Empirical Likelihood in Econometrics

by

Lauren Bin Dong
B.Sc., Fudan University, 1987
M.Sc., Fudan University, 1990

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

DOCTOR OF PHILOSOPHY

in the Department of Economics

Supervisor: Dr. David E. A. Giles

ABSTRACT

The principal purpose of this dissertation is to develop theoretical approaches to four selected topics using the maximum empirical likelihood (EL) method. These topics are (i) Testing for normality in a pure random data set; (ii) Testing for normality in regressions; (iii) The Behrens-Fisher problem; (iv) Testing for structural change in the coefficients in regressions. Our focus is mainly on the finite sample properties of the empirical likelihood type (EL-type) tests. In particular, we provide a detailed analysis of the sampling properties of the EL-type tests and we conduct comparisons of these properties with those of other commonly used tests in the literature, using Monte Carlo simulations. The conclusion is that the EL-type tests perform at least as well as some of the conventional tests and can be better than the other tests.

Keywords: Empirical Likelihood, Likelihood Ratio, Nonlinear Moment Conditions, Monte Carlo Simulations, Normality, Behrens-Fisher Problem, Structural Change.

Contents

Acknowledgement

This dissertation could not have been completed without the help of many people to whom I would like to express my deep gratitude.

First, I would like to express my deepest appreciation to my thesis supervisor, Professor David Giles. During the two years of working on the dissertation, Professor Giles has provided me with priceless suggestions, constructive criticism, and timely guidance. I am especially grateful for his initial suggestions on the topics and research methods. I would also like to thank Professor Ron Mittelhammer (Washington State University) for his comments and suggestions, and for providing me with some examples of Gauss codes.

Second, I would like to thank the professors in the Department of Economics at University of Victoria for their valuable support. Special thanks are extended to Dr. Graham Voss for his great help on Gauss coding problems and on using the LaTex editing software. I would also like to thank Dr. Don. Ferguson, Mr. Lief Bluck and Dr. van Kooten for their help in providing powerful computing facilities. I would like to acknowledge the financial support from Department of Economics in the form of University of Victoria Fellowships.

Special thanks are also extended to Professor Charles Meadow and Mr. Craig Barlow for their comments and help on proofreading my writings.

Finally, I would like to thank my family members and friends, particularly Lili and Tom, for their loving and prayerful support. Special thanks are extended to my daughter, Leanne, for her understanding and loving support at such a young age.

Chapter 1

Introduction

This dissertation is about applying the maximum empirical likelihood (EL) method to several selected topics in the area of econometrics. In the EL method literature to date, there have been very few research papers that deal with the applications and the theory of the EL method in the area of econometrics, and even fewer papers that relate to the properties of the empirical likelihood type (EL-type) tests in finite samples. The aim of this dissertation is to partially fill this gap. First, we will develop new theoretical approaches and solutions using the EL method for some selected topics. These topics are: (i) testing for normality in pure random data; (ii) testing for normality in the errors of a linear regression model; (iii) the Behrens-Fisher problem; and (iv) the related topic of testing for structural change in the coefficients of a regression model. Second, our main focus will be on testing in the context of the EL approach, rather than on estimation. The detailed sampling properties of EL-type tests for the above problems will be presented using the Monte Carlo simulation method. We will also provide detailed power comparisons for the EL-type tests and other commonly used tests to demonstrate the merits of the EL method.

The empirical likelihood method is a recently developed nonparametric technique for estimation and inference. The empirical likelihood method is very flexible. In forming the likelihood function, it is able to incorporate information from different data sources and knowledge arising from outside of a sample of data. The assumed form of the underlying data distribution is important in constructing a parametric likelihood function. The usual parametric likelihood methods, for example, the maximum likelihood method, give the best estimation and inference for the parameters of interest, at least asymptotically, if the specification of the underlying distribution is correct. The likelihood ratio test and the Wald test can be constructed based on the estimators and distributional assumptions to make useful inferences. A problem with parametric likelihood inference is that we may not know the correct distributional family to use, and there is usually not sufficient information to assume that a data set is from a specific parametric distribution family. The problem of mis-specification can cause likelihood based estimators to be asymptotically inefficient and even inconsistent, and inferences based on the wrongly specified underlying distribution can be completely inappropriate.

Alternative approaches used by researchers include other nonparametric methods, such as the method of moments and the bootstrap. These nonparametric methods provide point estimators, confidence intervals, and inferences that do not depend on strong distributional assumptions. However, each of theses methods has certain limitations, as is detailed in Section 2.2.2.

The empirical likelihood method is able to utilize the concept of the likelihood function without requiring a parametric specification for the underlying distribution of the data. At the same time it utilizes the information available in the form of moment conditions just as some other nonparametric methods do. It is able to bridge the gap between the two extremes, the parametric methods and the other non-parametric methods, and still is able to offer some asymptotic efficiency gain.

The concept of maximum empirical likelihood method and the main theory of this method were developed originally by Owen (1988, 1990, and 1991) in the statistics literature. Additional theoretical and applied contributions associated with the method have been subsequently made by DiCiccio, Hall, and Romano (1991) and Qin and Lawless (1994), and others. Owen (2001) and Mittelhammer et al. (2000) summarized the complete asymptotic properties for the EL estimator. The focus of the EL literature has been mainly on estimation and the coverage accuracy of the associated confidence intervals. For example, the empirical likelihood method is Bartlett correctable (DiCiccio, Hall, and Romano, 1991); the error of the coverage can be reduced from $O(n^{-1})$ to $O(n^{-2})$, *etc.* In our study we will focus on the sampling properties of the empirical likelihood type tests, namely the empirical likelihood ratio (ELR) test and the EL-type Wald test.

We make three contributions to the literature in this dissertation. First, we develop new theoretical approaches to each of the four selected topics using the EL method in the area of econometrics. Second, we provide a complete analysis of the sampling properties of the EL-type tests for a range of situations using Monte Carlo simulations. The analysis includes simulation studies of the actual sizes, the size-adjusted critical values, and the powers of the tests. Third, we present detailed comparisons of the finite sample properties of the EL-type tests and other conventional tests for the problems being considered.

The organization of the dissertation is as follows: Chapter 2 provides a brief review of the empirical likelihood method literature and the properties of the associated estimators and tests. This chapter also touches briefly on a comparison of the EL method with the parametric likelihood method (ML) and the generalized method of moments (GMM) approach. Several new theoretical applications of EL-based testing are then presented. Chapters 3 and 4 develop applications of the EL approach in testing for normality in pure random data sets and in the residuals from a regression model. Chapter 5 provides new theoretical results that use the empirical likelihood method to solve the well known Behrens-Fisher problem. The sampling properties of the empirical likelihood ratio (ELR) test are analyzed in detail across a range of situations. The behavior of the ELR test over the parameter space in the Behrens-Fisher problem is clearly illustrated. Section 6 provides the application of the EL approach to the closely related problem of testing for structural change in the parameters of a linear regression model. Chapter 7 provides some concluding remarks, suggestions and some directions for further research.

Chapter 2

The Empirical Likelihood Method

2.1 Introduction

The maximum empirical likelihood (EL) method is a recently developed nonparametric technique for conducting estimation and hypothesis testing. It is based on a concept known as the empirical likelihood function and the ratios of such functions, as will be defined later in this chapter. The maximum empirical likelihood method was established originally by Owen (1988). Developments of the theory and applications of the method in different areas are provided by Owen (1990 and 1991), Qin and Lawless (1991 and 1994), DiCiccio, Hall, and Romano (1991), and Kitamura (1997). The method has recently attracted some interest in the econometrics literature, but in that field there have been relatively few developments to date.

The purpose of this chapter is to give a brief introduction to the empirical likelihood method and a brief survey of the empirical likelihood method in the research areas that are related to this dissertation.

The outline of this chapter is as follows: Section 2.2 gives a brief introduction of the EL method and the way it is implemented. Section 2.3 discusses the properties of the method and the related literature. Section 2.4 briefly explains the relationship between the EL method and two other commonly used and closely related methods. Section 2.5 provides a short summary.

2.2 The Empirical Likelihood Method

The empirical likelihood approach to statistical estimation and inference is a distribution-free method that still incorporates the notions of the likelihood function and likelihood ratio. The motivation for using the EL method is two-fold. First, the method utilizes the concept of likelihood functions, which is very important. The likelihood method is very flexible. It is able to incorporate the information from different data sources and knowledge arising from outside of the sample of data. The assumption of the underlying data distribution is important in constructing a parametric likelihood function. The usual parametric likelihood methods lead to asymptotically best estimators and asymptotically powerful tests of the parameters if the specification of the underlying distribution is correct. The term "best" means that the estimator has the minimum asymptotic variance. The likelihood ratio test and the Wald test can be constructed based on the estimates and distributional assumptions to make useful inferences. A problem with parametric likelihood inference is that we may not know the correct distributional family to use and there is usually not sufficient information to assume that a data set is from a specific parametric distribution family. Mis-specification can cause likelihood based estimates to be inefficient and inconsistent, and inferences based on the wrongly specified underlying distribution can be completely inappropriate. Using the empirical likelihood method, we are able to avoid mis-specification problems that can be associated with parametric methods.

Second, using the empirical likelihood method enables us to fully employ the information available from the data in an asymptotically efficient way. It is well known that the general method of moments (GMM) approach uses the estimating equations to provide asymptotically efficient estimates for parameters of interest using the information constraints. The empirical likelihood method is able to use the same set of estimating equations together with the empirical likelihood function approach to provide the empirical likelihood estimates for the parameters. The empirical likelihood estimator is obtained in an operationally optimal way and is asymptotically as efficient as the GMM estimator. Details are given in Section 2.4.2.

2.2.1 Data in Hand

Consider a random data set of size n: $y_1, y_2, \ldots, y_n \in R^1$, which are i.i.d. with a common unknown distribution $F_0(y, \theta)$, where θ is a vector of unknown parameters. We assume that the y_i's are random scalars (but the following discussion still applies, with minor adjustments, to i.i.d. random vectors).

For each data point y_i, a probability parameter p_i is assigned, for $i = 1, 2, \ldots, n$. The p_i 's are subject to the usual probability constraints: $0 < p_i < 1$ and $\sum_{i=1}^n p_i = 1$. The empirical likelihood function is simply the product of the p_i's: $\prod_{i=1}^n p_i$.

The maximum empirical likelihood method is to maximize the objective function $\prod_{i=1}^n p_i$ by choosing the p_i's, subject to the probability constraints on the p_i's and unbiased moment constraints in the form of $E(h(y, \theta)) = 0$, where $h(y, \theta)$ is a general $m \times 1$ function of the data vector y and p-element parameter vector, θ.

2.2.2 Information in Hand

The information available is in the form of unbiased moment equations: $E(h(y, \theta)) = 0$, where $h(y, \theta) \in R^m$ is the set of moment functions of the data and the parameter vector θ that are functionally independent. The unbiased moment equations hold true only when evaluated at the true value of the parameter vector, θ_0, where $\theta \in R^p$. An example of the unbiased moment equations is the first moment equation: $Eh(y_i, \mu) = E(y_i - \mu) = 0$, where μ is the parameter for the mean of the underlying population. The moment equation holds true only at the true mean, μ_0.

The empirical analog of the unbiased moment equations has the form:

$$E_p(h(y, \theta)) \equiv \sum_{i=1}^{n} p_i h(y_i, \theta) = 0 \tag{2.1}$$

where p_i is the probability parameter assigned to the ith data point y_i. In our study, the moment equations come naturally from the data; we will match the sample moments with the population moments to construct the moment equations.

Generally, unbiased moment equations provide a nice connection between the data and the parameters of interest, and therefore, the underlying distribution. In economic theories, orthogonality conditions often arise from the optimizing behavior of agents and data structures. These conditions can be treated as unbiased estimating equations and used in the EL approach. That is, side information from possible sources can be incorporated easily into the EL approach.

The EL method requires some mild assumptions. It assumes only the existence of several unbiased moment conditions for the data. There is no requirement for a specific parametric family of distributions for the data. This feature helps the EL method to avoid the misspecification problem that can be encountered with parametric methods. The EL approach also gets around the dimensional limitation problem that is commonly faced by some other nonparametric methods, such as kernel regression. The EL method is able to integrate the likelihood concepts and the unbiased moment equations in an ideal format. It lies between the parametric and nonparametric methods and is able to yield some gain in estimator efficiency.

2.2.3 Objective Function

The objective function is the empirical likelihood function and it has the form: $\prod_{i=1}^{n} p_i$. The log empirical likelihood function is: $\sum_{i=1}^{n} \log p_i$. This objective function is then maximized subject to the moment restrictions. The maximum empirical likelihood approach is a constrained optimization problem and can be set up in the Lagrangian function form:

$$G = n^{-1}\sum_{i=1}^{n}\log p_i - \eta(\sum_{i=1}^{n} p_i - 1) - \lambda' \sum_{i=1}^{n} p_i h(y_i, \theta), \tag{2.2}$$

where $0 < p_i < 1$, and λ and η are the Lagrangian multiplier vectors.

Some manipulations of the first order conditions with respect to η and p_i lead to $\eta = 1$ and

$$p_i = n^{-1}(1 + \lambda' h(y_i, \theta))^{-1}. \tag{2.3}$$

The p_i's are functions of the Lagrangian multiplier vector λ, and the parameter vector. We notice that without the moment constraints the solution to the optimization problem is $p_i = 1/n$, and the maximized value of the empirical likelihood function is n^{-n}. The basic rationale behind the maximum empirical likelihood approach is to modify the weights from

n^{-1} to p_i on each data point such that the moment conditions $E(h(y_i, \theta)) = 0$ are satisfied. In other words, the EL method imposes the moment restrictions by appropriately re-weighting the data.

The maximization problem is high dimensional. The associated computational burden can be reduced by substituting the p_i's back into the Lagrangian function. The refined system involves only the elements of θ as the parameters and the elements of λ as the multipliers. Thus, the maximization problem over the p_i space is transformed into an optimization problem over the λ space of smaller dimension. The vector λ is constrained by the set that is associated with values of θ: $\Lambda(\theta) = \{\lambda : 1 + \lambda' h(y_i,\ \theta) \geq 1/n,\ i = 1,\ \ldots,\ n\}$ due to the constraints on the p_i's. The constraint $\lambda \in \Lambda(\theta)$ bounds the argument of the log function to be within the domain. This constraint has a linear inequality form and it requires the λ vector to be an element of the n open half spaces for each value of θ. Theoretically, this guarantees that a unique solution exists with probability approaching one when the moment conditions are satisfied. In practice, imposing the constraint is problematic and sometime researchers just ignore the constraint. If the constraint is not met, then the moment restrictions are severely violated in the data. In our study, we impose the constraint by checking if the estimated $\hat{p}_i$ lies in the range of $(0,\ 1)$. If the constraint is not satisfied, we alter the initial values of θ and λ and iterate for a valid solution.

The problem is nonlinear in the parameters. There is usually no closed form solution to the EL approach. Numerical methods are required for computing the numerical solutions, and details of the computational issues are discussed in the appendix at the end of this chapter. The numerical solutions to the nonlinear problem are denoted $\hat{\theta}$, $\hat{\lambda}$, and $\hat{p}_i$'s. These solutions provide us with the EL estimators for the parameters and the means to construct empirical likelihood type tests.

2.2.4 Empirical Likelihood Ratio Function

The Empirical Likelihood Ratio function (ELR) is, as the name indicates, usually defined as:

$$R(F) = L(F^c)/L(F^u), \tag{2.4}$$

where $L(F^c) = \prod_{i=1}^{n} p_i^c$ and $L(F^u) = \prod_{i=1}^{n} p_i^u$ are the values of the maximized empirical likelihood functions in the constrained and unconstrained cases. $R(F)$ is a multi-nomial likelihood ratio function that is supported on the distinct data points. A nonparametric version of Wilks' (1938) result holds for the $R(F)$ function, *i.e.* minus twice of the log empirical likelihood ratio is asymptotically χ^2 distributed under some mild conditions (Owen, 1988).

An application of the theory is as follows. Suppose there are j restrictions on the parameter vector θ, $c(\theta) = r$, where the vector r is known with certainty. The null hypothesis is that the constraints are true. The empirical likelihood ratio test statistic has the form of $R(F) = L(F^c)/L(F^u)$. If the null hypothesis is true, minus two times the log empirical likelihood ratio is asymptotically distributed $\chi^2_{(j)}$.

Another application of the theory is when we are interested in testing for the validity of the moment constraints. In this case, $L(\hat{F}) = \prod_{i=1}^{n} \hat{p}_i$ is the maximized value of the empirical likelihood function using the moment constraints. $L(F_n)$ is the maximum value of the likelihood function without the constraints, which equals n^{-n}. Minus two times the log empirical likelihood ratio function has the form:

$$
\begin{aligned}
-2\log R(\hat{F}) &= -2(\log L(\hat{F}) - \log L(F_n)) \\
&= -2\sum_{i=1}^{n} \log np_i \\
&= 2\sum_{i=1}^{n} \log(1 + \hat{\lambda}' h(y_i,\ \hat{\theta})).
\end{aligned}
$$

The log likelihood ratio statistic has a limiting distribution as follows:

$$-2\log R(\hat{F}) \xrightarrow{d} \chi^2_{(m-p)},$$

where m is the number of moment equations and p is the number of parameters, $m \geq p$.

These theoretical results can be used to test various hypotheses and to construct confidence intervals in different models. These theoretical results are the fundamental basis of the work in this dissertation. Here, our focus will be mainly on the properties of the EL type tests in finite samples. The asymptotic properties of the EL type tests have been provided by Owen (1988, 1990, and 1991) and Qin and Lawless (1994) and others. For finite samples, the actual size and the size-adjusted critical values of the tests will be simulated using the Monte Carlo technique. A complete analysis of the power properties of the tests across the full dimension of the parameter space will be presented and detailed power comparisons between different tests will also be provided at the same actual significance levels. One example is testing for normality in Chapters 3 and 4. In Chapter 5, the EL approach is used to provide a new solution to the well known Behrens-Fisher problem, namely testing for the equality of two normal means when the two variances are not known to be equal. This is another interesting application of the results. Details will be provided in the following chapters.

The maximum empirical likelihood method utilizes the concept of likelihood functions and the ratio of these functions. It exploits the information from data in an asymptotically efficient and operationally optimal way as compared with other nonparametric methods. Details will be given in section 2.4.2. Wilks' result in the context of the EL approach enables us to construct tests and confidence intervals in a way that is completely analogous to the ones associated with the standard parametric likelihood method.

2.3 Properties of the Empirical Likelihood Method

2.3.1 Flexibility and Adaptability

The EL method lies between the parametric likelihood methods and other nonparametric methods. It is able to bridge the gap between the two extremes. It requires mininal assumptions about the underlying process that has generated the data. It has the flexibility to incorporate information from different sources using the likelihood function approach and also utilizes the information from the data in the form of unbiased moment equations. It is applicable to various economic models. Qin (1993) applied the EL method to biased samples. Mittelhammer *et al.*(2003) applied the EL approach to the structural equation system when the parameters are over identified. Kitamura (1997) dealt with weakly dependent data processes using the EL method, *etc.* In this section, we will briefly summarize the properties of the EL method and categorize them into two categories: those associated with estimation and those associated with hypothesis testing.

2.3.2 Estimation

In the literature on the EL method, the focus has been mainly on point estimation and confidence region construction. Owen (1988, 1990 and 1991) developed a full theory for the EL method. He showed that the EL estimators have sampling properties that are analogous to those of the parametric maximum likelihood (ML) estimators in large samples. The first order asymptotic properties of the EL estimators are parallel to those for parametric ML. That is, the EL estimator is weakly consistent, asymptotically normal, and asymptotically most efficient within the class of estimators derived from linear combinations of the moment equations.

To ensure these properties of the EL estimator, we require some mild assumptions regarding the existence of certain unbiased moment conditions associated with the functions of the random sample and the parameters of interest. First, the moment function, $h(y,\theta)$, must be at least twice continuously differentiable with respect to the parameters. Second, the moment function itself, and its first and second derivatives must be bounded in a neighborhood of the true value, θ_0, of the parameter vector. Under these conditions, the EL estimate $\hat{\theta}$ has the following asymptotic distribution:

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\Sigma),$$

where

$$\Sigma = [E[\frac{\partial h(y,\theta)}{\partial \theta}\;|_{\theta_0}]E[h(y,\theta)h(y,\theta)'\;|_{\theta_0}]^{-1}E[\frac{\partial h(y,\theta)}{\partial \theta'}\;|_{\theta_0}]]^{-1}$$

The matrix of $E[\frac{\partial h(y,\theta)}{\partial \theta}\;|_{\theta_0}]$ should be of full rank. These conditions are relatively mild. They lead to EL estimators being consistent, asymptotically normal, and asymptotically efficient.

2.3.3 Empirical Likelihood Type of Wald Test

The usual Wald test statistic has a asymptotic distribution of χ^2 under the null hypothesis. For example, when we are interested in testing a set of j linear constraints, $c\theta = r$, the Wald test statistic has the form:

$$W = (c\tilde{\theta}-r)'\Sigma_c^{-1}(c\tilde{\theta}-r), \tag{2.5}$$

where Σ_c is the asymptotic variance-covariance matrix of the vector $(c\tilde{\theta} - r)$, and $\tilde{\theta}$ is the unconstrained maximum likelihood estimator of θ. Σ_c is usually unknown and a consistent estimator of Σ_c can be obtained using the maximum likelihood estimator. The test statistic W has a limiting distribution of $\chi^2_{(j)}$ if the constraints are valid.

Given the knowledge of the asymptotic distribution of the EL estimators, the EL estimator $\hat{\theta}$ of the parameter vector and the consistently estimated variance-covariance matrix of $(c\hat{\theta} - r)$, namely $\hat{\Sigma}_c$, an EL-type of Wald test statistic can be formed as follows:

$$EL_W = (c\hat{\theta} - r)'\hat{\Sigma}_c^{-1}(c\hat{\theta} - r). \tag{2.6}$$

The test statistic also has an asymptotic distribution of $\chi^2_{(j)}$ under the null hypothesis. The EL-type Wald test is useful in testing for the problem of structural change in regression models. Details will be provided in Chapter 6.

2.3.4 Efficiency of EL

EL estimators are operationally optimal and asymptotically efficient for the following two reasons. First, the EL method uses the likelihood approach which lends itself to an operational way of obtaining consistent estimators for the unknown parameters. Second, the EL method utilizes the information of the data in the form of unbiased moment conditions. The EL estimator has the same asymptotic variance-covariance matrix as the consistent and optimal estimators derived from linear combinations of the $m \times 1$ available unbiased moment functions. Therefore, the EL estimator is at least as efficient as the consistent estimator within the class of estimators that are derived from the linear combinations of the unbiased moment equations.

As Mittelhammer *et al.*(2000, page 294) point out:

> "...the EL method utilizes the moment conditions in a most optimal way. The linear combination of the estimation functions used in the EL approach is the best in the sense of defining a consistent estimator with minimum asymptotic covariance matrix in the class of estimators of defined as solutions to the $m \times 1$ transformed moment conditions."

Given the unbiased feature of the moment equations $Eh(y,\ \theta) = 0$, suppose $A(\theta)$ is a $p \times m$ weighting matrix associated with θ such that $h_A(y,\ \theta) \equiv A(\theta)h(y,\ \theta)$ is a $p \times p$ matrix that can be used to solve for θ. We note that $h_A(y,\ \theta)$ is a vector of functions that are linear combinations of the functions in the vector of $h(y, \theta)$, and $h_A(y,\ \theta)$ is unbiased. The asymptotic variance of the estimator from solving $Eh_A(y,\ \theta) = 0$ depends on the choice of $A(\theta)$. The optimal choice of the weighting matrix, $A^*(\theta)$, is the one that gives the estimator of θ, from $Eh_A(y,\ \theta) = 0$, that has the minimum asymptotic variance. Mittelhammer *et al.* (2000) show that the asymptotic covariance matrix of the EL estimator of θ is precisely the same as the asymptotic covariance matrix of the estimator associated with the optimally defined weighting matrix, provided that moment equations are unbiased. Thus, the EL estimator is equivalent, in an asymptotic way, to the most efficient estimator in the class of estimators that are derived from the linear combinations of the unbiased estimating functions.

Solution Uniqueness

There exists a unique solution for the optimum of the empirical likelihood function if the convex hull of $h(y_i, \theta)$ contains zero (Owen, 1988). For instance, for the mean of a distribution F, a unique solution exists, provided that the true value μ_0 is in the convex hull of the data set $y_1,\ y_2,\ \ldots,\ y_n$. This unique solution aspect of the EL method is an application of finding the extremum of a concave function over a convex domain.

In likelihood settings, there can be difficulties in passing from maximizing the log likelihood function to solving $Eh(y, \theta) = 0$. The set of solutions θ may include multiple global maxima and local maxima that are not global maxima. However, for the models with log concave likelihoods, such as the normal distribution and the binomial, the optimization problem does indeed provide the maximum likelihood estimate (Owen 2001, page 213). The EL method is on firmer ground in these cases. In our study of testing for normality, the log likelihood function is strictly concave, the constraint functions are well behaved. These features guarantee the solution uniqueness.

2.3.5 Inference

Statistical inference includes constructing confidence intervals and hypothesis testing. The EL literature has mostly focused on constructing confidence intervals and on ways to improve their coverage inaccuracy.

Increasing the Number of Moment Equations

The coverage accuracy of EL estimators can be improved by making use of more information. Qin and Lawless (1994) pointed out that the asymptotic covariance matrix of the EL estimator generally becomes "smaller" (in a matrix sense) as the number of functionally independent estimating equations on which it is based increases. In practice, this means that the larger the number of correct estimating equations used, generally, the greater the coverage accuracy of the EL confidence intervals become. The term "correct" means the additional moment equations that we put into the system should be unbiased and are functionally independent with each other and with the original ones.

In the context of hypothesis testing, this conjecture means that the larger the number of correctly specified moment equations used, the higher the power of tests based on the empirical likelihood method is likely to be. In Chapter 3, Section 3.5, we will provide empirical evidence to illustrate this feature of the empirical likelihood ratio test in the case of testing for normality. However, there are two aspects we should be aware. One is that there is a potential problem of infeasibility in finite samples in computational practice of the EL method. Given the constraints on the p_i's, a set of over-identified moment equations may not provide a valid solution for θ. The probability of this infeasibility is small. When we increase the number of correctly specified moment equations, this potential may increase. Second, everything comes with a cost. There is a trade-off between increasing the number of moment equations (the number of over-identified equations) and the computational difficulty of iterating for a valid numerical solution to the EL testing problem in finite samples.

Bartlett Correction

The Bartlett correction is a simple and empirical way of adjustment for the expected value of the log likelihood function. It is a way of increasing the coverage accuracy by an order of magnitude. The likelihood ratio test statistic has a limiting distribution of χ^2, under the null hypothesis. Using the χ^2 approximation to the distribution of the likelihood ratio test statistic in finite samples, one way to reduce the coverage error is called the Bartlett correction.

The confidence region of a statistic functional $T(F)$ is of the form:

$$C_{r,n} = \{T(F) \mid R(F) \geq r, F \ll F_n\}, \tag{2.7}$$

where $\text{Prob}(\chi^2_{(d)} \leq -2\log r) = 1 - \alpha$, and $F \ll F_n$ represents discrete empirical distributions that are supported on the data.

DiCiccio, Hall, and Romano (1991) show that the expansion of the log empirical likelihood ratio in terms of n^{-1} has the following form:

$$E[\log R(\theta)] = d(1 + a\, n^{-1}) + O(n^{-2}), \tag{2.8}$$

where d is the degrees of freedom of the limiting distribution $\chi^2_{(d)}$, the coefficient a of the term n^{-1} depends on the parameter θ and on the significance level, α, for the test. A simple adjustment for the expected value of the statistic will remove the term of order n^{-1} from the right-hand side, *i.e.* we correct the confidence region as:

$$C_{r,n} = \{T(F) \mid R(F) \geq r(1 + a\, n^{-1})^{-1}, F \ll F_n\}. \tag{2.9}$$

Then, the accuracy of the coverage of the EL confidence region is improved to $O(n^{-1})$ for a one-sided test and $O(n^{-2})$ for a two-sided test (DiCiccio *et al.* 1991).

The Bartlett correction is based solely on the discrepancies in the mean of the ELR test statistic, $-2\log(R(\theta))$, and takes no explicit account of the variance, the skewness or other higher moments. It is a first order correction. The Bartlett correction was originally established for parametric likelihood ratios. Within nonparametric methods, only the empirical likelihood estimate method is Bartlett correctable. Although we do not utilize the bootstrap method in this paper, one point worth noting is that the bootstrap method is not Bartlett correctable (DiCiccio *et al.*, 1991).

2.4 Comparison of EL With Other Commonly Used Methods

The EL method is relatively new technique for estimation and inference in econometrics. It has certain advantages over other commonly used methods such as the maximum likelihood method (ML) and the general method of moments (GMM) method. Here we briefly provide some comparisons.

2.4.1 EL and ML

John Tukey is alleged to have commented, *"It is better to be approximately right than exactly wrong"*. Usually, we do not have enough information to assume a specific parametric form for the underlying distribution of a data set. The EL approach allows us to pursue the problem by utilizing the most information available without introducing any mis-specification problems.

This idea is the key motivation for using the EL method rather than the parametric likelihood approach. The EL method has the ability to deal with data from different sources, just as the parametric likelihood methods do. The EL approach requires only mild assumptions on the existence of general estimating equations associated with functions of the random sample and the parameters of interest. The EL method can also directly incorporate equality or inequality restrictions on parameters by imposing them as side information in the optimization steps.

In our study we focus on certain zero-valued and functionally independent moment equations associated with functions of the random sample and the parameters of interest. The EL method has the advantage of easily incorporating any side information, as needed, into the approach in the form of moment equations. The EL method has the flexibility to increase or reduce the number of moment conditions that are used. This permits the analyst to deal with multiple pieces of information about an unknown distribution and the parameters of interest, as well as to deal with only those pieces of information that the analyst feels confident about. In later chapters, we will see the details of this and some examples.

The EL method is analogous to the ML method in many aspects. The asymptotic sampling properties of EL estimators and the properties of the ELR function are parallel to those of the ML estimator and the familiar likelihood ratio test. In particular, Wilks' result has a nonparametric version, that is, minus two times the log empirical likelihood ratio has a limiting distribution of χ^2.

2.4.2 Method of Moments and the EL

The Method of moments (MOM) estimator has been very popular in econometrics. It was originally developed by Pearson (1894 and 1902). The generalized method of moments (GMM) estimator is an extension of MOM when we have more moment restrictions than the number of unknown parameters (Hansen, 1982). GMM estimator begins with a set of first order or orthogonality conditions, $E[h(y_i, \theta)] = 0$. Let

$$g(\theta) \equiv \frac{1}{n} \sum_{i=1}^{n} h(y_i,\ \theta) = 0. \tag{2.10}$$

GMM estimation proceeds to a point estimator of θ by choosing $\hat{\theta}_g$ to minimize the quadratic form:

$$g(\theta)' A^{-1} g(\theta),$$

where A is a weighting matrix used whenever the dimension of the $h(y_i, \theta)$ function exceeds the dimension of the parameter vector θ. Hansen (1982) showed that the optimal choice of A is $\mathrm{A}^* = [E[g(\theta)g'(\theta)|_{\theta_0}]]^{-1}$, the inverse of the variance-covariance matrix of $g(\theta)$ evaluated at the true value of the parameter vector θ_0. Under a range of quite weak conditions the GMM estimator $\hat{\theta}_g$ is a consistent estimator of θ_0, and it is asymptotically normal with a limiting distribution:

$$\sqrt{n}(\hat{\theta}_g - \theta_0) \xrightarrow{d} N(0, n^{-1}(G' A^{*-1} G)^{-1}),$$

where $G' = \sum_{i=1}^{n} (\partial h(y_i, \theta)/\partial\theta|_{\theta_0})$. In practice, A^* is unknown. GMM estimation requires a two-step procedure with the first step iterating for the consistent and efficient estimator of A^*. The finite sampling properties of the GMM estimator depend on the estimation of A^*.

The EL method offers an operational way to optimally combine the unbiased moment equations. The common ground of the empirical likelihood method and the GMM method is that the two methods use the same set of over-determined and unbiased moment equations. The GMM method rectifies the over-determined nature of the moment conditions by minimizing the quadratic form of the set of the moment equations. This leads to a linear transformation from the set of over-identified moment equations into a set of just-identified moment equations. The EL method rectifies the nature of over-identification by appropriately reweighing the observations. The estimators of the unknown parameter from the two methods are first order asymptotically equivalent. In contrast to the GMM method, the EL method requires only one step; this is expected to result in improved finite sample properties of the estimators. The first order conditions of the objective function from the GMM approach are a type of moment conditions with the weights being $p_i = 1/n$. That is the GMM method can be considered as a special case of the EL approach.

The solutions using the MOM and the EL approaches are exactly the same when the dimension of the function $h(y_i, \theta)$ equals the dimension of the parameter space. The solution of the method of moments solves the constrained optimization problem of the empirical likelihood approach with weights $p_i = 1/n$. The maximized value of the empirical likelihood function equals n^{-n}. Hereafter, our focus will be on the cases when the number of moment equations is greater than the number of parameters, $m > p$.

The difficulty of using the GMM estimator lies in the area of choosing the appropriate weighting matrix $A(\theta)$. The optimal weighting matrix $A^*(\theta)$ is the one which maximizes the asymptotic efficiency of the GMM estimator. In practice, inference based on the GMM method suffers from poor finite sample properties (Hansen 1996).

Under some weak regularity conditions, and for the given set of unbiased moment equations, the EL estimator is asymptotically equivalent to the efficient GMM estimator within the class of GMM estimators. In addition to this, the EL method makes use of the empirical likelihood function of the data. This offers an increased chance of improved finite sample properties. It provides an operational way to obtain consistent estimators through imposing the moment restrictions by appropriately re-weighting the data. In econometric practice, for small samples, GMM estimators often have larger biases and/or variances relative to the EL estimators (Mittelhammer *et al.* 2003). The EL method is superior to the GMM in small samples in this sense.

The limiting distributions of $\hat{\theta}_{EL}$ and $\hat{\theta}_g$ allow asymptotic hypothesis tests and confidence regions to be constructed. The EL approach has the advantage of providing likelihood ratio statistics upon which tests and confidence intervals can easily be constructed.

2.5 Summary of the EL Method

In this chapter, we have briefly reviewed the maximum empirical likelihood method and the properties of the EL estimator. The EL method is a new nonparametric technique for estimation and inference in statistics. It is starting to draw increasing interest among econometricians. The way that the EL method utilizes the concepts of the likelihood function as well as moment conditions is attractive. It allows us to avoid the mis-specification problem that some parametric methods often incur, and it is superior to those methods in Tukey's sense: "*being approximately right is better than being absolutely wrong*". The EL method is asymptotically efficient; it offers an operational and optimal way to obtain a consistent and efficient estimator within the class of estimators that are derived from linear combinations of the unbiased moment equations.

The EL method is able to be applied to various probability models. It provides the basis for estimation, for hypotheses testing, and for the construction of confidence regions. Mittelhammer *et al.* (2003) have applied the EL-type estimation to structural equations models. Qin and Lawless (1991 and 1994) provided examples when the data are partially generated from biased distributions. Kitamura (1997) provides examples that show that the EL method can be applied to data that are independent but not identical, or that are weakly dependent.

In the following chapters, we develop some new theoretical approaches to the selected topics, and we derive theoretical results for the sampling properties of the EL type tests. A thorough analysis of the sampling properties of the empirical likelihood type tests in a full range of situations is provided through Monte Carlo simulations. Detailed comparisons of these properties of the EL type tests and other commonly used tests are presented. Testing for normality in pure random data sets is considered in Chapter 3. Chapter 4 details a natural extension of the technique to the problem of testing for normality in a regression model. In Chapter 5, we derive a new theoretical solution to the well known Behrens-Fisher problem using the EL method. We provide a unique way to utilize the data sets and the EL function to set up the EL approach for this problem. In Chapter 6, we apply the EL method to the problem of testing for structural change in a regression model. The Behrens-Fisher and the structural change in regression are closely related. The application in Chapter 6 is a natural extension of the technique used in Chapter 5. These four EL approaches in different problems demonstrate the merits of the EL method.

Appendix: Computing Issues of the EL Method

Introduction

As we have mentioned in Chapter 2, generally there is no closed form solution to the optimization problem using the empirical likelihood approach. Numerical methods and algorithms are mostly called for to obtain the empirical likelihood estimates and to conduct tests within this framework.

The maximum empirical likelihood method involves maximizing the empirical likelihood function, subject to the probability constraints and the moment constraints. The objective function to be maximized is concave and the domain is convex; therefore, theoretically, we are able to find the global maximum. This constrained optimization problem can be expressed in terms of the Lagrangian:

$$G = n^{-1}\sum_{i=1}^{n}\log p_i - \eta(\sum_{i=1}^{n} p_i - 1) - \lambda'\sum_{i=1}^{n} p_i h(y_i,\ \theta), \tag{2.11}$$

where $0 < p_i < 1$, and λ and η are the Lagrangian multiplier vectors.

The optimal value of η is unity and the p_i's can be expressed as functions of λ and θ in the following form:

$$p_i = n^{-1}(1 + \lambda' h(y_i,\ \theta))^{-1}. \tag{2.12}$$

Substituting this information back into the Lagrangian function we get an objective function with unknowns of λ and θ. The optimization problem becomes a minimization problem over $\lambda \in R^m$. The new problem is the convex dual of the original constrained maximization problem.

The common structure of the equation system is as follows. The first order conditions of the Lagrangian function with respect to the parameter θ, and the information constraints in the form of the moment equations are:

$$foc(\lambda,\ \theta) \ = \ 0 \tag{2.13}$$

$$\sum_{i=1}^{n} p_i(\lambda,\ \theta)[h(y_i,\ \theta)] \quad = \quad 0. \tag{2.14}$$

At this stage, there are two ways to continue. The first path is to solve the system directly for λ and θ.

- Step 1, The first order necessary conditions for the optimization of the Lagrangian function with respect to the parameter vector θ provide the first group of p equations.
- Step 2, The unbiased moment equations provide the second group of m equations.
- Step 3, A nonlinear equation system of $m + p$ equations is formed by pulling the two groups of equations together. We directly solve the system for the $m + p$ unknowns. The solutions are $\hat{\lambda}$ and $\hat{\theta}$.
- Step 4, Substitute $\hat{\lambda}$ and $\hat{\theta}$ into the formula for the p_i's, and we get the $\hat{p}_i$'s.
- Step 5, The empirical likelihood ratio statistic is formed using the estimated $\hat{p}_i$'s.

This direct approach of solving a nonlinear equation system is used by Mittelhammer *et al.* (2000). It is also the approach we have chosen to use throughout this dissertation. The second possibility is to concentrate out some of the parameters before we solve the system. The steps involved are:

- Step 1, Solve for the Lagrangian multiplier vector $\hat{\lambda}$ first as a numerical function vector of the parameter θ using the set of moment equations:

$$E_p(h(y, \theta)) = 0.$$

- Step 2, Substitute $\hat{\lambda}(\theta)$ into the formula of the p_i's to get the $\hat{p}_i(\hat{\lambda}(\theta),\ \theta)$'s which depend on the parameter θ only.
- Step 3, Substitute $\hat{\lambda}(\theta)$ and the $\hat{p}_i(\hat{\lambda}(\theta),\ \theta)$'s back into the Lagrangian function, and we get a system which involves only the parameter vector θ. Solve the system for $\hat{\theta}$, which is the EL estimate of θ.
- Step 4, Going back step by step, we get $\hat{\lambda}(\hat{\theta})$ and the $\hat{p}_i(\hat{\lambda}(\hat{\theta}),\ \hat{\theta})$'s.

- Step 5, Form the empirical likelihood ratio statistic using the estimated $\hat{p}_i$'s.

The second approach is an example of the *concentrating out the multipliers* method. The advantage of this approach is that we can avoid the saddle point problem that often occurs with constrained optimization problems. Mittelhammer *et al.* (2003) used this approach. Obviously, no matter which way we go, the system is nonlinear in the unknowns.

Numerical Methods and Algorithms

Numerical methods for solving a nonlinear equation system are required by the EL approach. As explained by Mittelhammer *et al.* (2003), for a specific problem, an algorithm is usually chosen according to the features of the problem. Usually it takes a great number of experiments to find out a suitable algorithm for the problem. There is usually no universally suitable algorithm for all types of problems in EL. In the Gauss package (Aptech System, Inc. 2002), the Eqsolve and the NLSYS routines are two useful nonlinear equation solving procedures. These two procedures worked well with the applications in the papers by Mittelhammer *et al.*(2000, 2003). Generally, these two procedures provide similar amount of accuracy for estimation and require a similar computing time in solving the same problem. We have chosen to use the Eqsolve procedure with certain modifications through this entire dissertation.

The Eqsolve procedure is a gradient based method for searching for the global maximum. One feature of this type of method is that the convergence speed is fast. However, it has some limitations: it may not converge to the global maximum; and it may yield multiple solutions. We have two ways to guard for the global maximum. One is that the Eqsolve procedure has a mechanism for checking for a global maximum built in it. Second, the objective function of the EL approach in this dissertation is concave over a convex domain, as explained in Section 2.3.4. These two aspects guarantee that if there is a solution, then the solution of the EL approach is the global maximum.

The Nelder-Mead optimization method, as the name indicated, was proposed by Nelder and Mead (1965). It is not as fast as Newton's method but it is robust and able to find a valid solution when Newton's method has trouble finding the global maximum. The Nelder-Mead method has been used in the empirical approach by Mittelhammer *et al.* (2003). The

Nelder-Mead method is one of the most frequently used deterministic search algorithms. It is effective for many problems, such as the structural equation estimation and inference (Mittelhammer *et al.*, 2003). However, it often performs poorly for difficult optimizations which are nonlinear and have many parameters (Price and Storn, 1997). We have experienced that the Nelder-Mead method does not work for the Behrens-Fisher problem in Chapter 5.

Random search methods, such as the Differential Evolution method and the Genetic Algorithm, for the global maximum are usually very slow, especially when there is a large number of unknown parameters. The computational work in this dissertation is mostly associated with the Monte Carlo simulations. In this context, we have chosen not to use any of the random search methods.

Summary

Generally speaking, the computational work for the applications considered in this dissertation using the empirical likelihood approach is challenging and interesting. The Gauss package does indeed provide good techniques. The empirical results using the Gauss package are satisfactory and meet our expectations.

As part of our future work in the area of computational methods, we would like to explore a broader array of computational approaches that can be used for solving non-linear equation systems in the EL context. These include various numerical techniques and algorithms that may provide us with time-efficient results with fewer practical difficulties associated with convergence.

Chapter 3

Testing for Normality

3.1 Introduction

The empirical likelihood method has a number of attractive properties, as we have seen in Chapter 2. The purpose of this chapter is two-fold: (i) to construct an empirical likelihood ratio (ELR) test of the hypothesis that the underlying population of a sample is normal; and (ii) to undertake a power comparison for the ELR test and four other commonly used tests for this problem. We will illustrate the application of the ELR test for pure random data, and in the next chapter we extend this to the errors of a linear regression model. If the ELR test has well controlled size, and power that is as good as the other tests considered, then, we can say that the ELR is a good test. Our findings show that the ELR test has good power properties and it is invariant with respect to the form of the information constraints. These results are also robust with respect to various changes in the parameters and to the form of the alternative hypothesis. We recommend the use of the ELR test for normality in this context.

One reason why we are interested in testing for normality of the data is that if the data are actually not normally distributed, the maximum likelihood estimator will be distorted because an incorrectly specified likelihood function is used. Other estimators may lack efficiency, at least in finite samples. In addition to this, the usual inference methods based on the assumption of normality, such as the t test and the F test, will in general be distorted. This is a problem of mis-specification. The methods based on the normality assumption may give *asymptotically reliable* inferences for the mean μ of the distribution because of the central limit theorem. If the data have a finite variance, confidence intervals with width based on the estimated variance may still be reliable *asymptotically*, but **not** in finite samples. Thus, there is a great interest in testing for normality in finite samples of pure random data.

Tests for normality are statistical inference procedures designed to test whether the underlying distribution of a random variable is normal. It is commonly known that a normal distribution has skewness coefficient $\alpha_3 = 0$ and kurtosis coefficient $\alpha_4 = 3$. The sample skewness and kurtosis statistics are excellent descriptive and inferential measures for evaluating normality. Any test based on skewness or kurtosis is usually called an omnibus test. An omnibus test is sensitive to various forms of departure from normality.

There are a handful of commonly used tests that can fulfil this testing objective. These include the Jarque-Bera (1980) test (JB), D'Agostino's (1971) D test, and Pearson's (1900) χ^2 goodness of fit test (χ^2 test). These are all omnibus tests. Using them separately gives us the opportunity of testing for departures from normality in different respects.

In this chapter, we will develop an empirical likelihood ratio test (ELR) for normality, and then use Monte Carlo simulations to compare the performance of the ELR test with its competitors. Random data sets are generated using the Gauss package (Aptech System Inc., 2002). For each replication, the same data set is used for all of the tests that we have chosen. The five tests, the ELR, the JB, the D test, the χ^2, and the χ^{2*} (the adjusted χ^2 test to be defined in the next section) are all asymptotic tests. The properties of the tests in finite samples are unknown, although some of them have received some previous consideration. We simulate their actual sizes and calculate their size-adjusted critical values. These results allow us to undertake a power comparison of the tests at the same actual significance levels. One exception is the D test. The actual critical values of the D test are taken from D'Agostino (1971 and 1972). The reason for this is given Section 3.2.3. The results of the experiments are presented in the appendix tables at the end of this chapter.

The outline of this chapter is as follows. Section 3.2 describes the tests that we consider. Section 3.3 gives the Monte Carlo experiments and the results of the tests for pure random data sets. The properties of the ELR test are analyzed in different dimensions. Section 3.4 provides our summary and conclusions.

3.2 Tests: ELR, JB, D, χ^2, and χ^{2*}

Consider a random data set of size n: $y_1, y_2, \ldots, y_n$ which is i.i.d. and has a common distribution $F_0(\theta)$ that is unknown. θ is the parameter vector of the underlying distribution. In the case of testing for normality, it is $\theta = (\mu, \sigma^2)'$. Our interest is to test for normality $H_0 : N(\mu, \sigma^2)$ using the information from the sample. The main focus of this section is to derive an empirical likelihood ratio (ELR) test. We have chosen other four commonly used tests in testing for normality as the competitors. They are the Jarque-Bera test, the D'Agostino's test, Peason's χ^2 goodness of fit test and the χ^{2*} test. The set up of each test is given below.

3.2.1 ELR Test

The empirical likelihood ratio test is based on empirical likelihood functions. First, we assign a probability parameter p_i to each data point y_i and then form the empirical likelihood function $L(F) = \prod_{i=1}^{n} p_i$. The maximum empirical likelihood method is to maximize the likelihood function subject to information constraints. These constraints arise from the data naturally: they are the moment equations and the probability constraints. Let $h(y, \theta)$ be the moment function vector. Under the null hypothesis that the data are from a normal distribution with mean μ and variance σ^2, the first four unbiased empirical moment equations, $E_p(h(y, \theta)) = 0$, have the form:

$$\sum_{i=1}^{n} p_i y_i - \mu = 0 \tag{3.1}$$

$$\sum_{i=1}^{n} p_i y_i^2 - (\mu^2 + \sigma^2) = 0 \tag{3.2}$$

$$\sum_{i=1}^{n} p_i y_i^3 - (\mu^3 + 3\sigma^2\mu) = 0 \tag{3.3}$$

$$\sum_{i=1}^{n} p_i y_i^4 - (\mu^4 + 6\sigma^2\mu^2 + 3\sigma^4) = 0. \tag{3.4}$$

The first term on the left hand of each equation is the sample moment; the second term is the population moment under the null hyposis H_0. We match the two terms to set up the moment equation. We denote this system of equations as $E_p(h(y, \theta)) = 0$. The probability constraints are the usual ones: $0 < p_i < 1$ and $\sum_{i=1}^{n} p_i = 1$.

The reasons that we have chosen to use the first four moment equations are as follows. First, we need at least three moment equations so that the number of moment equations is greater than the number of parameters. Second, we would like to make the various tests comparable. The ELR test should use four moment equations since the JB test uses the standardized first four moments.

We transform the objective function by taking the natural logarithm of the likelihood function. This is an affine transformation and it does not alter the location of the maximum of the objective function. The log empirical likelihood is of the form: $l(F) = \sum_{i=1}^{n} \log p_i$. The constrained optimization problem is then set up in the Lagrangian function form:

$$G = n^{-1} \sum_{i=1}^{n} \log p_i - \eta(\sum_{i=1}^{n} p_i - 1) - \lambda' E_p h(y, \theta). \tag{3.5}$$

The first order conditions of the Lagrangian function with respect to the parameter vector $\theta = (\mu,\ \sigma^2)'$ have the form:

$$\sum_{i=1}^{n} p_i(\lambda_1 + 2\mu\lambda_2 + 3(\mu^2 + \sigma^2)\lambda_3 + 4(\mu^3 + 3\sigma^2\mu)\lambda_4) = 0 \tag{3.6}$$

$$\sum_{i=1}^{n} p_i(\lambda_2 + 3\mu\lambda_3 + 6(\mu^2 + \sigma^2)\lambda_4) = 0, \tag{3.7}$$

where $p_i = n^{-1}(1 + \lambda' E_p(h(y_i,\ \theta)))^{-1}$. As we have seen in Chapter 2, the optimal value for the Lagrangian multiplier η is unity. Substituting the p_i's and η into the Lagrangian function, the original maximization problem over p_i's is transformed into a minimization problem over a smaller number of parameters, namely the elements of the vector λ.

With four moment equations and two first order conditions, the solution $\hat{\theta}$ and $\hat{\lambda}$ can be obtained using the nonlinear equation solver procedure, Eqsolve, in the Gauss package, as was discussed in the Appendix to Chapter 2. The log likelihood function here is log-concave, and the constraint functions are well behaved with positive coefficients associated with parameter terms. Therefore, the conditions for a unique solution are satisfied.

The EL estimator of the parameter vector is $\hat{\theta}$ and the estimated Lagrangian multiplier vector is $\hat{\lambda}$. Substituting these values into the formula for the p_i's, we get the $\hat{p}_i$'s as the estimated probability values for the y_i's. The estimated maximum value of the empirical likelihood function is $L(\hat{F}) = \prod_{i=1}^{n} \hat{p}_i$.

The null hypothesis and the alternative hypothesis for the ELR test are:

$$H_0 : y_i's \sim iidN(\mu, \sigma^2); \qquad H_a : \text{not } H_0.$$

The empirical likelihood ratio function has the form: $R(F) = \frac{L(F)}{L(F_n)}$, where F is the underlying distribution and $L(F_n) = n^{-n}$. Under the null hypothesis, minus two times the log empirical likelihood ratio has the limiting distribution:

$$-2 \log R(F) \xrightarrow{d} \chi^2_{(m-p)}$$

where m is the number of moment equations and p is the number of parameters of interest.

The value of the ELR test statistic based on the values of the restricted and unrestricted empirical likelihood functions is:

$$-2 \log R(\hat{\theta}) = -2 \log(L(\hat{F})/L(F_n)) \tag{3.8}$$

$$= 2 \sum_{i=1}^{n} \log(1 + \hat{\lambda}' h(y_i, \hat{\theta})). \tag{3.9}$$

The ELR test is an asymptotic test. The actual sizes of the ELR test for finite samples are unknown and are therefore computed using Monte Carlo simulations. We reject the null hypothesis when the value of the test statistic is greater than the critical value based on the asymptotic distribution of the test statistic. The total number of the rejections are counted and are divided by the number of replications, which gives us the actual rejection rate. This rejection rate is considered as the actual size of the test for this value of n, given that the number of the replication is large enough, say 10,000. The values of the ELR test statistic are stored and sorted in ascending order so that the percentiles of their empirical distribution can be determined. In this way we can obtain, say, 10%, 5%, 2% and 1%, size-adjusted critical values. In another words the size-adjusted critical values are the values of the test statistic when the actual sizes of the test equal the nominal significance levels. These critical values can then be used to simulate the power of the test in finite samples, by considering various forms of the alternative hypothesis

3.2.2 Jarque-Bera Test (JB)

The JB test was proposed by Jarque and Bera (1980). The JB test is based on the difference between the skewness and kurtosis of the data set $\{y_1,\ y_2,\ \ldots,\ y_n\}$ and those from the assumed normal distribution.

The null hypothesis and the alternative for the JB test are:

$$H_0 : y_i's \sim iidN(\mu, \sigma^2); \qquad H_a : \text{not } H_0.$$

The JB test statistic is:

$$JB = n(\frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24}), \tag{3.10}$$

where

$$\alpha_3 \equiv \frac{n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^3}{s^3} \tag{3.11}$$

$$\alpha_4 \equiv \frac{n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^4}{s^4} \tag{3.12}$$

$$s^2 \equiv n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2. \tag{3.13}$$

Here, $\bar{y}$ is the sample mean, and s^2, α_3 and α_4 are the second, third, and fourth sample moments about the mean, respectively. The JB statistic has an asymptotic distribution which is $\chi^2_{(2)}$ under the null hypothesis.

The JB test is known to have very good power properties in testing for normality; it is clearly easy to compute; and it is commonly used in the regression context in econometrics. One limitation of the test is that it is designed only for testing for normality, while the ELR test can be applied to test for any types of underlying distribution with some modification to the moment equations. As in the context of the ELR test, the Monte Carlo simulation technique is used to determine the size distortion of the JB test in finite samples, and to calculate its size-adjusted critical values which are then used to compute its power.

3.2.3 D'Agostino's Test (D)

The D test was originally proposed by D'Agostino (1971). It has been widely used for testing for normality.

Suppose $y_1, y_2, \ldots, y_n$ is the data set. $y_{1,n}, y_{2,n}, \ldots, y_{n,n}$ are the ordered observations, where $y_{1,n} \leq y_{2,n} \leq \ldots \leq y_{n,n}$. The D test statistic has the form:

$$D = \frac{T}{n^2 s}, \tag{3.14}$$

where s is the sample standard deviation, which is the square root of s^2 as defined in the context of the JB test, and $T = \sum_{i=1}^{n}\{i - \frac{n+1}{2}\}y_{i,n}$. If the sample is drawn from a normal distribution, then

$$E(D) = \frac{(n-1)\Gamma(\frac{n}{2} - \frac{1}{2})}{2\sqrt{2n\pi}\Gamma(\frac{n}{2})} \approx (2\sqrt{\pi})^{-1} \approx 0.28209479. \tag{3.15}$$

The asymptotic standard deviation of the D test statistic is:

$$asd(D) = (\frac{12\sqrt{3} - 37 + 2\pi}{24n\pi})^{\frac{1}{2}} \approx 0.02998598/\sqrt{n}. \tag{3.16}$$

The standardized D test statistic is:

$$D^* = \frac{D - E(D)}{asd(D)}, \tag{3.17}$$

and the null hypothesis and the alternative for the D test are:

$$H_0 : y_i's \sim iidN(\mu, \sigma^2); \qquad H_a : \text{not } H_0.$$

Under the null hypothesis, D^* is asymptotically distributed as $N(0,1)$. If the sample is drawn from a distribution other than normal, $E(D^*)$ tends to differ from zero. If the underlying distribution has greater than normal kurtosis, then, $E(D*) < 0$. If it has less than normal kurtosis, then, $E(D^*) > 0$. So to guard against all the possibilities, the test is a two-sided test.

The percentage points for sample sizes, $n = 30, 50, 70, 100$ are given by D'Agostino

(1972). They were constructed using Pearson curves fitted by moments and extensive simulations. The percentile points for larger sample sizes, $n = 150,\ 200,\ 500,\ 1000$, are provided by D'Agostino (1971) and they are based on Cornish-Fisher expansions. These percentile points were calculated and verified by (D'Agostino, 1972). In our study, instead of simulating critical values, we just use these published values. The D test is an omnibus test in the sense of being able to appropriately detect deviations from normality due either to skewness or to kurtosis.

The Shapiro-Wilks (1965) W test for normality is also known to be a relatively powerful test. The W test is based on the ratio of the best linear unbiased estimator of the population standard deviation to the sample variance. Appropriate weights for the ordered sample observations are needed in computing the numerator of the W test statistic and in computing the percentile points of the null distribution of W for small samples. Each sample size requires a new set of appropriate weights. The W test is also an omnibus test. It has power properties that are superior to those of the chi-squared goodness of fit test in many situations. However, the lately developed D test has power properties that compare favorably with the W test (D'Agostino, 1971). Shapiro and Wilks did not extend their test beyond samples of size 50. D'Agostino (1971) commented on the W test that there are a number of indications that it is best not to make such an extension, although subsequently Royston (1982) did extend the W test for normality to large samples. We have chosen not to include the W test in our study as it is known to have power similar to the D test, while being more difficult to implement computationally.

3.2.4 Pearson's χ^2 Goodness of Fit Test (χ^2)

Pearson's (1900) χ^2 goodness of fit test is the first constructive test in the statistics literature and is a commonly used nonparametric test. It is based on the discrepancies between observed and expected data frequencies. Consider a sample of independent observations of size n, $y_1,\ y_2,\ \ldots, y_n$, with a common distribution $F(y, \theta)$ unknown, where θ is the parameter vector. The null hypothesis is set to be:

$$H_0 : F(y, \theta) = F_0(y, \theta),$$

where F_0 is the distribution function of a particular specified distribution.

In our study, we first transform the data y_i's to be $x_i = \frac{y_i - \bar{\mu}}{\bar{\sigma}}$'s, where $\bar{\mu}$ and $\bar{\sigma}$ are the sample mean and sample deviation. Then, we specify F_0 as $N(0,\ 1)$. The sample of data is then classified into k mutually exclusive categories. The number of the categories, k, and the boundaries of the categories are determined in advance, independently of the data. Let p_{0i} denote the expected probability of an observation falling in the ith category, np_{0i} denote the expected frequencies, and n_i denote the observed frequencies, where $i = 1, 2, \ldots, k$.

The χ^2 test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_{0i})^2}{np_{0i}} \tag{3.18}$$

$$= \frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_{0i}} - n \tag{3.19}$$

and it has a limiting distribution $\chi^2_{(k-3)}$ if the null hypothesis is true.

The number of mutually exclusive categories k is supposed to be arbitrary and independent of the observed data. The asymptotic theory of the χ^2 test is valid no matter how the k categories are determined provided that they are determined without reference to the observations. There are some basic criteria that k should meet, for example $k < n$. Often, an additional restriction is imposed in practice on the choice of k. The resulting intervals should be such that $np_i \geq 5$, for all i. In this case we will denote the test statistic as χ^{2*}. The χ^2 and the χ^{2*} tests may not be applicable when the sample size is very small. Both the ELR test and the χ^2 tests are nonparametric and are applicable with testing for any type of underlying distributions.

3.3 Application to Pure Random Data

This section discusses a full set of Monte Carlo simulations applying the empirical likelihood method to testing for normality of the underlying data distribution. Power comparisons of the ELR test and the other tests discussed above are conducted. Pure random data sets are generated using the pseudo random data generating routine in the Gauss package.

3.3.1 Data Generating Process

The null hypothesis is that the underlying population has a distribution that is $N(\mu, \sigma^2)$. The four alternative distributions that we consider are: Lognormal (LN); $\chi^2_{(2)}$; Student $t_{(5)}$; and the Double Exponential distribution (DE). The four alternative distributions cover a wide range of distributions from symmetric and fat-tailed to skewed distributions. The Log-normal and the $\chi^2_{(2)}$ distributions are skewed to the right. The student $t_{(5)}$ is symmetric and fat-tailed. The Double Exponential is a symmetric and long-tailed distribution. These features are all relative to the normal distribution.

Without loss of generality, all of these distributions are standardized to have mean zero and variance unity. This serves only to fix the true values of the location and scale, possibly both unknown, and does not preclude inferences about those values. This approach is also taken by White and MacDonald (1980), for example.

1. Data for the standardized Lognormal distribution are generated by transforming the standard normal variable $z \sim N(0,1)$ to $y \sim LN(0,1)$:

$$y = \exp(z)/2.161197416 - 0.762873978. \tag{3.20}$$

2. Data for the standardized $\chi^2_{(2)}$ distribution are obtained from two independent standard normal variates: z_1 and z_2:

$$\chi^2_{(2)} = ((z_1^2 + z_2^2) - 2)/2. \tag{3.21}$$

3. Data for the standardized Student $t_{(5)}$ distribution are obtained by:

$$t_{(5)} = \frac{z\sqrt{3/5}}{\sqrt{\chi^2_{(5)}/5}}, \tag{3.22}$$

where $z \sim N(0,1)$, z and $\chi^2_{(5)}$ are independent of each other.

4. Data for the standardized Double Exponential distribution are obtained by:

$$y = \frac{x_1 - x_2}{2\sqrt{2}}, \tag{3.23}$$

where $x_i \sim \chi^2_{(2)}$, $i = 1,\ 2$ are independent of each other.

With the five tests and the four alternatives in hand, a full set of experiment can be conducted. Two particular questions are of interest. First, how do the five tests differ in terms of size distortion in finite samples? Second, how do the powers of the tests compare with each other across all the alternatives, once the size distortion has been taken into account?

3.3.2 Size Distortion

The size of a test is the rejection rate of the test statistic when the null hypothesis is true. All of the five tests are asymptotic tests and the sizes of the tests in finite samples are unknown. The actual sizes of all of the tests, except the D test, for finite samples are simulated and illustrated in Table 3.1. The size distortion is the difference between the actual size of the test and the nominal significance level. The size-adjusted critical values are the values that ensure that the actual sizes of a test equal the nominal significance levels based on the asymptotic distribution of the test statistic. Table 3.1 also provides the size-adjusted critical values for the four tests.

The percentile points, *i.e.* the size-adjusted critical values, for the D test, do not appear in the table but are taken from D'Agostino (1971 and 1972) since these values have been proved and verified to be very accurate for the D test over the years.

The true size of the ELR test is quite large for small samples. For example, the actual size is 34.94% when the nominal significance level is 10%, at $n = 30$. The sizes come down quickly and converge to the correct nominal levels as n increases, as would be expected.

The size of the JB test is much lower than the respective nominal level for small sample sizes. For example, the actual size is about 4.38% when the nominal significance level is 10% for $n = 30$. The sizes converge to the correct nominal levels when n grows.

The size distortion of the χ^2 test is smaller than that of the ELR test for small samples. However, it is worse than that of the ELR test when the sample size grows. In particular, the size distortion does not vanish as the sample size $n \to \infty$. This problem is avoidable if the adjusted chi-square goodness of fit test, the χ^{2*} test, is used. The χ^{2*} test is the χ^2 goodness of fit test adjusted for the expected frequencies in each category to be greater than or equal to five.

The ELR test is an asymptotic test with a limiting distribution of χ^2. The purpose of the Monte Carlo simulation study is to provide the actual distribution for the test statistic in finite samples. The fact that the size distortion of the ELR test is relative large indicates that the approximation of the finite sample distribution in small samples using the asymptotic χ^2 is relatively poor.

Owen (1990) suggested that, for small sample size n, we should replace $\chi^2_{(d)}$ with $\frac{(n-1)d}{(n-d)}$ times $F(d, n-d)$ for a better approximation. This would be very effective to reduce the size of the ELR test. For example, at $n = 30$, the following are the critical values of $\chi^2_{(2)}$ and $(n-1)d/(n-d)F(d, n-d)$, where $d = 2$ and $n = 30$:

α	$\chi^2_{(2)}$	$\frac{29\times 2}{28}F(2,28)$
10%	4.6052	5.1786
5%	5.9915	6.9354
2%	7.8240	9.3484
1%	9.2100	11.3141

However, we did not explore this point in this study.

3.3.3 Power Comparisons

Tables 3.2 to 3.5 give the power comparisons of the five tests for normality across certain alternatives. The five tests are the ELR, the JB, the D test, the χ^2, and the χ^{2*}. The four chosen alternative distributions are the Lognormal; the $\chi^2_{(2)}$; the Student $t(5)$; and the Double Exponential. Each alternative distribution is standardized to have zero mean and unit variance. The null distribution for all the tests is $N(\mu, \sigma^2)$ where the true values for the parameters are $\mu = 0$ and $\sigma^2 = 1$.

In order to conduct the power comparisons, we need to use the same standards for the different tests. The size-adjusted critical values are used for this purpose. We have seen that the five tests have different actual sizes in finite samples. By using the size-adjusted critical values, we are able to compare the power of the five tests at the same actual significance levels, 10%, 5%, 2%, and 1%.

Table 3.2 gives the results when the alternative distribution is Lognormal. The ELR test has the highest power among the tests for significance levels of 5%, 2%, 1% . The power of the JB is in the same range as that of the ELR test, especially for small sample sizes. For example, the power of the ELR test is 93.76% for $n = 30$ and an actual level of 5%, while the JB test has a power of 92.77%. Both the ELR and the JB tests are very powerful for this skewed alternative distribution. The powers of the two tests converge to 100% at $n = 100$. The power of the D test is inferior to that of the ELR and the JB tests for small sample sizes. The χ^2 and the χ^{2*} tests are not applicable for some of the smaller sample sizes. The powers of all of the tests converge to 100% as n grows, though more slowly for the χ^{2*} test than for the other ones.

Table 3.3 gives the results when the alternative distribution is $\chi^2_{(2)}$. The ELR test is the most powerful one among all of the five tests considered for the various sample sizes. The power is 93.07% when $n = 30$, compared with 92.33% for the JB test, when the significance level is 10%. It's power converges to 100% faster than for any of the other tests and it reaches 100% at $n = 50$. Again, the power of the D test and χ^{2*} tests are lower than those of the ELR and the JB tests.

When the alternative distributions are symmetric, as for the Student $t(5)$ and the Double Exponential distributions, all of the tests have quite low power. It is difficult for any test to detect this forms of departure from the normality. Tables 3.4 and 3.5 illustrate thess case. The JB test in this situation is the most powerful test. With a true significance level of 10%, its power is 37.7% and 46.5% against the $St(5)$ and the DE distributions, respectively, when $n = 30$; while the ELR test has a power of 10.6% and 13.2%, respectively, at a true significance level of 10%. The power of the ELR is even lower than that of the D test for small sample sizes. All of the three tests, the ELR, the JB, and the D, have higher power when the sample size reaches $n = 200$. The powers of the three tests are about 100% at $n = 500$. This indicates that the power of the ELR test improves fast over the sample size range of $n = 30$ to $n = 500$ even though it starts low at the small samples for the symmetric alternative distributions.

The relative good power properties of the ELR test result from the ability of the EL method to incorporate the most information available. For instance, in the context of testing for normality, using the first four moment equations, the EL method is able to take into account the information of the sample mean, the variance, the skewness, and the kurtosis. The JB test has the same advantage as the ELR test with four moment equations since the design of the JB test incorporates the standardized third and fourth sample moments. Moreover, the EL method naturally utilizes the likelihood function which may lead to some efficiency gain. Therefore, the ELR test exhibits some attractive features in the application of testing for normality.

To provide some guidance for practitioners in taking the advantage of the good power properties of the ELR test in finite samples, we would suggest that one could use the size-adjusted critical values that we have provided in this study when the values of one's parameters match the values that we have considered. In addition, it would be worthwhile to devote some future efforts to the provision of the size-adjusted critical values for a more extensive range of sample sizes.

3.3.4 Invariance of the ELR Test

In this section, we will show that the ELR test is robust with respect to changes in the functional form of the unbiased moment equations. Instead of using the first four raw moment equations, we consider the first four standardized central moment equations. The data are distributed i.i.d. $N(\mu,\ \sigma^2)$. We standardize the data such that, theoretically, the data will be i.i.d.$N(0,1)$.

The purpose of this section is to illustrate the property of invariance of the ELR test with respect to the form of the moment equations. The transformation from the raw moments to the standardized central moments is a smooth and nonlinear type of transformation in the parameter space. The raw and the standardized central moment equations have the form:

Raw moment conditions	Standardized central moment conditions
$E(y-\mu)=0$	$E(\frac{y-\mu}{\sigma})=0$
$E(y^2-(\mu^2+\sigma^2))=0$	$E(\frac{(y-\mu)^2}{\sigma^2}-1)=0$
$E(y^3-(\mu^3+3\sigma^2\mu))=0$	$E(\frac{(y-\mu)^3}{\sigma^3})=0$
$E(y^4-(\mu^4+6\mu^2\sigma^2+3\sigma^4))=0$	$E(\frac{(y-\mu)^4}{\sigma^4}-3)=0$

Using the standardized central moment conditions places the ELR test on the same basis as the JB test. The JB test uses the skewness and the kurtosis coefficients of the data which are in the form of standardized central moments.

Table 3.6 gives the actual sizes and size-adjusted critical values of the ELR test, and Table 3.7 gives the power of the ELR test using the first four standardized moment conditions about the mean with two unknown parameters. The null distribution is $N(\mu, \sigma^2)$. From Table 3.6, it is easy to see that the range and the pattern of the size distortion of the ELR test using standardized moment conditions about the mean are the same as the ones in the nonstandardized case. The size is approximately 33% at $n = 30$ and converges to the nominal level of 10% as n increases. In Table 3.7, the power of the ELR test is also in the same relative range as it is in the nonstandardized case. It is slightly higher at the lower actual significance levels for asymmetric alternatives and lower for symmetric ones, relatively.

Owen (2001, page 50) discusses the transformation invariance of EL. This relates to the fact that empirical likelihood confidence regions are invariant under one to one parameter transformations and are also invariant under one to one invertible transformations of data.

The empirical evidence of the invariance of the EL method that we have found in this study is that the distribution of the ELR test statistic in finite samples is invariant with respect to the functional form of the moment equations. There is an implicit connection between these two types of invariance of the EL method. The confidence regions and the power of the ELR test are two sides of the same coin. It would be worthwhile to explore the theoretic underpinnings of these findings in a future study. Indeed, a deeper understanding of this issue may also assist in preparing practical guidance for practitioners regarding size adjustment when applying the various EL-based tests.

3.3.5 The ELR Test with Increased Number of Moment Equations

Table 3.8 gives the sizes and the size-adjusted critical values for the ELR test using five moment equations, rather than four. Table 3.9 contains the power comparisons of the ELR tests, the ELR_4 and ELR_5 with the first four and the first five moment equations, and the JB test for small and medium sample sizes. In Tables 3.2 to 3.5, we have seen that the ELR test has very good power properties for large sample sizes over various types of alternative distributions and it is the most powerful test (among those considered) for small sample sizes against skewed alternative distributions. However, it is inferior to the JB test with respect to symmetric alternative distributions such as the $St(5)$ and the Double Exponential distributions. The purpose of this section is to see the improvement in the power of the ELR test that can be achieved by using an increased number of functionally independent moment equations.

Mittelhamer *et al.* (2000) conjectured that the power of the ELR test increases with the number of moment conditions. The purpose of this section is to provide empirical evidence that the power of the ELR test does increase as the number of moment conditions increases. Hopefully the power of the ELR test can be improved in the case of symmetric alternative distributions. However, we should be aware of the following three issues. First, for the $St(5)$ distribution, the integer-order moments exist only up to four at most. Second, there is small probability of the infeasible computational problem in small samples in the EL regime as we have mentioned in Chapter 2. That is the probability of the potential problem may increase as the number of the moment equations increases. Third, the increased degree of over-identification may cause an increase in the computing time for the method.

The null distribution is still the same, namely H_0 : $N(\mu, \sigma^2)$. We illustrate that for small samples, *i.e.* $n = 30$, the power of the ELR test does increase significantly, especially at small significance levels such as 1% and 2%. Table 3.9 shows that the power of the ELR_5 test has increased up to 17% in small sample sizes against skewed distributions, with especially large increases over the low significance levels for each alternative distribution. For example, for the alternative $\chi^2_{(2)}$ distribution, the increment is approximately 17% at the actual significance level of 1% and at $n = 30$, which is quite significant. For the symmetric alternative distribution, the Double Exponential, at $n = 30$ and $\alpha = 1\%$, the increment is about 3%. The results overall are consistent with the conjecture in Mittelhammer *et al.* (2000). Unfortunately, the power of the ELR_5 test is still lower than the JB test in small samples against the alternative distribution of the Double Exponential.

3.4 Computing Issues

The required computing work is efficient in testing for normality. The nonlinear equation solver, the Eqsolve algorithm, works very well in the application considered in this chapter. Each draw from the underlying population is valid in the sense that the sample of data is able to work well for all of the tests: the ELR test, the JB test, the D test, the χ^2 test, and the χ^{2*} test. If a sample draw from the underlying distribution could not provide a valid numerical solution for the ELR test either because the estimated $\hat{p}_i$ is not in the $(0,\ 1)$ range or because the iteration could not converge to provide a valid solution, then, the sample would be thrown away. There are few data sets being thrown away in small samples and no data sets being thrown away when the sample size is greater than fifty. That is, there is no selection bias when using the EL approach in this application.

The computing time in testing for normality is very reasonable. For example, it takes approximately one minute of processing time on a Pentium 4 2.0 GHZ personal computer to conduct a simulation experiment with $10,000$ replications to determine just the empirical size of the ELR test when $n = 30$. It takes about three minutes for $10,000$ replications and all of the five tests when $n = 30$.

3.5 Summary and Conclusions

In this chapter, we have developed an empirical likelihood ratio test to the problem of testing for normality in pure random data cases. Monte Carlo simulations are used to provide the actual sizes and the size-adjusted critical values for the ELR test and for four other tests. These critical values are used in computing the power of each test and conducting power comparisons between the tests. The empirical results provide evidence that the ELR test is a relatively powerful test. It is the most powerful test over asymmetric alternative distributions among all of the five tests considered here. For the symmetric alternative distributions, the power of the ELR test is slightly inferior to that of the JB test. The power of the ELR test can be improved by increasing the number of moment equations we use. The ELR test is invariant to the form of the moment equations. Overall, the ELR test for normality has good power properties, and it is quite easily implemented.

Appendix: Tables in Normality Test

Table 3.1: Size and Size-adjusted Critical Values for the Four Tests: ELR, JB, χ^2, and χ^{2*}

m :	10,000					$H_0 : N(\mu,\ \sigma^2)$					
n :	30	50	70	100	200	250	500	1,000	2,000	5,000	10,000
ELR test at nominal levels:											
10%	0.3494	0.3215	0.2910	0.2668	0.2142	0.2043	0.1650	0.1423	0.1235	0.1129	0.1051
5%	0.2490	0.2203	0.2044	0.1825	0.1438	0.1325	0.1030	0.0819	0.0695	0.0562	0.0554
2%	0.1534	0.1402	0.1251	0.1152	0.0843	0.0738	0.0572	0.0411	0.0297	0.0267	0.0209
1%	0.1049	0.0988	0.0866	0.0792	0.0562	0.0494	0.0360	0.0255	0.0149	0.0141	0.0111
Size-adjusted Critical Values:											
10%	9.3778	9.1638	8.7144	8.3184	7.2353	6.8555	6.0854	5.4348	5.1198	4.8842	4.7309
5%	12.1792	11.6566	11.1896	10.9248	9.5447	9.1718	8.2100	7.2881	6.6690	6.2568	6.1587
2%	15.4973	14.8743	14.4297	14.2330	12.7728	12.3614	11.2863	9.8560	8.6991	8.5000	7.9667
1%	17.7115	17.4968	17.1119	16.8595	15.5300	15.7825	13.6090	11.7415	10.4135	9.8666	9.5787
JB test at nominal levels:											
10%	0.0438	0.0543	0.0569	0.0633	0.0785	0.083	0.0865	0.0881	0.0937	0.0976	0.0984
5%	0.0294	0.0353	0.0366	0.0390	0.0472	0.0467	0.0475	0.0455	0.0455	0.0484	0.0491
2%	0.0192	0.0229	0.0229	0.0246	0.0268	0.0278	0.0221	0.0231	0.0196	0.0189	0.0186
1%	0.0147	0.0165	0.0176	0.0183	0.0191	0.0194	0.0147	0.0149	0.0110	0.0102	0.0106
Size-adjusted Critical Values:											
10%	2.7415	3.1072	3.3437	3.5628	4.0822	4.164	4.3016	4.3573	4.5162	4.5533	4.5620
5%	4.2229	4.8305	4.9965	5.2076	5.7969	5.8276	5.8854	5.7599	5.7988	5.9441	5.9562
2%	7.6184	8.3630	8.7360	8.8227	8.9930	9.1034	8.0600	8.2455	7.7146	7.7384	7.6884
1%	11.2731	12.4052	12.4829	12.6074	11.8813	12.5295	10.7754	10.4652	9.3900	9.2140	9.3332
χ^2 goodness of fit test at nominal levels:											
10%	0.1320	0.1115	0.1022	0.1071	0.1239	0.1324	0.1357	0.1366	0.1398	0.1361	0.1267
5%	0.0605	0.0596	0.0537	0.0654	0.0768	0.0837	0.0851	0.0833	0.0863	0.0866	0.0788
2%	0.0214	0.0262	0.0267	0.0355	0.0461	0.0514	0.0479	0.0487	0.0510	0.0504	0.0454
1%	0.0105	0.0149	0.0162	0.0241	0.0344	0.0371	0.0350	0.0352	0.0359	0.0361	0.0322
Size-adjusted Critical Values:											
10%	5.0784	9.3466	12.9873	19.0312	37.3835	45.7889	55.8576	57.7062	58.8775	59.5980	59.4830
5%	6.2818	11.3592	15.2381	22.3365	42.1121	51.1546	61.9765	63.2299	64.3628	65.5994	65.5305
2%	7.8279	13.6176	18.4865	26.2892	50.2357	59.7278	71.3387	73.1178	73.5503	75.5200	74.6156
1%	9.1078	15.4547	20.8135	30.9140	63.5271	74.2101	85.9508	88.8786	90.9052	95.0253	87.8157
χ^{2*} goodness of fit test at nominal levels:											
10%	–	0.1693	0.1316	0.1187	0.1079	0.1104	0.1082	0.1045	0.1064	0.1012	0.0991
5%	–	0.0888	0.0671	0.0616	0.0568	0.0552	0.0541	0.0538	0.0547	0.0519	0.0501
2%	–	0.0410	0.0279	0.0267	0.0230	0.0231	0.0209	0.0213	0.0210	0.0224	0.0207
1%	–	0.0212	0.0137	0.0132	0.0117	0.0115	0.0096	0.0115	0.0105	0.0113	0.0111
Size-adjusted Critical Values:											
10%	–	7.4047	9.1804	12.4743	22.8229	27.7303	38.1887	43.0001	46.4041	49.2212	50.7836
5%	–	9.3701	11.1657	14.6779	25.5606	30.8296	41.6029	46.7956	50.3262	53.4114	55.1265
2%	–	11.6226	13.6931	17.5926	29.0460	34.3290	45.7909	51.4873	54.6466	58.5462	60.1098
1%	–	13.2294	15.4926	19.5909	31.2562	37.3967	48.6595	55.1539	58.1896	62.3658	63.3077

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 3.2: Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			$H_a: Lognormal(0,1)$			
n	30	50	70	100	150	200	250
ELR test:							
10%	0.9811	0.9999	1	1	1	1	1
5%	0.9376	0.9991	1	1	1	1	1
2%	0.8368	0.9966	0.9999	1	1	1	1
1%	0.7212	0.9897	0.9997	1	1	1	1
JB test:							
10%	0.9854	0.9995	1	1	1	1	1
5%	0.9277	0.9970	1	1	1	1	1
2%	0.8081	0.9642	0.9976	1	1	1	1
1%	0.7037	0.9088	0.9878	0.9998	1	1	1
D test:							
10%	0.8905	0.9761	0.9963	0.9999	1	1	1
5%	0.8374	0.9592	0.9931	0.9996	1	1	1
2%	0.7579	0.9300	0.9852	0.9990	1	1	1
1%	0.6982	0.9016	0.9771	0.9979	1	1	1
χ^2 *test:*							
10%	–	–	0.8887	0.9662	0.9984	0.9997	1
5%	–	–	0.8434	0.9498	0.9973	0.9997	1
2%	–	–	0.7809	0.9252	0.9938	0.9994	0.9998
1%	–	–	0.7072	0.8957	0.9894	0.9980	0.9995
χ^{2*} *test:*							
10%	–	–	–	0.8991	0.9962	0.9997	1
5%	–	–	–	0.8236	0.9909	0.9997	1
2%	–	–	–	0.7041	0.9784	0.9990	1
1%	–	–	–	0.6178	0.9642	0.9984	1

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 3.3:Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			$H_a : \chi^2_{(2)}(0,1)$			
n	30	50	70	100	150	200	250
ELR test:							
10%	0.9307	0.9975	1	1	1	1	1
5%	0.8295	0.9929	0.9998	1	1	1	1
2%	0.6411	0.9734	0.9986	1	1	1	1
1%	0.4767	0.9383	0.9970	1	1	1	1
JB test:							
10%	0.9233	0.9974	0.9999	1	1	1	1
5%	0.7475	0.9661	0.9975	1	1	1	1
2%	0.5535	0.8240	0.9599	0.9990	1	1	1
1%	0.4331	0.6909	0.8824	0.9867	1	1	1
D test:							
10%	0.6536	0.8621	0.9467	0.9888	0.9995	0.9999	1
5%	0.5533	0.7941	0.9150	0.9775	0.9984	0.9997	1
2%	0.4480	0.6997	0.8561	0.9579	0.9957	0.9995	0.9999
1%	0.3819	0.6265	0.8021	0.9352	0.9929	0.9991	0.9999
χ^2 *test:*							
10%	–	0.6111	0.7521	0.8922	0.9830	0.9974	0.9998
5%	–	0.4812	0.6569	0.8364	0.9748	0.9960	0.9996
2%	–	0.3615	0.5310	0.7573	0.9497	0.9914	0.9988
1%	–	0.2949	0.4580	0.6649	0.9183	0.9771	0.9959
χ^{2*} *test:*							
10%	–	0.2669	0.4057	0.6569	0.9411	0.9927	0.9989
5%	–	0.1349	0.2537	0.5204	0.8874	0.9831	0.9975
2%	–	0.0637	0.1390	0.3640	0.8088	0.9659	0.9959
1%	–	0.0390	0.0927	0.2743	0.7345	0.9450	0.9925

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 3.4:Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			H_a : Student $t_{(5)}(0,1)$					
n	30	50	70	100	150	200	250	500	1,000
ELR test:									
10%	0.1060	0.1635	0.2394	0.3746	0.5818	0.7506	0.8502	0.9924	1
5%	0.0483	0.0874	0.1375	0.2464	0.4307	0.6346	0.7675	0.9829	1
2%	0.0208	0.0393	0.0659	0.1347	0.2752	0.4831	0.6330	0.9580	1
1%	0.0097	0.0196	0.0364	0.0786	0.1876	0.3704	0.4858	0.9303	0.9997
JB test:									
10%	0.3767	0.5033	0.6050	0.7182	0.8262	0.9037	0.9419	0.9972	1
5%	0.2913	0.4167	0.5212	0.6501	0.7801	0.8678	0.9183	0.9947	1
2%	0.2071	0.3222	0.4198	0.5454	0.6963	0.8017	0.8655	0.9901	1
1%	0.1575	0.2605	0.3501	0.4686	0.6257	0.7532	0.8191	0.9822	1
D test:									
10%	0.3157	0.4594	0.5745	0.7099	0.8369	0.9121	0.9563	0.9981	1
5%	0.2342	0.3663	0.4868	0.6314	0.7762	0.8710	0.929	0.9963	1
2%	0.1615	0.2753	0.3852	0.5330	0.6928	0.8080	0.8894	0.9906	1
1%	0.1251	0.2267	0.3224	0.4656	0.6334	0.7574	0.8541	0.9853	1
χ^2 *test:*									
10%	0.1885	0.2863	0.3758	0.4695	0.5853	0.6889	0.7463	0.9494	0.9990
5%	0.1220	0.2133	0.3040	0.3932	0.5230	0.6270	0.6879	0.9268	0.9983
2%	0.0774	0.1621	0.2396	0.3317	0.4465	0.5474	0.6225	0.8904	0.9956
1%	0.0604	0.1393	0.2010	0.2854	0.4033	0.4711	0.5461	0.8365	0.9880
χ^{2*} *test:*									
10%	–	0.1407	0.1671	0.1864	0.2148	0.2282	0.2308	0.3559	0.8340
5%	–	0.0735	0.0985	0.1089	0.1271	0.1460	0.1462	0.2532	0.7535
2%	–	0.0343	0.0477	0.0562	0.0697	0.0785	0.0844	0.1577	0.6365
1%	–	0.0229	0.0265	0.0367	0.0441	0.0531	0.0509	0.1121	0.5374

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 3.5:Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			H_a : *DoubleExponential*(0, 1)					
n	30	50	70	100	150	200	250	500	1,000
ELR test:									
10%	0.1318	0.2163	0.3484	0.5824	0.8286	0.9486	0.9843	1	1
5%	0.0720	0.1277	0.2222	0.4207	0.7102	0.8974	0.9650	0.9996	1
2%	0.0337	0.0665	0.1139	0.2636	0.5260	0.7986	0.9157	0.9992	1
1%	0.0188	0.0397	0.0635	0.1670	0.4147	0.6953	0.8371	0.9986	1
JB test:									
10%	0.4648	0.6266	0.7359	0.8541	0.9433	0.9805	0.9938	0.9998	1
5%	0.3620	0.5281	0.6541	0.7888	0.9126	0.9660	0.9895	0.9997	1
2%	0.2562	0.4059	0.5162	0.6744	0.8470	0.9326	0.9719	0.9994	1
1%	0.1866	0.3196	0.4279	0.5800	0.7764	0.8928	0.9458	0.9993	1
D test:									
10%	0.4540	0.6648	0.7972	0.9079	0.9773	0.9948	0.9990	1	1
5%	0.3418	0.5551	0.7128	0.8603	0.9612	0.9897	0.9977	1	1
2%	0.2367	0.4329	0.5968	0.7796	0.9258	0.9772	0.9945	1	1
1%	0.1784	0.3546	0.5125	0.7176	0.8923	0.9631	0.9906	1	1
χ^2 *test:*									
10%	0.2851	0.4421	0.5694	0.6980	0.8390	0.9124	0.9478	0.9993	1
5%	0.1956	0.3373	0.4747	0.5995	0.7688	0.8596	0.9154	0.9980	1
2%	0.1260	0.2504	0.3628	0.5002	0.6564	0.7621	0.8493	0.9925	1
1%	0.0950	0.2039	0.3040	0.4080	0.5739	0.6155	0.7289	0.9766	1
χ^{2*} *test:*									
10%	–	0.2588	0.3601	0.4755	0.6180	0.7253	0.7929	0.9816	1
5%	–	0.1483	0.2341	0.3432	0.4813	0.6094	0.6872	0.9610	1
2%	–	0.0758	0.1311	0.2070	0.3408	0.4614	0.5588	0.9218	0.9999
1%	–	0.0465	0.0832	0.1442	0.2464	0.3689	0.4488	0.8856	0.9999

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 3.6: Size and Size-adjusted Critical Values of the ELR Test Using Standardized Central Moment Equations

m	10,000			$H_0: N(\mu, \sigma^2)$							
n	30	50	70	100	150	200	500	1,000	2,000	5,000	10,000
ELR test at nominal levels:											
10%	0.3321	0.2852	0.2579	0.2399	0.2388	0.1904	0.1497	0.1273	0.1153	0.1113	0.1057
5%	0.2338	0.2004	0.1792	0.1625	0.1574	0.1162	0.0880	0.0729	0.0616	0.0563	0.0526
2%	0.1666	0.1393	0.1239	0.1162	0.1092	0.0789	0.0528	0.0411	0.0347	0.0284	0.0260
1%	0.1054	0.0855	0.0767	0.0711	0.0657	0.0462	0.0297	0.0205	0.0156	0.0111	0.0104
Size and Size-adjusted critical values:											
10%	9.43	8.55	8.17	7.95	7.71	6.48	5.64	5.15	4.87	4.82	4.72
5%	12.02	11.21	11.09	10.51	10.20	8.95	7.56	6.88	6.53	6.19	6.09
2%	15.71	14.39	14.74	14.15	13.49	12.19	10.27	9.28	8.67	8.04	7.87
1%	18.44	16.72	17.46	16.95	16.62	14.45	12.46	11.31	10.05	9.46	9.27

Notes to table: The data is standardized to be $x_i = (y_i - \mu)/\sigma$, for $i = 1, 2, \ldots, n$. The true value of $\theta = (\mu, \sigma^2)'$ is $(0, 1)'$. m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$.

Table 3.7: Power of The ELR Test Using Standardized Central Moment Equations

m	10,000			$H_0 : N(\mu, \sigma^2)$				
n	30	50	70	100	150	200	250	500
H_a : *Lognormal*								
10%	0.9787	1	1					
5%	0.9429	0.9991	1					
2%	0.8454	0.9958	1					
1%	0.7308	0.9916	0.9998					
$H_a : \chi^2_{(2)}$								
10%	0.9440	0.9984	1					
5%	0.8689	0.9945	1					
2%	0.7061	0.9802	0.9992					
1%	0.5634	0.9600	0.9975					
H_a : *Student* $t_{(5)}$								
10%	0.0759	0.1163	0.1850	0.3147	0.5781	0.7596	0.8917	0.9953
5%	0.0398	0.0593	0.0835	0.1862	0.4336	0.6283	0.8254	0.9874
2%	0.0168	0.0269	0.0342	0.0782	0.2814	0.4646	0.7203	0.9679
1%	0.0090	0.0157	0.0194	0.0409	0.1793	0.3628	0.6241	0.9442
H_a : *Double Exponential*								
10%	0.0876	0.1939	0.3501	0.5758	0.8399	0.9525	0.9898	0.9999
5%	0.0437	0.0996	0.1870	0.4116	0.7295	0.8999	0.9776	0.9999
2%	0.0188	0.0431	0.0843	0.2252	0.5703	0.8026	0.9450	0.9997
1%	0.0116	0.0255	0.0474	0.1273	0.4194	0.7147	0.9072	0.9993

Notes to table: The data is standardized to be $x_i = (y_i - \mu)/\sigma$, for $i = 1, 2, \ldots, n$. m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$.

Table 3.8: Size of the ELR Test with Five Moment Equations

m	10,000			$H_0 : N(\mu, \sigma^2)$				
n	30	50	70	100	150	200	250	500
ELR test at nominal levels:								
10%	0.4301	0.4216	0.3993	0.3817	0.3478	0.3283	0.2999	0.2546
5%	0.3349	0.3259	0.3099	0.2877	0.2655	0.2437	0.2156	0.1797
2%	0.2325	0.2345	0.2172	0.1976	0.1820	0.1631	0.1455	0.1139
1%	0.1786	0.1750	0.1641	0.1485	0.1383	0.1246	0.1076	0.0817
Size-adjusted Critical Values:								
10%	14.3394	14.2768	14.1680	13.7818	13.2607	12.6498	11.7367	10.4588
5%	18.3187	17.8126	17.6145	17.3194	17.0302	16.5965	15.4773	13.5068
2%	24.2292	22.5592	22.6917	22.1746	21.8368	21.8703	20.2833	18.1422
1%	30.3333	26.8072	25.9431	25.7542	25.2727	25.6406	24.2938	22.2712

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$.

Table 3.9: Power of the ELR Test Using Five Moment Equations

m	10,000		$H_0 : N(\mu, \sigma^2)$			
n	30	50			30	50
	H_a : *Lognormal*				$H_a : \chi^2_{(2)}$	
ELR_4 test:						
10%	0.9811	0.9999			0.9307	0.9975
5%	0.9376	0.9991			0.8295	0.9929
2%	0.8368	0.9966			0.6411	0.9734
1%	0.7212	0.9897			0.4767	0.9383
ELR_5 test:						
10%	0.9874	0.9965			0.9751	0.9997
5%	0.9642	0.9944			0.9352	0.9984
2%	0.9006	0.9868			0.8114	0.9950
1%	0.7810	0.9770			0.6465	0.9874
JB test:						
10%	0.9854	0.9995			0.9233	0.9974
5%	0.9277	0.9970			0.7475	0.9661
2%	0.8081	0.9642			0.5535	0.8240
1%	0.7037	0.9088			0.4331	0.6909

n	30	50	70	100	150	200	250	500
		H_a : *Double Exponential*						
ELR_4 test:								
10%	0.1318	0.2163	0.3484	0.5824	0.8286	0.9486	0.9843	1
5%	0.0720	0.1277	0.2222	0.4207	0.7102	0.8974	0.9650	0.9996
2%	0.0337	0.0665	0.1139	0.2636	0.5260	0.7986	0.9157	0.9992
1%	0.0188	0.0397	0.0635	0.1670	0.4147	0.6953	0.8371	0.9986
ELR_5 test:								
10%	0.1692	0.1659	0.2132	0.3582	0.6445	0.8398	0.9389	0.9998
5%	0.1025	0.0992	0.1219	0.2130	0.4672	0.7030	0.8599	0.9996
2%	0.0525	0.0513	0.0564	0.0956	0.2655	0.4887	0.7169	0.9959
1%	0.0303	0.0317	0.0352	0.0533	0.1640	0.3475	0.5824	0.9886
JB test:								
10%	0.4648	0.6266	0.7359	0.8541	0.9433	0.9805	0.9938	0.9998
5%	0.3620	0.5281	0.6541	0.7888	0.9126	0.9660	0.9895	0.9997
2%	0.2562	0.4059	0.5162	0.6744	0.8470	0.9326	0.9719	0.9994
1%	0.1866	0.3196	0.4279	0.5800	0.7764	0.8928	0.9458	0.9993

Notes to table: ELR_4 and ELR_5 are the ELR test with four and five moment equations. The degrees of freedom of the ELR_5 test is 3. The alternative of student $t_(5)$ is not applicable. m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$.

Chapter 4

Normality Testing in OLS Residuals

4.1 Introduction

In the classical regression model:

$$y = X\beta + \varepsilon, \tag{4.1}$$

we usually assume that the error term is normally distributed, $\varepsilon \sim N(0,\ \sigma^2 I)$, and it is not correlated with the regressors x. The regressor matrix X is non-stochastic and of full rank. Under these assumptions, the OLS estimator is the best unbiased estimator. It is also the maximum likelihood estimator. Thus, we can apply the usual t test and the F test, for linear restrictions on the parameters and we can make other useful inferences.

A test for the normality of the error term is essential. If the underlying distribution of the errors is actually not normal, the various tests based on the normality assumption do provide *asymptotically reliable* inferences about the parameter vector β because of the central limit theorem. However, in finite samples, the estimated covariance matrix of the coefficient estimators of the regression model would be inefficient and inferences based on the usual t test and the F test would be inappropriate. This is a so-called mis-specification problem.

Since the error term of the regression model, ε, is unobservable, we usually use the residuals $\hat{\varepsilon}$ from the linear regression model to replace the random error term for testing purposes. The main purpose of this chapter is to construct a new test for the normality of the error terms in an ordinary multiple linear regression model. This new test is an ELR test.

The literature on testing for normality in this context is vast. For example, see D'Agostino (1971 and 1972), Bera and Jarque (1980 and 1981), White and MacDonald (1980), and Huang and Bolch (1974). These papers contain details of alternative tests, further references to the literature, and Monte Carlo evidence regarding the relative performance of the tests. We have chosen four of these existing tests to conduct a power comparison analysis in this chapter.

We use the empirical likelihood method, and specifically the empirical likelihood ratio test, to develop a new test for the normality of the error term in the regression model. Monte Carlo simulations are then conducted to show the sampling properties of the ELR test. We conduct a power comparison for the ELR test with the four commonly used tests to parallel the results with those in Chapter 3.

The outline of this chaper is as follows. Section 4.2 sets up the model and derives the ELR test. The regression residuals are introduced and the differences between the BLUS residuals, which will be defined later in the chapter, and the OLS residuals are discussed. Section 4.3 illustrates the Monte Carlo simulations for the five tests. Here random data sets for the error terms are generated using the Gauss package for the null hypothesis and alternative distributions. The experimental results are presented in the Appendix at the end of this chapter. The discussion of the results is provided in Section 4.3.2. The last section 4.4 provides the summary and some conclusions.

4.2 The Model

The classical linear regression model has the form:

$$Y = X\beta + \varepsilon, \tag{4.2}$$

where Y is a $n \times 1$ vector of observed values of the random dependent variable, X is a known $n \times k$ regressor matrix of rank k, β is the $k \times 1$ vector of unknown parameters, and the error term ε is a vector of unobservable stochastic disturbances, assumed to be normally distributed with mean zero and a covariance matrix of a scalar times the identity matrix. That is $\varepsilon \sim N(0,\ \sigma^2 I)$.

The independent variables in the linear regression model are assumed to be "fixed in repeated samples". In the Monte Carlo experiment described in section 4.3, we achieve this by drawing these x variables independently from a uniform distribution and holding the regressor matrix X constant for each replication of the experiment based upon a given sample size.

4.2.1 OLS and BLUS residuals

The residual vector from the regression model (OLS) is a linear transformation of the error vector. It has the form:

$$\hat{\varepsilon} = M\varepsilon, \tag{4.3}$$

and has a distribution $\hat{\varepsilon} \sim N(0, \sigma^2 M)$, where $M = I - X(X'X)^{-1}X'$ is a $n \times n$ idempotent symmetric matrix with a rank of $n - k$. The coefficient vector β does not appear in this expression. Therefore, we have no need to consider the values and estimation of the coefficients. The covariance matrix of the residual vector is $\sigma^2 M$, which is not a diagonal matrix and is singular. Therefore, the elements of the residual vector are not independently distributed, *i.e.* they are not i.i.d..

Some researchers (*e.g.*,Huang and Bolch, 1974) prefer to use Theil's (1965, 1968) Best Linear Unbiased Scalar (BLUS) residuals to test for normality in the regression model. The BLUS residual vector is a linear transformation of the OLS residual vector. The BLUS residual vector, ε^*, is obtained from:

$$\varepsilon^* = A'\varepsilon, \tag{4.4}$$

where A is a $n \times (n-k)$ matrix and it is in the null space of X, *i.e.* $X'A = 0$, and $A'A = I_{n-k}$. We note that:

$$E(\varepsilon^*) = 0 \quad \text{and} \quad Var(\varepsilon^*) = \sigma^2 I_{n-k}. \tag{4.5}$$

The covariance matrix of the BLUS residual vector is diagonal and is of full rank. The BLUS residual vector is distributed $N(0,\ \sigma^2 I_{n-k})$ when the error term is from a normal distribution.

Huang and Bolch (1974) proved that *theoretically both the BLUS and the OLS residuals suffer from the common problem of lack of independence under the alternative hypothesis of non-normal disturbances.* The BLUS residuals are independent *if and only if* the error term is independent and normally distributed. Comparing the OLS and the BLUS residuals from the viewpoint of testing for normality, the OLS residuals are at least as good as the BLUS residuals when the underlying distribution is not normally distributed. This is relevant from the prospective of power considerations. Huang and Bolch (1974) report on Monte Carlo studies where the least squares residual vector $\hat{\varepsilon}$ led to a more powerful test than that obtained by using the BLUS residual vector ε^*. Thus, we focus on the least squares residuals from here on.

An additional problem in testing for normality in a regression model is that the probability distributions of the OLS residuals are always closer to the normal form than is the probability distribution of the disturbance, if the disturbance is not normal. White and McDonald (1980) show that the skewness (positive or negative) and the kurtosis of the OLS residuals will never exceed the skewness and kurtosis of the disturbance term. The residuals for small samples appear more normal than would the error term ε. This is called super-normality. Any test for normality using the residuals is more likely to fail in rejecting the null hypothesis when the null is false, than would be the case by using the error term itself (if this were in fact possible) in the construction of the test.

4.2.2 ELR Test

Consider the least squares residual vector $\hat{\varepsilon}$ derived from the regression model $Y = X\beta + \varepsilon$, where the disturbance term ε has an unknown distribution with mean zero. Our interest is to test for normality of the error term using the least squares residuals. The null hypothesis is H_0 : $\varepsilon_i's \sim iidN(0,\ \sigma^2)$, where $i = 1,\ 2,\ \ldots,\ n$. The corresponding $\hat{\varepsilon}_i$'s are used in the construction of the test.

For each $\hat{\varepsilon}_i$, in order to form the ELR test, a probability parameter p_i is assigned to it. The empirical likelihood function is $\prod_{i=1}^{n} p_i$. In order to detect any type of departure from normality, we take into account the third and the fourth moments of the residuals. The first four unbiased moment equations are in the following form:

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i = 0 \tag{4.6}$$

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i^2 - \sigma^2 = 0 \tag{4.7}$$

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i^3 = 0 \tag{4.8}$$

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i^4 - 3\sigma^4 = 0. \tag{4.9}$$

As in Chapter 3, we denote the moment equations as $E_p(h(\hat{\varepsilon},\ \theta)) = 0$. The usual probability constraints are: $0 < p_i < 1$ and $\sum_{i=1}^{n} p_i = 1$.

The Lagrangian function of the log empirical likelihood is of the form:

$$\max_{\{p_i,\ \lambda,\ \theta\}} n^{-1} \sum_{i=1}^{n} \log p_i - \eta(\sum_{i=1}^{n} p_i - 1) - \lambda' E_p h(\hat{\varepsilon}_i,\ \theta), \tag{4.10}$$

where $\theta = \sigma^2$, and λ is the vector of the Lagrangian multipliers. The optimal value for η is unity. The p_i 's can be solved as functions of λ and θ: $p_i = n^{-1}(1 + \lambda' E_p(h(\hat{\varepsilon}_i,\ \theta)))^{-1}$. Substituting this information into the Lagrangian, we get a minimization problem involving only the vector of the Lagrangian multipliers λ and the parameter σ^2. The first order condition with respect to the parameter σ^2 is:

$$\sum_{i=1}^{n} p_i(\lambda_2 + 6\sigma^2\lambda_4) = 0. \tag{4.11}$$

With the four moment equations and the first order condition, we have a system of five equations to be solved for the five unknowns. The EL estimators are $\hat{\lambda}$ and $\hat{\sigma}^2$. Substituting these back into the formula for the p_i's, we get the $\hat{p}_i$'s. The empirical likelihood ratio test statistic is formed as:

$$-2\log R(\sigma^2) = 2\log(1 + \hat{\lambda}' E_{\hat{p}} h(\hat{\varepsilon}_i, \hat{\sigma}^2)). \tag{4.12}$$

The limiting distribution of the test statistic is $\chi^2_{(d)}$ where the degrees of freedom, d, equals the number of the moment constraints less the number of the parameter, which is implies that $d = 3$.

For the JB test, the D test, and the χ^2 tests, every step is the same as in Chapter 3 except that we replace the y_i data with the $\hat{\varepsilon}_i$'s.

4.3 Monte Carlo Simulation

4.3.1 The Set Up of the Experiments

The model used here, $y = X\beta + \varepsilon$, is the one we discussed in Section 4.2 with four regressors, *i.e.*, $k = 4$. The regressor matrix X is constructed by first obtaining three $2,000 \times 1$ vectors of uniformly distributed random variables. Then we transform the vectors to have zero mean and theoretical unitary variance. By adding a vector of ones, the basic $2,000 \times 4$ regressor matrix is formed. The basic X regressor matrix is stored as a data file, and loaded into the program as needed. For those sample sizes, where n is smaller than $2,000$, the regressor matrices are simply obtained by taking the first n rows of the basic regressor X matrix. The regressor matrix for the same sample size is kept fixed for each replication. $10,000$ replications are made for the null and for each of the alternative distributions, given the various sample sizes from 30 to $2,000$. The largest sample size we consider is $2,000$ due to our computing capacity since for this portion of the study, we used a pentium 3 500MHZ personal computer.

The vector of the random disturbance ε is drawn from the null hypothesis distribution $N(0, \sigma^2 I)$ with the true value $\sigma^2 = 1$. In computing the power of a test, the error vectors of the random disturbances are drawn from the following four alternative distributions : Lognormal, $\chi^2_{(2)}$, Students $t(5)$, and Double Exponential respectively as in Chapter 3. In each case, the vector ε is standardized to have zero mean and unitary variance as for the pure random data case. This particular transformation does not result in any loss of generality. All of the tests that we have considered are invariant with respect to the mean and the variance of the error term ε.

For each replication, five normality tests are applied to the OLS residuals. The five tests considered here are the ELR, the JB, the D test, the χ^2 and the χ^{2*}, the same ones as discussed in Chapter 3. The ELR test and the two χ^2 goodness of fit tests are appropriate and readily applicable to the OLS residuals directly. Their asymptotic distributions are chi-squared as in the pure random data case. White and McDonald (1980) provided the theoretical evidence that the D test, and the $\sqrt{\alpha_3}$ and α_4 measures can be applied to the residuals in testing for the normality of the disturbances in regressions, where α_3 and α_4 are the skewness and kurtosis coefficients used in the JB test. The JB test is a smooth function of these two coefficients. Therefore, the JB test is applicable in testing for normality in a regression model. Thus, the five tests are able to maintain their asymptotic properties in testing normality in a regression model.

There are two aspects of these results that we should keep in mind. One is that the distribution of the OLS residuals is closer to normal than is the non-normal random error term itself. Thus, we would expect that any test for normality using regression residuals is more likely to fail in rejecting the null hypothesis when the null is false than would the case if the test were able to be constructed using the true random error term itself. Second, the distribution of the OLS residual vector $\hat{\varepsilon}$ depends on the distribution of the error term ε, the number of regressors k and the elements of the regressor matrix X, and the sample size n. Thus, the performance of any test for normality depends on these factors, ε, M, k, and n, as well. In the Monte Carlo experiments that are conducted here, we have simulated the size-adjusted critical values for the tests that we consider. These critical values are specific to the regressor matrices we have chosen; they are not applicable to other situations with a different regressor matrix X.

To provide some guidance to the researchers who may be interested in using the ELR test for normality in the context of regression, we will be providing, on the internet, a small library that contains two procedures. One procedure will intake a more general regressor matrix X and will calculate the size and the correct size-adjusted critical values. Another procedure will intake the X matrix and the associated size-adjusted critical values and will calculate the actual powers of the ELR test.

4.3.2 Experiment Results

Tables 4.1 to 4.5 in the Appendix of this chapter give the results of the Monte Carlo experiments in the application of testing for normality in a regression model.

Table 4.1 presents the size and the size-adjusted critical values for the ELR test, the JB test, and the two χ^2 tests. The size of a test is computed here as the rejection rate when the null hypothesis is true, given the nominal significance level. The size distortion is the difference between the empirical size of the test and the nominal significance level. We choose to illustrate the size at four nominal significance levels, 10%, 5%, 2%, 1%, in order to provide a broad picture of the sampling properties of the tests. The null hypothesis is $N(0,\ \sigma^2)$ where the true value of $\sigma^2 = 1$. The size and the size-adjusted critical values for the D test are not provided in this table. The reason for this is given in Section 3.2.3. The percentile points of the D test are taken from D'Agostino (1971 and 1972). For each experimental replication the same data set is used for the construction of all five tests to make the comparisons valid.

The size of the ELR test is larger than the nominal significance level, but it converges nicely to the nominal level as the sample size grows. For example, the size of the test changes from 28.11% to 6.02% when the sample size varies from 30 to 2000 at the nominal significance level of 5%.

The size of the JB test is lower than the nominal significance level; it converges to the nominal level as the sample size grows. For instance, the size of the test increases from 2.78% to 4.88% when the sample size varies from 30 to 2000 at the nominal level of 5%. Comparing the actual sizes of the ELR test and the JB test, it is clearly that the ELR test is always over-rejecting while the JB test is under-rejecting. In the context of testing for normality in regression residuals, given the fact that the residuals are closer to the normal than the error term itself, the situation seems to lend a certain advantage to the ELR test. Suppose there is an alternative distribution for the error term that is very close to normal, then the ELR test would have a higher possibility of rejecting the null hypothesis than would the JB test.

The size of the basic χ^2 test is closer to the nominal significance level than is the case for the ELR test for small samples. For example, the size is 5.09% at $n = 30$, and with a nominal significance level of 5%. However, it settles down to 9.1% when the sample size grows to 2000. The size distortion of the test does not vanish when n grows to this extent. This is avoidable if we adjust the χ^2 test to take into account the consideration that the observed frequencies in each category should be no less than five. The adjusted χ^2 test is denoted as the χ^{2*} test. The actual size of the χ^{2*} test is very close to the nominal significance level at all sample sizes. The size distortion vanishes when n grows. Comparing the four tests, the size of the χ^{2*} test is the best in the sense it is the closest to the nominal significance level. It exhibits the least size distortion, overall.

The power comparisons of the five selected tests are given in Tables 4.2 to 4.5. The simulated size-adjusted critical values are used in computing the power of each test. That is, each test is now applied at the same actual significance level by using the critical values that are appropriate for the particular value of n. The sample size ranges from 30 to 2000.

The alternative distributions are: the Lognormal, the $\chi^2_{(2)}$, the Student $t_{(5)}$, and the Double Exponential. Each alternative distribution is standardized to have mean zero and variance unity for comparison with the null distribution of $N(0,\ \sigma^2)$ where $\sigma^2 = 1$.

The power of the ELR test increases as the sample size grows, given any alternative distribution. For example, the power increases from 69.39% to 100% when n grows from 30 to 100, given that the alternative distribution is the Lognormal, and at a true significance level of 5%. All of other tests are consistent while the power of the χ^2 test converges more slowly than others. The powers of the first three tests, the ELR test, the JB test, and the D test, are close to each other for moderate samples. For instance, the powers of these tests are 97.91%, 99.14%, 93.37% at $n = 50$ and with a significance level of 5%. For the small sample size $n = 30$, the JB test is the most powerful one, and the ELR test is the second most powerful. The χ^2 tests are not applicable for some of the smaller sample sizes that are considered. Overall, all of the five tests have good power against the lognormal alternative distribution.

When the alternative distribution is $\chi^2_{(2)}$, the power of the ELR test is 55.29% at $n = 30$ and reaches 100% at $n = 100$ and at the nominal level of 5%. The power of the JB test is slightly better than that of the ELR test. The power of the tests in descending order is: the JB test, the ELR test, the D, the χ^2 and the χ^{2*} tests for all sample sizes.

Both the Lognormal and the $\chi^2_{(2)}$ are asymmetric alternative distributions. It is, relative to symmetric alternatives, easier for any of these tests to detect such a departure from the normality. All of these tests have high power for medium and large samples.

The Student $t_{(5)}$ and the Double Exponential (DE) alternative distributions are both symmetric with fatter tails relative to the normal distribution. It is difficult for any commonly used tests to detect for the departure from normality in such cases. All of the tests have lower power for small sample size $n = 30$ over these two alternatives than was the case with asymmetric alternative distributions. The power of the ELR test is only 5.68% for the $St(5)$ and is 6.8% for the DE at $n = 30$. For the alternative $St(5)$, the power of the D test overrides the JB test when $n > 150$. For the alternative distribution DE, the D test is the most powerful test among the five tests when $n > 30$. The power of the ELR test reaches 100% at $n = 1000$.

4.4 Conclusion

A new empirical likelihood ratio test for normality in a regression model is constructed in this chapter. Monte Carlo experiments are employed to simulate the actual sizes and the size-adjusted critical values for the ELR test and several other tests that we have selected. The size-adjusted critical values are then used to conduct power comparisons for all of the tests. The experiment results indicate that the actual size of the ELR test is still large relative to the nominal significance level. However, using the actual significance level, the ELR test has good power properties relative to other tests, and is recommended for use in regression context.

4.4.1 Future Work

As in Chapter 3, we notice that the ELR test has large size distortion for small samples. It would be interesting to explore the techniques that can reduce the size distortion and maintain the good power properties for the ELR approach. Testing for normality when the alternative distributions are symmetric is a challenge using the ELR test. It would be useful to explore ways to improve the power of the test in this context.

Appendix: Normality Tables in Regressions

Table 4.1: Size and Size-adjusted Critical Values for the Tests in the OLS Residuals

m	10,000		$H_0 : N(0, \sigma^2)$ where $\sigma^2 = 1$							
n	30	50	70	100	150	200	250	500	1,000	2,000
ELR test at nominal levels:										
10%	0.2638	0.2276	0.2063	0.1771	0.1615	0.1408	0.1344	0.1076	0.0776	0.0648
5%	0.1878	0.1584	0.1423	0.1241	0.1042	0.0924	0.0838	0.0640	0.0431	0.0345
2%	0.1185	0.0974	0.0885	0.0771	0.0633	0.0550	0.0492	0.0358	0.0202	0.0155
1%	0.0801	0.0674	0.0626	0.0520	0.0419	0.0368	0.0325	0.0238	0.0114	0.0089
Size-adjusted Critical Values:										
10%	10.43	9.70	9.25	8.68	7.98	7.49	7.27	6.47	5.61	5.19
5%	12.95	12.59	12.21	11.45	10.65	10.17	9.74	8.61	7.47	6.83
2%	16.64	16.27	16.11	15.47	14.42	13.87	13.28	11.80	9.85	9.20
1%	19.29	18.81	18.58	18.05	17.32	16.60	16.07	14.36	11.77	10.95
JB test:										
10%	0.0446	0.0552	0.0613	0.0671	0.0737	0.0748	0.0805	0.0881	0.0927	0.0963
5%	0.0284	0.0361	0.0382	0.0423	0.0438	0.0415	0.0463	0.0477	0.0515	0.0506
2%	0.0181	0.0240	0.0228	0.0252	0.0256	0.0240	0.0248	0.0223	0.0246	0.0231
1%	0.0126	0.0187	0.0161	0.0173	0.0179	0.0172	0.0174	0.0127	0.0146	0.0136
Size-adjusted Critical Values:										
10%	2.72	3.12	3.46	3.65	3.89	4.06	4.12	4.32	4.43	4.53
5%	4.29	4.94	5.14	5.42	5.62	5.52	5.78	5.91	6.05	6.02
2%	7.28	8.81	8.23	8.83	8.81	8.49	8.58	8.00	8.29	8.17
1%	10.32	12.73	11.20	12.04	11.54	12.33	11.76	10.13	10.31	9.99
χ^2 *goodness of fit test:*										
10%	0.0593	0.0662	0.0743	0.0880	0.1006	0.1063	0.1134	0.1193	0.1214	0.1270
5%	0.0262	0.0351	0.0404	0.0528	0.0628	0.0675	0.0720	0.0758	0.0798	0.0782
2%	0.0110	0.0170	0.0220	0.0296	0.0379	0.0441	0.0463	0.0485	0.0490	0.0482
1%	0.0052	0.0110	0.0147	0.0216	0.0284	0.0329	0.0350	0.0356	0.0341	0.0363
Size-adjusted Critical Values:										
10%	6.50	10.51	14.43	20.59	30.52	38.78	46.92	57.19	58.86	60.14
5%	8.11	12.73	16.99	23.96	34.98	44.23	52.68	63.90	65.47	66.54
2%	9.98	15.75	20.89	29.16	42.34	53.12	62.82	73.67	76.18	80.09
1%	11.79	18.46	24.99	36.86	51.78	67.35	79.34	85.79	90.22	97.87
χ^{2*} *goodness of fit test:*										
10%	–	0.0812	0.0800	0.0818	0.0820	0.0835	0.0859	0.0826	0.0799	0.0830
5%	–	0.0393	0.0400	0.0414	0.0415	0.0407	0.0406	0.0413	0.0395	0.0398
2%	–	0.0157	0.0151	0.0154	0.0161	0.0162	0.0168	0.0155	0.0177	0.0155
1%	–	0.0073	0.0079	0.0076	0.0082	0.0089	0.0095	0.0085	0.0087	0.0086
Size-adjusted Critical Values:										
10%	–	9.14	11.13	14.37	19.72	24.58	29.41	39.72	44.07	47.48
5%	–	11.08	13.37	16.87	22.41	27.58	32.42	43.42	47.92	51.34
2%	–	13.42	16.13	19.98	25.79	31.33	36.58	47.53	52.93	56.06
1%	–	15.34	18.19	21.68	28.48	33.75	39.74	50.94	56.17	59.50

Notes to table: m is the number of replications; n is the sample size. The ELR test uses the first four raw moment equations with one parameter. The χ^2 and χ^{2*} tests may not be applicable with small samples.

Table 4.2: Power of The Five Tests in the OLS Residuals

m	10,000		H_a : Lognormal				
n	30	50	70	100	150	200	250
ELR test at actual levels:							
10%	0.8637	0.9968	1	1	1	1	1
5%	0.7785	0.9874	0.9996	1	1	1	1
2%	0.6297	0.9647	0.9984	1	1	1	1
1%	0.5174	0.9355	0.9961	1	1	1	1
JB test:							
10%	0.9339	0.9978	1	1	1	1	1
5%	0.8715	0.9895	0.9999	1	1	1	1
2%	0.7670	0.9627	0.9981	1	1	1	1
1%	0.6837	0.9255	0.9956	0.9997	1	1	1
D test:							
10%	0.8262	0.9735	0.9970	0.9997	1	1	1
5%	0.7648	0.9585	0.9957	0.9994	1	1	1
2%	0.6898	0.9327	0.9915	0.9988	1	1	1
1%	0.6272	0.9095	0.9869	0.9986	1	1	1
χ^2 *goodness of fit test:*							
10%	–	–	–	0.9819	0.9984	0.9991	0.9992
5%	–	–	–	0.9696	0.997	0.9988	0.9991
2%	–	–	–	0.9486	0.9932	0.9982	0.9989
1%	–	–	–	0.9166	0.9879	0.9969	0.9984
χ^{2*} *goodness of fit test:*							
10%	–	–	–	–	–	0.9999	1
5%	–	–	–	–	–	0.9998	1
2%	–	–	–	–	–	0.9992	1
1%	–	–	–	–	–	0.9985	1

Notes to table: m is the number of replications; n is the sample size. The ELR test uses the first four raw moment equations with one parameter. The χ^2 and χ^{2*} tests may not be applicable with small samples.

Table 4.3: Power of The Five Tests in the OLS Residuals

m	10,000		$H_a : \chi^2_{(2)}$							
n	30	50	70	100	150	200	250	500	1,000	2,000
ELR test at actual levels:										
10%	0.7548	0.9803	0.9988	0.9999	1	1	1			
5%	0.6362	0.953	0.9963	0.9999	1	1	1			
2%	0.4774	0.8938	0.9881	0.9997	1	1	1			
1%	0.3674	0.8362	0.9782	0.9993	1	1	1			
JB test:										
10%	0.8217	0.9841	0.9991	0.9999	1	1	1			
5%	0.6855	0.9328	0.9929	0.9998	1	1	1			
2%	0.5103	0.8101	0.9616	0.9976	1	1	1			
1%	0.4025	0.6974	0.9165	0.9901	1	1	1			
D test:										
10%	0.5986	0.8434	0.9403	0.9891	0.9996	0.9999	1			
5%	0.5000	0.7854	0.9116	0.9801	0.9988	0.9999	1			
2%	0.3911	0.7025	0.8653	0.9656	0.9973	0.9998	1			
1%	0.3172	0.6352	0.8250	0.9498	0.9949	0.9997	1			
χ^2 *goodness of fit test:*										
10%	–	0.6409	0.8204	0.9546	0.9977	0.9997	1			
5%	–	0.5411	0.743	0.9256	0.9945	0.9993	1			
2%	–	0.4307	0.6343	0.865	0.9863	0.9983	0.9999			
1%	–	0.3519	0.5283	0.7523	0.9595	0.9935	0.9993			
χ^{2*} *goodness of fit test:*										
10%	–	–	–	0.9237	0.9960	0.9996	1			
5%	–	–	–	0.8729	0.9919	0.999	1			
2%	–	–	–	0.7976	0.9828	0.998	1			
1%	–	–	–	0.7481	0.9710	0.9967	0.9995			

Notes to table: m is the number of replications; n is the sample size. The ELR test uses the first four raw moment equations with one parameter. The χ^2 and χ^{2*} tests may not be applicable with small samples.

Table 4.4: Power of the Five Tests in the OLS Residuals

m	10,000		H_a : Student $t_{(5)}$						
n	30	50	70	100	150	200	250	500	1,000
ELR test at actual levels:									
10%	0.1008	0.1290	0.1987	0.3414	0.5541	0.7255	0.8291	0.9886	1
5%	0.0568	0.0681	0.1063	0.2174	0.4175	0.5931	0.7330	0.9777	1
2%	0.0239	0.0262	0.0433	0.1054	0.2505	0.4216	0.5837	0.9497	1
1%	0.0130	0.0153	0.0248	0.0651	0.1636	0.3159	0.4693	0.9161	0.9998
JB test:									
10%	0.3152	0.4632	0.5732	0.6962	0.8226	0.8975	0.9365	0.9963	1
5%	0.2332	0.3687	0.4933	0.6162	0.7654	0.8644	0.9121	0.9939	1
2%	0.1567	0.2693	0.3988	0.5120	0.6805	0.7975	0.8697	0.9896	1
1%	0.1171	0.2121	0.3375	0.4494	0.6221	0.7268	0.8224	0.9838	1
D test:									
10%	0.2533	0.4036	0.5453	0.6771	0.8226	0.9090	0.9479	0.9987	1
5%	0.1700	0.3168	0.4526	0.5902	0.7625	0.8671	0.9192	0.9964	1
2%	0.1080	0.2337	0.3546	0.4941	0.6764	0.8045	0.8798	0.9932	1
1%	0.0791	0.1821	0.2892	0.4302	0.6149	0.7556	0.8448	0.9882	1
χ^2 *goodness of fit test:*									
10%	–	0.2982	0.3660	0.4393	0.5349	0.6233	0.6810	0.9153	0.9979
5%	–	0.2184	0.2841	0.3620	0.4610	0.5504	0.6201	0.8767	0.9956
2%	–	0.1566	0.2146	0.2928	0.3911	0.4809	0.5522	0.8279	0.9884
1%	–	0.1271	0.1776	0.2414	0.3439	0.4216	0.4873	0.7765	0.9729
χ^{2*} *goodness of fit test:*									
10%	–	–	–	0.2221	0.2335	0.2404	0.2404	0.3603	0.8587
5%	–	–	–	0.1323	0.1476	0.1534	0.1561	0.2500	0.7758
2%	–	–	–	0.0675	0.0782	0.0822	0.0805	0.1540	0.6439
1%	–	–	–	0.0439	0.0463	0.0545	0.0472	0.0998	0.5473

Notes to table: m is the number of replications; n is the sample size. The ELR test uses the first four raw moment equations with one parameter. The χ^2 and χ^{2*} tests may not be applicable with small samples.

Table 4.5: Power of the Five Tests in the OLS Residuals

m	10,000	H_a : Double Exponential							
n	30	50	70	100	150	200	250	500	1,000
ELR test at actual levels:									
10%	0.1057	0.1695	0.3066	0.5374	0.8045	0.9299	0.9751	1	1
5%	0.0618	0.0893	0.1752	0.3750	0.6824	0.8606	0.9458	1	1
2%	0.0296	0.0397	0.0738	0.1978	0.4949	0.7389	0.8792	0.9995	1
1%	0.0170	0.0228	0.0442	0.1277	0.3648	0.6264	0.8084	0.9977	1
JB test:									
10%	0.3919	0.5738	0.7044	0.8340	0.9380	0.9763	0.9916	1	1
5%	0.2925	0.4692	0.6161	0.7645	0.9003	0.9604	0.9828	1	1
2%	0.1951	0.3383	0.5073	0.6523	0.8244	0.9248	0.9676	0.9999	1
1%	0.1419	0.2640	0.4300	0.5682	0.7680	0.8712	0.9413	0.9999	1
D test:									
10%	0.3453	0.5813	0.7514	0.8914	0.9713	0.9936	0.9988	1	1
5%	0.2409	0.4714	0.6611	0.8332	0.9489	0.9871	0.9963	1	1
2%	0.1530	0.3471	0.5396	0.7415	0.9115	0.9721	0.9902	1	1
1%	0.1083	0.2708	0.4577	0.6726	0.8725	0.9559	0.9844	1	1
χ^2 *goodness of fit test:*									
10%	–	0.4301	0.5381	0.6545	0.7771	0.8667	0.9152	0.9980	1
5%	–	0.3221	0.4310	0.5472	0.6848	0.7870	0.8588	0.9956	1
2%	–	0.2238	0.3187	0.4230	0.5489	0.6611	0.7510	0.9855	1
1%	–	0.1727	0.2483	0.3215	0.4381	0.5237	0.6165	0.9567	1
χ^2 *goodness of fit test:*									
10%	–	0.2733	0.3389	0.4272	0.5649	0.6623	0.7296	0.9744	1
5%	–	0.1827	0.2258	0.3013	0.4365	0.5387	0.6238	0.9489	1
2%	–	0.1101	0.1319	0.1809	0.3029	0.3947	0.4840	0.9036	1
1%	–	0.0707	0.0873	0.1351	0.2198	0.3124	0.3716	0.8508	1

Notes to table: m is the number of replications; n is the sample size. The ELR test uses the first four raw moment equations with one parameter. The χ^2 and χ^{2*} tests may not be applicable with small samples.

Chapter 5

The Behrens-Fisher Problem

5.1 Introduction

Testing the equality of the means of two normal populations when the variances are both unknown, and not known to be equal, is called the Behrens-Fisher (BF) problem. The Behrens-Fisher problem has been well known since the early 1930's. One reason for its fame is that it can be proven that there is no solution satisfying the classical criteria for good tests. That is, every invariant rejection region of fixed size for the problem must have some unpleasant properties (Zaman, 1996, p. 246). First best solutions that are uniformly most powerful and invariant either do not exist or have strange properties. We need to look for second best solutions.

In the literature associated with the Behrens-Fisher problem, there have been quite a few "solutions" proposed since the 1930's. For example, Fisher (1935), Welch-Aspin (1947), Cox (1950), Qin (1991) and Jing (1995) have all suggested different solutions. Lee and Gurland (1975) presented a detailed comparison of several selected tests and proposed a refined solution to the BF problem. These solutions can be classified into two categories, approximate tests and asymptotic tests.

The purpose of this chapter is to offer a new solution to the Behrens-Fisher problem. The approach that we use is based on the new nonparametric technique of the empirical likelihood (EL) method. We will present some new theoretical results to solve the BF problem. The way that we use the data and information in this chapter is distinct from that in the other chapters of this dissertation. Our EL approach makes efficient use of the information available. The EL approach ties together the estimation and testing procedures nicely. In addition to this, we conduct a power comparison of the empirical likelihood ratio (ELR) test and the Welch-Aspin (1947)test for the BF problem. The ELR test is an asymptotic test. Its finite-sample properties are simulated using Monte Carlo techniques. The sizes and the power of the ELR test in finite samples are given for a range of situations. The results indicate that the EL approach to the BF problem is both efficient and attractive.

There have been two empirical likelihood type (EL-type) approaches to the BF problem in the literature, that of Qin (1991) and that of Jing (1995). Details are provided in Section 5.2.2. The EL approach of this chapter is quite distinct from those from Qin and Jing in two respects. First, our way of using the data is different. We transform the first data set S_1 into a data set S_a that has the same theoretical distribution as the second data set S_2. Then we combine the transformed data set S_a and the second data set S_2 into a full data set S such that the data have a unique distribution. Second, we exploit the data information in an efficient way. We assume that the first five moments of the data exist and we use the five moment equations as constraints to set up the EL method. Then, we apply the usual EL approach as described in Chapter 2 to the full data set S. The ELR test is constructed for the equality of the two means without assuming the two variances are known or equal. Details of this are provided in Section 5.2.3.

The outline of this chapter is as follows: Section 5.2 gives a brief introduction to the conventional tests in the BF problem literature. The Welch-Aspin test and the approaches of Qin (1991) and Jing (1995) are discussed. Section 5.3 sets up the new EL approach to the BF problem. It provides the ELR test and the EL-type Wald test for the BF problem. Section 5.4 presents the Monte Carlo experiments and the results for the new EL approach. The size and the power of the ELR test in finite samples are analyzed in detail across a broad range of situations. Section 5.5 provides a summary and some conclusions.

5.2 Solutions to the Behrens-Fisher Problem

Suppose $x_{i1}, x_{i2}, \ldots, x_{in_i}, i = 1,2$ are the observed values of two random samples independently drawn from the normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Let

$$\rho^2 = \frac{\sigma_2^2}{\sigma_1^2}, \tag{5.1}$$

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \tag{5.2}$$

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2, \tag{5.3}$$

$$C_0 = \frac{\frac{\sigma_1^2}{n_1}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \frac{1}{1 + \rho^2 \frac{n_1}{n_2}}. \tag{5.4}$$

The problem in hand is to construct a test of the null hypothesis, $H_0 : \mu_1 = \mu_2$, against the alternative hypothesis, $H_a : \mu_1 \neq \mu_2$. Various tests have been developed by Fisher (1935), Welch and Aspin (1947), Cox (1950) and others. The critical regions corresponding to these conventional tests have the general form of:

$$|v| > V(\hat{C}), \tag{5.5}$$

where $v = (\bar{x}_1 - \bar{x}_2)(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^{-1}$, $\hat{C} = (\frac{s_1^2}{n_1})(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^{-1}$. $V(\hat{C})$ is a function of $\hat{C}$ and the preassigned nominal significance level α. The distribution of v is no longer Student-t when the two variances are not known to be equal. The critical regions depend on the variance parameters σ_i^2, $i = 1,2$ and sample sizes $(n_1,\ n_2)$. These solutions to the BF problem involve utilizing various means to approximate the distribution of the variable v and to control for the size distortion of the tests. Further details and discussion can be found in Lee and Gurland (1975).

5.2.1 Welch-Aspin Test

Welch (1947) and Aspin (1948) independently developed a higher order approximation, $V_{wa}(\hat{C})$, to the distribution of $V(\hat{C})$ in the terms up to f_i^{-2} and f_i^{-4}, where $f_i = n_i - 1$

and $i = 1,\ 2$. Their approach is referred to as the Welch-Aspin (WA) test. The WA test is a highly efficient solution to the BF problem (Weerahandi, 1987). However, the functional form of $V_{wa}(\hat{C})$ is lengthy, involving infinite series, and it is difficult to work with (Lee and Gurland, 1975). The actual sizes of the Welch-Aspin test in small samples are very close to the nominal significance levels. For example, at a nominal level of 5%, the size of the WA test lies between 4.98% and 5.02% for $(n_1,\ n_2) = (7,\ 7)$.

Lee and Gurland (1975) provided detailed comparisons of various tests that were proposed by Fisher (1935), Cox (1950), Welch (1937), Welch-Aspin (1947) and others. At the last stage of their comparison, all of the tests were eliminated from their tables, except the Welch-Aspin test because the WA test has very accurate size and good power properties. In addition to that, Lee and Gurland proposed a refined test, which we call the LG test, using two techniques: (i) a simple functional form to approximate the $V(C)$ function; and (ii) a T-transform to accelerate the convergence of the size and the power functions. The method in their research involves solving a minimization problem of the squared difference between the size function of the test and the preassigned nominal level, and the T-transform is not trivial. Lee and Gurland provided the refined results that were very close to the results of the WA test. For the reasons mentioned above, we choose to use the WA test and cite the results of the WA test from Lee and Gurland for our comparison with the ELR test for the case of $(n_1,\ n_2) = (7,\ 7)$ and $\alpha = 5\%$. Of course, we also consider other sample sizes and nominal significance levels when considering the ELR test. A detailed discussion is given in Section 5.3.

5.2.2 Approaches of Qin and Jing

As has been noted in earlier chapters, the EL approach makes use of likelihood functions and the moment conditions of the data without assuming a specific parametric form for the underlying distribution. It has the flexibility to incorporate various information about the data into the approach. This leads to the efficiency of the method. Applying the EL method to the Behrens-Fisher problem gives us a new way of using the data and it demonstrates that the EL method is a useful tool in solving various econometric problems.

Jing (1995) proposed a nonparametric version of the EL approach to the two-sample problem. In his paper, he showed that the nonparametric version of Wilks' theorem holds in the two-sample problem and that the solution is Bartlett correctable. The empirical likelihood ratio statistic has a limiting chi-squared distribution with an error of order $O(n^{-1})$ and with the Bartlett correction, the error is reduced to the order of $O(n^{-2})$, where n is the smaller one of the two sample sizes. A special case of his approach is a particular solution to the Behrens-Fisher problem.

The focus of Jing (1995) was primarily on the coverage accuracy and the Bartlett correction of the test. The approach he used works only when the null hypothesis is true. It means that the restriction that two means are equal must be binding, and the Lagrangian multiplier for the constraint must not equal zero. If the restriction is not binding, the solution of his approach reduces to a situation where the probability parameter $p_i = 1/n$, and the estimated empirical mean of the data is just the sample mean. His approach does not appear to offer the potential to deal with issues relating to power in the context of an ELR test.

Qin (1991) generalized Owen's empirical likelihood to a two-sample problem in which one sample is assumed to come from a distribution that is unknown and the other sample is assumed to come from a known distribution specified upto a parameter, $g_\theta(x_2)$. Qin's approach is a semi-parametric one by combining the EL method and the parametric likelihood method to the two-sample problem. Consider the following assumptions:

- $\mu_2 = \mu(\theta)$ is differentiable and $\mu'(\theta) \neq 0$
- the known density function $g_\theta(x_2)$ is differentiable three times with respect to θ
- $\int g_\theta(x_2)dx_2$ is twice differentiable
- The Fisher information matrix, $I(\theta) = E[\frac{\partial \log g_\theta}{\partial \theta}]^2$, satisfies $0 < I(\theta) < \infty$
- $\mid \frac{\partial^3}{\partial \theta^3} \log g_\theta(x_2) \mid$ is bounded.

Then, there exists an EL estimator $\hat{\theta}$ of the mean μ_1 that is more efficient than the sample mean $\bar{x}_1$. Qin theoretically proved that the empirical likelihood ratio has a limiting distribution of $\chi^2_{(1)}$ under the null hypothesis $\mu_1 = \mu_2 = \mu(\theta)$. The coverage accuracy is of the order $O_p(n^{-\frac{1}{2}})$. The connection between the two data sets is brought in by the single restriction that the two means are equal.

The EL approach developed in this chapter provides a new theoretical solution to the BF problem and hopefully it can overcome some of the shortcomings mentioned above. The focus of the following sections is on the properties of the ELR test in finite samples for the Behrens-Fisher problem.

5.3 The EL Approach

5.3.1 ELR Test

Suppose that we have two data sets: $S_i = \{x_{i1},\ x_{i2},\ \ldots, x_{in_i}\}$, $i = 1,\ 2$. They are independently drawn from two normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. The parameter vector is $\theta = (\mu_1,\ \mu_2,\ \sigma_1^2,\ \sigma_2^2)'$, which is unknown. Solving the BF problem involves constructing a test of the equality of the two population means without the knowledge of the variances. As noted above, the new approach described in this chapter is distinct from those of Jing (1995) and Qin (1991).

First, we transform S_1, into a data set that has the same theoretical distribution as S_2. Without losing generality, we assume that $n_1 \geq n_2$. The transformation has the following form:

$$t_j = (x_{1j} - \mu_1)(\rho^2)^{\frac{1}{2}} + \mu_2, \tag{5.6}$$

where $x_{1j} \in S_1$, and $\rho^2 = \sigma_2^2/\sigma_1^2$. Thus, $t_j \sim N(\mu_2,\ \sigma_2^2)$, $j = 1,\ 2,\ \ldots,\ n_1$.

The second data set S_2 remains unchanged. We denote: $t_{n_1+j} = x_{2j}$, where $x_{2j} \in S_2$, for $j = 1,\ 2, \ldots, n_2$. Let $n = n_1 + n_2$. Then, the full data set $S = \{t_1,\ t_2,\ \ldots,\ t_n\}$ is of size n and has a distribution of $N(\mu_2,\ \sigma_2^2)$.

The parameter $C_0 = (1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_1}{n_2})^{-1}$ is the one that is used in the BF literature (*e.g.*, Welch (1947), Lee and Gurland (1975)), where the n_i's are the sample sizes of the data sets, and the σ_i^2's are the population variance parameters, $i = 1,\ 2$. Obviously, $C_0 \in (0,\ 1)$, and it is a smooth function of the ratio of the variances, ρ^2, and the ratio of the sample sizes, n_1/n_2. In our study, we will keep the ratio of the sample sizes constant and vary the ratio of the variances such that the corresponding values of C_0 vary widely in the range of $(0,\ 1)$.

The next step is to apply the EL approach to the full data set S. A probability parameter

p_j is assigned to each data point t_j. The empirical likelihood function for the full data set is formed as $\prod_{j=1}^{n} p_j$. The EL approach is to maximize the empirical likelihood function subject to the probability constraints, $0 < p_j < 1$ and $\sum_{j=1}^{n} p_j = 1$, and the information constraints.

The information that we have is that the data are independent and are normally distributed with a mean μ_2 and a variance σ_2^2. We choose to use the ratio of the variances as one of the parameters, so the parameter vector becomes $\theta = (\mu_1, \mu_2, \rho^2, \sigma_2^2)'$. In addition, we choose to use the first five unbiased raw moment equations as the information constraints, so that the number of the moment equations is greater than the number of the parameters. We denote the information constraints as $E_p h(t_j, \theta) = 0$, which have the following form:

$$\sum_{j=1}^{n} p_j t_j - \mu_2 = 0 \tag{5.7}$$

$$\sum_{j=1}^{n} p_j t_j^2 - (\mu_2^2 + \sigma_2^2) = 0 \tag{5.8}$$

$$\sum_{j=1}^{n} p_j t_j^3 - (\mu_2^3 + 3\sigma_2^2\mu_2) = 0 \tag{5.9}$$

$$\sum_{j=1}^{n} p_j t_j^4 - (\mu_2^4 + 6\sigma_2^2\mu_2^2 + 3\sigma_2^4) = 0 \tag{5.10}$$

$$\sum_{j=1}^{n} p_j t_j^5 - (\mu_2^5 + 10\sigma_2^2\mu_2^3 + 15\sigma_2^4\mu_2) = 0. \tag{5.11}$$

The set up of the Lagrangian function is:

$$G = n^{-1}\sum_{j=1}^{n} \log p_j - \eta(\sum_{j=1}^{n} p_j - 1) - \lambda' \sum_{j=1}^{n} p_j h(t_j, \theta), \tag{5.12}$$

where λ is a $m \times 1$ vector, which together with the scalar η gives the Lagrangian multipliers. As illustrated in Chapter 2, η takes the value of unity and the p_j's can be expressed as the function of λ and θ:

$$p_j = n^{-1}(1 + \lambda' h(t_j, \theta))^{-1},\ j = 1, 2, \ldots, n \tag{5.13}$$

Substituting this information back into the Lagrangian function, we get a minimization problem over a reduced number of unknowns, λ and θ. Working out the first order conditions of the Lagrangian function with respect to the parameter vector θ requires some special care. We note that the values in the first portion of the data are functions of the unknown parameters. The first derivative of the data with respect to the parameters involves the following terms:

$$\frac{\partial t_j}{\partial \theta} = (-\rho,\ 1,\ (x_{1j} - \mu_1)\frac{1}{2}(\rho^2)^{-1/2},\ 0)',\quad j = 1,\ 2,\ \ldots,\ n_1, \tag{5.14}$$

where $x_{1j} \in S_1$. Taking into account this information, the actual first order conditions of the Lagrangian function with respect to the four unknown parameters $(\mu_1,\ \mu_2,\ \rho^2,\ \sigma_2^2)'$ can be represented as follows:

$$\sum_{j=1}^{n_1} p_j(\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4)(\rho^2)^{1/2} = 0$$

$$\sum_{j=1}^{n_1} p_j(\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4)$$
$$- \sum_{j=1}^{n} p_j(\lambda_1 + 2\lambda_2\mu_2 + 3\lambda_3(\sigma_2^2 + \mu_2^2) + 4\lambda_4(3\sigma_2^2\mu_2 + \mu_2^3) + 5\lambda_5(3\sigma_2^4 + 6\sigma_2^2\mu_2^2 + \mu_2^4)) = 0$$

$$\sum_{j=1}^{n_1} p_j(\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4)(t_j - \mu_1)\frac{1}{2}(\rho^2)^{-1/2} = 0$$

$$\sum_{j=1}^{n} p_j(\lambda_2 + 3\lambda_3\mu_2 + 6\lambda_4(\sigma_2^2 + \mu_2^2) + 10\lambda_5(3\sigma_2^2\mu_2 + \mu_2^3)) = 0.$$

Putting the five moment equations and these four first order conditions together, we get a system of nonlinear equations. We can solve the system using the nonlinear equation solver "Eqsolve" in the Gauss package. Suppose that $\hat{\lambda}$ and $\hat{\theta}$ are the EL estimators for the Lagrangian multiplier vector and parameter vector. The EL estimators for the probability parameters are the $\hat{p_j}$'s. The estimated maximum value of the likelihood function is obtained as $L(F^u) = \prod \hat{p}_j^u$, where u stands for the unconstrained model.

In order to construct an ELR test for the equality of the two means when the variances are unknown and are not known to be equal, we need the solutions to the constrained case. With the null hypothesis: $H_0 : \mu_1 = \mu_2$, we simply substitute the restriction $\mu_1 = \mu_2$ into the unconstrained case and get the constrained case. Here, the parameter vector becomes $\theta = (\mu,\ \rho^2,\ \sigma_2^2)'$ which is of dimension three, with $\mu = \mu_1 = \mu_2$. The five moment equations are:

$$\sum_{j=1}^{n} p_j t_j - \mu = 0 \tag{5.15}$$

$$\sum_{j=1}^{n} p_j t_j^2 - (\mu^2 + \sigma_2^2) = 0 \tag{5.16}$$

$$\sum_{j=1}^{n} p_j t_j^3 - (\mu^3 + 3\sigma_2^2\mu) = 0 \tag{5.17}$$

$$\sum_{j=1}^{n} p_j t_j^4 - (\mu^4 + 6\sigma_2^2\mu^2 + 3\sigma_2^4) = 0 \tag{5.18}$$

$$\sum_{j=1}^{n} p_j t_j^5 - (\mu^5 + 10\sigma_2^2\mu^3 + 15\sigma_2^4\mu) = 0. \tag{5.19}$$

The first derivative of the first portion of the data with respect to the parameters has the following form:

$$\frac{\partial t_j}{\partial \theta} = (-(\rho^2)^{1/2} + 1,\ (x_{1j} - \mu)\frac{1}{2}(\rho^2)^{-1/2},\ 0)',\ \ j = 1,\ 2,\ \ldots,\ n_1, \tag{5.20}$$

where $x_{1j} \in S_1$. The first order conditions of the Lagrangian function with respect to the three parameters are:

$$\sum_{j=1}^{n_1} p_j(\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4)((\rho^2)^{1/2} - 1)$$
$$- \sum_{j=1}^{n} p_j(\lambda_1 + 2\lambda_2\mu + 3\lambda_3(\sigma_2^2 + \mu^2) + 4\lambda_4(3\sigma_2^2\mu + \mu^3) + 5\lambda_5(3\sigma_2^4 + 6\sigma_2^2\mu^2 + \mu^4)) = 0$$
$$\sum_{j=1}^{n_1} p_j(\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4)(t_j - \mu)\frac{1}{2}(\rho^2)^{-1/2} = 0$$
$$\sum_{j=1}^{n} p_j(\lambda_2 + 3\lambda_3\mu + 6\lambda_4(\sigma_2^2 + \mu^2) + 10\lambda_5(3\sigma_2^2\mu + \mu^3)) = 0.$$

Solving the system of the nonlinear equations of the five moment equations and the three first order condition, we get $\hat{\lambda}^c$, $\hat{\theta}^c$, and then, the $\hat{p}_j^c$'s. The estimated maximum value of the empirical likelihood function under the constraint is formed as $L(F^c) = \prod_{j=1}^{n} \hat{p}_j^c$, where c stands for the constrained case.

The log likelihood ratio statistic:

$$\begin{aligned} -2\log R(F) &= -2\log\frac{L(F^c)}{L(F^u)} \\ &= 2\sum_{j=1}^{n}(\log(1 + \hat{\lambda^c}' h(t_j,\ \hat{\theta}^c)) - \log(1 + \hat{\lambda^u}' h(t_j,\ \hat{\theta}^u))) \end{aligned}$$

has a limiting distribution of $\chi^2_{(1)}$ if H_0 is true.

5.3.2 EL-type Wald Test

The EL estimator of the parameter vector is asymptotically efficient, and it has a limiting distribution of the following form:

$$\sqrt{n}(\hat{\theta}_{EL} - \theta_0) \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma = [E[\frac{\partial h(y,\theta)}{\partial \theta} |_{\theta_0}]E[h(y,\theta)h(y,\theta)' |_{\theta_0}]^{-1}E[\frac{\partial h(y,\theta)}{\partial \theta'} |_{\theta_0}]]^{-1}, \tag{5.21}$$

and θ_0 is the true value of θ. A consistent estimator of the asymptotic covariance matrix can be obtained using the EL estimator $\hat{\theta}$:

$$\hat{\Sigma} = [E[\frac{\partial h(y,\theta)}{\partial \theta} |_{\hat{\theta}}]E[h(y,\theta)h(y,\theta)' |_{\hat{\theta}}]^{-1}E[\frac{\partial h(y,\theta)}{\partial \theta'} |_{\hat{\theta}}]]^{-1}. \tag{5.22}$$

The estimated covariance matrix $\hat{\Sigma}$ is a 4×4 symmetric matrix. With this, we can easily obtain an EL-type Wald test for any linear restrictions of the parameters. Suppose $c\hat{\theta} - r = 0$ is a set of j restrictions. The EL-type Wald test has the form:

$$W = (c\hat{\theta} - r)'[c\hat{\Sigma}c']^{-1}(c\hat{\theta} - r), \tag{5.23}$$

and it has an asymptotic distribution of $\chi^2_{(j)}$, if the restrictions are valid.

For the BF problem, the restriction is simply $\mu_1 = \mu_2$, *i.e.*, $c = \{1, \ -1, \ 0, \ 0\}$, then, the EL-type Wald test has the form of:

$$w = (\mu_1 - \mu_2)^2(a_{11} + a_{22} - 2a_{12})^{-1}, \tag{5.24}$$

where $a_{ij}, \ i, \ j = 1, 2$ are the elements of the $\hat{\Sigma}$ matrix. The EL-type Wald test statistic w has an asymptotic distribution of $\chi^2_{(1)}$ under the null hypothesis.

We have explored the EL-type Wald test for the BF problem using Monte Carlo simulations. Intuitively, the Wald test should be computationally easier than the ELR test since the Wald test involves solving the unconstrained case only. However, our study finds that both the computing time and computational difficulty associated with the Wald test are greater than that associated with the ELR test. One possible reason for this is that the Wald test involves computing the consistently estimated covariance matrix for the parameter estimators, and this involves three matrix inversions. For an ill behaved problem, such as the Behrens-Fisher problem, these matrix inversions may cause some difficulties. The results of the Wald test are not as good as that of the ELR test in the BF context. Also the computing time is longer than that of the ELR test. Therefore, we choose to present the results of the ELR test only here, although the EL-type Wald test is considered further in the related material in the next chapter.

Testing for Normality

The Behrens-Fisher problem is the case of testing the equality of two means when two data sets are independently drawn from two normal distributions with general and unknown variances. It was categorized by Tukey (1954) as the fourth level problem of *normal sequences of growth in consideration in the area of comparing the typical values of two populations with the aid of a sample drawn from each* (Tukey, 1954, p. 713).

The EL approach that we described in Section 5.3.1 provided a new solution to the Behrens-Fisher problem. In addition to this, we can conduct an ELR test for *normality* of the underlying distributions of the two data sets. As described in Chapters 2 and 3, testing for the validity of the moment conditions provides a way of testing for normality.

Consider two data sets $S_i = \{x_{i1},\ x_{i2},\ \ldots,\ x_{in_i}\}$, $i = 1,\ 2$. Suppose they are drawn independently from two populations $F(\mu_i,\ \sigma_i^2)$ of the same family, where F is not known and is not necessarily normal. Applying the EL approach (unconstrained case) described in Section 5.3.1, we get $\hat{\lambda}$, $\hat{\theta}$, and the $\hat{p}_j$'s, where $j = 1,\ 2,\ \ldots,\ n$. With the estimated probability parameters, the $\hat{p}_j$'s, we can conduct a normality test for the underlying distribution. The log empirical likelihood ratio statistic for testing for normality is of the form:

$$-2\sum \log n\hat{p}_j = 2\log(1+\hat{\lambda}'h(t_j,\ \hat{\theta})). \tag{5.25}$$

It has a limiting null distribution of $\chi^2_{(1)}$, where the number of degrees of freedom equals the number of moment equation, five, less the number of parameters, four. The normality test for the underlying distribution ties this material nicely to that of Chapter 3.

Usually, we may be interested in using the technique that is described here to test for the normality of the two underlying populations. If we are satisfied with the results and accept the null hypothesis that the underlying populations are normal, then, we can continue to test for the Behrens-Fisher problem. The first step of testing for normality introduces a pretesting situation as described by Giles and Giles (1993). It is well known that sequential testing strategies can result in size and power distortions if the tests in question are not independent of each other. This issue is not explored in this dissertation.

Alternatively, we can conduction an ELR test for normality and an EL-type Wald test for the Behrens-Fisher problem simultaneously. Using the unrestricted model and applying the EL approach described in Section 5.3.1, we obtain the EL estimator of the parameter vector, $\hat{\theta}$, and the estimators of the probability parameter, $\hat{p}_i$'s. With the $\hat{p}_i$'s, we can implement the ELR test for normality for the underlying distribution as we described in Section 3.2.1. In the meantime, without being influenced by the testing for normality, we can compute the consistent estimator of the covariance matrix of $\hat{\theta}$ to perform the EL-type Wald test as we described in Section 5.3.2 for the Behrens-Fisher problem. If we accept the null hypothesis of the first test that the underlying distribution of the two populations is normal, then the Wald test is on the firm ground. This alternative approach seems to be able to effectively avoid the pretesting issue and it may provide better results. We will explore this alternative in the future research.

5.3.3 Advantages of the EL Approach

The empirical likelihood approach described in this chapter utilizes more information from the data sets than do the approaches suggested by other authors. In particular, we exploit the obvious relationship between the two data sets that they both are drawn from normal distributions. We make use of the first five unbiased moment equations of the data sets. In addition, we utilize the empirical likelihood functions. These techniques allow the EL approach of our design to reach a higher asymptotic efficiency for estimation and higher testing power for the new solution to the BF problem. If the ELR test had a well controlled size and good power, then we would say that the ELR test is a good solution to the BF problem and the EL approach is valuable. Monte Carlo simulations that are described in the following section provide an extensive analysis of the sampling properties of the ELR test.

5.4 Monte Carlo Experiments

For the Behrens-Fisher problem, the experimental design is to apply the empirical likelihood method to test for the equality of two means without the knowledge of the variances. Two random samples are generated independently from two normal distributions: $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with the true values of the parameters $(\mu_1,\ \mu_1,\ \sigma_1^2,\ \sigma_1^2)' = (1,\ 1,\ 1,\ \sigma_2^2)'$, where σ_2^2 varies with the variance ratio parameter $\rho^2 = \sigma_2^2/\sigma_1^2$. The true value of the parameter ρ^2 changes according to the values of $\{0.1,\ 0.5,\ 1,\ 2,\ 10\}$. The sample size pair ranges from (20, 10) to (250, 125). The ratio of the sample sizes of the two data sets is kept constant, $n_1/n_2 = 2$. These settings lead to the corresponding values of $C_0 = 1/(1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_1}{n_2})$, the parameter used in the literature, to take the values of $\{0.83,\ 0.5,\ .33,\ 0.2,\ 0.048\}$ which has a good coverage of the region (0, 1). Therefore, our experiments have a good coverage of the parameter space.

The computation time in solving the Behrens-Fisher problem is much longer than that of testing for normality. Two possible reasons are (i) the BF problem is much more difficult than the problem of testing for normality; (ii) the optimal values of the empirical likelihood functions are computed twice, one for the constrained case and one for the unconstrained case. We limit the number of replications in the Monte Carlo experiment to $5,000$ instead of $10,000$, for these reasons.

For the power comparisons, the non-centrality parameter is defined as:

$$\delta = \frac{\mu_1 - \mu_2}{(\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}}. \tag{5.26}$$

The null hypothesis is true when $\delta = 0$. The alternative hypotheses we consider are the cases when the parameter δ takes the values of $\{1,\ 1.5,\ 2,\ 2.5,\ 3,\ 4\}$, other than zero. We note that, in contrast with the non-centrality parameter used in Lee and Gurland (1975), the parameter δ in our design does not depend on the sample sizes $(n_1,\ n_2)$. Therefore, we are able to disentangle the effect upon the power of the ELR test that is arising from the non-centrality parameter δ, and that from sample sizes. The log empirical likelihood ratio statistic depends on the parameter ρ^2 and the sample size ratio n_1/n_2. Thus, the size and the power of the ELR test are also functions of these parameters. However, the asymptotic distribution of the ELR test statistic does not depend on these nuisance parameters. The ELR test statistic has an asymptotic distribution of $\chi^2_{(1)}$ if H_0 is true.

5.5 The Results

Table 5.1 in the Appendix the end of this chapter illustrates the sizes and the size-adjusted critical values of the ELR test for the Behrens-Fisher problem. We have chosen to use the four nominal significance levels: 10%, 5%, 2%, 1% in order to have a larger picture of the sampling properties of the ELR test. We use α to denote the significance level hereafter.

- The actual size of the ELR test is relatively large comparing to the nominal significance level. The size has an appropriate trend to converge to the correct nominal significance level when the sample size pair increases, given the value of ρ^2. For example, the size changes from 12.14% to 10.66% when the sample pair varies from (20, 10) to (250, 125) at the nominal significance level of 5% when $\rho^2 = 0.1$. In the future work, it would be worthwhile to provide a detailed example to illustrate the size convergence.

- The size of the ELR test is not very sensitive to the changes in the sample size pair at the nominal significance level of 10%.

- The size of the test decreases as the value of the parameter ρ^2 declines, given the values of the sample size pair and the significance level. For instance, the size changes from 27.36% to 12.14% when ρ^2 moves from 10 to 0.1 at $(n_1,\ n_2) = (20,\ 10)$ and $\alpha = 5\%$. The size distortion is the worst when $\rho^2 = 10$ but it is still within the range of our expectations for the ELR test, based on other experience.

Tables 5.2 to 5.4 provide the full analysis of the power of the ELR test across a range of values for the non-centrality parameter, δ, the parameter ρ^2, and the sample size pair $(n_1,\ n_2)$. The size-adjusted critical values are used to compute the powers of the test. These values allow us to evaluate the power of the ELR test at the actual significance levels. The findings are as follows.

- First, the power increases as the parameter δ increases, given the values of ρ^2, α, and $(n_1,\ n_2)$. For example, the power increases from 25.20% to 95.46% when the parameter δ varies from 1 to 4 at $\rho^2 = 0.1$ and $\alpha = 5\%$ for the sample size pair as small as $(n_1,\ n_2) = (20,\ 10)$. These are very encouraging results. They show that the ELR test has good power properties for quite small samples.

- Second, the power of the test increases with the sample size pair. For instance, the power increases from 49.26% to 99.94% when the sample size pair increases from (20, 10) to (250, 125) at $\rho^2 = 0.1$, $\alpha = 5\%$, and $\delta = 1.5$.

- Third, the power is higher when the value of ρ^2 is farther away from unity, given the values of the parameters of $(n_1,\ n_2)$, α, and δ. For example, the power changes from 30.16% to 90.28% when ρ^2 changes from 1 to 0.1; the power increases to 66.30% when the parameter ρ^2 moves from 1 to 10, holding $\alpha = 5\%$, $\delta = 2.5$, and $(n_1,\ n_2) = (20,\ 10)$ fixed. The rationale behind this result is as follows. When $\rho^2 = 1$, the variances of the two populations are unknown but equal. The test for $H_0:\ \mu_1 = \mu_2$ reduces to the usual t test. When ρ^2 deviates from unity, the sign of the Behrens-Fisher problem shows up; the power of the ELR test becomes high. That is testing for H_0 under the alternative hypotheses when the variances of the two populations are unknown and unequal, the ELR test has high power. The result tells us that the ELR test is able to capture the information of the BF problem that the samples were drawn from two populations with different variances that are unknown.

Overall, the power results are acceptably good. The ELR test is recommended to be applied to the BF problem.

5.5.1 Comparison of the WA and the EL Tests

As discussed in Section 5.2.1, the solution provided by the Welch-Aspin test results from directly solving the size function equation

$$P(|v| > V(\hat{C})) = \alpha \tag{5.27}$$

and approximating in the terms of, and up to, f_i^{-4}, where $f_i = n_i - 1$ is the number of degrees of freedom left for each data set, and $i = 1,\ 2$. The WA test is designed for small samples. It has well controlled size and good power. In Table 5.6, we show the experimental results of the WA test from Lee and Gurland (1975). The results were specifically for the case of $(n_1,\ n_2) = (7,\ 7)$, $\alpha = 5\%$, and $C_0 = \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$. To our knowledge, there is no other published information available to allow a comparison of the ELR and the WA for $n_i > 7, i = 1,\ 2$.

Table 5.6 presents a comparison of the two tests: the Welch-Aspin (WA) test and the ELR test for one specific case. The parameter ρ^2 takes the values of $\{9,\ 4,\ 2.33,\ 1.5,\ 1\}$ in corresponding to the values of the parameter C_0 used by Lee and Gurland (1975). The non-centrality parameter, δ, used in the ELR test is different from the non-centrality parameter, δ_0, used in their paper for the WA test. These two non-centrality parameters have the following relationship: $\delta = \delta_0/\sqrt{7}$. From the table, we see that the ELR test, unfortunately, performs poorly for this very small-sample situation.

The size distortion of a test is the difference between the actual size of the test and the nominal significance level. For the sample size pair as small as $(n_1,\ n_2) = (7,\ 7)$, the actual size of the ELR test obviously exceeds the nominal significance level but it is still within the usual size range of the ELR test. As usual, we compute the power of the ELR test at the actual level of 5% by using the simulated size-adjusted critical values. The power of the ELR test for such a small sample size is clearly inferior to that of the WA test. It is very low indeed. It is unfortunate that the ELR test is not able to show its merits for this extremely small sample size pair.

The ELR test for the Behrens-Fisher problem is an asymptotic test. The power performance is acceptably good when the sample size pair is as small as $(n_1,\ n_2) = (20, 10)$. However, we would not expect it to perform very well when the sample size pair is extremely small, such as $(n_1,\ n_2) = (7,\ 7)$. The power results from Monte Carlo experiment for this specific case (but with various significance levels) are presented in Table 5.7 of the Appendix at the end of this chapter.

Although this particular comparison between the ELR test and the WA test is rather disappointing, it must be kept in perspective. First, we are comparing our asymptotic test with one which is explicitly designed for small samples. Second, this comparison involves a particularly small sample size. The full experimental results for the ELR test solving the Behrens-Fisher problem that are presented in Tables 5.2 to 5.5 of the appendix at the end of this chapter show that the EL method is able to solve the BF problem, and that the ELR test has good power properties over a wide range of realistic situations.

5.6 Computational Issues

The computing work associated with our approach to solving the BF problem is challenging. As Owen (2001) stated, it is computationally challenging to optimize a likelihood function of either parametric or empirical type over some nuisance parameters with other parameters held fixed at test values.

There are two possible reasons for the difficulty. One reason is that any solution to the Behrens-Fisher problem must have some unpleasant properties (Zamen, 1996, page 246). The behavior of the ELR test, like other tests, has some unfavorable features over some areas in the parameter space.

The second possible reason comes from the design of the empirical likelihood approach. The nature of the EL method is that, in the neighborhood of the solution, the gradient matrix associated with the moment constraints will approach an ill-behaved state of being less than full rank (Mittelhammer *et al.*, 2003). This occurs by design because the basic rationale of the EL method is to modify the sample weights such that the over-identified m empirical moment equations can be satisfied in order to solve for the unique solutions of the p unknowns, where $m > p$. This creates instability in gradient-based constrained optimization algorithms regarding the representation of the feasible spaces and feasible directions for such problems.

Mittelhammer *et al.* (2003) used a concentrating out technique which utilized the NLSYS and the Nelder-Mead algorithms to fulfill the purpose. The technique worked for their problem very well. We have also explored the possibility of using the Nelder-Mead method in solving the Behrens-Fisher problem. The Nelder-Mead method did not work in our situation, unfortunately.

The approach used in this chapter is the "direct solve" method that we discussed in computation issues in the Appendix at the end of Chapter 2. The five moment equations and the first order conditions with respect to the parameters are used simultaneously to solve the nonlinear equation system. The Eqsolve procedure in the Gauss package is employed to solve the system for the numerical solutions.

There are quite a few samples drawn from the underlying distributions that can not be used to solve the nonlinear system in computing the power of the ELR test. This is a typical example of the potential infeasibility in practise that happens in implementing the EL method.

When this happens, we reuse every sample that does not work for the Eqsolve procedure by altering the initial values of the parameters and implementing the procedure again until the solution is found. In addition to this, we apply the essential idea of the Differential Evolution method into the data reuse process by using a random search direction that is formed from the difference between two random vectors to search for the global maximum. The new initial value for the parameter vector is then in the form: $\theta_a = \hat{\theta} + s(\theta_1 - \theta_2)$, where $\hat{\theta}$ is the unsuccessfully estimated value of θ and s is a step size taking values of $\{0.4,\ 1,\ 2\}$ accordingly. These techniques work very well in keeping the samples that we have generated such that we can approximate the exact distribution in finite samples for the ELR test statistic. By doing so we effectively avoid throwing away data sets casually. Therefore, we avoid the problem of selection bias.

In addition to this, the empirical size distortion of the ELR test is effectively improved using this data reuse technique. The computing time is lengthened and it is significantly longer than the EL approach in the testing for normality case. Overall, the EL approach using the Eqsolve algorithm works for the BF problem. The results are acceptably good.

To give an indication of how difficult it is to solve the BF problem using the ELR method, we present some computing times used in the Monte Carlo experiments. It takes 10 hours to compute the empirical size for the ELR test when $(n_1,\ n_2) = (20, 10)$, $\rho^2 = 2$, and the number of replications is 5,000 using a Pentium 4, 2.4 GHZ PC. In computing the power of the ELR test, it takes the same machine 27 hours for the case of $(n_1,\ n_2) = (20, 10)$, $\rho^2 = 0.5$, the non-centrality parameter $\delta = 1$, the number of replications is 5,000. The computing time is much longer when the parameter ρ^2 is away from unity. The worst case is when $\rho^2 = 10$.

5.7 Summary and Conclusions

We have developed an new theoretical approach using the EL method to solve the Behrens-Fisher problem in this chapter. The fact that the EL method is able to solve the BF problem is important. It shows the flexibility of the EL method in solving various problems in statistics and econometrics. Second, a full range of Monte Carlo experiments are conducted to provide the sampling properties of the ELR test. The actual sizes and the size-adjusted critical values in finite samples are simulated. The size-adjusted critical values are used to conduct the analysis in the power properties of the ELR test. The empirical results provide the evidence that the ELR test has good sampling properties across different parameter dimensions: the variance ratio parameter, the sample size pair, and the non-centrality parameter.

Generally, the size-adjusted critical values that we provided in Table 5.1 are ready to be used by researchers provided that the values of the parameter ρ^2 and the sample size pair are conformable with the ones in our study.

5.7.1 Future Work

We have noted that the computing time in solving the BF problem is significantly longer than in the application of testing for normality. The size of the ELR test statistic is still large, in general. It would be interesting to explore some techniques that could reduce the size distortion and could still maintain the good power properties of the ELR test.

Appendix: Tables of the Behrens-Fisher Problem

Table 5.1: Size and Size-adjusted Critical Values of the ELR Test

$(n_1,\ n_2)$:	(20, 10)	(60, 30)	(100, 50)	(250, 125)	(20, 10)	(60, 30)	(100, 50)	(250, 125)
	$\rho^2 = 0.1$				$\rho^2 = 0.5$			
10%	0.1838	0.1868	0.1832	0.1878	0.2202	0.2202	0.2266	0.1964
5%	0.1214	0.1124	0.1084	0.1066	0.1482	0.1468	0.1448	0.1218
2%	0.0712	0.0636	0.0580	0.0554	0.0924	0.0852	0.0896	0.0668
1%	0.0488	0.0428	0.0404	0.0342	0.0686	0.0584	0.0626	0.0424
Size-adjusted Critical Values:								
10%	4.3067	4.1787	4.0088	3.9780	5.1494	4.8877	4.9112	4.3564
5%	6.5523	6.1960	5.8025	5.6186	7.7853	7.1960	7.4334	6.1392
2%	10.7818	10.1350	9.0218	8.2500	12.4479	10.4005	11.1566	9.3809
1%	14.2266	13.2669	12.8602	10.3353	16.5913	13.6268	14.1973	11.9212
	$\rho^2 = 1$				$\rho^2 = 2$			
10%	0.2424	0.2538	0.2536	0.2332	0.2926	0.3014	0.2780	0.2448
5%	0.1698	0.1788	0.1748	0.1466	0.2136	0.2110	0.1966	0.1592
2%	0.1100	0.1126	0.1068	0.0828	0.1404	0.1422	0.1298	0.0888
1%	0.0814	0.0776	0.0710	0.0530	0.1046	0.1044	0.0966	0.0570
Size-adjusted Critical Values:								
10%	5.7988	5.8435	5.5645	4.7971	6.8276	6.7529	6.4580	5.1142
5%	8.9674	7.9674	7.5631	6.7822	10.8456	9.3973	9.3711	7.1188
2%	14.3630	11.8099	10.5112	9.8197	17.7402	13.1147	13.3604	10.0495
1%	18.6082	14.3278	13.6295	12.0383	22.9714	15.9200	16.7853	12.1949
	$\rho^2 = 10$							
10%	0.3564	0.3306	0.2872	0.2376				
5%	0.2736	0.2574	0.2066	0.1576				
2%	0.1892	0.1826	0.1384	0.0948				
1%	0.1516	0.1428	0.1044	0.0636				
Size-adjusted Critical Values:								
10%	9.25930	8.4782	6.7945	5.2401				
5%	15.0960	12.3244	10.4183	7.4935				
2%	23.8243	18.3507	16.2891	11.1926				
1%	31.6749	21.8045	21.2274	14.8243				

Notes to table: The number of replications is 5,000. The sample sizes are the pair $(n_1,\ n_2)$. $\rho^2 = \sigma_2^2/\sigma_1^2$.

Table 5.2: Power of the ELR Test for the Behrens-Fisher Problem

	(n_1, n_2)											
	$\rho^2 = 0.1$						$\rho^2 = 0.5$					
δ:	1	1.5	2	2.5	3	4	1	1.5	2	2.5	3	4
	$(20, 10)$											
10%	0.3826	0.6786	0.9032	0.9684	0.9750	0.9768	0.1688	0.2945	0.4942	0.7038	0.8408	0.9508
5%	0.2520	0.4926	0.7764	0.9028	0.9388	0.9546	0.0994	0.1816	0.3101	0.4562	0.608	0.8166
2%	0.1396	0.2262	0.4122	0.5364	0.5910	0.6896	0.0542	0.1028	0.1674	0.2344	0.3078	0.5022
1%	0.0906	0.1226	0.2122	0.2836	0.3298	0.4316	0.0371	0.0743	0.1226	0.1562	0.2104	0.3314
	$(60, 30)$											
10%	0.6086	0.9314	0.9926	0.9924			0.2870	0.5748	0.9148	0.9832	0.9906	
5%	0.4800	0.8748	0.9870	0.9846			0.1800	0.4256	0.8350	0.9692	0.9848	
2%	0.2934	0.7292	0.9724	0.9764			0.1034	0.2710	0.6926	0.9376	0.9798	
1%	0.2088	0.5964	0.9530	0.9740			0.0682	0.1812	0.5422	0.8764	0.9688	
	$(100, 50)$											
10%	0.6976	0.9812	0.9978				0.3180	0.7362	0.9768	0.9962		
5%	0.5992	0.9608	0.9970				0.1956	0.5988	0.9494	0.9924		
2%	0.4452	0.9074	0.9930				0.0950	0.3978	0.8850	0.9840		
1%	0.2972	0.8092	0.9848				0.0584	0.2792	0.8032	0.9742		
	$(250, 125)$											
10%	0.8254	0.9996					0.4216	0.9450	0.9992			
5%	0.7512	0.9994					0.3130	0.9052	0.9986			
2%	0.6308	0.9972					0.1828	0.8040	0.9962			
1%	0.5474	0.9952					0.1112	0.7026	0.9928			

Notes to table: The number of replications is 5,000. The sample sizes are the pair $(n_1,\ n_2)$. $\rho^2 = \sigma_2^2/\sigma_1^2$. δ is the non-centrality parameter.

Table 5.3: Power of the ELR Test for the Behrens-Fisher Problem

	(n_1, n_2)											
	$\rho^2 = 1$						$\rho^2 = 2$					
δ:	1	1.5	2	2.5	3	4	1	1.5	2	2.5	3	4
						(20, 10)						
10%	0.1564	0.2170	0.3456	0.5116	0.6886	0.9050	0.1593	0.1956	0.2672	0.4400	0.6366	0.8352
5%	0.0972	0.1306	0.2078	0.3016	0.4634	0.7318	0.1044	0.1303	0.1750	0.3016	0.4612	0.6564
2%	0.0606	0.0898	0.1342	0.1872	0.2806	0.4624	0.0658	0.0950	0.1378	0.2354	0.3731	0.5456
1%	0.0420	0.0700	0.1112	0.1584	0.2312	0.3878	0.0440	0.0800	0.1240	0.2188	0.3522	0.5272
						(60, 30)						
10%	0.2076	0.4050	0.7574	0.9458	0.9846	0.9941	0.1926	0.289	0.6084	0.8770	0.9772	0.9968*
5%	0.1440	0.2878	0.6310	0.8962	0.9748	0.9914	0.1342	0.1804	0.4504	0.7666	0.9462	0.9942
2%	0.0870	0.1584	0.4192	0.7648	0.9374	0.9850	0.0869	0.0995	0.2876	0.6190	0.8794	0.9871
1%	0.0664	0.1144	0.3082	0.6614	0.8930	0.9798	0.0648	0.0716	0.2170	0.5232	0.8082	0.9683
						(100, 50)						
10%	0.2364	0.5486	0.9086	0.9918			0.1988	0.4190	0.8054	0.9774	0.9970	
5%	0.1572	0.4284	0.8452	0.9828			0.1303	0.2693	0.6634	0.9472	0.9950	
2%	0.0980	0.2770	0.7322	0.9612			0.0814	0.1380	0.4756	0.8756	0.9886	
1%	0.0698	0.1766	0.5894	0.9290			0.0620	0.0808	0.3459	0.7996	0.9776	
						(250, 125)						
10%	0.3104	0.8450	0.9964				0.2744	0.7362	0.9878	0.9996		
5%	0.2082	0.7514	0.9914				0.1844	0.6116	0.9746	0.9988		
2%	0.1152	0.6086	0.9738				0.1024	0.4572	0.9432	0.9982		
1%	0.0776	0.5048	0.9572				0.0682	0.3624	0.9084	0.9976		

Notes to table: The number of replications is 5,000. The sample sizes are the pair $(n_1,\ n_2)$. $\rho^2 = \sigma_2^2/\sigma_1^2$. δ is the non-centrality parameter. Results with a star (*) are the ones that the number of replications is less than 5,000 due to computing difficulties.

Table 5.4: Power of the ELR tTest for the Behrens-Fisher Problem

(n_1, n_2)	$\rho^2 = 10$					
δ:	1	1.5	2	2.5	3	4
(20, 10)						
10%	0.2056	0.2586	0.4852	0.7378	0.8340	0.8556
5%	0.1512	0.2084	0.4182	0.6630	0.7815	0.8088
2%	0.1046	0.1782	0.3714	0.6250	0.7482	0.7864
1%	0.0764	0.1572	0.3422	0.5928	0.7240	0.7712
(60, 30)						
10%	0.1900	0.2572	0.7734	0.9894	0.9970	0.9945*
5%	0.1386	0.2060	0.7144	0.9762	0.9856	0.9872
2%	0.0798	0.1594	0.6150	0.8982	0.9277	0.9287
1%	0.0600	0.1270	0.5270	0.8072	0.8543	0.8793
(100, 50)						
10%	0.1522	0.3674	0.9114	0.9995*	0.9997*	1.0000*
5%	0.0812	0.2214	0.8544	0.9987	0.9995	1.0000
2%	0.0516	0.1446	0.7945	0.9960	0.9973	1.0000
1%	0.0450	0.1296	0.7734	0.9922	0.9940	0.9968
(250, 125)						
10%	0.2388	0.6006	0.9977*			
5%	0.1276	0.4794	0.9931			
2%	0.0544	0.3242	0.9851			
1%	0.0300	0.2248	0.9747			

Notes to table: The number of replications is 5,000. The sample sizes are the pair $(n_1,\ n_2)$. $\rho^2 = \sigma_2^2/\sigma_1^2$. δ is the non-centrality parameter. Results with a star (*) are the ones that the number of replications is less than 5,000 due to computing difficulties.

Table 5.6: Size and Power Comparison of the Welch-Aspin and the ELR Tests for $(n_1, n_2) = (7, 7)$, $\alpha = .05$

δ_0	0		1		2		3		4	
C_0	V_{wa}	EL	V_{wa}	EL	V_{wa}	EL	V_{wa}	EL	V_{wa}	EL
0.1	0.0501	0.2846	0.2301	0.0522	0.5628	0.0494	0.8521	0.0594	0.9729	0.0660
0.2	0.0500	0.2306	0.2349	0.0554	0.5753	0.0582	0.8631	0.0604	0.9767	0.0874
0.3	0.0500	0.1864	0.2385	0.0640	0.5855	0.0764	0.8722	0.0886	0.9799	0.1164
0.4	0.0498	0.1608	0.2406	0.0608	0.5920	0.0794	0.8782	0.1008	0.9819	0.1486
0.5	0.0498	0.1510	0.2413	0.0762	0.5942	0.1048	0.8803	0.1380	0.9826	0.2000

Notes to table: The number of replications is 5,000. $C_0 = (\frac{\sigma_1^2}{n_1})(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})^{-1}$, $C_0 = \{.1, .2, .3, .4, .5\}$ corresponds to $\rho^2 = \{9, 4, 2.33, 1.5, 1\}$. $\delta_0 = (\mu_1 - \mu_2)(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})^{-1} = \delta\sqrt{7}$. $\delta_0 = \{1,\ 2,\ 3,\ 4\}$ is equivalent to $\delta = \{.378,\ .7559,\ 1.1339,\ 1.5119\}$

Table 5.7: Size and Power of the ELR Test for $n_1 = n_2 = 7$

ρ^2	α	$\delta = 0$	$\delta = 0.3780$	$\delta = 0.7559$	$\delta = 1.1334$	$\delta = 1.5119$
1	10%	0.2014	0.1030	0.0992	0.1266	0.1612
	5%	0.1510	0.0522	0.0494	0.0594	0.0660
	2%	0.1072	0.0184	0.0194	0.0210	0.0218
	1%	0.0850	0.0082	0.0098	0.0082	0.0070
1.5	10%	0.2062	0.1134	0.1232	0.1274	0.1770
	5%	0.1608	0.0554	0.0582	0.0604	0.0874
	2%	0.1182	0.0204	0.0224	0.0208	0.0332
	1%	0.0966	0.0106	0.0112	0.0088	0.0102
2.33	10%	0.2338	0.1184	0.1398	0.1562	0.2058
	5%	0.1864	0.0640	0.0764	0.0886	0.1164
	2%	0.1400	0.0206	0.0282	0.0276	0.0374
	1%	0.1110	0.0094	0.0158	0.0106	0.0150
4	10%	0.2838	0.1216	0.1500	0.1756	0.2672
	5%	0.2306	0.0608	0.0794	0.1008	0.1486
	2%	0.1686	0.0278	0.0370	0.0434	0.0566
	1%	0.1366	0.0110	0.0168	0.0172	0.0244
9	10%	0.3384	0.1438	0.1974	0.2652	0.3754
	5%	0.2846	0.0762	0.1048	0.1380	0.2000
	2%	0.2140	0.0242	0.0386	0.0362	0.0508
	1%	0.1760	0.0126	0.0174	0.0180	0.0240

Notes to table: The number of replications is 5,000. $C_0 = \{.1, .2, .3, .4, .5\}$ corresponds to $\rho^2 = \{9, 4, 2.33, 1.5, 1\}$. $\delta_0 = \delta\sqrt{7}$. $\delta_0 = \{1,\ 2,\ 3,\ 4\}$ is equivalent to $\delta = \{.378,\ .7559,\ 1.1339,\ 1.5119\}$

Chapter 6

Testing for Structural Change

6.1 Introduction

There has been great deal of interest in testing for the equality of regression coefficients (*i.e.*, the absence of structural change) in two linear regressions when the disturbance variances are unequal. We assume that these two linear regressions are classical regression models, so that the two error terms are independent, and each disturbance term satisfies the classical assumptions such as normality, homoscedasticity, and serial independence. The Behrens-Fisher problem that we discussed in Chapter 5 is a special case of the problem of structural change when there is only one regressor, a constant, in each of the regression models. The usual Chow test (Chow, 1960) assumes equal variances for the errors of the models. Toyoda (1974) showed that the usual Chow test of the coefficients of two regression model is misleading if the two variances are unequal and the sample sizes are small. The first test for the problem of structural change in the linear regression model when the error variance may also change was developed by Jayatissa (1977). We refer to it as the J test. The J test wisely makes use of the information in hand. The J test is an exact test and the test statistic has an exact F distribution if the null hypothesis is true. Watt (1979) and Honda (1982) proposed a Wald test for this problem and provided evidence that the Wald test is preferred to the J test for the case when the number of regressors is greater than one. The Wald test is an asymptotic test, of course. Weerahandi (1987) developed another exact test which makes use of the empirical significance level, the p-value. We refer to this test here as the WEE test. Zaman (1996) highly recommended the WEE test and discussed the test in detail since

Weerahandi's approach introduced a new idea to the econometrics testing literature.

The purpose of this chapter is as follows. First, we develop a new solution to solve the problem of testing for structural change in a linear regression model when the variance of the error term is not necessarily constant. The approach that we take is to apply the empirical likelihood method (EL). The way that we use the data allows us to utilize the most information to hand. In addition to this, the approach we use ties the estimation and testing issues together nicely. The EL approach used in this chapter also allows us to conduct tests for normality in the presence of the heteroscedasticity of the error terms in a regression model.

Second, we conduct a power comparison between the EL-type tests and several selected conventional tests that are used in testing for structural change in the regression model. Monte Carlo simulations are employed to provide the empirical size and the size-adjusted critical values in finite samples. The empirical powers of the tests are computed using the size-adjusted critical values to ensure that every test is being considered at the same actual significance level.

We develop a new approach to the problem of testing for structural change and a means of testing for normality under the heteroscedasticity in a regression model in this chapter. The outline of this chapter is as follows. Section 6.2 gives a brief outline of the existing tests that we have chosen to consider. Section 6.3 provides the set up of the EL approach that we use. Section 6.4 discusses the Monte Carlo experiment and associated results. A summary and our conclusions are provided in Section 6.5.

6.2 Tests for Structural Change under Heteroscedasticity

Our primary task is to test for the equality of the coefficient vector in two linear regressions when the disturbance variances are not known to be equal. The framework for this problem has the form:

$$Y_i = X_i\beta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2 I_{n_i}), \quad i = 1, 2 \tag{6.1}$$

where Y_i and X_i are $n_i \times 1$ and $n_i \times k$ observation matrices, β_i are $k \times 1$ coefficient vectors, and ε_i are $n_i \times 1$ error vectors. We assume that $E(\varepsilon_1 \varepsilon_2') = 0$ and that each of the regressor matrices is non-random and of full column rank. The least squares estimators of β_i are:

$$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'Y_i, \ i = 1, \ 2. \tag{6.2}$$

The least squares residual vectors are

$$\hat{\varepsilon}_i = M_i Y_i, \tag{6.3}$$

where M_i is the idempotent symmetric matrix and has the form:

$$M_i \ = \ I_{n_i} - X_i(X_i'X_i)^{-1}X_i'; i = 1, \ 2. \tag{6.4}$$

The matrix M_i can be decomposed into $Z_i Z_i'$, where Z_i is the $n_i \times (n_i - k)$ eigenvector matrix of M_i corresponding to the unit roots and have the properties: $Z_i'X_i = 0$ and $Z_i'Z_i = I_{n_i-k}$; $i = 1, \ 2$.

A type of BLUS residual vector (Theil, 1965 and 1968) is formed using the Z_i matrix:

$$\varepsilon_i^* = Z_i'\hat{\varepsilon}_i; i = 1, \ 2. \tag{6.5}$$

The BLUS residuals have the distribution: $\varepsilon_i^* \sim N(0, \sigma_i^2 I_{n_i-k})$. These residuals are independent and identically distributed if the error terms are from normal distributions.

The difference of the two least squares estimators of the β_i vectors is distributed as follows:

$$\hat{\beta}_1 - \hat{\beta}_2 \sim N(\delta, \Sigma), \tag{6.6}$$

where $\delta = \beta_1 - \beta_2$, and $\Sigma = \sigma_1^2(X_1'X_1)^{-1} + \sigma_2^2(X_2'X_2)^{-1}$. A solution to the problem of testing for structural change is just a test of the hypothesis that $H_0 : \ \beta_1 - \beta_2 = 0$ based on the estimated covariance matrices.

There are two types of solutions to this problem of testing for structural change: the exact and asymptotic tests. The Jayatissa and the Weerahandi tests are exact tests in which the exact distributions of the test statistics are known under the null hypothesis. The Wald and the empirical likelihood type tests are asymptotic ones, where the asymptotic null distributions are known but the actual null distributions in finite samples are unknown. We have chosen the Jayatissa test, the Weerahandi test, and the Wald test for comparison purposes. The reason that these tests are chosen is that they are the most commonly used tests in the econometrics literature associated with testing for this type of structural change.

6.2.1 Jayatissa Test (J)

Jayatissa (1977) proposed a test (the J test) in which the test statistic has an exact central F distribution under the null hypothesis of no structural change. The virtue of this test is that the probability of an incorrect rejection of the null hypothesis does not depend on the values of the nuisance parameters, the variances $\sigma_i^2, i = 1,\ 2$. The J test has been the corner stone and benchmark in the literature on testing regression vector equality in the presence of heteroscedasticity. The J test makes the wise use of the transformed regression residuals, ε_i^*, and the decomposition of the matrices $(X_i'X_i)^{-1} = Q_i'Q_i$ where Q_i are $k \times k$ matrices. When the numbers of observations from two regressions are not equal: $n_1 \neq n_2$, suppose $n_1 > n_2$, then, the ε_1^* vector is truncated to have a length of n_2. If n_2/k is not an integer, then, the data are truncated again in order to form the J test statistic. The criticisms of other authors arise from the fact that the J test does not use all of the data efficiently. It involves throwing away some of the data. It also lacks uniqueness, for there are different decomposition methods that could be used for $(X_i'X_i)^{-1}$. The J test requires the minimum sample size, *i.e.*, $\min((n_1 - k)/k, (n_2 - k)/k) > 1$. Watt (1979) and Honda (1982) have discussed these issues in more detail.

6.2.2 Weerahandi Test (WEE)

Weerahandi (1987) initiated a new approach to the testing procedure for the structural change problem. We refer to this test as the WEE test. The WEE test yields a particular type of exact solution to the problem. It is an exact test based on the observed level of significance, the p-value. The test is to reject the null hypothesis if the p-value is too small, for instance, smaller than a preassigned significance level. The computational work associated with the construction of the WEE test is not too difficult. It requires only an one-dimensional numerical integration over a parameter $R = \frac{\hat{\varepsilon}_1'\hat{\varepsilon}_1}{\sigma_1^2} / (\frac{\hat{\varepsilon}_1'\hat{\varepsilon}_1}{\sigma_1^2} + \frac{\hat{\varepsilon}_2'\hat{\varepsilon}_2}{\sigma_2^2})$. The parameter R has a distribution of Beta$((n_1 - k)/2,\ (n_2 - k)/2)$, under the null hypothesis. The observed significance level is obtained from the formula:

$$p = 1 - E^R(F_{k,T}(V)) \tag{6.7}$$

where $F_{k,T}$ is the cumulative distribution of F with degrees of freedom $(k,\ T)$,

$$V = \frac{T}{k}\hat{\delta}'(\frac{SSR_1}{R}(X_1'X_1)^{-1} + \frac{SSR_2}{1-R}(X_2'X_2)^{-1})^{-1}\hat{\delta}, \tag{6.8}$$

$T = n_1 + n_2 - 2k$, $\hat{\delta} = \hat{\beta}_1 - \hat{\beta}_2$, and SSR_i are the sums of squared residuals, $i = 1,\ 2$.

The WEE test performs well for small samples. The two parameters of the Beta distribution that are involved in computing the WEE test statistic depend on the sample sizes $(n_1,\ n_2)$. When the sample sizes are larger, these two parameters become large; the integration over the space for R, which is $(0,\ 1)$, yields a result very close to zero; and then the calculated p-value becomes close to one. Thus, the WEE test fails to reject any hypothesis when the sample sizes are large.

The p-value approach is useful for some problems with nuisance parameters, such as the problem of structural change with the σ_i^2 as nuisance parameters. The probability of an incorrect rejection of the null hypothesis depends on the observations and the nuisance parameters. It is not fixed in advance. To make testing on the basis of the p-value comparable to the fixed level testing, we can choose to reject the null hypothesis whenever the p-value is less than the preassigned nominal significance levels. Weerahandi's p-value approach often yields a useful and clear solution while the fixed level testing does not (Zaman, 1996, p. 247).

6.2.3 Wald Test

Watt (1979) and Honda (1982) proposed a Wald test under the inequality of the two variances. The test statistic has the form:

$$w = (\hat{\beta}_1 - \hat{\beta}_2)'(\hat{\sigma}_1^2(X_1'X_1)^{-1} + \hat{\sigma}_2^2(X_2'X_2)^{-1})^{-1}(\hat{\beta}_1 - \hat{\beta}_2) \tag{6.9}$$

where $\hat{\sigma}_i^2 = \frac{\hat{\epsilon}_i'\hat{\epsilon}_i}{n_i - k}$, $i = 1, 2$ are the usual unbiased least squares estimators of the variances of the error terms. The Wald test is obviously easy to compute and the test statistic has an asymptotic distribution of $\chi^2_{(k)}$. Watt (1979) and Honda (1982) provided comparisons of the size and the power of the Wald test and the J test. They pointed out that only when the number of regressors is one, $k = 1$, is the J test preferred to the Wald test. For $k > 1$, the Wald test always outperforms the J test in terms of higher power. The limitation of these two studies are essentially two-fold. First, when the power of the Wald test was calculated, the number of rejections was counted with reference to the asymptotic distribution of the test rather than the actual distribution of the test in finite samples. Second, Watt(1979) and Honda (1982) considered an "ad hoc $W2$" test, this being the Wald test applied at the 2.5% significance level, but used in this case to approximate a 5% level test. They provided the actual size and power results for this test.

In the Monte Carlo experiments in this chapter, we simulate the actual sizes and the size-adjusted critical values for various nominal significance levels for the Wald test, and use these critical values to compute the power of the Wald test and conduct power comparisons with other tests at the same actual significance levels. Details are provided in Section 6.4.

6.2.4 Empirical Likelihood Method in a Regression Model

The theory associated with applying the EL method to a regression model for the estimation of the coefficient vector β was established by Owen (1990 and 1991). As illustrated in Mittelhammer *et al.* (2000, p. 306), the unbiased moment equations used in the EL approach are of the form:

$$E(h(Y,\beta)) = E(X'(Y - X\beta) = 0. \tag{6.10}$$

The number of moment equations equals the number of the parameters. The solution from the equation system solves the maximization problem of the empirical likelihood with the weights $p_j = n^{-1}$, where j stands for the jth observation. The likelihood function achieves its maximum, $L(F_n) = n^{-n}$. The EL estimator of the coefficient vector β is precisely the same as the least squares estimator, since the moment equations coincide with those equations used in the least squares estimation method.

In the context of the classical linear regression model with the assumptions of homoscedasticity and a multivariate normal distribution for the error term, the least squares estimator of the coefficient vector β is $\hat{\beta}^{LS}$; it is unbiased and most efficient.

When the homoscedasticity assumption is dropped, $\hat{\beta}^{LS}$ is still unbiased but it is inefficient. The variance of $\hat{\beta}^{LS}$ is no longer consistently estimated by $(X'X)^{-1}\hat{\sigma}^2$. However, the asymptotic covariance matrix estimator using the EL approach is still asymptotically efficient even under heteroscedasticity. The EL estimated covariance matrix $\hat{\Sigma}$ of the EL estimator $\hat{\beta}$ has the form:

$$\hat{\Sigma} = [n^{-1}(X'X)^{-1}(\sum_{j=1}^{n}\hat{p}_j(y_j - x_j'\hat{\beta})^2 x_j' x_j)(X'X)^{-1}]^{-1}. \tag{6.11}$$

There is a close analogy between $\hat{\Sigma}^{-1}$ and White's (1980) heteroscedasticity-robust estimate of the covariance matrix of $\hat{\beta}^{LS}$. We can easily see that the EL method is able to capture the information associated with the possible presence of heteroscedasticity.

When the regressor matrix X is non-stochastic, the EL approach to the regression model actually becomes more complicated than when random regressors are allowed. The set of moment equations for each observation has the form:

$$h(y_j, \beta) = x_j'(y_j - x_j\beta), \quad \text{for } j = 1, \ldots, n. \tag{6.12}$$

It is unbiased, $E(h(y_j, \beta)) = 0$, but the covariance matrix of $h(y_j,\ \beta)$ varies with each observation:

$$cov(h(y_j, \beta)) = \sigma^2 x_j' x_j; j = 1,\ \ldots,\ n. \tag{6.13}$$

That is, the $h(y_j,\ \beta)$ are not identically distributed for all j.

Theorem 2 in Owen (1991) provides a solution to the situation when the data are not identically distributed. We denote: $cov(h(y_j, \beta)) = \Phi_j$, and $V_n = n^{-1}\sum_{j=1}^{n}\Phi_j$. ξ^S and ξ^L are the smallest and largest eigenvalues of V_n. The following assumptions are made:

1. $\lim_{n\to\infty} P(0 \in ch\{h(y_1,\ \beta),\ \ldots,\ h(y_n,\ \beta)\}) = 1$, where $ch\{\}$ denotes the convex hull of the data;

2. $n^{-2}\sum_{j=1}^{n} E\frac{\|h(y_j,\beta)\|^4}{\xi^{L2}} \to 0$, as $n \to \infty$;

3. $\frac{\xi^S}{\xi^L} \geq c > 0$, $\forall n \geq k$;

Under these assumptions, minus two times the log empirical likelihood ratio function,

$$-2\log R(\beta) = -2(\log L(\hat{\beta}^c) - \log L(\hat{\beta})^u), \tag{6.14}$$

has a limiting distribution of $\chi^2_{(d)}$, where d is the number of restrictions.

The assumption that the data are i.i.d. is relaxed. This relaxation is essential to handle the regression models with non-random regressors. The Lindeberg-Levy central limit theorem is replaced by the Lindeberg-Feller central limit theorem to deal with the asymptotics in this non-i.i.d. case. The largest eigenvalue of V_n is used to scale the problem. With this theory, we are able to set up the EL approach for the problem of testing for structural change in a regression model.

6.2.5 EL Approach

For the two linear regression models in the problem of structural change, the EL estimators, $\hat{\beta}_i$, of the coefficient vectors are the same as the least squares estimators. From the regression model using the $\hat{\beta}_i$'s, we obtain two least squares residual vectors: $\hat{\varepsilon}_i = Y_i - X_i\hat{\beta}_i$, and these residual vectors are distributed as: $\hat{\varepsilon}_i \sim N(0,\ M_i\sigma_i^2)$, where M_i are the idempotent matrices described in Section 6.2; $i = 1,\ 2$.

The objective of this section is to develop a EL type test for the equality of the two coefficient vectors. The null hypothesis is:

$$H_0 :\ \beta_1 = \beta_2.$$

We know that the distribution of $\hat{\delta} = \hat{\beta}_1 - \hat{\beta}_2$ is $N(0,\ \sigma_1^2(X_1'X_1)^{-1} + \sigma_2^2(X_2'X_2)^{-1})$ under the null hypothesis and under the classical assumptions for each of the two regression models. Suppose the X_i matrices are non-stochastic. Then the possible efficiency gain of the EL approach could come from the EL estimators of σ_i^2's. We hope that the EL estimators of the σ_i^2's would be more efficient than the least squares estimators given the fact that the EL approach utilizes both the likelihood functions and the information available in terms of the data distribution and the equality of two coefficient vectors.

The data at hand are the two sets of least squares residuals $\hat{\varepsilon}_i,\ i = 1,\ 2$. Since the

OLS residuals are not independently distributed, we first transform the OLS residuals $\hat{\varepsilon}_i$ into a type of BLUS residuals, $\varepsilon_i^* \sim N(0,\ \sigma_i^2 I_{n_i-k})$. The transformation used here is the one described in Section 6.2, and it is the same one used in the Jayatissa test. The Z_i are the $n_i \times (n_i - k)$ eigenvector matrices of M_i matrices corresponding to the unit roots. The BLUS residual vectors $\varepsilon_i^* = Z_i'\hat{\varepsilon}_i$ have the distribution $N(0, \sigma_i^2 I_{n_i-k})$, for $i = 1,\ 2$.

The data transformation technique that we have described in Chapter 5 will be applied here to the two sets of the residuals. We transform the residual vector ε_1^* to a vector V_1 that has the same distribution as ε_2^*; we combine the two sets of the residuals V_1 and ε_2^* to form a full set of residuals that are i.i.d.; then we apply the EL approach to the full set of residuals. The EL approach that we develop here allows us to achieve three objectives sequentially. (i) We can obtain the EL estimators of the two variance parameters that are more efficient; (ii) With these estimators, we can construct a EL-type Wald test for the structural change problem; (iii) We can conduct a ELR test for normality of the disturbance terms in the presence of possible heteroscedasticity. Testing for normality in this context ties the material nicely to the analysis in Chapters 3 and 4.

The steps associated with implementing the EL approach to the problem of testing for structural change are as follows.

Step 1. Transform the residual vector ε_1^* to have the same distribution as of the residual vector ε_2^* using the formula:

$$V_1 = \varepsilon_1^*(\rho^2)^{\frac{1}{2}} \tag{6.15}$$

where $\rho^2 = \sigma_2^2/\sigma_1^2$. Then, $V_1 \sim N(0,\ \sigma_2^2 I_{n_1-k})$. Let

$$V_2 = \varepsilon_2^*. \tag{6.16}$$

Stacking the two vectors, V_1 and V_2 on top of each other, we get the full set of the residuals $V = \{v_1,\ v_2,\ \ldots,\ v_T\}'$, where $T = n_1 + n_2 - 2k$. The residual vector V has a distribution $N(0,\ \sigma_2^2 I_T)$.

Assign a probability parameter p_j to v_j, the jth element of the residual vector V. The empirical likelihood function that is supported on the data is formed by $\prod_{j=1}^T p_j$. Maximizing the empirical likelihood function $\prod_{j=1}^T p_j$ subject to the probability constraints and the moment constraints is the conventional EL method. The Lagrangian function of the log empirical likelihood function has the form:

$$G = T^{-1}\sum_{j=1}^{T}\log p_j - \eta(\sum_{j=1}^{T} p_j - 1) - \lambda' \sum_{j=1}^{T} p_j h(v_j, \theta) \tag{6.17}$$

where $E[h(v_j, \theta)] = 0$ is the set of the first four unbiased moment equations for the residual vector V. The empirical version of $E[h(v_j, \theta)] = 0$ has the form:

$$\sum_{j=1}^{T} p_j v_j = 0 \tag{6.18}$$

$$\sum_{j=1}^{T} p_j v_j^2 - \sigma_2^2 = 0 \tag{6.19}$$

$$\sum_{j=1}^{T} p_j v_j^3 = 0 \tag{6.20}$$

$$\sum_{j=1}^{T} p_j v_j^4 - 3\sigma_2^4 = 0. \tag{6.21}$$

The parameter vector is $\theta = (\rho^2,\ \sigma_2^2)'$. The optimal value of the Lagrangian multiplier η is unity. The p_j's have the expression:

$$p_j = T^{-1}(1 + \lambda' h(v_j, \theta))^{-1},\ j = 1,\ 2,\ \ldots,\ T. \tag{6.22}$$

Note that the elements in the first portion of the vector V are functions of the parameters; the first order derivative of these elements with respect to the parameter has the form:

$$\frac{\partial v_j}{\partial \theta} = (\varepsilon_{1j}^{*}(\rho^2)^{-0.5}0.5,\ 0)', \tag{6.23}$$

for $j = 1,\ 2,\ \ldots,\ n_1 - k$. The first order conditions of the Lagrangian function with respect to the parameters have the form:

$$\sum_{j=1}^{n_1-k} p_j(\lambda_1 + 2\lambda_2\varepsilon_{1j}^{*} + 3\lambda_3\varepsilon_{1j}^{*2} + 4\lambda_4\varepsilon_{1j}^{*3})(\rho^2)^{-0.5}0.5 = 0 \tag{6.24}$$

$$\sum_{j=1}^{T} p_j(\lambda_2 + 6\lambda_4\sigma_2^2) = 0. \quad (6.25)$$

Solving the equation system of the four moment equations and the two first order conditions, we get the EL estimators $\tilde{\sigma}_2^2$, $\hat{\rho}^2$, and $\hat{\lambda}$. Then we obtain the probability parameter estimators, the $\hat{p}_j$'s, using the formula of the p_j's, for $j = 1, 2, \ldots, T$. The estimator of the parameter σ_1^2 can be recovered from: $\tilde{\sigma}_1^2 = \tilde{\sigma}_2^2/\hat{\rho}^2$.

Step 2. With the EL estimators of the variance parameter in hand, we can conduct an EL-type Wald test for structural change. The EL-type Wald test statistic has the form:

$$EL_W = (\hat{\beta}_1 - \hat{\beta}_2)'(\tilde{\sigma}_1^2(X_1'X_1)^{-1} + \tilde{\sigma}_2^2(X_2'X_2)^{-1})^{-1}(\hat{\beta}_1 - \hat{\beta}_2) \quad (6.26)$$

where $\tilde{\sigma}_i^2$ are the EL estimators of σ_i^2, $i = 1, 2$. The test statistic has an asymptotic distribution of $\chi^2_{(k)}$ under the null hypothesis.

Step 3. Test for the normality

With the estimators of the probability parameters, $\hat{p}_j$'s, we can easily set up an empirical likelihood ratio test for normality in the error terms. The log empirical likelihood ratio statistic has the form:

$$-2\log R(\hat{\theta}) = -2\sum_{j=1}^{T} \log T\hat{p}_j, \quad (6.27)$$

and it has a limiting distribution of $\chi^2_{(2)}$. This is a ELR test for normality in the context of the problem of structural change. That is, the EL approach we described above can be used to test for normality in a regression model in the presence of heteroscedasticity, and this provides a useful connection between various topics discussed in this dissertation.

In a general situation, we may be interested in using the technique that is described here to test for normality of the two underlying populations. If we are satisfied with the results and accept the null hypothesis that the underlying populations are normal, then, we can continue to test for the problem of structural change in regression. As was noted in Chapter 5, this can give rise to "preliminary test" distortions (Giles and Giles, 1993). This is an issue that we do not pursue further here. Alternatively, the settings here are similar to that in Chapter 5, we can conduct the two tests, testing for normality and structural change, simultaneously without stepping into the pretesting issue.

6.3 Monte Carlo Experiments

The design of the Monte Carlo experiments for the problem of testing for structural change is based on the regression models:

$$Y_1 = \beta_{11} + \beta_{12}x_1 + \varepsilon_1 \tag{6.28}$$

$$Y_2 = \beta_{21} + \beta_{22}x_2 + \varepsilon_2. \tag{6.29}$$

The disturbance vector ε_i has a distribution of $N(0,\ \sigma_i^2 I_{n_i})$; n_i is the number of observations for the ith model, and $i = 1,\ 2$. The error terms in the two regression models are independent with each other and they are independent with the regressor variables x_i. The variance ratio parameter of the two error terms is $\rho^2 = \sigma_2^2/\sigma_1^2$. The true value of the parameter ρ^2 changes according the values $\{0.1,\ 0.5,\ 1,\ 2,\ 10\}$. The true value of the parameter σ_1^2 equals unity, the true value of the parameter σ_2^2 varies with the parameter ρ^2.

The number of replications is chosen to be $5,000$ rather than 10,000 due to our computing capacity. The sample size pair $(n_1,\ n_2)$ ranges from (20, 10), (60, 30), (100, 50), to (250, 125). The ratio of the sample sizes is kept constant at two. The entire design is parallel to the one used in Chapter 5.

The coefficient vector for the first regression model is $\beta_1 = \begin{pmatrix} \beta_{11} \\ \beta_{12} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. $\beta_2 = \begin{pmatrix} \beta_{21} \\ \beta_{22} \end{pmatrix}$ is the coefficient vector of the second linear regression model in the problem. Under the null hypothesis, we have $\beta_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Otherwise, this vector is $\beta_2 = \begin{pmatrix} 1 \\ \beta_{22} \end{pmatrix}$, where β_{22} varies with the non-centrality parameter δ.

In computing the power of the test, the non-centrality parameter is chosen to be:

$$\delta = [(\beta_1 - \beta_2)'(\sigma_1^2(X_1'X_1)^{-1} + \sigma_2^2(X_2'X_2)^{-1})^{-1}(\beta_1 - \beta_2)]^{\frac{1}{2}}. \tag{6.30}$$

With only two regressors in our setting, this can be simplified to

$$\delta = [(1 - \beta_{22})^2 d]^{\frac{1}{2}}, \tag{6.31}$$

where

$$d = (\frac{\sigma_1^2}{\sum_{t=1}^{n_1} x_{1t}^2 - (\sum_{t=1}^{n_1} x_{1t})^2} + \frac{\sigma_2^2}{\sum_{t=1}^{n_2} x_{2t}^2 - (\sum_{t=1}^{n_2} x_{2t})^2}). \tag{6.32}$$

Thus, the parameter $\beta_{22} = 1 - \delta/d^{\frac{1}{2}}$. It changes with the non-centrality parameter, the true value of which is varied according to $\{1,\ 2,\ 3,\ 4\}$.

6.3.1 Regressor Matrix X

The regressor matrix has the form of $X = \{1,\ x\}$ in our setting. It is kept fixed in each replication. That is, X is non-stochastic. The x vector is designed to come from two different generating process.

The first case is when the x vector is generated following a stationary AR(1) process as follows:

$$x_t = \rho_1 x_{t-1} + u_t, \quad t = 1, \ldots, n \tag{6.33}$$

where $x_0 = 0$ and the u_t are distributed i.i.d. $N(0,\ 1-\rho_1^2)$. The true value of the parameter ρ_1 equals 0.5. The results are invariant to this value. We first generate the x vector of size $n + 300$, and then we discard the first 300 observations so as to eliminate the effect of x_0. We keep the remainder of the x vector fixed and partition it into the sample sizes $(n_1,\ n_2)$ as needed. We then join the x vector with a column of ones to form the regressor matrices. The regressor matrices X_1 and X_2 are non-stochastic.

The second case is when the x vector is generated from a uniform distribution: $U(0,\ 1)$. Steps corresponding to those described above are taken to ensure that the matrices X_1 and X_2 are non-stochastic. In this second setting, the true values of the elements of the β_i's are chosen to be different from the first setting. The true values of the coefficient vectors are $\beta_i = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ under the null hypothesis. Under the alternative hypothesis, the true values of the vector β_2 is $\beta_2 = \begin{pmatrix} 0 \\ \beta_{22} \end{pmatrix}$, and $\beta_{22} = 1 - \delta/d^{\frac{1}{2}}$ varies with the non-centrality parameter as was described in the first case.

In these Monte Carlo experiments, we simulate the size-adjusted critical values for the tests we consider. These critical values depend on the regressor matrices we have choose; they are not universally applicable to other situations, although the methods and tests are.

Generally, the EL-type tests are asymptotic ones, to conduct an EL-type test, one needs to use the actual significance levels to compute the powers for the test. In the future work, we would provide a small library in Gauss-code that would intake a general X matrix and output the size, size-adjusted critical values and the power of the EL test to help researchers in a practical way.

6.4 Experiment Results

The tables in the Appendix at the end of this chapter provide the complete comparisons of the sizes, the size-adjusted critical values, and the powers for the four tests: the EL_W test, the Wald test, the WEE test, and the J test, across a wide range of situations. We have chosen to use the four nominal significance levels: 10%, 5%, 2%, 1% so that we can have a larger picture of the sampling properties of the tests. As described in Section 6.3, the parameter ρ^2 takes the values of $\{0.1,\ 0.5,\ 1,\ 2,\ 10\}$ and the sample size pair increases following the pattern of $(n_1,\ n_2) = \{(20,\ 10),\ (60,\ 30),\ (100,\ 50),\ (250,\ 125)\}$.

In each replication, the same data set is used in the application of all of the four tests. The J test is an exact test with a known distribution in finite samples, but we have simulated the sizes and the size-adjusted critical values; and we use these critical values to compute the powers just to make them comparable with those for the other two asymptotic tests. For the WEE test, only the sizes are provided since the WEE test is an exact test that uses the observed significance level, the p-value, for inference purposes. The concept of the size-adjusted critical values is not applicable for the WEE test. In computing the powers for the WEE test, the observed significance levels, rather than the size-adjusted critical values, under the alternative hypotheses are used.

The first group of tables, Tables 6.1 to 6.13, provides the experiment results for the first case when the regressor x is generated following an AR(1) process, as described in the previous section. Tables 6.1 to 6.3 present the sizes and the size-adjusted critical values for the three tests: the EL_W test, the Wald test, and the J test.

The sizes of the EL_W test are slightly lower than the nominal significance levels. These sizes converge to the correct nominal levels as the sample size pair $(n_1,\ n_2)$ grows. For example, the size changes from 3.74% to 4.72% when the sample pair grows from (20, 10) to

(250, 125) at the nominal significance level of 5%, with $\rho^2 = 0.1$. The size distortion of a test is the difference between the actual size of the test and the nominal significance level. The size distortion of the EL_W test is small and it changes with the value of the parameter ρ^2; it actually grows as the parameter ρ^2 varies from 0.1 to 10. For instance, the size distortion is -1.26% at $\rho^2 = 0.1$ and it is 21.6% at $\rho^2 = 10$ for the sample pair $(n_1, n_2) = (20, 10)$ and at $\alpha = 5\%$. The size distortion is the worst when $\rho^2 = 10$. However, the size distortion is within the range of our expectation for the EL-type tests and it vanishes as the sample sizes grow.

The size of the Wald test has the same patterns as the EL_W test. The size of the test converges to the correct nominal level as the sample size pair grows. The size slightly increases with the value of the ρ^2 parameter. For the cases when $\rho^2 = 2$ and $\rho^2 = 10$, the size distortion of the test is less severe than that of the EL_W test. In another words, the size of the Wald test is better for the cases when $\rho^2 = 2$ and $\rho^2 = 10$ than is that of the EL_W test.

The size-adjusted critical values of both the EL_W and the Wald tests are associated with the regressor matrices that we use. They are not universally applicable to the situation with a different regress matrix X. To provide researchers a convenience in using the EL_W test, we would provide a small library in Gauss code that contains two procedures in the future work. One is used to take in a regressor matrix and to output the actual size and size-adjusted critical values; another is used to take in the regressor matrix and size-adjusted critical values and to output the power of the EL_W test.

The sizes of the J test and the WEE test are very close to the nominal levels. In addition, they are robust against the changes in the value of the parameter ρ^2 and changes in the sample sizes. These outcomes result from the fact that both of the J test and the WEE test are exact tests. The WEE test is not applicable for the sample pair (250, 125) for the reasons explained in Section 6.2.2.

Tables 6.4 to 6.13 provide the power comparisons for the four selected tests. The power of the three tests, the EL_W, the Wald, and the J test are computed using the respective size-adjusted values. That is, we compare the powers of the tests at the same actual levels. The power of the WEE test is computed using the observed significance level, the p-value. We reject the null hypothesis when the p-value is smaller than the same actual levels used for other tests.

The power of the EL_W test grows as the non-centrality parameter δ increases, given the sample size pair and the value of the parameter ρ^2. For example, the power grows from 16.44% to 56.82% as δ increases from 1 to 4, given $\rho^2 = 0.1$, the sample size pair $(n_1, n_2) = (20, 10)$, and at the actual size level of $\alpha = 5\%$. The power of the test increases as the sample size grows, given the the values of δ and ρ^2. For example, the power increases from 56.82% to 70.5% when (n_1, n_2) increases from (20, 10) to (250, 125), given $\delta = 4$, $\rho^2 = 0.1$, and at the actual size level of $\alpha = 5\%$. The power of the test changes very little as the parameter ρ^2 varies, given the sample size pair, the non-centrality parameter, and the actual size level.

The power performances of the four tests are very close. We can barely say which one is better than another. For the case when $(n_1, n_2) = (20, 10)$ and $\rho^2 = 0.1$, the power comparison in descending order is: the Wald test, the WEE test, the EL_W test, and the J test. For the case of larger sample size pair $(n_1, n_2) = (250, 125)$ with $\rho^2 = 0.1$, the order is: the Wald test, the EL_W test, and the J test. The WEE test is not applicable in the cases whenever $(n_1, n_2) = (250, 125)$. The order of the power for the tests changes when the parameter ρ^2 varies, but the difference among them is minor. All the four tests have good power properties in testing for the problem of structural change in the regression model.

Table 6.14 to 6.17 provide the sizes and the size-adjusted critical values for the four tests when the regressor x is obtained from a Uniform distribution and kept fixed in each replication. The values of the parameter ρ^2 and the sample size pair (n_1, n_2) change in the same way as described in the first case. Table 6.18 to 6.27 illustrate the power comparisons for the tests of the cases when the non-centrality parameter δ varies. The comparison results are similar to those in the first case.

6.5 Summary and Conclusions

We have applied the empirical likelihood method to the problem of testing for absence of structural change in a regression model in this chapter. First, we have illustrated the flexibility of the EL approach that the EL method can be applied to different problems in econometrics. Second, the Monte Carlo simulations and the results for the detailed power comparisons among the four tests indicate that the EL-type test is as powerful as other conventional tests. The comparisons are made across the full dimensions of the parameter space, including different designs for the regressor matrices. Overall, the EL approach provides good results and the EL-type tests have good power properties.

Appendix: Tables of Structural Change

Table 6.1: Size and Size Adjusted Critical Values for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.1$				$\rho^2 = 0.5$			
$(n_1,\ n_2)$	(20, 10)	(60, 30)	(100, 50)	(250, 125)	(20, 10)	(60, 30)	(100, 50)	(250, 125)
EL-W test at nominal levels:								
10%	0.0730	0.0730	0.0704	0.0840	0.0858	0.0980	0.0888	0.0972
5%	0.0374	0.0383	0.0371	0.0416	0.0462	0.0552	0.0448	0.0480
2%	0.0180	0.0160	0.0160	0.0154	0.0228	0.0240	0.0154	0.0208
1%	0.0110	0.0092	0.0090	0.0090	0.0128	0.0130	0.0080	0.0104
Size-Adjusted Critical Values:								
10%	3.9241	4.0676	4.0206	4.3125	4.3119	4.5473	4.3753	4.5177
5%	5.4822	5.3601	5.3769	5.6873	5.7367	6.1951	5.7596	5.9109
2%	7.6182	7.2349	7.2311	7.3603	8.0932	8.2175	7.3601	7.9367
1%	9.3748	8.8698	8.7977	9.1190	10.185	10.005	8.6382	9.2574
Wald test :								
10%	0.1280	0.1090	0.1032	0.1040	0.1283	0.1090	0.0995	0.0994
5%	0.0728	0.0620	0.0522	0.0526	0.0728	0.0584	0.0509	0.0490
2%	0.0370	0.0252	0.0231	0.0225	0.0342	0.0294	0.0194	0.0216
1%	0.0214	0.0124	0.0138	0.0126	0.0190	0.0180	0.0110	0.0100
Size-Adjusted Critical Values:								
10%	5.2138	4.8003	4.6500	4.7040	5.2206	4.7854	4.5933	4.5964
5%	7.0280	6.4164	6.1223	6.0811	6.9428	6.4037	6.0578	5.9422
2%	9.3464	8.3834	8.3975	7.9536	9.0290	8.9807	7.7791	8.0267
1%	11.1943	10.0002	9.9763	9.7063	10.8544	10.451	9.3119	9.2080
J test :								
10%	0.0982	0.1070	0.0985	0.1056	0.0984	0.1042	0.1010	0.0927
5%	0.0480	0.0488	0.0478	0.0522	0.0492	0.0562	0.0488	0.0466
2%	0.0200	0.0216	0.0218	0.0222	0.0194	0.0240	0.0180	0.0220
1%	0.0076	0.0092	0.0102	0.0118	0.0108	0.0150	0.0088	0.0112
Size-Adjusted Critical Values:								
10%	5.3889	2.8113	2.5242	2.4551	5.3855	2.7956	2.5564	2.3237
5%	9.3401	3.7716	3.3628	3.1857	9.5194	4.0153	3.3741	3.1102
2%	18.475	5.4456	4.7514	4.319	18.4115	5.7280	4.5575	4.2390
1%	27.3093	6.4371	5.6666	5.1832	32.1514	7.3551	5.4789	5.0438
WEE test :								
10%	0.0948	0.0992	0.0961	–	0.0859	0.0960	0.0930	–
5%	0.0463	0.0516	0.0488	–	0.0386	0.0496	0.0463	–
2%	0.0162	0.0200	0.0216	–	0.0122	0.0225	0.0162	–
1%	0.0072	0.0102	0.0112	–	0.0060	0.0124	0.0078	–
Computing time in minutes								

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.2: Size and Size-Adjusted Critical Values for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 1$				$\rho^2 = 2$			
Sample size:	(20, 10)	(60, 30)	(100, 50)	(250, 125)	(20, 10)	(60, 30)	(100, 50)	(250, 125)
EL-W test at nominal levels:								
10%	0.1196	0.1240	0.1090	0.1168	0.1638	0.1584	0.1318	0.1308
5%	0.0651	0.0724	0.0598	0.0620	0.1108	0.0960	0.0796	0.0746
2%	0.0325	0.0352	0.0278	0.0276	0.0662	0.0462	0.0360	0.0398
1%	0.0200	0.0196	0.0160	0.0148	0.0446	0.0292	0.0188	0.0225
Size-Adjusted Critical Values:								
10%	4.9605	5.1270	4.8045	4.9273	6.3331	5.8789	5.3248	5.3178
5%	6.7269	6.9166	6.4057	6.4209	8.7909	7.6497	7.1144	7.0886
2%	9.1073	9.1581	8.4132	8.6556	12.1633	10.4332	9.1453	9.5255
1%	11.360	10.9984	10.5341	10.1076	14.7807	12.5284	10.4337	11.0906
Wald test :								
10%	0.1258	0.1114	0.1022	0.1044	0.1368	0.1206	0.1086	0.1072
5%	0.0734	0.0606	0.0566	0.0538	0.0834	0.0644	0.0578	0.0592
2%	0.0364	0.0284	0.0242	0.0231	0.0448	0.0292	0.0228	0.0282
1%	0.0208	0.0152	0.0148	0.0118	0.0298	0.0172	0.0102	0.0132
Size-Adjusted Critical Values:								
10%	5.1752	4.8401	4.6440	4.6913	5.4119	5.0095	4.7581	4.7812
5%	7.0252	6.4137	6.2856	6.1137	7.5290	6.5304	6.2902	6.4230
2%	9.2898	8.5605	8.3555	8.0896	10.4506	8.9098	8.0356	8.4791
1%	11.0669	10.0879	10.0416	9.6211	13.3108	10.7222	9.2518	10.0419
J test:								
10%	0.1014	0.1096	0.1000	0.1052	0.0964	0.1036	0.1000	0.1046
5%	0.0458	0.0550	0.0526	0.0540	0.0500	0.0548	0.0518	0.0584
2%	0.0191	0.0224	0.0222	0.0228	0.0202	0.0222	0.0218	0.0256
1%	0.0084	0.0106	0.0095	0.0106	0.0118	0.0108	0.0094	0.0138
Size-Adjusted Critical Values:								
10%	5.5241	2.8717	2.5468	2.4589	5.2738	2.8077	2.5473	2.4564
5%	9.0994	3.9826	3.5018	3.2342	9.5241	3.9280	3.4469	3.2940
2%	18.0297	5.5664	4.7575	4.3470	19.0044	5.5976	4.7747	4.3624
1%	26.4114	6.8010	5.6176	5.0388	35.5952	6.8281	5.5962	5.2980
WEE test :								
10%	0.0806	0.0972	0.0960	–	0.0878	0.1028	0.0985	–
5%	0.0386	0.0496	0.0504	–	0.0422	0.0504	0.0504	–
2%	0.0126	0.0202	0.0208	–	0.0156	0.0224	0.0172	–
1%	0.0056	0.0094	0.0110	–	0.0082	0.0108	0.0080	–

Notes to the table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1, 1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.3: Size and Size-Adjusted Critical Values for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 10$			
Sample size:	(20, 10)	(60, 30)	(100, 50)	(250, 125)
EL-W test at nominal levels:				
10%	0.2960	0.2046	0.1754	0.1544
5%	0.2174	0.1330	0.1092	0.0890
2%	0.1442	0.0822	0.0598	0.0454
1%	0.1070	0.0546	0.0378	0.0254
Size-Adjusted Critical Values:				
10%	9.5921	7.0449	6.2411	5.7256
5%	14.1282	9.5328	8.4754	7.4453
2%	20.9046	13.7238	11.4052	9.9942
1%	26.0257	16.1886	13.7603	11.8519
Wald test :				
10%	0.1542	0.1172	0.1074	0.1086
5%	0.0952	0.0670	0.0558	0.0572
2%	0.0562	0.0324	0.0272	0.0216
1%	0.0396	0.0196	0.0152	0.0098
Size-Adjusted Critical Values:				
10%	5.8456	4.9903	4.7658	4.8786
5%	8.2013	6.7091	6.2682	6.2549
2%	11.5971	9.1366	8.6370	8.0732
1%	14.5764	10.845	9.8575	9.1729
J test:				
10%	0.0948	0.1026	0.1000	0.1060
5%	0.0488	0.0508	0.0482	0.0532
2%	0.0191	0.0222	0.0176	0.0222
1%	0.0094	0.0126	0.0078	0.0088
Size-Adjusted Critical Values:				
10%	5.2953	2.7676	2.5447	2.4775
5%	9.3683	3.8351	3.3853	3.2253
2%	18.1224	5.5513	4.5569	4.3196
1%	29.8879	7.2861	5.2979	4.8404
WEE test :				
10%	0.0938	0.1006	0.0978	–
5%	0.0466	0.0512	0.0488	–
2%	0.0170	0.0218	0.0228	–
1%	0.0086	0.0102	0.0088	–

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.4: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.1$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2528	0.4384	0.5802	0.6984	0.3154	0.5114	0.6682	0.7808
5%	0.1644	0.3006	0.4372	0.5682	0.2092	0.3788	0.5352	0.6792
2%	0.0948	0.1780	0.3008	0.4074	0.1180	0.2410	0.3664	0.523
1%	0.0616	0.1222	0.2198	0.3074	0.0688	0.1608	0.2664	0.3978
Wald test :								
10%	0.2890	0.5016	0.6438	0.7688	0.3184	0.5266	0.6924	0.8004
5%	0.1812	0.3602	0.5062	0.6480	0.2124	0.3880	0.5522	0.6954
2%	0.0966	0.2322	0.3660	0.5074	0.1262	0.2662	0.3992	0.5562
1%	0.0674	0.1702	0.2756	0.4060	0.0792	0.1852	0.3000	0.4474
J test :								
10%	0.2150	0.3217	0.4218	0.5224	0.2940	0.4758	0.6200	0.7382
5%	0.1146	0.1952	0.2524	0.3234	0.1932	0.3486	0.4818	0.6162
2%	0.0458	0.0910	0.1170	0.1646	0.0976	0.1974	0.3036	0.4348
1%	0.0286	0.0522	0.0702	0.1019	0.0678	0.1456	0.2298	0.3396
WEE test :								
10%	0.3022	0.4940	0.6478	0.7516	0.3096	0.5168	0.6954	0.7976
5%	0.1942	0.3526	0.5092	0.6310	0.2034	0.3852	0.5646	0.6932
2%	0.1034	0.2164	0.3370	0.4704	0.1130	0.2592	0.4140	0.5498
1%	0.0640	0.1434	0.2508	0.3534	0.0709	0.1834	0.3166	0.4412
Computing time in minutes								

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.5: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.1$							
$(n_1,\ n_2)$	(100, 50)				(250, 150)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3366	0.5145	0.6778	0.7986	0.3258	0.5322	0.6824	0.8014
5%	0.2166	0.3822	0.5452	0.6898	0.2114	0.3938	0.5578	0.7050
2%	0.1134	0.2454	0.3836	0.5394	0.1278	0.2676	0.4182	0.5716
1%	0.0672	0.1620	0.2878	0.4138	0.0708	0.1810	0.3026	0.4408
Wald test :								
10%	0.3466	0.5302	0.6956	0.8136	0.3232	0.5397	0.6904	0.8070
5%	0.2272	0.4016	0.5732	0.7072	0.2188	0.4082	0.5696	0.7148
2%	0.1156	0.2472	0.4024	0.5506	0.1298	0.2764	0.4302	0.5832
1%	0.0678	0.1714	0.3002	0.4438	0.0796	0.1862	0.3222	0.4600
J test:								
10%	0.3286	0.4994	0.6574	0.7808	0.3104	0.5192	0.6742	0.7916
5%	0.2142	0.3684	0.5288	0.6646	0.2148	0.3960	0.5510	0.6929
2%	0.1090	0.2178	0.3492	0.4852	0.1130	0.2452	0.3926	0.5334
1%	0.0704	0.1540	0.2674	0.3796	0.0720	0.1674	0.2965	0.4274
WEE test :								
10%	0.3192	0.5430	0.6856	0.8058	–	–	–	–
5%	0.2146	0.4134	0.5740	0.7082	–	–	–	–
2%	0.1198	0.2723	0.4194	0.5679	–	–	–	–
1%	0.0758	0.1932	0.3217	0.4700	–	–	–	–

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.6: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.5$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2918	0.4362	0.5921	0.7054	0.3230	0.5135	0.6848	0.8016
5%	0.1988	0.3182	0.4584	0.5814	0.2064	0.3640	0.5402	0.6866
2%	0.1016	0.1876	0.3020	0.4168	0.1130	0.2368	0.3884	0.5326
1%	0.0604	0.1206	0.2084	0.3044	0.0662	0.1544	0.2808	0.4128
Wald test :								
10%	0.3192	0.4842	0.6542	0.7622	0.3248	0.5175	0.6894	0.8058
5%	0.2022	0.3478	0.5158	0.6444	0.2116	0.3794	0.5530	0.7018
2%	0.1180	0.2348	0.3708	0.5102	0.1056	0.2164	0.3736	0.5224
1%	0.0728	0.1642	0.2826	0.4088	0.0660	0.1554	0.2846	0.4214
J test:								
10%	0.2326	0.3226	0.4292	0.5130	0.3000	0.4672	0.6298	0.7510
5%	0.1224	0.1770	0.2516	0.3196	0.1772	0.3084	0.4616	0.5896
2%	0.0516	0.0834	0.1214	0.1572	0.0878	0.1696	0.2738	0.4028
1%	0.0210	0.0370	0.0606	0.0782	0.0500	0.0994	0.1794	0.2752
WEE test :								
10%	0.2700	0.4812	0.6212	0.7392	0.3222	0.5185	0.6770	0.8018
5%	0.1660	0.3390	0.4756	0.6018	0.2120	0.3850	0.5457	0.6888
2%	0.0792	0.1900	0.3118	0.4324	0.1144	0.2534	0.3872	0.5340
1%	0.0428	0.1254	0.2132	0.3236	0.0732	0.1758	0.2910	0.4272

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.7: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.5$							
$(n_1,\ n_2)$	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3230	0.5290	0.6924	0.7956	0.3406	0.5368	0.6952	0.8234
5%	0.2156	0.4064	0.5674	0.6944	0.2340	0.4102	0.5726	0.7310
2%	0.1312	0.2770	0.4304	0.5662	0.1254	0.2602	0.4162	0.5816
1%	0.0882	0.1966	0.3412	0.4726	0.0866	0.1958	0.3292	0.4848
Wald test :								
10%	0.3322	0.5407	0.7042	0.8098	0.3430	0.5400	0.6954	0.8266
5%	0.2238	0.4154	0.5848	0.7094	0.2384	0.4158	0.5812	0.7348
2%	0.1358	0.2868	0.4436	0.5788	0.1258	0.2642	0.4214	0.5878
1%	0.0866	0.1960	0.3442	0.4782	0.0903	0.2072	0.3428	0.5030
J test:								
10%	0.3056	0.4996	0.6566	0.7680	0.3410	0.5387	0.6939	0.8238
5%	0.2058	0.3796	0.5286	0.6614	0.2264	0.4006	0.5584	0.7138
2%	0.1130	0.2410	0.3738	0.5082	0.1192	0.2532	0.3936	0.5496
1%	0.0736	0.1692	0.2864	0.4082	0.0767	0.1776	0.3026	0.4440
WEE test :								
10%	0.3368	0.5336	0.6896	0.8020	–	–	–	–
5%	0.2260	0.3996	0.5698	0.6996	–	–	–	–
2%	0.1300	0.2662	0.4188	0.5604	–	–	–	–
1%	0.0806	0.1874	0.3217	0.4626	–	–	–	–

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.8: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 1$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2914	0.4582	0.6004	0.7154	0.3197	0.5158	0.6802	0.7796
5%	0.1876	0.3230	0.4616	0.5762	0.1981	0.3724	0.5374	0.6667
2%	0.1044	0.1991	0.3120	0.4216	0.1080	0.2374	0.3886	0.5228
1%	0.0596	0.1318	0.2090	0.3050	0.0626	0.1578	0.2868	0.4136
Wald test :								
10%	0.3170	0.5018	0.6652	0.7675	0.3270	0.5264	0.6852	0.7884
5%	0.2058	0.3595	0.5244	0.6400	0.2120	0.3904	0.5552	0.6818
2%	0.1228	0.2366	0.3731	0.4966	0.1114	0.2428	0.3982	0.5312
1%	0.0808	0.1708	0.2824	0.3978	0.0716	0.1728	0.3044	0.4382
J test:								
10%	0.2178	0.3134	0.4306	0.5038	0.2800	0.4726	0.6153	0.7250
5%	0.1324	0.1845	0.2768	0.3368	0.1719	0.3156	0.4594	0.5814
2%	0.0562	0.0792	0.1292	0.1668	0.0906	0.1864	0.2985	0.4018
1%	0.0366	0.0484	0.0842	0.1094	0.0560	0.1224	0.2098	0.2962
WEE test :								
10%	0.2708	0.4652	0.6213	0.7326	0.3138	0.5076	0.6818	0.7886
5%	0.1690	0.3207	0.4766	0.5852	0.2054	0.3778	0.5467	0.6854
2%	0.0842	0.1844	0.3076	0.4180	0.1206	0.2484	0.3872	0.5380
1%	0.0494	0.1144	0.2046	0.3034	0.0728	0.1724	0.2868	0.4340

Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.9: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 1$							
(n_1, n_2)	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3380	0.5394	0.6850	0.7907	0.3178	0.5185	0.6942	0.7950
5%	0.2196	0.3996	0.5554	0.6770	0.2088	0.3934	0.5754	0.6998
2%	0.1268	0.2600	0.4078	0.5306	0.1080	0.2534	0.4006	0.5437
1%	0.0696	0.1535	0.2854	0.3963	0.0738	0.1820	0.3078	0.4464
Wald test :								
10%	0.3502	0.5544	0.6986	0.8074	0.3250	0.5264	0.7052	0.8020
5%	0.2248	0.4110	0.5674	0.6908	0.2124	0.4006	0.5874	0.7072
2%	0.1312	0.2660	0.4160	0.5400	0.1156	0.2660	0.4220	0.5570
1%	0.0766	0.1732	0.3158	0.4300	0.0746	0.1865	0.3206	0.4566
J test:								
10%	0.3268	0.5215	0.6632	0.7658	0.3081	0.5090	0.6802	0.7854
5%	0.2064	0.3718	0.5248	0.6350	0.2008	0.3758	0.5492	0.6814
2%	0.1116	0.2306	0.3605	0.4756	0.1012	0.2486	0.3956	0.5225
1%	0.0728	0.1634	0.2744	0.3744	0.0728	0.1806	0.3110	0.4318
WEE test :								
10%	0.3252	0.5180	0.6818	0.8054	–	–	–	–
5%	0.2252	0.3886	0.5584	0.7028	–	–	–	–
2%	0.1280	0.2530	0.4192	0.5629	–	–	–	–
1%	0.0835	0.1766	0.3120	0.4578	–	–	–	–

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.10: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 2$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2630	0.4188	0.5500	0.6788	0.3000	0.4960	0.6344	0.7712
5%	0.1590	0.2778	0.3966	0.5292	0.1991	0.3741	0.5162	0.6656
2%	0.0754	0.1580	0.2406	0.3524	0.1080	0.2376	0.3654	0.5080
1%	0.0460	0.0990	0.1610	0.2504	0.0674	0.1584	0.2723	0.4069
Wald test :								
10%	0.3138	0.4916	0.6375	0.7622	0.3090	0.5132	0.6518	0.7870
5%	0.1872	0.3348	0.4756	0.6156	0.2060	0.3844	0.5334	0.6766
2%	0.0938	0.1996	0.3042	0.4430	0.1076	0.2370	0.3706	0.5162
1%	0.0482	0.1120	0.1922	0.3062	0.0638	0.1596	0.2688	0.4058
J test:								
10%	0.2260	0.3396	0.4290	0.5308	0.2914	0.4726	0.6098	0.7392
5%	0.1130	0.1872	0.2436	0.3206	0.1780	0.3260	0.4516	0.5921
2%	0.0478	0.0774	0.1106	0.1526	0.0910	0.1865	0.2814	0.4122
1%	0.0218	0.0344	0.0526	0.0694	0.0528	0.1290	0.2006	0.3090
WEE test :								
10%	0.2728	0.4596	0.6072	0.7332	0.3258	0.5178	0.6812	0.7910
5%	0.1592	0.3154	0.4568	0.5914	0.2218	0.3834	0.5542	0.6796
2%	0.0812	0.1752	0.2944	0.4114	0.1222	0.2396	0.3970	0.5264
1%	0.0468	0.1070	0.1956	0.2965	0.0734	0.1666	0.3094	0.4196

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.11: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 2$							
$(n_1,\ n_2)$	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3120	0.5128	0.6724	0.7947	0.3086	0.5110	0.6844	0.7802
5%	0.1988	0.3714	0.5400	0.6792	0.1928	0.3756	0.5406	0.6647
2%	0.1244	0.2488	0.3962	0.5464	0.0956	0.2372	0.3766	0.5072
1%	0.0886	0.1892	0.3207	0.4676	0.0634	0.1652	0.2926	0.4198
Wald test :								
10%	0.3272	0.5340	0.6934	0.8108	0.3187	0.5195	0.6934	0.7902
5%	0.2148	0.3898	0.5668	0.7060	0.1946	0.3806	0.5524	0.6774
2%	0.1342	0.2718	0.4300	0.5834	0.1038	0.2444	0.3943	0.5298
1%	0.0970	0.2110	0.3449	0.4960	0.0622	0.1676	0.3018	0.4286
J test:								
10%	0.3120	0.5088	0.6640	0.7828	0.3106	0.5108	0.6808	0.7804
5%	0.2008	0.3696	0.5188	0.6610	0.1948	0.3736	0.5376	0.6596
2%	0.1058	0.2268	0.3536	0.4868	0.1036	0.2346	0.3906	0.5164
1%	0.0748	0.1628	0.2760	0.4022	0.0624	0.1534	0.2800	0.4036
WEE test :								
10%	0.3246	0.5098	0.6872	0.7924	–	–	–	–
5%	0.2140	0.3776	0.5628	0.6874	–	–	–	–
2%	0.1168	0.2520	0.4136	0.5484	–	–	–	–
1%	0.0720	0.1810	0.3144	0.4452	–	–	–	–

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.12: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 10$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2698	0.4334	0.5852	0.6822	0.3066	0.5002	0.6350	0.7432
5%	0.1578	0.2788	0.4026	0.5074	0.2002	0.3750	0.5008	0.6203
2%	0.0709	0.1496	0.2312	0.3187	0.0980	0.2210	0.3248	0.4314
1%	0.0386	0.0926	0.1506	0.2198	0.0596	0.1548	0.2452	0.3434
Wald test :								
10%	0.2910	0.4588	0.6236	0.7348	0.3236	0.5292	0.6692	0.7868
5%	0.1830	0.3250	0.4634	0.5748	0.2116	0.3920	0.5346	0.6690
2%	0.0984	0.1968	0.3026	0.4060	0.1084	0.2442	0.3720	0.4994
1%	0.0556	0.1276	0.2026	0.2954	0.0654	0.1702	0.2770	0.4008
J test:								
10%	0.2206	0.3330	0.4366	0.5222	0.3032	0.4910	0.6336	0.7403
5%	0.1162	0.1920	0.2662	0.3164	0.1954	0.3502	0.4784	0.5986
2%	0.0482	0.0900	0.1283	0.1528	0.0893	0.1950	0.2942	0.4002
1%	0.0254	0.0462	0.0664	0.0826	0.0476	0.1086	0.1874	0.2652
WEE test :								
10%	0.2670	0.4568	0.5846	0.7034	0.3126	0.5038	0.6686	0.7866
5%	0.1613	0.3020	0.4346	0.5588	0.2082	0.3642	0.5380	0.6762
2%	0.0774	0.1706	0.2684	0.3662	0.1150	0.2304	0.3832	0.5198
1%	0.0468	0.1070	0.1780	0.2484	0.0748	0.1612	0.2838	0.4114

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.13: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 10$							
(n_1, n_2)	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3110	0.4950	0.6604	0.7710	0.3120	0.5148	0.6664	0.7776
5%	0.1976	0.3584	0.5188	0.6524	0.2114	0.4004	0.5514	0.6746
2%	0.1064	0.2300	0.3574	0.4944	0.1224	0.2610	0.3956	0.5188
1%	0.0634	0.1572	0.2728	0.3850	0.0806	0.1896	0.3060	0.4222
Wald test :								
10%	0.3282	0.5238	0.6948	0.8034	0.3140	0.5185	0.6776	0.7922
5%	0.2180	0.3918	0.5704	0.6969	0.2116	0.4069	0.5672	0.6934
2%	0.1104	0.2474	0.3980	0.5330	0.1302	0.2842	0.4298	0.5560
1%	0.0806	0.1892	0.3237	0.4538	0.0912	0.2240	0.3570	0.4840
J test:								
10%	0.3134	0.4940	0.6684	0.7762	0.3172	0.5188	0.6744	0.7856
5%	0.2034	0.3646	0.5312	0.6637	0.2084	0.3948	0.5500	0.6744
2%	0.1136	0.2384	0.3872	0.5058	0.1160	0.2617	0.4042	0.5152
1%	0.0772	0.1776	0.3114	0.4252	0.0872	0.2068	0.3340	0.4566
WEE test :								
10%	0.3282	0.5258	0.6864	0.8042	–	–	–	–
5%	0.2212	0.3916	0.5544	0.6994	–	–	–	–
2%	0.1258	0.2610	0.4008	0.5508	–	–	–	–
1%	0.0804	0.1850	0.2970	0.4422	–	–	–	–

Notes to table: Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$, under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.14: Size and Size-Adjusted Critical Values for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.1$				$\rho^2 = 0.5$			
(n_1, n_2)	(20, 10)	(60, 30)	(100, 50)	(250, 125)	(20, 10)	(60, 30)	(100, 50)	(250, 125)
EL-W test at nominal levels:								
10%	0.0660	0.0718	0.0782	0.0916	0.1102	0.0902	0.0978	0.1014
5%	0.0374	0.0354	0.0383	0.0472	0.0648	0.0420	0.0470	0.0480
2%	0.0180	0.0120	0.0150	0.0196	0.0308	0.0172	0.0214	0.0196
1%	0.0122	0.0047	0.0060	0.0088	0.0190	0.0090	0.0106	0.0108
Size-Adjusted Critical Values:								
10%	3.7750	3.8969	4.1094	4.4349	4.8774	4.4135	4.5564	4.6314
5%	5.2994	5.3515	5.4618	5.8975	6.6703	5.6767	5.8985	5.9091
2%	7.4650	6.9453	7.1417	7.7820	8.9708	7.4310	7.9519	7.7614
1%	9.7809	7.9671	8.4766	8.9547	10.9034	8.9247	9.2294	9.3805
Wald test :								
10%	0.1198	0.1080	0.1062	0.1104	0.1180	0.1036	0.1038	0.1019
5%	0.0674	0.0564	0.0542	0.0594	0.0672	0.0534	0.0520	0.0480
2%	0.0316	0.0231	0.0212	0.0246	0.0342	0.0208	0.0216	0.0206
1%	0.0196	0.0122	0.0114	0.0138	0.0196	0.0108	0.0106	0.0102
Size-Adjusted Critical Values:								
10%	4.9596	4.7238	4.6984	4.8243	4.9918	4.6876	4.6647	4.6487
5%	6.7220	6.2828	6.1744	6.3649	6.7224	6.1036	6.0306	5.9311
2%	9.1258	8.0539	8.0013	8.2947	9.1083	7.9028	8.1867	7.8829
1%	10.7119	9.6299	9.4989	9.8070	10.8487	9.2949	9.3284	9.2365
J test:								
10%	0.0978	0.0998	0.0984	0.1116	0.0970	0.0992	0.1036	0.0984
5%	0.0504	0.0488	0.0497	0.0602	0.0516	0.0450	0.0490	0.0484
2%	0.0176	0.0196	0.0185	0.0242	0.0182	0.0158	0.0182	0.0198
1%	0.0094	0.0086	0.0098	0.0132	0.0090	0.0074	0.0076	0.0100
Size-Adjusted Critical Values:								
10%	5.3412	2.7328	2.5295	2.5358	5.2671	2.7255	2.5849	2.3612
5%	9.6550	3.7804	3.3983	3.3486	9.6543	3.7144	3.3958	3.1068
2%	17.3237	5.2890	4.5679	4.3895	17.2202	4.8989	4.5007	4.1591
1%	29.0342	6.4562	5.5921	5.2787	28.3156	6.1546	5.4440	4.9743
WEE test :								
10%	0.0842	0.0950	0.0988	–	0.0764	0.0916	0.0958	–
5%	0.0383	0.0488	0.0494	–	0.0346	0.0429	0.0462	–
2%	0.0134	0.0170	0.0188	–	0.0115	0.0144	0.0200	–
1%	0.0068	0.0092	0.0090	–	0.0050	0.0076	0.0082	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.15: Size and Size-Adjusted Critical Values for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 1$				$\rho^2 = 2$			
(n_1, n_2)	(20, 10)	(60, 30)	(100, 50)	(250, 125)	(20, 10)	(60, 30)	(100, 50)	(250, 125)
EL-W test at nominal levels:								
10%	0.1686	0.1250	0.1208	0.1140	0.2184	0.1468	0.1424	0.1296
5%	0.1056	0.0662	0.0634	0.0596	0.1486	0.0859	0.0859	0.0714
2%	0.0630	0.0278	0.0298	0.0258	0.0940	0.0438	0.0404	0.0325
1%	0.0429	0.0144	0.0156	0.0138	0.0686	0.0256	0.0254	0.0176
Size-Adjusted Critical Values:								
10%	6.1735	5.0768	4.9858	4.8674	7.5273	5.5747	5.5483	5.2841
5%	8.6057	6.5641	6.5777	6.4220	10.7483	7.5126	7.2799	6.7662
2%	12.4093	8.3297	8.6862	8.4486	14.9988	10.0055	9.8494	8.8505
1%	14.8839	9.8460	10.4269	10.1336	17.6106	11.7867	11.382	10.6668
Wald test :								
10%	0.1354	0.1162	0.1072	0.0964	0.1436	0.1140	0.1100	0.1062
5%	0.0792	0.0602	0.0522	0.0520	0.0884	0.0578	0.0578	0.0520
2%	0.0414	0.0228	0.0238	0.0204	0.0512	0.0274	0.0252	0.0185
1%	0.0260	0.0128	0.0114	0.0104	0.0320	0.0146	0.0136	0.0100
Size-Adjusted Critical Values:								
10%	5.42350	4.9473	4.7550	4.5587	5.6213	4.8503	4.8098	4.7030
5%	7.40080	6.3320	6.1357	6.0367	7.8950	6.3658	6.3470	6.0525
2%	10.1167	8.1745	8.1948	7.8897	10.3846	8.5207	8.2424	7.7123
1%	12.7445	9.7742	9.5291	9.2376	13.3478	9.9669	9.7975	9.2089
J test :								
10%	0.1022	0.1076	0.0995	0.0954	0.0982	0.106	0.1032	0.0976
5%	0.0522	0.0524	0.0484	0.0476	0.0466	0.0509	0.0524	0.0488
2%	0.0216	0.0188	0.0198	0.0188	0.0176	0.0216	0.0230	0.0194
1%	0.0092	0.0104	0.0094	0.0100	0.0102	0.0092	0.0126	0.0092
Size-Adjusted Critical Values:								
10%	5.5808	2.8581	2.5453	2.3395	5.4016	2.8224	2.5889	2.3602
5%	9.8341	3.893	3.3770	3.0961	9.0499	3.8455	3.4679	3.1198
2%	20.0004	5.2806	4.6482	4.0770	17.0544	5.5745	4.8583	4.0907
1%	30.1025	6.7266	5.6375	4.9625	30.9161	6.5544	6.0671	4.8730
WEE test :								
10%	0.0846	0.1014	0.0980	–	0.0900	0.0966	0.1008	–
5%	0.0383	0.0476	0.0468	–	0.0446	0.0470	0.0514	–
2%	0.0150	0.0166	0.0190	–	0.0132	0.0184	0.0188	–
1%	0.0078	0.0086	0.0090	–	0.0078	0.0094	0.0098	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.16: Size and Size-Adjusted Critical Values for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 10$			
$(n_1,\ n_2)$	(20, 10)	(60, 30)	(100, 50)	(250, 125)
EL-W test at nominal levels:				
10%	0.3336	0.2074	0.1802	0.1446
5%	0.2606	0.1332	0.1116	0.0844
2%	0.1922	0.0762	0.0606	0.0408
1%	0.1486	0.0526	0.0388	0.0225
Size-Adjusted Critical Values:				
10%	11.9227	7.0128	6.2510	5.6206
5%	17.1694	9.4087	8.4526	7.3271
2%	25.1885	12.9078	11.2381	9.7585
1%	32.9852	15.8220	13.4227	11.9317
Wald test :				
10%	0.1618	0.1180	0.1104	0.1002
5%	0.1084	0.0622	0.0596	0.0496
2%	0.0634	0.0304	0.0258	0.0200
1%	0.0440	0.0176	0.0130	0.0098
Size-Adjusted Critical Values:				
10%	6.2595	4.9847	4.7794	4.6069
5%	8.7252	6.5938	6.4056	5.9851
2%	12.9218	8.9203	8.3061	7.7744
1%	16.0356	10.5972	10.1592	9.0934
J test:				
10%	0.1046	0.1000	0.1019	0.0982
5%	0.0568	0.0442	0.0484	0.0450
2%	0.0228	0.0160	0.0191	0.0182
1%	0.0126	0.0082	0.0100	0.0080
Size-Adjusted Critical Values:				
10%	5.6697	2.7335	2.5698	2.3715
5%	10.4569	3.6688	3.3904	3.0405
2%	21.2871	5.1100	4.6080	4.0787
1%	36.7252	6.3380	5.6577	4.8048
WEE test :				
10%	0.1042	0.0998	0.0982	–
5%	0.0500	0.0488	0.0509	–
2%	0.0198	0.0198	0.0188	–
1%	0.0092	0.0095	0.0106	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.17: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.1$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3006	0.4820	0.6150	0.7148	0.3323	0.5274	0.6697	0.7902
5%	0.1886	0.3292	0.4754	0.5832	0.2084	0.3850	0.5276	0.6590
2%	0.1024	0.2039	0.3170	0.4178	0.1200	0.2556	0.3910	0.5278
1%	0.0526	0.1194	0.2004	0.2934	0.0868	0.1984	0.3176	0.4510
Wald test :								
10%	0.3166	0.5264	0.6782	0.7764	0.3380	0.5368	0.6806	0.8016
5%	0.2012	0.3822	0.5392	0.6606	0.2202	0.4074	0.5558	0.6914
2%	0.1100	0.2334	0.3736	0.5106	0.1310	0.2856	0.4176	0.5679
1%	0.0738	0.1748	0.2936	0.4204	0.0858	0.1978	0.3184	0.4612
J test :								
10%	0.2148	0.3286	0.4332	0.5294	0.3128	0.4926	0.6316	0.7460
5%	0.1044	0.1766	0.2466	0.3288	0.1998	0.3439	0.4806	0.6040
2%	0.0472	0.0876	0.1174	0.1786	0.1070	0.2128	0.3230	0.4324
1%	0.0220	0.0468	0.063	0.0964	0.0698	0.1444	0.2368	0.3284
WEE test :								
10%	0.2754	0.4840	0.6340	0.7512	0.3190	0.528	0.6854	0.7964
5%	0.1660	0.3464	0.4954	0.6200	0.2144	0.3966	0.5654	0.6906
2%	0.0880	0.2120	0.3340	0.4518	0.1200	0.2572	0.4158	0.5510
1%	0.0524	0.1398	0.2390	0.3404	0.0740	0.1816	0.3232	0.4460

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.18: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.1$							
$(n_1,\ n_2)$	(100, 50)				(250, 150)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3152	0.5302	0.6889	0.7970	0.3197	0.5218	0.6806	0.7892
5%	0.2076	0.3996	0.5628	0.6929	0.2039	0.3840	0.5440	0.6788
2%	0.1266	0.2644	0.4170	0.5506	0.1120	0.2518	0.3996	0.5308
1%	0.0874	0.1896	0.3192	0.4466	0.0782	0.1894	0.3071	0.4448
Wald test :								
10%	0.3190	0.5360	0.6939	0.8070	0.3197	0.5286	0.6840	0.7927
5%	0.2124	0.4076	0.5668	0.7016	0.2066	0.3906	0.5530	0.6836
2%	0.1312	0.2774	0.4296	0.5622	0.1196	0.2610	0.4132	0.5492
1%	0.0890	0.1994	0.3276	0.4650	0.0750	0.1848	0.3076	0.4480
J test:								
10%	0.3094	0.5000	0.6664	0.7774	0.3038	0.5054	0.6622	0.7794
5%	0.2052	0.3764	0.5358	0.6496	0.1988	0.3746	0.5312	0.6618
2%	0.1180	0.2422	0.3734	0.5030	0.1124	0.2410	0.3824	0.5110
1%	0.0724	0.1664	0.2710	0.3894	0.0704	0.1656	0.2822	0.4068
WEE test :								
10%	0.3262	0.5320	0.7008	0.7988	–	–	–	–
5%	0.2186	0.4036	0.5718	0.7004	–	–	–	–
2%	0.1298	0.2678	0.4292	0.5692	–	–	–	–
1%	0.0820	0.1864	0.3306	0.4674	–	–	–	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.19: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.5$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2954	0.4742	0.6160	0.7422	0.3316	0.5410	0.6834	0.8114
5%	0.1858	0.3396	0.4700	0.6054	0.2344	0.4160	0.5742	0.7191
2%	0.1090	0.2176	0.3312	0.4656	0.1408	0.2846	0.4370	0.5910
1%	0.0722	0.1530	0.2446	0.3554	0.0876	0.2058	0.3398	0.4886
Wald test :								
10%	0.3240	0.5240	0.6770	0.7966	0.3323	0.5414	0.6852	0.8154
5%	0.2110	0.3872	0.5336	0.6732	0.2248	0.4124	0.5702	0.7158
2%	0.1176	0.2552	0.3800	0.5190	0.1364	0.2874	0.4366	0.5896
1%	0.0777	0.1822	0.2898	0.4222	0.0892	0.2138	0.3482	0.4962
J test:								
10%	0.2128	0.3348	0.4346	0.5374	0.3034	0.4834	0.6256	0.7604
5%	0.1132	0.1824	0.2556	0.3322	0.1948	0.3494	0.4952	0.6240
2%	0.0570	0.0872	0.1356	0.1855	0.1174	0.2336	0.3660	0.4812
1%	0.0282	0.0482	0.0724	0.1060	0.0728	0.1524	0.2592	0.3622
WEE test :								
10%	0.2662	0.4518	0.6250	0.7398	0.3068	0.5125	0.6774	0.7914
5%	0.1608	0.3128	0.4778	0.5974	0.1988	0.3754	0.5516	0.6852
2%	0.0800	0.1796	0.2996	0.4104	0.1096	0.2476	0.4020	0.5242
1%	0.0448	0.1106	0.1928	0.2890	0.0696	0.1678	0.3054	0.4252

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.20: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 0.5$							
$(n_1,\ n_2)$	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3353	0.5356	0.7078	0.8038	0.3372	0.5410	0.6979	0.8178
5%	0.2302	0.4218	0.5996	0.7098	0.2304	0.4220	0.5920	0.7284
2%	0.1258	0.2784	0.4360	0.5688	0.1368	0.2834	0.4502	0.5910
1%	0.0882	0.2076	0.3464	0.4764	0.0835	0.1874	0.3352	0.4778
Wald test :								
10%	0.3292	0.5316	0.7050	0.8024	0.3392	0.5402	0.7006	0.8206
5%	0.2248	0.4138	0.5890	0.7036	0.2286	0.4232	0.5914	0.7296
2%	0.1202	0.2612	0.4180	0.5512	0.1340	0.2782	0.4440	0.5876
1%	0.0852	0.1988	0.3482	0.4698	0.0884	0.1998	0.3469	0.4898
J test:								
10%	0.3074	0.4982	0.6656	0.7644	0.3388	0.5387	0.6924	0.8129
5%	0.1990	0.3662	0.5336	0.6528	0.2236	0.4102	0.5699	0.7086
2%	0.1148	0.2382	0.3874	0.5078	0.1252	0.2552	0.4152	0.5648
1%	0.0722	0.1648	0.2829	0.3986	0.0786	0.1774	0.3142	0.4502
WEE test :								
10%	0.3210	0.5264	0.6830	0.8036	–	–	–	–
5%	0.2150	0.3992	0.5586	0.7054	–	–	–	–
2%	0.1208	0.2693	0.4176	0.5774	–	–	–	–
1%	0.0736	0.1952	0.3242	0.4678	–	–	–	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.21: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 1$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2774	0.4536	0.6024	0.7160	0.3278	0.5250	0.6726	0.7934
5%	0.1676	0.3066	0.4416	0.5574	0.2262	0.3996	0.5588	0.6954
2%	0.0828	0.1638	0.2636	0.3756	0.1442	0.2834	0.4452	0.5856
1%	0.0522	0.1124	0.1930	0.2890	0.1002	0.2026	0.3540	0.4920
Wald test :								
10%	0.2974	0.4806	0.6464	0.7576	0.3202	0.5182	0.6717	0.7906
5%	0.1858	0.3414	0.4940	0.6250	0.2252	0.3934	0.5642	0.6962
2%	0.1032	0.2046	0.3328	0.4606	0.1332	0.2710	0.4372	0.5734
1%	0.0556	0.1336	0.2326	0.3352	0.0910	0.1826	0.3343	0.4694
J test:								
10%	0.1981	0.3098	0.4152	0.5030	0.2950	0.4644	0.6124	0.7432
5%	0.1054	0.1708	0.2428	0.3126	0.1888	0.3204	0.4832	0.6098
2%	0.0432	0.0675	0.1014	0.1400	0.1104	0.2026	0.3268	0.4448
1%	0.0242	0.0376	0.0614	0.0816	0.0651	0.1232	0.2194	0.3206
WEE test :								
10%	0.2664	0.4402	0.5956	0.7221	0.3232	0.5070	0.6652	0.7848
5%	0.1606	0.3002	0.4472	0.5748	0.2120	0.3721	0.5427	0.6756
2%	0.0842	0.1712	0.2832	0.3970	0.1190	0.2424	0.3867	0.5284
1%	0.0420	0.1076	0.1910	0.2824	0.0730	0.1676	0.2926	0.4252

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.22: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 1$							
$(n_1,\ n_2)$	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3362	0.5350	0.6844	0.7967	0.3390	0.5416	0.6878	0.8080
5%	0.2248	0.4140	0.5620	0.6962	0.2274	0.4062	0.5629	0.6996
2%	0.1318	0.2800	0.4192	0.5639	0.1228	0.2668	0.4198	0.5564
1%	0.0854	0.2050	0.3180	0.4626	0.0762	0.1870	0.3176	0.4416
Wald test :								
10%	0.3280	0.5294	0.6780	0.7917	0.3492	0.5494	0.6959	0.8149
5%	0.2248	0.4122	0.5620	0.6956	0.2310	0.4118	0.5689	0.7066
2%	0.1266	0.2706	0.4040	0.5540	0.1258	0.2770	0.4308	0.5706
1%	0.0884	0.2008	0.3252	0.4686	0.0834	0.2052	0.3412	0.4744
J test:								
10%	0.3196	0.5130	0.6494	0.7688	0.3416	0.5374	0.6899	0.8012
5%	0.2088	0.3744	0.5185	0.6538	0.2254	0.4048	0.5598	0.6954
2%	0.1112	0.2332	0.3585	0.4878	0.1232	0.2716	0.4194	0.5466
1%	0.0684	0.1586	0.2630	0.3786	0.0792	0.1816	0.3111	0.4346
WEE test :								
10%	0.3234	0.5118	0.6996	0.8076	–	–	–	–
5%	0.2104	0.3830	0.5788	0.7040	–	–	–	–
2%	0.1252	0.2518	0.4238	0.5618	–	–	–	–
1%	0.0792	0.1738	0.3302	0.4546	–	–	–	–

Notes to table: The regressor x is from uniform distribution $U(0,1)$. Number of replication is 5000. The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.23: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 2$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2942	0.4540	0.5961	0.7151	0.3286	0.5182	0.6778	0.7992
5%	0.1736	0.3020	0.4350	0.5644	0.2136	0.3780	0.5450	0.6876
2%	0.0908	0.1794	0.2806	0.3968	0.1204	0.2532	0.3990	0.5386
1%	0.0630	0.1293	0.2154	0.3190	0.0840	0.1816	0.3108	0.4402
Wald test :								
10%	0.3046	0.4760	0.6246	0.7410	0.3276	0.5128	0.6802	0.8014
5%	0.1868	0.3226	0.4688	0.5970	0.2210	0.3836	0.5586	0.7024
2%	0.1138	0.2186	0.3358	0.4616	0.1230	0.2498	0.4062	0.5442
1%	0.0651	0.1364	0.2304	0.3276	0.0850	0.1858	0.3168	0.4576
J test:								
10%	0.2226	0.3376	0.4258	0.5205	0.2896	0.4624	0.6268	0.7428
5%	0.1266	0.2008	0.2650	0.3390	0.1855	0.3246	0.4774	0.6056
2%	0.0538	0.0946	0.1303	0.1748	0.0932	0.1784	0.2964	0.4120
1%	0.0228	0.0484	0.0572	0.0872	0.0651	0.1282	0.2226	0.3318
WEE test :								
10%	0.2816	0.4462	0.5946	0.7168	0.3328	0.5192	0.6727	0.7836
5%	0.1648	0.3114	0.4346	0.5688	0.2200	0.3834	0.5416	0.6690
2%	0.0859	0.1778	0.2744	0.3792	0.1178	0.2450	0.3782	0.5150
1%	0.0496	0.1062	0.1772	0.2716	0.0750	0.1754	0.2798	0.4136

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.24: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 2$							
$(n_1,\ n_2)$	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3232	0.5174	0.6882	0.7906	0.3256	0.5292	0.6796	0.7952
5%	0.2200	0.3932	0.5610	0.6936	0.2292	0.4082	0.5672	0.6946
2%	0.1196	0.2514	0.4056	0.5520	0.1300	0.2832	0.4156	0.5564
1%	0.0832	0.1962	0.3320	0.4634	0.0790	0.1956	0.3136	0.4390
Wald test :								
10%	0.3170	0.5242	0.6899	0.8004	0.3353	0.5426	0.6954	0.8026
5%	0.2144	0.3900	0.5659	0.6982	0.2314	0.4226	0.5824	0.7032
2%	0.1220	0.2644	0.4172	0.5686	0.1408	0.2984	0.4432	0.5828
1%	0.0776	0.1902	0.3252	0.4616	0.0910	0.2132	0.3432	0.4730
J test:								
10%	0.3088	0.4950	0.6632	0.7682	0.3392	0.5440	0.6896	0.8006
5%	0.1962	0.3595	0.5302	0.6530	0.2258	0.4138	0.5669	0.6912
2%	0.0958	0.2178	0.3454	0.4836	0.1290	0.2778	0.4230	0.5480
1%	0.0508	0.1336	0.2340	0.3605	0.0848	0.1986	0.3217	0.4468
WEE test :								
10%	0.3300	0.5410	0.6750	0.7940	–	–	–	–
5%	0.2146	0.4170	0.5436	0.6909	–	–	–	–
2%	0.1252	0.2738	0.4012	0.5486	–	–	–	–
1%	0.0801	0.1896	0.3030	0.4434	–	–	–	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Table 6.25: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 10$							
$(n_1,\ n_2)$	(20, 10)				(60, 30)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.2662	0.4030	0.5446	0.6450	0.2996	0.4938	0.6254	0.7478
5%	0.1628	0.2732	0.3982	0.4970	0.1986	0.3604	0.5002	0.6260
2%	0.0800	0.1570	0.2436	0.3298	0.1058	0.2180	0.3536	0.4708
1%	0.0429	0.0952	0.1544	0.2242	0.0650	0.1474	0.2640	0.3646
Wald test :								
10%	0.2808	0.4270	0.5820	0.6939	0.3114	0.5100	0.6622	0.7810
5%	0.1734	0.2970	0.4390	0.5500	0.2048	0.3731	0.5332	0.6700
2%	0.0784	0.1678	0.2640	0.3605	0.1088	0.2372	0.3832	0.5130
1%	0.0502	0.1086	0.1824	0.2634	0.0680	0.1668	0.2920	0.4058
J test:								
10%	0.2154	0.3050	0.4062	0.4874	0.2985	0.4812	0.6238	0.7506
5%	0.1112	0.1671	0.2344	0.2908	0.1950	0.3546	0.5006	0.6258
2%	0.0412	0.0694	0.1014	0.1290	0.1030	0.2204	0.3434	0.4480
1%	0.0206	0.0364	0.0514	0.0644	0.0641	0.1426	0.2486	0.3366
WEE test :								
10%	0.2842	0.4556	0.5846	0.7018	0.3227	0.5220	0.6674	0.7832
5%	0.1702	0.3071	0.4296	0.5450	0.2158	0.3830	0.5460	0.6654
2%	0.0842	0.1698	0.2610	0.3634	0.1138	0.2438	0.3942	0.5195
1%	0.0492	0.1092	0.1676	0.2474	0.0730	0.1654	0.2945	0.4076

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter.

Table 6.26: Power Comparison for the Four Tests: EL-W, Wald, J, and WEE

	$\rho^2 = 10$							
$(n_1,\ n_2)$	(100, 50)				(250, 125)			
δ:	1	2	3	4	1	2	3	4
EL-W test at actual size levels:								
10%	0.3200	0.5172	0.6608	0.7760	0.3187	0.5220	0.6812	0.7966
5%	0.2128	0.3842	0.5250	0.6566	0.2158	0.3980	0.5636	0.6918
2%	0.1264	0.2466	0.3878	0.5210	0.1206	0.2604	0.4134	0.5360
1%	0.0818	0.1762	0.2992	0.4222	0.0712	0.1776	0.3002	0.4242
Wald test :								
10%	0.3266	0.5328	0.6782	0.7966	0.3318	0.5452	0.7070	0.8169
5%	0.2142	0.3956	0.5464	0.6842	0.2260	0.4206	0.5996	0.7158
2%	0.1346	0.2634	0.4124	0.5528	0.1378	0.2862	0.4528	0.5870
1%	0.0766	0.1774	0.3006	0.4316	0.0908	0.2112	0.3544	0.4956
J test:								
10%	0.3071	0.5098	0.6538	0.7732	0.3248	0.5342	0.6948	0.8044
5%	0.2092	0.3782	0.5235	0.6578	0.2260	0.4188	0.5866	0.7044
2%	0.1176	0.2366	0.3654	0.4954	0.1258	0.2678	0.4306	0.5634
1%	0.0706	0.1544	0.2693	0.3776	0.0810	0.1912	0.3342	0.4670
WEE test :								
10%	0.3094	0.5195	0.6826	0.8064	–	–	–	–
5%	0.2044	0.3864	0.5534	0.6959	–	–	–	–
2%	0.1120	0.2506	0.4000	0.5488	–	–	–	–
1%	0.0718	0.1848	0.3004	0.4432	–	–	–	–

Notes to table: The regressor x is from a uniform distribution $U(0,1)$. Number of replications is 5,000. Sample sizes are the pair (n_1, n_2). $\rho^2 = \sigma_2^2/\sigma_1^2$. The true values for the two coefficient vectors are: under the null hypothesis: $\beta_1 = \beta_2 = (1,1)'$; under the alternative: $\beta_2 = (1, \beta_{22})'$, where β_{22} varies according to the non-centrality parameter. The WEE test is not applicable with large sample sizes.

Chapter 7

Conclusions

The objectives of this dissertation are two-fold. One is to develop and apply new procedures based on the maximum empirical likelihood method in the context of econometrics. The second objective is to analyze the sampling properties of various empirical likelihood type tests.

The maximum empirical likelihood (EL) method stands out among all of the parametric and nonparametric methods in the literature of statistics for the following two reasons. One is that the EL method overcomes the possible mis-specification problem that is incurred by the parametric methods. The second reason is that the EL method utilizes the likelihood function and unbiased moment equations, and this leads to improved estimation efficiency and potentially improved power for tests. These two aspects of the EL method show the great potential that the EL method can have if it is applied in various areas of econometrics. The EL literature has mostly focused on the estimation side of the story. Far less attention has been paid to testing issues in the context of the EL method in econometrics. The research reported in this dissertation fills some of these gaps. The focus is on testing problems. The problems that we have considered are the ones that are of special importance in econometrics.

The dissertation starts with two problems in statistics. One is testing for the normality of the underlying population. The normality assumption is very fundamental and very important in regression models in econometrics. There is a need to test for the normality in regressions. The second problem is so-called testing for the Behrens-Fisher problem, which has a long and challenging history. Here, one is concerned with testing the equality of two normal means when the variances are unknown and unequal. This problem is extended to that testing for structural change in regressions. In all of these cases, our focus is on the sampling properties of the EL type tests. Monte Carlo simulation experiments are employed to provide the empirical sizes and the powers for the tests that we have considered.

Our principal contributions are as follows.

- First, we have developed new theoretical approaches to the selected topics using the maximum empirical likelihood method. These topics are: (i) testing for normality in pure random data sets; (ii) testing for normality in a regression model; (iii) solving the Behrens-Fisher problem; and (iv) testing for the problem of structural change in regressions. In particular, we have established the forms of the EL-type tests, including the empirical likelihood ratio (ELR) test and EL-type Wald test, in each topic.
- Second, we have provided a complete analysis of the sampling properties of the EL-type tests using the Monte Carlo simulation technique. These include the empirical sizes, the size-adjusted critical values, and the powers of the tests.
- Third, we have provided detailed comparisons of the sampling properties of the EL-type tests and other conventional tests.

The empirical likelihood ratio test is applied to testing for normality in pure random data (Chapter 3). The comparison of the sampling properties of the ELR test is conducted with the Jarque-Bera test, D'Agostino test, χ^2 goodness of fit tests. The ELR is found to be the most powerful of these tests in testing for normality against the asymmetric alternative distributions that we have considered. The technique of testing for normality is extended to a linear regression model in Chapter 4. The finite-sample size-adjusted critical values for the ELR test in the regression context are specific to the regressor matrices used. They may not be applicable to a regression model with a different regressor matrix.

Chapter 5 presents a new theoretical solution to the well known Behrens-Fisher problem using the empirical likelihood method. The ratio of the variances of the two data sets of the problem is a key parameter. Changes in the variance ratio alter the behavior of any solution to the Behrens-Fisher problem. In our approach, we present a unique way of using the variance ratio parameter, and the common features of the two data sets. The power properties of the empirical likelihood ratio test are carefully analyzed across a range of situations. We have investigated the way that the power of the ELR test varies with the ratio of the two variances, the sample size pair of the data sets, and the non-centrality parameter. Chapter 6 contains a natural extension of the approach in Chapter 5 to testing for the problem of structural change in a regression model. The sampling properties of the EL-type Wald test are compared against various conventional tests: the Jayattisa test, the Weerahandi test, and the usual Wald test. The Monte Carlo experiment method is used to simulate the actual sizes and the size-adjusted critical values for the EL-type Wald test and the three other tests. These size-adjusted critical values are used to compute the powers for the tests at actual significance levels. Again these values are specific to the regressor matrices we have chosen; they are not applicable to a regression model with a different regressor matrix.

The following major conclusions emerge from the research reported upon in this dissertation:

- The EL method is very flexible. The four applications that we have considered using the EL method in this dissertation provide strong evidence that the empirical likelihood method is flexible enough to be applied to various other problems in econometrics.
- The EL method is efficient. The four applications also show that the EL method can incorporate information from many sources. The fact that the EL method utilizes the likelihood function and unbiased moment equations enables it to have this efficiency.
- The EL type tests have good sampling properties. The results of the simulation studies demonstrate that the empirical likelihood type tests have very good sampling properties. It is recommended that the EL type tests can be used in various testing purposes.

The computational work associated with the EL method can be very challenging. In solving the Behrens-Fisher problem, the computing difficulties are clearly presented. As we have seen in the appendix of Chapter 2, nonlinear equation solving algorithms provide ways to implement the EL method. There should be some optimal algorithms that can reduce the computational difficulties. Exploring other numerical methods in this respect would be very interesting.

Given the promising properties and attractive features of the EL method, and also the good sampling properties of the EL-type tests, it seems that there is a lot of potential to explore the application of the EL method to various other problems in econometrics.

Several such possibilities come to mind, including:

- It would be interesting to construct EL-type tests for the validity of the other assumptions (in addition to normality) that are traditionally made about the error term in a regression model. These include the assumptions of homoscedasticity and serial independence, for example.

- Testing for more general types of structural breaks in a regression model would be another interesting topic. In real data, there may be more than one structural break. The ability to deal with more than one structural change is needed for any test. Applying the EL method to this type of model could be very fruitful.

- The EL method may be useful in the analysis of models involving stochastic and deterministic switching regimes. Certain types of regime-switching models and models of markets in disequilibrium involve econometric issues which essentially amount to dealing with mixtures of normal disturbances. It would be interesting to undertake an EL-type analysis of normal mixtures, and extend this to these types of econometric models.

- Applying the EL method to GARCH models for high-frequency time series data would also be very interesting.

- Given the immense impact that the analysis of non-stationary time series has had on the econometrics literature during the past twenty years, it would appear to be worthwhile to investigate the usefulness of the EL method for testing for unit roots and co-integration.

In short, there are many avenues in econometrics that could be explored fruitfully using these new tools. The results reported in this dissertation represent a small step down this path. There is no doubt that the empirical likelihood techniques for estimation and testing will find an important place in econometric analysis in the future.

References

Ali, M. M., and J. L. Silver (1985), "Test for Equality between Sets of Coefficients in Two Linear Regressions Under Heteroscedasticity", *Journal of the American Statistical Association*, 80, 730 - 735.

Bera, A. and Y. Bilias (2002), "The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis", *Journal of Econometrics*, 107, 51 - 86.

Bera, A. and C. Jarque (1980), "Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals" *Economics Letters*, 6, 255-259.

Bera, A. and C. Jarque (1981) "Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals: Monte Carlo Evidence" *Economics Letters*, 7, 313-318.

Cox, D. R. (1984), "Effective Degrees of Freedom and the Likelihood Ratio Test", *Biometrika*, 71, 487 - 493.

D'Agostino, R. B. (1971), "An Omnibus Test of Normality for Moderate and Large Size Samples", *Biometrika*, 58, 341 - 348.

D'Agostino, R. B. (1972), "Small Sample Probability Points for the D Test of Normality", *Biometrika*, 59, 219-221

DiCiccio, T., P. Hall, and J. Romano (1989), "Comparison of Parametric and Empirical Likelihood Functions", *Biometrika*, 76, 465 - 476.

DiCiccio, T. J. and J. P. Romana (1989), "On Adjustments Based on the Signed of the Empirical Likelihood Ratio Statistic", *Biometrica*, 76, 447-456.

DiCiccio, T., P. Hall, and J. Romano (1991), "Empirical Likelihood is Bartlett-Correctable", *Annals of Statistics*, 19, 1053 - 1061.

Duong, Q. P., and R. W. Shorrock (1996), "On Behrens-Fisher Solutions", *The Statistician*, 45, 57-63

Fenstan, G. U. (1983), "A Comparison between the U and V Tests in the Behrens-Fisher Problem", *Biometrika*, 70, 300 - 302.

Gauss 5.0 for Windows NT (2002), *Aptech System, Inc.*, Maple Valley WA.

Giles, J. A. and D. E. A. Giles (1993), "Pre-Testing Estimation and Testing in Econometrics: Recent Developments", *Journal of Economic Surveys*, 7, 145 - 197.

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029 - 1054.

Honda, Y. (1982), "On Tests of Equality Between Sets of Coefficients in Two Linear Regressions When Disturbance Variances Are Unequal", *The Manchester School*, 49, 116 - 125.

Huang, C. J. and B. W. Bolch (1974), "On the Testing of Regression Disturbances for Normality", *Journal of the American Statistical Association*, 69, 330-335.

Imben, G. W., R. H. Spady, and P. Johnson (1998), "Information Theoretic Approaches To Inference in Moment Condition Models", *Econometrica*, 66, 333 - 357.

Jayatissa, W. A. (1977), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions When Disturbance Variances Are Unequal", *Econometrica*, 45, 1291 - 1292.

Jing, B. Y. (1995), "Two-Sample Empirical Likelihood Method", *Statistics and Probability*

Letters, 24, 315 - 319.

Kitamura, Y. (1997), "Empirical Likelihood Methods with Weakly Dependent Processes", *Annals of Statistics*, 25, 2084 - 2102.

Kendall, M. and A. Stuart (1963, 1969, and 1977), *The Advanced Theory of Statistics*, Charles Griffin & Company Limited, London.

Lee, A. F. S. and J. Gurland (1975), "Size and Power of Tests for Equality of Means of Two Normal Populations", *Journal of the American Statistical Association*, 70, 933 - 944.

Mittelhammer, R., G. Judge, and D. Miller (2000), *Econometric Foundations*, Cambridge University Press, Cambridge.

Mittelhammer, R., G. Judge, and R. Schoenberg (2003), "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods", *Working Paper*, Washington State University, University of California, Berkeley, and Aptech Systems, Inc.

Nelder, J. A. and R. Mead (1965), "A Simplex Method for Function Minimization", *Computer Journal*, 7, 308 - 313.

Ohtani, K. and T. Toyoda (1985), Small Sample Properties of Tests of Equality Between Sets of Coefficients in Two Linear Regressions Under Heteroscedasticity, *International Economic Review*, 26, 37 - 43.

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional", *Biometrika*, 75, 237 - 249.

Owen, A. B. (1990), "Empirical Likelihood Ratio Confidence Region", *The Annals of Statistics*, 18, 90 - 120.

Owen, A. B. (1991), "Empirical Likelihood for Linear Models", *The Annals of Statistics*, 19, 1725 - 1747.

Owen, A. B. (2001), *Empirical Likelihood*, Chapman & Hall/CRC, New York.

Pearson, K. (1894), "Contribution to the Mathematical Theory of Evolution", *Philosophical Transactions of the Royal Society of London*, Series A 186, 343 - 414.

Pearson, K. (1900) "On the Criterion That a Given System of Deviations From the Probable In the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen From Random Sampling", *Philosophical Magazine Series*, 50, 157 - 175.

Pearson, K. (1902), "On the Systematic Fitting of Curves to Observations and Measurements, Parts I and II", *Biometrika*, 1, 265 - 303; 2, 1 - 23.

Pearson, K. (1936), "Method of Moments and Method of Maximum Likelihood", *Biometrika*, 28, 34 - 59.

Price, K. and R. Storn (1997), "Differential Evolution", *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, April 1997 Algorithms.

Qin, J. (1991), "Likelihood and Empirical Likelihood Ratio Confidence Intervals in Two Sample Semi-parametric Models", *Technical Report Series University of Waterloo*, Stat-91-6.

Qin, J. (1993), "Empirical Likelihood in Biased Sample Problems", *Annals of Statistics*, 21, 1182 - 1196.

Qin, J. and J. Lawless (1991), "Empirical Likelihood and General Estimating Equations I", *Technical Report Series University of Waterloo*, Stat-91-10.

Qin, J. and J. Lawless (1994),"Empirical Likelihood and General Estimating Equations", *Annals of Statistics*, 22, 300 - 325.

Qin, J. and J. Lawless (1995), "Estimating Equations, Empirical Likelihood and Constraints on Parameters", *Canadian Journal of Statistics*, 23, 145 - 159.

Royston, J. P. (1982), "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples", *Applied Statistics*, 31, 115 - 124.

Shapiro, S. S. and M. B. Wilk (1965), "An Analysis of Variance Test for Normality (Complete Samples)", *Biometrika*, 52, 591 - 611.

Theil, H. (1965), "The Analysis of Disturbances in Regression Analysis", *Journal of the American Statistical Association*, 60, 1067 - 1079.

Theil, H. (1968), "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances", *Journal of the American Statistical Association*, 63, 242 - 251.

Tukey, J. W. (1954), "Unsolved Problems of Experimental Statistics", *Journal of the American Statistical Association*, 49, 706 - 731.

Watt, P. A. (1979), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions When Disturbance Variances Are Unequal: Some Small Sample Properties", *The Manchester School*, 47, 391 - 396.

Weerahandi, S. (1987), "Testing Regression Equality with Unequal Variances", *Econometrica*, 55, 1211 - 1215.

Welch, B. L. (1947), "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved", *Biometrika*, 34, 28 - 35.

White, H. (1980), "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica*, 48, 817 - 838.

White, H. and G. MacDonald (1980), "Some Large-Sample Tests for Nonnormality in the Linear Regression Model", *Journal of the American Statistical Association*, 75, 16 - 31.

Zaman, A. (1996), *Statistical Foundations for Econometric Techniques*, Academic Press, New York.