India", *Review of Income and Wealth*, 50, 51-68.

Crompton, P. (2000), "Extending the Stochastic Approach to Index Numbers", *Applied Economics Letters*, 7, 367-371.

Daniels, H. E. (1954), "Saddlepoint Approximations in Statistics", *the Annals of Mathematical Statistics*, 25, 631-650.

Daniels, H. E. (1980), "Exact Saddlepoint Approximations", *Biometrika*, 67, 59-63.

Daniels, H. E. (1983), "Saddlepoint Approximations for Estimating Equations", *Biometrika*, 70, 89-96.

Daniels, H. E. (1987), "Tail Probability Approximations", *International Statistical Review*, 55, 37-48.

Davies, R. B. (1973), "Numerical Inversion of a Characteristic Function", *Biometrika*, 60, 415-417.

Davis, L. (1984), "Comments on a Paper by T. Amemiya on Estimation in a Dichotomous Logit Regression Model", *Annals of Statistics,* 12, 778-782.

Diebold, F. X., Husted, S. and Rush, M. (1991), "Real Exchange Rates under the Gold Standard", *Journal of Political Economy,* 99, 1252-1271.

Diewert, W. E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics*, 4, 115-145.

Diewert, W. E. (1978), "Superlative Index Numbers and Consistency in Aggregation", *Econometrica*, 46, 883-900.

Diewert, W. E. (1981), "The Economic Theory of Index Numbers: A Survey", in *Essays in the Theory and Measurement of Consumer Behaviour (in Honour of Richard Stone*), Deaton, A. (ed.), Cambridge University Press, New York, 163-208.

Diewert, W. E. (1992), "Fisher Ideal Output, Input, and Productivity Indexes Revisited", *Journal of Productivity Analysis,* 3, 211-248.

Diewert, W. E. (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Discussion Paper No. 04-16, Department of Economics, The University of British Columbia.

Diewert, W. E. (2005), "Weighted Country Product Dummy Variable Regressions and Index Number Formulae", *Review of Income and Wealth,* 51, 561-570.

Divisia, F. (1925), "'L'Indice Monetaire et la Theorie de la Monnaie", *Revue d'Economic Politique*, 39, 980-1008.

Easton, G. S. and Ronchetti, E. (1986), "General Saddlepoint Approximations with Applications to L Statistics", *Journal of the American Statistical Association*, 81, 420-430.

Edgeworth, F. Y. (1925), *Papers Relating to Political Economy* (London), Vol.1 198-259.

Edison, H. J. and Fisher, E. O. (1991), "A Long-Run View of the European Monetary System", *Journal of International Money and Finance*, 10, 53-70.

Efron, B. (1978), "Regression and ANOVA with Zero-One Data: Measures of Residual Variation", *Journal of the American Statistical Association*, 73, 113-212.

Eichorn, W. and Voeller, J. (1983), "Axiomatic Foundation of Price Indexes and Purchasing Power Parities", *Price Level Measurement,* Statistics Canada, Ottawa.

El-Gamal, M.A. and Ryu, D. (2006), "Short-Memory and the PPP Hypothesis", *Journal of Economic Dynamics and Control*, 30, 361-391.

Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996), "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 813-836.

Engel, C. and Morley, J. C. (2001), "The Adjustment of Prices and the Adjustment of the Exchange Rate", NBER Working Paper No.8550.

Feuerverger, A. (1989), "On the Empirical Saddlepoint Approximation", *Biometrika*, 76, 57-464.

Field, C. and Ronchetti, E. (1990), *Small Sample Asymptotics,* Institute of Mathematical Satistics.

Fisher, F. M. and Shell, K. M. (1972), *the Economic Theory of Price Indices*, Academic Press, New York.

Fisher, I. (1922), *The Making of Index Numbers*, Houghton-Muffin, Boston.

Fraser, D. A. S., Reid, N. and Wong, A. (1991), "Exponential Linear Models: A Two-Pass Procedure for Saddlepoint Approximation", *Journal of the Royal Statistical Society*, B, 53, 483-492.

Frisch, R. (1936), "Annual Survey of General economic Theory: The Problem of Index Numbers", *Econometrica*, 4, 1-38.

Gatto, R. (1996), "Saddlepoint Approximations of Marginal Densities and Confidence Intervals in the Logistic Regression Measurement Error Model", *Biometrics,* 52, 1096-1102.

Gatto, R. and Ronchetti, E. (1996), "General Saddlepoint Approximations of Marginal Densities and Tail Probabilities", *Journal of the American Statistical Association*, 91, 666-673.

Ghosh, J. K. and Sinha, B. K. (1981), "A Necessary and Sufficient Condition for Second Order Admissibility with Applications to Berkson's Bioassay Problem", *Annals of Statistics,* 9, 1334-1338.

Giles, D. E. A. (2001), "A Saddlepoint Approximation to the Distribution Function of the Anderson-Darling Test Statistic", *Communications in Statistics*, B, 30, 899-905.

Giles, D. E. A. and McCann, E. (1994), "Price Indices: Systems Estimation and Tests", *Journal of Quantitative Economics*, 10, 219-225.

Glen, J. H. (1992), "Real Exchange Rates in the Short, Medium, and Long Run", *Journal of International Economics,* 33, 147-166.

Godambe, V. P. (1955), "A Unified Theory of Sampling from Finite Populations", *Journal of the Royal Statistical Society, B,* 17, 269-278.

Goustis, C. and Casella, G. (1999), "Explaining the Saddlepoint Approximation", *American Statistician*, 53, 216-224.

Grilli, V. and Kaminsky, Z. G. (1991), "Nominal Exchange Rate Regimes and the Real Exchange Rate", *Journal of Monetary Economics*, 27, 191-212.

Grubb, D. and Symons, J. (1987), "Bias in Regressions with a Lagged-Dependent Variable", *Econometric Theory*, 3, 371-386.

Hensher, D. A., J. M. Rose and W. H. Greene (2005), *Applied Choice Analysis*, Cambridge University Press, Cambridge.

Hill, R. J. (1997), "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities", *Review of Income and Wealth*, 43, 49-69.

Hughes, G. A. and Savin, N. E. (1994), "Is the Minimum Chi-Square Estimator the Winner in Logit Regression? ", *Journal of Econometrics,* 61, 345-366.

Inder, B. (1986), "An Approximation to the Null Distribution of the Durbin-Watson Statistic in Models Containing Lagged Dependent Variables", *Econometric Theory*, 2, 413-428

Jensen, J. L. (1988), "Uniform Saddlepoint Approximations", *Advances in Applied Probability*, 20, 622-634.

Jing, B. Y. and Feuerverger, A. and Robinson, J. (1994), "On the Bootstrap Saddlepoint Approximations", *Biometrika*, 81, 211-215.

Jing, B. and Robinson, J. (1994), "Saddlepoint Approximations for Marginal and Conditional Probabilities of Transformed Variables", *Annals of Statistics*, 22, 1115-1132.

Johansen, S. (1988), "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control,* 12, 231-254.

Kadane, J. (1971), "Comparison of k-Class Estimators when the Disturbances are Small", *Econometrica,* 39, 723-37.

Kilian, L. and Zha, T. (1999), "Quantifying the Half-Life of Deviations from PPP: The Role of Economic Priors", *Federal Reserve Bank of Atlanta Working Paper 99-21.*

King, M. L. (1979), "Some Aspects of Statistical Inference in the Linear Regression Model", The PHD dissertation, the University of Canterbury Christchurch, New Zealand.

Kiviet, J. F. and Phillips, G. D. A. (1993), "Alternative Bias Approximations in Regression with a Lagged-Dependent Variable", *Econometric Theory*, 9, 62-80.

Kiviet, J. F. and Phillips, G. D. A. (1994), "Bias Assessment and Reduction in Linear Error-Correction Models", *Journal of Econometrics*, 63, 215-243.

Kiviet, J. F. and Phillips, G. D. A. (2005), "Moment Approximation for Least-Squares Estimators in Dynamic Regression Models with a Unit Root", *Econometrics Journal*, 8, 115-142.

Kiviet, J. F., Phillips, G. D. A. and Schipp, B. (1995), "The Bias of OLS, GLS and ZEF Estimators in Dynamic Seemingly Unrelated Regression Models", *Journal of Econometrics*, 69, 241-266.

Knight, J. L. and Satchell, S. E. (1992), "The Exact Distribution of the Maximum Likelihood Estimators for the Linear Regression Negative Exponential Model", mimeo., Department of Economics, University of Western Ontario.

Koop, J. C. (1986), "Estimating Variance of a Consumer Price Index and Some Comments on Inference", *Journal of Official Statistics*, 2, 74-76.

Kott, P. S.(1984), "A Superpopulation Theory Approach to the Design of Price Index Numbers with Small Sampling Biases", *Journal of Business and Economic Statistics*, 2, 83-90.

Lahey Computer Systems (1992), *Lahey F77L Fortran Programmer's Reference Manual*, Lahey Computer Systems, Incline Village, NV.

Laspeyres, E. (1871), "Die Berechnung Einer Mittleren Waaren-Preissteigerung", *Jahrbucher Fur Nationalokonomie Und Statistik*, 296-314.

Lewis, P. A. W. (1961), "Distribution of the Anderson-Darling Statistic", *Annals of Mathematical Statistics*, 32, 1118-1124.

Li, J. (2005), "Small Sample Properties of Discrete Choice Model Estimators Based on Symmetric and Asymmetric Cumulative Distribution Functions", M.A. Extended Essay, Department of Economics, University of Victoria.

Lieberman, O. (1994a), "Saddlepoint Approximation for the Distribution of a Ratio of Quadratic Forms in Normal Variables", *Journal of the American Statistical Association*, 89, 924-928.

Lieberman, O. (1994b), "Saddlepoint Approximation for the Least Squares Estimator in First-Order Autoregression", *Biometrika*, 81, 807-811.

Lieberman, O. (1997), "The Effect of Nonnormality," *Econometric Theory, 13, 52-78.*

Lloyd, P. J. (1975), "Substitution Effects and Biases in Nontrue Price Indices", *American Economic Review*, 65,301-313.

Lopez, C., Murray, C. J. and Papell, D. H., (2005) "More Powerful Unit Root Tests and the PPP Puzzle", University of Cincinnati Economics Working Paper No. 2003-07.

Lothian, J. R. (1997), "Multi-country Evidence on the Behavior of Purchasing Power Parity Under the Current Float", *Journal of International Money and Finance*, 16, 19-35.

Lothian, J. R. and Taylor, M. P. (1996), "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries", *Journal of Political Economy* 104, 488-509.

Lugannani, R. and Rice, S. O. (1980), "Saddlepoint Approximation for the Distribution of the Sum of Independent Random Variables", *Advances in Applied Probability*, 12, 475-90.

Mackinnon, J. G. and Smith, A. A. (1998), "Approximate Bias Correction in Econometrics", *Journal of Econometrics,* 85, 205-230.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

Magnus, J. R. (1986), "The Exact Moments of a Ratio of Quadratic Forms in Normal Variables," *Annals d'Economie et de Statistique*, 4, 95-109.

Marsaglia, G. and Marsaglia, J. C. W. (2004) "Evaluating the Anderson-Darling Distribution", *http://www.jstatsoft.org/v09/i02/ad.pdf.*

Meese, R. A. and Rogoff, K. S. (1988), "Was it Real? The Exchange-Rate Interest Differential Relation over the Modern Floating Rate Period", *Journal of Finance*, 43, 933-948.

Monti, A. C. and Rogoff, E. (1993), "On the Relationship between Empirical Likelihood and Empirical Saddlepoint Approximation for Multivariate M-Estimators", *Biometrika*, 80, 329-338.

Murray, C. J. and Papell, D. H. (2002), "The Purchasing Power Parity Persistence Paradigm", *Journal of International Economics*, 56, 1-19.

Murray, C. J. and Papell, D. H. (2005), "Do Panels Help Solve the Purchasing Power Parity Puzzle?", *Journal of Business & Economic Statistics*, 23, 410-415

Nagar, A. L. (1959), "The Bias and Moments Matrix of the General k-Class Estimators of the Parameters in Structural Equations", *Econometrica,* 27, 575-595.

Paasche, H. (1874), "Ueber die Preisentwicklung der Letzten Jahre nach den Hamburger Borsennotirungen", *Jahrbucher fur Nationalokonomie und Statistik*, 168-178.

Papell, D. H. (1997), "Searching for Stationarity: Purchasing Power Parity under the Current Float", *Journal of International Economics*, 43, 313-332.

Papell, D. and Theodoridis, H. (1998), "Increasing Evidence of Purchasing Power Parity over the Current Float", *Journal of International Money and Finance*, 17, 41-50.

Park, H. J., Fuller, W. A. (1995), "Alternative Estimators and Unit Root Tests for the Autoregressive Process", *Journal of Time Series Analysis*, 16, 415-429.

Pedroni, P. (2004), "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis", *Econometric Theory*, 20, 597-625.

Phillips, P. C. B. (1978) "Edgeworth and Saddlepoint Approximations in the First-Order Noncircular Autogression", *Biometrika*, 65, 91-98.

Prasada Rao, D. S. (1997), "Aggregation Methods for International Comparison of Purchasing Power Parities and Real Income: Analytical Issues and Some Recent Developments", *Proceedings of the International Statistical institute*, 51, 197-200.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd Ed, Cambridge University Press, Cambridge.

Quantitative Micro Software (2004), *EViews 5 User's Guide*, Quantitative Micro Software, Irvine CA.

Raj, B. and Ullah, A. (1981), *Econometrics: A Varying Coefficients Approaach*, Croom Helm, London.

Reid, N. (1988), "Saddlepoint Methods and Statistical Inference", (with discussion), *Statistical Science*, 3, 213-238.

Reid, N. (1991), "Approximations and Asymptotics", in *Statistical Theory and Models, Essays in Honor of D. R. Cox,* Chapman and Hall, 287-334.

Rilstone, P., Srivatsava, V. K. and Ullah, A. (1996), "The Second Order Bias and MSE of Nonlinear Estimators", *Journal of Econometrics,* 75, 369-395.

Rilstone, P. and Ullah, A. (2002), "Sampling Bias in Heckman's Sample Selection Estimator", in Y. P. Chaubey (ed.), *Recent Advances in Statistical Methods*, World Scientific, Hackensack NJ.

Robinson, J. (1982), "Saddlepoint Approximations for Permutation Tests and Confidence Intervals", *Journal of the Royal Statistical Society*, B, 44, 91-101.

Rogoff, K. (1996), "The Purchasing Power Parity Puzzle", *Journal of Economic Literature*, 34, 647-68.

Routledge, R. and Tsao, M. (1997), "On the Relationship Between Two Asymptotic Expansions for the Distribution of Sample Mean and its Applications", *the Annals of Statistics*, 25, 2200-2209.

Samuelson, P. A. and Swamy, S. (1974), "Invariant Economic Index Numbers and Canonic Duality: Survey and Synthesis", *American Economic Review*, 64, 566-593.

Selvanathan, E. A. (1988), "Stochastic Approach to Index Numbers: Extensions", Discussion Paper 88.15, Department of Economics, University of Western Australia.

Selvanathan, E. A. (1989), "A Note on the Stochastic Approach to Index Numbers", *Journal of Business and Economic Statistics*, 7, 471-474.

Selvanathan, E. A. (1991), "Standard Errors for Laspeyres and Paasche Index Numbers", *Economics Letters*, 35, 35-38.

...

Selvanathan, E. A. (1993), "More on Laspeyres Index Price", *Economics Letters*, 43, 157-162.

Selvanathan, E. A. and Prasada Rao, D. S. (1994), *Index Numbers: A Stochastic Approach*, London, Macmillan.

Sinclair, C. D. and Spurr, B. D. (1988), "Approximations to the Distribution Function of the Anderson-Darling Test Statistic", *Journal of the American Statistical Association*, 83, 1190-1191.

Skovgaard, I. (1987), "Saddlepoint Expansions for Conditional Distributions", *Journal of Applied Probability*, 24, 875-887.

Smith, M. D. (1989), "On the Expectation of a Ratio of Quadratic Forms in Normal Variables", *Journal of Multivariate Analysis*, 31, 244-257.

Spady, R. H. (1991), "Saddlepoint Approximation for Regression Models", *Biometrika*, 78, 879-889.

Srivastava, V. K. and Giles, D. E. A. (1987), *Seemingly Unrelated Regression Equations Models*, Marcel Dekker, New York.

Taylor, W. F. (1953), "Distance Functions and Regular Best Asymptotically Normal Estimates", *Annals of Mathematical Statistics*, 24, 85-92.

Taylor, M. P. and Sarno, L. (1998), "The Behavior of Real Exchange Rates during the Post Bretton Woods Period", *Journal of International Economics*, 46, 281-312.

Taylor, M. P., Peel, D. A. and Sarno, L. (2001), "Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles", *International Economic Review*, 42, 1015-1042.

Ullah, A. (2004), *Finite Sample Econometrics,* Oxford University Press, Oxford.

Ullah, A. and Srivastava, V. K. (1994), "Moments of the Ratio of Quadratic Forms in Non-Normal Variables with Econometric Examples", *Journal of Econometrics*, 62, 129-141.

Ullah, A. and Srivastava, V. K. and Roy, N. (1995), "Moments of the Function of Non-Normal Random Vector with Applications to Econometric Estimators and Test Statistics," *Econometric Reviews*, 4, 459-471.

Whistler, D., White, K. J., Wong, S. D. and Bates, D. (2001), *Shazam: The Econometrics Computer Program Version 9*, Northwest Econometrics.

Wold, H. and Jureen, L. (1953), *Demand Analysis: A Study in Econometrics*, Wiley, New York.

Wood, A. T. A.; Booth, J. G. and Butler, R. W. (1993) "Saddlepoint Approximations to the CDF of Some Statistics with Nonnormal Limit Distributions", *Journal of the American Statistical Association*, 88, 680-686.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge MA.

Wu, Y. (1996), "Are Real Exchange Rates Nonstationary? Evidence from Panel-Data Tests", *Journal of Money, Credit, and Banking*, 28, 54-63.

Zolotarev, V. M. (1961), "Concerning a Certain Probability Problem", *Theory of Probability and its Applications*, 6, 201-204.