BOOTSTRAP AND ITS APPLICATIONS

theory and Evidence

A Master Thesis Presented, by

Mustafa Cenk Tire

to

The Department of Economics and The Institute of Economics
and Social Sciences of
Bilkent University

in partial fulfillment for the degrees of

MASTER OF ECONOMICS

BILKENT UNIVERSITY
ANKARA
September
1995

BOOTSTRAP AND ITS APPLICATIONS

Theory and Evidence

A Master Thesis Presented by

Mustafa Cenk TIRE

to

The Department of Economics and The Institute of Economics and Social Sciences of Bilkent University in partial fulfillment for the degrees of

MASTER OF ECONOMICS

BILKENT UNIVERSITY
ANKARA
September
1995

Mustafa Cenk TIRE
tarafından bağışlanmıştır

I certify that I have read this thesis and in my opinion, it is fully adequate, in scope and in quality, as a thesis for degree of Master of Economics

Prof. Dr Asad ZAMAN

I certify that I have read this thesis and in my opinion, it is fully adequate, in scope and in quality, as a thesis for degree of Master of Economics

Assoc. Prof. Pranesh KUMAR

I certify that I have read this thesis and in my opinion, it is fully adequate, in scope and in quality, as a thesis for degree of Master of Economics

Assoc. Prof. Syed MAHMUD

Approved by the Institute of Economics and Social Sciences

Director:

ABSTRACT

BOOTSTRAP AND ITS APPLICATIONS
Theory and Evidence

Mustafa Cenk TIRE

Master of Economics
Supervisor : Prof. Dr. Asad ZAMAN

September, 1995

This thesis mainly discusses the theory and applications of an estimation technique called **Bootstrap**. The first part of the thesis focuses on the accuracy of Bootstrap in density estimation by comparing Bootstrap with another estimation technique called Normal approximation based on central limit theorem. The theoretical analysis on this issue shows that Bootstrap is always, at least as good as, and in some cases better than, the Normal approximation. This analysis has been supported by empirical analysis.

Later parts of the thesis are devoted to the applications of Bootstrap. Two examples for these applications, Bootstrapping F-test in dynamic models and using Bootstrap in common factor restrictions have been extensively discussed. The performance of Bootstrap has been investigated separately and interpreted precisely. Bootstrap has worked well in F-test application, but it has been dominated by other tests such as Likelihood Ratio test, Wald test; in common factor restrictions.

Keywords: Bootstrap, Monte Carlo, F-test, Common Factor Restrictions, Likelihood Ratio, Wald test.

ÖZ

BOOTSTRAP TEKNİĞİ VE UYGULAMALARI

Teori ve Kanıt

Mustafa Cenk TİRE

Yüksek Lisans Tezi, İktisat Bölümü
Tez Danışmanı: Prof. Dr. Asad ZAMAN

Eylül, 1995

Bu çalışmada Bootstrap tekniği ve onun uygulamaları incelenmiştir. Tezin ilk bölümünde, Bootstrap tekniğini, merkezi limit teoremine dayalı Normal yaklaşım tekniğiyle karşılaştırıp, tekniğin dağılım tahminindeki doğruluk oranı üzerinde durulmuştur. Bu konudaki teorik analiz, Bootstrap tekniğinin en az Normal yaklaşım tekniği kadar iyi, hatta bazı durumlarda daha iyi, çalıştığını göstermiştir. Bu analiz, daha sonra deneysel analiz ile desteklenmiştir.

Çalışmanın sonraki bölümleri, Bootstrap uygulamalarına adanmıştır. Bu uygulamalara iki örnek olan, dinamik modellerde uygulanan F testinde Bootstrap kullanımı ve ortak faktör kısıtlamalarında Bootstrap kullanımı, açıkça incelenmiştir. Bootstrap tekniğinin performansı her iki uygulamada ayrı ayrı incelenip, yorumlanmıştır. F testi uygulamasında, Bootstrap tekniği iyi çalışmasına karşın; ortak faktör kısıtlamalarında, olasılık oranı testi, Wald testi gibi testlere oranla daha kötü sonuçlar vermiştir.

Anahtar Kelimeler: Bootstrap, Monte Carlo, F testi, Ortak Faktör Kısıtlamarı, Olasılık Oranı testi, Wald testi.

Acknowledgement

I would like to thank Prof. Dr. Asad ZAMAN for his supervision and guidance during the development of this thesis. I would also like to thank Assoc. Prof. Dr. Pranesh KUMAR and Assoc. Prof. Dr. Syed MAHMUD for reading and evaluating this thesis.

Contents

1
Introduction

Estimation of the distribution of statistics based on the observed data has been developed during recent years. People began to estimate the distribution of the whole data by just analyzing a specified size of sample data taken from actual whole data. Although the techniques of estimation is becoming more complex, there are some techniques which are rather simple and efficient. However, we have to check how much these techniques are efficient relative to other techniques. These efficiency criteria should be investigated rather deeply and if possible, analytical and theoretical suggestions should be supported by empirical work.

The issue presented here is the pros and cons of an estimation technique called **Bootstrap**. The bootstrap technique covers a wide variety of applications. Therefore, here, it is intended to narrow its scope, and two applications of Bootstrap are going to be presented to the reader.

The central feature of Bootstrap is to resample the initial sample by drawing one observation with probability 1/n where n is the size of the sample with replacement. The repetition of this sample asymptotically composes a distribution function which converges to the probability distribution function of the actual data. Nevertheless, the rate of convergence needs to be considered. Because, it has been seen , and later, proven in following chapters, that Bootstrap works quite well in some cases relative to some other estimation methods, e.g. Normal approximation based on CLT. The accuracy of

Bootstrapping depends on some conditions. These conditions will be extensively discussed later.

In later discussions, we will concentrate on some applications of Bootstrap technique. Using Bootstrap in calculating critical value of F-test and Common Factor test, can be regarded as two good examples. This thesis also focuses on accuracy .of the Bootstrap. In addition, we reconsider all Bootstrap applications when raw data have some leverage points inside. The effect of these leverage points is separately discussed later. Furthermore, real life examples are included in some applications in order to bring theories to facts.

The plan of this work is briefly described as follows. The next section is devoted the literature survey of Bootstrap and maximum likelihood. Chapter 2 extensively discusses in what conditions Bootstrap gives better results than Normal approximation. The applications of Bootstrap, i.e Bootstrapping F-test when the model is dynamic and using Bootstrap in common factor restrictions, are clearly discussed in Chapter 3 and 4 respectively. Final chapter is a summary of this thesis.

It should be noted that **GAUSS**, a special statistical programming language, was used for computer simulations during analyses and these programs are included in Appendix part.

1.1 Maximum Likelihood Estimates

Given a sample, $X_1,...X_n \sim f(x,\theta')$ where random variables are i.i.d and θ' is a parameter belonging to the sample space Θ, we define maximum likelihood estimate as the value of θ' which makes the probability of the observed sample as large as possible. We have some necessary and sufficient conditions under which the sequence of estimates, $\bar{\theta}_n$ converges to true parameter, θ as n approaches to infinity. If the sequence of estimates converges to the true parameter, it is called *Consistent.* The conditions for consistency are given below;

Definition 1.1:(identification) *Suppose the set of possible densities of y is given by $f(y,\theta)$ for $\theta \in \Theta$. The parameter θ is called identified if for any two parameters θ_1 and θ_2 such that $\theta_1 \neq \theta_2$, there is some set A such that $P(y \in A \mid \theta_1) \neq P(y \in A \mid \theta_2)$*

Definition 1.2: *Suppose x_i is an i.i.d sequence of random variables. The log normal likelihood function ρ, ($\rho(x_i,\theta) = \log f(x_i,\theta)$) is dominated if*

$$E\ Sup_{\theta \in \Theta}\, \rho(x_i,\theta) < \infty$$

Using the above definitions, we can state an useful theorem.

Theorem 1.1: (Wald) *Let x_i be an i.i.d sequence of random variable with common density $f(x,\theta')$ and suppose $\bar{\theta}_n$ is a sequence of approximate maxima of the joint likelihood $L(\theta) = \Pi^n_{i=1} f(x_i,\theta)$ for $\theta \in \Theta$ a compact parameter space. Then $\bar{\theta}_n$ converges almost surely to θ' provided that θ is identified and the log normal likelihood is dominated*[1] .

[1] See Zaman (1996)

Depending on above theorem, we can state maximum likelihood is consistent. However, we need following lemma to show that θ' is always unique maximum of $E\rho(x_1,\theta)$.

Lemma 1.1:(information inequality) *For all θ such that $f(x_i,\theta)$ is a different density from $f(x_i,\theta')$, we have the inequality*[2]

$$E\log f(x_i,\theta) < E\log f(x_i,\theta')$$

This is a sketch of a proof. We can state that ML is strongly consistent if theorem 1.1 and lemma 1.1 hold.

Maximum likelihood estimate of θ', denoted as $\bar{\theta}$, is defined as follows,

$$\Pi^n_{i=1} f(x_i, \bar{\theta}) = \mathrm{Sup}_{\theta\in\Theta} \Pi^n_{i=1} f(x_i,\theta')$$

The above equation shows that maximum likelihood estimator is an estimator which maximizes the probability of observed sample.

Now, suppose that we have a standard linear regression model, $\mathbf{y=X\beta+\varepsilon}$, where the error term, ε, is distributed as Normal with mean 0 and variance $\sigma^2\mathbf{I}$. Then, the maximum likelihood estimates of β and σ^2 are found easily. Since it is known that $\varepsilon\sim N(0,\sigma^2I)$, by the property of Normal distribution, it is easily seen that y is distributed as $y\sim N(X.\beta,\sigma^2I)$[3]. Mathematically, maximizing values $\bar{\beta}$ and $\bar{\sigma}^2$ are found by differentiating likelihood function,

[2] See Zaman (1996)
[3] Calculations are straightforward and, therefore, omitted.

$$L(.) = \prod_{i=1}^{n} f(x_i, y_i, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y_i - x_i.\beta)^2) \qquad \textbf{(1.1)}$$

This yields the following,

$$\partial L/\partial\beta=0 \Rightarrow X'y-X'X\beta=0 \Rightarrow \bar{\beta}=(X'X)^{-1}.X'y \qquad \textbf{(1.2)}$$
$$\partial L/\partial\sigma^2=0 \Rightarrow \bar{\sigma}^2=1/N \,\|y-X.\ \bar{\beta}\|^2$$

Since these two estimates are independent and distribution of $\bar{\beta}$ and $\bar{\sigma}^2$ form an exponential family[4], it is easily shown that $\bar{\beta}$ and $\bar{\sigma}^2$ are functions of the complete sufficient statistics. This implies that they are MVUE. There is no unknown parameter in the formula of these functions.

1.2 Monte Carlo Simulation

Monte Carlo procedure for estimating the distribution of maximum likelihood consists of some straightforward steps. These are,

Step1: Choosing a random sample of size n.

Step2: Calculating maximum likelihood estimates of chosen sample.

Step3: Repeating step 1 and step 2 sufficient number times and keeping track of these estimates, construct a histogram of them.

Although, Monte Carlo simulation consists of trivial steps in algorithm, it needs a powerful computer. The repetition size plays a crucial role in accuracy. It has been proven that the accuracy of Monte Carlo simulation increases as repetition size increases. Therefore, choosing optimal size is quite important. Because, too much repetition causes

---

[4] See Zaman (1996)

more time in calculation or may result with out of memory during simulation. Conversely, insufficient repetition size damages the accuracy of methods which uses Monte Carlo. Usually, 1000 iteration can be regarded as sufficient. Therefore, throughout the analyses in this thesis, Monte Carlo repetition size will be taken as 1000.

1.3 Bootstrap

When Bootstrap was not used, there was a popular mean and variance estimation method for unknown distribution which is called **Jacknife** and introduced by **Quenouille** and **Tukey**. However, **B.Efron(1979)** introduced a method which is simple but more widely applicable than Jacknife, called **Bootstrap**. This method gives not only more accurate results than Jacknife but also correctly estimates the cases where Jacknife is known to fail.[5] e.g. the variance of the sample median

1.3.1 Theory of Bootstrap

Suppose that we have a random sample of size n which is observed from a completely unspecified probability distribution F.

$x_1, \ldots, x_n \sim \text{iid } F$

If R(**X**,F) is defined as given a specified random variable, it is intended to estimate the sampling distribution of R on the basis of the observed data and unknown distribution F. This was also the principle of Jacknife.

Traditional Jacknife theory focuses on two particular choices of R. One is, finding some parameter of interest, $\theta(F)$ such as mean, correlation or standard deviation of F. and the second is, finding an estimator of $\theta(F)$, say t(x), e.g. sample mean, sample correlation and sample standard deviation.

[5] See Efron (1979)

Then the choice of R(**X**,F) can be defined as

$$R(\mathbf{X},F)=t(x)-\theta(F) \qquad \textbf{(1.3)}$$

In second approach, Equation 1.3 is modified by injecting bias and variance of estimator into R term,

$$R(\mathbf{X},F)=(\, t(x)-bias(t)-\theta(F))/\,(var(t))^{1/2} \qquad \textbf{(1.4)}$$

However, Bootstrap does not need such complex estimation. Furthermore, these variables do not play any special role in Bootstrap theory. Bootstrap method consists of simple steps in principle.

Step 1: Construct the sample probability distribution $\bar{F}$, putting mass 1/n at each point $x_1,\ldots,x_n$.

Step 2: With $\bar{F}$ fixed, draw a random sample of size n from F say $x^1*,\ldots,x^n* \sim$iid $\bar{F}$. Define $\mathbf{X}*=(x^1*,\ldots,x^n*)$

Step 3: Approximate the sampling distribution of R(**X**,F) by the Bootstrap distribution of

$$R*=R(\mathbf{X*},\ \bar{F})$$

The key issue for Bootstrap technique is the approximation of F by $\bar{F}$. If this fails, this technique may lead some unexpected results.

The difficult part of Bootstrap procedure is actual calculation of the Bootstrap distribution. Efron (1979) presented three methods.

1. Direct theoretical calculation.

2. Monte Carlo approximation to Bootstrap. Repeated realizations of X* are generated by taking random samples of size n from $\bar{F}$, say $x*^1,\ldots,x*^N$ and histogram of the corresponding values $R(\mathbf{x}*^1, \bar{F})$, $R(\mathbf{x}*^2, \bar{F}),\ldots R(\mathbf{x}*^N, \bar{F})$ are taken as an approximation to the actual Bootstrap distribution.

3. Taylor series expansion methods can be used to obtain the approximate mean and variance of Bootstrap distribution of R^*.

We plan to use second method for calculation of Bootstrap distribution. In standard regression model problem, there are two unknowns; β and distribution F of ε. However, Monte Carlo method requires these two unknowns to be known. In Bootstrapping, by the help of maximum likelihood estimators, these unknowns can be estimated. If it is assumed that F is Normal with mean 0 and variance σ^2, then F distribution is reduced to a parametric family. This kind of Bootstrap is called *Parametric Bootstrap*.

For *Nonparametric Bootstrap*, these assumptions about F are relaxed and we can use empirical distribution of ε_i to estimate F by using observed ε_i's.

1.3.2 Remarks

Accuracy of Bootstrap methods depends on certain conditions. Each of these conditions play important role in theory. Therefore, they all should be satisfied for obtaining correct result. This section clearly discusses these conditions.

Continuity:[6] Assume the regression model $y=X\beta+\varepsilon$ where β and distribution F of ε are unknown. The distribution of $n^{-1/2}.(\bar{\beta}-\beta)$ is denoted as $\psi(F)$ where $\bar{\beta}$ term is estimated parameter of β. The continuity principle requires that $\psi(F)$ should be continuous for all values of F, i.e. actual density. Because, Bootstrap procedure derives the distribution of $n^{-1/2}.(\beta^*-\bar{\beta})$ depending on $\psi(\bar{F})$ where $\bar{F}$ is close to F. Therefore, Bootstrap completely fails at the points where $\psi(F)$ is discontinuous.

Centered Residuals:[7] Before resampling in Bootstrap procedure, residuals should be centered. It is done simply by substracting the mean of estimated residuals from each

[6] See Zaman (1996)
[7] See Freedman (1981)

estimated residual. (i.e. $\bar{\varepsilon}_i - \bar{\mu}$ where $\bar{\mu}=1/n \, \Sigma^n_{i=1} \, \bar{\varepsilon}_i$). What happens if the residuals are not centered before resampling ? Suppose the constant vectors are neither included in nor orthogonal to the column space of X, then distribution of $n^{-1/2}.(\beta^* - \bar{\beta})$ incorporates a bias term which is random (depending on ε_1 ,...., ε_n) and which in general has a degenerate normal limiting distribution. Because of this case , Bootstrap will usually fail. Note that constants are usually included, so this is not of importance in most regression models.

2
Asymptotic Efficiencies of Bootstrap and Normal Approximation Based On Central Limit Theorem

2.1 Preliminaries

So far, the theory of Bootstrap has been investigated. In this chapter, we discuss when Bootstrap works better than CLT Approximation. That is, in what sense, we can conclude that Bootstrap gives more accurate results than CLT Approximation.

Before passing to analysis, it is convenient to check whether maximum likelihood for linear regression model is consistent or not.

In the model, we restrict ourselves to choose the maximum likelihood estimators $\bar{\beta}$ and $\bar{\sigma}$ from compact sets **B** and Σ. Therefore, along real line, these sets are bounded. The compactness of sets provides **E** $\mathrm{Sup}_{\beta \in B, \sigma \in \Sigma}\, f(x,y, \bar{\beta}, \bar{\sigma}) < +\infty$. Note that we have to put a light on the case $\bar{\sigma}^2 = 0$. In this case, the estimator indicates that the random variable exactly equals to its mean. Probability distribution function becomes $f(x,y, \bar{\beta},0)=0.\infty$ which is undetermined. However, by the help of a calculus rule,(i.e. *L'Hospital Rule*), we see that exponential term converges to 0 faster than the first term of the Normal distribution formula. Therefore, it is convenient to write

$$\lim_{\bar{\sigma}\to 0} f(x,y, \bar{\beta}, \bar{\sigma}^2)=0$$

which satisfies the dominance of $E\rho(x,y, \bar{\beta}, \bar{\sigma}^2) < \infty$ for all $\bar{\beta} \in B$ and $\bar{\sigma} \in \Sigma$ where B and Σ are compact sets.

By the help of information inequality, discussed before, we can conclude that maximum likelihood is strongly consistent for linear regression model and $\bar{\beta}_n$ converges to true parameter β almost surely.

The next section, briefly, discusses the theoretical analyses of Bootstrap and Normal approximation based on CLT. The performances of both methods will be discussed by the help of Edgeworth Expansion.

2.2 Edgeworth Expansion

In this section, we will analyze an important question. Under which conditions Bootstrap and CLT Approximation give accurate results, i.e. the results which are close to results found by Monte Carlo, has been discussed by many authors. **Navidi (1989)** and **Singh (1981)** theoretically proved that Bootstrap estimation gives more accurate results than that for Normal Approximation based on CLT under certain conditions.

I intended to go on this proof more formally, and theoretically show that Bootstrap converges to true distribution rather faster than Normal Approximation. Edgeworth Expansion is a useful tool for theoretical proof.

In standard linear regression model, $\mathbf{y=X.\beta+\varepsilon}$, maximum likelihood estimate of $\bar{\beta}$ gives us that $\mathbf{E}\ \bar{\beta}=\beta$ and $\mathrm{Var}(\bar{\beta})=\sigma^2(X'X)^{-1}$. Then, we can write,

$$E\frac{(X'X)^{1/2}.(\hat{\beta}-\beta)}{\sigma}=0 \quad \text{and} \quad Var((1/\sigma)(X'X)^{1/2}.(\bar{\beta}-\beta))=1 \qquad \textbf{(2.1)}$$

Maximum likelihood estimate for β is found as $\bar{\beta}=(X'X)^{-1}X.y$. In our analysis, X and y are tx1 vectors, denoting observed sample.(t) is the sample size of observation which makes maximum likelihood estimate a scalar not a vector. Now, let's find out n maximum likelihood estimates of β by Monte Carlo by choosing different samples, X_i and y_i, with same size, t, from whole data. ($\bar{\beta}_i=(X_i'X_i)^{-1}.X_i.y_i$ where i=1...n). Note that, n is the Monte Carlo sample size which is quite different from observation sample size,t. Then by the help of *Kolmogorov's IID Strong Law*, since **E** $\bar{\beta}<\infty$, we can write the mean of Monte Carlo sample as $\boldsymbol{\beta}=1/n\ \Sigma^n_{i=1}\ \bar{\beta}_i$. Let's define $S_n=n^{1/2}\ (X'X)^{1/2}\ (\boldsymbol{\beta}-\beta)\ /\sigma$ or more explicitly.

$$S_n=\frac{\sum_{i=1}^{n}(X_i\ 'X_i)^{1/2}(\hat{\beta}_i-\beta)}{\sqrt{n.\sigma^2}} \qquad \textbf{(2.2)}$$

Central limit theorem which is the part of Normal Approximation method, indicates that S_n converges to Normal distribution with mean 0 and variance 1. Mathematically, we can prove this by the help of *Cumulant Generating Function.*

Cumulant Generating Function is defined as $K(\theta,X)=$ *log* **E** $e^{\theta x}$, in other words, it is the logarithm of moment generating function.[8]

Proceeding with proof, cumulant generating function of S_n will give us,

[8] The properties of Cumulant Generating Function will not be discussed here. However, some of them will be used through out the proof.

$$K\left(\frac{\sum_{i=1}^{n}(X_i'X_i)^{1/2}(\hat{\beta}_i-\beta)}{\sqrt{n\sigma^2}},\theta\right) = K\left(\sum_{i=1}^{n}\frac{(X'_i X_i)^{1/2}(\hat{\beta}_i-\beta)}{\sigma},\frac{\theta}{\sqrt{n}}\right) \quad \textbf{(2.3)}$$

It is seen that maximum likelihood estimate, $\bar{\beta}_i$ completely depends on regressors X_i. From the property of cumulant generating function, Eq. 2.3 can be written as;

$$K(.,.) = \sum_{i=1}^{n} K\left(\frac{(X_i'X_i)^{1/2}.(\hat{\beta}_i-\beta)}{\sigma},\frac{\theta}{\sqrt{n}}\right) \quad \textbf{(2.4)}$$

Also, we state that β term is obtained by errors which are i.i.d. Therefore, we proposed that $\bar{\beta}_i$ terms are also i.i.d. This information provides us to write above sum as below.

$$K(.,.) = n.K\left(\frac{(X_i'X_i)^{1/2}.(\hat{\beta}_i-\beta)}{\sigma},\frac{\theta}{\sqrt{n}}\right) \quad \textbf{(2.5)}$$

Taylor series expansion of cumulant generating function at the right hand size of Eq. 2.5 around zero will gives us,

$K(.,.)=n.\{\kappa_1[\,(X_i'X_i)^{1/2}.(\,\bar{\beta}_i-\beta)/\sigma\,].\,\theta.\,.n^{-1/2} + 1/2.\,\kappa_2[(X_i'X_i)^{1/2}.(\,\bar{\beta}_i-\beta)/\sigma].\,\theta^2.n^{-1} + ..\}$ **(2.6)**
where κ_j term is j^{th} cumulant of $(X_i'X_i)^{1/2}(\,\bar{\beta}_i-\beta)/\sigma$ and defined as

$$\kappa_j = \frac{\partial^j K(.,.)}{(\partial\theta)^j}\Big|_{\theta=0}$$

These cumulants keep special features inside. $\kappa_1(.)$ gives the mean of $(X_i'X_i)^{1/2}.(\,\bar{\beta}i-\beta)/\sigma$ which is zero in Eq.2.1. Therefore, first term of the Taylor expansion disappears. Furthermore, $\kappa_2(.)$ gives the variance which is 1. (See Eq. 2.1). Hence, Eq. 2.6 becomes,

$$K(.,.)=1/2\,.\,\theta^2+ 1/3!\,.\kappa_3.\theta^3.n^{-1/2} + 1/4!.\,\kappa_4.\theta^4.n^{-1}+ ....$$

$$K(.,.)=1/2\,\theta^2 + 1/3!.\,\kappa_3.\theta^3.n^{-1/2} + O(n^{-1}) \qquad \textbf{(2.7)}$$

The first term is a familiar term which is cumulant generating function of ~N(0,1). Therefore, by uniqueness property of cumulant generating function , we can conclude that S_n asymptotically converges to ~N(0,1). Note that the rest of Taylor series converges to zero as $n \to \infty$. This is the main propety of Normal approximation. Furthermore, third cumulant, κ_3 term, which catches the skewness of distribution play crucial role in approximation for n is not large enough. Fourth term, κ_4, which measures kurtosis can also affect the approximation. However, this thesis mainly focuses on skewness part which is more important than kurtosis. Therefore, in later discussions, skewness part will be considered.

Later, it will be discussed that skewed distributions, $\kappa_3 \neq 0$, causes CLT method to converge Monte Carlo distribution rather slow. However, for symmetric distributions (i.e. $\kappa_3=0$) , Eq 2.7 shows that distribution converges to standard Normal distribution faster. The order of approximation changes from $O(n^{-1/2})$ to $O(n^{-1})$.

Now, let's try to estimate the distribution of S_n by Edgeworth expansion.

$$K(S_n,\theta)=1/2.\theta^2 + 1/3!\,.\kappa_3.\theta^3.\, n^{-1/2} + 1/4!.\,\kappa_4.\,\theta^4.\, n^{-1} + \ldots$$

Converting Eq. 2.7 into moment generating function and using power series expansion of e^x, we, consequently, reach that

$$P(S_n<x) = \Phi(x) + n^{-1/2}.\, q(x).\, \phi(x) + O(n^{-1})$$

The terms Φ, ϕ, q are standard Normal distribution, standard Normal density and an even quadratic polynomial respectively. The rest of terms are written in order of approximation form. It means that the rest of terms are converging to some bounded constant, $M<\infty$, with order of n^{-1}.

Now, suppose that

$$S_n^* = n^{1/2}.\ (X'X)^{1/2}(\beta^* - \bar{\beta})\ /\ \bar{\sigma} \text{ where } \beta^* = 1/n.\ \Sigma^n_{i=1}\ \beta_i^* \qquad \textbf{(2.8)}$$

β_i^* terms are obtained by Monte Carlo simulation of Bootstrap. Similarly, Edgeworth expansion for S_n^* gives us

$$P(S_n^*<x) = \Phi(x) + n^{-1/2}.\ \bar{q}(x).\ \phi(x) + O_p(n^{-1}) \qquad \textbf{(2.9)}$$

After constructing the distribution of maximum likelihood estimate of β and Bootstrap estimation of β, the accuracy of CLT Approximation and Bootstrap can be calculated by order of approximation, i.e. the rate of convergence can be treated as accuracy of method. So, accuracy of CLT Approximation is calculated as follows.

$$H(.) = P(S_n<x) - \Phi(x) = n^{-1/2}.q(x).\phi(x) + O(n^{-1}) = O(n^{-1/2}) \qquad \textbf{(2.10)}$$

This means H term goes to some bounded constant , supposing $q(x).\phi(x)=M<\infty$, of order $n^{-1/2}$. For Bootstrap accuracy, we simply subtract distribution of Maximum likelihood estimation from that of Bootstrap distribution. i.e.

$$H(.) = P(S_n^*<x) - P(S_n<x) = n^{-1/2}.\ (\ \bar{q}(x) - q(x))\ .\ \phi(x) + O(n^{-1}) \qquad \textbf{(2.11)}$$

Hall (1992) indicated in his book that $\bar{q}-q=O_p(n^{-1/2})$[9]. This modifies Equation 2.11 as

[9] See Hall, pg 83

$$H(.)=P(S_n^*<x)-P(S_n<x)= O_p(n^{-1}) \qquad \textbf{(2.12)}$$

If we compare order of approximation of CLT Approximation and Bootstrap estimation, it is seen that Bootstrap converges much faster than CLT Approximation. Therefore, we can easily conclude that Bootstrap gives more accurate results asymptotically.

Next section is devoted to empirical analysis of this comparison. It has been divided into two sections. In theoretical work, it has been proven that in all cases, Bootstrap works as good as CLT Approximation. Furthermore, in most cases, it works better, i.e. converges faster than CLT Approximation. Therefore, we will concentrate on which cases Bootstrap works as good as or better than CLT Approximation. Actually, these cases completely depend on the distribution of unknown parameters. Skewness of distribution plays crucial role in performance of estimation methods. So, let's investigate the cases where errors are symmetrically and asymmetrically distributed.

Before passing through empirical analysis, it is necessary to follow two guidelines suggested by **Hall** and **Wilson (1991)**. Because, these guidelines provide good performance in many important statistical problems. The first guideline is to resample $\beta^*-\bar{\beta}$ not $\beta^*-\beta$. This has the effect of increasing power in the hypothesis testing. The second guideline is to base the test on the Bootstrap distribution of $(X'X)^{1/2}.(\beta^*-\bar{\beta})/\sigma^*$, not on the Bootstrap distribution of $(X'X)^{1/2}.(\beta^*-\bar{\beta})/\bar{\sigma}$. With this guideline, we reduce the error in the level of significance. The device of dividing by σ^* is known as *Bootstrap Pivoting* which provides the statistic $(X'X)^{1/2}(\beta^*-\bar{\beta})/\sigma^*$ asymptotically pivotal.[10] Considering these guidelines, we can investigate our model depending on shape of distribution.

[10] See Hall and Wilson, 1991 for details

2.3 Symmetrically Distributed Errors

In this model, we assume that errors are i.i.d. and distributed symmetrically. An example for symmetric distribution is Uniform distribution. Therefore, for computer simulation , it is convenient to use errors which are ~U(-0.5,0.5). The distribution of errors is plotted in Figure 2.1

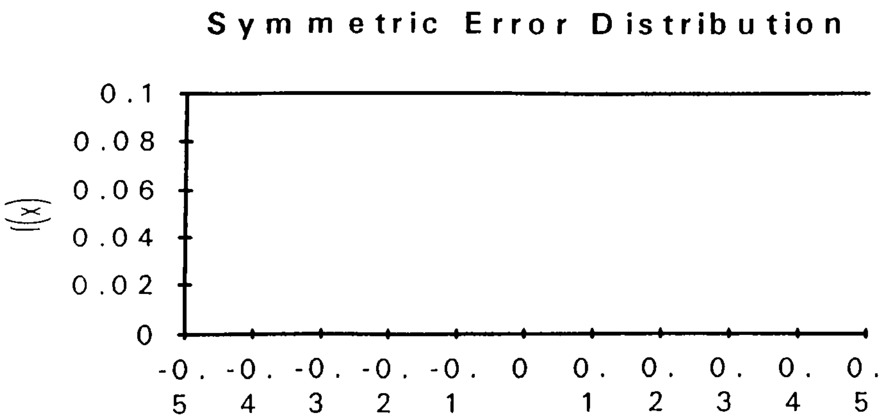

Figure 2.1 Probability distribution of errors

When distribution of errors is symmetric, the ML estimate of β which is $\bar{\beta}=(X'X)^{-1}.X'y$ has a symmetric distribution. Actually, this condition completely, depends on the positive definiteness and nonsingularity of X'X term. Symmetric distribution of $\bar{\beta}$ causes $E(\bar{\beta}-E\bar{\beta})^3$ to be zero, since there is no skewed part in symmetric distributions. This is actually same thing with $\kappa_3=0$. This helps Normal approximation based on CLT to converge Monte Carlo distribution faster than before. The order of approximation changes to $O(n^{-1})$. (See Eq. 2.10). Furthermore, we have derived that order of approximation for Bootstrap is $O(n^{-1})$. (See Eq.2.12). This means, generally, Bootstrap works at least as good as CLT approximation even the distribution is symmetric. If we choose our errors which are Normal with mean 0 and variance σ^2, then Monte Carlo becomes useless. Because, for Normal errors, Normal approximation based on CLT is exact. Therefore, there is no need to calculate even Monte Carlo in this case.

In the analysis of this part, we will use errors which are distributed uniformly between -0.5 and 0.5. Mean of errors is 0. Therefore, residuals are centered. We will depict the distribution of S_n which is Monte Carlo distribution, (see Eq.2.2); S_n* which is Bootstrap distribution (see Eq.2.8), and standard Normal distribution which represents Normal approximation based on CLT. Figure 2.2 depicts these curves. The curve with small lines (_ _ _) shows Monte Carlo. Solid line represents Bootstrap and curve with dots represents CLT approximation. Since Monte Carlo and Bootstrap give broken curves due to construction of Histogram, we can not assess which method is better. However, in later section, we will introduce an efficient smoothing procedure.

2.4 Asymmetrically Distributed Errors

It is interesting to check what happens to the model when errors are asymmetrically distributed. We can assign many asymmetric distribution for errors. For this case, a simple trick is used for constructing asymmetric distribution. The distribution of errors is composed by chopping distribution into half and derived errors in each half according to ~N(0,1) and ~N(0,9) respectively. Hence we can construct a distribution which is asymmetric for usage in modelling.

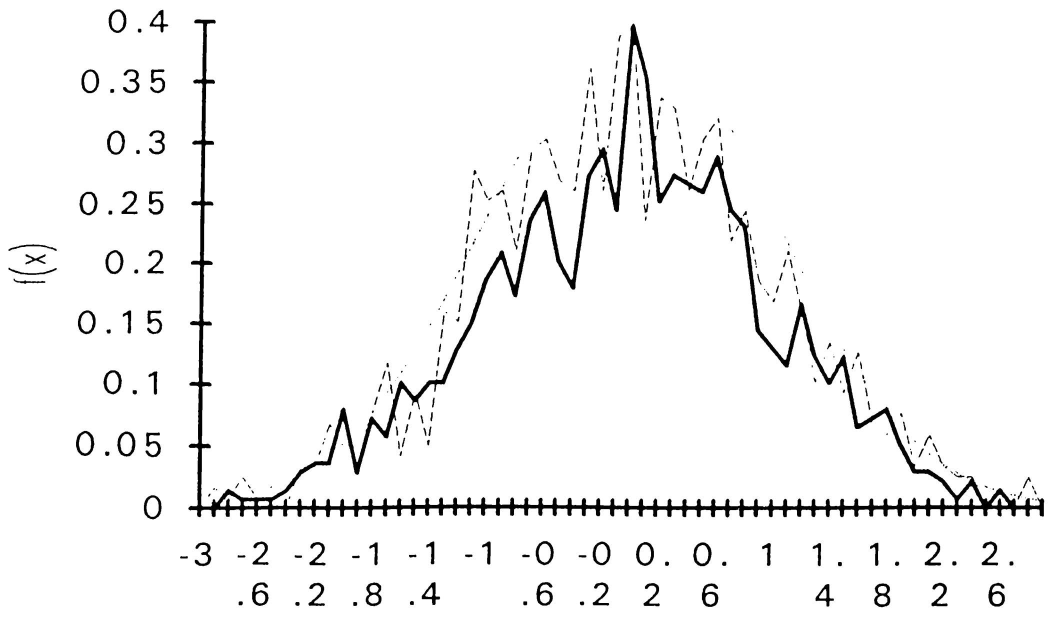

Figure 2.2 Unsmoothed distribution of S_n, S^*_n and ϕ for errors ~U(-0.5,0.5)

Actually, mean of errors for this case inevitably shifted away from zero a little bit. This violates the centered residuals condition and so , it should be corrected. We can eliminate this drawback by subtracting all residuals from their mean. This will provide us to have centered residuals for the analysis. Figure 2.3 shows the distributed errors which is asymmetric.

Asymmetric distributed errors cause estimated β term to be distributed asymmetrically too. Since $\kappa_3 \neq 0$ for asymmetric distributions because of skewness, it is supposed that CLT Approximation diverges from accurate results. Because, structure of Normal Approximation based on CLT is not suitable to catch skewness part of unknown distribution. In this case, Bootstrap should work better and catch this asymmetry part where CLT Approximation fails.

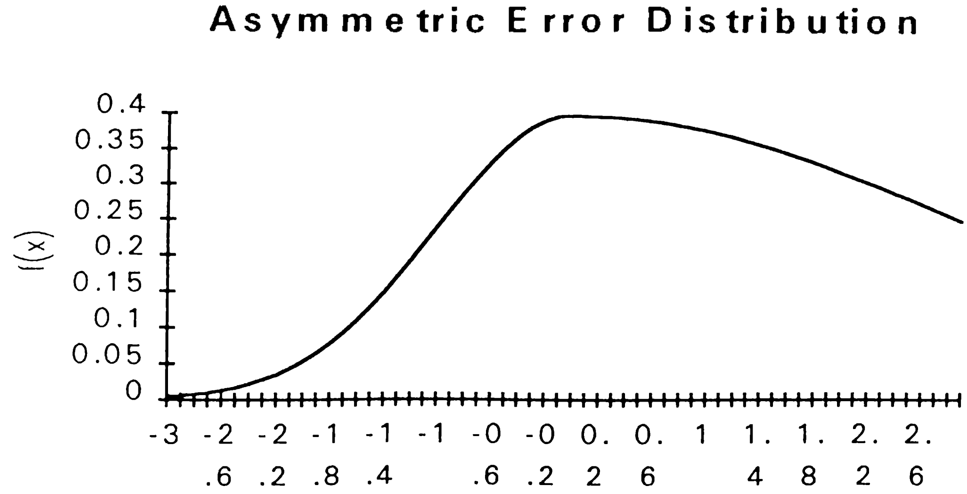

Figure 2.3

Considering above assumptions, we have to face that Bootstrap curve should be closer to Monte Carlo distribution curve. Figure 2.4 sketches the distributions of Monte Carlo, Bootstrap and standard Normal for asymmetric distribution case.

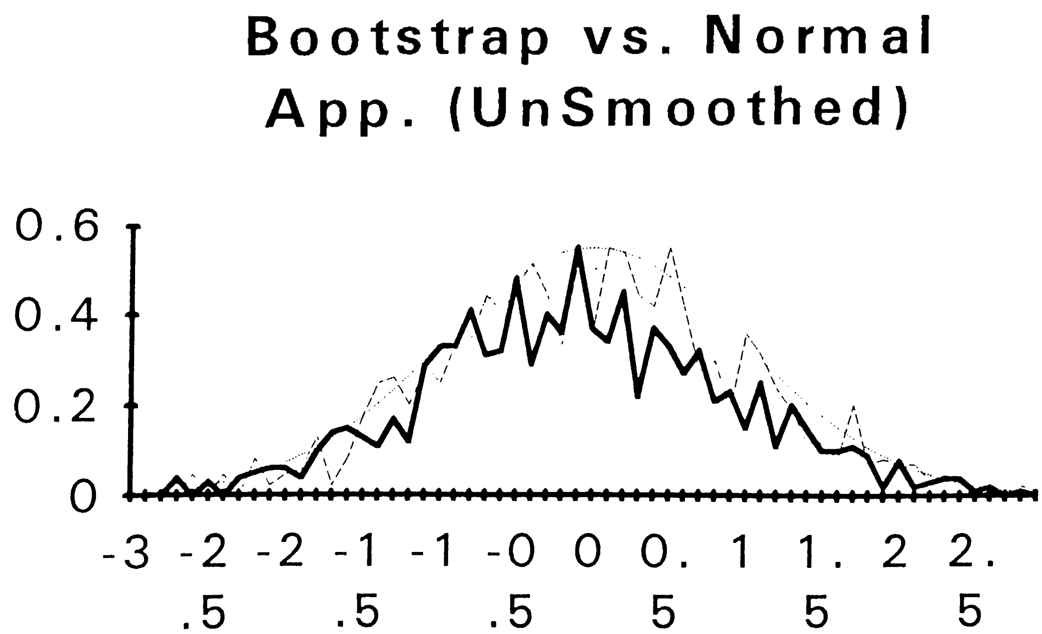

Figure 2.4 Distributions of S_n, S^*_n and ϕ for asymmetrically distributed errors

It is seen that these broken lines do not give clear idea about how good Bootstrap estimation works from above figure. It is necessary to smooth these curves. The following section discusses how to smooth these curves.

2.5 Smoothing Procedure

So far, the comparison of Normal Approximation and Bootstrap estimation was analytically investigated and computer simulations were performed. However, in the light of curves derived by Monte Carlo simulations, it is difficult to predict which method is better than the other. Here, it is necessary to introduce a smoothing procedure to give clear vision of comparison of these curves. **B.W.Silverman, 1986,** in his book called, *"Density Estimation For Statistics and Data Analysis"* introduced a smoothing procedure which is quite convenient for our model.

The problem of choosing how much to smooth is of crucial importance in density estimation. It should never be forgotten that the appropriate choice of smoothing parameter will always be influenced by the purpose for which the density estimate is to be used.

Before passing through smoothing procedure, it should be chosen an appropriate **kernel** for unknown distribution. Kernel estimates should satisfy the condition

$$\int_{-\infty}^{\infty} K(x).dx = 1$$

where usually, but not always, K will be a symmetric probability density. Silverman proposes many different kernel estimates to smooth density, but which kernel will be used is an important question to be considered. Below table **(Table 2.1)** gives the different types of kernels and their efficiencies which is calculated out of 1.

Kernel	Efficiency
Epanechnikov	1
Biweight	0.9939
Triangular	0.9859
Gaussian	0.9512
Rectangular	0.9295

Table 2.1 Some kernels and their efficiencies

It is seen from above table that all these kernels are quite efficient for estimation. In this model, **Gaussian** kernel is chosen for smoothing procedure. The formula of the Gaussian kernel is defined as

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)$$

Note that it has same formula with standard Normal density. This will provide convenience in our analysis. Because, the distribution of estimated β is converted to a distribution which has mean 0 and variance 1.

The working principle of smoothing procedure is as follows. The kernel estimator is a sum of "bumps" placed at the observations. Therefore, kernel function, K determines the shape of bumps while the window width, denote as h, determines their width. When h tends to zero, kernel function spikes at the observations, but while h becomes large, all details, spurious and others, are obscured. Therefore, after choosing the suitable kernel, an appropriate window width should be calculated. Silverman has calculated the optimal window width for Gaussian kernel as[11]

[11] See Silverman, pg 45

$$h=1.06\,\sigma\, n^{-1/5}$$

In our model we proposed that variance is 1. Furthermore, sample size, n, has been chosen 100 before. Hence, appropriate window width is found as 0.422. Based on this values, the smoothing density function is defined as

$$f(x)=\frac{1}{nh}\sum_{i=1}^{n}K(\frac{x-X_i}{h})$$

where n and h are sample size and optimal window width respectively. X_i's are observations in sample data. Consequently, we can reach a smooth distribution by injecting Monte Carlo simulation results into above function.

Figure 2.5 and 2.6 clearly depict the smoothed versions of S_n, S^*_n and ϕ distribution for symmetrically and asymmetrically distributed errors respectively. Now, we can investigate

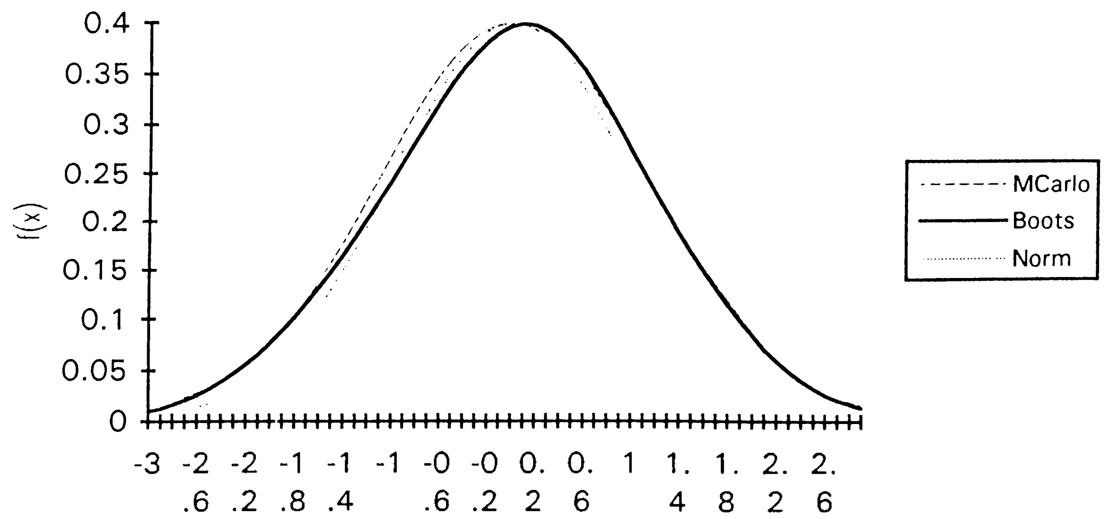

Figure 2.5 Smoothed distribution of S_n, S^*_n and ϕ for errors~U(-0.5, 0.5)

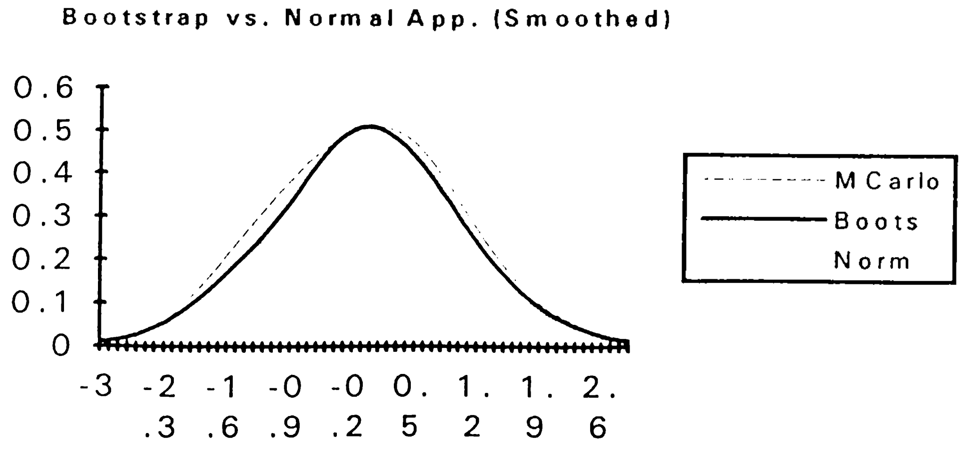

Figure 2.6 Distribution of S_n, S^*_n and ϕ for asymmetrically distributed errors

the effect of changes in regressor X. So far, when we use asymmetrically distributed errors, we have taken regressor X as 90% -1 or 1 and 10% -3.3 or 3.3. This provides X to be nonsymmetric. It can be interesting to investigate the behavior of curves when regressor X is symmetric which takes either -1 or 1. Figure 2.7 depicts the case when X is symmetric.

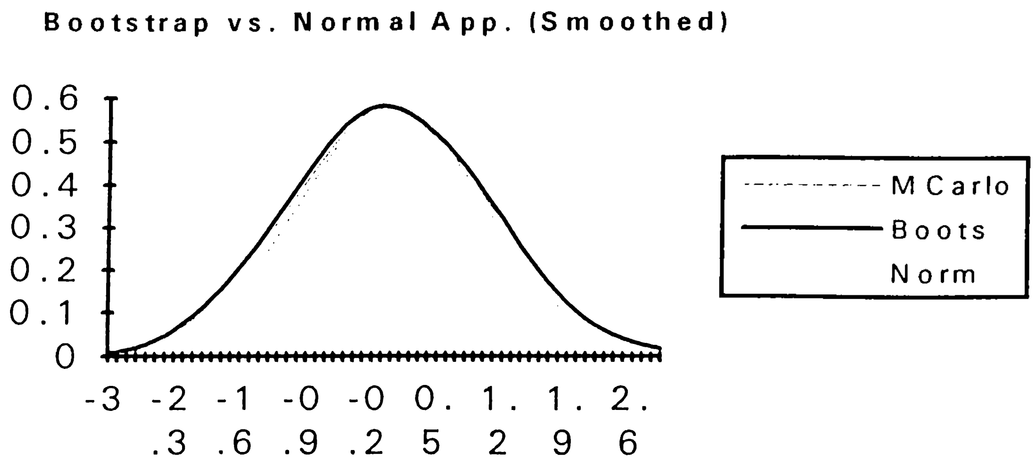

Figure 2.7 Distribution of S_n, S^*_n and ϕ when X is either -1 or 1

It is seen that when we use symmetric regressors, X, both Bootstrap and CLT approximation curves approach to Monte Carlo curve. Actually, from the figures, it may be difficult to predict which curve is closer than the other. Figure 2.8, 2.9 and 2.10 depict the difference between Bootstrap and Monte Carlo; CLT approximation and Monte Carlo.

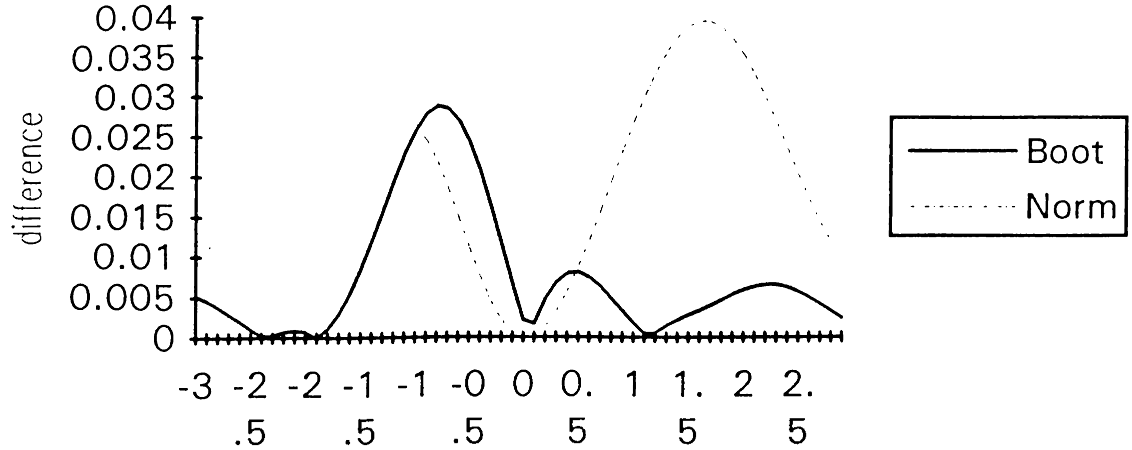

Figure 2.8 Differences of curves to Monte Carlo when errors are ~U(-0.5, 0.5)

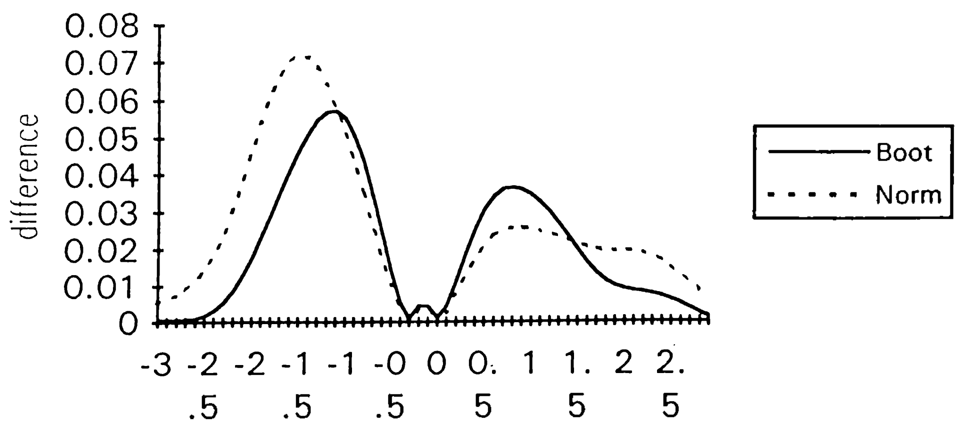

Figure 2.9 Differences of curves to Monte Carlo when errors are asymmetrically distributed and X is asymmetric

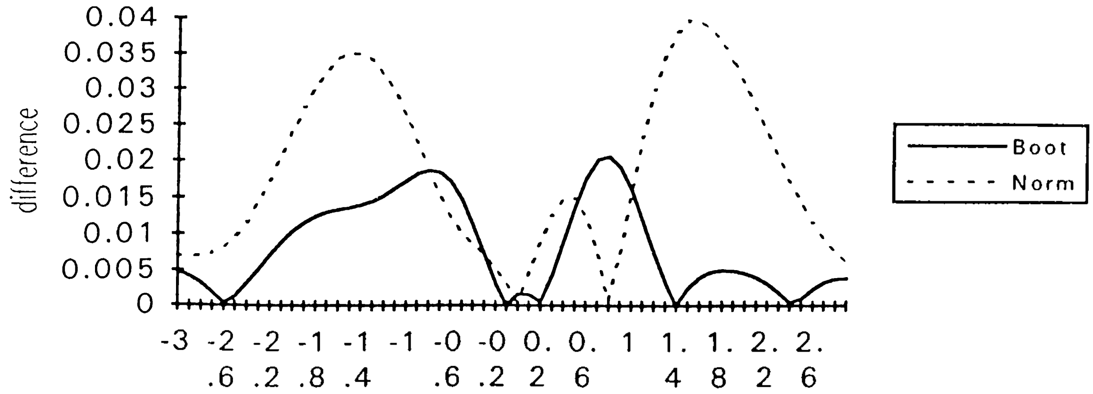

Figure 2.10 Difference of curves when errors are asym. distributed and X is symmetric

We will introduce a key word, **maximum gap** which is helpful in our analysis. It is defined as

$$\mathrm{Sup}_{\beta\in B}\,|\,f^{\bar{\beta}} - f^{\beta^*}\,| \quad \text{or} \quad \mathrm{Sup}_{\beta\in B}\,|\,f^{\bar{\beta}} - \Phi| \qquad \textbf{(2.13)}$$

Eq. 2.13 will provide a tool to comment on efficiencies of estimation methods. In Figures 2.8, 2.9, 2.10, it is seen that Bootstrap works better than CLT Approximation. There are some points where CLT approximation curve is closer to Monte Carlo curve than Bootstrap. However, in most of the points, Bootstrap approaches to Monte Carlo curve more than CLT approximation. In our analysis, the maximum gaps between curves clearly show that Bootstrap gives more correct results than CLT approximation. Because of this, we can conclude that Bootstrap works better than Normal Approximation based on CLT. Below table **(Table 2.2)** indicates the methods and their maximum gaps for both symmetric and asymmetric distributions.

	Bootstrap	**Normal App.**
Symmetric Distribution	0.027182	0.037718
Asym. Distr. when X is Symm.	0.020672	0.039569
Asym. Distr. when X is Asymm.	0.057052	0.071729

Table 2.2 Methods and their maximum gaps

Maximum gaps also clearly validate that Bootstrap is more suitable to use. It will be misleading to use CLT approximation instead of Bootstrap to estimate distribution.

2.6 An Example

The empirical analysis showed that the results are consistent with the results of theoretical approach. However, it should be considered whether our analysis can be applied to real life statistics. The following analysis investigates an example data which is taken from the *Economic Report of the President, 1984 pq 261.*

The data contain per capita disposable income (Y) and per capita consumption expenditures for the period 1929- 1984. It is intended to estimate the consumption function for United states from the data. However, during estimation, it is found that there are some outliers in data. During 1942-1945, the observations deviate from estimated results quite a lot. Therefore, these observations were all disregarded.

The analyses which were done before, derived that true beta term is 0.885 and intercept is 85.725. Based on these values , the formula of fitted line is found as;

$$C_t = 85.725 + 0.885\, Y_t$$

We will perform our analysis in the light of these considerations. In this case, we will take intercept as known and try to estimate the coefficient in front of disposable income for sake of simplicity. Otherwise, X matrix becomes tx2 matrix which brings more work to do. Figure 2.11 clearly depicts the curves of S_n, S^*_n and ϕ based on actual data. Monte Carlo curve (S_n) is solid line. Bootstrap curve (S^*_n) is dotted line and CLT approximation curve is cutted line.

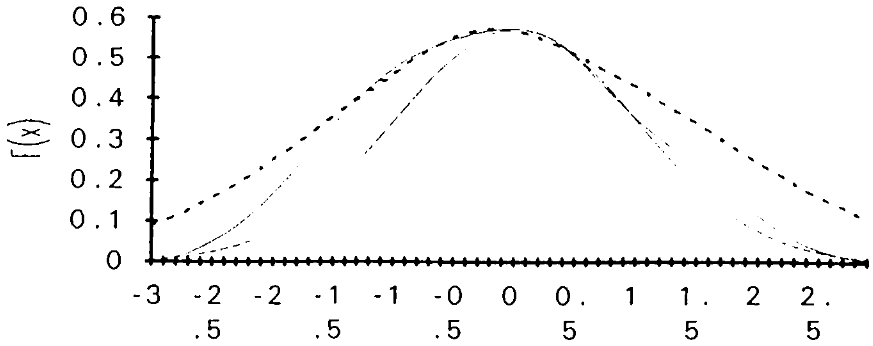

Figure 2.11 Distribution of S_n, S^*_n, ϕ based on actual data

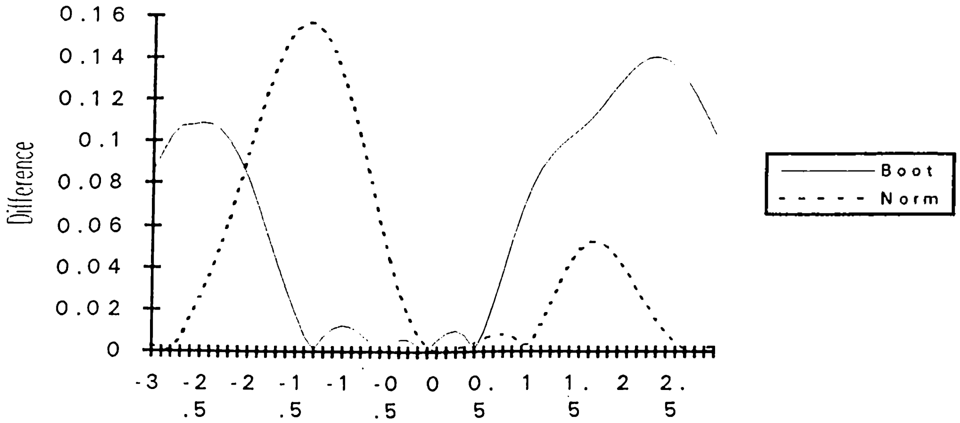

Figure 2.12 Difference of curves to Monte Carlo in real data

In Figure 2.11, the actual (Monte Carlo) distribution seems quite far away from being symmetric. We have proven that Bootstrap gives closer results than CLT approximation when distribution is asymmetric. In Figure 2.12, we see that Bootstrap is closer than CLT approximation in negative scale. But, on the other side, CLT approximation gives closer results than Bootstrap. Therefore, we have to look the maximum gap of curves in order to understand which method is better than the other in overall look. We found that maximum gap between Monte Carlo and CLT approximation is 0.312485 which is higher than the gap between Monte Carlo and Bootstrap (i.e. 0.280967). Therefore, we can conclude that Bootstrap works better than CLT Approximation considering overall performance. This example validates our analysis on artificial data. We have found that in asymmetric distributions, Bootstrap gives closer results than Normal approximation based on CLT.

At the end of our first application analysis, it is useful to check the behavior of Bootstrap when data contain leverage point. We inserted a good leverage point to the data through matrix X and replaced one value which was 1 with 10.

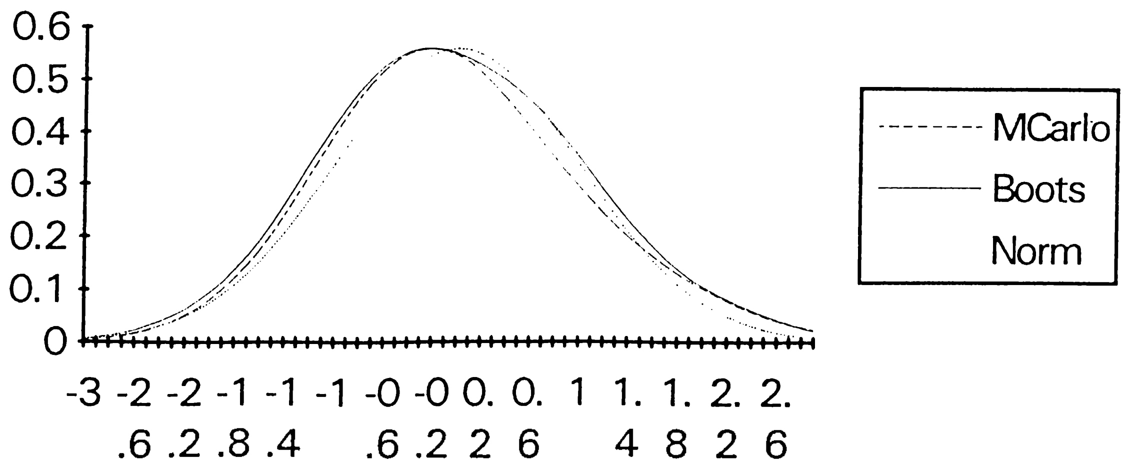

Figure 2.12 Distribution of S_n, S^*_n and ϕ when data contain leverage point

From Figure 2.12, when data contain leverage points, the maximum gap between Bootstrap and Monte Carlo curves increases to 0.058985. The gap between CLT Approximation and Monte Carlo also increases to 0.059551. As a result, we can easily conclude that leverage point slightly affects the accuracy of both Bootstrap and CLT Approximation. Nevertheless, Bootstrap still gives close results to Monte Carlo relative to CLT Approximation.

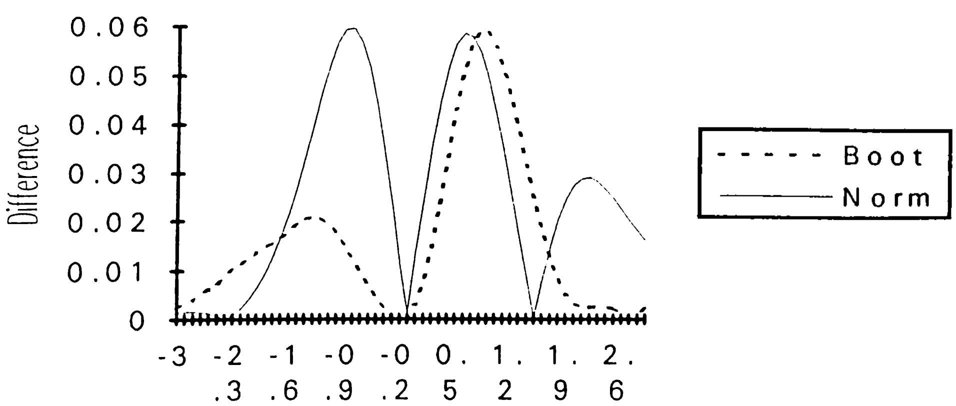

Figure 2.13 Difference of curves to Monte Carlo when data contain leverage point

3
Using Bootstrap in F-test

In the previous chapter, we have investigated the accuracy of Bootstrapping and Normal Approximation based on CLT.

3.1 Bootstrapping F-test when Model is 1st order ADL

In this chapter, it is intended to investigate some further applications of Bootstrap. In this section, we will mainly focus on using Bootstrap in F-test. F-test is a technique to test the significance of regressors. F-test works exactly when the regressors are nonstochastic, i.e. $Y_t = \beta . X_t + \varepsilon_t$. However, it has been proven that F-test asymptotically works when dynamic models are used. Precisely, F-test does not give true values when sample sizes are low. In this case, we can not use the table values as critical values of F-test. Monte Carlo simulation results are the true critical values. On the other hand, these results are not available to the experimenter, because he should know the values of true parameters (a_0, a_1,a_2) in order to compute critical values by Monte Carlo.

At this point, we use Bootstrap in order to compute critical values for different sample sizes. At first, we construct our model which is first order ADL model, i.e. $Y_t = a_0 . X_t + a_1 . Y_{t-1} + a_2 . X_{t-1} + \varepsilon_t$ and then , test whether Bootstrap would give close results to Monte Carlo simulation of 95% significance level of F-test. In our testing hypothesis, null hypothesis is $a_1=0$ and $a_2=0$ which restricts our dynamic model. On the other hand, alternative hypothesis is against null hypothesis, i.e. $a_1\neq 0$ and $a_2\neq 0$ which is unrestricted form of the model.

The F-test statistic is derived as follows.

$$F=\frac{(\|Y_t-\tilde{a}_0.X_t\|^2-\|Y_t-\hat{a}_0.X_t-\hat{a}_1.Y_{t-1}-\hat{a}_2.X_{t-1}\|^2)/2}{\|Y_t-\hat{a}_0.X_t-\hat{a}_1.Y_{t-1}-\hat{a}_2.X_{t-1}\|^2/(t-4)} \quad \textbf{(3.1)}$$

In the above equation, tilda (~) above a_0 denotes restricted maximum likelihood estimation of the coefficient in front of the X_t term under restriction $a_1=0$ and $a_2=0$. Meanwhile, the terms which have hat(^) above, are unrestricted maximum likelihood estimators. The critical value is the measure which depends on the significance level of test and helps us to reject or accept the null hypothesis. If F-test gives smaller value than critical value then we accept null hypothesis otherwise we should reject it. The aim of this thesis is to check that whether Bootstrap can be used in calculating critical value of F-test or not. If so, this will give us an important facility in calculating critical value. That means, the necessity of knowing the coefficient of regressors disappears and estimated coefficients can also reach same results that we are seeking.

During this experiment, we will use 95% as significance level and hold Monte Carlo iteration size enormously high such as 60.000 in order to find the exact value of critical value. After experiment, table 3.1 indicates the results for critical value of F-test according to Monte Carlo, Bootstrap and Table (Asymptotic) depending on different sample sizes.

SAMPLE SIZE	MONTE CARLO	BOOTSTRAP	TABLE (Asymp.)
10	4.67794546	4.5606907	5.14
25	3.344504	3.3125865	3.47
50	3.1633513	3.1257652	3.21
100	3.0855887	3.0663446	3.08

Table 3.1 Critical Values of F-test for different sample sizes

It is seen that Monte Carlo simulation results and asymptotic values become closer to each other when sample size is over 50. However, what we are concerned is whether Bootstrap approaches to exact values before size is 50. It is absolutely seen that Bootstrap technique also gives closer results to Monte Carlo results than table values when sample size is below 50. We observe that until sample size 50, we can use Bootstrap technique to find the critical value of F-test. However, for the cases when size is bigger than 50, it is preferable to use asymptotic values (i.e. table values) since the problem in F-test because of using dynamic models disappears. Moreover, after sample size 50, table values are more correct than Bootstrap values. Therefore, Bootstrap should be used when sample size is below 50 and left when size is over 50. Actually, sample size 50 may not be standard for F-test. It can change depending on regressor X. In later sections, we will concentrate on the effect of change in regressors.

3.1.1 Leverage Points Effect

We can also check the effect of existance of leverage points in the data set. This is actually checking the dependence of Bootstrap technique on matrix X. Because, if Bootstrap diverges away from correct results when there are leverage points, then we have to leave using Bootstrap in F-test completely. Leverage points can be easily generated by replacing the values of robust data X with some uncorrelated values by hand. For the sake of simplicity, we will investigate the effect of good leverage points. Since, bad leverage points also affect the result of Monte Carlo which is regarded as true value, we can not comment on rest of the analysis correctly.

During the experiment, we inserted 10 value into matrix X where each variable in X was distributed as Normal with mean 0 and variance 1. If we rerun our program according to this consideration, we will get the following results;

SAMPLE SIZE	MONTE CARLO	BOOTSTRAP	TABLE (Asymp.)
10	4.5910419	4.5226294	5.14
25	3.3464298	3.3185360	3.47
50	3.1128431	3.0782419	3.21
100	3.0447362	3.0168968	3.08

Table 3.2 Critical values of F-test when data contain leverage points

When data contains leverage points inside, we see that for small sample size, the performance of Bootstrap still looks quite good. As a result, Bootstrap can again be used in low sample sizes.Furthermore, leverage points do not change critical value of F statistic much, and the accuracy of Bootstrap stays still in considerable range. Therefore, generally, we can use Bootstrap in low sample sizes such as 10-50. However, table values that means asymptotic results gives better results when sample size is high such as over 50. Furthermore, existance of leverage points does not affect the analysis.

This opens a new idea to be investigated. Since leverage points do not affect critical value much , then it may be interesting to analyze the changes in critical value of F-test when all regressor X changes. In this analysis, we chose sample size as 20 and search for the critical values of F-test depending on different regressor X. Here, we dropped Monte Carlo iteration size from 60,000 to 10,000. Because, it is not needed to put much emphasis on very accurate results. In the first run, we created regressor X in which values are distributed as Cauchy. In second run, we changed X to the floor of uniformly distributed numbers between 0 and 5 and finally, in the third run, we created X distributed as ~N(0,16). Based on these regressors, Monte Carlo simulation gave critical value of F-test as below,

Matrix X	**Critical Value of F statistic for Sample 20**
Cauchy	3.3874558
~ Floor [UNIFORM(0,5)]	3.3741046
~ N(0,16)	3.4185837

Table 3.3 Critical value of F statistic for different regressors X

Table 3.3 shows that critical value of F-test does not completely depend on regressor X. The critical values are not changing too much as regressor X completely changes. This proves that F-test does not depend on regressor X much. Therefore, we can use F test whatever data X is.

3.2 Bootstrapping F-test when Model is 2nd Order ADL

After investigating the features of Bootstrapping in first order ADL model, it is also intended to seek the changes in results when model is changed to second order ADL, i.e. $Y_t=a_0.X_t+a_1.X_{t-1}+b_1.Y_{t-1}+a_2.X_{t-2}+b_2.Y_{t-2}+\varepsilon_t$.

Now, if null and alternative hypothesis are constructed as follows;

H_0: $a_2=b_2=0$ vs. H_1:$a_2\neq0$ and $b_2\neq0$

The aim of including second order ADL model in the analysis, is to make the model more dynamic. In this case, the required sample size for correct value of F statistic should increase. Furthermore, we will check the performance of Bootstrapping when the order of lag increases. After experiment, the critical values of F test based on second order ADL model are derived at below table.

SAMPLE SIZE	TABLE	MONTE CARLO	BOOTSTRAP
10	5.79	11.857767	12.74724
15	4.10	5.5250613	5.0368244
25	3.49	3.8339534	3.9817073
50	3.22	3.4521521	3.5058512
100	3.11	3.0218225	3.1034317

Table 3.4 Critical values of F when model is second order ADL

According to table 3.4, we see that there are two interesting results. The first one is, when sample size is 10, there is a big difference between asymptotic value and Monte Carlo value. Table value is absolutely incorrect. However, as sample size increases to 15, there occurs a sharp decrease in difference between asymptotic result and Monte Carlo result. Second interesting issue is, when model is changed to second order ADL model which is more dynamic than before, table values (which are asymptotic results) do not give correct results until sample size is over 100. Actually, even at sample size 100, Bootstrap gives more accurate result than asymptotic value.

Therefore, when our model is second order ADL, we should again use Bootstrap instead of table values for low sample sizes (10-100). Table values become correct when sample size is sufficiently high (e.g. >100).

So far, we have investigated the usage of Bootstrap technique in F-test. Consequently, Bootstrap gave better results than asymptotic values of F-test when sample size is low. This is because of dynamic nature of the model. It is known that F-test is asymptotically true in dynamic models. Therefore, we can use Bootstrap in low sample sizes where F-test does not work.

4
Using Bootstrap in Common Factor Restrictions

4.1 Using Bootstrap in COMFAC restrictions

In this chapter, we will use Bootstrap to test common factor restrictions. Suppose that our model is taken as follows:

$$Y_t=\alpha.X_t+\beta.Y_{t-1}+\gamma.X_{t-1}+\varepsilon_t \qquad \textbf{(4.1)}$$

To test common factor restriction, we propose testing hypotheses as follows;

$$H_0:\gamma=-\alpha.\beta \quad \text{vs.} \quad H_1:\gamma\neq-\alpha.\beta$$

Null hypothesis denotes that model has common factor, while alternative one denotes that model does not have a common factor and can be treated as linear one. The term γ can be regarded as actual coefficient of regressor X_{t-1}. The comparison criteria for tests with common factor should be reconsidered. It is appropriate to use **power** curves of tests to test their efficiencies. Since there is a possibility of being nonlinear for our model, we can not use the F-test for this case. Therefore, we intend to use likelihood ratio test for testing nonlinear models. In this test, LR statistic would give the ratio of likelihood functions of the model for two different hypotheses. For the null hypothesis, it can be difficult to guess the parameters, (α,β) simultaneously which maximize the likelihood function. Therefore, it is more appropriate to find these parameters by fixing one and finding the other parameter which maximizes likelihood, and then, changing the fixed parameter according to both parameters maximize likelihood function at global maximum value. Equation can be written as follows,

$$LR = \frac{Sup\|Y_t - \tilde{\alpha}.X_t - \tilde{\beta}.Y_{t-1} + \tilde{\alpha}.\tilde{\beta}.X_{t-1}\|^2}{Sup\|Y_t - \hat{\alpha}.X_t - \hat{\beta}.Y_{t-1} - \hat{\gamma}.X_{t-1}\|^2} \qquad \textbf{(4.2)}$$

In Eq.4.2, the parameters which have tilda (~) above, denotes the restricted maximum likelihood estimates. On the other hand, the parameters which have hat (^) above, denotes the unrestricted maximum likelihood estimates.

Throughout this analysis, we will define Bootstrap as follows. At first, the parameters, $\alpha^{\sim}$, $\beta^{\sim}$ are estimated by restricted ML. Furthermore, $\alpha^{\wedge}$, $\beta^{\wedge}$ and $\gamma^{\wedge}$ are estimated by unrestricted ML. Then, dependent variable, Y_t, is produced by using estimated parameters. We denote new dependent variable as Y_t*. Bootstrapped LR is obtained from Eq.4.2 by replacing Y_t* instead of Y_t.

Finally, it is worthwhile to discuss about another testing technique for common factor, called **Wald Test**[12]. This test is quite simple and rather efficient. The testing hypotheses for the Wald test are as follows ;

$$\text{Ho: } W = 0 \quad \text{vs} \quad H_1\text{: } W \neq 0$$

where W stands for Wald statistic. Actually, there are many derivation formula for W. In this thesis, we will simply find W as follows,

Let's define f(.) as

$$\mathbf{f(.)} = \bar{\gamma} + \bar{\alpha}.\ \bar{\beta}$$

[12] See G.Kemp (1991) for comlex Wald tests with application to COMFAC tests

These three parameters are derived from unrestricted maximum likelihood estimation. By dividing f(.) value into its standard deviation , Wald statistic can be derived easily.

$$W = f(.) / [\, var (f(.)) \,]^{1/2}$$

In empirical analysis part, we constructed our model as follows: $Y_t=\alpha.X_t+\beta.Y_{t-1} - \alpha.\beta.X_{t-1}$. As it is seen that the coefficient in front of X_{t-1} is the multiplication of other coefficients of other regressors with minus sign. Therefore, there is a common factor in the model. This model was firstly discussed by **Sargan**. Afterwards, **Hendry and Mizon** put some empirical analysis on this model. However, so far, there has been not so much analyses on using Bootstrap method on common factor models.

We will investigate the efficiencies of three different methods, *Likelihood Ratio, Bootstrap and Wald.* As we mentioned before, we are going to use simple Wald statistic. Reader may refer to studies of Hendry and Mizon[13] for more complex Wald statistics. Our hypotheses are;

$$H_0:\gamma+\alpha.\beta=0 \quad \text{vs.} \quad H_1:\gamma+\alpha.\beta\neq0$$

During the experiment, we have derived that 95% critical values for LR, Bootstrap and Wald are 4.0612399, 5.1852510 and 1.6411099 respectively. Actually, these critical values are not standard for all tests. They also depend on regressor X. That means there is no a certain critical value to be looked for as in F-test. Therefore, we have to use other comparison mechanisms to check the performance of tests. It is convenient to use **Power Curves** of each test.

We have used **Neyman-Pearson** to construct Power Envelope. We obtain Neyman-Pearson by dividing the probability density of Y_t under alternative hypothesis by the probability density of Y_t under null hypothesis. Now, our testing hypotheses become,

[13] See Mizon and Hendry (1980) for more information

$$\text{Ho: } \gamma+\alpha.\beta=0 \qquad \text{vs.} \qquad \text{H}_1\text{: } \gamma + \alpha.\beta = \xi > 0$$

In figure 4.1, Power curve of each test has been derived.

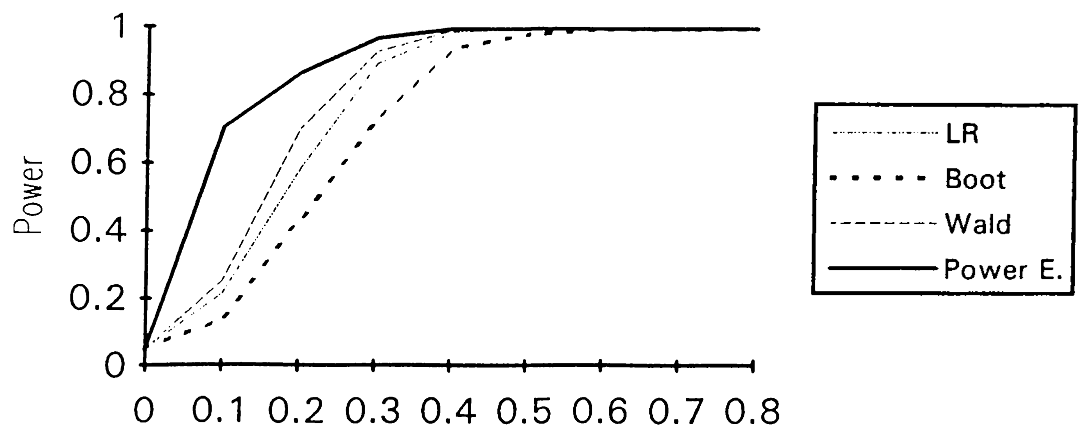

Figure 4.1 Power Curves of Wald, LR and Bootstrap

In the above figure, Wald test gives the closest power curve to the power envelope. That means Wald test provides more powerful tool in common factor restriction models. There is no need to use Bootstrap or LR test. In Figure 4.2, we can see the differences of tests to power envelope. The difference between Wald test and power envelope is minimum. LR works worser than Wald test but better than Bootstrap. Bootstrap is the worst test in our analysis.

So far, we have generated regressor X as values between -5 and 5. To see leverage point effect, by replacing one value of X with 20, we generated a good leverage point in data set. For this case, LR and Wald test give approximately same performance. Again, Bootstrap is worst test in the analsysis. Figure 4.3 depicts the difference between tests and power envelope.

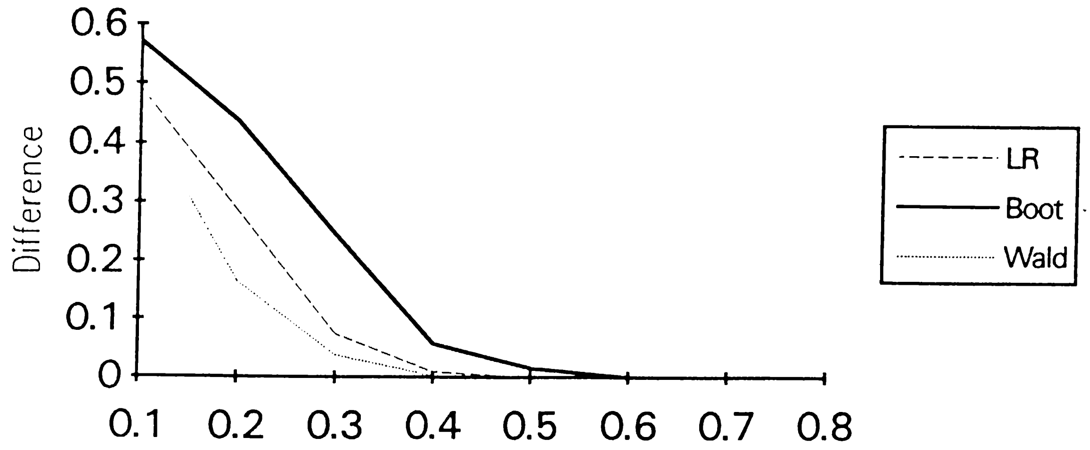

Figure 4.2 Differences between tests and power envelope

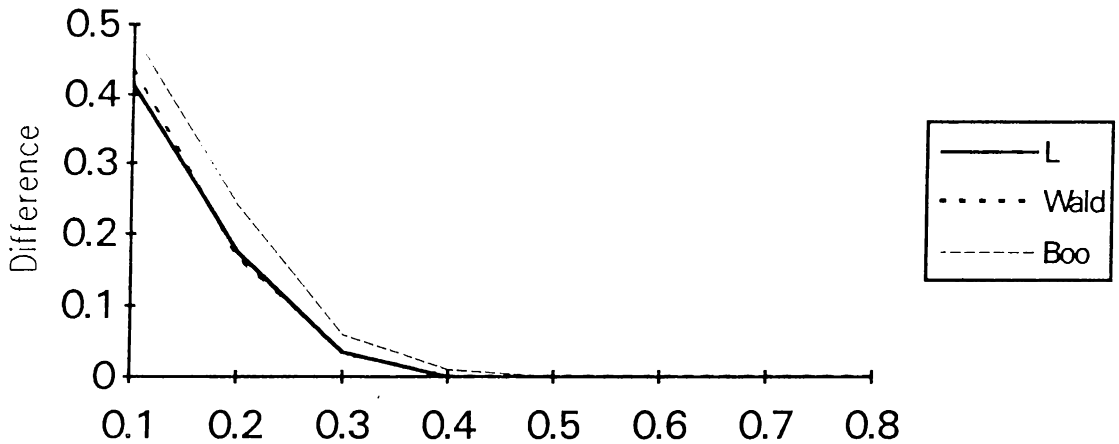

Figure 4.3 Differences between tests and power envelope when data have leverage point

Figure 4.3 shows that Wald and LR give very close power curves when data contain leverage point. The differences between these tests and power envelope are almost same. We see that Bootstrap still works bad relative to other tests.

In next section, we will analyze a new technique which increases the performance of Bootstrap.

4.2 Bootstrap-Bartlett Correction

During empirical analysis, we have found that straightforward Bootstrap test gives worser answers than other tests. Furthermore, power curve calculation of Bootstrap seems to be cumbersome both for programmer and computer. We will introduce a technique which brings easiness and accuracy in calculation. It is called **Bootstrap-Bartlett Correction**. This technique decreases necessary iteration size considerably low and finds more accurate results with respect to straightforward Bootstrap.

Before passing to issue, we know that -2 log LR asymptotically converges to χ^2_f where f is the degrees of freedom[14]. Since COMFAC puts one restriction on parameters, it can be shown that LR test is asymptotically Chi-Square distribution with degrees of freedom 1. If we sufficiently obtain LR statistics, the average of them will converge to 1 since the mean of Chi-Square distribution is 1. However, for low iteration sizes, this value may be away from 1. This is also true for using Bootstrap in LR statistic. We will use this fact in Bartlett correction. We define this correction by using following algorithm.

Step1: Generate 100 H=-2 log LR by using Bootstrap. (i.e. H^{*1}, H^{*2},....,H^{*100})

Step2: Average these and get the expected value of that sample.(Say $M=(H^{*1}+...+H^{*100})/100$).

Step3: Define new H, say H', as H'= H / M.

Step4: Note that H' will automatically have expected value 1 which matches the large sample value. Finally, reject null hypothesis if H'>c where c is critical value of χ^2_1 .

Now, we will analyze its empirical application. Before, passing Bootstrap Bartlett issue, let's investigate the asymptotic behavior of LR. If we check for the mean of LR for

---

[14] This can be easily found by Taylor series expansion around 0

different sizes, we see that as size increases, mean of LR gradually converges to 1 (See Table 4.1)

SAMPLE SIZE	MEAN
50	1.5616565
100	1.4489513
500	1.3886276
1000	1.2477456

Table 4.1 Mean of LR depending on sample size

Now, if we apply Barlett correction method on Bootstrap, we provide our sample to converge to Chi-Square with degrees of freedom 1 faster than straightforward Bootstrap. By using Bartett correction, we both decrease the iteration size from 1000 to 100 which brings speed in calculation and obtain more accurate result than straightforward Bootstrap. It is seen that mean of -2 log LR is still larger than 1 when sample size is 1000. (See table 4.1). Therefore, -2 log LR hasn't converge to Chi-Square distribution yet. However, by Bootstrap-Bartlett correction, we have obtained the results which have mean 1 which is same as Chi-Square distribution with degrees of freedom 1. Therefore, we have approached to Chi-Square distribution faster than before. Because of this, critical value of Bootstrap which was 5.18 in straightforward Bootstrap decreased to 3.7677999 which is quite close to critical value of χ^2_1 which is 3.84. From the theoretical analysis, we should obtain more close curves to the power envelope. When we rerun the program according to Bartlett correction, we will get the power curves of Bootstrap-Bartlett and straightforward Bootstrap as;

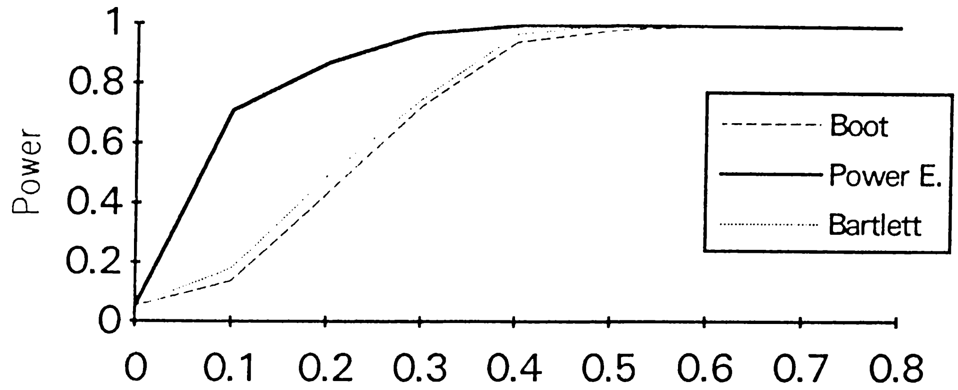

Figure 4.4 Power curve of Bootstrap-Bartlett and StrFwd. Bootstrap

In Figure 4.4, we see that power curve of Bootstrap has been improved a little bit. That means, Bartlett correction increases the power curve of Bootstrap. The difference between power envelope and Bootstrap decreases. (See Figure 4.5). However, this decrease is not so large to dominate Wald test.

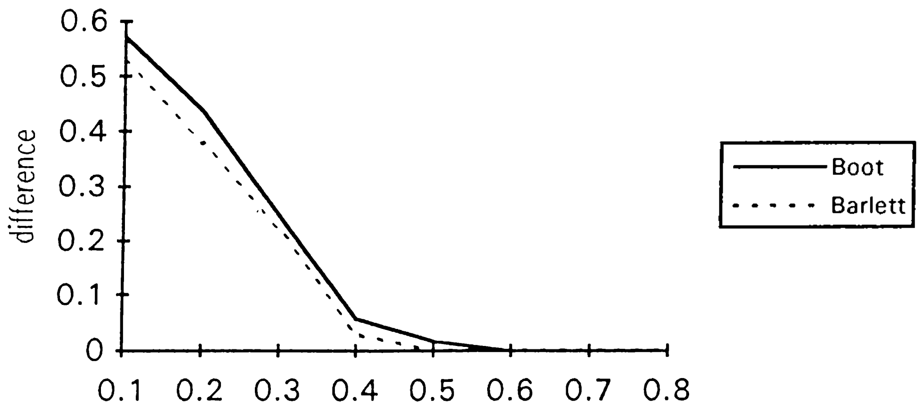

Figure 4.5 Differences between Bootstrap (StrFwd,Bartlett) powers and power envelope

5
Summary

Bootstrap has been investigated many times and all properties of it has beed derived throughout the analyses before. The aim of this thesis was also to inform the reader about features of Bootstrapping.

Here, the performance of Bootstrap theory was totally depended upon maximum likelihood estimates. Therefore, second part of introduction was devoted to inform reader what maximum likelihood and Bootstrapping mean.

For the accuracy of methods, there are some necessary and sufficient conditions that should be satisfied before analysis. These conditions were clearly discussed.

First aim of this thesis was to investigate the accuracy of estimation methods, Bootstrap and Normal Approximation based on central limit theorem. In fact, since true density is not known, it should have been predicted. Therefore, Monte Carlo distribution of maximum likelihood, was regarded as closest distribution to true distribution.

The first analysis was the comparison of Monte Carlo and Bootstrap distributions; and, Monte Carlo and Normal Approximation based on CLT. Before empirical issue, theoretical analysis was held and derived that Bootstrap converges to Monte Carlo distribution faster than CLT Approximation asymptotically. This means Bootstrap is always at least as good as, and in some cases better than, the classical Normal Approximation based on CLT. Of course, this theoretical approach should be validated by an empirical analysis. Therefore, two cases, *symmetric and asymmetric unknown distribution* , were investigated here. The problem of broken curves due to Monte Carlo simulation was solved by density smoothing technique proposed by **Silverman (1986)**. The conclusion was quite consistent with theoretical approach, i.e. Bootstrap works at least as good as CLT Approximation when distribution is symmetric. Furthermore, for asymmetric distributions, CLT Approximation was dominated by Bootstrap.

After this analysis, the application areas of Bootstrap are widened. One is the using Bootstrap in F-test. The performance of Bootstrap was measured by using critical values of F-test. Since critical value of F-test cannot be found for low sample sizes when model is dynamic, it is aimed to use Bootstrap in these sample sizes. As a result, it is obtained that Bootstrap gives close results to actual ones (i.e. the results obtained by Monte Carlo). Furthermore, it is seen that as sample size increases, Monte Carlo simulation result slightly converges to table value which is asymptotic value, (Size is ∞) as it should be. This analysis showed that Bootstrap can be used in F-test for considerable sample sizes when asymptotic values are misleading. The accuracy of F-test when regressor X is changed was discussed during analysis. Furthermore, the performance of Bootstrap was investigated when order of lag in model increases and data contain leverage points.

Using Bootstrap technique for the cases when model is nonlinear, was the last analysis of this thesis. COMFAC model was taken as an example for this part. The coefficient in front of one regressor was produced by multiplication of other coefficients. Because of nonlinerity of the model, new comparison mechanism was introduced, i.e. power curves of each tests. During the analysis, LR, simple Wald test and Bootstrap were used. Theoretically, LR should have worked better. However, since our model was an autoregressive model, Wald test gave better results than LR. Bootstrap was the worst test in this analysis. This leaded us to introduce a correction mechanism, called **Bootstrap-Bartlett Correction** which increases the performance of straightforward Bootstrap. Furthermore, this tool provided us efficiency in power curve computation time of Bootstrap. However, the results were not better than the results of Wald test.

Besides explanatory figures and tables during analysis, a tool which is called maximum gap was developed to measure how accurate the methods were. If maximum gap is low, that means method gives close results to actual results, then method is considered as good.

Consequently, the analyses presented in this thesis can be extended for other different applications of Bootstrap. We have derived that Bootstrap can be considered as suitable technique in some areas. The performance evaluation of Bootstrap shows the place where Bootstrap can be used in analysis.

References

Efron, Bradley, Bootstrap Methods: Another Look at the Jacknife, *Annals of Statistics,* 1979,7, 1-26

Efron, Bradley, The Jacknife, the Bootstrap and Other Sampling Plans, *SIAM, Philadelphia,* 1982

Franke, J. and Hardle, W. On Bootstrapping Kernal Spectral Estimates, *Annals of Statistics,* 1992,**20**, 121-145

Freedman, D.A. Bootstrapping Regression Models,*Annals of Statistics,* 1981, **9**, 1218-1228

Hall,Peter Using Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems, *Journal of Multivariate Analysis,* 1990, **7**, 177-203

Hall, Peter The Bootstrap and Edgeworth Expansion, *Springer-Verlag, Newyork, Inc.* 1992

Hall, Peter and Wilson, Susan Two Guidelines for Bootstrap Hypothesis Testing, *Biometrics,* 1991, **47**, 757-762

Kemp, Gordon C.R. On Wald tests for Globally and Locally Quadratic Restrictions, *Journal Of Econometrics,* 1991, **50**, 257-272

Maddala, G.S. Introduction to Econometrics, *Macmillian Publishing Company, NewYork,* 1988

Mammen, Enno When Does Bootstrap Work ? *Springer-Verlag, Newyork,Inc.* 1992

Mizon, Grayham and Hendry, David An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification, *Review of Economic Studies,* 1980, 21-45

Navidi, William Edgeworth Expansions for Bootstrapping Regression Models, *Annals of Statistics*, 1989,**17**, 1472-1478

Silverman, B.W. Density Estimation for Statistics and Data Analysis, *Chapman and Hall*, 1986

Singh Kesar On the Asymptotic Accuracy of Efron's Bootstrap,*Annals of Statistics*, 1981, **9**, 1187-1195

Zaman, Asad Statistical Foundations of Econometric Techniques, *Academic Press*, 1996, forthcoming.

```
/*    MUSTAFA CENK TIRE  */
   /*   MASTER THESIS */

/* MAIN BLOCK FOR BOOTSTRAP AND NORMAL APP. COMPARISON */


/* graph*/
library pgraph;
graphset;
trubeta=1;
x=zeros(100,1);grid=zeros(100,1);mmax=zeros(3,1);smax=zeros(3,1);
y=zeros(100,1);count=zeros(60,1);newt=zeros(60,1);
monteval=zeros(1000,1);
itersize=1000;
/* generate x */
i=1;
do while i<=90;
   chk1=rndu(1,1);
   if chk1<0.5;x[i]=-1;else;x[i]=1;endif;
   i=i+1;
endo;
do while i<=100;
   chk2=rndu(1,1);
   if chk2<0.5;x[i]=-3.3;else;x[i]=3.3;endif;
   i=i+1;
endo;

iter=1;
"Calculating Monte Carlo";
do while iter<=itersize;
   /* generate errors */
   err=generate(100);
   /* built y  where beta is 1*/
   y=x+err;
   betahat=(x'*y)/(x'*x);
   sigmahat=((y-betahat.*x)'*(y-betahat.*x))/100;
   monteval[iter]=(sqrt(x'*x)*(betahat-trubeta))/sqrt(sigmahat);
   level=31+floor(10*monteval[iter]);
   if iter/100==floor(iter/100);iter/100;endif;
   if level>0 and level<61;count[level]=count[level]+1;endif;
   iter=iter+1;
endo;
count=count./100;
err=generate(100);
"Calculating Bootstrap";
{bootcnt,value}=boot(1,x,err,1000);
bootcnt=bootcnt./100;
newt=seqa(-3,0.1,60);
ndens=(1/sqrt(2*pi))*exp(-0.5*newt.^2);

sndens=ndens;
begwind;
makewind(12,6.855,0,0,0);    @ Main graph window @
setwind(1);
aa=newt;
scount=smooth(monteval,1);
sbootcnt=smooth(value,2);

title("\202Unsmoothed Bootstrap, Monte Carlo and Standard Normal\201");
ylabel("Distribution");


xy(aa,count);
```

```
xy(newt,bootcnt);
xy(newt,ndens);
endwind;
begwind;
makewind(12,6.855,0,0,0);     @ Main graph window @
setwind(1);

title("\202Smoothed Bootstrap, Monte Carlo and Standard Normal\201");
ylabel("Distribution");

xy(aa,scount);
xy(newt,sbootcnt);
xy(newt,sndens);
endwind;
```

```
/*  PROCEDURE FOR BOOTSTRAP  */


proc (2) = boot(trubeta,x,err,rep);
local t,bstar,cnt,count,estar,y,bhat,sigmahat,er,level,ystar,dum,value;
count=zeros(60,1);
value=zeros(1000,1);
y=x*trubeta+err;
t=rows(x);
bhat=(x'*y)/(x'*x);
dum=x'*bhat;
er=y-dum';
sigmahat=((y-bhat.*x)'*(y-bhat.*x))/100;
cnt=1;
do until cnt>=rep;
  estar=er[1+floor(t*rndu(t,1))];
  ystar=(x'*bhat+estar')';
  if cnt/100==floor(cnt/100);cnt/100;endif;
  bstar=(x'*ystar)/(x'*x);
  value[cnt]=(sqrt(x'*x)*(bstar-bhat))/sqrt(sigmahat);
  level=31+floor(10*(sqrt(x'*x)*(bstar-bhat))/sqrt(sigmahat));
  if level>0 and level<61;count[level]=count[level]+1;endif;
  cnt=cnt+1;
endo;
retp(count,value);
endp;




/* this proc generates errors */
proc generate(n);
local err,i,l,e;
i=0;err=zeros(n,1);
do until i>=n;
i=i+1;
  l=rndu(1,1);
  if l>.5;
  e=-abs(rndn(1,1));
  else;
  e=abs(3*rndn(1,1));
  endif;
err[i]=e;
endo;
retp(err);
endp;
```

```
/*  This Procedure smooths the density curves of Monte Carlo
    and Bootstrap

*/


proc smooth(value,indx);
local fc,i1,i2,sum,t;
cls;
fc=zeros(60,1);
y=seqa(-3,0.1,60);
h=1.06*100^(-1/5);
i1=1;
if indx==1;
"MonteCarlo is Smoothing.....";
else;
"Bootstrap is smoothing......";
endif;

do while i1<=60;
  i2=1;
  sum=0;
  if (60-i1)/10==floor((60-i1)/10);(60-i1)/10;endif;
  do while i2<=1000;
    t=(y[i1]-value[i2])/h;
    sum=sum+(1/sqrt(2*pi))*exp(-(1/2)*t^2);
    i2=i2+1;
  endo;
  fc[i1]=(1/(1000*h))*sum;
  i1=i1+1;
endo;
retp(fc);
endp;
```

```
/* MUSTAFA CENK TIRE */
/* F-test vs. Bootstrap */

/* These are the table (asymptotic) values for 95% f test */
/* for sample size 10,25,50 and 100 respectively */


/* generate 1st order ADL */

beta=1;        /* Coefficient of X */
sigma=1;       /* Error Std. Deviation */
T=25;          /* Time Series Length                            */
x=rndn(t,1); /* GENERATE AND FIX THE REGRESSORS X ONCE AND FOR ALL */


MCSS=1000;                /* Monte Carlo Sample Size */
y=beta*x+sigma*rndn(t,1);  /* This is the original Data */
Fstat=zeros(1000,1);
tops=fstat;
fn rss(y,x)=sumc((y-x*(y/x))^2);


j=1; do while j<MCSS+0.1;
      ys=beta*x+sigma*rndn(t,1);  /* This is the Monte Carlo repeat */
      lev=floor(t/2);temp=x[lev];
      x[lev]=10; /* LEVERAGE POINT */
      rs0=rss(ys[2:t],x[2:t]);
      x1=x[2:t]~ys[1:t-1]~x[1:t-1];
      rs1=rss(ys[2:t],x1);
      f=((rs0-rs1)/2)/(rs1/(t-4));
      fstat[j]=f;
      x[lev]=temp;
j=j+1; endo;

fstat=sortc(fstat,1);
c95=fstat[950];
c95;
/* Note that sample size is t-1 and there are three estimated
   parameters, so denominator has (t-1)-3 = t-4 degrees of freedom */

/* Get Bootstrap Critical Value--- note, this is for a particular Y and
   X --- It will NOT apply to different y and x  */
x[lev]=10;
bh=y/x;                       /* Bhat is OLS estimate on ORIGINAL DATA   */
sigh=sqrt(rss(y,x)/(t-1));  /* SigHat is MVU est. of sigma (ORIG.DATA) */
BOOT=1000;                    /* BootStrap Sample Size                   */
Fboot=zeros(1000,1);
j=1; do while j<BOOT+0.1;
     yst=bh*x+sigh*rndn(t,1);  /* Generates Bootstrap Sample of y UNDER NULL *
     rs0=rss(yst[2:t],x[2:t]);
     x1=x[2:t]~yst[1:t-1]~x[1:t-1];
     rs1=rss(yst[2:t],x1);
     f=((rs0-rs1)/2)/(rs1/(t-4));
     fboot[j]=f;
j=j+1; endo;
fboot=sortc(fboot,1);
b95=fboot[950];
"Boot est of Crit Val is= ";;b95;
"P-val for this is= ";;cdffc(b95,2,t-4);

wait;
```

```
/*

At this point we have 95% critical value of F based on
MCSS 1000. Now to get 60,000 MCSS.
*/

Range=fstat[975]-fstat[925];
grid=seqa(fstat[925],range/100,101);
count=0*grid;
i=1; do while i<60000;
     ys=beta*x+sigma*rndn(t,1);   /* This is the Monte Carlo repeat */
     rs0=rss(ys[2:t],x[2:t]);
     x1=x[2:t]~ys[1:t-1]~x[1:t-1];
     rs1=rss(ys[2:t],x1);
     fstat=((rs0-rs1)/2)/(rs1/(t-4));
     count=count+(grid .< fstat);
     i;
i=i+1; endo;

count=count/60000;
ii=1;
do while ii<=100;
 if count[ii]<0.05;limit=ii;ii=100;endif;
 ii=ii+1;
endo;
MCCV=grid[limit];
MCCV;   /* Monte Carlo 95% critical value of F */
```

```
/* MUSTAFA CENK TIRE */
/* LR, BOOTSTRAPPED LR, WALD */

/* This program calculates the power curves of each test */

IterSize=1000;

Power=0;fpass=1;iter=1;data=ZEROS(8,4);

/* generate x */
x=floor(rndu(100,1)*10-5);
/* Creates Output File */
output file=Lr_Wald.out reset;
output off;

proc level(lr,cv);
  local i,lev,flag;
  i=1;flag=0;
  do while (lr[i]<=cv) and (i<iterSize);i=i+1;flag=1;endo;
  if flag==1;lev=(IterSize-i)/IterSize;else;lev=1;endif;
  retp(lev);
endp;




do while power<=0.8;
totgama=0;totbeta=0;

/* Sample size */
t=25;

/* True paramaters */
beta=1;gama=0.8;



MCiter=1;
Wald=zeros(IterSize,1);
LR=zeros(IterSize,1);
do while MCiter<=IterSize;

   /* Model Yt=B.Xt+G.(Yt_1-B.Xt_1) + e */
   y=zeros(t,1);
   y0=0;
   err=rndn(t,1);
   xp=zeros(t,1);
   xp[2:t]=beta*x[2:t]+(power-(gama*beta))*x[1:t-1]+err[2:t];
   xp[1]=0;
   y=recserar(xp,y0,gama);


   /* Non-linear case (Ho) */
   betahat=0.5;gamahat=0;maxx=-100000;
   do while betahat<=1.5;
     gamahat=0;
     fterm=y[2:t]-betahat*x[2:t];
     sterm=y[1:t-1]-betahat*x[1:t-1];

     /* ML estimate of gamma^ by fixing beta^ */
     gamahat=fterm/sterm;
     pp=-betahat*gamahat;
     lpart=y[2:t]-betahat*x[2:t]-gamahat*y[1:t-1]-pp*x[1:t-1];
     l=-(1/2)*(lpart'*lpart);
```

```
        /* find parameters that maximizes likelihood */
        if l>maxx;maxx=l;maxbeta=betahat;maxgama=gamahat;endif;
        betahat=betahat+0.1;
      endo;

      /* Linear case (H1) */
      xpart=x[2:t]~y[1:t-1]~x[1:t-1];
      OLS=y[2:t]/xpart;
      EC=y[2:t]-OLS[1]*x[2:t]-OLS[2]*y[1:t-1]-OLS[3]*x[1:t-1];
      l1=-(1/2)*(EC'*EC);
      /* 2Log LR */
      lamda=-2*maxx+2*l1;

      LR[MCiter]=lamda;

      /* Simple Wald Estimation */
      /* W(O1,O2,O3)= O1+O2.O3 = 0 */
      Wald[MCiter]=OLS[3]+OLS[1]*OLS[2];


      MCiter=MCiter+1;
      totgama=totgama+maxgama;
      totbeta=totbeta+maxbeta;
   endo;
   Wald=sortc(Wald,1);
   LR=sortc(LR,1);

   /* Check for W/STDC(W)= ? */
   Wald=Wald./stdc(Wald);

   /* Get critical value in first loop then construct power */
   /* in later loops */

   if fpass==1;
      LrMC=LR[Itersize*0.95];
      WaldMC=Wald[Itersize*0.95];
   else;
      WDvalue=level(Wald,WaldMc);
      LRvalue=level(lr,LrMc);
   endif;


   /* Bootstrap */

   /* Use average of estimators */
   maxb=totbeta/IterSize;maxg=totgama/IterSize;
   if power==0;mmaxg=maxg;mmaxb=maxb;
   shat=sumc((y[2:t]-mmaxb*x[2:t]-mmaxg*y[1:t-1]+mmaxb*mmaxg*x[1:t-1])^2)/(t-1);
   endif;
   MCiter=1;

   LR=zeros(IterSize,1);
   do while MCiter<=IterSize;

     /* MODEL */
     y=zeros(t,1);
     y0=0;
     err=rndn(t,1);
     xp[2:t]=mmaxb*x[2:t]+(power-(mmaxg*mmaxb))*x[1:t-1]+shat*err[2:t];
     xp[1]=0;
     y=recserar(xp,y0,mmaxg);

     /*  Ho  */
     betahat=0.5;gamahat=0;maxx=-100000;
     do while betahat<=1.5;
        gamahat=0;
```

```
      fterm=y[2:t]-betahat*x[2:t];
      sterm=y[1:t-1]-betahat*z[1:t-1];

      /* ML estimate of gamma' by fixing beta' */
      gamahat=(fterm'*sterm)/(sterm'*sterm);

      pp=-betahat*gamahat;
      lpart=y[2:t]-betahat*x[2:t]-gamahat*y[1:t-1]-pp*x[1:t-1];
      l=-(1/(2*shat))*(lpart'*lpart);
      if l>maxx;maxx=l;maxbeta=betahat;maxgama=gamahat;endif;
      betahat=betahat+0.1;
   endo;

   /* H1 */
   xpart=x[2:t]~y[1:t-1]~x[1:t-1];
   OLS=y[2:t]/xpart;
   EC=y[2:t]-OLS[1]*x[2:t]-OLS[2]*y[1:t-1]-OLS[3]*x[1:t-1];
   l1=-(1/(2*shat))*(EC'*EC);
   lamda=-2*maxx+2*l1;
   LR[MCiter]=lamda;
   MCiter=MCiter+1;
endo;
LR=sortc(LR,1);
if fpass==1;
   Boot=LR[IterSize*0.95];
   cls;
   "CRITICAL VALUES ARE :";
   "Likelihood Ratio:";;LrMC;
   "BootStrap        :";;Boot;
   "Wald Statistic   :";;WaldMC;
   "Power curve is being produced....";
   "          GAMMA          LR POWER          BOOTS. POWER          WALD POWER";
else;
   output file=Lr_Wald.out on;
   BTvalue=level(lr,Boot);
   Data[iter,.]=POWER~LRvalue~BTvalue~WDvalue;
   Data[iter,.];
   iter=iter+1;
   output off;
endif;


fpass=0;
power=power+0.1;
endo;

Power0={0 0.05 0.05 0.05};
data=Power0|data;
locate 24,25;
"Press Any Key to Continue";
WaitC;
Library Pgraph;
Graphset;
begwind;
makewind(12,6.855,0,0,1);    @ Main graph window @
setwind(1);

title("\202Power Curves of LR, Bootstrap and Wald \201");
xlabel("Gamma");
ylabel("Power");
xtics(0,0.8,0.1,0.1);
ytics(0,1.01,0.05,0.05);
xy(data[.,1],data[.,2]);
xy(data[.,1],data[.,3]);
xy(data[.,1],data[.,4]);
endwind;
```