ASYMPTOTIC EXPANSIONS FOR TEST STATISTICS AND TESTS FOR NORMALITY BASED ON ROBUST REGRESSION

A Ph. D. Dissertation

by
A ÖZLEM ÖNDER

Department of
Economics
Bilkent University
Ankara
June 1999

ASYMPTOTIC EXPANSIONS FOR TEST STATISTICS
AND
TESTS FOR NORMALITY BASED ON ROBUST REGRESSION

The Institute of Economics and Social Sciences
of
Bilkent University

by

A. ÖZLEM ÖNDER

In Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY IN ECONOMICS

in

THE DEPARTMENT OF
ECONOMICS
BİLKENT UNIVERSITY
ANKARA

June 1999

I certify that I have read the thesis and have found that is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Economics

Prof. Dr. Asad Zaman
Supervisor

I certify that I have read the thesis and have found that is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Economics

Asst. Prof. Dr. Kıvılcım Metin
Examining Committee Member

I certify that I have read the thesis and have found that is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Economics

Asst Prof. Dr. Metin Karadağ
Examining Committee Member

I certify that I have read the thesis and have found that is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Economics

Asst. Prof. Dr. Mehmet Caner
Examining Committee Member

I certify that I have read the thesis and have found that is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Economics

Asst. Prof. Dr. Aslıhan Salih
Examining Committee Member

Approval of the Institute of Economics and Social Sciences

Prof. Dr. Ali L. Karaosmanoğlu
Director

ABSTRACT

ASYMPTOTIC EXPANSIONS FOR TEST STATISTICS

AND

TESTS FOR NORMALITY BASED ON ROBUST REGRESSION

A. Özlem Önder

Ph. D. Department of Economics

Supervisor: Prof.Dr. Asad Zaman

June 1999

This dissertation focuses on two different topics in econometrics. The first one is presented in Chapter 2 and is related to higher order asymptotic theory. The power of the Lagrange multiplier, Wald and likelihood ratio tests for the first order autoregressive model is compared through the approximations to the distributions of these three tests. The adequacy of the approximation is examined. The Wald and likelihood ratio tests are found to have superior performance than the Lagrange multiplier test. The comparisons are done according to stringency of the test statistics.

As a second topic in Chapter 3, the dissertation examines the use of residuals from robust regression instead of OLS residuals in test statistics for the normality of the errors. According to simulation results their improvement over standard normality tests is found only in specialized circumstances. The applications on real data set show these conditions occur often enough in practice.

Key Words: Asymptotic Expansion, Autoregressive Model, Stringency, Normality Test, Robust Regression

ÖZET

TEST İSTATİSTİKLERİ İÇİN ASİMPTOTİK AÇILIMLAR VE GÜÇLÜ REGRESYONA DAYALI NORMAL DAĞILIM TESTLERİ

A. Özlem Önder

Doktora, Ekonomi Bölümü

Tez Yöneticisi: Prof.Dr. Asad Zaman

Haziran 1999

Bu çalışma ekonometrinin iki farklı konusunu incelemektedir. Bunlardan ilki 2. bölümde yer almaktadır ve yüksek dereceden asimptotik teoriyle ilgilidir. Lagrange çarpanı, Wald ve olabilirlik oranı testlerinin birinci dereceden otoregresif modelde güçleri, söz konusu testlerin dağılımlarına yaklaştırım yoluyla karşılaştırılmaktadır. Yaklaştırımların yeterlilik dereceleri incelenmiştir. Wald ve olabilirlik oranı testlerinin, Lagrange çarpanı testinden daha üstün performansa sahip olduğu gözlenmiştir. Testler arası karşılaştırma sıkılık kriterine göre gerçekleştirilmiştir.

İkinci konu olarak 3. bölumde tez, hata terimlerinin normal dağılımıyla ilgili test istatistiklerinde olağan en küçük kareler artıkları yerine, güçlü regresyon artıklarının kullanımının etkisini incelemektedir. Simulasyon sonuçlarına göre teknik, standart kullanılan normal dağılım testlerinden ancak belli koşullar altında üstün performans göstermektedir. Gerçek veri setiyle yapılan uygulamalar bu koşulların gerçekte yeterli sıklıkta görüldüğünü göstermektedir.

Anahtar Kelimeler: Asimptotik açılım, Otoregressif Model, Sıkılık, Normal Dağılım Testi, Güçlü Regresyon

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Asad Zaman for his guidance, help and support throughout the development of the thesis. He has not only supervised my thesis but also helped me overcome my problems which I had encountered while doing my research. I am grateful to Kıvılcım Metin for her help and encouragement I thank to Metin Karadağ for his careful reading and valuable comments.

I would like to thank Erdem Başçı, Sıdıka Başçı, Aziz Sağlam and Aykut Lenger for their help.

I also thank to my professors and research assistants from Ege University, for their understanding and support.

I would like to thank my family for their encouragements. I owe the greatest dept to my husband Abdullah who has provided me with constant support and patience. Without him it would be difficult to suceed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

This dissertation examines two different topics in econometrics. The first part of the dissertation is related to the higher order asymptotic theory, which is presented in Chapter 2. For most of the econometric problems the exact distributions of estimators and test statistics are not known. Then as a remedy we can rely on asymptotic theory. The limiting distribution of a statistic can be used to infer an approximate distribution for the statistic in a finite sample. The central limit theorem provides us such an approximation. But as it is presented in the literature small sample accuracy of this kind of approximations are often not accurate enough. This situation directed econometricians to benefit from the approximation techniques devised by mathematical statisticians. Edgeworth and sadllepoint expansion are some of these methods. It is called as higher order theory since according to these methods the first few terms of asymptotic expansions of distributions is used instead of the first order normal or chi-squared approximations.

Although there is an extensive literature about higher order asymptotic theory in mathematical statistics, the use of these techniques among econometricians have been slow. The techniques devised were dealing with expansions of distributions of sums of independent and identically distributed random vectors at the beginning. The expansions was in general for the univariate statistics. As Rothenberg (1984a) points out, the increased interest about the issue among econometricians forced the statisticians to generalization of expansions to more complex problems, like multidimonsional cases with dependency, nonlinearity etc. The approximation to these complicated problems requires difficult computations, but the advances in this area are still continuing with the new applications in econometrics.

In this direction, in this study we attempt to get better approximation by the application of Edgeworth expansions to some test statistics. We have used the classical trio of the test statistics, namely, likelihood ratio (LR), Wald and Lagrange multiplier (LM). However, firstly, we present the higher order asymptotic theory in Section 2.1. Considering the complexity of the topic, this section is attempted to be presented at a simple level. It especially covers the techniques used through this chapter in the dissertation. Therefore it provides the formulas related to the expansions of one dimensional statistics. Some preliminary notation and definitions are also presented in this section.

Section 2.2 presents the literature study. Although there have been a few surveys about econometric applications of asymptotic expansions to the best of our knowledge, increased interest about the subject has given rise to new researchs to emerge. The studies in this area are related to a wide variety of problems such as autoregressive and simultaneous equation models. In this section we only present the previous studies that are related to approximations to test statistics.

Section 2.3 provides the definitions of the test statistics and techniques to compare the performance of test statistics. The concept of stringency will be introduced. The theoretical results related to efficiency of test statistics is presented. Asymptotic equivalence of the test statistics will be presented as a result of the first order theory and the higher order efficiency of test statistics will be discussed.

After that point the dissertation presents an application to a simple model. Section 2.4 provides the statistics for testing the first order autocorrelation in a stationary process. If there is a uniformly most powerful test (UMP), than it is not possible to improve upon this test. So it is unnecessary to use other test statistics. Hovewer, UMP test does not exist in all hypothesis testing problems. Also, we can include test for autocorrelation in this category. In this context, different tests are proposed for these types of hypothesis testing problems. LM, Wald and LR tests are among them.

For the test statistics, firstly we calculate the empirical critical values and powers through Monte Carlo simulations and then try to get good approximations for reasonable sample sizes. The comparisons of test statistics have been conducted by using accurate approximation formulas instead of empirical ones. Our comparison method is stringency. Regarding to this the shortcomings of the tests are compared.

According to the first order theory the test performance of the LM, Wald and LR tests are equivalent. Engel (1984) shows that the asymptotic local power curve of the three tests are the same. If they are asymptotically equivalent the choice between the test statistics should be according to ease of computation. But the higher order theory suggest the performance of the three tests are differing and the results of the higher order theory are in favor of the LR test.

The aim of this dissertation, therefore, is to see the situation for the test of first order autocorrelation. The significance of this thesis is mainly to clarify the finite sample performances of the test statistics for the first order autocorrelation model. Also, this study presents the formulas for the critical values and the power curves of each tests, so that this results could be replicated without the need of Monte Carlo simulations. We hope that the formulas for the critical values could be used for applied studies to test for first order autocorrelation.

Chapter 3 develops a normality test based on robust regression and discusses its improvement on other normality tests. The consequences of violation of the normality assumption of regression residuals are known, and many test statistics devised for testing the normality of residuals. See for example Pierce and Gray (1982), White and McDonald (1980) and Pearson et.al. (1977), Urzua (1996) and Jarque and Bera (1987) for the descriptions some of these tests and discussions about their powers.

One difficulty related to normality tests is that the residuals are not directly observable. So the tests developed for these purposes depending on estimated residuals which are the ordinary least squares (OLS) ones. On the other hand, it is also known that OLS estimators can highly be influenced by the outliers. In other words, outliers and nonnormal errors may easily be masked in an OLS analysis.

For the identification of the outliers some robust techniques are developed. For instance, Least Trimmed Squares (LTS) estimator, which is introduced by Rousseeuw (1984), is one of them. Since the robust regression reveals the outliers, it may provide a clearer indication about lack of normality of residuals. The main idea in this chapter is to use residuals from a robust estimator (LTS) instead of OLS residuals as the basis for normality tests. Although it seems a clear idea, there does not seem to be such a study in the literature. In this context, we can say that this part of the thesis fills this gap in the literature and presents the effect of the use of robustified normality tests through simulations and applications on some real data set.

The normality tests developed by Jarque and Bera (1980) and Doornik and Hansen (1994) are very popular regarding regression applications. We have selected these two tests in our analysis as the standard normality tests. Section 3.1 presents these statistics, with their explicit formulas. As in most of the normality tests the statistics depend on sample skewness and kurtosis. Section 3.2 introduces the LTS estimator. Also, the general algorithm to find the estimator is presented. In Section 3.3, some motivating examples are presented. However, first the calculations of critical values are conducted through Monte Carlo experiments. Section 3.4 develops the simulation study. In this section the comparisons of the tests are presented. The situations in which the robust tests have improvement over the standard normality tests are examined. Section 3.5 presents some applications to data sets from economic literature.

Finally Chapter 4 is devoted to concluding remarks regarding to the success of higher order approximations in our application and the success robust tests of normality. Also, the recommendations for the directions of further research are given in this chapter.

CHAPTER II
ASYMPTOTIC EXPANSIONS FOR TEST STATISTICS

2.1 Higher Order Asymptotic Theory

2.1.1 Introduction

Exact finite sample distributions of estimators and test statistics are not available in most of the cases. Then a solution is to rely on asymptotic theory. Several approximation methods devised for this purposes. Techniques for approximating probability distributions have been studied by mathematical statisticians since the nineteenth century, and there is an extensive literature on this subject.

It is possible to obtain approximate distribution of an estimator or test statistic as the sample size becomes large. The central limit theorem provides us with approximation to distribution of estimator. Similar approximations are used for test statistics although the limiting distribution is often chi-squared rather than normal. It is also possible to get better approximations through higher order asymptotic expansions. Edgeworth expansion and sadllepoint expansion are the two well known methods for obtaining the higher order approximation to distribution functions. There are also different approximation methods (see Rothenberg (1984a) for the alternative methods). But as Magdalinos (1992) points out, these ad hoc methods may lead to more accurate but less interpretable results relative to asymptotic approximations, so may not be suitable for theoretical work.

We will give the emphasis to the results related to univariate distribution function, since the econometric application of the asymptotic expansions will be for the univariate case in the dissertation.

2.1.2 Preliminary Notation and Definitions

We will adhere closely to the notation presented here throughout the dissertation.

The cumulative distribution function (CDF) is denoted by F and the corresponding distribution function is f. The CDF of a standard normal random variable is represented by Φ and its distribution function by ϕ. Normal distribution with mean μ and variance σ will be denoted by $N(\mu, \sigma)$. The notation $\mathcal{X}_k^2$ will be used for chi-squared distribution with k degrees of freedom.

Let X be a random variable. $E(X)$ represents expected value of X, while $Var(X)$ the variance of X. With reference to a sequence $\{X_i\}$ of random vectors, the abbrevation *i.i.d.* will stand for "independent and identically distributed". The probabilty of an event will be indicated by $P(.)$.

Definition 1 *Given two sequences of real numbers* $\{a_n\}$ *and* $\{u_n\}$ *we say that* a_n **is of order** u_n*, denoted* $a_n = O(u_n)$*, if there exist a constant* $M > 0$ *such that* $|a_n/u_n| < M$ *for all n. Clearly* $a_n = O(1)$*, if the sequence* a_n *is bounded.*

Definition 2 *Given two sequences of real numbers* $\{a_n\}$ *and* $\{u_n\}$ *we say that* a_n **is of lower order than** u_n*, denoted* $a_n = o(u_n)$*, if*

$$lim_{n\to\infty}\left|\frac{a_n}{u_n}\right| = 0.$$

Clearly $a_n = o(u_n)$ *implies that* $a_n = O(u_n)$.

Definition 3 *(Convergence in Distribution)*

X_n **converges in distribution** *to a random variable* X *with distribution function F(X), if* $lim_{n\to\infty}|F_n(X) - F(X)| = 0$ *at all continuity points of F(X). This is written as* $X_n \xrightarrow{d} X$

Definition 4 *If* X *is a scalar random variable with distribution function F, its* **characteristic function** *is defined as* $\psi(t) = Eexp\{itX\}$, *where t is real, E represents the expectation with respect to the distribution of X, and* $i = \sqrt{-1}$

Definition 5 *The* **r-th moment** *of* X *is given by the r-th derivative of* $i^{-r}\psi(t)$ *evaluated at zero;*

$$E(X^r) = i^{-r}\psi^{(r)}(0).$$

Definition 6 *The function* $K(t,X) = K(t) = log\psi(t)$ *is called the* **cumulant generating function***(CGF). The r-th derivative of* $i^{-r}K(t)$ *evaluated at zero, is called the* **r-th cumulant** *of* X *and is denoted by;*

$$\kappa_r = i^{-r}K^{(r)}(0).$$

Remark: κ_1 is the mean and κ_2 is the variance.

2.1.3 Central Limit Theorem

The central limit theorem is the basic theorem of the asymptotic theory, through which it is possible to approximate the distribution of many statistic as normal.

Theorem 1 *(Lindeberg-Levy Central Limit Theorem) Let* $\{X_i\}$ *be i.i.d. with mean* μ *and finite variance* σ^2. *Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} N(0,\sigma^2).$$

Proof: Rao (1973:127)

It is possible to generalize this result to non-identical distributions and multivariate case (see Serfling (1980) for the central limit theorems in general cases).

Berry-Essen theorem gives an explanation to the accuracy of the central limit theorem. It is related to the difference between the exact distribution of the standardized statistic and the standard normal distribution.

Theorem 2 *(Berry-Essen)*

Suppose $X_1 ... X_n$ *are i.i.d. random variables with* $EX_i = 0, EX_i^2 = \sigma^2$*and* $E|X_i|^3 = \rho < \infty$. *Let* $F_n(t)$ *be the CDF of* $S_n = (\sum_{i=1}^{n} X_i)/\sqrt{n}\sigma)$.

$$\sup_t |F_n(t) - \Phi(t)| \leq \frac{0.7975\rho}{\sigma^3\sqrt{n}}.$$

Proof: see Bhattacharya and Rao(1976:110).

2.1.4 Edgeworth Expansion

The central limit theorem provides us approximation of test statistics. This is certainly a powerful tool but unfortunately in many cases these approximations are poor in quality and does not provide a good accuracy unless the sample size is very large. Many techniques have been devised to increase the accuracy of the approximation to the test statistics.

Edgeworth expansion is one of these methods. This is an expansion in powers of $n^{-1/2}$. The asymptotic approximation of the central limit theorem is the leading term in the Edgeworth expansion. So it can be seen an extension of large-sample techniques based on central limit theorem. The central limit theorem provides us an approximation of order $O(n^{-1/2})$. We will call it as a first order approximation. Besides we will develop Edgeworth approximation accurate of order $O(n^{-3/2})$ and call it as third order approximation in this section. More generally a k-th order approximation will be one of order $O(n^{-k/2})$ In order to make the approximation we will benefit from the Taylor expansion.

Theorem 3 *(Taylor Expansion)*

Let the function g have n-th derivative $g^{(n)}$ everywhere in the open interval (a,b) and (n-1)-th derivative $g^{(n-1)}$ continuous in the closed interval [a, b]. Let $x \in [a,b]$. For each point $y \in [a,b]$, $y \neq x$, there exists a point z interior to the interval joining x and y such that

$$g(y) = g(x) + \sum_{k=1}^{n-1} \frac{g^{(k)}(x)}{k!}(y-x)^k + \frac{g^{(n)}(z)}{n!}(y-x)^n.$$

Edgeworth expansions requires the use of the higher order cumulants. We will need the following preliminary results related to CGF's and cumulants.

Lemma 1 *(Properties of CGF) Let K(t,X) be cumulant generating function of X. It has the following properties;*

$$\begin{aligned} K(t, aX+b) &= itb + K(at, X), \\ K(t, \sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} K(t, X_i), \\ K(t, X) &= \sum_{i=1}^{\infty} \kappa_j(X)(it)^2/j!. \end{aligned}$$

Proof: The first two equations follow from the definiton of cumulant generating function, the last one can be shown by Taylor expansion.

Lemma 2 *(Properties of the cumulants) Let $\kappa_j(x)$ be the j th cumulant of X.*

$$\kappa_j(aX) = a^j \kappa_j(X),$$

$$\begin{aligned}\kappa_j(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} \kappa_j(X_i), \\ \kappa_j(X+b) &= \kappa_j(X) \quad if \quad j > 1, \\ \kappa_1(X+b) &= \kappa_1(x) + b.\end{aligned}$$

Proof: Follows from the properties of cumulant generating functions.

Remarks: Let μ be the mean and σ the standard deviation of X. From the properties of cumulants follows that, the cumulants of standardized variable are $\kappa_j((X-\mu)/\sigma) = \kappa_j(X)/\sigma^j$ for $j > 2$, while $\kappa_1 = 0$ and $\kappa_2 = 1$ for the standardized variable. The third cumulant of the standardized variable called as *skewness* and the fourth cumulant as *kurtosis.* For a standardized variable κ_3 is the third moment and the κ_4 is the fourth moment less three. For a normal random variable all cumulants greater than two are zero.

Let $X_1 ... X_n$ be i.i.d. random variables with common density function f and $E(X_i) = 0$ and $Var(x_i) = 1$ and X_i posseses derivatives up to the fourth order. Let ψ be characteristic function associated with f, then the CGF $log\psi$ can be expanded around zero as a power series through Taylor expansion:

$$log\psi(t) = \frac{1}{2}(it)^2 + \frac{1}{6}\kappa_3(it)^3 + \frac{1}{24}\kappa_4(it)^4 + ...,$$

where κ_r is the r-th cumulant of f.

The standardized sum $T_n = \sum X_i/\sqrt{n}$ also has zero mean and variance one; let f_n and ψ_n be its density and characteristic functions. Then,

$$log\psi_n(t) = nlog\psi(t/\sqrt{n}).$$

This follows from Lemma 1. By the Taylor expansion of the CGF,

$$nlog\psi(t/\sqrt{n}) \quad = \quad \frac{1}{2}(it)^2 + \frac{1}{6\sqrt{n}}\kappa_3(it)^3 + \frac{1}{24n}\kappa_4(it)^4 + ...$$

Taking the exponents,

$$\begin{aligned}\psi_n(t) &= exp(log\psi_n(t)), \\ &= e^{-t^2/2} e^{\frac{1}{6\sqrt{n}}\kappa_3(it)^3 + \frac{1}{24n}\kappa_4(it)^4 + \ldots}.\end{aligned}$$

Since e^x has the expansion $1 + x + \frac{1}{2}x^2 + \ldots$, if we expand the second term,

$$\psi_n(t) = e^{-t^2/2}\left(1 + \frac{1}{6\sqrt{n}}\kappa_3(it)^3 + \frac{3\kappa_4(it)^4 + \kappa_3^2(it)^6}{72n} + \ldots\right). \qquad (1)$$

Lemma 3 *If f has characteristic function $\psi(t)$, then the first derivative f' has characteristic function $\psi_{f'}(t) = -it\psi(t)$.*

Proof: Let $u(x) = e^{itx}$ and $dv(x) = f'(x)dx$, from integration by parts

$$\begin{aligned}\psi_{f'}(t) &= \int_{-\infty}^{\infty} u(x)dv(x) = u(x)v(x)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v(x)u'(x)dx, \\ &= -\int_{-\infty}^{\infty} ite^{itx}f(x)dx = -it\psi(t).\end{aligned}$$

From this result by induction follows that, if f has the characteristic function $\psi(t)$, then the r-th derivative f^r has characteristic function $(-it)^r\psi(t)$.

Lemma 4 *(Fourier inversion) Let $\psi(t)$ be characteristic function for T_n. The density function f_n for T_n can be written as:*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ixt}\psi(t)dt. \qquad (2)$$

Theorem 4 *(Edgeworth Density)*

Suppose X_i are i.i.d. with mean 0 and variance 1, and κ_3 and κ_4 the skewness and kurtosis, respectively. A third order approximation to density $T_n = \sum X_i/\sqrt{n}$ is

$$f(x) \approx \phi(x) + \frac{1}{6\sqrt{n}}(-1)^3\kappa_3\phi^{(3)}(x) + \frac{1}{24n}(-1)^4\kappa_4\phi^{(4)}(x) + \frac{1}{72n}(-1)^6\kappa_3^2\phi^{(6)}(x).$$

Proof: The result follows through application the Fourier inversion formula in Lemma(4) to characteristic function in Equaton (1) and dropping the higher order terms. By using the Lemma (3) and the fact that characteristic function of $\phi(x)$ is $e^{(-t^2/2)}$ the desired result is obtained.

Definition 7 *A Hermite polynomial of degree r is*

$$H_r(x) = (-1)^r \frac{\phi^{(r)}(x)}{\phi(x)}.$$

Using the Hermite polynomials, we can write the third order approximation to the density T_n as:

$$f(x) \approx \phi(x)\left(1 + \frac{\kappa_3}{6\sqrt{n}} H_3(x) + \frac{3\kappa_4 H_4(z) + \kappa_3^2 H_6(x)}{72n}\right). \tag{3}$$

It is easily checked that,
$H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3,$
$H_5(x) = x^5 - 10x^3 + 15x, \quad H_6(x) = x^6 - 15x^4 + 45x^2 - 15.$

Integration of the Edgeworth density in Equation (3) gives us the approximation to the cumulative distribution function:

$$F_n(x) \approx \Phi - \phi(x)\left(\frac{\kappa_3 H_2(x)}{6\sqrt{n}} + \frac{3\kappa_4 H_3(z) + \kappa_3^2 H_5(x)}{72n}\right). \tag{4}$$

The approximation in Equation (4) is called as *Edgeworth-A expansion*. Note that the Edgeworth expansion is not a probability density since it does not integrate to one and can take negative values. The following formula avoids this problem.

Theorem 5 *(Edgeworth-B Expansion) Another third order approximation to the cumulative distribution function is given by*

$$F_n(x) \approx \Phi\left(x - \frac{\kappa_3(x^2 - 1)}{6\sqrt{n}} + \frac{3\kappa_4(3x - x^3) + \kappa_3^2(8x^3 - 14x)}{72n}\right). \tag{5}$$

Proof: Let a, b, be unknown quantities. By Taylor expansion of $P = \Phi(x + a/\sqrt{n} + b/n)$,

$$\begin{aligned} P &\approx \Phi(x) + \left(\frac{a}{\sqrt{n}} + \frac{b}{n}\right)\phi(x) + \frac{1}{2}\left(\frac{a}{\sqrt{n}} + \frac{b}{n}\right)^2 \phi^{(1)}(x) \\ &= \Phi(x) + \frac{a}{\sqrt{n}}\phi(x) + \frac{b - xa^2/2}{n}\phi(x) + O(n^{-3/2}). \end{aligned}$$

Matching the terms of order $1/\sqrt{n}$ to the ones in Equation (4) we see that, $a = -\kappa_3 H_2(x)/6$. If we put the value of Hermite polynomial, we get the value of a. Matching the terms of order 1/n and solving for b, yields the formula of the theorem.

From Edgeworth-B expansion it is possible to have an inversion, called as the Cornish-Fisher expansion. The inversion is related to a representation of a random variable by asymptotic expansion in terms of a standard normal variable.

Corollary 1 *(Cornish-Fisher Inversion)*

Let c^*_α be the upper α quantile, i.e. the solution for fixed α of the equation $F(c^*_\alpha) = \alpha$, where $\Phi(c_\alpha) = \alpha$. Let $a(x), b(x)$ be as defined in Theorem 5, where $a'(x)$ be the first derivative of a. Then,

$$c^*_\alpha = c_\alpha - \frac{a(c_\alpha)}{\sqrt{n}} + \frac{a'(c_\alpha)a(c_\alpha) - b(c_\alpha)}{n}.$$

is a third order approximation. $P(T_n < c^*_\alpha) - \alpha$ goes to zero at rate $O(n^{-3/2})$.

Proof: Let $c^*_\alpha = c_\alpha + f/\sqrt{n} + g/n$. Then,

$$P(T_n < c^*_\alpha) = \Phi\left(c^*_\alpha + \frac{a(c^*_\alpha)}{\sqrt{n}} + \frac{b(c^*_\alpha)}{n}\right)$$

Substitute c^*_α. Note that $a(c^*_\alpha) \approx a(c_\alpha) + fa'(c_\alpha)/\sqrt{n}$ and $b(c^*_\alpha) \approx b(c_\alpha)$, where these approximations gives the required degree of accuracy. So,

$$P(T_n < c^*_\alpha) = \Phi\left(c_\alpha + \frac{f + a(c_\alpha)}{\sqrt{n}} + \frac{g + b(c_\alpha) + fa'(c_\alpha)}{n}\right).$$

Since $\Phi(c_\alpha) = \alpha$, we must set the higher terms to zero to get the desired accuracy. So, $f = -a(c_\alpha)$ and $g = a(c_\alpha)a'(c_\alpha) - b(c_\alpha)$. We can get the approximation by substituting the values of f and g.

Expansion under General Conditions

Until this point we have discussed asymptotic expansions of independent and identically distributed random variables. Following the procedure in Rothenberg (1984a) it is possible to develop higher order approximations for more general cases. Now we will make a summary of his approach.

Let T_n be a standardized statistic having stochastic expansion:

$$T_n = X_n + \frac{A_n}{\sqrt{n}} + \frac{B_n}{n} + \frac{R_n}{n\sqrt{n}}.$$

where X_n, A_n, B_n are sequences of random variables with limiting distributions as n tends to infinity. Suppose R_n is stocastically bounded and the limiting distribution of X_n is $N(0,1)$. Let $T' = X_n + A_n n^{-1/2} + B_n n^{-1}$. Suppose, T' has finite moments up to high order and its r-th cumulant is order of $n^{n(2-r)/2}$, where r is greater than 2. Furthermore, suppose the mean and variance can be written as:

$$\begin{aligned} E(T') &= \frac{a}{\sqrt{n}} + o(n^{-1}), \\ Var(T') &= 1 + \frac{b}{n} + o(n^{-1}). \end{aligned}$$

where a and b depend on the moments of X_n, A_n and B_n. The restandardized variable is,

$$T^* = \frac{T' - a/\sqrt{n}}{\sqrt{1 + b/n}}.$$

Its third and fourth moments are

$$\begin{aligned} E(T^*)^3 &= \frac{c}{\sqrt{n}} + o(n^{-1}), \\ E(T^*)^4 &= 3 + \frac{d}{n} + o(n^{-1}). \end{aligned}$$

here $c/\sqrt{n}$ and d/n is the approximate third and fourth cumulants of T' respectively. It is clear that

$$P(T' \leq x) = P\left(T^* \leq \frac{x - a/\sqrt{n}}{\sqrt{1 + b/n}}\right).$$

if we expand the term in denominator in Taylor series and dropping the terms of order $n^{3/2}$ we get

$$P(T' \leq x) = P\left(T^* \leq x - \frac{a}{\sqrt{n}} - \frac{bx}{2n} + o(n^{-1})\right).$$

Now it is possible to approximate $P(T \leq x)$ by using Edgeworth-B expansion in Equation (6) replacing c with κ_3 and d with κ_4.

$$P(T_n \leq x) \approx \Phi\left(x + \frac{\gamma_1 + \gamma_2 x^2}{6\sqrt{n}} + \frac{\gamma_3 x + \gamma_4 x^3}{72n}\right).$$

where

$$\begin{aligned} \gamma_1 &= c - 6a; \quad \gamma_3 = 9d - 14c^2 - 36b + 24ac; \\ \gamma_2 &= -c; \quad\quad \gamma_4 = 8c^2 - 3d. \end{aligned}$$

In a similar way we can get Edgeworth-A expansion.

It is possible to generalize the results for the one-dimensional Edgeworth expansion to the multivariate case. It is again through the expansion of characteristic function and Fourier inversion. Barndorff-Nielson and Cox (1979) and Skovgaard (1986), McCullagh (1987) are some of the studies about the multivariate Edgeworth expansion. For the use of Edgeworth expansion for different statistics under different condititons see Barndorff-Nielson and Cox (1989).

2.1.5 Saddlepoint Approximation

In general the Edgeworth expansions provides a good approximation in the center of the density but can be inaccurate in the tails. *Saddlepoint approximation*, which is called also as *tilted Edgeworth expansion*, gives more accurate results especially in tails of the distributions. It was developed by Daniels (1954). Daniels (1956) and Phillips (1978) have applied the method to autocorrelation statistics. Since we don't benefit from this tecnique in our analysis, we will not explain it in detail. Field and Ronchetti (1990) explain the saddlepoint technique including different applications. Jensen (1995) contains many results, applications and a survey of literature. The general idea can be summarized as follows. First the distribution is recentered at the point where the distribution is to be approximated. Then the Edgeworth expansion is used locally at that point and then the results are transformed back in terms of the original density. One difficulty about this approximation is it requires knowledge of the cumulant function and is more complex relative to Edgeworth expansion.

2.2 Literature Survey

Although there is an extensive literature about the higher order approximation theory by mathematical statisticians, the application to the econometrics not have a long history. Since 1970's there has been increasing interest about this issue. Today asymptotic expansions are used in different topics of econometrics. Phillips (1980) and Rothenberg (1984a) made a survey of econometric applications. But since that time many new applications take place in the literature. The emphasis in the present survey will be on the econometric applications of asymptotic expansions for test statistics. But first of all we will investigate the studies about higher order asymptotic expansion of test statistics.

There are some survey papers that evaluate the performance of testing procedures through the higher order asymptotic expansions under local alternative. Bickel (1974), Pfanzagl (1980), and Rothenberg (1982, 1984a) are some of these studies. The formal expansions are given for the univariate case in Chibisov (1974) and Pfanzagl (1973). The main results contained in Pfanzagl and Wefelmeyer (1978). For the multivariate case there are the studies of Peers (1971), Hayakawa (1975,1977), Harris and Peers (1980), Hayakawa and Puri (1985). Cordeiro et.al.(1994) develop formulae for test statistics in generalized linear models. Hayakawa (1977) and Hayakawa and Puri (1985) present also the asymptotic expansions of some test statistics for testing the hypothesis against fixed alternatives.

While much of the literature examined the case of independent and identically distributed random variables, some papers extended the calculations to certain time series settings. Taniguchi (1985) derives the asymptotic expansion of some test statistics for a gaussian autoregressive moving average process.

Cribari-Neto and Zarkos (1995) obtain Bartlett type corrections to Wald, LM, and LR test statistics for the multivariate regression model. They compare their results with the size corrected critical values through simulation. They found in general their approach more effective.

For the econometric applications of higher order approximations, it is necessary to know the validity of expansion. Chandra and Ghosh (1979,1980), Sargan (1980) presents the theory of the validity of Edgeworth expansions in the i.i.d case. Sargan(1976), Sargan and Satchell (1986) and Phillips (1977b) make general extension of the validity theory to time series. Magdalinos (1992) develops a method of proving the validity of Edgeworth type approximation in econometrics.

The econometric applications of higher order asymptotic expansions of test statistics are conducted according to the applications of general formulas devised by the studies above to econometric problems. Now we will list some of these studies.

Autoregressive Model:

P.C.B. Phillips has been one of the pioneers in applying asymptotic expansion to the first order autoregressive AR(1) model. Phillips (1977a) obtained the Edgeworth expansion of the least squares estimator and associated t ratio test statistics for the AR(1) model with zero mean for a stationary process. Then comparisons between the exact, first and second order approximations of the lest squares estimators are made for the finite sample sizes. The exact distributions are calculated by numerical integration.

Tanaka (1983) extends the results of Phillips (1977a) through getting higher order expansion for the estimator of AR(1) model with unknown constant mean. He also obtains t ratio tests based on these estimators. The exact distributions are calculated through the Monte Carlo simulations and some comparisons between exact and approximate distributions are made. According to the Phillips' and Tanaka's studies the higher order asymptotic approximations are not satisfactory for finite sample distributions (for a sample size of 20-30) of estimators and test statistics. For a less stable model less satisfactory results are obtained.

Rothenberg (1984a) examines autoregressive models with a stable root in his survey. In this study, the approximations related to the autoregressive models seems to be satisfactory for small values of the estimators. Edgeworth approximations are not so accurate for small samples and small values of estimator.

There are also other approaches to approximate sample statistics from autoregressive models. Daniels (1956), Durbin (1980), Phillips (1978) and Wang (1992) derived saddlepoint approximation for the least squares coefficient. Phillips's (1978) compares Edgeworth expansions with saddlepoint expansion. Both approximations are found unsatisfactory for small sample sizes and large least squares coefficient. Saddlepoint approximation is not defined in the tails for small sample sizes and Edgeworth approximation distorts substantially in the tails. Wang (1992) extends Lugannani & Rice's (1980) saddlepoint approximation method to the problem and shows that his approximation performs better relative to other approximation methods for the distribution including the extreme tail and sample size as small as 10.

The case of unit root attracted most of the researcher and there has been a lot of studies about the asymptotic distribution of the test statistics for the autregressive models having unit root as the null hypothesis. Dickey and Fuller (1979, 1981) and Evans Savin(1981,1984) are some of these researches. Phillips (1987a) generalized these results by generalizing the assumptions on error terms. The major results of these studies is that the asymptotic distribution have a discontinuity around one and therefore the finite sample performance of the statistics are poor. Therefore Phillips (1987b) develops an asymptotic theory for a first order AR model covering the possibility of a root near unity. Perron (1989), Nabeya and Tanaka (1990) and Perron (1991) tabulate the limiting distribution of the least squares estimator in a AR(1) model where the true model is near-integrated. Nabeya and Tanaka (1990) and Perron (1991) examine the limiting power of unit root tests. Abadir (1993) derives the closed forms for the distribution of conventional statistics to derive asymptotic power functions of some unit root tests.

Related to higher order approximation Phillips (1987a) developes asymptotic expansions for the unit root case. But the improvement is not examined. Larsson (1995) derives saddlepoint approximation of some test statistics in near-integrated AR processes. Some exact formulas for the distribution functions are also presented. Simulation and numerical calculations show that in most of the cases the approximations work reasonably well.

Simultaneous Equation models:

Rothenberg (1984a) presents some examples of higher order asymptotic distribution for simultaneous equation models. Related to the higher order power comparisons of test statistics he suggests that Wald, LR and LM test are asymptotically efficient after size correction. In more complex, full information models the power functions for the three tests cross. Edgeworth expansion to the distribution functions of some test statistics under the null hypothesis are given in Sargan (1975, 1980).

Morimune (1989) examines the properties of the t tests associated with ordinary least squares, two-stage least squares, limited information maximum likelihood estimators in a structural equation. He benefits from asymptotic expansions of t statistics to find out deviations of real sizes from nominal sizes theoretically. The asymptotic expansions confirm the results of Monte Carlo experiments. The limiting t distribution are not found appropriate as null distribution for some cases. Modified t test statistics were proposed as a byproduct from the asymptotic expansion.

Magdalinos (1994) is concerned with the relative performance of the several tests for the admissibility of overidentifiying instrumental variables. He calculates the size corrections to the order T^{-1}, where T is the sample size. The local power function of the size corrected tests is found to be same to the order T^{-1}.

There are different test statistics for testing the overidentifiability conditions on a structural equation in simultaneaus equation system. Kunitomo et.al. (1983) and Magdalinos (1988) derives the distribution of test statistics through asymptotic expansions. Magdalinos (1988) compare the performance of the tests using higher order local power of the tests.

Magdalinos and Symeonides (1996b) reinterpreted the tests of overidentification restrictions as the test of overidentifying orthogonality conditions for the simultaneous equation models. The third order local power of various tests under the alternative of false orthogonality conditions is derived and the formulations found to be the same.

Bootstrap:

It is known that in many econometric problems bootstrapping can give greater accuracy than asymptotic normal distribution. Through Edgeworth expansion it is possible to show the accuracy of the bootstrap estimates and tests (see Zaman (1996) for details related to the subject). Rayner (1990) derives estimates of p values and critical values through bootstraping the standardized estimator of the coefficients in the normal linear regression models where the error precision matrix depends on unknown parameters with error $o(T^{-1})$. It is also shown that through the bootstrapping of Rothenberg's (1984b) variance adjusted statistics it is possible to obtain test statistics with errors $o(T^{-2})$.

Hall and Horowitz (1996) get improved critical values for the test of overidentifying restrictions and t test based on generalized method of moments through bootstrapping the test statistics. By the use of Edgeworth expansion it is shown that bootstrap provides improvement over first order approximation of the asymptotic theory. The Monte Carlo experiments also support this result.

McManus et.al (1994) suggest like Hall and Horowitz (1996), to use the bootstrap critical values for a better approximation. They show that the approximation of the asymptotic theory is poor for finite sample sizes, when the partial adjustment model with autoregressive errors is nearly nonidentified.

Heteroskedasticity:

Maekawa (1988) apply general formulas of Hayakawa (1975) and Harris and Peers (1980) for the LR, LM and Wald tests to detecting heteroscedasticty in regression models. He gets asymptotic expansion of the non-null distribution of the three tests up to $O(T^{-1/2})$. Through power comparsion, none of statistic is found uniformly superior to the others. But if the moments of the explanatory variable are stable over the whole sample period, then the asymptotic power up to $O(T^{-1/2})$, is equivalent for the three tests.

The asymptotic theory provides poor approximation for the null hypothesis of homoskedasticity. In order to improve the approximation, Attfield (1991) made Barlett corrections to the LR test statistic for different types of heteroskedacticity in the linear model.

A similar work is conducted by Honda (1988). By using the general formula for the LM test developed by Harris (1985), he provides the formula for the size correction to the LM test for heteroskedasticity. Through Monte Carlo experiment, he indicates the improvement in the accuracy of the size of the test. The use of size corrected LM test increases the power of the standard LM test.

Further Applications:

For the statistics where error covariance matrix is nonscalar and depends on a set of unknown parameters, exact analysis is difficult and asymptotic approximations takes place. Rothenberg (1984c) has used Edgeworth expansions to evaluate different testing procedures for the regression coefficient of the normal linear model with unknown covariance matrix. General formulae for the multiparameter Wald, LR and LM tests are derived and the test statistics are compared. Rothenberg obtained adjusted critical values so that the three tests have the same size. The third order approximate local power functions indicate that when null hypothesis is one dimensional, all three tests are equally powerful. When the hypothesis is multidimensional none of the tests is uniformly powerful than the others.

Magee (1989) has applied Rothenberg's (1984c) size correction to F test of linear hypothesis in the linear regression model with AR(1) errors. The simulation results are in favor of these corrections.

Magdalinos and Symeonides (1996a) derives alternative ciritical values for Rothenberg's (1984c) testing problem, using Edgeworth expansion based on F and t distributions which are locally exact,i.e. they reduce to the exact critical values when the error covariance matrix known up to a multiplicative factor. They also suggest instead of size correction to use Cornish-Fisher corrections.

The simulation results are found in favor of the locally exact Cornish-Fisher corrections.

Similarly Rothenberg (1988) deals with the problem of testing regression coefficients in models with unknown error covariance matrix. He applies the approach in Rothenberg (1984c) for one dimensional hypotheses to get more interpretable results through relatively simpler higher order approximations. Approximate local power functions are derived for these tests. He applies the approximations to two examples, one involving heteroscedasticity and the other autocorrelation. He concludes that size and power correction terms can be large even in very simple models where the first order asymptotic theory might be expected to work well.

There are a lot of estimators and tests in the form of a ratio of quadratic forms in normal variables and their exact distributions are not known. Marsh (1998) derives saddlepoint approximations for the distributions of a ratio of noncentral quadratic forms in normal variables. He also presents an application to a simple case, F tests in the linear regression model to increase the accuracy of the size and power calculations. The calculations demonstrate that approximate size calculations are in fact exact, whereas those for powers are of high accuracy.

Although the saddlepoint approximation is an extremely accurate method for approximating probability distributions, as presented in Marsh (1998) it is difficult to apply. Lieberman (1994) proposes a theoretically justified approximation to the saddlepoint expansion to circumvent these difficulties. He makes an application to the Durbin-Watson test statistic. It is found that the approximation of saddlepoint expansion is satisfactory for sample sizes 40 or more.

The diagnostic checks that seek the evidence of misspecification is important for parametric models. There are conventional moment based test statistics for misspecification. Chesher and Smith (1997) examines potential to use likelihood ratio test to detect misspecification of parametric densities. The advantage of the use of LR test is stated as its suitable structure for Bartlett correction, unlike conventional moment tests. After obtaining Barlett correction of test statistics they present the approximate power of the test under the local alternative. Monte Carlo experiments suggest that LR test can perform well relative to conventional moment based tests.

The information matrix test was introduced by White (1982) to detect the misspecification of likelihood functions. It is suggested that information matrix test have finite sample distribution which is poorly approximated by its asymptotic chi-squared distribution. Chesher and Spady (1991) presents Monte Carlo experiments which confirm this situation and as a remedy proposes to use Cornish-Fisher expansion. For application these approximations information matrix test is defined as efficient score test. As the special case of information matrix test the results for the full information matrix, heteroscedasticty and nonnormality tests presented. For moderate sample sizes in the range of 100 to 250 the $O(T^{-1})$ approximation to the distribution function are found substantially better than chi-squared approximation.

Wald statistics which are based on different but algebrically equivalent restrictions have the same asymptotic distribution under the null. But the studies show that these statistics may be divergent in small samples. Phillips and Park (1988) tries to explain this phenomena by using a general Edgeworth expansion of Wald statistics. By various examples from the literature they show that the finite sample distribution of the Wald statistics for testing nonlinear restriction can depend substantially on the algebric form of the restriction.

Gurmu and Trivedi (1992) is an example related to the application of asymptotic expansions in the analysis of discrete count data. They derive adjustment factors of the overdispersion tests for truncated Poisson regression models. Through Monte Carlo investigation it is shown that asymptotic expansions improve the performance of these tests.

2.3 LR, Wald and LM tests

Empirical verification of hypothesis is very important in economics. The hypothesis testing is an important tool of this kind of analysis. In this section the most commonly used test procedures: the LR, Wald and LM tests will be discussed. But first, some definitions about hypothesis testing will be presented.

2.3.1 Definitions

Hypothesis testing has only two outcomes. A statement of the hypothesis is defined as null hypothesis (H_0). If the data fall in a particular region of sample space called the critical region then the test is said to reject the null hypothesis. Since there are only two possible outcomes, there are two ways such a procedure can be in error:

Definition 8 **Type I error** *occurs when the null hypothesis rejected when it is true.* **Type II error** *occurs when the null hypothesis incorrectly accepted.*

Definition 9 *For any test the probability of Type I error is called as the* **size** *of the test, which is denoted as α and also called as* **significance level**.

Definition 10 **The power** *of a test is the probability of rejecting null when it is false. So;*

$$power = 1 - \delta,$$

where δ denotes the Type II error.

Definition 11 *Let the data x are generated by a density function $f(x, \theta)$ and the null hypothesis be $H_0 : \theta \in \Theta_0$ versus the alternative $H_1 : \theta \in \Theta_1$. For any test T, suppose $R(T, \theta)$ be the probability of rejecting the null when θ is the true parameter. Let $\mathcal{T}_\alpha$ be the set of all tests of size α. For any $\theta_1 \in \Theta_1$, the maximum possible power any test of size α can attain is given by the* **power envelope** β_α^* *defined as follows:*

$$\beta_\alpha^*(\theta_1) = \sup_{T \in \mathcal{T}_\alpha} R(T, \theta_1).$$

The **shortcoming** S *of a test $T \in \mathcal{T}_\alpha$ at some alternative hypothesis $\theta_1 \in \Theta_1$ is the gap between the performance of T and the power envelope;*

$$S(T, \theta_1) \equiv \beta_\alpha^*(\theta_1) - R(T, \theta_1).$$

We can define also the shortcoming of a test over all $\theta_1 \in \Theta_1$ as;

$$S(T, \Theta_1) = \sup_{\theta_1 \in \Theta_1} S(T, \theta_1).$$

$S(T, \Theta_1)$ measures the maximum gap between the power envelope and the power of a given test, and will be referred to as the shortcoming of the test. A test having the smallest possible shortcoming in the set $\mathcal{T}_\alpha$ of all tests of level α, is called a **most stringent** *test of level α for testing H_0 against H_1*

2.3.2 Three Test Procedures

In this section a general formulation of the three tests will be presented.

Let the data x are generated by a joint density function $f(x, \theta_0)$ under the null hypothesis and by $f(x, \theta)$ with $\theta \in R^k$ under alternative. Let $\hat{\theta}$ be the unconstrained maximum likelihood (ML) estimator and let $\tilde{\theta}$ be the ML estimator subject to the constraint imposed by the null hypothesis. The log-likelihood is defined as $L(x, \theta) = log f(x, \theta)$ and the score as $S(x, \theta) = \partial L(x, \theta)/\partial \theta$. The information matrix is defined as,

$$\mathcal{I}(\theta) = -E\frac{\partial^2 L}{\partial\theta\partial\theta'}(\theta).$$

Let $\hat{\theta}$ be ML estimator. Then it is possible to show that,

$$Var(\hat{\theta}) = \mathcal{I}^{-1}(\theta).$$

The Wald statistic is,

$$W = (\hat{\theta} - \theta_0)'\mathcal{I}(\hat{\theta})(\hat{\theta} - \theta_0),$$

was introduced by Wald (1943).

The LR test was suggested by Neyman and Pearson (1928). It is based on the difference between maximum of the likelihood under null and alternative hypothesis. So, the LR statistic is

$$LR = \frac{f(x, \theta_0)}{f(x, \hat{\theta})}.$$

Clearly, LR takes values in interval [0,1]. Equivalently, the test may be carried out in terms of statistic

$$\mathcal{L} = -2log(LR).$$

which turns out to be more convenient for asymptotic considerations.

While the Wald test is based on unconstrained ML estimator, *LM test* is based on the constrained ML estimator. If the constraint is valid, the constrained ML estimator should be close to the point that maximizes loglikelihood. Therefore, the slope of the loglikelihood function should be close to zero at the constrained ML estimator. It is first suggested by Rao (1947) and since depend on the score functions called also Rao's score function. It can be shown that, the score has mean zero and variance $\mathcal{I}(\theta_0)$ under the null. Therefore

$$LM = S(x, \theta_0)'\mathcal{I}(\theta_0)S(x, \theta_0,).$$

reject the null for high values of the statistic.

Theorem 6 *Under general conditions, the statistics* W, $\mathcal{L}$, LM, *converge in distribution to* $\mathcal{X}^2$ *distribution with k degrees of freedom under the* H_0.

Proof: see Serfling (1980:155).

The LR test is difficult to compute since it uses the ML estimator from both constrained and unconstrained maximization of the loglikelihood. But for Wald test unconstrained ML estimator and for LM test the constrained ML estimator is needed. In complex model LR test may therefore very difficult to compute relative to the Wald and the LM tests. But the increased use of computers make LR test also applicable.

2.3.3 Approaches to Compare Test Statistics

There are a lot of procedure to compare the test statistics. Serfling (1980) in Chapter 10 discusses six different approaches. These are Pitman, Chernoff, Bahadur, Hodges and Lehmann, Hoeffding, Rubin and Sethuraman. There are also other approaches. Local methods compare tests in a shrinking neighborhood of the null hypothesis. Pitman efficiency is one of them and the most popular and widely used criterion. The nonlocal methods compare test performance at a fixed element in the null and the alternative. Bahadur efficiency one of the nonlocal methods, looks at the rate at which the power goes to one.

The concept of stringency is introduced by Wald (1942) to compare the performances of test statistics. The definition of the concept is presented in Section 2.3.1 The comparisons of the test statistics in Section 2.4 will be done according to stringency.

2.3.4 Asymptotic Equivalence of the Test Statistics and Higher Order Efficiency

Engel (1984) shows that the Wald, Likelihood Ratio and Lagrange multiplier tests statistics are distributed $\mathcal{X}^2$ asymptotically under the null hypothesis and have the same non-central $\mathcal{X}^2$ distribution under local alternative for testing multivariate hypothesis. Furthermore it is found that these tests are asymptotically locally most powerful invariant tests. So these statistics are called as asymptotically equivalent.

If the three tests are asymptotically equivalent the choice of them be according to ease of computation. But finite sample behavior of the tests especially in small samples are found different in some of the studies (see for example Bernt and Savin (1977) and Evans and Savin (1982)).

Since higher order expansions produce better approximation for small sample sizes, the validity of the results related to asymptotic equivalence of test statistics may be questioned through higher order approximation methods. Rothenberg (1984a) uses third order approximation to the local power function to compare the performances of LM, Wald and LR statistics. He benefits from Edgeworth approximation in his calculations. His results mainly depend on Pfanzagl and Wefelmeyer (1978). First he gets the adjusted critical values for the three tests by using Cornish Fisher expansion derived in Section 2.1.4. Then insert them to the third order power function. So it is possible to compare third order power function with the third order power envelope. As a result it is found that, each test are tangent to the power envelope at different points. The LM dominates all others when power is approximately α, the Wald dominates other when power is approximately $1-\alpha$, the LR test dominates at power approximately one-half, but none of the tests dominates each other. Amari et al. (1987) get similar result through the third order approximation of test statistics (see Figure 7 in Amari et.al (1987:5)).

The analysis done for the multiparameter case have the following result: For the normal linear model the power surfaces crosses and no one is uniformly superior. In nonlinear models LR test has optimal power characteristics in central region of the power curve (see Rothenberg (1984a) for details).

The nonlocal approaches also supports the LR test. Hoeffding (1965) and Brown (1975), Bahadur (1966), Kallenberg (1982) establishes the superiority of LR to other tests.

2.4 An Auotocorrelation Example

The purpose of this section is to compare the performance of LM, LR and the Wald test for testing autocorrelation by using higher order expansion of the test statistics.

Consider the stationary AR(1) model:

$$y_t = \rho y_{t-1} + \varepsilon_t \qquad \text{for } t = 1, 2, ..., T$$

where $y_0 \sim N(0,1)$ and $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ and $\sigma^2 = 1 - \rho^2$.

We test the null hypothesis $\rho = 0$ versus alternative $\rho = \rho_1$ for some fixed $\rho > 0$.

Let $y = (y_0, y_1, ..., y_T)'$.The likelihood function can be written as:

$$l(y) = l(y_0)l(y_1|y_0)...l(y_t|y_{(t-1)}, y_{(t-2)}...y_0)$$

So it is equal to

$$l(y) = \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi(1-\rho^2)}} exp - \frac{(y_t - \rho y_{t-1})^2}{2(1-\rho^2)} \right) \times \left(\frac{1}{\sqrt{2\pi}} exp - \frac{y_0^2}{2} \right)$$

The Neyman-Pearson (NP) statistic is;

$$NP(0, \rho_1) = \frac{l(y|\rho = \rho_1)}{l(y|\rho = 0)}$$

$$NP(0, \rho_1) = \frac{(1-\rho_1^2)^{(-T/2)} exp(-\frac{1}{2(1-\rho_1^2)} \sum_{t=1}^{T} (y_t - \rho_1 y_{t-1})^2)}{exp(-\frac{1}{2} \sum_{t=1}^{T} y_T^2)}$$

Dropping the constant term and taking logs the Neyman Pearson test takes the following form:

$$\begin{aligned} NP^*(0, \rho_1) &= -\frac{1}{2(1-\rho_1^2)} \sum_{t=1}^{T} (y_t^2 - 2\rho_1 y_t y_{t-1} + \rho_1^2 y_{t-1}^2) + \frac{1}{2} \sum_{t=1}^{T} y_t^2, \\ &= \frac{\rho_1}{(1-\rho_1^2)} \left(\sum_{t=1}^{T} y_t y_{t-1} - \rho_1 \sum_{t=0}^{T} y_t^2 + 0.5\rho_1 (y_0^2 + y_T^2) \right). \end{aligned}$$

Dropping the constant $\rho_1/(1-\rho_1^2))$, NP statistic becomes;

$$NP^{**}(0, \rho_1) = \sum_{t=1}^{T} y_t y_{t-1} - \rho_1 \left(\sum_{t=0}^{T} y_t^2 - 0.5(y_0^2 + y_T^2) \right).$$

In order to get higher order approximation of the NP statistics we need to standardize it. Using the following Lemma's it is possible to get the formula of the first and second cumulants of NP.

Lemma 5 *Let Λ be a real diagonal matrix, and $u \sim N(0, I)$. The cumulants $\kappa_h(h = 1, 2, ...)$ of the distribution $u'\Lambda u$ are*

$$\kappa_h(u'\Lambda u) = 2^{(h-1)}(h-1)! tr\Lambda^h.$$

Proof: see Magnus(1978:203).

Lemma 6 *Let A be a symmetric matrix, and $\varepsilon \sim N(0, V)$, where V is positive definite. The cumulants $\kappa_h(h = 1, 2, ...)$ of the distribution $\varepsilon' V \varepsilon$ are*

$$\kappa_h(\varepsilon' V \varepsilon) = 2^{(h-1)}(h-1)! tr(AV)^h.$$

Proof: Since V is positive definite, there exist a unique, positive definite and symmetric matrix $V^{1/2}$ such that $V^{1/2}V^{1/2} = V$. Let T be an orthogonal matrix such that

$$T'V^{1/2}AV^{1/2}T = \Lambda,$$

where Λ is a diagonal matrix containing the eigenvalues of $V^{1/2}AV^{1/2}$ on its diagonal. Then

$$\varepsilon' A \varepsilon = (\varepsilon' V^{-1/2}T)(T'V^{-1/2}AV^{-1/2})(T'V^{-1/2}\varepsilon) = u'\Lambda u.$$

Hence

$$\kappa_h(u'\Lambda u) = \kappa_h(\varepsilon' V \varepsilon),$$

and

$$tr\Lambda^h = tr(T'V^{(1/2)}AV^{(1/2)}T)^h = tr(AV)^h.$$

The last equality follows from properties of trace.

NP statistic can be written as follows:

$$NP^{**}(0, \rho_1) = y'My \quad \text{for} \quad y \sim N(0, \Sigma),$$

$\Sigma_{i,j} = \rho^{|i-j|}$ where $\Sigma_{i,j}$ denotes the i,j'th element of the covariance matrix.

$$M = -\rho_1 I + 0.5\rho_1(E_{11} + E_{T+1,T+1}) + 0.5J,$$

where I is an $(T+1)\times(T+1)$ identity matrix; $E_{i,i}$ is an $(T+1)\times(T+1)$ matrix which consist ones at $E[i,i]$ and zeros everywhere. J is a $(T+1)\times(T+1)$ symmetric matrix, which has ones on the first sub and super diagonals and zeros at the rest.

According to Lemma 6, $\kappa_j(y'My) = 2^{j-1}(j-1)!tr((\Sigma M)^j)$ The first cumulant and second cumulant under the null hypothesis can be expressed in simple formulas.

$$\begin{aligned}
\kappa_1(\rho,\rho_1) &= T(\rho-\rho_1), & (6)\\
\kappa_1(0,\rho_1) &= -T\rho_1, & (7)\\
\kappa_2(0,\rho_1) &= 2tr((\Sigma M)^2) = T + 2(T-0.5)\rho_1^2. & (8)
\end{aligned}$$

But for the higher order cumulants the formulas becomes too complex.

Then the NP statistic is standardized so that;

$$NP'(0,\rho_1) = \frac{NP^{**}(0,\rho_1) + T\rho_1}{\sqrt{T + 2(T-0.5)\rho_1{}^2}}. \qquad (9)$$

2.4.1 Critical Values

We considered 5% significance level tests. In order to obtain sample value of NP' statistic, y_t for $t = 0, 1, ..., T$ are generated under the null hypothesis;

where $y_t \overset{iid}{\sim} N(0,\sigma^2)$ under the null. The time series length T for NP' statistics begins from 1 until 100.

We get NP' statistics for 20 different ρ_1; where $\rho_1 = 0.05, 0.10, ..., 1$ from 1 to 100 number of observation. So $NP'(0,\rho_1)$ is a (100×20) matrix.

Our first choice for the critical values was with reference to its asymptotic distribution to give a nominal size of 0.05. According to central limit theorem NP' is distributed standard normal. For each ρ_1 with a 10.000 Monte Carlo sample size we counted the number of times $NP'(0,\rho_1) \geq 1.645$ where 1.645 is 5% critical value for standard normal density. So we get empirically calculated probability of $P(NP'(0,\rho_1) \geq 1.645)$.

The asymptotic distribution of the statistic should converge to standard normal distribution. But as it can be seen form Figure 1 the convergency is accuring slowly. The empirical size reached to 0.045 for $\rho < 0.5$ and to 0.04 for $\rho > 0.5$, from 30 to 100 observation. In order to improve the accuracy we applied second order approximation to critical values (cv), so obtained adjusted critical values. Since by Edgeworth expansion

$$P(NP' < cv) = \Phi(cv - \frac{\kappa_3'(cv^2-1)}{6}),$$

it is possible to say that

$$cv - \kappa_3'(cv^2-1)/6 \approx 1.645.$$

Solving the equation

$$cv = \frac{6 - \sqrt{36 - 4\kappa_3'(6 \times 1.645 - \kappa_3')}}{2\kappa_3'} \qquad (10)$$

we get the critical values, where cv is a (100×20) matrix, κ_3 is the third cumulant of $NP'(0,\rho_1)$ can be written as;

$$\kappa_3' = \frac{\kappa_3(0,\rho_1)}{\kappa_2(0,\rho_1)^{3/2}},$$

follows from properties of cumulants stated in Lemma 2.

Our main concern was to express cv as a function, which depends on ρ_1 and t. So it can easily be replicated without using simulations. For a fixed t, depending on ρ_1, cv has some quadratic form.

Looking OLS results

$$\hat{cv}(t,\rho_1) = \alpha(t) + \beta_1(t)\rho_1 + \beta_2(t){\rho_1}^2 + \beta_3(t){\rho_1}^3 \quad t = 1, 2, ..., 100$$

is found to be a good fit of the cv for each t. Various other functional forms were also checked, but they didn't performed better than the estimation given above evidenced by high R^2 and t values of the coefficients. So we get α and β_i's as the regression result, where each coefficient is a (100×1) matrix.

But we also have to add time period t as a regressor to the estimation. According to asymptotic theory;

$$\lim_{t\to\infty} \beta_i(t) = 0 \ \text{ for } \ i = 0,1,2,3$$

where $\beta_0 = \alpha - 1.645$.

We estimated

$$\beta_i(t) = \frac{b}{(t-c)^a} + \epsilon_t \ \ i = 0,1,2,3 \ \ t = 1, 2, ..., 100$$

through estimating the coefficients a, c which are the best fits for β_i by regression. Finally the function for 5% critical values takes the following form,

$$\begin{aligned} cv'(t,\rho_1) &= 1.645 + 0.106(t-10)^{-5/4} - 3.885(t+15)^{-0.5}\rho_1 \\ &\quad +4.749(t+10)^{-0.5}{\rho_1}^2 - 1.867(t+5)^{-0.5}{\rho_1}^3 \end{aligned}$$

The critical values calculated above was very close to critical values found in Equation (10). The maximum absolute error from 20 to 100 observation is 0.004 and the error drops to 0.002 from 30 to 100 observation. The theoretically calculated critical values is presented in Figure 3.

2.4.2 The Power

After the calculations of critical values, it is possible to find the power empirically. By using Monte Carlo simulation, we have found $P(NP' > cv(t,\rho_1)|\rho)$, where $\rho = 0, 0.05, ..., 1$. Here the Monte Carlo sample size is 10.000. So we get a (100×420) matrix. The first 20 colums of the matrix consist of the empirical power of NP' statistics for $\rho = 0$ and $\rho_1 = 0.05, 0.10, ..., 1$ respectively. This is the empirical size of second order approximation (see Figure 2). The maximum difference from 0.05 is ± 0.005 beginnig from 50 observation and the empirical size is converging to 0.05 as the time series length increases.

It is also possible to get power curve through theoretical formula. We benefit in this case from Edgeworth approximation. In order to apply the Edgeworth-B expansion we have to standardize NP' in Equation (9).

$$NP'' = \frac{NP' - E(NP'|\rho)}{\sqrt{Var(NP'|\rho)}},$$

where

$$\begin{aligned} E(NP'|\rho) &= \frac{E(NP|\rho)) - \kappa_1(0,\rho_1)}{\sqrt{\kappa_2(0,\rho_1)}}, \\ &= \frac{T\rho}{\sqrt{T + 2(T-0.5)\rho_1^2}}, \end{aligned}$$

$$
\begin{aligned}
Var(NP'|\rho) &= \frac{\kappa_2(\rho, \rho 1)}{\kappa_2(0, \rho_1)}, \\
&= \frac{\kappa_2(\rho, \rho 1)}{\sqrt{T + 2(T - 0.5)\rho_1^2}}.
\end{aligned}
$$

The results follows from Equations (6)-(7).

It is clear that

$$P(NP' > cv'(t, \rho_1)|\rho) = P(NP'' > cv''(t, \rho_1)|\rho)$$

where

$$cv'' = \frac{cv' - E(NP'|\rho)}{\sqrt{Var(NP'|\rho)}}.$$

Now it is possible to apply second order Edgeworth expansion for the power as:

$$P(NP'' > cv''(t, \rho_1)|\rho) = 1 - F(cv'') = 1 - \Phi(cv'' - \frac{\kappa_3''((cv'')^2 - 1)}{6}), \quad (11)$$

where Φ is the cumulative distribution function of a standard normal random variable and κ_3'' is the third cumulant of NP'' We have also applied the third order approximation to the power as

$$1 - F(cv'') = 1 - \Phi(cv'' - \frac{\kappa_3''((cv'')^2 - 1)}{6} + \frac{\kappa_4''(3cv'' - (cv'')^3)}{24} + \frac{\kappa_3''(8(cv'')^3 - 14cv'')}{72}.$$

After the calculation of empirical and theoretical power, we checked the difference. The second order expansion found to be a better approximation to empirical power. The maximum absolute error is around 0.02 from 30 to 100 observation for $\rho < 0.5$ and is decreasing as t increases. From 60 to 100 observation the maximum absolute error is around 0.01 for $\rho < 0.5$.

Now it remains to find a formula for theoretical power. Define

$$a(\rho, \rho 1) = cv'' - \frac{\kappa_3''((cv'')^2 - 1)}{6}.$$

Some of the terms in matrix a was too small. In order to simplify the problem of finding a functional form, the matrix elements smaller than -2.5 replaced as -2.5. This is possible since if $x < -2.5$ then $1 - \Phi(x) \approx 1$.

2.4.3 The Power Envelope

It is possible to get power envelope from the power of the NP test. Let $amaxp$ consist of 1,21,42,63,...420'th colums of matrix a. $1 - \Phi(amaxp)$ will give the maximum power. Following the steps to get the approximate formula for critical values in subsection 2.4.1, we approximated $amaxp$ and get a formula for power envelope.

So we have found the approximate formula for the power envelope ;

$$amaxp'(t, \rho_1) \;\; = \;\; 1.645 + 270t^{-3} - 0.6(t + 10)^{0.6}\rho_1,$$

which is a good approximation to $amaxp(t, \rho_1)$. The difference between $1 - \Phi(amaxp)$ and $1 - \Phi(amaxp')$ checked. The maximum absolute error is found 0.019 from 40 to 100 observation, 0.016 from 50 to 100 observation, 0.014 from 60 to 100 observation (see Figure 4 and Figure 5).

2.4.4 Lagrange Multiplier Test

Lagrange multiplier or locally most powerful test is

$$LM = \lim_{\rho_1 \to 0} NP(0, \rho_1)$$

In order to obtain a formula for LM statistics we get the *alm* matrix, which consist of 1,21,41,...401 colums of matrix *a*.

Following the same procedure to get the formula of critical values in section 2.4.1, we have the formula for $almp(t, \rho_1)$ as ;

$$\begin{aligned} alm'(t,\rho_1) &= 1.645 - 531.5t^{-3} - 0.62936(t+7)^{0.6}\rho_1 + 0.00649(t+46)^{1.2}\rho_1^2 \\ &\quad +0.685(t-14)^{0.45}\rho_1^4 \end{aligned}$$

The difference between $1-\Phi(alm)$ and $1-\Phi(alm')$ checked for $\rho < 0.8$. Maximum absolute error is found 0.014 from 30 to 100 observation (see Figure 6 and 7).

2.4.5 The Wald Test

For the higher order approximation of Wald test we used the standardized statistic developed by Rothenberg (1984a), which can be written as:

$$W = \frac{\sqrt{T}(\hat{\rho}-\rho)}{\sqrt{1-\rho^2}}$$

where

$$\hat{\rho} = \frac{y'y_{-1}}{y'_{-1}y_{-1} + (y_T^2 - y_0^2)/2} \tag{12}$$

here $y = (y_1, ..., y_T)'$ and $y_{-1} = (y_0, ...y_{T-1})$. $\hat{\rho}$ is a modified ML estimator, which has the property that always taking values in the interval (-1,1).

Following his procedure let $\alpha = \rho/\sqrt{1-\rho^2}$

$$W = \frac{\sqrt{T}\hat{\rho}}{\sqrt{1-\rho^2}} - \sqrt{T}\alpha.$$

Since $y = \rho y_{-1} + \varepsilon$, $\sigma^2 = 1 - \rho^2$ and using Equation (12) we get,

$$W = \Big(\frac{y'_{-1}\varepsilon}{\sigma^2\sqrt{T}} - \frac{\alpha(y_T^2 - y_0^2)}{2\sigma^2\sqrt{T}}\Big)\Big(\frac{y'_{-1}y_{-1}}{\sigma^2 T} + \frac{y_T^2 - y_0^2}{2\sigma^2 T}\Big)^{-1}.$$

So

$$W = \left(X - \frac{\alpha Z}{\sqrt{T}}\right)\left(1 + \frac{K}{\sqrt{T}}\right)^{-1}, \tag{13}$$

where

$$X = \frac{y'_{-1}\varepsilon}{\sigma^2\sqrt{T}}; \quad K = \frac{y'_{-1}y_{-1} + (y_T^2 - y_0^2)/2 - T\sigma^2}{\sigma^2\sqrt{T}}; \quad Z = \frac{y_T^2 - y_0^2}{2\sigma^2}.$$

By expanding the second term in paranthesis at equation (13) in Taylor series and dropping terms of order $T^{3/2}$

$$\begin{aligned} W' &= (X - \frac{\alpha Z}{\sqrt{T}})\left(1 - \frac{K}{\sqrt{T}} + \frac{K^2}{T}\right), \\ &= X - \left(\frac{XK + \alpha Z}{\sqrt{T}}\right) + \frac{XK^2 + \alpha KZ}{T}. \end{aligned}$$

The approximate first four cumulants are;

$$E(W') = \frac{-2\alpha}{\sqrt{T}}; \quad Var(W') = 1 + \frac{7\alpha^2 - 2}{T};$$

$$\kappa_3 = \frac{-6\alpha}{\sqrt{T}}; \quad \kappa_4 = \frac{6(10\alpha^2 - 1)}{T}.$$

The Edgeworth-B approximation to the distribution function is

$$P(W \leq w) = \Phi\left[w + \frac{\alpha(w^2+1)}{\sqrt{T}} + \frac{w(1+4\alpha^2) + w^3(1+6\alpha^2)}{4T}\right]. \tag{14}$$

It is possible to find the %5 critical values from Equation (14) putting $\rho = 0$ So we get the critival values from the following equation

$$F(cv) = \Phi(cv + \frac{(cv + cv^3)}{4T}) = 0.95.$$

where cv is a (100×1) matrix. By using the Cornish Fisher expansion, it is possible to obtain the critical values

$$cv' = c_\alpha - \frac{(c_\alpha + c_\alpha^3)}{4T}$$

where $c_\alpha = 1.645$. The maximum absolute difference of cv' from cv is 0.0034 for 30 to 100 observations. We calculate also $P(\sqrt{(T)}\hat{\rho} > cv'|H_0)$ through Monte Carlo. It is approximately 0.05, which has an error less than ±0.005 from 7 to 100 observations. Following the same procedure of the former sections, we can get the empirical power $P(\sqrt{(T)}\hat{\rho} > cv'|H_1)$.

The theoretical power can be calculated by the Equation (14) as

$$1 - P(W \leq cv''|\rho_1) = 1 - \Phi\left[cv'' + \frac{\alpha((cv'')^2+1)}{\sqrt{T}} + \frac{cv''(1+4\alpha^2) + (cv'')^3(1+6\alpha^2)}{4T}\right]$$

here

$$cv'' = \frac{(cv' - \sqrt{T}\rho)}{\sqrt{1-\rho^2}}$$

follows from the fact that

$$P(\sqrt{T}\hat{\rho} > cv'|\rho_1) = P\left(\frac{\sqrt{T}(\hat{\rho}-\rho)}{\sqrt{1-\rho^2}} > \frac{cv' - \sqrt{T}\rho)}{\sqrt{1-\rho^2}}\Big|\rho_1\right)$$

Maximum absolute error of theoretical power is 0.0156 for 30 to 100 observation For $\rho_1 < 0.5$ it is less than 0.01. The maximum absolute error is 0.01 for 40 to 100 observation (see Figure 8 and Figure 9).

2.4.6 The Likelihood Ratio Test

LR statistic is defined as;

$$\lambda = \frac{l(y|\rho = 0)}{l(y|\rho = \hat{\rho})}$$

$$\lambda = \frac{exp(-\frac{1}{2}\sum_{t=1}^{T} y_T^2)}{(1-\hat{\rho}^2)^{(-T/2)} exp(-\frac{1}{2(1-\hat{\rho}^2)}\sum_{t=1}^{T}(y_t - \hat{\rho}y_{t-1})^2)}$$

where $\hat{\rho}$ denotes the ML estimate of ρ, calculated according to the formula in Section 2.4.5 as,

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2 + (y_T^2 - y_0^2)/2}.$$

It follows that,

$$
\begin{aligned}
-2log\lambda &= -Tlog(1-\hat{\rho}^2) + \sum_{t=1}^{T} y_t^2 - \frac{1}{1-\hat{\rho}^2}\sum_{t=1}^{T}(y_t^2 - 2\hat{\rho}y_t y_{t-1} + \hat{\rho}^2 y_{t-1}^2) \\
&= -Tlog(1-\hat{\rho}^2) + \frac{2\hat{\rho}}{1-\hat{\rho}^2}\left(\sum_{t=1}^{T} y_t y_{t-1} - \hat{\rho}\left(\sum_{t=0}^{T} y_t^2 - 0.5(y_0^2 + y_T^2)\right)\right).
\end{aligned}
$$

Our hypothesis testing problem is one sided with inequality of the constraints on the regression coefficient. The asymptotic approximations stated in Theorem 6 is not applicable to this problem. But the asymptotic distribution of the LR statistic in the presence of boundary constraint discussed in some researchs. Self and Liang (1987), Gourieroux et.al. (1982) are some of these studies. Instead of using asymptotic approximations, we have benefit curve fitting technique to get approximation to the empirically calculated distributions in this section.

For the calculation of critical values as in the former sections we conducted Monte Carlo simulation as the first step. The simulation designed as follows. We set $\rho = 0$. Generated 1000 samples of y_t under null hypothesis. We calculated the LR statistics. $B = \sum_{t=1}^{T} y_t y_{t-1} < 0$ is replaced by 0. Since if $B < 0$, ML estimator becomes negative. So for $B < 0$ we have LR statistics equalt to 0. We sort the LR statistics from low to high.

The 880th,890th,...950th..,990th observations are kept as initial estimates of critical value (cv). Then we calculated $P(LR > cv|H_0)$ for each cv. Finally through interpolation by using these set of cv's and related probabilities, we tried to find cv' which give the 5% critical value of LR test. These procedure repeated from 1 to 100 time series length. So cv' is an (100×1) matrix.

As the next step we approximate a formula for the empirically calculated critical values. Looking OLS results,

$$cv'' = 2.64 + 0.00064T$$

found to be a good approximation to cv'. We checked $P(LR > cv''|H_0)$. It is around 0.05 with an error $\pm$ 0.005 from 10 to 100 observation. To obtain the power we calculate $P(LR > cv''|H_1)$ again through Monte Carlo.

For the derivation the approximate power curve, first we obtained alr, a (100×20) matrix where

$$P(LR > cv''|H_1) = 1 - \Phi(alr). \tag{15}$$

This is done through computer programming. With the same techniques in Section 2.4.1 we tried to find a formula for alr. Looking the OLS results

$$alr' = 1.645 - 0.6(T + 17)^{0.6}\rho$$

found to be a good fit.

Inserting alr' instead of alr in Equation (15), we get the approximate power curve of LR test. Maximum absolute error of approximation is 0.023 form 40 to 100 observation and 0.016 from 50 to 100 observation (see Figure 10 and Figure 11). So we derived the approximation to LR statistics without the use of higher order formulas.

2.4.7 Comparisons

After obtaining the theoretical approximations, we compared the performances of LM, Wald and LR test, according to stringency defined in Section 2.3.1 Figure 12 and Figure 13 present the power curves of the LM test for 50 and 100 observations The figures are drawn according to the theoretical formulas. It is observed that the power envelope of the LM test shift upwards and the shortcoming of the test decreases as the number of observation increases.

One property related to power of LM test is, it dips down near $\rho = 1$. This makes the test weak in terms of stringency.

Figure 14 and 15 present the power curves for the Wald test. It is almost same with the maximum power curve. For the LR test looking to Figure 16 and 17 we observe a similar result.

Figure 18 provides the shortcomings of the LM, Wald and LR tests according to their emprically calculated power curves. Figure 19 compares the performance of the three tests this time by using their theoretical approximate power curves. The conclusions obtained from the two approaches are the same. The shortcoming of the LM test is around 30% for 30 time series length and decreasing as the number of observation increases. It becomes around 5% when the number of observation reaches to 100. The Wald and LR test has equal performance according to stringency. The shortcomings are less than 1% for both of the tests. There is a strict conclusion that the LR and Wald tests outperform the LM test by testing the first order autocorrelation coefficient.

CHAPTER III
TESTS FOR NORMALITY BASED ON ROBUST REGRESSION

In this chapter we will examine the effects of using residuals from robust regression instead of OLS residuals in test statistics for the normality of errors. The first part of this chapter will introduce the Jarque Bera and Doornik Hansen tests, which are two well known tests for normality. Robust estimator LTS, used for the normality tests is described in Section 3.2. In Section 3.3 the test statistics based on LTS estimator is presented and some examples from Rousseeuw and Leroy data sets will be reported. In Section 3.4 we conduct simulations. Finally in Section 3.5 we apply the test statistics to data sets from economic literature.

3.5 Diagnostic Tests for Normality Based on Least Squares Residuals

The violation of the normality assumptions in regression residuals may have important results. Therefore many test statistics devised to test the normality of residuals. Jarque and Bera (1987) (JB) and Doornik and Hansen (1994) (DH) tests are the most popular among these tests. In this section we will introduce these tests.

Definition 12 *Let x be a random variable, μ denote the mean and $\mu_i = E(x - \mu)^i$ be the i'th central moment of x. The skewness and kurtosis are defined as:*

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} \ \ and \ \ \beta_2 = \frac{\mu_4}{\mu_2^2}$$

Sample counterparts are defined by;

$$\bar{x} = \frac{1}{T}, \ \ m_i = \frac{1}{T}\sum_{t=1}^{T}(x_i - \bar{x})^i,$$

Sample skewness and kurtosis is;

$$\sqrt{(b_1)} = \frac{m_3}{m_2^{3/2}} \ \ and \ \ b_2 = \frac{m_4}{m_2^2}$$

The JB test is based on a weighted average of the sample skewness and kurtosis

$$JB = T(\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24})$$

JB (1987) show that the test performs quite well relatively other tests, available in the literature. They show that the test is a Lagrange multiplier test if the alternatives to the normal distribution are in the Pearson family. As pointed in Section 2.3.2, the LM test is *asymptotically* chi-squared. But as Urzua (1996) points out the performance of the tests depens on the use of ciritical values determined by Monte Carlo simulation, since the convergence of the distribution to chi-squared $\mathcal{X}_2^2$ distribution is slow.

Doornik and Hansen (1994) developed an omnibus test based on Shenton and Bowman (1977) for normality. Let z_1 and z_2 be denote the transformed skewness and kurtosis. Doornik and Hansen suggest that the transformation creates statistics which are much closer to standard normal. So,

$$DH = z_1^2 + z_2^2$$

is *approximately* chi-squared $\mathcal{X}_2^2$. They show that their formula is closer to chi-squared than the JB formula. The explicit formula of z_1 and z_2 is presented in Appendix.

3.6 A Robust Estimator: LTS

The OLS estimators are effective under the hypothesis of normality of errors. But there are cases where this hypothesis is not satisfied. The data may include outliers. In that cases the nonnormality of the errors are not easily identified by the OLS residuals, because nonnormal errors may easily be masked by the OLS analysis. As an example consider the Figure 20, where five points lie nearly on a straight line, and one point is an outlier. One outlier make OLS estimate to misbehave. The OLS line has a negative slope and the residual of OLS does not reveal the outlier.

The robust estimators are devised as a remedy to this situation. These estimators are robust in the sence that they work well in the failure of the normality of errors assumption. They are not so strongly affected by outliers. The least trimmed squares (LTS) estimator is one of these robust estimators. It is introduced by Rousseeuw (1984). This has the property of being highly resistant to a relatively large proportion of outliers.

Definition 13 *Let $r_t^2(\beta)$ be the squared residual of the t-th observation. Let $r_{(1)}^2(\beta) \leq r_{(2)}^2(\beta) \cdots \leq r_{(T)}^2(\beta)$ be a rearrangement of the residuals in increasing order. The 50% trimmed LTS estimator is defined to be the value of β minimizing $\sum_{t=1}^{T/2} r_{(t)}^2$, where T is the sample size.*

Let S be any subsample of original sample, for the 50%trimmed LTS estimator it consist half of the data ,so its size is T/2. Let $\beta(S)$ and SSR(S) be the least squares estimator and sum of squared residuals for this subsample.

Let S^* be the optimal subsample of original sample, means that $\beta(S^*)$ is the LTS estimator. From the definition the following corrolary follows: Since S^* is optimal if one inside element is swapped with an outside element then SSR should increase.

Computation of the LTS is difficult as well as other robust estimators. But through following algorithm described in Zaman (1996), it is possible to compute LTS estimator. We choose any subsample S with size (T/2) from the original sample. We calculate SSR(S) for this subsample. Then we begin to swap each inside element with an outside element. If the SSR is reduced we start over with the new selected subsample containing the swapped element. Since the SSR is reduced at each step, we will reach a subset (S^*), where no further swap produces any reduction. The least squares estimator $\beta(S^*)$ will be the LTS estimator.

For details regarding the technique and its properties, see Rousseeuw and Leroy (1987), and also Chapter 5 of Zaman (1996).

3.7 The Normality Tests with LTS Estimator

As discussed in Section 3.1 both JB and DH are calculated on the basis of OLS residuals in standard analysis. But as we pointed out in Section 3.2, the OLS estimator can highly be influenced by the outliers. Since the robust estimators reveals the outliers, it may provide a clearer indication about lack of normality of residuals. In this section we will use the residuals of LTS instead of OLS residuals for JB and DH tests and try to see the effect on some real data sets which includes outliers. But first we calculated the significance points for each tests.

We conducted a Monte Carlo simulation to obtain the significance points of the JB and DH statistics and their robust versions (called as JB^* and DH^*). n observations are generated from N(0,1) distribution. Following Jarque and Bera (1987) the regressor matrix X consist of a column of ones and three colums of uniform random numbers. These matrix is same for each replications. Since the estimated error $\hat{\epsilon} = (I - X(X'X)^{-1}X')\epsilon$ does not depend on true β. The dependent variables can be calculated regardless of the value taken by β. After having the dependent variables and regressor matrix, it is possible to get OLS and LTS estimators. LTS estimators are calculated according the algorithm discussed in Section 3.2. JB, DH and their robust versions was calculated from the estimated residuals. This procedure is repeated 10.000 times; the α(10.000)th largest value is our α significance point. The significance points of the tests are presented in Table 1

Table 1: Significance Points for Four Tests for Normality

n	JB			JB^*			DH			DH^*		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	2.36	3.63	9.07	4.49	7.83	18.06	4.42	5.84	8.99	6.55	8.40	13.11
50	3.21	4.83	12.94	4.78	7.77	18.99	4.39	5.91	9.49	5.28	7.24	11.65

*10.000 replication.

Rousseeuw and Leroy (1987) use some data sets for their analysis related to robust regression. Through these data sets they show the high influence of outliers on OLS estimates and the benefits of the use of robust estimators. As a preliminary test of the validity of our basic idea, we conducted a study of tests for normality on five of these data sets.

Table 2: The Values of the Normality Test Statistics from Rousseeuw and Leroy Data Sets

Series	n	JB	JB^*	DH	DH^*
Brain	28	1.93	26.24	2.65	27.88
Cloud	19	2.23	10.62	2.95	15.05
Salinity	28	0.03	33.46	1.20	16.20
Stackloss	21	0.14	6.87	1.68	7.51
Aircraft	23	0.17	57.88	1.26	12.30

*Rousseeuw and Leroy (1987) from p. 57-96-82-76-154.

It is clear from Table 2 that the techniques uses OLS residuals fails to reject the null hypothesis normality with both JB^* and DH^* statistics. But the statistics with LTS residuals reject the null at 10% significance level. For most of the cases rejection at 1% significance level is observed.

Since the presence of big outliers (corresponding to lack of normality) in these series is well known, we see that JB and DH tests, conducted on the basis of OLS residuals can lead to wrong conclusions. This result directed us towards to examine the situation with simulated data.

3.8 Power Comparisons

We investigated the power of the tests as in Jarque and Bera (1980) with four alternative distributions; beta(3,2), Student's t (5), gamma (2,1) and log-normal. Additionally we used Cauchy distribution, since its distribution has heavy tails. Following a similar procedure for the calculations of critical values, we generated the true residuals from the listed alternative distributions and obtained normality test statistics accordingly. For a Monte Carlo sample size 10.000, we count the number of times the statistic exceeds the critical values given in Table 1. Dividing this number to 10.000, gives an estimate of the power. The results are presented in Table 3.

Much to our surprise JB and DH based on conventional OLS residuals outperformed JB* and DH*. The increase of the number of observations and the change of the percentage trimmed did not change the results, therefore not reported here. The alternative distributions used frequently for power comparison of tests for normality is not able to explain the situation with Rousseeuw and Leroy data sets.

Table 3: Estimated Power for n=20

Series	JB	JB^*	DH	DH^*
Beta	0.06	0.12	0.08	0.10
t	0.26	0.23	0.25	0.20
Gamma	0.39	0.29	0.37	0.29
Lognormal	0.74	0.60	0.74	0.62
Cauchy	0.82	0.76	0.83	0.77

*10.000 replications, $\alpha = 0.10$.

Therefore we conduct new simulations through the use of mixture of normal distributions as the alternative to the null of normality. In the examples related to failure of OLS in Section 3.2 we observe that the outliers lie in clusters. (For an example see Rousseeuw and Leroy (1987) p.58). In order to see the effect of this situation the new simulations are conducted as follows. The regressor matrix consist of one column of ones and one column of uniform random variables. Then the matrix is sorted from low to high. In that way, the outlier generated next to each other would also lie next to each other. The significance points related to new regressor matrix are presented in Table 4. This time the significance points for the LTS tests are calculated for 20% trimmed LTS test statistics.

Table 4: Significance Points for Four Normality Tests with New Regressor Matrix

n	JB			JB^*			DH			DH^*		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	2.34	3.58	8.92	3.34	5.15	12.23	4.43	5.91	9.32	5.05	6.57	10.31
50	3.17	5.13	12.78	3.82	6.50	14.55	4.50	5.90	10.06	4.92	6.40	10.64

*10.000 replication.

In order to make a comparison we generate also outliers in random places in samples. The first case, where the outliers lie in clusters is called as Case A and the second case, where they are in random places as Case B. In Table 5 the results for 20 and 50 observation are presented. In all cases 80% of the true residuals are generated from N(0,1) and 20% of them from normal distributions with different means and variances. So for 20 observations 4 and for 50 observations 10 outliers are generated. We matched the number of outliers with the trimmed part of the statistics. Through this way the efficiency loss of LTS procedure due to trimming the observations will be minimized. So the LTS tests are 20% trimmed.

The results are in favor of LTS tests, if the outliers are generated through the shift of mean and they are next to each other. The power increase over OLS tests is reached as much as 25% with 20 observation. But looking to Case B the same improvement are not observable. So if the outliers are in random places then the LTS is not so powerful. Also if the outliers has high variances LTS tests does not improve OLS tests substantially. This can be again support our views that LTS performs better if the outliers are in clusters. With balanced outliers, the OLS estimators is not much affected by the outliers, and hence OLS residuals are similar to robust residuals. This also explains why the Cauchy did not lead improved performance for JB^* and DH^* as we had expected. Outliers generated by a mean shift all lie on the one side of the regression and hence much more effective in sistematically distorting OLS estimates.

Table 5: Power Comparisons of Mixture of Normal Alternatives with 20% Outliers under Different Conditions

n=20		Case A	Case B		Case A	Case B
N(5,1)	JB	0.31	0.74	DH	0.31	0.75
	JB^*	0.44	0.63	DH^*	0.56	0.80
N(10,1)	JB	0.42	0.94	DH	0.45	0.94
	JB^*	0.68	0.98	DH^*	0.70	0.99
N(0,9)	JB	0.54	0.54	DH	0.55	0.53
	JB^*	0.54	0.54	DH^*	0.58	0.53
N(0,16)	JB	0.72	0.71	DH	0.73	0.71
	JB^*	0.73	0.71	DH^*	0.76	0.71
n=50		**Case A**	**Case B**		**Case A**	**Case B**
N(5,1)	JB	0.59	1.00	DH	0.61	1.00
	JB^*	0.86	1.00	DH^*	0.87	1.00
N(10,1)	JB	0.69	1.00	DH	0.74	1.00
	JB^*	0.75	1.00	DH^*	0.87	1.00
N(0,9)	JB	0.85	0.83	DH	0.85	0.83
	JB^*	0.85	0.83	DH^*	0.85	0.82
N(0,16)	JB	0.96	0.95	DH	0.96	0.96
	JB^*	0.96	0.95	DH^*	0.97	0.96

*10.000 replicaitons, α= 0.10.

3.9 Applications

Through simulations we have shown that the residuals from robust regression lead to improvement in rather specialized circumtances. In order to asses the use of these tests on economic data sets, we made a few applications.

Table 6: The Values of the Normality Test Statistics from Tansel (1993) and Metin (1998) Models

	Tansel (1993)				Metin (1998)	
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2
n/k	28/6	28/5	29/5	28/7	33/17	33/11
JB	0.72	1.27	0.83	1.30	0.73	1.48
JB^*	44.68	25.15	6.99	7.07	1245.57	14.06
DH	0.41	1.76	0.62	1.69	0.24	1.85
DH^*	35.19	7.23	6.68	9.07	582.74	20.96

*k is the number of regressors.

LTS statistics are 50% trimmed.

First we applied our robust tests to models used by Tansel (1993) investigating the cigarette demand for Turkey. We repeated the regressions and computed the normality test statistics. The results are presented in Table 6. Using the critical values from Table 1 it is seen that JB test fail to reject normality for all of the models as in Tansel (1993) We reached the same conclusion with the DH test. But if we use robust tests, the decisions of tests is different in most of the cases. JB^* and DH^* reject the null for Model 1 at 1% significance level. JB^* test reject the null for Model 2 at 1%, DH^* test at 10% significance level. Both LTS tests reject the null for Model 3 at 10% significance level. JB^* test reject the null for Model 4 at 10%, DH^* test at 5% significance level. According to Orhan and Zaman(1999) these data sets includes outliers, which implies the true residuals are not distributed normally. Their results support the inference drawn from our robust tests.

Metin (1998) analyzes the relationship between inflation and the budget deficit in Turkish economy. We replicated the first and second model there and get the normality test results presented in Table 6. By using the critical values in Table 1 it is seen that the DH test statistics fail to reject the null hypothesis of normality for both of the models as in Metin (1998). Through JB statistic result we also reach the same conclusion. But the results of robust test are just the opposite. For Model 1 The JB^* and DH^* reject the normality at 1% significance level. For Model 2 the JB^* and DH^* reject the null at 5% and 1% significance level respectively. Again the robust residuals suggest significant nonnormality in the errors which are not detected by OLS regression analysis.

CHAPTER IV
CONCLUSION

In this dissertation, we attempted to compare the power of the LM, Wald and LR tests for the first order autocorrelation model through the approximations to the distributions of these three tests firstly.

Accurate approximations to test statistics becomes great importance if the exact distribution of the statistic is not known. Asymptotic theory provides us approximation which are not accurate enough in most of the cases. In this dissertation we have shown how asymptotic expansion techniques can be succesfully applied to some test statistics for a first order autoregressive model. We obtained the statistics first empirically, then compared with our theoretical approximations. We benefit from Edgeworth approximation through the dissertation. Our approximations has good accuracy for the sample size as small as 30. It is always possible to apply other approximation techniques and get better approximations for the smaller sample sizes. One of them is the Saddlepoint approximation technique presented in Section 2.1.5.

One difficulty for the study of higher order approximation, it is a new research area and theoretical formulas derived by statisticians are valid for quite general cases. In order to apply in econometrics tedious calculations are needed. This is also the reason for its slow spread in econometric applications. But once the suitable formulation for the econometric problem obtained, it is possible to derive good approximations for many statistics. As it is presented in Section 2.2, the studies of higher order approximations have satisfactory results in different topics of econometrics. But there are still many interesting subjects for further research.

In this dissertation we have obtained first the empirical distributions of the power envelope and three test statistics to compare the adeqacy of the approximations. Hovewer, we get the corrected ciritical values before the derivation of the powers. This step was very essential, since the application of inaccurate critical values to a test affects the power of the test and could be misleading. We have adjusted the critical values so that the true size of the test can become close enough to nominal size of the tests. The nominal size in our experiments fixed to 5% through the analysis.

As far as the approximation of the power envelope and the power curves of the test statistics are concerned, we have found out that the second order approximation is accurate enough for our purposes in general. With regard to the LR test statistics, we have tried to fit a functional form and so obtained our approximation as a result.

We have compared the performance of test statistics in the dissertation. According to the first order asymptotic theory, the local power of the LR, Wald and LM test are the same. However first empirically and afterwards through approximate power functions we have shown that the performances of the three tests are different. In this respect, we have found that, Wald and LR statistic have a strong superiority, becauese the stringency of these tests are found less than 1%. On the other hand, the stringency of the LM test is too high for small number of observations and reaches to 5% as the number of observation reaches to 100. So we suggest strongly the use of LR and Wald test for the test of first order autocorrelation as far as the stringency results are concerned.

Our technique to approximate test statistics can be applied in the different reserch areas of econometrics. Hovewer, we recommend that unit root case may be examined as a first attempt, which has both economic and statistical implications and attracted attention in the literature. Although many test statistics have been devised to test the null hypothesis that a time series posess unit root, the asymptotic approximation of test statistic is not extensively examined until now.

Another research topic could be to compare the performances of the test statistics for the first order autocorrelation of the residuals. Especially the LM test for the first order autocorrelation, which is called as Durbin Watson test, is very famous amongst these tests. Regarding autocorrelation of residuals, exact distribution of these tests also not known and the comparisons through asymptotic approxiamtion to test statistics are not conducted in the literature as far as we are concerned.

In the second part of the dissertation we have suggested to use robust estimators instead of the OLS estimator for the test of normality of regression residuals. Regarding this we have shown by using real data that the tests using OLS residuals may cause wrong conclusions.

Also through our simulations we compared the performance of the two approaches. In simulation study the choose of alternative distribution has become importance. We have conducted simulation by using the framework of Jarque and Bera (1987) firstly. With the use of the alternative distributions in that study, power improvement over the standard normality (JB and DH) tests could not be observed.

Afterwards we conducted new simulations, allowing for a larger set of alternatives by using mixture of normal distributions. The results implies that, when the outliers are clustered they have big distortion effect on the OLS estimators and these situation lead to maximum improvement for tests based on robust estimators. We have also shown on two applications, inference drawn from the two approaches is differing.

Although the simulation result show the improvement of the robust test under specialized situations, through real data it is stated that these situation occur often enough in practice. This result support our views about the benefit to use tests for normality based on robust regression as a diagnostic test.

SELECT BIBLIOGRAPHY

Abadir, K.M. 1993. "On the Asymptotic Power of Unit Root Tests," *Econometric Theory* 9:189-221.

Amari, S. L., O. E. Barndorff-Nielson, R. E. Kass, S.S. Lauritzen and C. R. Rao. 1987. *Differential Geometry in Statistical Inference*, Lecture Notes-Monograph Series 10, Institute of Mathematical Statistics, Hayward, California.

Attfield, C.L.F. 1991. "A Bartlett Adjustments to the Likelihood Ratio Test for Homoskedasticity in the Linear Model," *Economics Letters* 37:119-123.

Bahadur, R.R. 1966. *An Optimal Property of the Likelihood Ratio Statistics*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1:13-26.

Barndorff-Nielson, O. and D. R. Cox. 1979. "Edgeworth and Sadle-Point Approximations with Statistical Applications," *Journal of the Royal Statistical Society B* 41:279-312.

—–. 1989. *Asymptotic Techniques for Use in Statistics*, London: Chapman and Hall.

Bernt, E. and N. E. Savin. 1977. "Conflict Among Criteria for Testing Hypothesis in the Multivariate Linear Regression Models," *Econometrica* 45:1263-1277.

Bhattacharya, R. N. and R. R. Rao. 1976. *Normal Approximation and Asymptotic Expansion, Wiley, Springer-Verlag*, New York.

Bickel, P. 1974. "Edgeworth Expansions in Nonparametric Statistics," *Annals of Statistics* 2:1-20.

Brown, L. D. 1973. "Estimation with Incompletely Specified Loss Functions," *Journal of American Statistical Association* 70:417-427.

Chandra, T.K. and J. K. Ghosh. 1979. "Valid Asymptotic Expansions for the Likelihood Ratio Statistics and Other Perturbed Chi-Square Variables," *Sankhya Series A* 41:22-47.

—. 1980. "Valid Asymptotic Expansions for the Likelihood Ratio Statistics and Other Statistics Under Contiguous Alternatives," *Sankhya Series A* 42:170-184.

Chesher, A. and R.J. Smith. 1997. "Likelihood Ratio Specification Tests," *Econometrica* 65(3):627-646.

Chesher, A. and R.Spady. 1991. "Asymptotic Expansions of the Information Matrix Test Statistic", *Econometrica* 59:787-815.

Chibisov, D. M. 1974. "Asymptotic Expansions for Some Asymptotically Optimal Tests," In J.Hajek,ed., *Proceedings of the Prague Symposium on Asymptotic Statistics* 2:37-68.

Cordeiro, G. M., D.A. Botter and S.L.D.P. Ferrari. 1994. "Nonnull Asymptotic Distributions of Three Classic Criteria in Generalised Linear Models," *Biometrika* 81(4):709-720.

Cribari-Neto, F. and S. Zarkos. 1995. "Improved Test Statistics for Multivariate Regression," *Economics Letters* 49:113-120.

Daniels, H. E. 1954. "Saddlepoint Approximations in Statistics," *Annals of Mathematical Statistics* 25:631-650.

—. 1956. "The Approximate Distribuiton of Serial Correlation Coefficients," *Biometrika* 43:169-185.

Dickey, D. A. and W. A. Fuller. 1979. "Distribution of Estimators for Autoregressive Time Series with a Unit Root," *Journal of American Statistical Association* 74:427-431.

—. 1981. "Likelihood Ratio Statistics for Autoregressive Statistic with a Unit Root," *Econometrica* 49:1057-1072.

Doornik, J. A. and H. Hansen. 1994. "An Omnibus Test for Univariate and Multivariate Normality," Working Paper, Nuffield College, Oxford, 1-16.

Durbin, J. 1980. "The Approximate Distribution of Partial Serial Correlation Coefficients Calculated from Residuals from Regression on Fourier Series," *Biometrika* 67(2):335-346.

Engle, R. F. 1984. "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," In Z Grilliches and M.D. Intriligator eds., *Handbook of Econometrics*, Volume II,Elsevier Science Publishers.

Evans, G. B. A. and N. E. Savin. 1981. "Testing for Unit Roots 1," *Econometrica* 49:1057-1072.

—. 1982. "Conflict Among the Criteria Revisited," *Econometrica* 50(3):737-748.

—. 1984. "Testing For Unit Roots:2," *Econometrica*, 52:1241-1269.

Field, C. and E. Ronchetti. 1990. *Small Sample Asymptotics* IMS Lecture Notes-Monograph Series, USA.

Gouieroux, C., H. Alberto and A. Monfort. 1982. "Likelihood Ratio Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica* 50(1):63-80.

Gurmu, S. and P.K.Trivedi. 1992. "Overdispersion Tests for Truncated Poisson Regression Models," *Journal of Econometrics* 54:347-370.

Hall, P. and J.L. Horowitz. 1996. "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica* 64(4):891-916.

Harris, P. 1985. "An asymptotic Expansion for the Null Dsitribution of the Efficient Score Statistics," *Biometrika* 72:653-659.

Harris, P. and H. W. Peers. 1980. "The Local Power of the Efficient Scores Test Statistics," *Biometrika* 67:525-529.

Hayakawa, T. 1975. "The Likelihood Ratio Criterion for a Composite Hypothesis Under a Local Alternative," *Biometrika* 62(2):451-460.

—. 1977. "The Likelihood Ratio Criterion and the Asymptotic Expansion of its Distribution," *Ann. Inst. Statist. Math. A* 29:359-378.

Hayakawa, T. and M. L. Puri. 1985. "Asymptotic Expansions of the Distributions of Some Test Statistics," *Ann. Inst. Statist. Math. A* 37:95-108.

Hoeffding, W. 1965. "Asymptotically Optimal Tests for Multinomial Distributions," *Annals of Mathematical Statistics* 36:369-412.

Honda, Y. 1988. "A Size Correction to the Lagrange Multiplier Test for Heteroskedasticity," *Journal of Econometrics* 38:375-386.

Jarque, C. M. and A. K. Bera. 1987. "A test for normality of observations and regression residuals," *International Statistical Review* 55:163-172.

Jensen, J.K. 1995. *Saddlepoint Approximations*, Oxford Statistical Science Series 16, Oxford:Clarendon Press.

Kallenberg, W. C. M. 1982. "Chernoff Efficiency and Deficiency," *Annals of Statistics* 10:583-594.

Kunitomo, N., K. Morimune and Y.Tsukuda. 1983. "Asymptotic Expansions of the Distributions of the Test Statistics for Overidentifying Restrictions in a System of Simultaneous Equations," *International Economic Review* 24:199-215.

Larsson, R. 1995. "The Asymptotic Distributions of some Test Statistics in Near-Integrated Ar Processes," *Econometric Theory* 11:306-330.

Lieberman, O. 1994. "On the Approximation of Saddlepoint Expansions in Statistics," *Econometric Theory* 10:900-916.

Lugannani, R. and S.O. Rice. 1980. "Saddlepoint Approximation for the Distribution of Sum of Independent Random Variables," *Adv. Appl. Prob* 12:475-490.

Maekawa, K. 1988. "Comparing the Wald, LR and LM Tests for Heteroscedasticity in a Linear Regression Model," *Economics Letters* 26:37-41.

Magdalinos, M.A. 1988. "The Local Power of The Tests of Overidentifying Restrictions," *International Economic Review* 29:509-525.

—–. 1992. "Stochastic Expansions and Asymptotic Approximations," *Econometric Theory* 8:343-367.

Magdalinos, M.A. 1994. "Testing Instrument Admissibility:Some Refined Asymptotic Results," *Econometrica* 62:373-403.

Magdalinos, M.A and S.D.Symeonides. 1996a. "Alternative Methods of Size Correction for some GLS Test Statistics:The Case of Heteroscedasticity," Technical Report, Athens School of Economics.

——. 1996b. "A Reinterpretation of the Tests of Overidentifying Restrictions," *Journal of Econometrics* 73:325-353.

Magee, L. 1989. "An Edgeworth Test Size Correction For The Linear Model With AR(1) Errors," *Econometrica* 57:661-674.

Magnus, J, R. 1978. "The Moments of Products of Quadratic Forms in Normal Variables," *Statistica Neederlandica* 32:201-210.

Marsh, P.W.N. 1998. "Saddlepoint Approximations for Noncentral Quadratic Forms," *Econometric Theory* 14:539-559.

McCullagh, P. 1987. *Tensor Methods in Statistics* New York.Chapman and Hall

McManus, D. A., J.C. Nankervis and N.E. Savin. 1994. "Multiple Optima and Asymptotic Approximations in the Partial Adjustment Model," *Journal of Econometrics* 62:91-128.

Metin, K. 1998. "The Relationship Between Inflation and the Budget Deficit in Turkey," *Journal of Business & Economic Statistics* 16(4):412-422.

Morimune, K. 1989. "*t* Test in A Structural Equation," *Econometrica*, 57:1341-1360.

Nabeya, S. and K.Tanaka. 1990. "Limiting Power of Unit-Root Tests in Time-Series Regression," *Journal of Econometrics* 46:247-271.

Neyman, J. and E. S. Pearson. 1928. "On the Use of Interpretation of Certain Test Criteria for the Purpose of Statistical Inference," *Biometrika* 20A:175-240, 263-294.

Orhan, M. and A. Zaman. 1999. "Econometric Applications of High-Breakdown Robust Regression Techniques," Technical Report, Bilkent University, 06533, Bilkent, Ankara.

Pearson, E. S., D'Agostina R. B. and K.O. Bowman. 1977. "Tests for Departure From Normality: Comparison of Powers," *Biometrika*, 64(2):231-246.

Peers, H. W. 1971. "Likelihood Ratio and Associated Test Criteria," *Biometrika*, 58(3):577-587.

Perron, P. 1989. "The Calculation of the Limiting Distribution of the Least-Squares Estimator In a Near-Integrated Model," *Econometric Theory* 5:241-255.

—. 1991. "A Continuous Time Approximation to the Unstable First-Order Autoregressive Process:The Case Without an Intercept," *Econometrica* 59:211-236.

Pierce, D. A. and R. J. Gray. 1982. "Testing Normality of Errors in Regression Models," *Biometrika* 69(1):233-236.

Pfanzagl, J. 1973. "Asymptotically Optimum Estimation and Test Procedures," In J. Hajek ed., *Proceedings of the Prague Symposium on Asymptotic Statistics* Vol 2, 201-272.

—. 1980. "Asymptotic Expansions in Parametric Statistical Theory,"In. P. R. Krishniah ed., *Developments in Statistics* Academic Press.

Pfanzagl, J. and W. Wefelmeyer. 1978. "An Asymptotically Complete Class of Tests," *Zeitschrift fr Wahrscheinlichkeitstheorie verw. Gebiete* 45:49-72.

Phillips, P. C. B. 1977a. "Approximations to Some Finite Sample Distributions Associated with a First-Order Stochastic Difference Equation," *Econometrica*, 45(2):463-485.

—. 1977b. "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Economic Estimators," *Econometrica* 45(6):1517-1534.

—. 1978. "Edgeworth and Saddlepoint Approximations in the First Order Noncircular Autoregression," *Biometrika* 65(1):91-98.

—. 1980. "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume," *Review of Economic Studies* 47:183-224.

—. 1987a. "Time Series Regression With a Unit Root," *Econometrica* 55:277-301.

—. 1987b. "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika* 74:535-547.

Phillips, P.C.B and J.Y. Park. 1988. "On the Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica* 56(5):1065-1083.

Rao, C. R. 1947. "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Application to Problems of Estimators," *Proc. Comp. Phil. Soc.* 44:50-57.

—–. 1973. *Linear Statistical Inference and Its Applications* 2nd ed., Wiley, New York.

Rayner, R. K. 1990. "Bootstrap Tests for Generalized Least Squares Regression Models," *Economics Letters* 34:261-265.

Rothenberg, T. J. 1982. "Comparing Alternative Asymptotically Equivalent Tests," In :W. Hildenbrand,ed., *Advances in Econometrics* Cambridge: Cambridge University Press.

—–. 1984a. "Approximating the disitributions of Econometric Estimators and Test Statistics," In: Z Grilliches and M.D. Intriligator eds., *Handbook of Econometrics*, Volume II, Elsevier Science Publishers.

—–. 1984b. "Approximate Normality of Generalized Least Squares Estimates," *Econometrica* 52:811-825.

—–. 1984c. "Hypothesis Testing in Linear Models When The Error Covariance Matrix Is Nonscalar," *Econometrica* 52:827-843.

—–. 1988. "Approximate Power Functions for Some Robust Tests of Regression Coefficients," *Econometrica* 56(5):997-1019.

Rousseeuw, P. J. 1984. "Least Median Squares Regression," *J. Am. Statist. Ass*, 79:871-880.

Rousseeuw, P. J. and A. M. Leroy. 1987. *Robust regression & Outlier Detection* John Wiley & Sons. Inc.

Sargan, J. D. 1975. "Gram-Charlier Approximations Applied to t Ratios of k-Class Estimators," *Econometrica* 43:327-346.

—–. 1976. "Econometric Estimators and the Edgeworth Approximation," *Econometrica* 44(3):421-448.

—–. 1980. "Some Approximations to the Distribution of Econometric Criteria Which are Asymptotically Distributed As Chi-Squared," *Econometrica* 48:1107-1139.

Sargan, J.D. and S.E. Satchell. 1986. "A Theorem of Validity for Edgeworth Expansions," *Econometrica* 54(1):189-213.

Self, S.G. and K. Liang. 1987. "Asymptotic Properties of Maximum Likelihood

Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *J. A. Stat. Ass.*, 82(398):605-610.

Serfling, R. J. 1980. *Approximation Theorems in Mathematical Statistics* John Wiley & Sons.

Shenton, L. R. and K. O. Bowman. 1977. "A Bivariate Model for the Distribution" of $\sqrt{b_1}$ and b_2, *J. Am. Statist. Ass.* 72:206-211.

Skovgaard, I.M. 1986. "On Multivariate Edgeworth Expansions," *International Statistical Review* 54:169-186.

Tanaka, K. 1983. "Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean," *Econometrica* 51(4):1221-1231.

Tansel, A. 1993. "Cigarette Demand, Health Scares and Education in Turkey," *Applied Economics* 25:521-529.

Taniguchi, M. 1985. "An Asymptotic Expansion for the Distribution of the Likelihood Ratio Criterion for a Gaussian Autoregressive Moving Average Process Under a Local Alternative," *Econometric Theory* 1:73-84.

Urzua, C. M. 1996. "On the Correct Use of Omnibus Tests for Normality," *Economics Letters* 53:247-251.

Wald, A. 1942. "On the Principles of Statistical Inference," *Notre Dame Mathematical Lectures*, 1, Indiana.

——. 1943. "Test of Statistical Hypothesis Concerning Several Parameters When the Number of Observations is Large," *Trans. Amer. Math. Sos.* 54:426-482.

Wang, S. 1992. "Tail Probability Approximations in the First-Order Noncircular Autoregression," *Biometrika* 79:431-434.

White, H. 1982. "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50:1-25.

White, H. and G. M. Macdonald. 1980. "Some Large Sample Tests for Nonnormality in the Linear Regression Model," *J. Am. Statist. Ass.* 75:16-28.

Zaman, A. 1996. *Statistical Foundations for Econometric Technique* Academic Press.

APPENDIX A

The Explicit Formula of z_1 and z_2 for DH test

z_1 is transformed from $\sqrt{b_1}$, as follows;

$$\begin{aligned}
\beta &= \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \\
\omega^2 &= -1+(2\beta-1))^{1/2}, \\
\delta &= \frac{1}{\left(log(\sqrt{\omega^2})\right)^{1/2}}, \\
y &= \sqrt{b_1}\left(\frac{\omega^2-1}{2}\frac{(n+1)(n+3)}{6(n-2)}\right)^{1/2}, \\
z_1 &= \delta log\left(y+(y^2+1)^{1/2}\right).
\end{aligned}$$

z_2 is transformed from b_2 as follows;

$$\begin{aligned}
\delta &= (n-3)(n+1)(n^2+15n-4), \\
a &= \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta}, \\
c &= \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta}, \\
k &= \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12\delta}, \\
\alpha &= a+b_1c, \\
X &= (b_2-1-b_1)2k, \\
z_2 &= \left(\left(\frac{X}{2\alpha}\right)^{1/3}-1+\frac{1}{9\alpha}\right)(9\alpha)^{1/2}.
\end{aligned}$$

APPENDIX B

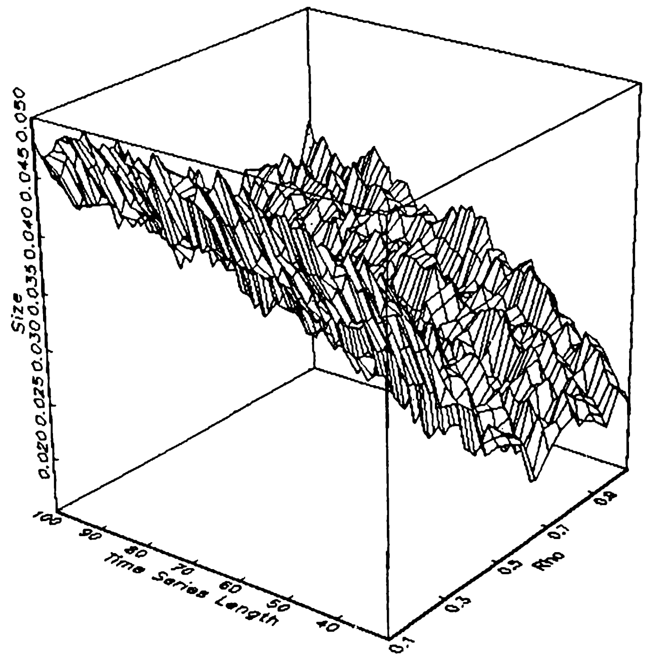

Figure 1: Empirical Size of First Order Approximation

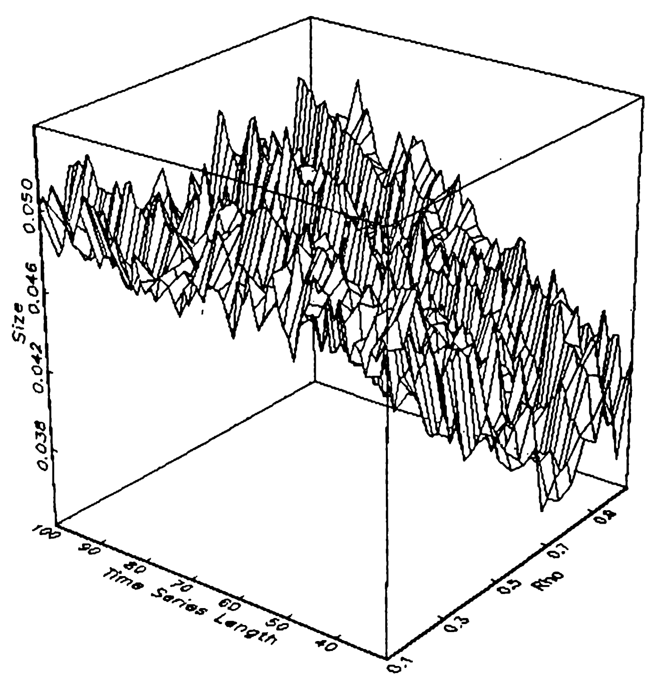

Figure 2: Empirical Size of Second Order Approximation

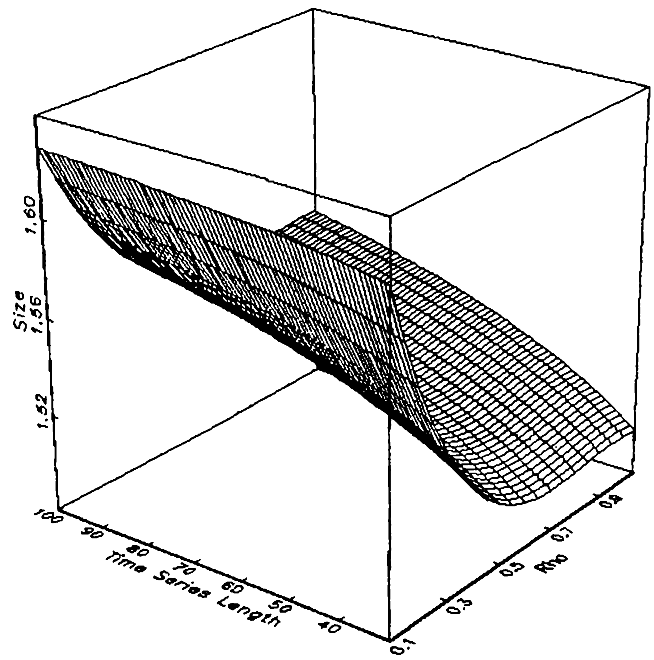

Figure 3: Theoretical Approximation to Critical Value

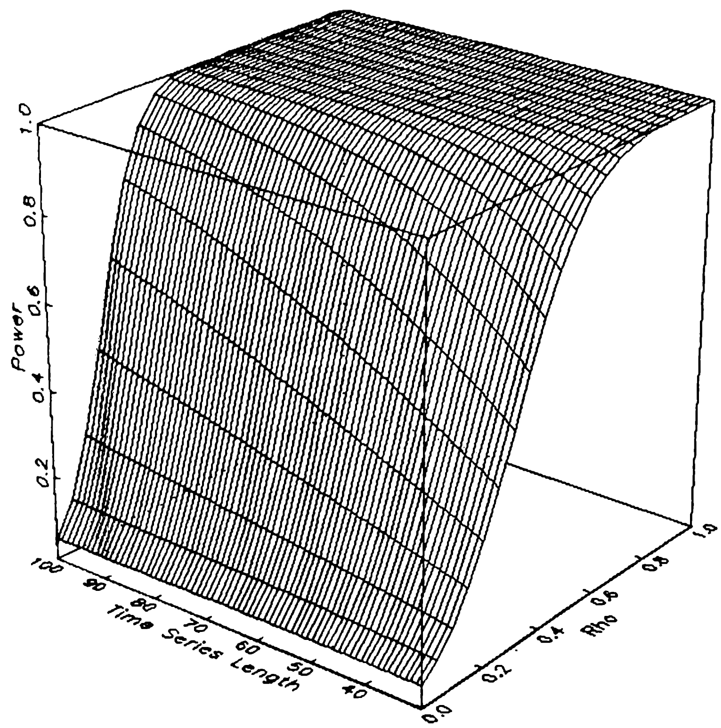

Figure 4: Theoretical Approximation to Power Envelope

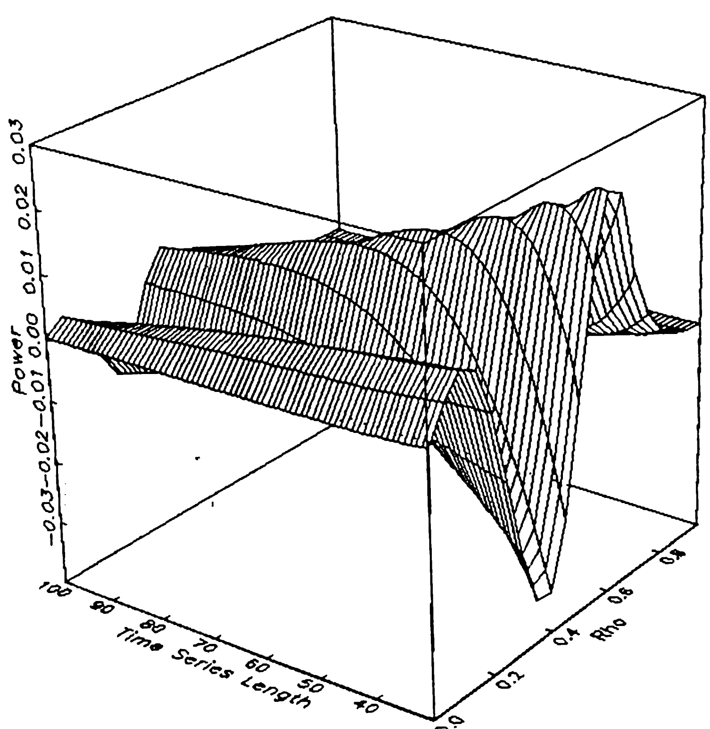

Figure 5: Approximation Error of Power Envelope

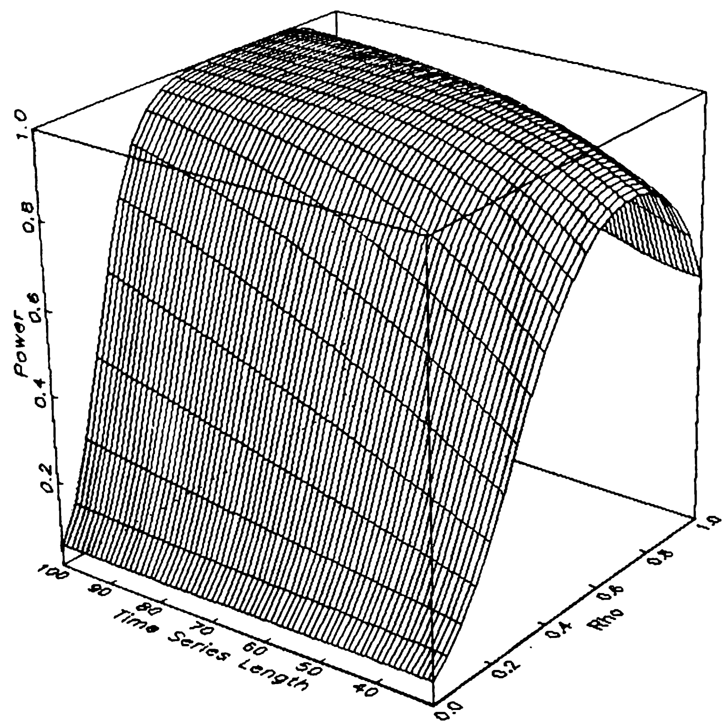

Figure 6: Theoretical Approximation to LM Test

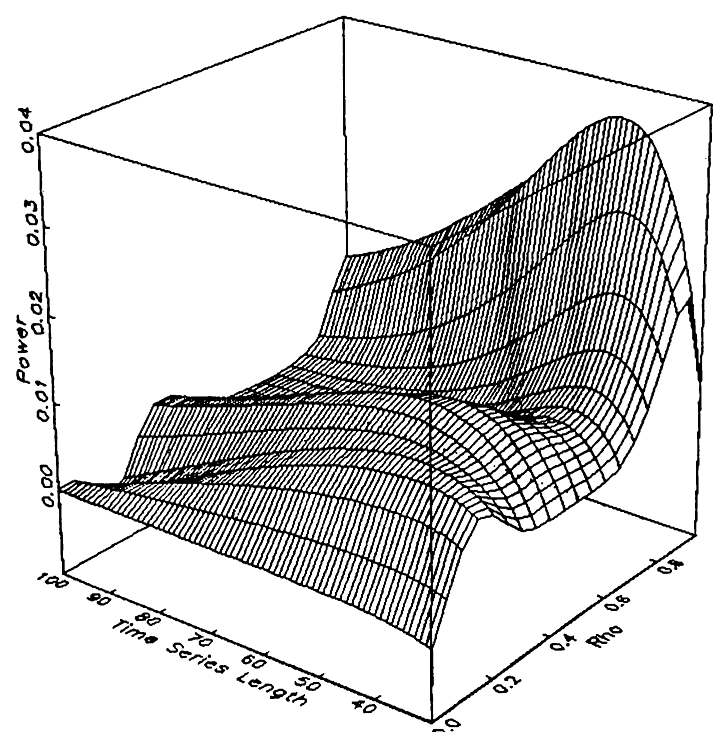

Figure 7: Approximation Error of LM Test

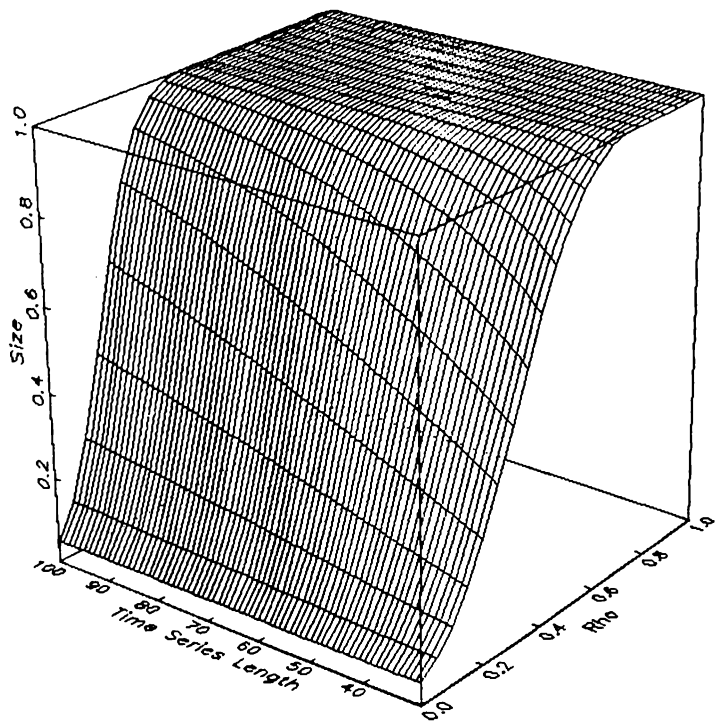

Figure 8: Theoretical Approximation to Wald Test

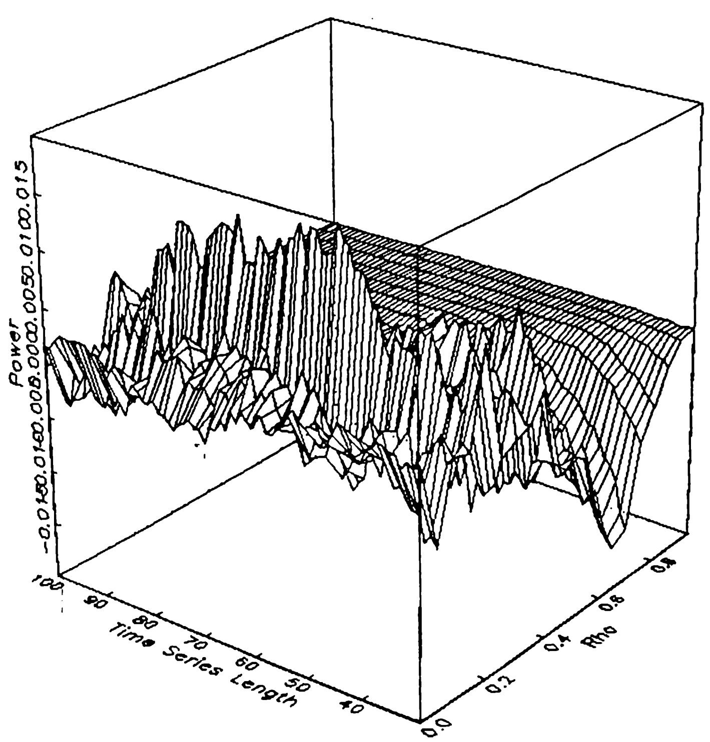

Figure 9: Approximation Error of Wald Test

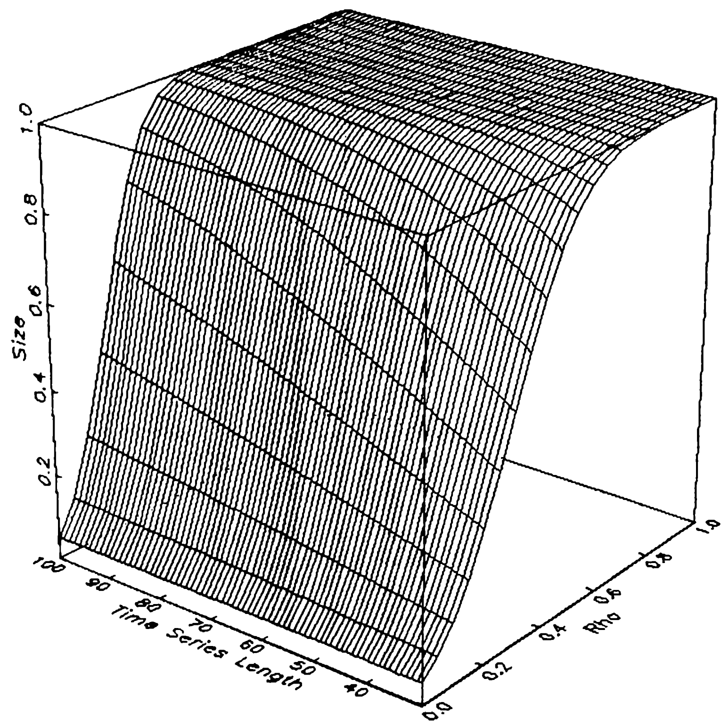

Figure 10: Theoretical Approximation to LR Test

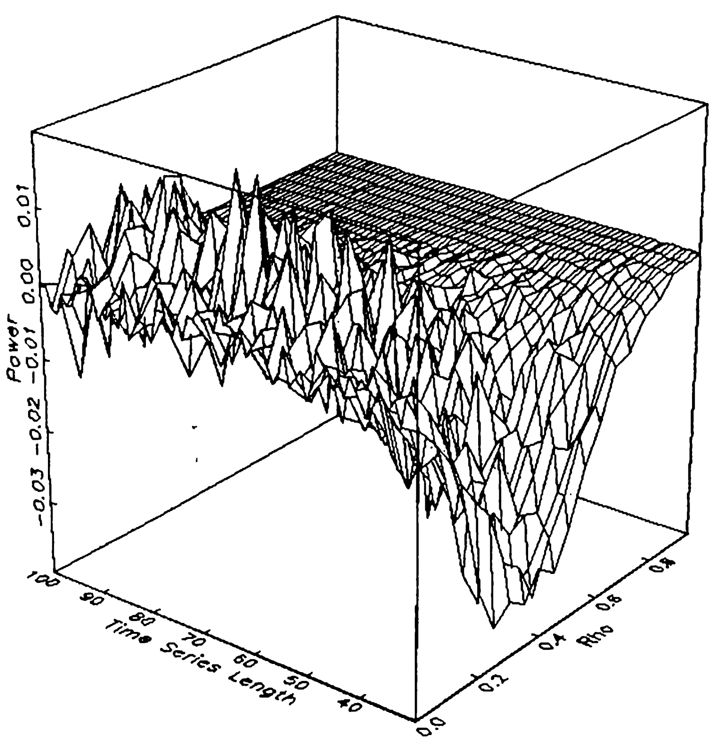

Figure 11: Approximation Error of LR Test

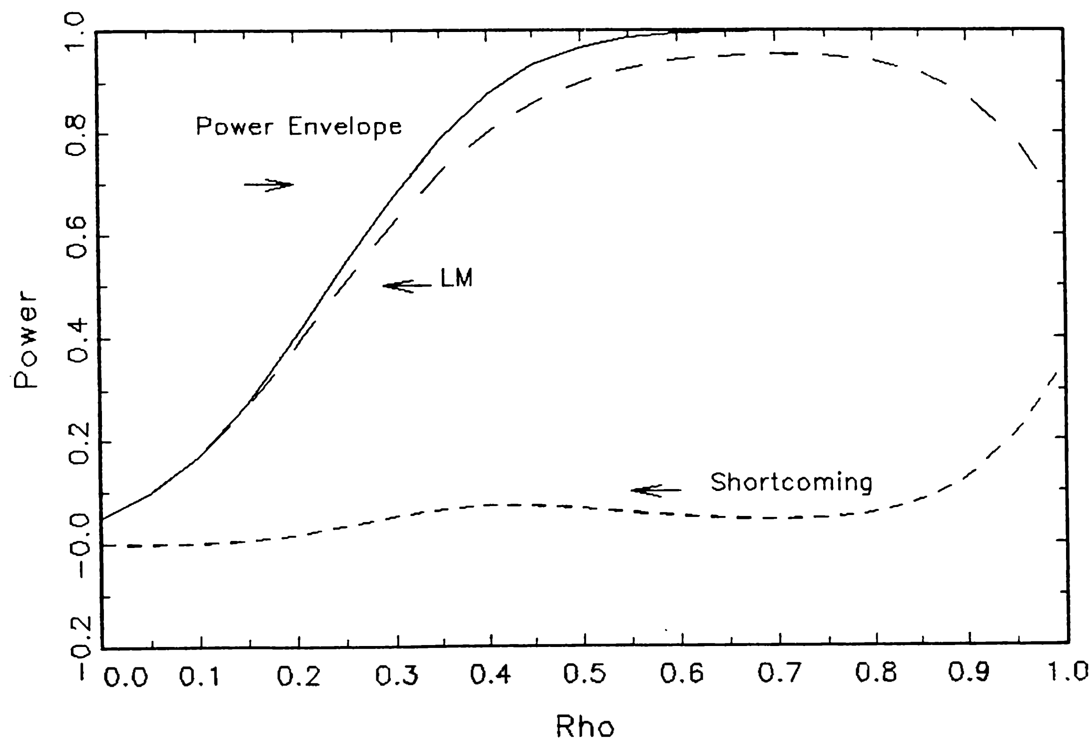

Figure 12: Power Envelope vs. LM for T=50

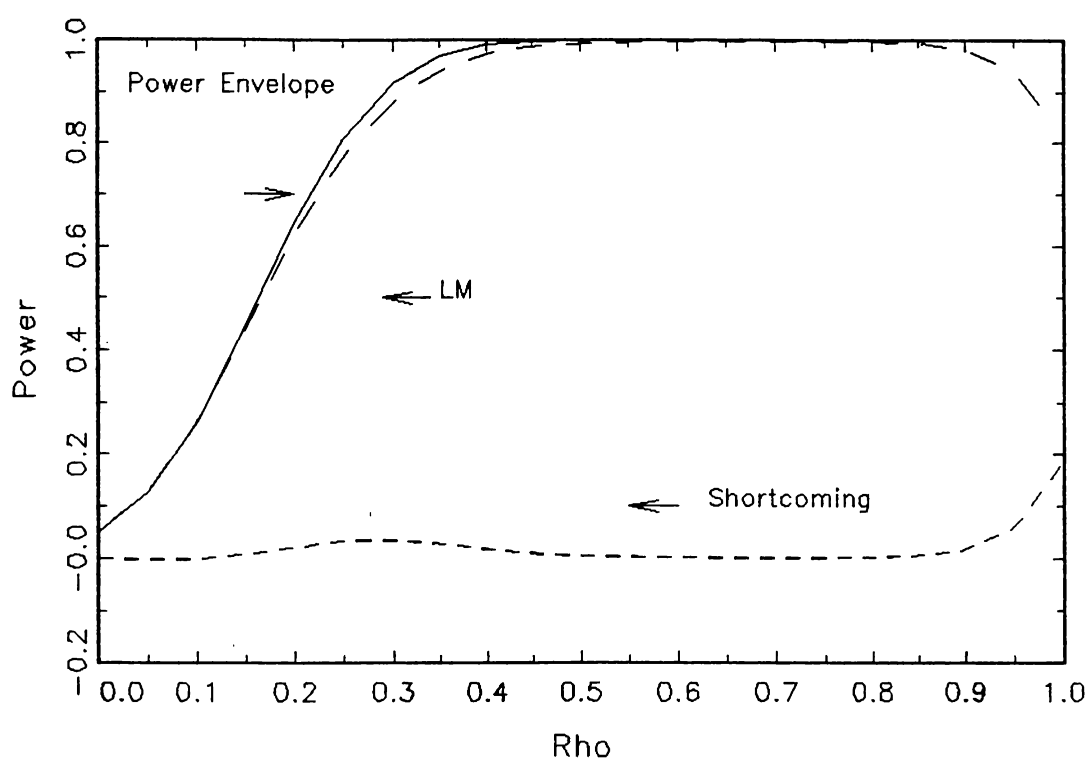

Figure 13: Power Envelope vs. LM for T=100

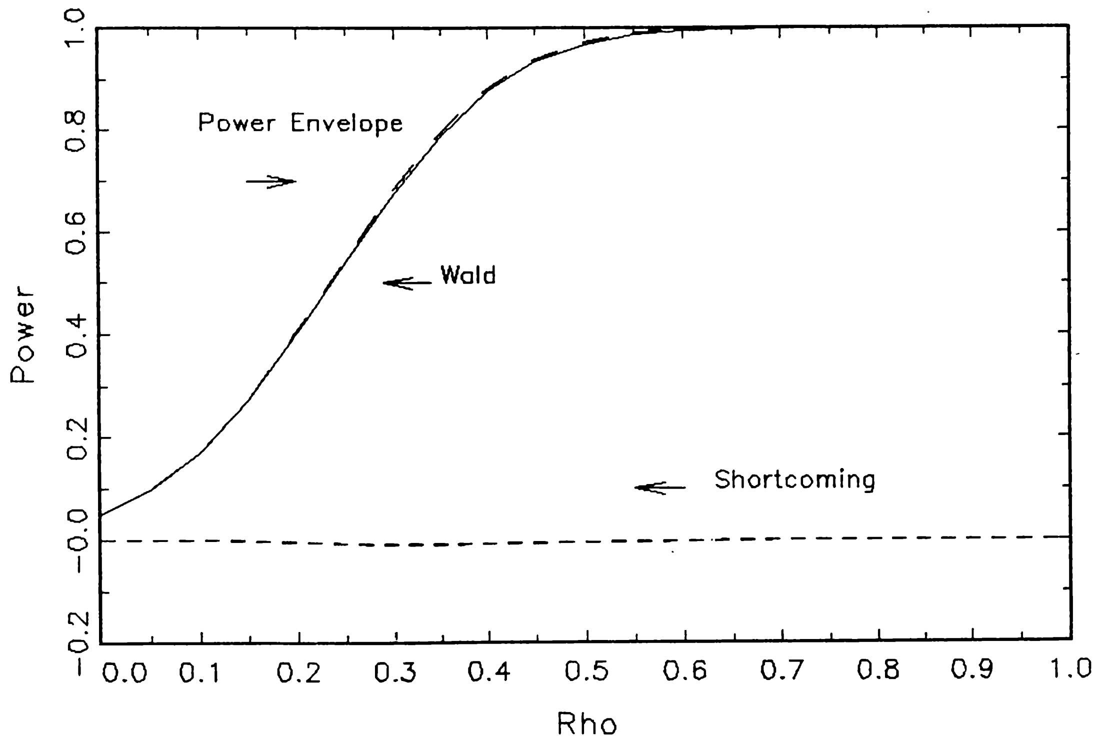

Figure 14: Power Envelope vs. Wald for T=50

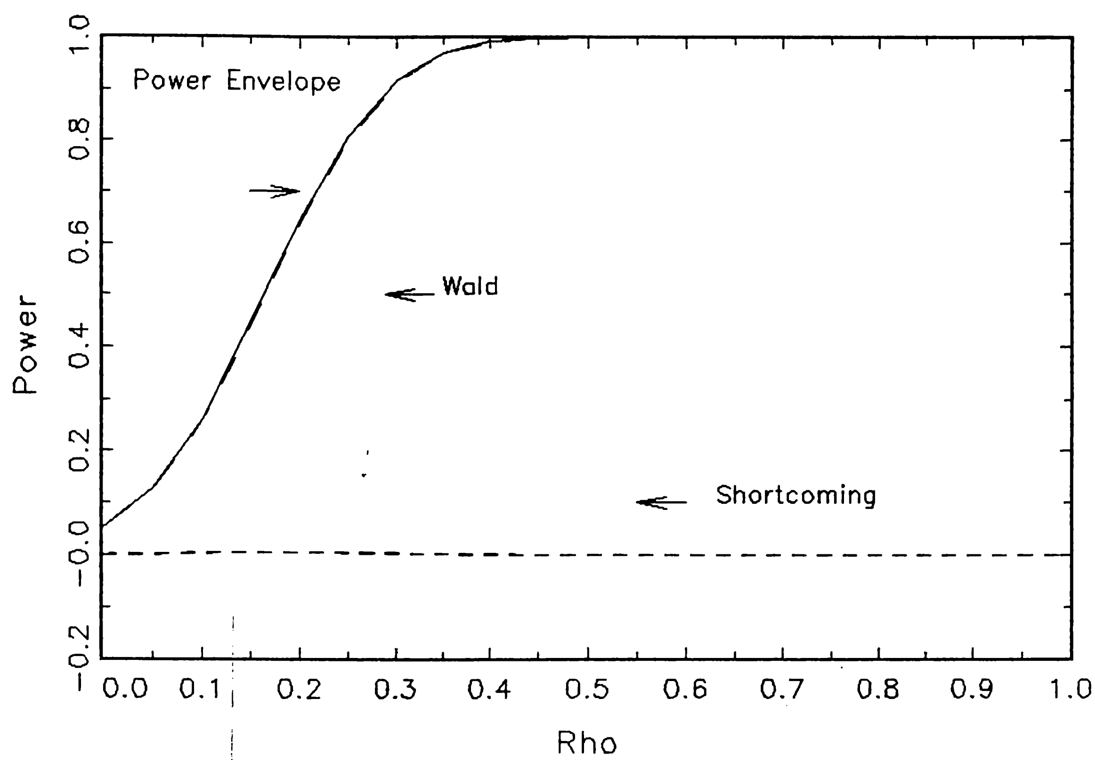

Figure 15: Power Envelope vs. Wald for T=100

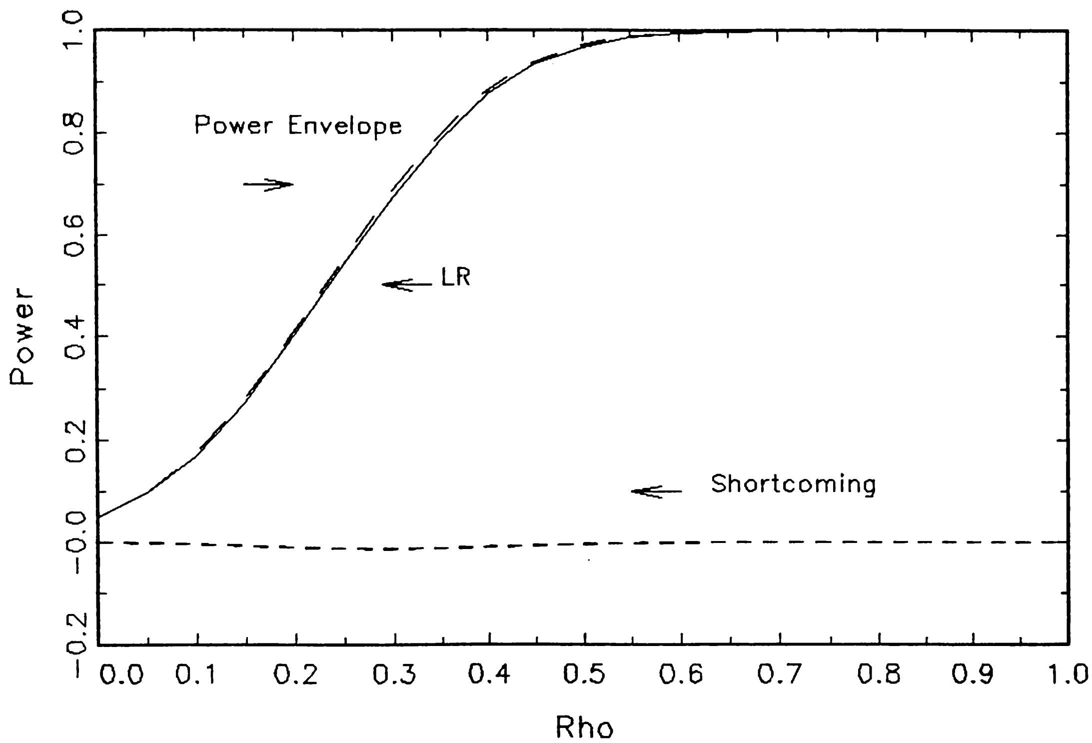

Figure 16: Power Envelope vs. LR for T=50

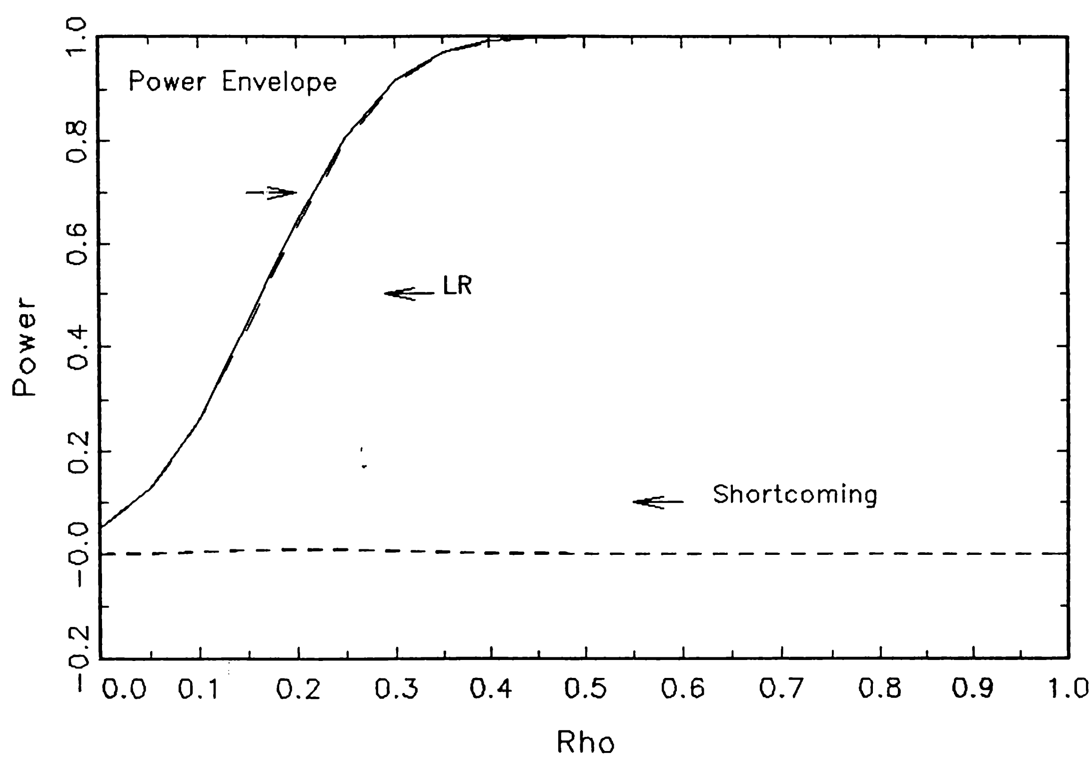

Figure 17: Power Envelope vs. LR for T=100

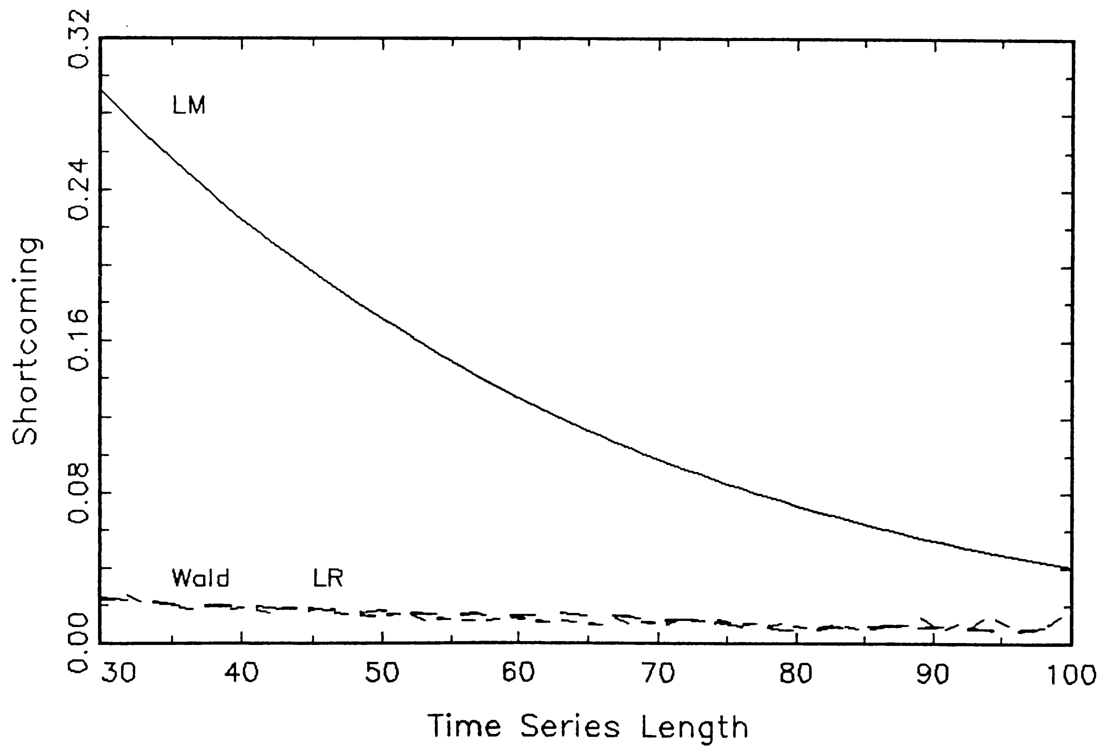

Figure 18: Empirical Shortcoming of LM, Wald and LR

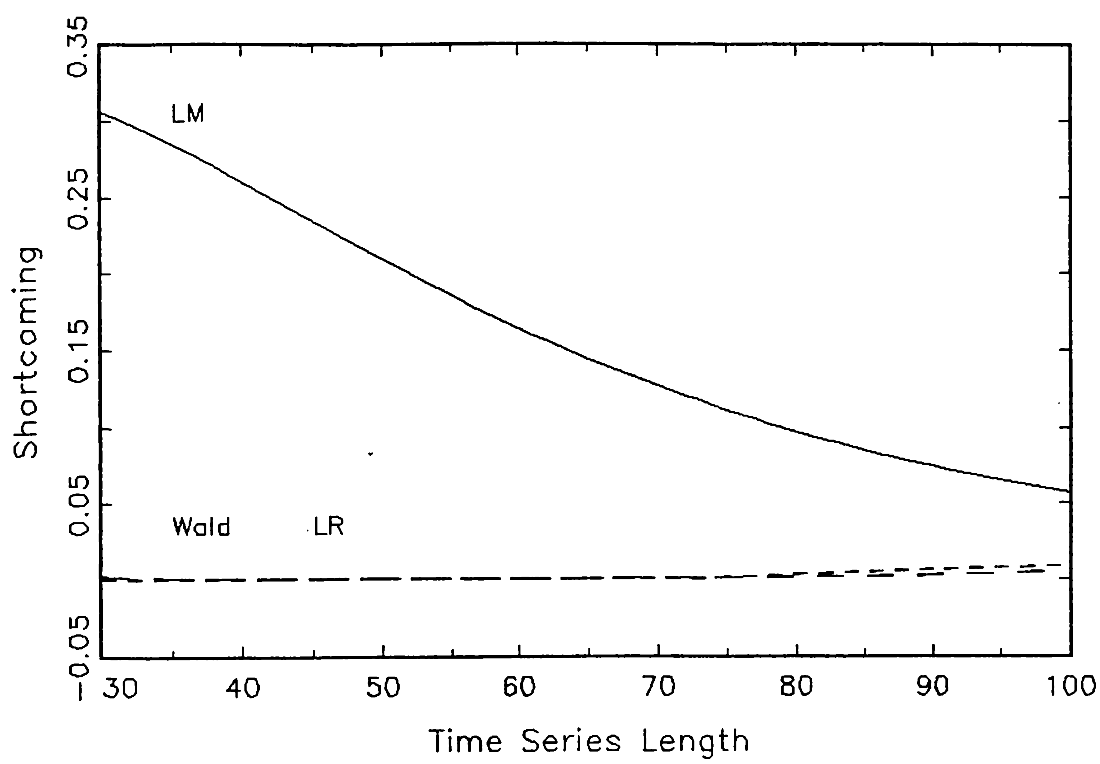

Figure 19: Theoretical Approximate Shortcomings of LM, Wald, and LR

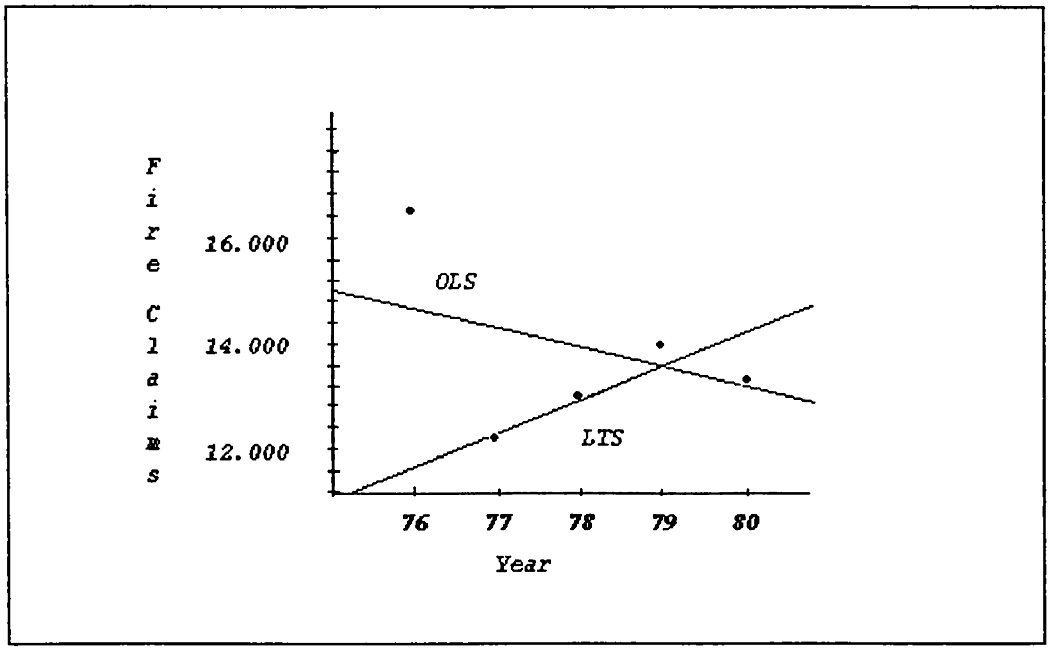

Figure 20: Fire Claims in Belgium