DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

TESTING UNIT ROOT USING BOOTSTRAP METHOD

by
Emel TUĞ

September, 2013
İZMİR

TESTING UNIT ROOT
USING
BOOTSTRAP METHOD

A Thesis Submitted to the

Graduate School of Natural and Applied Sciences of Dokuz Eylül University

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics Programme

by

Emel TUĞ

September, 2013

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"TESTING UNIT ROOT USING BOOTSTRAP METHOD"** completed by **EMEL TUĞ** under supervision of **Assoc. Prof. Dr. AYLİN ALIN** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Aylin ALIN

Supervisor

Doç.Dr. Esin FIRUZAN

(Jury Member)

Prof.Dr. M.Vedat PAZARLIOĞLU

(Jury Member)

Prof.Dr. Ayşe OKUR
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

There are three people without whom this thesis might not have been written, and I want to express my feelings about them.

I would like to express the deepest appreciation to my supervisor, Assoc. Prof. Aylin ALIN, who gave me a chance to work together in the field of Time Series Analysis which is determined by my insistence. We both searched books and articles, tried to understand the theoretical proofs intensively, prepared the simulation programme with a real enthusiasm. She wanted to obtain satisfactory results from the programme as much as me. Without her guidance, patience, and persistent help, this thesis would not have been possible.

I would also like to thank my family, my little son and my husband. They let me to continue my education at the university. They accepted to spare time without me at the weekends while I was studying. They tolerated my tension especially during the time of exams. Without their support and encouragement, I would not manage to be successful. I am very blessed to have them in my life.

Emel Tuğ

TESTING UNIT ROOT USING BOOTSTRAP METHOD

ABSTRACT

The aim of this study is to give general information about the bootstrap and the time series and, to evaluate the performance of the bootstrap unit root test which has drawn much attention especially in economics and other related fields.

In this study, first of all, the concept of bootstrap is given; implementation of the bootstrap for both independent and dependent data are told; the fields where the bootstrap method are used to obtain asymptotic refinements are given; the Edgeworth Expansions and the Cornish-Fisher Expansions on which the proof that the bootstrap provides asymptotic refinements is based are told; the Sufficient Bootstrap which provides an important reduction in the sample size is told briefly. Later, the concept of time series and, the fundamental concepts which are necessary to understand the logic of time series are given; representations of the time series processes are showed; the stationarity and the nonstationarity situations are told. After giving brief information about the other concepts of time series such as model selection criteria, the unit root processes which are the basic concept for this thesis are told in details. The most popular unit root tests, Dickey-Fuller tests and Phillips-Perron tests are examined and the intuition behind these tests is given.

Three different methods are compared for their powers on the unit root tests: *Asymptotic*, *bootstrap*, and *sufficient bootstrap* methods. Independent and dependent residuals have been studied separately. Finally, the concluding remarks obtained as a result of the simulation study are listed.

Keywords: Bootstrap, asymptotic refinement, Edgeworth expansion, Cornish-Fisher expansion, sufficient bootstrap, time series, stationarity, nonstationarity, unit root process, residual, Dickey-Fuller test, Phillips-Perron test.

BOOTSTRAP YÖNTEMİ İLE BİRİM KÖK TESTİ

ÖZ

Bu çalışmanın amacı, bootstrap ve zaman serisi hakkında genel bilgi vermek ve özellikle ekonomide ve diğer ilgili alanlarda fazla dikkat çeken bootstrap birim kök testinin performansını değerlendirmektir.

Bu çalışmada, ilk olarak, bootstrap kavramı verilir; bootstrap' in hem bağımsız hem de bağımlı verilere uygulanışı anlatılır; bootstrap yönteminin asimptotik netlikler elde etmek amacıyla kullanıldığı alanlar verilir; bootstrap' in asimptotik netlikler vermesine ilişkin kanıtın dayandırıldığı Edgeworth açılımları ve Cornish-Fisher açılımları anlatılır; örneklem ölçümünde önemli bir azalma sağlayan Sufficient Bootstrap kısaca anlatılır. Daha sonra, zaman serisi kavramı ve zaman serisi mantığını anlamak için gerekli olan temel kavramlar verilir; zaman serisi süreçlerinin sunumları gösterilir; durağan olma ve durağan olmama durumları anlatılır. Model seçim kriteri gibi zaman serisinin diğer kavramları hakkında kısa bir bilgi verdikten sonra, bu tez için temel kavramlar olan birim kök süreçleri detaylarıyla anlatılır. En popüler birim kök testleri, Dickey-Fuller testleri ve Phillips-Perron testleri incelenir ve bu testlerin arkasındaki mantık verilir.

Üç farklı yöntem birim kök testleri üzerindeki güçleri açısından karşılaştırılırlar: *Asymptotic*, *bootstrap*, ve *sufficient bootstrap* yöntemleri. Bağımsız ve bağımlı artıklar ayrı incelenirler. Sonunda, simulasyon çalışması sonucunda elde edilen çıkarsamalı ifadeler listelenir.

Anahtar kelimeler: Bootstrap, asimptotik netlik, Edgeworth açılımı, Cornish-Fisher açılımı, sufficient bootstrap, zaman serisi, durağan olma, durağan olmama, birim kök süreci, artık, Dickey-Fuller testi, Phillips-Perron testi.

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Horowitz (2001) defines that "the bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data" (p. 3161). In bootstrap technique, you behave the original sample as if it is the population itself. Generally, bootstrap provides more accurate approximations compared to those of first-order asymptotic theory. However, this accuracy depends on whether the data are a random sample from a distribution or a time series.

> If the data are *i.i.d.*, independently and identically distributed, the bootstrap can be implemented by sampling the data randomly with replacement or by sampling a parametric model of the distribution of the data....The situation is more complicated when the data are a time series because bootstrap sampling must be carried out in a way that suitably captures the dependence structure of the data generation process (DGP) (Härdle, Horowitz, & Kreiss, 2001, p. 1).

Their work also shows that the errors made by the bootstrap converge to zero more slowly when the data are a time series than they are a random sample. For implementing the bootstrap technique for a time series, a several methods have been developed, such as the block bootstrap, the sieve bootstrap etc. All these techniques have been developed to obtain more accurate approximations.

Wei (2006) defines a time series as an ordered sequence of observations. The observations in a time series are dependent or correlated, and therefore the order of the observations is important. Hence, statistical procedures and techniques that rely on independence assumption are no longer applicable, and different methods are needed. The body of statistical methodology available for analyzing time series is referred to as time series analysis. To understand the time series analysis technique, it is compulsory to understand the concept of stochastic process well since the developed theory for the time series analysis is based on the stochastic processes.

The unit root hypothesis has drawn much attention for the past three decades, especially in economics and other related field. Chang & Park (2000) point out that "the hypothesis has an important implication on, in particular, whether or not the shocks to an economic system have a permanent effect on the future path of the economy" (p. 379). It is known that many of important economic and financial time series display unit root characteristics. Phillips & Perron (1988) states that "formal statistical tests of the unit root hypothesis are of additional interest to economists because they can help to evaluate the nature of the nonstationarity that most macroeconomic data exhibit" (p. 335). The tests developed by Dickey & Fuller (1979, 1981) are the most commonly used. However, Chang & Park (2000) state that "the tests by Said-Dickey and Phillips-Perron are often preferred to the Dickey-Fuller tests in practical applications, since they do not require any particular parametric specification and yet are applicable for a wide class of unit root models" (p. 380). The disadvantage of all these tests is to have considerable size distortions in finite samples where the bootstrap method may perform better. In this thesis, performance of the bootstrap method has been investigated under finite samples.

In Chapter 2, the bootstrap technique is defined. Implementation of the bootstrap technique to both independent and dependent data is told. The bootstrap iteration and the intuition behind this technique are given. The bootstrap principle and the Edgeworth Expansion on which the proof that the bootstrap provides asymptotic refinements is based are told giving the theoretical information. The concept of the sufficient bootstrapping is given. In Chapter 3, basic definitions connected with the time series analysis are given. How the time series processes are represented as an autoregressive (AR), a moving average (MA), and a mixed autoregressive and moving average (ARMA) models are showed. Both stationary and nonstationary time series models are told giving their basic properties. The unit root processes, the most popular unit root tests and the intuition behind these tests are given. The bootstrap unit root tests and accuracy of these tests are told. In Chapter 4, simulation results are presented. Asymptotic, bootstrap, and sufficient bootstrap methods are compared as regards their powers on the unit root tests. Independent and dependent

residuals have been studied separately. Finally, in Chapter 5, the concluding remarks obtained as a result of the simulation study are listed.

The explanations and theoretical proofs in Chapter 2 are based on Horowitz (2001) and Boik (2006). The explanations and theoretical proofs in Chapter 3 are based on Wei (2006) and Enders (1948). The notation used for time series models is the same with Wei (2006). MATLAB R2011b is used for the simulation study. The codes of the programme are given in Appendices.

CHAPTER TWO
BOOTSTRAP

2.1 Definition of the Bootstrap

Horowitz (2001) defines that "the bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data" (p. 3161). In bootstrap technique, you behave the original sample as if it is the population itself. Hall (1992) explains the idea behind the bootstrap by comparing the population mean and the sample mean. He states that estimation of a functional of a population distribution *F*, such a population mean

$$\mu = \int x \, dF(x) \tag{2.1.1}$$

is done by employing the same functional of the empirical (or sample) distribution function, which is the sample mean

$$\bar{X} = \int x \, d\hat{F}(x). \tag{2.1.2}$$

The empirical distribution is a function which assigns the same probability to each of the sample individuals. The term of functional may be described as follows: If $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$, then $\hat{\mu}$ is a *function* of *x*. If $= \int x \, dF(x)$, then μ is a *functional* which takes away the distribution function to a real value.

Hall (1992) states how the bootstrap statistics are calculated more easily as follows:

> Efron (1979) also showed that in many complex situations, where bootstrap statistics are awkward to compute, they may be approximated by Monte Carlo "resampling". That is, same-size resamples may be drawn repeatedly from the original sample, the value of a statistic computed for each individual resample,

and the bootstrap statistic approximated by taking an average of an appropriate function of these numbers. The approximation improves as the number of resamples increases (Hall, 1992, p. 1).

Hall (1992) explains the bootstrap principle by a Russian "matryoshka" doll and the number of freckles on its face. In this thesis it is given the short summary of the theory in his book.

Let n_0 be the number of freckles on the biggest doll' s face, and n_1 be the number of freckles on the second biggest doll' s face, and n_2 be the number of freckles on the third biggest doll' s face, and etc. Let us try to estimate n_0. Since the second biggest doll is smaller than the biggest doll, only considering n_1 is likely to be an underestimate of n_0. However, because the ratio of n_1 to n_2 should be close to the ratio of n_0 to n_1, that is, $n_1/n_2 \approx n_0/n_1$,

$$\hat{n}_0 = n_1^2/n_2 \tag{2.1.3}$$

might be a reasonable estimate of n_0. By the same reason, while F_0 is the population distribution function and F_1 is the sample distribution function, the population equation may be defined as follows:

$$E\{f_t(F_0, F_1)|F_0\} = 0. \tag{2.1.4}$$

This is defined as **the population equation** since if it is solved exactly, the properties of the population should be known. However, since these properties are unknown, an approximate solution for this equation may be found by the sample equation

$$E\{f_t(F_1, F_2)|F_1\} = 0. \tag{2.1.5}$$

This is defined as **the sample equation** since if the sample distribution F_1 is known, this equation can be solved exactly.

This sense may be defined as **the bootstrap principle**.

Hall (1992) shows that the bootstrap principle may be used for the bias reduction, since as B which is the number of bootstrap samples goes to infinity, the parameter estimator obtained from the bootstrap samples behaves like the parameter estimator obtained from the original sample. As B goes to infinity, the mean of estimators obtained from the bootstrap samples equals to the estimator obtained from the original sample. He states that while the actual variance may have increased a little as a result of bootstrap bias reduction, the first-order asymptotic formula for variance has not changed. See Hall (1992) for the details.

The bootstrap has been the object of much research in statistics since its development by Efron (1979). Horowitz (2001) states that "under mild regularity conditions, the bootstrap yields an approximation to the distribution of an estimator or test statistic that is at least as accurate as the approximation obtained from first-order asymptotic theory" (p. 3161). Such improvements are called asymptotic refinements. This is resulted from the ability of the bootstrap on bias reduction and mean-square-error. Thus, the bootstrap provides a way to substitute computation for mathematical analysis if calculating the asymptotic distribution of an estimator or statistic is difficult.

However, there are some restrictions for using the bootstrap. The bootstrap technique may be used to estimate the probability distribution of an asymptotically pivotal statistic or the critical value of a test based on an asymptotically pivotal statistic whenever such a statistic is available. On the other hand, the bootstrap technique should not be used to estimate the probability distribution of a non-asymptotically-pivotal statistic such as a regression slope coefficient if an asymptotically pivotal statistic is available. See Horowitz (2001) for the details.

2.2 Pivotal Statistics and Consistency of the Bootstrap

The statistic whose distribution is dependent of the population parameters is called *pivotal*. Pivotal statistics are not available in most econometric applications. Many econometric statistics are *asymptotically pivotal* in the meaning of the statistic whose asymptotic distribution does not depend on unknown population parameters, or asymptotically normally distributed. Horowitz (2001) states that "if an estimator is asymptotically normally distributed, then its asymptotic distribution depends on at most two unknown parameters, the mean and the variance, that can often be estimated without great difficulty" (p. 3164).

Horowitz (2001) defines the consistency of the bootstrap with details. Roughly speaking, as $n \to \infty$ if the bootstrap estimator which is given as an approximation to the exact finite-sample CDF of T_n is converges in probability to the asymptotic CDF of T_n, then the bootstrap is said to be consistent.

On the other hand, in the cases of the heavy-tailed distributions, the distribution of the square of the sample average, the distribution of the maximum of a sample, the bootstrap is inconsistent. Also, the bootstrap does not consistently estimate the distribution of a parameter estimator when the true parameter point is on the boundary of the parameter space. The details are given in Horowitz (2001).

2.3 Asymptotic Refinements

In applied econometrics, the bootstrap provides a higher-order asymptotic approximation to the distribution of a statistic for many situations. To explain the refinements resulted from the bootstrap method, it is assumed that the data are a simple random sample from some distribution.

Many important econometric estimators, including maximum-likelihood and generalized-method-of-moments estimators, are either functions of sample moments or can be approximated by functions of sample moments with an

approximation error that approaches zero rapidly as the sample size increases (Horowitz, 2001, p. 3172).

Let's assume that, the inferential problem is to obtain a point estimate of a univariate parameter θ that can be expressed as a smooth function of a vector of population moments. Also assume that θ can be estimated consistently by substituting population moments with sample moments in the smooth function.

2.3.1 Bias Reduction

In the case of inference with a sample, the bias is caused by not knowing all values in the population and so not knowing the true population distribution. Because, in the bootstrap method we treat the original sample as if they were the population, we can calculate the difference between the estimator obtained from the original sample and the estimator obtained from the bootstrap sample. Hence, we can add the bias which is resulted from the bootstrap sampling to the estimator obtained from the original sample. As a result, the bias reduction is verified. Now, the new estimator is called as the bias-corrected estimator. Whereas the bias obtained by the first-order asymptotic approximations is $O(n^{-1})$, the bias obtained by the bootstrap approximation is $O(n^{-2})$.

To be specific, let X be a random vector, and set $\mu = E(X)$. Assume that the true value of θ is $\theta_0 = g(\mu)$, where g is a known, continuous function. Suppose that the data consist of a random sample $\{X_i: i = 1, \dots, n\}$ of X. Then θ is estimated consistently by

$$\theta_n = g(\bar{X}). \tag{2.3.1}$$

Monte Carlo procedure for computing the bootstrap bias estimator, B_n^*, is given in Horowitz (2001) as follows:

B1: Use the estimation data to compute θ_n.

B2: Generate a bootstrap sample of size n by sampling the data randomly with replacement. Compute $\theta_n^* = g(\bar{X}^*)$.

B3: Compute $E^*(\theta_n^*)$ by averaging the results of many repetitions of step B2. Set $B_n^* = E^*(\theta_n^*) - \theta_n$.

(Horowitz, 2001, p. 3174).

The criteria in choosing the number of repetitions, m, of step B2 is that to choose m sufficiently large that the estimate of $E^*(\theta_n^*)$ does not change significantly if m is increased further. See Horowitz (2001) for the details. Andrews and Buchinsky (2000) discuss more formal methods for choosing the number of bootstrap replications.

2.3.2 The Distributions of Statistics

The proof that the bootstrap provides asymptotic refinements is based on an Edgeworth expansion of a sufficiently high-order Taylor-series approximation to T_n. Hence, it is necessary to explain Smooth Function Model and Cramer Condition at this stage.

SFM (Smooth Function Model): (i) $T_n = n^{1/2}[H(\bar{Z}) - H(\mu_Z)]$, *where H(z) is 6 times continuously partially differentiable with respect to any mixture of components of z in a neighbourhood of* μ_Z. (ii) $\partial H(\mu_Z) \neq 0$. (iii) *The expected value of the product of any 16 components of Z exists.*

Assumption SFM insures that *H* has derivatives and *Z* has moments of sufficiently high order to obtain the Taylor series and Edgeworth expansions that are used to obtain a bootstrap approximation to the distribution of T_n that has an error of size $O(n^{-2})$... See Hall (1992a, pp. 52-56; 238-259) for a statement of the regularity conditions needed to obtain various levels of asymptotic and bootstrap approximations (Horowitz, 2001, p. 3176).

Cramer condition:

$$\lim_{\|t\|\to\infty} sup\ |E[exp(it'Z)]| < 1 \qquad (2.3.2)$$

where$= \sqrt{-1}$.

Cramer condition is satisfied if the distribution of Z has a non-degenerate absolutely continuous component in the sense of the Lebesgue decomposition.

Under the assumption of the smooth function model, the first-order asymptotic approximations to the exact finite-sample distribution of T_n make an error of size $O\left(n^{-1/2}\right)$. If T_n is not an asymptotically pivotal statistic, then the bootstrap has an error of size $O\left(n^{-1/2}\right)$ almost surely, which is the same as the size of the error made by the first-order asymptotic approximations. However, if T_n is an asymptotically pivotal statistic, then the bootstrap has an error of size $O(n^{-1})$. Thus, in this case, the bootstrap is more accurate than the first-order asymptotic theory for estimating the distribution of a smooth asymptotically pivotal statistic.

Whereas the error made by the first-order asymptotic approximations to the symmetrical distribution function is $O(n^{-1})$, the error made by the bootstrap approximation is $O\left(n^{-3/2}\right)$ in the case of asymptotically pivotal statistic. These errors are $O\left(n^{-1/2}\right)$ and $O(n^{-1})$, respectively, for the approximation to the one-sided distribution function.

2.3.3 Bootstrap Critical Values for Hypothesis Tests

Let T_n be a statistic for testing a hypothesis H_0 about the sampled population. Assume that under H_0, T_n is asymptotically pivotal and satisfies assumptions of Smooth Function Model and Cramer condition.

In the case of a symmetrical and two-sided test, if T_n is an asymptotically pivotal statistic, then the asymptotic critical value approximates the exact finite sample critical value with an error of size $O(n^{-1})$. In contrast, the bootstrap critical value for the same test differs from the exact, finite-sample critical value by $O\left(n^{-3/2}\right)$ almost surely. Hence, bootstrap gives more correct critical value.

In the case of a symmetrical and two-sided test, when the test statistic is asymptotically pivotal, the difference between the nominal and true Rejection Probabilities (RP) is $O(n^{-1})$ with the asymptotic critical value. However, the nominal RP with a bootstrap critical value differs from the true RP by $O(n^{-2})$. The bootstrap does not achieve the same accuracy for one-tailed tests. For such tests, the difference between the nominal and true RP's with asymptotic critical values is $O\left(n^{-1/2}\right)$, whereas the difference with a bootstrap critical value is usually $O(n^{-1})$. For the details, see Hall (1992, p. 102-103).Horowitz (2001) states that "tests based on statistics that are asymptotically chi-square distributed behave like symmetrical, two-tailed tests" (p. 3183). It is necessary to remind that if the distribution of T_n is symmetrical about 0, then equal-tailed and symmetrical tests are the same. Otherwise, they are different.

2.3.4 Confidence Intervals

Let T_n be asymptotically pivotal and satisfy assumptions of Smooth Function Model and Cramer condition.

When the asymptotic critical value is used, the true and nominal coverage probabilities of a symmetrical and two-sided confidence intervals differ by $O(n^{-1})$, whereas they differ by $O(n^{-2})$ when the bootstrap critical value is used. With asymptotic critical values, the true and nominal coverage probabilities of for one-sided and equal-tailed confidence intervals differ by $O\left(n^{-1/2}\right)$, whereas the differences are $O(n^{-1})$ with bootstrap critical values. In special cases such as the slope coefficients of homoscedastic, linear, mean-regressions, the differences with bootstrap critical values are $O\left(n^{-3/2}\right)$ as mentioned in Horowitz (2001).

2.3.5 The Importance of Asymptotically Pivotal Statistics

When the bootstrap techniques applied to statistics that are not asymptotically pivotal, it can't provide higher-order approximations to their distributions. Horowitz (2001) states that "the errors of bootstrap estimates of the distributions of statistics that are not asymptotically pivotal converge to zero at the same rate as the errors made by first-order asymptotic approximations" (p. 3185). However, it is possible to obtain higher-order approximations to the distributions of statistics that are not asymptotically pivotal through the use of bootstrap iteration [Beran (1987,1988); Hall(1992)] or bias-correction methods [Efron (1987)]. On the other hand, Horowitz (2001) also states that "bias correction methods are not applicable to symmetrical tests and confidence intervals", and "bootstrap iteration is highly computationally intensive, which makes it unattractive when an asymptotically pivotal statistic is available" (p. 3185).

2.3.6 Recentering

Horowitz (2001) explains the importance of recentering for the bootstrap with theoretical details. Roughly speaking, implementing the moment condition which is not hold in the population but the bootstrap samples, makes the bootstrap estimator of the distribution of the statistic for testing the overidentifying restrictions inconsistent. Because of this problem, the bootstrap method does not give asymptotic refinements. To solve this problem, recentering procedure is implied.

2.4 Dependent Data

Using independent bootstrap samples, asymptotic refinements with dependent data can't be obtained. Hence, in the case of working with dependent data, Horowitz (2001) states that "bootstrap sampling must be carried out in a way that suitably captures the dependence of the data-generation process" (p. 3188). This section describes several methods for doing this.

2.4.1 Methods for Bootstrap Sampling with Dependent Data

Horowitz (2001) states that "bootstrap sampling that captures the dependence of the data can be carried out relatively easily if there is a parametric model, such as an ARMA model, that reduces the data-generation process to a transformation of independent variables" (p. 3188). In this case and under suitable regularity conditions, the bootstrap has properties that are essentially the same as they are when the data are i.i.d. See Andrews (1999) and Bose (1988, 1990). However, when there is no parametric model that reduces the data-generation process to independent sampling from some probability distribution, the bootstrap can be implemented using the Block Bootstrap and the Sieve Bootstrap methods.

2.4.1.1 The Block Bootstrap

This method includes dividing the data into blocks and sampling the blocks randomly with replacement. The block bootstrap is important in GMM estimation with dependent data, because the moment conditions on which GMM estimation is based usually do not specify the dependence structure of the GMM residuals.

The blocks may be non-overlapping [Carlstein (1986)] or overlapping [Hall (1985), Künsch (1989), Politis and Romano (1994)]. To describe these blocking methods more precisely, let the data consist of observations $\{X_i: i = 1, \dots, n\}$. With non-overlapping blocks of length l, block 1 is observations $\{X_j: j = 1, \dots, l\}$ block 2 is observations $\{X_{l+j}: j = 1, \dots, l\}$, and so forth. With overlapping blocks of length l, block 1 is observations $\{X_j: j = 1, \dots, l\}$, block 2 is observations $\{X_{j+1}: j = 1, \dots, l\}$ and so forth. The bootstrap sample is obtained by sampling blocks randomly with replacement and laying them end-to-end in the order sampled. It is also possible to use overlapping blocks with lengths that are sampled randomly from the geometric distribution [Politis and Romano (1994)]. The block bootstrap with random block lengths is also called the *stationary bootstrap* because the resulting bootstrap data series is stationary, whereas it is not

with overlapping or non-overlapping blocks of fixed (non-random) lengths (Horowitz, 2001, p. 3189).

Regardless of whether the blocks are overlapping or non-overlapping, the block length must increase with increasing sample size *n* to make bootstrap estimators of moments and distribution functions consistent (Carlstein 1986, Künsch 1989, Hallet al. 1995). The block length must also increase with increasing *n* to enable the block bootstrap to achieve asymptotically correct coverage probabilities for confidence intervals and rejection probabilities for tests. When the objective is to estimate a moment or distribution function, the asymptotically optimal block length may be defined as the one that minimizes the asymptotic mean-square-error of the block bootstrap estimator. When the objective is to form a confidence interval or test a hypothesis, the asymptotically optimal block length may be defined as the one that minimizes the ECP (the error in the coverage probability) of the confidence interval or ERP (the error in the rejection probability) of the test. The asymptotically optimal block length and the corresponding rates of convergence of block bootstrap estimation errors, ECP's and ERP's depend on what is being estimated (e.g., bias, a one-sided distribution function, a symmetrical distribution function, etc.) (Härdle et al., 2002, p. 9).

Hall, Horowitz, & Jing (1995) showed that with either overlapping or non-overlapping blocks with non-random lengths, the asymptotically optimal block-length is $l \sim n^r$, where $r = 1/3$ for estimating bias or variance, $r = 1/4$ for estimating a one-sided distribution function, and $r = 1/5$ for estimating a symmetrical distribution function. Hall et al. (1995) also show that overlapping blocks provide somewhat higher estimation efficiency than non-overlapping ones. The efficiency difference is likely to be very small in applications, however. (Horowitz, 2001, p. 3190).

Lahiri (1999) investigated the asymptotic efficiency of the stationary bootstrap. He states that at least in terms of asymptotic RMSE (the root-mean-square estimation

error), the stationary bootstrap is unattractive relative to the block bootstrap with fixed-length blocks.

> Implementation of the block bootstrap in an application requires a method for choosing the block length with a finite sample. Hall et al. (1995) describe a subsampling method for doing this when the block lengths are non-random. The idea of the method is to use subsamples to create an empirical analogous of the mean-square error of the bootstrap estimator of the quantity of interest (Horowitz, 2001, p. 3190).

> Hall et al. (1995) and Lahiri (1999) have compared the estimation errors made by the overlapping- and non-overlapping-blocks bootstraps....They find that the bootstrap is less accurate with non-overlapping blocks because the variance of the bootstrap estimator is larger with non-overlapping blocks than with overlapping ones. The bias of the bootstrap estimator is the same for non-overlapping and overlapping blocks. It should be noted, however, the differences between the AMSE's (the asymptotic mean-square-error) with the two types of blocking occurs in higher-order terms of the statistics of interest and, therefore, is often very small in magnitude (Härdle et al., 2002, p. 17).

2.4.1.2 The Sieve Bootstrap

This method has been proposed by Kreiss (1992) and Bühlmann (1997). In this method, the infinite-order autoregression is replaced by an approximating autoregression with a finite-order that increases at a suitable rate as $n \to \infty$. Horowitz (2001) defines the procedure as "the coefficients of the finite-order autoregression are estimated, and the bootstrap is implemented by sampling the centered residuals from the estimated finite-order model" (p. 3190). Bühlmann (1997) gives conditions under which this procedure yields consistent estimators of variances and distribution functions.

2.4.2 Companion Stochastic Process

In their study, Kreiss & Paparoditis (2011) investigate which bootstrap procedures asymptotically work (are *consistent* or *valid* for short) for what kind of statistics and why this is the case or, in the negative case, why it is not the case. By the phrase *the bootstrap asymptotically works* they mean that the approximation error of the bootstrap distribution for the standardized distribution of the estimator converges to zero as the sample size increases to infinity. They explain the key points regarding to this aim as follows:

> One key point is to work out what a specific bootstrap procedure really mimics and another one is to investigate what features of the underlying data generating process necessarily have to be mimicked in order to be able to lead to a consistent bootstrap method. Of course the latter question is not only related to the underlying data generating mechanism but also to the statistic and parameter of interest. To be a little bit more precise, let us assume that we are interested in the expectation of the underlying process and that we consider the mean of our observed data as an estimator. Under rather mild assumptions on the dependence structure of an underlying stationary process we obtain that the asymptotic distribution of the mean only depends on the whole autocovariance function (to be precise, the sum overall autocovariances) or equivalently on the spectral density evaluated at zero frequency. This means that for a consistent bootstrap procedure in such a situation it suffices to correctly imitate the second-order properties of the underlying process. For consistency of a bootstrap proposal it is not necessary to mimic further parts of the possibly much more complicated dependence structure of the data. However, it may be advantageous to mimic features of the dependence structure beyond second-order properties in order to improve the finite sample size behaviour of the bootstrap approximation of the distribution of the estimator of interest. Anyway it is of general interest to know what a specific bootstrap procedure really does. To shed light on this property we introduce for some bootstrap procedures so-called *companion* processes of the underlying processes (Kreiss & Paparoditis, 2011, p. 358).

They define the companion process as a process which only depends on the underlying stochastic process and the particular bootstrap procedure, and not on the statistics and parameter of interest. If for a particular bootstrap method and a statistic of interest, the asymptotic distribution of the relevant test statistic does not change if we switch from the underlying process to its companion process, then this bootstrap method asymptotically works for these specific estimators. Otherwise, the particular bootstrap method is not able to lead to valid results. Briefly, if the limit distribution of the test statistic obtained from the underlying process and the limit distribution of the test statistic obtained from the companion process are the same, bootstrap asymptotically works. The adequate condition is that both processes have the same first-order and second-order probabilistic qualifications. Hence, a proper bootstrap procedure should be able to mimic at least the necessary parts of the dependence structure of the data generating process. See Kreiss & Paparoditis (2011) for the details.

2.5 Bootstrap Iteration

If an asymptotic pivot is not available, asymptotic refinements can be obtained by applying the bootstrap to the bootstrap-generated asymptotic-pivot. The computational procedure is called *bootstrap iteration* or *prepivoting* because it entails drawing bootstrap samples from bootstrap samples as well as using the bootstrap to create an asymptotically pivotal statistic. Beran (1987) explains how to use prepivoting to form confidence regions. Hall (1986) describes an alternative approach to bootstrap iteration. The computational procedure for carrying out prepivoting and bootstrap iteration is given by Beran (1988).

Iteration can also be explained through the Russian martyoshka doll example. This approach is used in Hall (1992). In this thesis the results are given more explicitly. The iterative procedure is given until the third iteration as follows:

$$\frac{n_0}{n_1} \cong \frac{n_1}{n_2} \cong \frac{n_2}{n_3} \cong \frac{n_3}{n_4}$$

Here, n_0 may be thought of a multiple of n_1; that is, $n_0 = tn_1$ for some $t > 0$, or

$$n_0 - tn_1 = 0$$

which is called the population equation. Then the sample equation may be defined as

$$n_1 - tn_2 = 0$$

$$n_0 - tn_1 = 0 \implies t = n_0/n_1 \implies n_0 = tn_1 = t\hat{n}_{01}$$

$$n_1 - tn_2 = 0 \implies \hat{t}_{01} = n_1/n_2 \implies \hat{n}_{01} = \hat{t}_{01}n_1 = (n_1/n_2)n_1 = n_1^2/n_2$$

or

$$n_0/n_1 = n_1/n_2 \implies n_0n_2 = n_1^2 \implies \hat{n}_{01} = n_1^2/n_2.$$

Now, the new population equation is

$$n_0 - t\frac{n_1^2}{n_2} = 0$$

and the new sample equation is

$$n_1 - t\frac{n_2^2}{n_3} = 0\,.$$

The second iteration is started with

$$n_1/n_2 = n_2/n_3 \implies n_1n_3 = n_2^2$$

$$\frac{n_0n_2}{n_1^2} = \frac{n_1n_3}{n_2^2} \implies \hat{n}_{02} = \frac{n_1^3n_3}{n_2^3}$$

$$n_1 - t\frac{n_2^2}{n_3} = 0 \implies \hat{t}_{02} = \frac{n_1 n_3}{n_2^2}.$$

Now, the new population equation is

$$n_0 - t\frac{n_1^3 n_3}{n_2^3} = 0$$

and the new sample equation is

$$n_1 - t\frac{n_2^3 n_4}{n_3^3} = 0\,.$$

The third iteration is started with

$$n_2/n_3 = n_3/n_4 \implies n_2 n_4 = n_3^2$$

$$\frac{n_1 n_3}{n_2^2} = \frac{n_2 n_4}{n_3^2} \implies \hat{n}_1 = \frac{n_2^3 n_4}{n_3^3}$$

$$n_1 - t\frac{n_2^3 n_4}{n_3^3} = 0 \implies \hat{t}_{03} = \frac{n_1 n_3^3}{n_2^3 n_4}$$

$$\hat{n}_{03} = \hat{t}_{03}\hat{n}_{02} = \frac{n_1 n_3^3}{n_2^3 n_4}\frac{n_1^3 n_3}{n_2^3} = \frac{(n_1 n_3)^4}{n_2^6 n_4}.$$

Now, the new population equation is

$$n_0 - t\frac{(n_1 n_3)^4}{n_2^6 n_4} = 0$$

and the new sample equation is

$$n_1 - t\frac{(n_2 n_4)^4}{n_3^6 n_5} = 0\,.$$

Bootstrap iteration proceeds in an entirely analogous manner. For the details see Hall (1992, p. 21).

2.6 Special Problems

The bootstrap provides asymptotic refinements because it amounts to a one-term Edgeworth expansion which will be explained in the following subsection. Horowitz (2001) states that "the bootstrap cannot be expected to perform well when an Edgeworth expansion provides a poor approximation to the distribution of interest" (p. 3212). Besides, this technique does not perform well when the variance estimator used for Studentization has a high variance itself. Horowitz (2001) states that "in such cases Studentization is carried out with an estimator of the variance of an estimated variance" (p. 3212). Other problem is the behaviour of the bootstrap when the null hypothesis is false. See Horowitz (2001) for the details.

2.7 The Bootstrap and Edgeworth Expansion

2.7.1 Principles of Edgeworth Expansion

Classical statistical theory uses the expansions to provide analytical corrections similar to those that the bootstrap gives by numerical means. The arguments showing that the bootstrap yields asymptotic refinements are based on **Edgeworth expansion** of $(T_n \leq z)$. Therefore, in this section, it is concentrated on this expansion. This expansion, derived by Edgeworth in 1905, relates the pdf of a standard normal random variable Y to the $P(T_n \leq z)$ using the Chebyshev-Hermite polynomials. Before giving this expansion and the Cornish-Fisher expansion, special mathematical functions connected with this expansion are told briefly. These functions are Characteristic Function, Moment Generating Function, Cumulant Generating

Function, Hermite Polynomials, and Power Series. The theoretical explanations presented in this section are based on Boik (2006).

2.7.1.1 The Characteristic Function

The characteristic function of the random variable *Y* is

$$\phi_Y(t) = E\left(e^{it\prime Y}\right) = \int e^{it\prime y}\, dF_Y(y) \tag{2.7.1}$$

where $i = \sqrt{-1}$ and t is a fixed *p*-vector.

The characteristic function is the expectation of

$$e^{it\prime Y} = \cos(t'Y) + i\, \sin(t'Y). \tag{2.7.2}$$

The probability density function of *Y* is defined as

$$f(x) = \frac{d}{dx} F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it\prime y}\phi_Y(t)dt. \tag{2.7.3}$$

This equation is the result of *the inversion theorem* (See the equation 2.7.29).

2.7.1.2 The Moment Generating Function

The moment generating function of the random variable *Y* is

$$M_Y(t) = E(e^{t\prime Y}) = \int e^{t\prime y}\, dF_Y(y) \tag{2.7.4}$$

where t is a fixed *p*-vector.

2.7.1.3 The Cumulant Generating Function

Let *Y* be a scalar random variable whose MGF is $M_Y(t)$. If $ln[M_Y(t)]$ is expanded in a Taylor series around $t = 0$, the result is called the cumulant generating function

$$C_Y(t) = \ln[M_Y(t)] = \sum_{r=1}^{\infty} \frac{t^r}{r!}\kappa_r \tag{2.7.5}$$

where κ_r' s are named as **cumulant**.

2.7.1.4 Relations between CF, MGF and CGF

$$\frac{d}{dt}M_Y(t) = \frac{1}{i}\frac{d}{dt}\phi_Y(t) = \kappa_1 = E(Y) \quad , \quad \text{for } t = 0 \tag{2.7.6}$$

and

$$\frac{d^2}{dt^2}M_Y(t) = \frac{1}{i^2}\frac{d^2}{dt^2}\phi_Y(t) = \kappa_2 = Var(Y) \quad , \quad \text{for } t = 0. \tag{2.7.7}$$

2.7.1.5 Hermite Polynomials

Let *z* be a scalar and denote the standard normal pdf by $\varphi(z)$

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}. \tag{2.7.8}$$

The r^{th} Hermite polynomial, $H_r(z)$ is defined as

$$H_r(z) = \frac{(-1)^r}{\varphi(z)}\frac{d^r\varphi(z)}{dz^r} \tag{2.7.9}$$

for $r = 1, 2, \ldots$

$$H_0(z) = 1, H_1(z) = -\frac{\varphi'(z)}{\varphi(z)}, H_2(z) = \frac{\varphi''(z)}{\varphi(z)}, \ldots$$

To generate these polynomials efficiently, note that

$$\varphi(z-t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} e^{\frac{2zt}{2}} e^{-\frac{t^2}{2}} = \varphi(z) e^{tz-\frac{t^2}{2}}. \qquad (2.7.10)$$

If $\varphi(z-t)$ is expanded around $t = 0$ the equations below are obtained:

$$\varphi(z-t) = \varphi(z) + \varphi'(z)(-t) + \frac{1}{2!}\varphi''(z)(-t)^2 + \ldots \approx \varphi(z) + \varphi'(z)(-t)$$

$$\varphi(z-t) = \varphi(z) - t\varphi'(z) + \frac{t^2}{2!}\varphi''(z) - + \ldots$$

$$\frac{d\varphi(z-t)}{dt} = \frac{d}{dt}\varphi(z) - \left[\varphi'(z) + \frac{d}{dt}\varphi'(z)t + \varphi''(z)t + \frac{d}{dt}\varphi''(z)\frac{t^2}{2!}\right]. \qquad (2.7.11)$$

For $t = 0$, Equation (2.7.11) equals to

$$\frac{d\varphi(z-t)}{dt} = -\varphi'(z) = -\frac{d}{dz}\varphi(z)$$

and

$$\frac{d^j\varphi(z-t)}{(dt)^j} = (-1)^j\varphi^{(j)}(z) = (-1)^j\frac{d^j\varphi(z)}{(dz)^j}.$$

By the chain rule,

$$\varphi(z-t) = \varphi(z) + (-1)^1 t\varphi'(z) + (-1)^2\frac{t^2}{2!}\varphi''(z) + \ldots + (-1)^n\frac{t^n}{n!}\varphi^{(n)}(z) + \cdots$$

$$\varphi(z-t) = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}t^j\,\varphi^{(j)}(z) = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}t^j\frac{d^j\varphi(z)}{(dz)^j}$$

$$\varphi(z-t) \;=\; \sum_{j=0}^{\infty} \frac{(-1)^j}{\varphi(z)} \frac{d^j \varphi(z)}{(dz)^j} \frac{t^j}{j!} \varphi(z) \;=\; \sum_{j=0}^{\infty} \frac{t^j}{j!} H_j(z)\, \varphi(z)$$

$$\varphi(z-t) \;=\; \varphi(z) \sum_{j=0}^{\infty} \frac{t^j}{j!} H_j(z).$$

Due to the Equation (2.7.10);

$$e^{tz-\frac{t^2}{2}} \;=\; \sum_{j=0}^{\infty} \frac{\left(tz-\frac{t^2}{2}\right)^j}{j!} \;=\; \sum_{j=0}^{\infty} \frac{t^j}{j!} H_j(z).$$

Now match coefficients and solve for $H_r(z)$. The result is

$$H_r(z) \;=\; z^r \;-\; \frac{P_{r,2}\, z^{r-2}}{2^1 1!} \;+\; \frac{P_{r,4}\, z^{r-4}}{2^2 2!} \;-\; \frac{P_{r,6}\, z^{r-6}}{2^3 3!} \;+\; -\; \ldots \qquad (2.7.12)$$

where

$$P_{r,m} \;=\; \begin{cases} \dfrac{r!}{(r-m)!} & r \;\geq m\,, \\ 0 & r \;< m\,. \end{cases} \qquad (2.7.13)$$

The first six Hermite polynomials are the following:

$$H_0(z) \;= 1$$

Since

$$\varphi'(z) \;=\; \frac{1}{\sqrt{2\pi}} \frac{-2z}{2} e^{-\frac{z^2}{2}} \;=\; -z\varphi(z)$$

the first Hermite polynomial is

$$H_1(z) \;=\; -\frac{\varphi'(z)}{\varphi(z)} \;=\; \frac{z\varphi(z)}{\varphi(z)} \;= z. \qquad (2.7.14)$$

Since

$$\varphi''(z) = -\frac{1}{\sqrt{2\pi}}\left[e^{-\frac{z^2}{2}} - \frac{2z}{2} z\, e^{-\frac{z^2}{2}}\right] = (z^2 - 1)\varphi(z)$$

the second Hermite polynomial is

$$H_2(z) = \frac{\varphi''(z)}{\varphi(z)} = \frac{(z^2 - 1)\varphi(z)}{\varphi(z)} \; z^2 - 1. \qquad (2.7.15)$$

The rest polynomials are as follows:

$$H_3(z) = z^3 - 3z \qquad H_4(z) = z^4 - 6z^2 + 3 \qquad (2.7.16)$$

$$H_5(z) = z^5 - 10z^3 + 15z \quad H_6(z) = z^6 - 15z^4 + 45z^2 - 15 \quad (2.7.17)$$

The following two theorems connected with the Edgeworth Expansion.

Theorem 2.7.1 (*Orthogonal Properties of* H_r). *If* $H_r(z)$ *is the* r^{th} *Hermite polynomial, then*

$$\int_{-\infty}^{\infty} H_m(z)\, H_n(z)\varphi(z)dz = \begin{cases} 0 & m \neq n, \\ m! & m = n. \end{cases} \qquad (2.7.18)$$

(Boik, 2006, p. 117).

Theorem 2.7.2. Denote the r^{th} *Hermite polynomial by* $H_r(z)$ *and denote the standard normal pdf by* $\varphi(z)$. *Then*

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}(-it)^r e^{-\frac{t^2}{2}}dt = (-1)^r H_r(z)\varphi(z), \qquad \text{for } r = 1, 2, \ldots \quad (2.7.19)$$

and

$$\int_{-\infty}^{z} H_r(u)\varphi(u)du \begin{cases} \Phi(z) & r = 0, \\ -H_{r-1}(z)\varphi(z) & r \geq 1. \end{cases} \quad (2.7.20)$$

(Boik, 2006, p. 118).

2.7.1.6 Power Series

The power series is the standard method for solving linear *Ordinary Differential Equations* with *variable* coefficients. A power series is an infinite series of the form

$$\sum_{m=0}^{\infty} a_m(x - x_0)^m = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \ldots \quad (2.7.21)$$

Here, x is a variable, $a_0, a_1, a_2, \ldots$ are constants, called the *coefficients* of the series, x_0 is a constant, called the *center* of the series. The Taylor series of a function is a kind of power series. A Taylor expansion of function $f(x)$ about the value a is defined as

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \ldots + \frac{f^{(n)}(a)(x-a)^n}{n!} + R_n. \quad (2.7.22)$$

Here, R_n defines the remainder part of the expansion. If $a = 0$, then this expansion is named as the MacLaurin series

$$f(x) = f(0) + f'(0)x + \frac{f''(0)x^2}{2!} + \ldots + \frac{f^{(n)}(0)x^n}{n!} + R_n. \quad (2.7.23)$$

2.7.1.7 Edgeworth Expansions

An expansion, derived by Edgeworth in 1905, that relates the pdf, f, of a random variable, *X*, having expectation 0 and variance 1, to the probability density function of a standard normal distribution, using the Chebyshev-Hermite polynomials. In this

section, a theorem connected with the edgeworth expansion for the sample mean and its proof are given.

Theorem 2.7.3 (*Edgeworth*). *If the cdf of Y is continuous and differentiable, the the pdf and cdf of the sample mean are*

$$f_{\bar{Y}}(\bar{y}) = \frac{\sqrt{N}}{\sigma}\varphi(z)\left[1 + \frac{\rho_3}{6\sqrt{N}}H_3(z) + \frac{\rho_4}{24N}H_4(z) + \frac{\rho_3^2}{72N}H_6(z) + O\left(N^{-3/2}\right)\right]$$

$$F_{\bar{Y}}(\bar{y}) = \Phi(z) - \varphi(z)\left[\frac{\rho_3}{6\sqrt{N}}H_2(z) + \frac{\rho_4}{24N}H_3(z) + \frac{\rho_3^2}{72N}H_5(z) + O\left(N^{-3/2}\right)\right]$$

where $z = \frac{\bar{y} - \mu}{\sigma/\sqrt{N}}$, $\rho_j = \frac{\kappa_j(Y)}{\sigma^j}$, $\kappa_j(Y)$is the j^{th} cumulant of Y.

(Boik, 2006, p. 119).

Proof: Suppose that Y_i are i.i.d. for $i = 1, \ldots, N$ with mean μ and variance σ^2 and

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{N}}. \tag{2.7.24}$$

Then the characteristic function of Z is

$$\phi_Z(t) = E\left(e^{itZ}\right) = E\left(e^{it\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{N}}\right)}\right) = E\left(e^{\frac{it\bar{Y}}{\sigma/\sqrt{N}}}e^{-\frac{it\mu}{\sigma/\sqrt{N}}}\right)$$

$$\phi_Z(t) = E\left(e^{\frac{it\sum_{i=1}^{n}Y_i}{\sigma\sqrt{N}}}\right)exp\left\{\frac{-\sqrt{N}\,it\mu}{\sigma}\right\} = \left[E\left(e^{iY\frac{t}{\sigma\sqrt{N}}}\right)\right]^N exp\left\{\frac{-\sqrt{N}\,it\mu}{\sigma}\right\}$$

$$\phi_Z(t) = \left[\phi_Y\left(\frac{t}{\sigma\sqrt{N}}\right)\right]^N exp\left\{\frac{-\sqrt{N}\,it\mu}{\sigma}\right\} \tag{2.7.25}$$

since $\phi_Y(t) = E\left(e^{itY}\right)$.

Now, consider the cumulant generating function,

$$CGF_Z(t) \;=\; \ln \phi_Z(t) \;=\; \ln\left(\left[\phi_Y\left(\frac{t}{\sigma\sqrt{N}}\right)\right]^N exp\left\{\frac{-\sqrt{N}\,it\mu}{\sigma}\right\}\right)$$

$$CGF_Z(t) \;= N \ln\left[\phi_Y\left(\frac{t}{\sigma\sqrt{N}}\right)\right] - \frac{\sqrt{N}\,it\mu}{\sigma} \;= N\left(\ln\left[\phi_Y\left(\frac{t}{\sigma\sqrt{N}}\right)\right] - \frac{it\mu}{\sigma\sqrt{N}}\right) \quad (2.7.26)$$

$$\ln\left[\phi_Y\left(\frac{t}{\sigma\sqrt{N}}\right)\right] \;= \sum_{j=1}^{\infty}\left(\frac{it}{\sigma\sqrt{N}}\right)^j \frac{\kappa_j(Y)}{j!}$$

since

$$CGF_Z(t) \;=\; \ln \phi_Y(t) = \sum_{j=1}^{\infty}\frac{(it)^j}{j!}\kappa_j.$$

Now, since $\kappa_1(Y) \;= E(Y) \;=\; \mu$, for $j \;= 1$

$$\left(\frac{it}{\sigma\sqrt{N}}\right)^1 \frac{\kappa_1(Y)}{1!} \;=\; \frac{it\mu}{\sigma\sqrt{N}}$$

Equation (2.7.26) may be rewritten as

$$CGF_Z(t) \;= N\left(\sum_{j=2}^{\infty}\left(\frac{it}{\sigma\sqrt{N}}\right)^j \frac{\kappa_j(Y)}{j!} + \frac{it\mu}{\sigma\sqrt{N}} - \frac{it\mu}{\sigma\sqrt{N}}\right)$$

$$CGF_Z(t) \;= N\sum_{j=2}^{\infty}\left(\frac{it}{\sigma\sqrt{N}}\right)^j \frac{\kappa_j(Y)}{j!}. \quad (2.7.27)$$

If this function is expanded, the equation below is obtained:

$$CGF_Z(t) \;= N\left[\left(\frac{it}{\sigma\sqrt{N}}\right)^2 \frac{\kappa_2(Y)}{2!} + \left(\frac{it}{\sigma\sqrt{N}}\right)^3 \frac{\kappa_3(Y)}{3!} + \left(\frac{it}{\sigma\sqrt{N}}\right)^4 \frac{\kappa_4(Y)}{4!} + \ldots\right]$$

$$CGF_Z(t) = -\frac{t^2}{2} + \frac{(it)^3\kappa_3(Y)}{\sigma^3 6\sqrt{N}} + \frac{(it)^4\kappa_4(Y)}{\sigma^4 24N} + O\left(N^{-3/2}\right). \qquad (2.7.28)$$

Here, instead of the remainder part of the expansion, $O\left(N^{-3/2}\right)$ is written since, if one more term was showed in the expansion, this term would be

$$\frac{(it)^5\kappa_5(Y)}{\sigma^5 120N^{3/2}}$$

and

$$\lim_{N\to\infty} \frac{\frac{1}{N^{3/2}}\left(\frac{(it)^5\kappa_5(Y)}{\sigma^5 120}\right)}{\frac{1}{N^{3/2}}} = \frac{(it)^5\kappa_5(Y)}{\sigma^5 120}$$

the expansion would be bounded by a value which is different from zero.

Because of the inversion theorem, it is well known that

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\phi_Z(t)dt \qquad (2.7.29)$$

where

$$z = \frac{\bar{y} - \mu}{\sigma/\sqrt{N}}.$$

With a different presentation, the equation above may be rewritten as follows:

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz} exp\{CGF_Z(t)\}dt$$

By substitution of the expansion above,

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz} exp\left\{-\frac{t^2}{2} + \frac{(it)^3\kappa_3(Y)}{\sigma^3 6\sqrt{N}} + \frac{(it)^4\kappa_4(Y)}{\sigma^4 24N} + O\left(N^{-3/2}\right)\right\}dt$$

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz} exp\left\{-\frac{t^2}{2} + \frac{(it)^3}{6\sqrt{N}}\frac{\kappa_3(Y)}{\sigma^3} + \frac{(it)^4}{24N}\frac{\kappa_4(Y)}{\sigma^4} + O\left(N^{-3/2}\right)\right\} dt$$

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz} e^{-\frac{t^2}{2}}\left\{1 + \frac{(it)^3\rho_3}{6\sqrt{N}} + \frac{(it)^4\rho_4}{24N} + \frac{(it)^6\rho_3^2}{72N} + O\left(N^{-3/2}\right)\right\} dt$$

where

$$\rho_j = \frac{\kappa_j(Y)}{\sigma^j}$$

the equation below is obtained:

$$f_Z(z) = \varphi(z)\left[1 + \frac{\rho_3}{6\sqrt{N}}H_3(z) + \frac{\rho_4}{24N}H_4(z) + \frac{\rho_3^2}{72N}H_6(z) + O\left(N^{-3/2}\right)\right] \quad (2.7.30)$$

using *Theorem 2.7.2*.

The pdf of *Y* is obtained by transforming from *Z* to *Y*. The cdf of *Z* is obtained by integrating the pdf:

$$F_Z(z) = \int_{-\infty}^{z} \varphi(u)\left[1 + \frac{\rho_3}{6\sqrt{N}}H_3(u) + \frac{\rho_4}{24N}H_4(u) + \frac{\rho_3^2}{72N}H_6(u) + O\left(N^{-3/2}\right)\right] du$$

$$F_Z(z) = \Phi(z) - \varphi(z)\left[\frac{\rho_3}{6\sqrt{N}}H_2(z) + \frac{\rho_4}{24N}H_3(z) + \frac{\rho_3^2}{72N}H_5(z) + O\left(N^{-3/2}\right)\right] \quad (2.7.31)$$

since

$$\int_{-\infty}^{z} \varphi(u)H_0(u)\,du = \int_{-\infty}^{z} \varphi(u)\,du = \Phi(z).$$

To obtain cdf of the sample mean, use

$$F_{\bar{Y}}(\bar{y}) = P(\bar{Y} \leq \bar{y}) = P(Z \leq z) \text{ where } z$$
$$= \frac{\bar{y} - \mu}{\sigma/\sqrt{N}}. \qquad (2.7.32)$$

Hall (1992) explains the terms in Edgeworth expansion of the sample mean as follows:

> Third and fourth cumulants κ_3 and κ_4 are referred to as *skewness* and *kurtosis* respectively. The term of order $N^{-1/2}$ corrects the basic Normal approximation for the main effect of skewness, while the terms of order N^{-1} corrects for the main effect of kurtosis and the secondary effect of skewness (Hall, 1992, p. 45).

2.7.1.8 Cornish-Fisher Expansions

Cornish & Fisher (1937) constructed an expansion so that the percentiles of the distribution of Z (or Y) can be expressed in terms of the percentiles of the $N(0,1)$ distribution and vice-versa. First, however, a preliminary result is required.

Theorem 2.7.4. Denote the 100α *percentile of*

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{N}}$$

by y_α *and denote the* 100α *percentile of the* $N(0,1)$ *distribution by* z_α*. Suppose that a valid Edgeworth expansion for the distribution of Z exists. Then,*

$$y_\alpha - z_\alpha = O\left(N^{-1/2}\right)$$

(Boik, 2006, p. 120).

Proof: It follows from the Edgeworth expansion that

$$F_Z(a) = \Phi(a) + O\left(N^{-1/2}\right), \forall a .$$

Specifically,

$$\alpha = \Phi(z_\alpha) = F_Z(y_\alpha) \quad \text{and} \quad F_Z(y_\alpha) = \Phi(y_\alpha) + O\left(N^{-1/2}\right).$$

Accordingly,

$$\alpha = F_Z(y_\alpha) = \Phi(y_\alpha) + O\left(N^{-1/2}\right) = \Phi(z_\alpha)$$

$$\Longrightarrow \Phi(z_\alpha) - \Phi(y_\alpha) = \int_{y_\alpha}^{z_\alpha} \varphi(u)du = O\left(N^{-1/2}\right). \qquad (2.7.33)$$

Since

$$P(y_\alpha \leq X \leq z_\alpha) = F(z_\alpha) - F(y_\alpha) = \int_{y_\alpha}^{z_\alpha} f(x)dx$$

and since

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \text{ where } \quad y = \frac{x-\mu}{\sigma}$$

$$A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \Longrightarrow A^2 = \frac{1}{2\pi} \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right)$$

$$A^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{y^2+t^2}{2}} dydt$$

$$A^2 = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\infty} r\, e^{-\frac{r^2}{2}} dr = \frac{1}{2\pi} \int_0^{2\pi} \left(-\frac{1}{e^\infty} + \frac{1}{e^0} \right) d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1$$

where $y = r\sin\theta$, $t = r\cos\theta$, $dydt = rdrd\theta, 0 < r < \infty, 0 < \theta < 2\pi.$

$$A = 1 \Longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = 1 \Longrightarrow \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$$

Note that

$$\left|\int_{y_\alpha}^{z_\alpha} \varphi(u)du\right| \leq |z_\alpha - y_\alpha| \sup \varphi(\zeta) = |z_\alpha - y_\alpha|\left(\frac{1}{\sqrt{2\pi}}\right) \quad (2.7.34)$$

which implies that

$$\frac{|z_\alpha - y_\alpha|}{\sqrt{2\pi}} = O\left(N^{-1/2}\right) \Longrightarrow y_\alpha - z_\alpha = O\left(N^{-1/2}\sqrt{2\pi}\right) = O\left(N^{-1/2}\right). \quad (2.7.35)$$

Theorem 2.7.5 (*Cornish-Fisher*). *Denote the* 100α *percentile of the distribution of*

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{N}}$$

by y_α *and denote the* 100α *percentile of the* $N(0,1)$ *distribution by* z_α. *Then,*

$$y_\alpha = z_\alpha + \frac{\rho_3}{6\sqrt{N}}(z_\alpha^2 - 1) + \frac{\rho_4}{24N}(z_\alpha^3 - 3z_\alpha) - \frac{\rho_3^2}{36N}(2z_\alpha^3 - 5z_\alpha) + O\left(N^{-3/2}\right)$$

$$z_\alpha = y_\alpha - \frac{\rho_3}{6\sqrt{N}}(y_\alpha^2 - 1) - \frac{\rho_4}{24N}(y_\alpha^3 - 3y_\alpha) + \frac{\rho_3^2}{36N}(4y_\alpha^3 - 7y_\alpha) + O\left(N^{-3/2}\right)$$

Furthermore, the above expansions hold for all $\alpha \in (0,1)$. *If* α *is a random variable with a* $Unif(0,1)$ *distribution, then* z_α *is a realization of a* $N(0,1)$ *random variable having the same distribution as*

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{N}}.$$

Accordingly,

$$P(Z^* \leq z) = \Phi(z) + O\left(N^{-3/2}\right)$$

where

$$Z^* = Z - \frac{\rho_3}{6\sqrt{N}}(Z^2 - 1) - \frac{\rho_4}{24N}(Z^3 - 3Z) + \frac{\rho_3^2}{36N}(4Z^3 - 7Z)$$

(Boik, 2006, p. 121).

Proof: It is well known that

$$\alpha = \Phi(z_\alpha) = F_Z(y_\alpha)$$

and

$$F_Z(y_\alpha) = \Phi(y_\alpha) + O\left(N^{-1/2}\right).$$

$$F_Z(y_\alpha) = \Phi(y_\alpha) - \varphi(y_\alpha)\left[\frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) + \frac{\rho_4}{24N}H_3(y_\alpha) + \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)\right]$$

$$\Phi(z_\alpha) - \Phi(y_\alpha) = -\varphi(y_\alpha)\left[\frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) + \frac{\rho_4}{24N}H_3(y_\alpha) + \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)\right]$$
$$= \alpha$$

Expand $\Phi(z_\alpha)$ in a Taylor series around $z_\alpha = y_\alpha$:

$$\Phi(z_\alpha) = \frac{(z_\alpha - y_\alpha)^0}{0!}\Phi(y_\alpha) + \frac{(z_\alpha - y_\alpha)^1}{1!}\left(\frac{d}{dz_\alpha}\Phi(y_\alpha)\right)$$
$$+ \frac{(z_\alpha - y_\alpha)^2}{2!}\left(\frac{d^2}{(dz_\alpha)^2}\Phi(y_\alpha)\right) + o(|z_\alpha - y_\alpha|^3). \qquad (2.7.36)$$

By the Equation (2.7.9),

$$\frac{d}{dz_\alpha}\Phi(z_\alpha) = \varphi(z_\alpha), \quad \frac{d^2}{(dz_\alpha)^2}\Phi(z_\alpha) = -H_1(z_\alpha)\varphi(z_\alpha)$$

the Equation (2.7.36) may be rewritten as follows:

$$\Phi(z_\alpha) = \Phi(y_\alpha) + \varphi(y_\alpha)(z_\alpha - y_\alpha) - \frac{1}{2}\varphi(y_\alpha)H_1(y_\alpha)(z_\alpha - y_\alpha)^2 + O\left(N^{-3/2}\right)$$

or

$$\Phi(z_\alpha) = \Phi(y_\alpha) + \varphi(y_\alpha)\left[(z_\alpha - y_\alpha) - \frac{1}{2}H_1(y_\alpha)(z_\alpha - y_\alpha)^2\right] + O\left(N^{-3/2}\right). \quad (2.7.37)$$

Substituting the expansion for $\Phi(z_\alpha)$ into the expression for $\Phi(z_\alpha) - \Phi(y_\alpha)$,

$$\Phi(z_\alpha) - \Phi(y_\alpha) = \varphi(y_\alpha)(z_\alpha - y_\alpha) - \frac{1}{2}\varphi(y_\alpha)H_1(y_\alpha)(z_\alpha - y_\alpha)^2 + O\left(N^{-3/2}\right)$$

is obtained. Since

$$F_Z(z) = \Phi(z) - \varphi(z)\left[\frac{\rho_3}{6\sqrt{N}}H_2(z) + \frac{\rho_4}{24N}H_3(z) + \frac{\rho_3^2}{72N}H_5(z) + O\left(N^{-3/2}\right)\right]$$

The Equation below is may be written:

$$F_Z(y_\alpha) - \Phi(y_\alpha) = \Phi(z_\alpha) - \Phi(y_\alpha) = -\varphi(y_\alpha)\left[\frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) + \frac{\rho_4}{24N}H_3(y_\alpha) + \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)\right].$$

Considering the Equation above and the Equation (2.7.37), the equality below may be written:

$$(z_\alpha - y_\alpha) - \frac{1}{2}H_1(y_\alpha)(z_\alpha - y_\alpha)^2 = -\frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)$$

$$(z_\alpha - y_\alpha) = \frac{1}{2}H_1(y_\alpha)(z_\alpha - y_\alpha)^2 - \frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right).$$

To calculate $(z_\alpha - y_\alpha)$, it is necessary to find $(z_\alpha - y_\alpha)^2$

$$(z_\alpha - y_\alpha)^2 = \left[\frac{1}{2}H_1(y_\alpha)(z_\alpha - y_\alpha)^2 - \frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)\right]^2$$

$$(z_\alpha - y_\alpha)^2 = \frac{\rho_3^2}{36N}\left[H_2(y_\alpha)\right]^2 + O\left(N^{-3/2}\right).$$

Considering $(z_\alpha - y_\alpha)^2$and $(z_\alpha - y_\alpha)$, the equality below is obtained:

$$(z_\alpha - y_\alpha) = \frac{1}{2}H_1(y_\alpha)\frac{\rho_3^2}{36N}[H_2(y_\alpha)]^2 - \frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)$$

$$\Longrightarrow z_\alpha = y_\alpha + H_1(y_\alpha)\frac{\rho_3^2}{72N}[H_2(y_\alpha)]^2 - \frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right).$$

The expansions for z_α and for Z^* that were claimed in the statement of the theorem are obtained by explicitly writing out the Hermite polynomials. To obtain the expansion for y_α , write y_α as

$$y_\alpha = \delta_0 + \frac{1}{\sqrt{N}}\delta_1 + \frac{1}{N}\delta_2 + O\left(N^{-3/2}\right). \qquad (2.7.38)$$

Substitute this expansion into the right-hand-side of

$$0 = y_\alpha - z_\alpha + H_1(y_\alpha)\frac{\rho_3^2}{72N}[H_2(y_\alpha)]^2 - \frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)$$

and then collect terms of same order. The result is

$$0 = -z_\alpha + \delta_0 + \frac{1}{\sqrt{N}}\delta_1 + \frac{1}{N}\delta_2 + H_1(y_\alpha)\frac{\rho_3^2}{72N}[H_2(y_\alpha)]^2 - \frac{\rho_3}{6\sqrt{N}}H_2(y_\alpha) - \frac{\rho_4}{24N}H_3(y_\alpha) - \frac{\rho_3^2}{72N}H_5(y_\alpha) + O\left(N^{-3/2}\right)$$

$$0 = (\delta_0 - z_\alpha) + \frac{1}{\sqrt{N}}\left[\delta_1 - \frac{\rho_3}{6}H_2(y_\alpha)\right] + \frac{1}{N}\left[\delta_2 + \frac{\rho_3^2}{72}H_1(y_\alpha)[H_2(y_\alpha)]^2 - \frac{\rho_3}{3}\delta_0\delta_1 - \frac{\rho_4}{24}H_3(y_\alpha) - \frac{\rho_3^2}{72}H_5(y_\alpha)\right] + O\left(N^{-3/2}\right).$$

When the counterparts of Hermite polynomials in the above equation are written with notation δ,

$$0 = (\delta_0 - z_\alpha) + \frac{1}{\sqrt{N}}\left[\delta_1 - \frac{\rho_3}{6}(\delta_0^2 - 1)\right] + \frac{1}{N}\left[\delta_2 + \frac{\rho_3^2}{72}\delta_0(\delta_0^2 - 1)^2 - \frac{\rho_3}{3}\delta_0\delta_1 - \frac{\rho_4}{24}(\delta_0^3 - 3\delta_0) - \frac{\rho_3^2}{72}(\delta_0^5 - 10\delta_0^3 + 15\delta_0)\right] + O\left(N^{-3/2}\right)$$

since

$$H_1(z) = z \implies H_1(y_\alpha) = \delta_0$$

$$H_2(z) = z^2 - 1 \implies H_2(y_\alpha) = \delta_0^2 - 1$$

$$H_3(z) = z^3 - 3z \implies H_3(y_\alpha) = \delta_0^3 - 3\delta_0$$

$$H_5(z) = z^5 - 10z^3 + 15z \implies H_5(y_\alpha) = \delta_0^5 - 10\delta_0^3 + 15\delta_0.$$

The right-hand-side of the above equation is zero for all values of ρ_3 and ρ_4 if and only if the $O\left(N^{-i/2}\right)$ term is zero for $i = 0, 1, 2$. Accordingly,

$$\delta_0 - z_\alpha = 0 \Leftrightarrow \delta_0 = z_\alpha$$

$$\delta_1 - \frac{\rho_3}{6}(\delta_0^2 - 1) = 0 \Leftrightarrow \delta_1 = \frac{\rho_3}{6}(\delta_0^2 - 1) = \frac{\rho_3}{6}(z_\alpha^2 - 1)$$

$$\left[\delta_2 + \frac{\rho_3^2}{72}\delta_0(\delta_0^2 - 1)^2 - \frac{\rho_3}{3}\delta_0\delta_1 - \frac{\rho_4}{24}(\delta_0^3 - 3\delta_0) - \frac{\rho_3^2}{72}(\delta_0^5 - 10\delta_0^3 + 15\delta_0)\right] = 0 \Leftrightarrow$$

$$\delta_2 = \left[-\frac{\rho_3^2}{72}\delta_0(\delta_0^4 - 2\delta_0^2 + 1) + \frac{\rho_3}{3}\delta_0\delta_1 + \frac{\rho_4}{24}(\delta_0^3 - 3\delta_0) + \frac{\rho_3^2}{72}(\delta_0^5 - 10\delta_0^3 + 15\delta_0)\right]$$

$$\delta_2 = \frac{\rho_4}{24}(\delta_0^3 - 3\delta_0) - \frac{\rho_3^2}{36}(4\delta_0^3 - 7\delta_0) = \frac{\rho_4}{24}(z_\alpha^3 - 3z_\alpha) - \frac{\rho_3^2}{36}(2z_\alpha^3 - 5z_\alpha).$$

Using these obtained equalities, the equation below is obtained:

$$y_\alpha = z_\alpha + \frac{\frac{\rho_3}{6}(z_\alpha^2 - 1)}{\sqrt{N}} + \frac{\frac{\rho_4}{24}(z_\alpha^3 - 3z_\alpha) - \frac{\rho_3^2}{36}(2z_\alpha^3 - 5z_\alpha)}{N} + O\left(N^{-3/2}\right). \quad (2.7.39)$$

2.7.2 An Edgeworth View of the Bootstrap

A key assumption for bootstrap to accurately approximate the "continuity correction" terms in an Edgeworth expansion is that the sampling distribution would typically be required to satisfy Cramér's condition. Hence, the performance of the bootstrap about this concept is valid under the "smooth function model". Hall (1992) defines bootstrap as a device for correcting an Edgeworth expansion or a Cornish-

Fisher expansion for the first error term, due to the main effect of skewness. He also explains this matter as follows:

> When used correctly, the bootstrap approximation effectively removes the first error term in an Edgeworth expansion, and so its performance is generally an order of magnitude better than if only the "0^{th} order" term, usually attributable to Normal approximation, had been accounted for (Hall, 1992, p. 108).

These corrections may be done for skewness and kurtosis, not just skewness. However, since a finite Edgeworth expansion is generally not a monotone function, these corrections do not always perform particularly well. Hence, a second Edgeworth correction is sometimes resulted in inferior coverage probability.

2.8 Sufficient Bootstrap

While the bootstrap which is used in this section in the meaning of "conventional bootstrap" is seen as a special case of simple random sampling with replacement where the sample size n becomes equal to the population size N, as defined by Singh & Sedory (2011), sufficient bootstrap technique is based on retaining only distinct individual responses. Hence, in a sufficient bootstrap sample, units never appear more than once. However, if more than one unit have the same value in a sufficient bootstrap sample, they are seen as different units. So, they all can be in the same sufficient bootstrap sample. Singh & Sedory (2011) states that "this is especially important when estimating a proportion with the proposed estimator where the outcome variable is a Bernoulli variate which only takes on the values 0 and 1" (p. 1634). It should be taken into consideration that the word "sufficient" is not tightly connected with sufficiency in terms of likelihood perspective. Singh & Sedory (2011) evaluate the performance of the sufficient bootstrap over the conventional bootstrap by considering a population consisting of $N = 3$ units. Their results are summarized and listed as follows:

1. The relative bias in the sufficient bootstrapping estimator of standard deviation is much less than that based on conventional bootstrapping.
2. While conventional bootstrapping underestimates the standard deviation, sufficient bootstrapping overestimates the standard deviation.
3. The estimator of the coefficient of variation based on the proposed sufficient bootstrapping has less relative bias compared to the value based on the conventional bootstrapping.

They presented the theoretical formulations for the expected value and the variance of the sufficient bootstrap estimate for the mean.

Singh & Sedory (2011) carry out a simulation study considering two different situations. Firstly, they check whether the estimators of the mean, the variance, the standard deviation and the coefficient of variation for a quantitative variable are more unbiased or not when using the sufficient bootstrapping rather than the conventional bootstrapping. They state that since the sufficient bootstrapping gives more unbiased estimation values for all the parameters mentioned above, as a result of using sufficient bootstrapping these parameters have significantly smaller mean squared errors than that in the case of conventional bootstrapping, and also the estimators obtained by the sufficient bootstrap have higher percent relative efficiencies than the estimators obtained by the conventional bootstrapping. Moreover, they check whether the estimator of the population proportion for a qualitative variable is more unbiased or not when using the sufficient bootstrapping. They state that both estimators obtained by two different bootstrap methods have nearly equal percent relative biases. However, when considering their mean squared errors, since the estimators obtained by using the sufficient bootstrapping have smaller mean squared errors, the percent relative efficiency of this estimator is higher. Hence, the sufficient bootstrap method should be preferred instead of the conventional bootstrap especially if the number of distinct units in a sufficient bootstrap sample follows the Feller (1957) distribution. In this thesis, the performance of the sufficient bootstrapping for the unit root test will be studied.

CHAPTER THREE
TIME SERIES ANALYSIS

3.1 Basic Definitions

> A time series is an ordered sequence of observations. Although the ordering is usually through time, particularly in terms of some equally spaced time intervals, the ordering may also be taken through other dimensions, such as space.... A time series, such as electric signals and voltage that can be recorded continuously in time, is said to be continuous. A time series, such as interest rates, yields, and volume of sales, which are taken only at specific time intervals, is said to be discrete (Wei, 2006, p. 1).

The observations in a time series are dependent or correlated, and therefore the order of the observations is important. Hence, statistical procedures and techniques that rely on independence assumption are no longer applicable, and different methods are needed. The body of statistical methodology available for analyzing time series is referred to as time series analysis.

A time series that exhibit variation about a fixed level are said to be *stationary in the mean*. A time series that exhibit an overall upward or downward trend are said to be *nonstationary in the mean*. In addition, if the variance of the series increases as the level of the series increases, that time series are said to be *nonstationary in the variance*. A time series that exhibit a regular increase or decrease in the same periods of a year that is a time series containing seasonal variation are called *seasonal time series*. The seasonal time series are also a kind of nonstationary time series. Nonstationary time series can be reduced to stationary series by proper transformations.

The time series approach which uses autocorrelation and partial autocorrelation functions to study the evolution of a time series through parametric models is known as *time domain analysis*. An alternative approach which uses spectral functions to

study the nonparametric decomposition of a time series into its different frequency components is known as *frequency domain analysis*.

Univariate time series analysis deals with the observations of only one variable. *Multivariate time series analysis* involves simultaneous observations on several variables.

3.2 Fundamental Concepts

A ***stochastic process*** is a family of time indexed random variables $Z(w,t)$, where *w* belongs to a sample space and *t* belongs to an index set. For a fixed t, $Z(w,t)$ is a random variable. For a given w, $Z(w,t)$, as a function of t, is called a sample function or *realization*. The population that consists of all possible realizations is called the *ensemble* in stochastic processes and time series analysis. Thus, a time series is a realization or sample function from a certain stochastic process (Wei, 2006, p. 6).

For a finite set of random variables $\{Z_{t_1},\ Z_{t_2}, \dots, Z_{t_n}\}$ from a stochastic process $\{Z(w,t): t\ = 0, \pm1, \pm2, \dots\}$, the *n*-dimensional distribution function is defined by

$$F_{Z_{t_1},\ \dots\ ,Z_{t_n}}(x_1, \dots, x_n)\ = P\{w: Z_{t_1}\ \leq\ x_1, \dots, Z_{t_n}\ \leq\ x_n\} \qquad (3.2.1)$$

where x_i, $i\ = 1, \dots, n$ are any real numbers. A process is said to be first-order stationary in distribution if its one-dimensional distribution function is time invariant, i.e., if $F_{Z_{t_1}}(x_1)\ =\ F_{Z_{t_1+k}}(x_1)$ for any integers t_1, k and $t_1\ + k$; second-order stationary in distribution if $F_{Z_{t_1},Z_{t_2}}(x_1, x_2)\ =\ F_{Z_{t_1+k},Z_{t_2+k}}(x_1, x_2)$ for any integers t_1, t_2, k, $t_1\ + k$ and $t_2\ + k$; and n^{th}-order stationary in distribution if

$$F_{Z_{t_1},\dots,Z_{t_n}}(x_1, \dots, x_n)\ =\ F_{Z_{t_1+k},\dots,Z_{t_n+k}}(x_1, \dots, x_n) \qquad (3.2.2)$$

for any n-tuple $(t_1, \dots, t_n)$ and k of integers. A process is said to be *strictly stationary* if (3.2.2) is true for any n, i.e., $n = 1, 2, \dots$

For a given real-valued process, that is the process which assumes only real values, $\{Z_t : t = 0, \pm 1, \pm 2, \dots\}$, the mean function of the process is defined as

$$\mu_t = E(Z_t) \tag{3.2.3}$$

the variance function of the process is defined as

$$\sigma_t^2 = E[(Z_t - \mu_t)^2] \tag{3.2.4}$$

the covariance function between Z_{t_1} and Z_{t_2} is defined as

$$\gamma(t_1, t_2) = E\left[\left(Z_{t_1} - \mu_{t_1}\right)\left(Z_{t_2} - \mu_{t_2}\right)\right] \tag{3.2.5}$$

the correlation function between Z_{t_1} and Z_{t_2} is defined as

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}. \tag{3.2.6}$$

For a strictly stationary process with the first two moments finite, the covariance and the correlation between Z_t and Z_{t+k} depend only on the time difference k.

Wei (2006) explains the importance of the autocorrelation functions and the partial autocorrelation functions in time series as follows:

> A stochastic process is said to be a normal or Gaussian process if its joint probability distribution is normal. Because a normal distribution is uniquely characterized by its first two moments, strictly stationary and weakly stationary are equivalent for a Gaussian process.... Like other areas in statistics, most time

series results are established for Gaussian processes. Thus, the autocorrelation functions and the partial autocorrelation functions... become fundamental tools in time series analysis (Wei, 2006, p. 10).

The covariance between Z_t and Z_{t+k} is defined as

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E[(Z_t - \mu)(Z_{t+k} - \mu)] \tag{3.2.7}$$

the correlation between Z_t and Z_{t+k} is defined as

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \tag{3.2.8}$$

where $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$. As functions of k, γ_k is called the ***autocovariance function*** and ρ_k is called the ***autocorrelation function*** (ACF) in time series analysis.

For a stationary process, the autocovariance function γ_k and the autocorrelation function ρ_k have the following properties:

1. $\gamma_0 = Var(Z_t)$; $\rho_0 = 1$.
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$.
3. $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$ for all k, i.e., γ_k and ρ_k are even functions and hence symmetric about the lag $k = 0$ since the time difference between Z_t and Z_{t+k}, and Z_t and Z_{t-k} are being the same.
4. γ_k and ρ_k are positive semidefinite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0 \tag{3.2.9}$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0 \tag{3.2.10}$$

for any set of time points $t_1, t_2, \dots, t_n$ and any real numbers $\alpha_1, \alpha_2, \dots, \alpha_n$.

The ***partial autocorrelation function*** is defined as the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the intervening variables $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$ has been removed, that is,

$$Corr(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1}) . \tag{3.2.11}$$

ϕ_{kk} has become a standard notation for the partial autocorrelation between Z_t and Z_{t+k} in time series literature. As a function of k, ϕ_{kk} is usually referred to as the partial autocorrelation function (PACF).

A process $\{a_t\}$ is called a ***white noise process*** if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean $E(a_t) = \mu_a$, usually assumed to be 0, constant variance $Var(a_t) = \sigma_a^2$ and $\gamma_k = Cov(a_t, a_{t+k}) = 0$ for all $k \neq 0$. The basic phenomenon of the white noise process is that its ACF and PACF are identically equal to zero. This process plays the role of an orthogonal basis in the general vector and function analysis. A white noise process is Gaussian if its joint distribution is normal.,

3.3 Stationary Time Series Models

3.3.1 Autoregressive Processes

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \dots + \phi_p \dot{Z}_{t-p} + a_t \tag{3.3.1}$$

or

$$\phi_p(B) \dot{Z}_t = a_t \tag{3.3.2}$$

where $\phi_p(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$ and $\dot{Z}_t = Z_t - \mu$.

The process is always invertible. To be stationary, the roots of $\phi_p(B) = 0$ must lie outside of the unit circle. The AR processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock. Yule (1927) used an AR process to describe the phenomena of sunspot numbers and the behaviour of a simple pendulum.

For an autoregressive process, the ACF ρ_k tails off as a mixture of exponential decays or damped sine waves depending on the roots of $\phi_p(B) = 0$. Damped sine waves appear if some of the roots are complex. The PACF ϕ_{kk} vanishes after lag p.

The first-order autoregressive process is named as AR(1) process

$$(1 - \phi_1 B)\dot{Z}_t = a_t \tag{3.3.3}$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t \,. \tag{3.3.4}$$

The AR(1) process is sometimes called the Markov process since the distribution of $\dot{Z}_t$, given $\dot{Z}_{t-1}, \dot{Z}_{t-2}, \ldots$ is exactly the same distribution of $\dot{Z}_t$, given $\dot{Z}_{t-1}$.

In terms of a stationary AR(1) process, it is always referred to the case in which the parameter value is less than 1 in absolute value. The ACF of the AR(1) process decays exponentially (dies-down). The PACF of the AR(1) process shows a positive or negative spike at lag 1 depending on the sign of the parameter and then cuts off.

3.3.2 Moving Average Processes

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{3.3.5}$$

or

$$\dot{Z}_t = \theta_q(B) a_t \tag{3.3.6}$$

where $\theta_q(B) = \left(1 - \theta_1 B - \cdots - \theta_q B^p\right)$ and $\dot{Z}_t = Z_t - \mu$.

The process is always stationary. To be invertible, the roots of $\theta_q(B) = 0$ must lie outside of the unit circle. The MA processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time. The process arose as a result of the study by Slutzky (1927) on the effect of the moving average of random events.

For a MA process, the ACF ρ_k will vanish after lag q. The PACF ϕ_{kk} tails off as a mixture of exponential decays or damped sine waves depending on the roots of $\theta_q(B) = 0$. Damped sine waves appear if some of the roots are complex.

The first-order moving average process is named as MA(1) process

$$\dot{Z}_t = (1 - \theta_1 B)a_t \tag{3.3.7}$$

or

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} \tag{3.3.8}$$

In terms of an invertible MA(1) process, the case in which the parameter value is less than 1 in absolute value is referred. The ACF of the MA(1) process shows a positive or negative spike at lag 1 depending on the sign of the parameter and then cuts off. The PACF of the MA(1) process decays exponentially (dies-down).

3.3.3 The Dual Relationship Between AR(p) and MA(q) Processes

A finite-order stationary AR(p) process corresponds to an infinite-order MA process, and a finite-order invertible MA(q) process corresponds to an infinite-order AR process. This dual relationship between the AR(p) and MA(q) processes also exists in the autocorrelation and partial autocorrelation functions. The AR(p) process has its autocorrelations tailing off and partial autocorrelations cutting off, but the MA(q) process is the inverse (Wei, 2006, p. 57).

3.3.4 Autoregressive Moving Average ARMA(p,q) Processes

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t \tag{3.3.9}$$

where $\phi_p(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$ and $\theta_q(B) = \left(1 - \theta_1 B - \cdots - \theta_q B^p\right)$.

For the process to be invertible, the roots of $\theta_q(B) = 0$ have to lie outside of the unit circle. To be stationary, the roots of $\phi_p(B) = 0$ have to lie outside of the unit circle. Also, two polynomials, $\phi_p(B) = 0$ and $\theta_q(B) = 0$, have to share no common roots. Here, p and q are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively.

> The autocorrelation function of an ARMA(p,q) model tails off after lag q just like an AR(p) process, which depends only on the autoregressive parameters in the model. The first q autocorrelations $\rho_q, \rho_{q-1}, \ldots, \rho_1$, however, depend on both autoregressive and moving average parameters in the model and serve as initial values for the pattern. This distinction is useful in model identification.... Because the ARMA process contains the MA process as a special case, its PACF will also be a mixture of exponential decays or damped sine waves depending on the roots of, $\phi_p(B) = 0$ and $\theta_q(B) = 0$ (Wei, 2006, p. 59).

It is necessary to note that the sample phenomenon of a white noise series implies that the underlying model is either a random noise process or an ARMA process with its AR and MA polynomials being nearly equal. The assumption of no common roots between $\phi_p(B) = 0$ and $\theta_q(B) = 0$ in the mixed model is needed to avoid this confusion. Briefly, if in an ARMA model, at least one of the roots of, $\phi_p(B) = 0$ and $\theta_q(B) = 0$ is the same, that process changes to a white noise process.

In modelling time series, the models constructed with only a finite number of parameters are used. It is useful to know that for a fixed number of observations, the more parameters in a model, the less efficient is the estimation of the parameters.

Hence, other things being equal, in general, a simpler model is chosen to describe the phenomenon. This modelling criteria is the principle of parsimony in model building recommended by Tukey (1967) and Box and Jenkins (1976).

3.4 Nonstationary Time Series Models

Compared to the class of covariance stationary processes, nonstationary time series can occur in many different ways. They could have nonconstant means μ_t, or time-varying second moments such as nonconstant variance σ_t^2, or both of these properties.

3.4.1 Nonstationarity in the Mean

In this section, two classes of models that are useful in modelling time series nonstationary in the mean are introduced.

3.4.1.1 Deterministic Trend Models

The mean function of a nonstationary process could be represented by a deterministic trend of time. In such a case, a standard regression model might be used to describe the phenomenon. For example, if the mean function μ_t follows a linear trend, $\mu_t = \alpha_0 + \alpha_1 t$, then the deterministic linear trend model

$$Z_t = \alpha_0 + \alpha_1 t + a_t \tag{3.4.1}$$

with the a_t being a zero mean white noise series can be used. For a deterministic quadratic mean function, $\mu_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$, the model

$$Z_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + a_t \tag{3.4.2}$$

can be used. See Wei (2006), for the details.

3.4.1.2 Stochastic Trend Models and Differencing

If different parts of a time series behave very much alike except for their difference in the local mean levels, this kind of nonstationary behaviour is named as *homogeneous nonstationary* (Box & Jenkins, 1976). In this circumstance, integration can be a solution. Wei (2006) defines *integrated process* as follows:

> If a time series Z_t is nonstationary and its d^{th} difference, $\Delta^d Z_t = (1-B)^d Z_t$ is stationary and also can be represented as a stationary ARMA(p, q) process, then Z_t is said to follow an ARIMA(p, d, q) model. The Z_t in this case is referred to as an integrated process or series (Wei, 2006, p. 186).

In terms of the ARMA models, the process is nonstationary if some roots of its AR polynomial do not lie outside of the unit circle. The local behaviour of this kind of homogeneous nonstationary series is independent of its level. This kind of series can be reduced to a stationary series by taking a suitable difference of the series. In other words, the series $\{Z_t\}$ is nonstationary, but its d^{th} differenced series, $\{(1-B)^d Z_t\}$ for some integer $d \geq 1$, is stationary. Wei (2006) explains this kind of nonstationary series considering $d = 1$.

$$(1-B)Z_t = a_t \tag{3.4.3}$$

or

$$Z_t = Z_{t-1} + a_t\,. \tag{3.4.4}$$

The level of the series at time t is

$$\mu_t = Z_{t-1} \tag{3.4.5}$$

which is subject to the stochastic disturbance at time $(t-1)$. Wei (2006) states that for $d \geq 1$ the mean level of the process Z_t changes through time stochastically, and this process is characterized as having a stochastic trend.

Briefly, for the deterministic trend models, the mean level of the series is determined only by the time variable, t. On the other hand, for the stochastic trend models, the mean level of the series is determined by the previous observations; i.e., by Z_{t-1} for $d = 1$, by Z_{t-2} for $d = 2$, etc.

3.4.2 Autoregressive Integrated Moving Average (ARIMA) Models

In this section, two classes of ARIMA models which are useful in describing various homogeneous nonstationary time series are introduced.

3.4.2.1 The General ARIMA Model

The differenced series $(1-B)^d Z_t$ usually follows the general stationary ARMA(p,q) process in (3.3.9). Then, for integrated processes, instead of (3.3.9), the model below is used:

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t \qquad (3.4.6)$$

where the stationary AR operator $\phi_p(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$ and the invertible MA operator $\theta_q(B) = \left(1 - \theta_1 B - \cdots - \theta_q B^p\right)$ share no common factors.

The parameter θ_0 plays very different roles for $d = 0$ and $d > 0$. When $d = 0$, the original process is stationary, and... θ_0 is related to the mean of the process, i.e., $\theta_0 = \mu\left(1 - \phi_1 - \ldots - \phi_p\right)$. When $d \geq 1$, however, θ_0 is called the deterministic trend term and,... is often omitted from the model unless it is really needed (Wei, 2006, p. 72).

θ_0 is often omitted when $d \geq 1$, since for large t, it can become very dominating so that it forces the series to follow a deterministic pattern.

To show why the intercept term causes a deterministic trend, let consider the model

$$Z_t = \alpha + Z_{t-1} + a_t \tag{3.4.7}$$

or

$$Z_t - Z_{t-1} = \Delta Z_t = \alpha + a_t . \tag{3.4.8}$$

By successive substitutions, this process may be defined as

$$\begin{aligned} Z_t &= \alpha + Z_{t-1} + a_t = 4\alpha + Z_{t-4} + \sum_{i=1}^{4} a_{t-i+1} \\ &= t\alpha + Z_0 + \sum_{i=1}^{t} a_{t-i+1} \end{aligned} \tag{3.4.9}$$

since

$$Z_{t-1} = \alpha + Z_{t-2} + a_{t-1},\ Z_{t-2} = \alpha + Z_{t-3} + a_{t-2}, \ldots$$

In (3.4.9),

$$Deterministic\ trend\ term = t\alpha$$

$$Stochastic\ trend\ term = Z_0 + \sum_{i=1}^{t} a_{t-i+1} .$$

Enders (1948) explains the importance of knowing the form of the trend as follows:

> As the forecast horizon expands, the precise form of the trend becomes increasingly important. Stationarity implies the absence of a trend and long-run mean reversion. A deterministic trend implies steady increases (or decreases) into the infinite future. Forecasts of a series with a stochastic trend converge to a steady level... The nature of the trend may have important theoretical formulations (Enders, 1948, p. 261).

3.4.2.2 The Random Walk Model

In (3.4.6), if $p = 0$, $d = 1$, and $q = 0$, then we have the well-known *random walk* model, which is shown in the section 3.4.1.2.

$$(1 - B)Z_t = a_t \tag{3.4.10}$$

or

$$Z_t = Z_{t-1} + a_t . \tag{3.4.11}$$

This model has been widely used to describe the behaviour of the series of a stock price. This behaviour is similar to following a drunken man whose position at time t is his position at time $(t - 1)$ plus a step in a random direction at time t.

> The random walk model is the limiting process of the $AR(1)$ process $(1 - \phi B)Z_t = a_t$ with $\phi \to 1$. Because the autocorrelation function of the AR(1) process is $\rho_k = \phi^k$, as $\phi \to 1$, the random walk model phenomenon can be characterized by large, nonvanishing spikes in the sample ACF of the original series $\{Z_t\}$ and insignificant zero ACF for the differenced series $\{(1 - B)Z_t\}$ (Wei, 2006, p. 72).

When the random walk model has a nonzero constant term,

$$(1 - B)Z_t = \theta_0 + a_t \tag{3.4.12}$$

or

$$Z_t = Z_{t-1} + \theta_0 + a_t \tag{3.4.13}$$

it is named as *the random walk model with drift*.

3.4.3 Nonstationarity in the Variance and the Autocovariance

Differencing can be used to reduce a homogeneous nonstationary time series to a stationary time series. However, the nonstationarity of the heterogeneous series is not

because of their time-dependent means but because of their time-dependent variances and autocovariances. To reduce these types of nonstationarity, transformations are needed.

3.4.3.1 Variance and Autocovariance of the ARIMA Models

A process that is stationary in the mean is not necessarily stationary in the variance and the autocovariance. A process that is nonstationary in the mean, however, will also be nonstationary in the variance and the autocovariance.

In Wei (2006), it is shown that the ARIMA model whose mean function is time-dependent is also nonstationary in its variance and autocovariance functions. There, the following results are established:

1. The variance, $Var(Z_t)$, of the ARIMA process is time dependent, and $Var(Z_t) \neq Var(Z_{t-k})$ for $k \neq 0$.
2. The variance $Var(Z_t)$ is unbounded as $t \to \infty$.
3. The autocovariance $Cov(Z_{t-k}, Z_t)$ and the autocorrelation $Corr(Z_{t-k}, Z_t)$ of the process are also time dependent and hence are not invariant with respect to time translation. In other words, they are not only functions of the time difference *k* but are also functions of both time origin *t* and the original reference point n_0.
4. If *t* is large with respect to n_0, then... $Corr(Z_{t-k}, Z_t) \simeq 1$. Because $|Corr(Z_{t-k}, Z_t)| \leq 1$, it implies that the autocorrelation function vanishes slowly as *k* increases.

(Wei, 2006, p. 83).

3.4.3.2 Variance Stabilizing Transformations

Differencing is not an appropriate solution to transform the series which are stationary in the mean but are nonstationary in the variance to stationary series. It is

very common for the variance of a nonstationary process to change as its level changes. Thus,

$$Var(Z_t) = c\, f(\mu_t) \tag{3.4.14}$$

for some positive constant c and function f. Wei (2006) shows to find a function T so that the transformed series, $T(Z_t)$, has a constant variance. He approximates the desired function by a first-order Taylor series about the point μ_t as follows:

Let

$$T(Z_t) \simeq T(\mu_t) + T'(\mu_t)(Z_t - \mu_t) \tag{3.4.15}$$

where $T'(\mu_t)$ is the first derivative of $T(Z_t)$ evaluated at μ_t. Now

$$\begin{aligned} Var[T(Z_t)] &\simeq [T'(\mu_t)]^2 Var(Z_t) \\ &= c\,[T'(\mu_t)]^2 f(\mu_t) \end{aligned} \tag{3.4.16}$$

Thus, in order for the variance of $T(Z_t)$ to be constant, the variance stabilizing transformation $T(Z_t)$ must be chosen so that

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}} \tag{3.4.17}$$

Equation (3.4.17) implies that

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t \tag{3.4.18}$$

For example, if the standard deviation of a series is proportional to the level so that $Var(Z_t) = c\,\mu_t^2$, then

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t) \tag{3.4.19}$$

Hence, a logarithmic transformation (the base is irrelevant) of the series, $\ln(Z_t)$, will have a constant variance (Wei, 2006, p. 84).

More generally, the power transformation

$$T(Z_t) = \frac{Z_t^{\lambda} - 1}{\lambda} \tag{3.4.20}$$

introduced by Box and Cox (1964) can be used. Table 3.1 shows some commonly used values of λ and their associated transformations.

Table 3.1 Values of λ and their associated transformations

Values of λ (lambda)	**Transformation**
-1.0	$(Z_t)^{-1}$
-0.5	$(Z_t)^{-1/2}$
0.0	$\ln(Z_t)$
0.5	$(Z_t)^{1/2}$
1.0	Z_t (no transformation)

Some important remarks as regards differencing and variance stabilizing transformations are given in Wei (2006) as follows:

1. The variance stabilizing transformations introduced above are defined only for positive series. This definition, however, is not as restrictive as it seems because a constant can always be added to the series without affecting the correlation structure of the series.
2. A variance stabilizing transformation, if needed, should be performed before any other analysis such as differencing.
3. Frequently, the transformation not only stabilizes the variance, but also improves the approximation of the distribution by a normal distribution (Wei, 2006, p. 86).

3.5 Forecasting

Taking into account of the available observations, making predictions for the values which may be realize in the future is called as *forecasting*. Wei (2006) states that "forecasting is essential for planning and operation control in a variety of areas such as production management, inventory systems, quality control, financial planning, and investment analysis" (p. 88). Forecasting may be also considered as one of the most important goals of the time series analysis.

> In forecasting, our objective is to produce an optimum forecast that has no error or as little error as possible, which leads us to the minimum mean square error forecast. This forecast will produce an optimum future value with the minimum error in terms of the mean square error criterion (Wei, 2006, p. 88).

Wei (2006) gives details about the minimum mean square error forecasts for ARMA and ARIMA models. During the forecasting process, when new observations are obtained, they are taken into account of for updating the previous forecasts.

3.6 Model Identification

Wei (2006) defines that "model identification refers to the methodology in identifying the required transformations, such as variance stabilizing transformations and differencing transformations, the decision to include the deterministic parameter θ_0 when $d \geq 1$, and the proper orders of p and q for the model" (p. 108). The goal is to match patterns in the sample ACF and sample PACF, with the known patterns of the population ACF and population PACF. Steps of the model identification can be summarized as follows:

***Step* 1:** The time series data are plotted to check the series containing a trend, seasonality, outliers, nonconstant variances, and other nonnormal and nonstationary phenomena, and if it is necessary, the appropriate transformations are determined. As regards the transformations, Wei (2006) stresses that "because variance-stabilizing

transformations such as the power transformation require non-negative values and differencing may create some negative values, we should always apply variance-stabilizing transformations before taking differences" (p. 109).

***Step* 2:** If differencing the series is necessary, the sample ACF and the sample PACF are computed to determine the degree of differencing which makes the series stationary. For example, if the sample ACF decays very slowly, whereas the sample PACF cuts off after lag 1, it indicates the differencing is needed. In this condition, the unit root test may be proposed.

> Some authors argue that the consequences of unnecessary differencing are much less serious than those of under-differencing, but do beware of the artifacts created by over-differencing so that unnecessary over-parameterization can be avoided (Wei, 2006, p. 109)

Briefly, you should use more parameter than you need instead of less parameter than you need. However, it should be noted that unnecessary over-parameterization decreases both the degrees of freedom and the efficiency of the estimators.

***Step* 3:** After the series are transformed properly to be stationary, you should compute the sample ACF and PACF to identify the orders of p and q. In practice, the needed orders are usually less than or equal to 3. Table 3.2 summarizes the important results for selecting p and q.

> To identify a reasonably appropriate ARIMA model, ideally, we need a minimum of $n = 50$ observations, and the number of sample lag-k autocorrelations and partial autocorrelations to be calculated should be about $n/4$ (Wei, 2006, p. 109).

***Step* 4:** When $d > 0$, you should test whether or not the deterministic trend term, θ_0, is necessary. Usually, θ_0 is included in the initial model, and if the preliminary estimation result is not significant, it is discarded at the final model estimation.

For the theoretical details and the empirical examples, see Wei (2006).

Table 3.2 Characteristics of theoretical ACF and PACF for stationary processes

Process	ACF	PACF
$AR(p)$	Tails off exponential decay or damped sine wave	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Tails off exponential decay or damped sine wave
$ARMA(p,q)$	Tails off after lag$(q-p)$	Tails off after lag$(p-q)$

3.7 Parameter Estimation, Diagnostic Checking, and Model Selection Criteria

3.7.1 Parameter Estimation

In time series analysis, several methods are used to estimate the parameters. In this thesis, some of them are mentioned briefly.

Method of moments estimators are found by equating the sample moments to the corresponding population moments, and solving the resulting system of simultaneous equations. These estimators are usually called Yule-Walker estimators. Wei (2006) points out that the moment estimators are not recommended for final estimation results and should not be used if the process is close to being nonstationary or noninvertible.

The Maximum Likelihood Estimator (MLE) is the one for which the probability of observing the corresponding sample is maximum. In general, the MLE is a good point estimator, possessing some of the optimality properties.

The ordinary least squares (OLS) estimation developed for linear regression models can also be used in time series analysis. However, Wei (2006) points out that the OLS estimator for the parameter of an explanatory variable in a regression model

will be inconsistent unless the error term is uncorrelated with the explanatory variable. For $ARMA(p,q)$models, this condition usually does not hold except when $q = 0$. Hence, the other estimation methods discussed above are more efficient and commonly used in time series analysis.

Chang & Park (2000) state that "the Yule-Walker method may be preferred to the OLS method in small samples, since it always yields an invertible autoregression; see, for example, Brockwell & Davis (1991, section 8.1, 8.2)" (p. 390).

3.7.2 Diagnostic Checking

After parameter estimation, whether the model assumptions are satisfied or not have to be checked. This matter is explained in Wei (2006) as follows:

> To check whether the errors are normally distributed, one can construct a histogram of the standardized residuals $\hat{a}_t/\hat{\sigma}_a$ and compare it with the standard normal distribution using the chi-square goodness of fit test or even Tukey's simple five-number summary. To check whether the variance is constant, we can examine the plot of residuals. To check whether the residuals are approximately white noise, we compute the sample ACF and sample PACF (or IACF) of the residuals to see whether they do not form any pattern and are all statistically insignificant, i.e., within two standard deviations if $\alpha = 0.05$ (Wei, 2006, p. 152-153).

3.7.3 Model Selection

> In time series analysis, for a given data set, when there are multiple adequate models, the selection criterion is normally based on summary statistics from residuals computed from a fitted model or on forecast errors calculated from the out-sample forecasts. The latter is often accomplished by using the first portion of the series for model construction and the remaining portion as a holdout period for forecast evaluation (Wei, 2006, p. 156).

Below, some model selection criteria based on residuals are introduced briefly.

3.7.3.1 Akaike's AIC

Akaike (1973,1974) introduced an information criterion to evaluate the quality of the model building. The criterion has been called AIC (Akaike's information criterion) in the literature and is defined as

$$AIC(M) = -2\ln[maximum\ likelihood] + 2M \qquad (3.7.1)$$

where M is the number of parameters in the model. For the ARMA model and n effective number of observations, it is showed in Wei (2006) that the AIC criterion reduces to

$$AIC(M) = n\ln\hat{\sigma}_a^2 + 2M\,. \qquad (3.7.2)$$

The optimal order of the model is chosen by the value of M, which is a function of p and q, so that $AIC(M)$ is the minimum.

Shibata (1976) has shown that the AIC tends to overestimate the order of the autoregression.

3.7.3.2 Akaike's BIC

Akaike (1978, 1979) has developed a Bayesian extension of the minimum AIC procedure, called the Bayesian information criterion (BIC), which takes the form

$$BIC(M) = n\ln\hat{\sigma}_a^2 - (n-M)\ln\left(1-\frac{M}{n}\right) + M\ln n + M\ln\left[\left(\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2} - 1\right)/M\right] \qquad (3.7.3)$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, M is the number of parameters, and $\hat{\sigma}_z^2$ is the sample variance of the series.

Through a simulation study Akaike (1978) has claimed that the BIC is less likely to overestimate the order of the autoregression.

Chang & Park (2000) gives information about the usage of AIC and BIC as follows:

> If it is known that the true model is generated by a finite order autoregression, the order selection based on BIC is consistent, and therefore, it might be preferred. Such a case, however, is rare in practical applications. True model is unknown, and not likely to be given exactly by a finite order autoregression. We may thus use AIC, in favour of BIC, since it leads to asymptotically efficient choice of the optimal order for a class of infinite order autoregressive processes (Brockwell & Davis, 1991).... Using BIC instead of AIC generally gives higher rejection probabilities under both the null and alternative hypotheses. A reversed tendency has been observed when we increase the number of maximum lag length. The use of AIC with no restriction on the maximum lag length yields the lowest rejection probabilities. The highest rejection probabilities are observed with the application of BIC with smallest maximum lag length. However, the choice of the selection criteria and the maximum lag length do not seem to affect the discriminatory powers of the tests. Their affects are rather uniform regardless of the presence or absence of the unit root (Chang & Park, 2000, p. 390-391).

3.8 Unit Root Processes

3.8.1 Definition and Importance of Unit Root Tests

The unit root hypothesis has drawn much attention for the past three decades, especially in economics and other related fields. Chang & Park (2000) point out that "the hypothesis has an important implication on, in particular, whether or not the shocks to an economic system have a permanent effect on the future path of the economy" (p. 379). It is known that many of important economic and financial time series display unit root characteristics. Phillips & Perron (1988) explain the importance of the unit root tests in economy as follows:

> One major field of application where the hypothesis of a unit root has important implications is economics. This is because a unit root is often a theoretical implication of models which postulate the rational use of information that is available to economic agents.... Formal statistical tests of the unit root hypothesis are of additional interest to economists because they can help to evaluate the nature of the nonstationarity that most macroeconomic data exhibit. In particular, they help in determining whether the trend is stochastic, through the presence of a unit root, or deterministic, through the presence of a polynomial time trend (Phillips & Perron, 1988, p. 335).

The tests by Dickey & Fuller (1979, 1981) are most commonly used. However, Chang & Park (2000) state that "the tests by Said-Dickey and Phillips-Perron are often preferred to the Dickey-Fuller tests in practical applications, since they do not require any particular parametric specification and yet are applicable for a wide class of unit root models" (p. 380). The disadvantage of all these tests is to have considerable size distortions in finite samples. Therefore, in this thesis whether the bootstrap method can improve their finite sample performance or not is investigated.

In Section 3.4 it is mentioned that a nonstationary time series can often be reduced to a stationary time series by differencing. The question of whether a series should be differenced is equal to the question of whether a series has a unit root. First of all it is useful to summarize the basic properties of both a covariance stationary series and a nonstationary series. These properties are given in Enders (1948) as follows:

We know that a covariance stationary series:

1. Exhibits mean reversion in that it fluctuates around a constant long-run mean.
2. Has a finite variance that is time-invariant.
3. Has a theoretical correlogram that diminishes as lag length increases.

.... To aid in the identification of a nonstationary series, we know that:

1. There is no long-run mean to which the series returns.
2. The variance is time-dependent and goes to infinity as time approaches infinity.
3. Theoretical autocorrelations do not decay but, in finite samples, the sample correlogram dies out slowly (Enders, 1948, p. 212).

Let consider a series is generated from the following first-order process:

$$Z_t = \phi_1 Z_{t-1} + a_t \tag{3.8.1}$$

where $\{a_t\}$ is generated from a white-noise process.

Why the usual *t*-test cannot be used to test the unit root hypothesis is explained in Enders (1948). He firstly considers to test $H_0: \phi_1 = 0$ and $H_1: \phi_1 \neq 0$. Under the null hypothesis, (3.8.1) can be estimated using OLS. The fact that a_t is a white-noise process and $|\phi_1| < 1$ guarantee that the $\{Z_t\}$ sequence is stationary and the estimate of the coefficient is efficient. Calculating the standard error of the estimate of coefficient, the researcher can use a *t*-test to determine whether the coefficient is significantly different from zero. However, if the hypotheses are established as $H_0: \phi_1 = 1$ and $H_1: |\phi_1| < 1$, under the null hypothesis, the $\{Z_t\}$ sequence is generated by the nonstationary process

$$Z_t = \sum_{i=1}^{t} a_i \,. \tag{3.8.2}$$

Now, under the null hypothesis, the variance becomes infinitely large as t increases. Then, it is inappropriate to use classical statistical methods to estimate and perform significance tests on the coefficient. The process returns to a random walk process. The first-order autocorrelation coefficient in a random walk model is

$$\rho_1 = [(t-1)/t]^{0.5} < 1 \,.$$

Enders (1948) states that "since the estimate of the coefficient is directly related to the value of ρ_1, the estimated value of the coefficient is biased to be below its true value of unity" (p. 213). Then, the estimated model will mimic that of a stationary AR(1) process with a near unit root. In this condition, the usual *t*-test cannot be used to test H_0: $\phi_1 = 1$. Dickey and Fuller (1979, 1981) devised a procedure to formally test for the presence of a unit root.

3.8.2 Unit Roots in a Regression Model

Enders (1948) explains the unit root situation for a regression model considering the regression equation below

$$y_t = a_0 + a_1 z_t + e_t \,. \tag{3.8.3}$$

Under the assumptions of the classical regression model, both the $\{y_t\}$ and $\{z_t\}$ sequences are stationary and the errors have a zero mean and finite variance. However, in the presence of nonstationary variables, there might be **spurious regression** as called by Granger and Newbold (1974). A spurious regression has a high R^2, *t*-statistics that appear to be significant but the results are without any economic meaning. The regression output "look good" because the least-squares estimates are not consistent and the customary tests of statistical inference do not hold. Granger & Newbold (1974) provide a detailed examination to survey the

consequences of violating the stationarity assumption by generating two sequences, $\{y_t\}$ and $\{z_t\}$, as *independent* random walks using the formulas:

$$y_t = y_{t-1} + \epsilon_{yt}$$

and

$$z_t = z_{t-1} + \epsilon_{zt}$$

where both ϵ_{yt} and ϵ_{zt} are white-noise processes independent of each other.

Surprisingly, while the Equation (3.8.3) is necessarily meaningless since both processes are independent random walks, and any relationship between two variables is spurious, Granger & Newbold (1974) have rejected the true null hypothesis, that is $a_1 = 0$, in approximately 75% of the time compared to true nominal 5% significance level. Moreover, the regressions usually had very high R^2 values and the estimated residuals exhibited a high degree of autocorrelation. This result is meaningless since the assumption that the error term is a unit root process is inconsistent with the distributional theory underlying the use of OLS. This problem will not disappear in large samples. To see the details, see Granger & Newbold (1974) or Enders (1948).

Enders (1948) gives the four cases for the econometrician to be careful in working with nonstationary variables as follows:

- **Case1:** Both $\{y_t\}$ and $\{z_t\}$ are stationary. When both variables are stationary, the classical regression model is appropriate.
- **Case2:** The $\{y_t\}$ and $\{z_t\}$ sequences are integrated of different orders. Regression equations using such variables are meaningless....
- **Case3:** The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order and the residual sequence contains a stochastic term. This is the case in which the regression is spurious. The results from such spurious regressions are meaningless in that all errors are permanent. In this case, it is often recommended that the regression equation be estimated in first differences....

- **Case4:** The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order and the residual sequence is stationary. In this circumstance, $\{y_t\}$ and $\{z_t\}$ are **cointegrated** (Enders, 1948, p. 219).

3.8.3 Some Useful Limiting Distributions

For the simple $AR(1)$ model

$$Z_t = \phi Z_{t-1} + a_t \tag{3.8.4}$$

with $t = 1, \dots, n$ and $Z_0 = 0$, where the a_t is the Gaussian $N(0, \sigma_a^2)$ white noise process, the unit root process implies a test for a random walk model. The alternative is that the series is stationary. That is $H_0: \phi = 1$ and $H_1: |\phi| < 1$.

Wei (2006) gives the test statistic formula connected with the OLS estimation as follows:

$$\hat{\phi} = \frac{\sum_{t=1}^{n} Z_{t-1} Z_t}{\sum_{t=1}^{n} Z_{t-1}^2}. \tag{3.8.5}$$

Under the hypothesis $H_0: \phi = 1$, $Z_t = Z_{t-1} + a_t$,

$$\hat{\phi} = \frac{\sum_{t=1}^{n} Z_{t-1} Z_t}{\sum_{t=1}^{n} Z_{t-1}^2} = \frac{\sum_{t=1}^{n} Z_{t-1}(Z_{t-1} + a_t)}{\sum_{t=1}^{n} Z_{t-1}^2} = \frac{\sum_{t=1}^{n} Z_{t-1}^2 + \sum_{t=1}^{n} Z_{t-1} a_t}{\sum_{t=1}^{n} Z_{t-1}^2}$$

$$\hat{\phi} = 1 + \frac{\sum_{t=1}^{n} Z_{t-1} a_t}{\sum_{t=1}^{n} Z_{t-1}^2} \Rightarrow \hat{\phi} - 1 = \frac{\sum_{t=1}^{n} Z_{t-1} a_t}{\sum_{t=1}^{n} Z_{t-1}^2}.$$

Thus, multiplying the equation above with $n^{-1}/n^{-2} = n$,

$$n(\hat{\phi} - 1) = \frac{n^{-1} \sum_{t=1}^{n} Z_{t-1} a_t}{n^{-2} \sum_{t=1}^{n} Z_{t-1}^2} \tag{3.8.6}$$

$$\xrightarrow{D} \frac{\frac{1}{2}\sigma_a^2\{[W(1)]^2 - 1\}}{\sigma_a^2 \int_0^1 [W(x)]^2\, dx} = \frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\int_0^1 [W(x)]^2\, dx}. \qquad (3.8.7)$$

Ferretti & Romo (1996) states that "the limit distribution of $\hat{\phi}$ is different for the three possible cases: stationary, unstable, and explosive; it is normal for the stationary case and nonnormal for the two nonstationary cases" (p. 849). For instance, in the unstable case, $\phi = 1$, it is known that

$$T = \frac{(\hat{\phi}_n - 1)(\sum_{t=1}^n Z_{t-1}^2)^{1/2}}{(\hat{\sigma}_a^2)^{1/2}}$$

$$\xrightarrow{D} \frac{\frac{1}{2}\sigma_a^2\{[W(1)]^2 - 1\}}{\left\{\sigma_a^2 \int_0^1 [W(x)]^2 dx\right\}^{1/2} (\sigma_a^2)^{1/2}} = \frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int_0^1 [W(x)]^2 dx\right\}^{1/2}}. \qquad (3.8.8)$$

In this section, it is shown that this test statistic converges weakly to the limit distribution above as $n \to \infty$, where $\{W(t)\}$ is the standard Brownian motion on [0, 1]. The same approach is used with Wei (2006) who follows the approach of Chan & Wei (1988) for developing a desired test statistic. The theoretical approaches are given only for AR(1) model without a constant term. Wei (2006) gives the Wiener process and its connection with the test statistic of Dickey-Fuller as follows:

A process,$W(t)$, is continuous if its time index t belongs to an interval of a real line. For distinction, we often write a continuous process as $W(t)$ rather than W_t. A process $W(t)$ is said to be a Wiener process (also known as Brownian motion process) if it contains the following properties:

1. $W(0) = 0$;
2. $E[W(t)] = 0$;
3. $W(t)$ follows a nondegenerate normal distribution for each t; and

4. $W(t)$ has independent increments, i.e. $[W(t_2) - W(t_1)]$ and $[W(t_4) - W(t_3)]$ are independent for any nonoverlapping time intervals (t_1, t_2) and (t_3, t_4).

With no loss of generality, we consider t in the closed interval between 0 and 1, i.e., $t \in [0,1]$. Furthermore, if for any t, $W(t)$ is distributed as $N(0,t)$, then the process is also called standard Brownian motion.

Given the i.i.d. random variables a_t, for $t = 1, \ldots, n$, with mean 0 and variance σ_a^2, define

$$F_n(x) = \begin{cases} 0 & \text{for } 0 \le x < 1/n \,, \\ a_1/\sqrt{n} & \text{for } 1/n \le x < 2/n \,, \\ (a_1 + a_2)/\sqrt{n} & \text{for } 2/n \le x < 3/n \,, \\ \quad . & \\ \quad . & \\ \quad . & \\ (a_1 + a_2 + \ldots + a_n)/\sqrt{n} & \text{for } x = 1 \,. \end{cases}$$

That is,

$$F_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[xn]} a_t \tag{3.8.9}$$

where $x \in [0,1]$ and $[xn]$ represents the integer portion of (xn). Now,

$$F_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[xn]} a_t = \frac{\sqrt{[xn]}}{\sqrt{n}} \frac{1}{\sqrt{[xn]}} \sum_{t=1}^{[xn]} a_t \,. \tag{3.8.10}$$

Because, as $n \to \infty, \left(\sqrt{[xn]}/\sqrt{n}\right) \longrightarrow \sqrt{x}$ and $\sum_{t=1}^{[xn]} a_t/\sqrt{[xn]}$ converges to a $N(0, \sigma_a^2)$ random variable by the central limit theorem, it follows that $F_n(x)$ converges in distribution to $\sqrt{x}\, N(0, \sigma_a^2) = N(0, x\sigma_a^2)$ as $n \to \infty$, which will be

denoted by $F_n(x) \xrightarrow{D} N(0, x\sigma_a^2)$. In the following discussion, we also use the notation $X_n \xrightarrow{P} X$ to indicate the convergence of X_n to X in probability as $n \to \infty$.

It can be easily seen that the limit of the sequence of the random variable $F_n(x)/\sigma_a$ can be described by a Wiener process, i.e.,

$$\frac{F_n(x)}{\sigma_a} \xrightarrow{D} W(x) \tag{3.8.11}$$

or

$$F_n(x) \xrightarrow{D} \sigma_a W(x) \tag{3.8.12}$$

where $W(x)$ at time $t = x$ follows an $N(0, x)$. Specifically,

$$F_n(1) = \sum_{i=1}^{n} \frac{a_t}{\sqrt{n}} \xrightarrow{D} \sigma_a W(1) \tag{3.8.13}$$

where $W(1)$ follows an $N(0,1)$.

Let $Z_t = a_1 + a_2 + \ldots + a_t$ and $Z_0 = 0$. We can rewrite $F_n(x)$... as

$$F_n(x) = \begin{cases} 0 & \text{for } 0 \le x < 1/n\ , \\ Z_1/\sqrt{n} & \text{for } 1/n \le x < 2/n\ , \\ Z_2/\sqrt{n} & \text{for } 2/n \le x < 3/n\ , \\ \cdot & \\ \cdot & \\ \cdot & \\ Z_n/\sqrt{n} & \text{for } x = 1\ . \end{cases} \tag{3.8.14}$$

Then...

$$\frac{Z_n}{\sqrt{n}} \xrightarrow{D} \sigma_a W(1) \tag{3.8.15}$$

or

$$\frac{Z_n^2}{n} \xrightarrow{D} \sigma_a^2 [W(1)]^2 \ . \tag{3.8.16}$$

Also, it is clear that the integral $\int_0^1 F_n(x)\, dx$ is simply the sum of the area...

$$\int_0^1 F_n(x)\, dx \ = \ \frac{1}{n}\frac{Z_1}{\sqrt{n}} \ + \ \frac{1}{n}\frac{Z_2}{\sqrt{n}} + \ \ldots + \ \frac{1}{n}\frac{Z_{n-1}}{\sqrt{n}} \ = \ n^{-3/2} \sum_{t=1}^{n} Z_{t-1} \, .$$

Thus...

$$n^{-3/2} \sum_{t=1}^{n} Z_{t-1} \ = \ \int_0^1 F_n(x)\, dx \xrightarrow{D} \sigma_a \int_0^1 W(x)\, dx \ . \tag{3.8.17}$$

Similarly,

$$\int_0^1 [F_n(x)]^2\, dx \ = \ \frac{1}{n}\left(\frac{Z_1}{\sqrt{n}}\right)^2 \ + \ \frac{1}{n}\left(\frac{Z_2}{\sqrt{n}}\right)^2 \ + \ \ldots \ + \ \frac{1}{n}\left(\frac{Z_{n-1}}{\sqrt{n}}\right)^2 \ = \ n^{-2} \sum_{t=1}^{n} Z_{t-1}^2$$

and

$$n^{-2} \sum_{t=1}^{n} Z_{t-1}^2 \ = \ \int_0^1 [F_n(x)]^2\, dx \xrightarrow{D} \sigma_a^2 \int_0^1 [W(x)]^2\, dx \ . \tag{3.8.18}$$

Next,

$$Z_t^2 \ = \ (Z_{t-1} \ + \ a_t)^2 \ = \ Z_{t-1}^2 \ + \ 2Z_{t-1}a_t \ + \ a_t^2$$

$$Z_{t-1}a_t \ = \ \frac{1}{2}[Z_t^2 \ - \ Z_{t-1}^2 \ - \ a_t^2]$$

and summing from 1 to n gives

$$\sum_{t=1}^{n} Z_{t-1}a_t \ = \ \frac{1}{2}[Z_n^2 \ - \ Z_0^2] \ - \ \frac{1}{2}\sum_{t=1}^{n} a_t^2 \ = \ \frac{1}{2}Z_n^2 \ - \ \frac{1}{2}\sum_{t=1}^{n} a_t^2 \, .$$

Therefore,

$$n^{-1}\sum_{t=1}^{n} Z_{t-1}a_t \;=\; \frac{1}{2}\left[\frac{Z_n^2}{n}\right] \;-\; \frac{1}{2}\left[\sum_{t=1}^{n}\frac{a_t^2}{n}\right]$$

$$\xrightarrow{D} \frac{1}{2}\sigma_a^2[W(1)]^2 \;-\; \frac{1}{2}\sigma_a^2 \;=\; \frac{1}{2}\sigma_a^2\{[W(1)]^2 - 1\} \qquad (3.8.19)$$

which follows from... that $[\sum_{t=1}^{n} a_t^2/n] \xrightarrow{P} \sigma_a^2$ (Wei, 2006, p. 186-189).

Now (3.8.6) converges in distribution to (3.8.7) following (3.8.18),(3.8.19) and that under H_0, (3.8.4) becomes a random walk model that can be written as

$$Z_t \;=\; a_1 + a_2 + \ldots + a_t\,. \qquad (3.8.20)$$

Proof for $\sum_{t=1}^{n} Z_{t-1}a_t \;=\; \frac{1}{2}[Z_n^2 \;-\; Z_0^2] \;-\; \frac{1}{2}\sum_{t=1}^{n} a_t^2 \;=\; \frac{1}{2}Z_n^2 \;-\; \frac{1}{2}\sum_{t=1}^{n} a_t^2$

$$\sum_{t=1}^{n} Z_{t-1}a_t \;=\; \frac{1}{2}\sum_{t=1}^{n} Z_t^2 \;-\; \frac{1}{2}\sum_{t=1}^{n} Z_{t-1}^2 - \frac{1}{2}\sum_{t=1}^{n} a_t^2$$

$$\sum_{t=1}^{n} Z_{t-1}a_t \;=\; \frac{1}{2}\left[\sum_{t=1}^{n} Z_t^2 \;-\; \sum_{t=1}^{n} Z_{t-1}^2\right] - \frac{1}{2}\sum_{t=1}^{n} a_t^2$$

$$\sum_{t=1}^{n} Z_{t-1}a_t \;=\; \frac{1}{2}[(Z_1^2 \;+\; Z_2^2 \;+\; \ldots + \; Z_n^2) - (Z_0^2 \;+\; Z_1^2 \;+\; \ldots + \; Z_{n-1}^2)] - \frac{1}{2}\sum_{t=1}^{n} a_t^2$$

$$\sum_{t=1}^{n} Z_{t-1}a_t \;=\; \frac{1}{2}[Z_n^2 \;-\; Z_0^2] \;-\; \frac{1}{2}\sum_{t=1}^{n} a_t^2 \;=\; \frac{1}{2}Z_n^2 \;-\; \frac{1}{2}\sum_{t=1}^{n} a_t^2$$

where $Z_0 \;= 0$.

It is known that the normal and the *t*-distributions cannot be used when the parameter value equals to 1. Wei (2006) also explains the reason why the limiting distribution of the test statistic above is skewed to the left as follows:

> $W(1)$ is known to be an $N(0,1)$ random variable. Hence $[W(1)]^2$ follows the chi-square distribution with one degree of freedom, i.e., $\chi^2(1)$. The probability that a $\chi^2(1)$ random variable is less than 1 is .6827. Because the denominator is always positive, the probability that $n(\hat{\phi} - 1) < 0$ approaches .6827 as n becomes large. The OLS estimator $\hat{\phi}$ clearly underestimates the true value in this case, and the limiting distribution of $n(\hat{\phi} - 1)$ is clearly skewed to the left. As a result, the null hypothesis is rejected only when $n(\hat{\phi} - 1)$ is really too negative, i.e., much less than the rejection limit when the normal or the *t*-distribution is used (Wei, 2006, p. 190).

The percentiles for the empirical distribution of $n(\hat{\phi} - 1)$ in (3.8.20) were constructed by Dickey (1976) using the Monte Carlo method and reported in Fuller (1996, p. 641).

The commonly used *t*-statistic under H_0 is

$$T = \frac{\hat{\phi} - 1}{S_{\hat{\phi}}} = \frac{\hat{\phi} - 1}{[\hat{\sigma}_a^2 (\sum_{t=1}^{n} Z_{t-1}^2)^{-1}]^{1/2}} \tag{3.8.21}$$

where

$$S_{\hat{\phi}} = \left[\hat{\sigma}_a^2 \left(\sum_{t=1}^{n} Z_{t-1}^2 \right)^{-1} \right]^{1/2} \text{ and } \hat{\sigma}_a^2 = \sum_{t=1}^{n} \frac{\left(Z_t - \hat{\phi} Z_{t-1} \right)^2}{(n-1)}.$$

Now, the equation (3.8.6) can be rewritten as follows:

$$n(\hat{\phi} - 1) \left[n^{-2} \sum_{t=1}^{n} Z_{t-1}^2 \right]^{1/2} \left[n^{-2} \sum_{t=1}^{n} Z_{t-1}^2 \right]^{1/2} = n^{-1} \sum_{t=1}^{n} Z_{t-1} a_t$$

$$n(\hat{\phi} - 1)\left[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2\right]^{1/2} = (\hat{\phi} - 1)\left[\sum_{t=1}^{n} Z_{t-1}^2\right]^{1/2} = \frac{n^{-1}\sum_{t=1}^{n} Z_{t-1}a_t}{[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2]^{1/2}}.$$

Hence,

$$T = \frac{\hat{\phi} - 1}{(\hat{\sigma}_a^2)^{1/2}[\sum_{t=1}^{n} Z_{t-1}^2]^{-1/2}} = \frac{(\hat{\phi} - 1)[\sum_{t=1}^{n} Z_{t-1}^2]^{1/2}}{(\hat{\sigma}_a^2)^{1/2}}$$

$$T = \frac{n^{-1}\sum_{t=1}^{n} Z_{t-1}a_t}{[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2]^{1/2}(\hat{\sigma}_a^2)^{1/2}} \tag{3.8.22}$$

$$\xrightarrow{D} \frac{\frac{1}{2}\sigma_a^2\{[W(1)]^2 - 1\}}{\left\{\sigma_a^2\int_0^1[W(x)]^2\,dx\right\}^{1/2}(\sigma_a^2)^{1/2}} \quad \frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int_0^1[W(x)]^2\,dx\right\}^{1/2}} \tag{3.8.23}$$

which follows from (3.8.18), (3.8.19) and that $\hat{\sigma}_a^2$ is a consistent estimator of σ_a^2.

Wei (2006) states that "the nature of the distribution of T is similar to that of $n(\hat{\phi} - 1)$; we reject H_0 if T is too large negatively" (p. 191). The percentiles for the empirical distribution of T were also constructed by Dickey (1976) using the Monte Carlo method and reported in Fuller (1996, p. 642).

The use of the assumed initial value $Z_0 = 0$ is purely for the convenience of deriving a limiting distribution; in actual data analysis, we normally use only real data points are used; hence (3.8.18) is computed only for $t = 2, \ldots, n$ (Wei, 2006, p. 191)

3.8.4 Dickey-Fuller Tests

Dickey and Fuller (1979) consider three different regression equations that can be used to test for the presence of a unit root:

$$\Delta Z_t = \gamma Z_{t-1} + a_t \tag{3.8.24}$$

$$\Delta Z_t = \alpha + \gamma Z_{t-1} + a_t \tag{3.8.25}$$

$$\Delta Z_t = \alpha + \delta t + \gamma Z_{t-1} + a_t \tag{3.8.26}$$

where $\gamma = \phi - 1$. Hence, testing the hypothesis $\phi = 1$ is equivalent to testing the hypothesis $\gamma = 0$.

The first regression model is a pure random walk model, the second adds an *intercept* or *drift* or *constant* term, and the third includes not only a drift but also a linear time trend.

The parameter of interest in all the regression equations above isγ. Having $\gamma = 0$ means that the $\{Z_t\}$ sequence contains a unit root. To obtain the estimated value of γ and associated standard error, one (or more) of the equations above are estimated using OLS. Using the formula which was given in the previous section, the T value is calculated. This value is compared with the appropriate critical value reported in the Dickey-Fuller tables. If the calculated T value is less than the critical value, that is more negative than the critical value, H_0 is rejected. Otherwise, it is understood that this process is a unit root process. This methodology and the decision mechanism is precisely the same for all forms of the regression equations given above. It should be emphasized that the critical values here do not depend on whether an intercept and/or time trend is included in the regression equation. In their Monte Carlo study, Dickey and Fuller (1979) found that the critical values for $\gamma = 0$ depend on the form of the regression and sample size. As in most hypothesis tests, for any given level of significance, the critical values of the T-statistic decrease as sample size increases. These critical values are unchanged if (3.8.24), (3.8.25), and (3.8.26) are replaced by the following autoregressive processes:

$$\Delta Z_t = \gamma Z_{t-1} + \sum_{i=2}^{p} \beta_i \, \Delta Z_{t-i+1} + a_t \tag{3.8.27}$$

$$\Delta Z_t = \alpha + \gamma Z_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t \qquad (3.8.28)$$

$$\Delta Z_t = \alpha + \delta t + \gamma Z_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t . \qquad (3.8.29)$$

See Enders (1948) for the details. See also Dickey and Fuller (1981) for the additional *F*-statistics to test **joint** hypotheses on the coefficients.

The intuition behind the test is as follows. If the series Z_t is stationary (or trend stationary), then it has a tendency to return to a constant (or deterministically trending) mean. Therefore large values will tend to be followed by smaller values (negative changes), and small values by larger values (positive changes). Accordingly, the level of the series will be a significant predictor of next period's change, and will have a negative coefficient. If, on the other hand, the series is integrated, then positive changes and negative changes will occur with probabilities that do not depend on the current level of the series; in a random walk, where you are now does not affect which way you will go next (Wikipedia (a)).

The extensions of the Dickey-Fuller tests which are named as Augmented Dickey-Fuller (ADF) Tests uses the same procedure, and remove all the structural affects (autocorrelation) in the time series.

3.8.5 Extensions of the Dickey-Fuller Tests

Since it is not possible to represent all time-series processes by the first-order autoregressive process, it is possible to extend the Dickey and Fuller tests for higher-order equations such as (3.8.27), (3.8.28), and (3.8.29). In this thesis, the extension is showed for the *p*th-order autoregressive process which is also considered in Enders (1948).

The p^{th} order autoregressive process is defined as,

$$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_{p-2} Z_{t-p+2} + \phi_{p-1} Z_{t-p+1} + \phi_p Z_{t-p} + a_t . \quad (3.8.30)$$

To understand the methodology of the **augmented Dickey-Fuller** test, add the term $(\phi_p Z_{t-p+1} - \phi_p Z_{t-p+1})$ into the equation above:

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + \underline{\phi_{p-1} Z_{t-p+1} + \phi_p Z_{t-p+1}} + \underline{\phi_p Z_{t-p} - \phi_p Z_{t-p+1}} + a_t$$

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + (\phi_{p-1} + \phi_p) Z_{t-p+1} - \phi_p (Z_{t-p+1} - Z_{t-p}) + a_t$$

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + \phi_{p-2} Z_{t-p+2} + (\phi_{p-1} + \phi_p) Z_{t-p+1} - \phi_p \Delta Z_{t-p+1} + a_t .$$

Next, add the term $[(\phi_{p-1} + \phi_p) Z_{t-p+2} - (\phi_{p-1} + \phi_p) Z_{t-p+2}]$ to obtain

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + \underline{\phi_{p-2} Z_{t-p+2} + (\phi_{p-1} + \phi_p) Z_{t-p+2}} + \underline{(\phi_{p-1} + \phi_p) Z_{t-p+1} - (\phi_{p-1} + \phi_p) Z_{t-p+2}} - \phi_p \Delta Z_{t-p+1} + a_t$$

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + (\phi_{p-2} + \phi_{p-1} + \phi_p) Z_{t-p+2} - (\phi_{p-1} + \phi_p)(Z_{t-p+2} - Z_{t-p+1}) - \phi_p \Delta Z_{t-p+1} + a_t$$

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + (\phi_{p-2} + \phi_{p-1} + \phi_p) Z_{t-p+2} - (\phi_{p-1} + \phi_p) \Delta Z_{t-p+2} - \phi_p \Delta Z_{t-p+1} + a_t .$$

In the third step, the equation below is obtained

$$Z_t = \alpha + \phi_1 Z_{t-1} + \ldots + (\phi_{p-3} + \phi_{p-2} + \phi_{p-1} + \phi_p) Z_{t-p+3} - (\phi_{p-2} + \phi_{p-1} + \phi_p) \Delta Z_{t-p+3} - (\phi_{p-1} + \phi_p) \Delta Z_{t-p+2} - \phi_p \Delta Z_{t-p+1} + a_t .$$

At the end of the step of $(p-1)$, the equation below is obtained

$$Z_t = \alpha + (\phi_1 + \phi_2 + \ldots + \phi_p) Z_{t-1} - (\phi_2 + \phi_3 + \ldots + \phi_p) \Delta Z_{t-1} - (\phi_3 + \phi_4 + \ldots + \phi_p) \Delta Z_{t-2} - \ldots - (\phi_{p-1} + \phi_p) \Delta Z_{t-p+2} - \phi_p \Delta Z_{t-p+1} + a_t$$

$$Z_t = \alpha + \sum_{i=1}^{p} \phi_i Z_{t-1} - \sum_{i=2}^{p} \phi_i \Delta Z_{t-1} - \sum_{i=3}^{p} \phi_i \Delta Z_{t-2} - \ldots - \sum_{i=p-1}^{p} \phi_i \Delta Z_{t-p+2} - \phi_p \Delta Z_{t-p+1} + a_t$$

$$Z_t = \alpha + \sum_{i=1}^{p} \phi_i Z_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t$$

where

$$\beta_i = -\sum_{j=1}^{p} \phi_j .$$

Next, add and subtract Z_{t-1} to obtain

$$Z_t - Z_{t-1} = \alpha - Z_{t-1} + \sum_{i=1}^{p} \phi_i Z_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t$$

$$\Delta Z_t = \alpha + Z_{t-1} \left(\sum_{i=1}^{p} \phi_i - 1 \right) + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t$$

$$\Delta Z_t = \alpha + \gamma Z_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t \qquad (3.8.31)$$

where

$$\gamma = \sum_{i=1}^{p} \phi_i - 1 .$$

In (3.8.31), the coefficient of interest is γ; if $\gamma = 0$, this process has a unit root.

The intuition behind the test is that if the series is integrated then the lagged level of the series (Z_{t-1}) will provide no relevant information in predicting the change in Z_t besides the one obtained in the lagged changes (ΔZ_{t-k}). In that case the $\gamma = 0$, null hypothesis is not rejected (Wikipedia (b)).

Said & Dickey (1984) have shown that the Dickey-Fuller procedure remains valid asymptotically for a general ARIMA (p, 1, q) process in which p and q are of unknown orders, that is for an autoregressive integrated moving average process of the indicated order provided that the lag length in the autoregression increases with the sample size, T, at a controlled rate less than $T^{1/3}$.

Enders (1948) states that "the Dickey-Fuller tests assume that the errors are independent and have a constant variance, and this assumption raises four important problems related to the fact that we do not know the true data-generating process" (p. 225). He explains these problems as follows:

First, the true data-generating process may contain both autoregressive and moving average components. We need to know how to conduct the test if the order of the moving average terms (if any) is unknown. Second, we cannot properly estimate γ and its standard error unless all the autoregressive terms are included in the estimating equation.... However, the true order of the autoregressive process is usually unknown to the researcher, so that the problem is to select the appropriate lag length. The third problem stems from the fact that the

Dickey-Fuller test considers only a single unit root. However, a p^{th} order autoregression has p characteristic roots; if there are $m \leq p$ unit roots, the series needs to be differenced m times to achieve stationarity. The fourth problem is that it may not be known whether an intercept and/or time trend belongs in p^{th} order autoregressive process (Enders, 1948, p. 225-226).

Because, an invertible MA model can be transformed into an autoregressive model, the procedure can be generalized to allow for moving average components. It is known that

$$\phi(B)Z_t = \theta(B)a_t \tag{3.8.32}$$

and

$$a_t = \frac{\phi(B)}{\theta(B)}Z_t = \pi(B)Z_t\,. \tag{3.8.33}$$

Since $\pi(B)$ will generally be an infinite-order polynomial, Equation (3.8.31) can be rewriten as follows:

$$\Delta Z_t = \gamma Z_{t-1} + \sum_{i=2}^{\infty} \beta_i \Delta Z_{t-i+1} + a_t\,. \tag{3.8.34}$$

Enders (1948) reminds that an infinite-order autoregression like (3.8.34) cannot be estimated using a finite data set. Said & Dickey (1984) have shown that an unknown ARIMA (p, 1, q) process can be well approximated by an ARIMA (n, 1, 0) autoregression of order no more than $T^{1/3}$. Thus, the first problem can be solved by using a finite-order autoregression to approximate (3.8.34).

Now, the second problem which is taken into account of in Section 3.2.4 is the number of parameters in the model. For a fixed number of observations, the more parameters in a model, the less efficient is the estimation of the parameters. Enders (1948) points out this matter as follows:

> Including too many lags reduces the power of the test to reject the null of a unit root since the increased number of lags necessitates the estimation of additional parameters and a loss of degrees of freedom.... On the other hand, too few lags will not appropriately capture the actual error process, so that γ and its standard error will not be well estimated (Enders, 1948, p. 226-227).

For this problem, the approach which is generally preferred is to start with a relatively long lag length and pare down the model by the usual *t*-test and/or *F*-tests.

As regards the problem of multiple unit roots, Dickey & Pantula (1987) suggest a simple extension of the basic procedure if more than one unit root is suspected. The procedure is that when exactly one root is suspected, the Dickey-Fuller procedure is used to estimate an equation such as $\Delta Z_t = \alpha + \gamma Z_{t-1} + a_t$, and when two roots are suspected, estimate the equation

$$\Delta^2 Z_t = \alpha + \beta_1 \Delta Z_{t-1} + a_t \qquad (3.8.35)$$

and as follows. If β_1 is significantly different from zero, it is understood that the $\{Z_t\}$ sequence is integrated of order 2. Otherwise, the process is detected for one unit root.

For the fourth problem, it is proposed to use the information criterions some of which are told in the Section 3.7. See Enders (1948) for more details.

Phillips & Perron (1988) state that Dickey-Fuller tests show some problems when the parameter value is close to 1, which is named as near-integrated process, and the reason of this problem is that "the sample moments of a near-integrated time series converge weakly to corresponding functional of a diffusion process rather than standard Brownian motion" (p. 342).

3.8.6 Phillips-Perron Tests

The Phillips-Perron test statistics are modifications of the Dickey-Fuller *t*-statistics that take into account of the less restrictive nature of the error process. While the Dickey-Fuller tests assume the errors to be statistically independent and to have a constant variance, the Phillips-Perron tests allow the errors/disturbances to be weakly dependent and heterogeneously distributed. Consider the following regression equations to understand the procedure:

$$Z_t = \alpha^* + \phi^* Z_{t-1} + u_t \tag{3.8.36}$$

and

$$Z_t = \tilde{\alpha} + \tilde{\phi}_1 Z_{t-1} + \tilde{\phi}_2 (t - T/2) + u_t \tag{3.8.37}$$

where T is the number of observations. Enders (1948) states that "the disturbance term u_t is such that $E(u_t) = 0$, but there is no requirement that the disturbance term is serially uncorrelated or homogeneous" (p. 229). The critical values for the Phillips-Perron statistics are precisely those given for the Dickey-Fuller tests. Enders (1948) explains why the two equations above are not as simple as their appearance. For example, let $\{u_t\}$ sequence be generated by the autoregressive process $u_t = [\theta(B)/\zeta(B)]a_t$, where $\theta(B)$ and $\zeta(B)$ are polynomials in the lag operator. Given this form of the error process, Equation (3.8.36) may be written in the form used in the Dickey-Fuller tests; that is,

$$\zeta(B) Z_t = \alpha^* \zeta(B) + \phi^* \zeta(B) Z_{t-1} + \theta(B) a_t \tag{3.8.38}$$

or

$$(1 - \phi^* B)\, \zeta(B) Z_t = \alpha^* \zeta(B) + \theta(B) a_t\,. \tag{3.8.39}$$

Monte Carlo studies have shown that in the presence of *negative* moving average terms, the Phillips-Perron test tends to reject the null of a unit root whether or not the actual data-generating process contains a negative unit root. It is preferable to use the ADF test when the true model contains negative moving average terms

and the Phillips-Perron test when the true model contains positive moving average terms (Enders, 1948, p. 232).

Enders (1948) proposes that since the true-generating process is never known, instead of choosing one test, both of these unit root tests should be used. He states that "if they reinforce each other, you can have confidence in the results" (p. 233).

Phillips & Perron (1988) state that unmodified versions of the Dickey-Fuller tests are valid asymptotically in the presence of some heterogeneity in the innovation sequence provided the innovations are martingale differences and as the sample size $T \longrightarrow \infty$,

$$Z_T(r) \Rightarrow W(r) \quad \text{weakly converges}$$

where $W(r)$ is the standard Brownian motion process, and

$$Z_T(r) = T^{-1/2}\sigma^{-1}S_{[Tr]} = T^{-1/2}\sigma^{-1}S_{j-1}$$

for $(j-1)/T \leq r \leq j/T\ (j = 1, \ldots, T)$ where $[Tr]$ denotes the integral part of Tr, and

$$\sigma^2 = E(a_1^2) + 2\sum_{k=2}^{\infty} E(a_1 a_k)$$

and $S_t = a_1 + \ldots + a_t$.

In probability theory, a **martingale** is model of a fair game where knowledge of past events never helps predict future winnings. In particular, a martingale is a sequence of random variables (i.e., a stochastic process) for which, at a particular time in the realized sequence, the expectation of the next value in the sequence is equal to the present observed value even given knowledge of all prior observed values at a current time. An unbiased random walk (in any number of dimensions) is an example of a martingale. Briefly,

$$E(Z_{t+1}|Z_t, Z_{t-1}, \dots, Z_1) = Z_t, \qquad \text{then Z is a martingale.}$$

In probability theory, a **martingale difference sequence (MDS)** is related to the concept of the martingale. A stochastic series Z is an MDS if its expectation with respect to the past is zero. If Z_t is a martingale, then $X_t = Z_t - Z_{t-1}$ will be an MDS (Wikipedia (c)).

3.8.7 Problems in Testing for Unit Roots

These problems can be divided into two groups: The power of the test, and the determination of the deterministic regressors.

3.8.7.1 Power of the Test

The **power** of a test is equal to the probability of rejecting a false null hypothesis (i.e., Power = 1 – Type II error). The Dickey-Fuller and the Phillips-Perron tests have low statistical power in that they often cannot distinguish between true unit-root processes ($\gamma = 0$) and near unit-root processes (γ is close to zero). This is called *the near observation equivalence problem*. Hence, these tests will too often indicate that a series contains a unit root. Moreover, Enders (1948) states that "they have little power to distinguish between trend stationary and drifting processes. In finite samples, any trend stationary process can be arbitrarily well approximated by a unit root process, and a unit root process can be arbitrarily well approximated by a trend stationary process" (p. 251-252). See Enders (1948) who shows that a trend stationary process can be made to mimic a unit root process arbitrarily well.

Monte Carlo studies indicate that when the true data-generating process is stationary but has a root close to unity, the one-step ahead forecast from a differenced model are usually superior to the forecasts from a stationary model (Enders, 1948, p. 254).

3.8.7.2 Determination of the Deterministic Regressors

Since the actual data-generating process is unknown, it may be sensible to use the most general of the models to test the hypothesis $\gamma = 0$, that is,

$$\Delta Z_t = \alpha + \delta t + \gamma Z_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Z_{t-i+1} + a_t .$$

Enders (1948) states that "if the true process is a random walk process, this regression should find that $\alpha = \gamma = \delta = 0$" (p. 254). However, this way may cause some problems. For example, the additional estimated parameters reduce degrees of freedom and the power of the test. As a result of this situation, the process is evaluated as a unit root process by mistake. For the problems and solutions concerning unit root tests are explained in Enders (1948).

3.8.8 Structural Change

Economic structural change is defined as a long-term shift in the fundamental structure of an economy, which is often linked to growth and economic development. It is an economic condition that occurs when an industry or market changes how it functions or operates.

Enders (1948) shows that "when there are structural breaks, the various Dickey-Fuller and Phillips-Perron test statistics are biased toward the nonrejection of a unit root" (p. 233). Perron (1989) proved with a Monte Carlo study that the bias of the Dickey-Fuller tests becomes more pronounced as the magnitude of the break increased. Mostly used economic procedure to test for unit roots in the presence of a structural break involves splitting the sample into two parts and using Dickey-Fuller tests on each part. The problem with this procedure which is also mentioned in Enders (1948) is that the degrees of freedom for each of the resulting regressions are diminished. It is preferable to have a single test based on the full sample. See Perron

(1989) for the alternative procedure. Moreover, Perron & Vogelsang (1992) show how to test for a unit root when the precise date of the structural break is unknown.

3.8.9 AR(1) Process Including both a Constant Term and a Linear Time Trend and as well as an Autoregressive Error

In this section, for an AR(1) model including both a constant term and a linear time trend is examined. Moreover, the errors are not supposed to be i.i.d. This approach is presented in DeJong et al. (1992) and the theoretical proofs are given with more details.

Let the time series $\{Z_t\}$ be a stochastic process generated by the linear model

$$Z_t = \alpha_0 + \alpha_1 t + X_t \tag{3.8.40}$$

and the first-order autoregressive (AR) process

$$X_t = \beta X_{t-1} + a_t \tag{3.8.41}$$

where the innovation sequence $\{a_t\}$ is i.i.d. $N(0, \sigma^2)$, and X_0 is an unknown constant. This model can be interpreted as a random walk about a linear trend when $\beta = 1$ and as an asymptotically stationary AR(1) process about a linear trend when $|\beta| < 1$.

$$X_{t-1} = \beta X_{t-2} + a_{t-1}$$

$$X_{t-2} = \beta X_{t-3} + a_{t-2}\,.$$

At the second step, the equation below is obtained:

$$X_{t-1} = \beta(\beta X_{t-3} + a_{t-2}) + a_{t-1} = \beta^2 X_{t-3} + \beta a_{t-2} + a_{t-1}$$

$$X_{t-3} = \beta X_{t-4} + a_{t-3}$$

$$X_{t-1} = \beta^2(\beta X_{t-4} + a_{t-3}) + \beta a_{t-2} + a_{t-1}.$$

At the third step, the equation below is obtained:

$$X_{t-1} = \beta^3 X_{t-4} + \beta^2 a_{t-3} + \beta a_{t-2} + a_{t-1}.$$

With successive substitutions, the equation below is obtained:

$$X_{t-1} = \beta^{t-1} X_0 + \sum_{k=0}^{t-1} \beta^{t-k+1} a_k + a_{t-1}$$

$$Z_{t-1} = \alpha_0 + \alpha_1(t-1) + \left[\beta^{t-1} X_0 + \sum_{k=0}^{t-1} \beta^{t-k+1} a_k + a_{t-1}\right]$$

$$Z_t = \alpha_0 + \alpha_1 t + \beta\left[\beta^{t-1} X_0 + \sum_{k=0}^{t-1} \beta^{t-k+1} a_k + a_{t-1}\right] + a_t.$$

Multiply the both sides of Z_{t-1} by β:

$$\beta Z_{t-1} = \alpha_0\beta + \alpha_1 t\,\beta - \alpha_1\beta + \beta^t X_0 + \sum_{k=0}^{t-1} \beta^{t-k+2} a_k + \beta a_{t-1}$$

$$\begin{aligned} Z_t - \beta Z_{t-1} + \beta Z_{t-1} & \\ &= \alpha_0 - \alpha_0\beta + \alpha_1 t - \alpha_1 t\,\beta + \alpha_1\beta + \beta^t X_0 - \beta^t X_0 \\ &+ \sum_{k=0}^{t-1} \beta^{t-k+2} a_k - \sum_{k=0}^{t-1} \beta^{t-k+2} a_k + \beta a_{t-1} - \beta a_{t-1} + \beta Z_{t-1} \\ &+ a_t \end{aligned}$$

$$Z_t = \alpha_0 - \alpha_0\beta + \alpha_1 t - \alpha_1 t\,\beta + \alpha_1\beta + \beta Z_{t-1} + a_t$$

$$Z_t = \alpha_0(1-\beta) + \alpha_1\beta + t\alpha_1(1-\beta) + \beta Z_{t-1} + a_t$$

$$Z_t = \gamma + t\delta + \beta Z_{t-1} + a_t$$

where

$$\gamma = \alpha_0(1 - \beta) + \alpha_1\beta \quad \text{and} \quad \delta = \alpha_1(1 - \beta).$$

If $\beta = 1$,

$$Z_t = \alpha_1 + Z_{t-1} + a_t \tag{3.8.42}$$

and $\delta = 0$. Equation (3.8.42) is a rearrangement of the quasi-first-difference transform of (3.8.40). The coefficients of interest are α_0, α_1, and β; Equation (3.8.42) is viewed as the reduced form of (3.8.40)-(3.8.41), and the coefficients γ and δ are treated as reduced-form parameters.

3.8.10 Bootstrap Unit Root Tests

Swensen (2003) states the reason why the bootstrap procedures give generally more accurate results than the procedures of asymptotic approximations as follows:

> Bootstrap procedures offer an opportunity to take into account such factors as sample size, various specifications on the initial condition, and the distribution of the errors. They may therefore have more accurate finite-sample properties than procedures making use of asymptotic approximations, where such elements typically do not enter (Swensen, 2003, p. 32).

Unit root test is the one for which the bootstrap procedures give satisfactory results.

> One important aspect to design a bootstrap procedure for testing purposes is that the procedure should be able to reproduce the sampling behaviour of the test statistic under the null hypothesis (e.g., unit root integration), whether the observed series obeys the null hypothesis or not.... The theory developed for bootstrapping unit root tests in an autoregressive (AR) context has been concerned mainly with the large-sample behaviour of the methods proposed under the assumption that the null hypothesis is true (Paparoditis & Politis, 2005, p. 545).

There are several proposed tests concerning with unit root. Many of them are based on the restricted residuals, the rest is based on the unrestricted residuals.

Ferretti & Romo (1996) propose bootstrap tests for unit roots in first-order autoregressive models and they establish their asymptotic validity both for independent and for autoregressive errors. They state that "in this case, the bootstrap methodology directly approaches the asymptotic distribution, making unnecessary the usual corrections due to dependence of innovations" (p. 849). Their procedure is identical to that based on unrestricted residuals, whereas for higher-order autoregressions, it leads to a two-step procedure. With a Monte Carlo study, they state that "for the case of independent innovations the power of the test is not lower in any case than the power of the other methods and it is more powerful for small samples" (p. 858). These mentioned methods are such as the Box-Pierce statistic, the likelihood ratio test statistic, Dickey & Fuller test statistic, Phillips & Perron test statistics. They also state that "if the errors have an autoregressive structure and the AR parameters are negative, their bootstrap tests... improve previous procedures for small samples and behave similarly for large ones" (p. 858). Hence, Ferretti & Romo (1996) propose that for small sample sizes to use these bootstrap tests should be preferred.

For the unit root tests, Chang & Park (2000) consider the sieve bootstrap which is based on an approximation of an infinite dimensional and nonparametric model by a sequence of finite dimensional parametric models, and which allows order to increase with the sample size. Clearly, it is the most natural bootstrap procedure for the tests. Chang & Park (2000) proposed test statistics of S_n^* and T_n^* which are the bootstrap counterparts of S_n and T_n of Dickey-Fuller. They state that the choice of the initial value Z_0^* for (Z_t^*) does not affect the asymptotics as long as it is stochastically bounded, however, it may affect the finite sample performance of the bootstrap. If the mean or linear time trend is maintained in the formula below and the unit root test is performed using the demeaned or detrended data, then the effect of the initial value Z_0^* of the bootstrap sample would disappear and $Z_0^* = 0$.

$$Y_t = \alpha + Z_t \qquad \text{or} \qquad Y_t = \alpha + \delta t + Z_t$$

where

$$Z_t = \phi Z_{t-1} + a_t \, .$$

With a simulation study Chang & Park (2000) establish that "the bootstrap tests are found to have finite sample sizes that are generally much closer to their nominal values, especially for models with large negative moving average coefficients" (p. 394). See Chang & Park (2000) for the details.

Swensen (2003) considers the power functions of bootstrap unit root tests based on differences and on unrestricted residuals in the case of a first-order AR process with i.i.d. errors. Swensen shows that for sequences of local alternatives approaching the null at a rate equal to the inverse of the sample size, the power functions are the same as for ordinary unit root tests.

Paparoditis & Politis (2005) consider a new proposal based on unrestricted residuals. Their study shows that "bootstrap procedures based on differencing the observed series suffer from power problems as compared with bootstrap procedures based on unrestricted residuals" (p. 545). In their study, they investigate the behaviour of the different bootstrap approaches under the null and under fixed alternatives and, they analytically compare their relative performance. They consider the AR process

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t$$

where $\phi_p \neq 0$. Paparoditis & Politis (2005) assume that $\{a_t\}$ is a sequence of i.i.d. random variables with mean 0 and $0 < \sigma_a^2 < \infty$. They define the summation of the parameters as $\rho = \sum_{j=1}^{p} \phi_j$. Paparoditis & Politis (2005) compare the two different bootstrap proposals: based on restricted residuals and based on unrestricted residuals. They state that a difference in the limiting behaviour of the two bootstrap proposals appears only if the alternative is true, that is Z is a stationary process. In this case both bootstrap statistics converge for $\rho > 1$ to different limits. Furthermore, for

$\rho = 1$, the limiting distributions are identical. In this case both bootstrap-based tests are asymptotically equivalent.

Paparoditis & Politis (2005) also investigate how the differences in the limiting behaviour of the two bootstrap procedures affect their relative performance under the alternative. They state theoretically that both tests are consistent in the sense that their power approaches unity as the sample size and/or the deviation from the null increases. However, with probability tending to 1, the power of the bootstrap test based on unrestricted residuals is bounded from below and uniformly in ρ by the power of the test based on differences. Briefly, the power of the test based on unrestricted residuals has much more power.

CHAPTER FOUR
NUMERICAL RESULTS

In this chapter, results for a group of simulation study will be presented to show the performance of df-AR Unit Root test which is the basic concept of this thesis. Three different methods are compared as regards their powers on the unit root tests: *Asymptotic*, *bootstrap*, and *sufficient bootstrap* methods. Two different situations are taken into consideration: (1) The residuals are supposed to be independent from each other. (2) The residuals are supposed to be dependent on each other. With graphs and tables, the presentation is supported.

For this study, 5% is taken as the nominal significance level. The power values obtained are based on 10,000 Monte Carlo simulations and 1,000 bootstrap replications.

4.1 Independent Residuals

In this section, the residuals are supposed to be independent from each other. Also, to show the effect of the distribution of the residuals on the power of the test, two different distributions are taken into consideration: $N(0,1)$ normal distribution, and $LOGN(0,1)$ lognormal distribution. To see the sample size effect, four different sample sizes are taken: $30,\ 50,\ 100$ and 250. Since this study is done to evaluate whether this test captures the unit root situation correctly or not, the parameter values are taken as forming a unit root process or a near-unit root process. Instead of taking the same parameter values for each sample sizes, the method using by Swensen (2003) is used. Swensen (2003) uses the recursion given in (3.8.4). He tests $H_0: \phi = 1$ against the local alternative $H_0: \phi = 1 + c/n$ with $c = 0, -1, -2, -5, -10, -15, -20$. He considers only the situation where the random variables a_t are independent and identically distributed with $N(0,1)$. He shows that as the sample size increases, the fraction of c/n goes to zero and the power of the test is good enough even in the near-unit root process.

To evaluate powers of the tests, the smaller values of c are also used for the lognormal distribution. Also to see the results of the situation where the process lost the property of *causality*, $\phi > 1, c = 1$ and $c = 10$ are used for both $N(0,1)$ normal distribution and $LOGN(0,1)$ lognormal distribution. The causality means that Z_t is expressible in terms of a_s, $s \leq t$.

The algorithm for the simulation study is given below:

*Step*1: For the standard normal distributed residuals, draw a random sample of n residuals from $N(0,1)$ distribution and define this vector as Z_t (For the lognormal with 0 mean and 1 variance distributed residuals, draw a random sample of n residuals from $LOGN(0,1)$ distribution).

*Step*2: Generate $AR(1)$ series by regressing Z_t with the parameter value $\phi = 1 + c/n$ where c's has the values defined above, and define this vector as X.

*Step*3: Calculate the unit root test statistic T given by the formula 3.8.22.

*Step*4: Calculate the new residuals from the regression equation and define this vector as zt_new.

*Step*5: Recenter the vector zt_new.

*Step*6: Count the number of acceptances (#(T> the critical value)) and rejections.

*Step*7: Collect the original sample test statistics in a vector.

*Step*8: Generate B bootstrap sample of size n from the recentered vector zt_new and define this matrix as zt_star.

*Step*9: Generate $AR(1)$ series by regressing zt_star under the assumption of $\boldsymbol{\phi = 1}$, and define this matrix as X_star.

*Step*10: Take all elements in the first column of X_star to be equal to the first value of X, since in this way, the confidence interval obtained for $\boldsymbol{\phi hat\_star}$ includes the real parameter value ϕ.

*Step*11: Calculate the bootstrap unit root test statistic T_star by the formula 3.8.22.

*Step*12: Calculate the bootstrap critical value by finding the $(1-\alpha)^{th}$ quantile of the ordered test statistics of the bootstrap samples.

*Step*13: Repeat Step1-Step12, S times.

*Step*14: Calculate the mean of the bootstrap critical values obtained from all simulations.

*Step*15: For each of the simulations, compare the original test statistic with the mean of the bootstrap critical values.

*Step*16: Count the original test statistics which are smaller than the mean of the bootstrap critical values.

*Step*17: Calculate the significance level for asymptotic DF test by dividing the number of test statistics smaller than DF critical value to the number of simulations S.

*Step*18: Calculate the significance level for bootstrapped DF test by dividing the number obtained in Step 16 to the number of simulations S.

For sufficient bootstrap df-AR Unit Root test, the algorithm is same except only the unique values of zt_star have been used for Step 9.

The parameter values are listed in Table 4.1.

Table 4.1 The values of c and ϕ used for the simulation study for df-AR unit root test with i.i.d. $N(0,1)$ and $LOGN(0,1)$ distributed residuals.

N		VALUES FOR NORMAL N(0,1) DISTRIBUTED RESIDUALS												
30	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.33	1.03	1.00	0.97	0.93	0.83	0.67	0.50	0.33				
50	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.20	1.02	1.00	0.98	0.96	0.90	0.80	0.70	0.60				
100	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.10	1.01	1.00	0.99	0.98	0.95	0.90	0.85	0.80				
250	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.04	1.004	1.00	0.996	0.992	0.98	0.96	0.94	0.92				
N		VALUES FOR LOGNORMAL LOGN(0,1) DISTRIBUTED RESIDUALS												
30	c	10	1	0	-1	-2	-5	-10	-15	-20	-24			
	ϕ	1.33	1.03	1.00	0.97	0.93	0.83	0.67	0.50	0.33	0.20			
50	c	10	1	0	-1	-2	-5	-10	-15	-20	-30	-40		
	ϕ	1.20	1.02	1.00	0.98	0.96	0.90	0.80	0.70	0.60	0.40	0.20		
100	c	10	1	0	-1	-2	-5	-10	-15	-20	-30	-40	-50	
	ϕ	1.10	1.01	1.00	0.99	0.98	0.95	0.90	0.85	0.80	0.70	0.60	0.50	
250	c	10	1	0	-1	-2	-5	-10	-15	-20	-30	-40	-50	-100
	ϕ	1.04	1.004	1.000	0.996	0.992	0.98	0.96	0.94	0.92	0.88	0.84	0.80	0.60

Table 4.2 Empirical rejection probabilities of df-AR unit root test with i.i.d. $N(0,1)$ and $LOGN(0,1)$ distributed residuals (**M:** method, ϕ**:** parameter, **A:** asymptotic method, **B:** bootstrap method, **S:** sufficient bootstrap method)

n	**M**	**NORMAL N(0,1) DISTRIBUTED RESIDUALS**												
30	ϕ	**1.33**	**1.03**	**1.00**	**0.97**	**0.93**	**0.83**	**0.67**	**0.50**	**0.33**				
	A	**0.00**	**0.03**	**0.05**	**0.08**	**0.13**	**0.34**	**0.79**	**0.97**	**1.00**				
	B	**0.00**	**0.03**	**0.05**	**0.07**	**0.11**	**0.30**	**0.75**	**0.96**	**1.00**				
	S	**0.00**	**0.02**	**0.04**	**0.07**	**0.11**	**0.28**	**0.71**	**0.95**	**1.00**				
50	ϕ	**1.20**	**1.02**	**1.00**	**0.98**	**0.96**	**0.90**	**0.80**	**0.70**	**0.60**				
	A	**0.00**	**0.03**	**0.05**	**0.08**	**0.12**	**0.33**	**0.78**	**0.97**	**1.00**				
	B	**0.00**	**0.03**	**0.04**	**0.07**	**0.11**	**0.31**	**0.75**	**0.97**	**1.00**				
	S	**0.00**	**0.02**	**0.04**	**0.07**	**0.11**	**0.30**	**0.72**	**0.95**	**1.00**				
100	ϕ	**1.10**	**1.01**	**1.00**	**0.99**	**0.98**	**0.95**	**0.90**	**0.85**	**0.80**				
	A	**0.00**	**0.03**	**0.05**	**0.08**	**0.12**	**0.32**	**0.77**	**0.97**	**1.00**				
	B	**0.00**	**0.03**	**0.05**	**0.07**	**0.11**	**0.31**	**0.75**	**0.97**	**1.00**				
	S	**0.00**	**0.02**	**0.05**	**0.07**	**0.11**	**0.30**	**0.72**	**0.95**	**1.00**				
250	ϕ	**1.04**	**1.004**	**1.00**	**0.996**	**0.992**	**0.98**	**0.96**	**0.94**	**0.92**				
	A	**0.00**	**0.03**	**0.05**	**0.08**	**0.12**	**0.33**	**0.77**	**0.97**	**1.00**				
	B	**0.00**	**0.03**	**0.05**	**0.08**	**0.12**	**0.32**	**0.76**	**0.97**	**1.00**				
	S	**0.00**	**0.03**	**0.05**	**0.08**	**0.12**	**0.32**	**0.73**	**0.95**	**1.00**				
n	**M**	**LOGNORMAL LOGN(0,1) DISTRIBUTED RESIDUALS**												
30	ϕ	**1.33**	**1.03**	**1.00**	**0.97**	**0.93**	**0.83**	**0.67**	**0.50**	**0.33**	**0.20**			
	A	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.11**	**0.51**	**0.82**			
	B	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.08**	**0.44**	**0.79**			
	S	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.07**	**0.25**			
50	ϕ	**1.20**	**1.02**	**1.00**	**0.98**	**0.96**	**0.90**	**0.80**	**0.70**	**0.60**	**0.40**	**0.20**		
	A	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.13**	**0.83**	**0.99**		
	B	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.11**	**0.81**	**0.99**		
	S	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.17**	**0.76**		
100	ϕ	**1.10**	**1.01**	**1.00**	**0.99**	**0.98**	**0.95**	**0.90**	**0.85**	**0.80**	**0.70**	**0.60**	**0.50**	
	A	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.20**	**0.82**	**0.99**	
	B	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.20**	**0.81**	**0.99**	
	S	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.09**	**0.44**	
250	ϕ	**1.04**	**1.004**	**1.000**	**0.996**	**0.992**	**0.98**	**0.96**	**0.94**	**0.92**	**0.88**	**0.84**	**0.80**	**0.60**
	A	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.05**	**0.32**	**1.00**
	B	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.05**	**0.33**	**1.00**
	S	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.98**

Table 4.2 shows that for $N(0,1)$ distributed residuals, df-AR test captures the unit root situation, $\phi = 1$, with an exact level equals to the given nominal level 5% for

even the sample size $n = 30$. While the parameter value decreases, the power of the test increases as expected. This table shows that as the sample size increases, all types of df-AR Test have powers well enough for even the near-unit root process, since for $n = 250$ when the parameter value is 0.96 which is very close to 1, the power of the test is 0.73. For $N(0,1)$ distributed residuals, the test reaches the power value of 1 at $\phi = 0.92$ and $\phi = 0.33$ for the sample sizes 250 and 30, respectively. Figure 4.1 and Figure 4.2 show the effect of the sample size on powers of the tests.

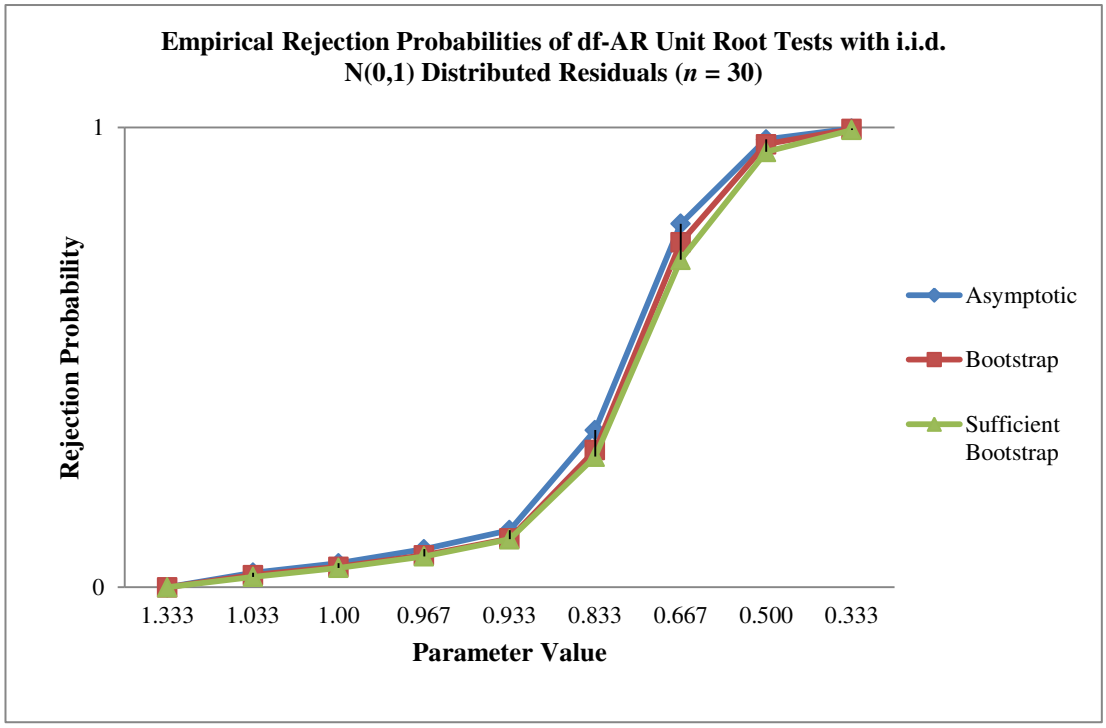

Figure 4.1 Empirical rejection probabilities of df-AR unit root tests with i.i.d. $N(0,1)$ distributed residuals $(n = 30)$

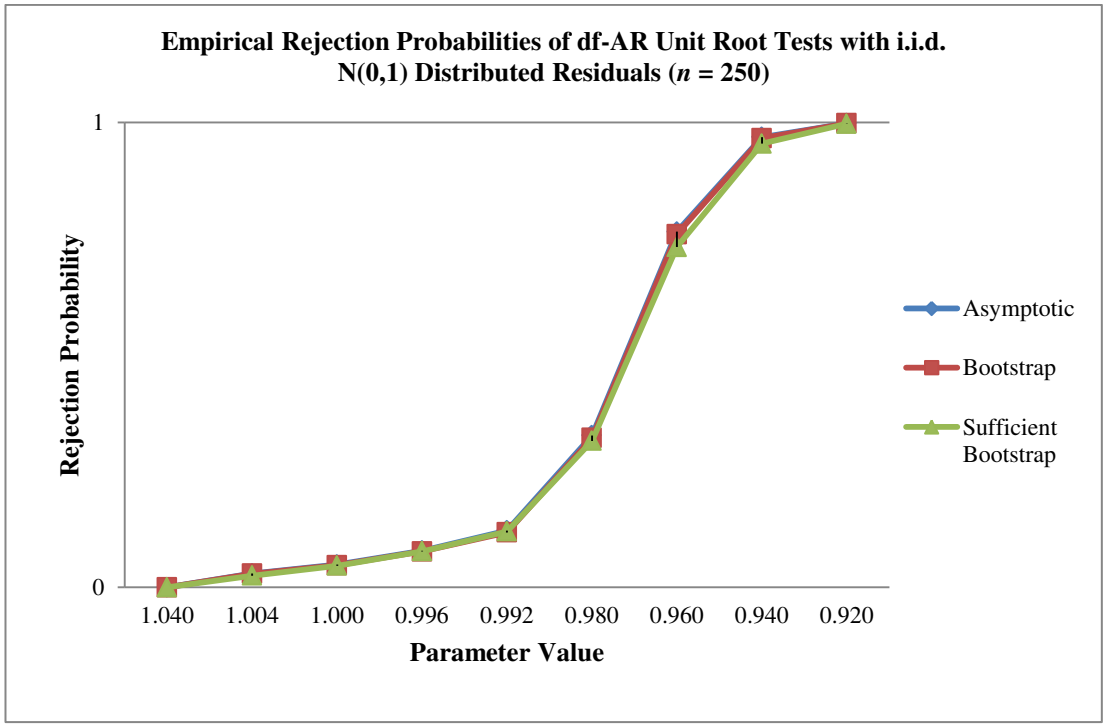

Figure 4.2 Empirical rejection probabilities of df-AR unit root tests with i.i.d. $N(0,1)$ distributed residuals $(n = 250)$

In Section 2.8, it is mentioned that in a sufficient bootstrap sample, units never appear more than once. Hence, by the sufficient bootstrap method, an important reduction in the sample size can be provided. Table 4.3 shows the percentiles of sample size reduction for the sample sizes used in this study..

Table 4.3 Comparing of the sample sizes used by the asymptotic and the bootstrap method with the sample sizes used by the sufficient bootstrap method

Sample Size used by Asymptotic and Bootstrap methods	**Sample Size used by Sufficient Bootstrap method**	**Percentiles of Sample Size reduction**
30	20.15	*32.8 %*
50	32.79	*34.4 %*
100	64.40	*35.6 %*
250	159.22	*36.3 %*

To evaluate whether a method is successful in testing a hypothesis or not, one indicator is the speed of its power being move away from the nominal level of the test. Figure 4.3 and Figure 4.4 show that for i.i.d. $N(0,1)$ distributed residuals, as the sample size increases, the differences between power and 0.05 are nearly the same for two different bootstrap methods.

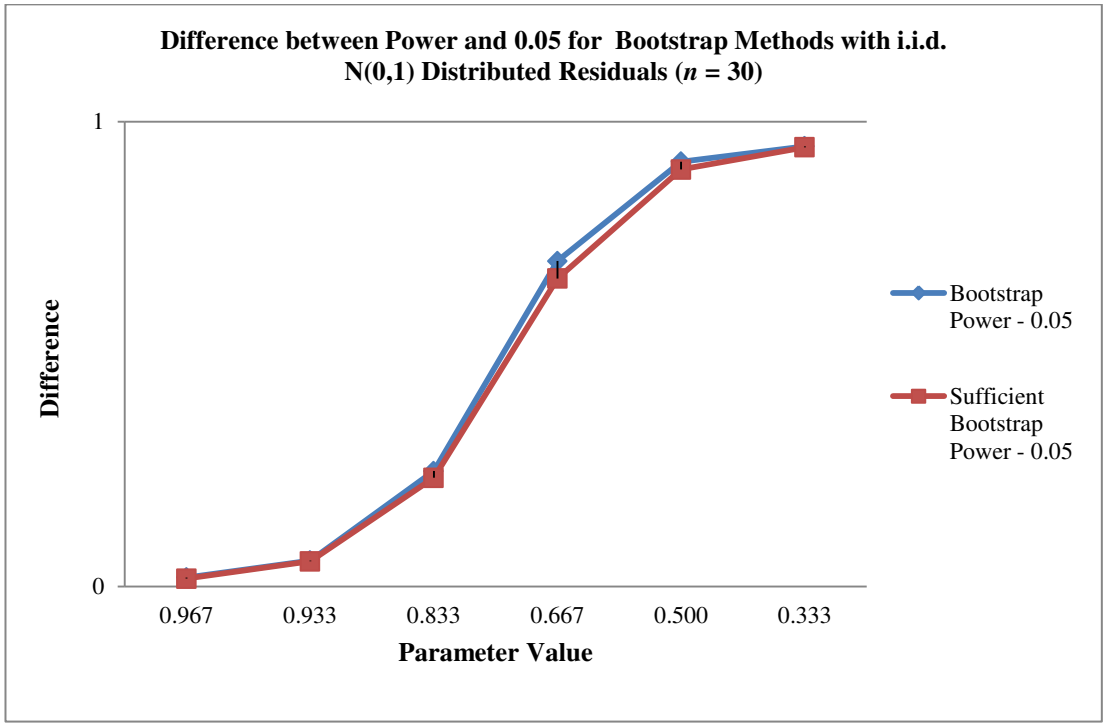

Figure 4.3 Difference between power and 0.05 for bootstrap methods with i.i.d. $N(0,1)$ distributed residuals $(n = 30)$

Briefly, for i.i.d. $N(0,1)$ distributed residuals, especially as the sample size increases, the satisfactory results are obtained. Even the sufficient bootstrap method with almost 35% reduction in sample size gives values nearly as well as given by the asymptotic method.

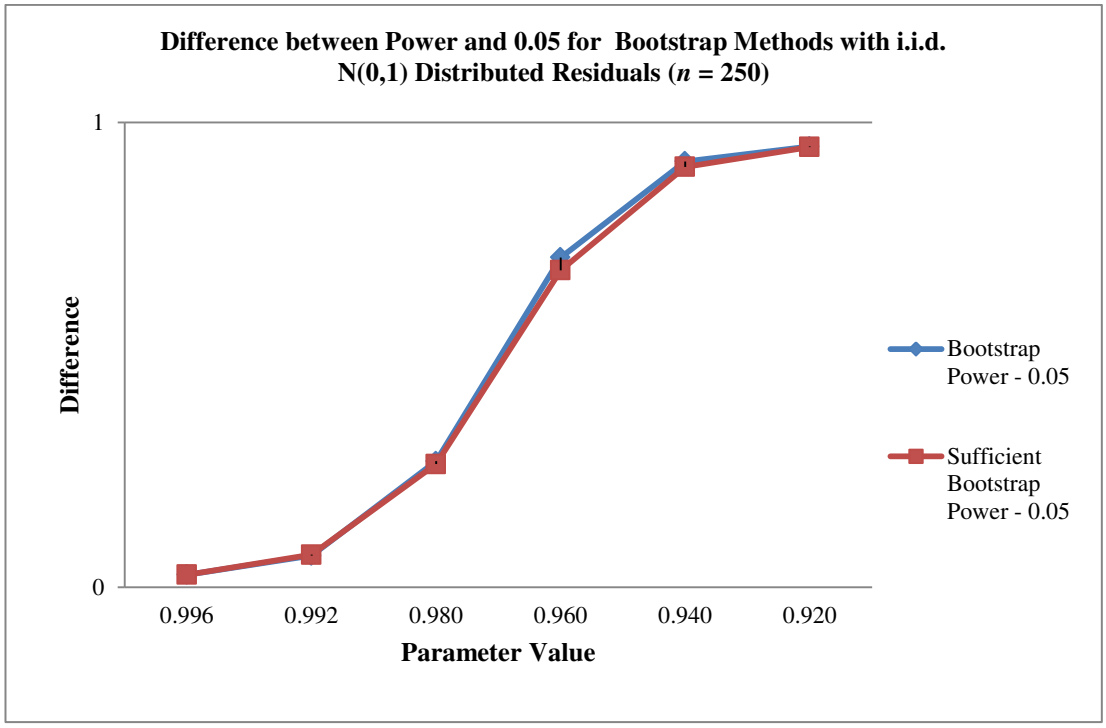

Figure 4.4 Difference between power and 0.05 for bootstrap methods with i.i.d. $N(0,1)$ distributed residuals ($n = 250$)

According to the results in Table 4.2, unlike $N(0,1)$ case, under $LOGN(0,1)$ all three tests do not perform well. For all sample sizes, when $\phi = 1$, the exact level of the test is 0. This situation does not change until the parameter value being move away from the unity. For instance, while for $N(0,1)$ distributed residuals and $n = 250$, the unit power is obtained for $\phi = 0.92$; for $LOGN(0,1)$ distributed residuals and $n = 250$, the unit power is obtained for $\phi = 0.60$. Moreover, for $LOGN(0,1)$ distributed residuals the power of the test increases very fast after a specific parameter value, while for $N(0,1)$ distributed residuals this increment is slow. Besides, as the sample size increases two methods, *the asymptotic* and *the bootstrap*, give the same power values, whereas *the sufficient bootstrap* does not give satisfactory results compared with the others. Like $N(0,1)$ distributed residuals case, as $\phi > 1$, the power of the test is zero. Figure 4.5 and Figure 4.6 show the effect of the sample size on powers of the tests.

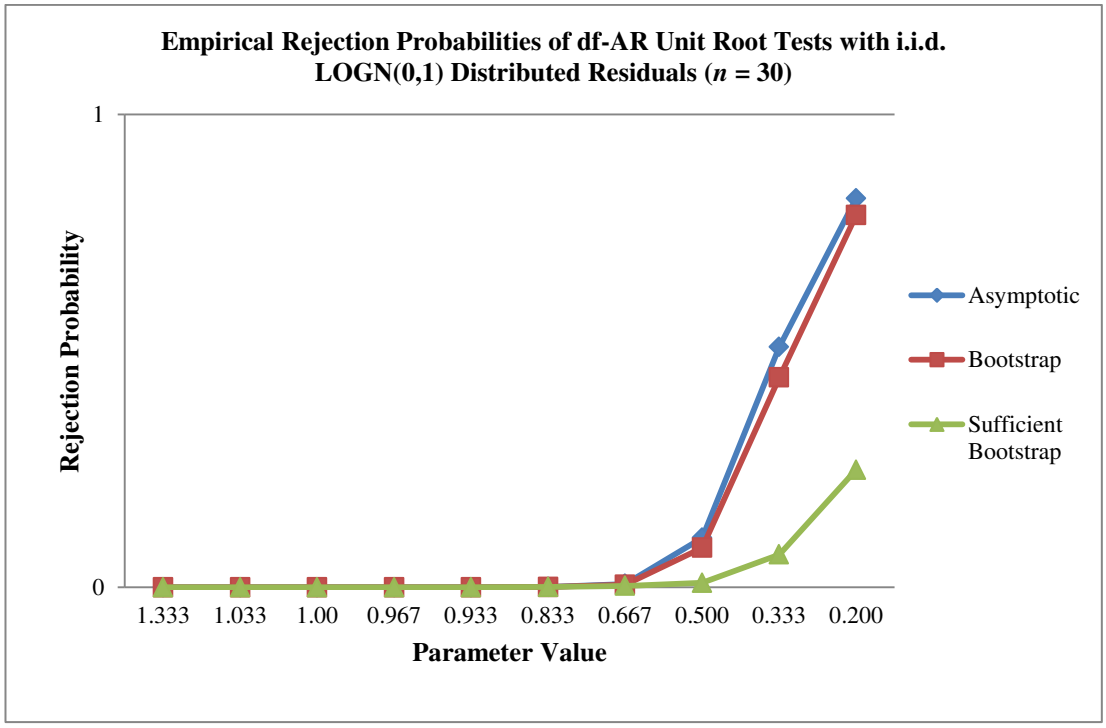

Figure 4.5 Empirical rejection probabilities of df-AR unit root tests with i.i.d. $LOGN(0,1)$ distributed residuals $(n = 30)$

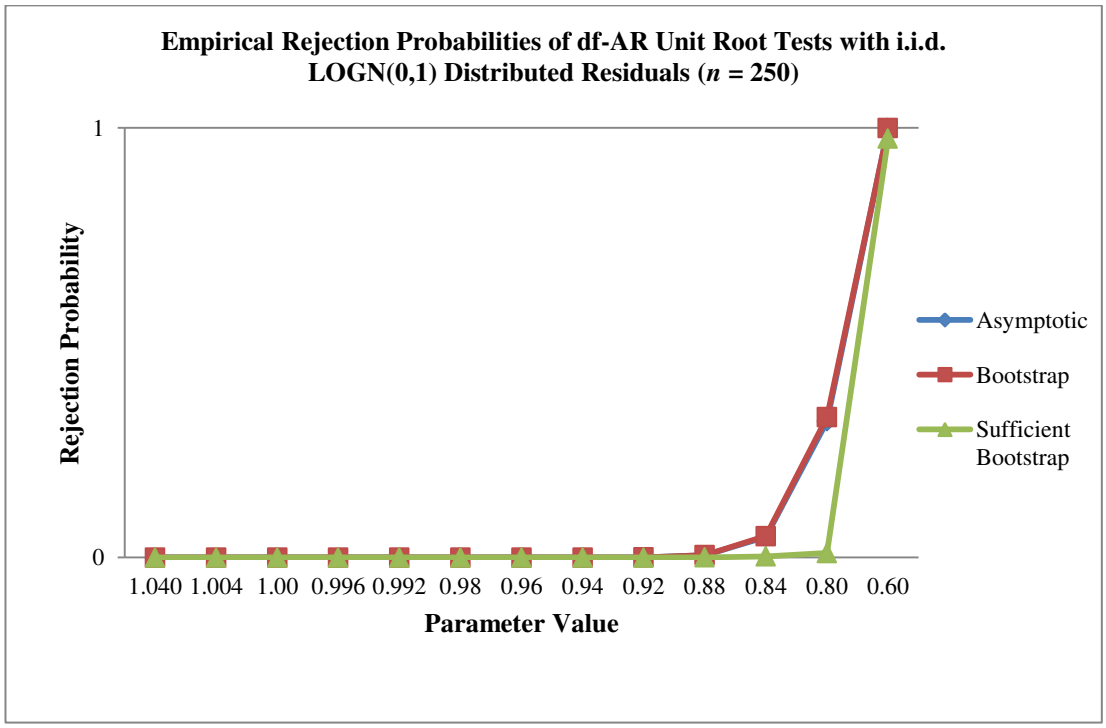

Figure 4.6 Empirical rejection probabilities of df-AR unit root tests with i.i.d. $LOGN(0,1)$ distributed residuals $(n = 250)$

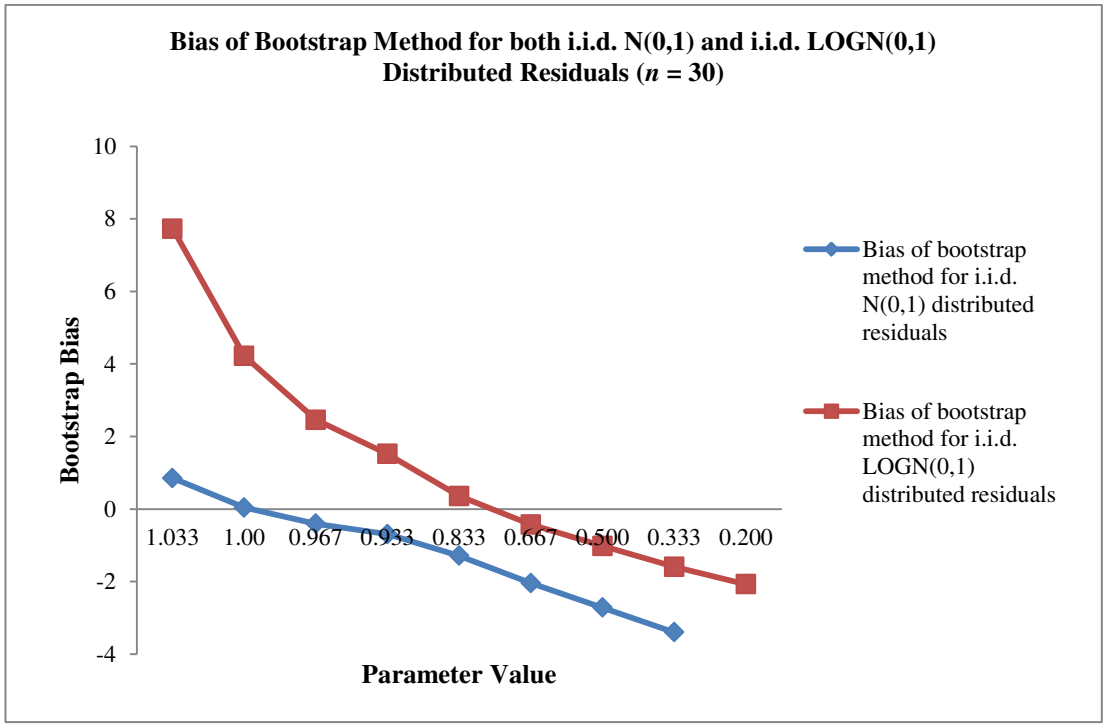

Figure 4.7 Bias of bootstrap method for both i.i.d. $N(0,1)$ and i.i.d. $LOGN(0,1)$ distributed residuals $(n = 30)$

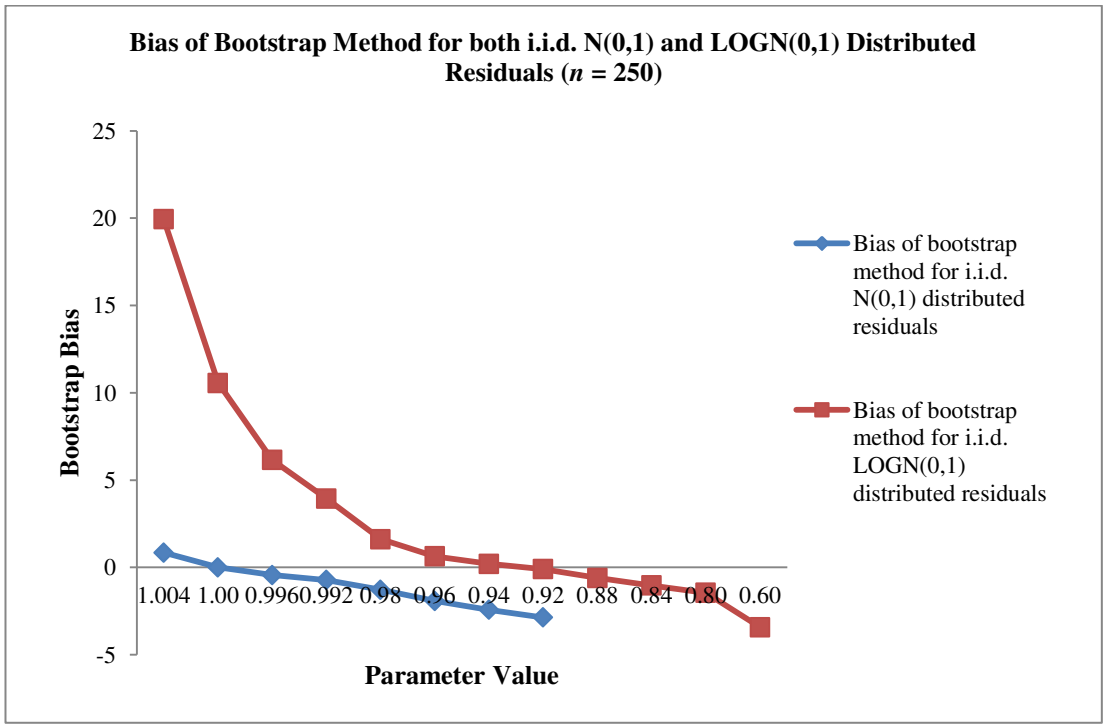

Figure 4.8 Bias of bootstrap method for both i.i.d. $N(0,1)$ and i.i.d. $LOGN(0,1)$ distributed residuals $(n = 250)$

The bias of the bootstrap method has also been calculated. The procedure is as follows: For the original sample, the tests statistic is calculated. Later, for $1{,}000$ bootstrap samples chosen from that original sample, the test statistics are calculated. Mean of all these test statistics of the bootstrap samples is subtracted from the test statistic of the original sample. For each simulation, one value is calculated. At the end of all simulations, mean of these all values is calculated and named as *bootstrap bias*.

Figures 4.7 and 4.8 illustrate the bias of bootstrap method for sample sizes of 30 and 250, respectively. For $n = 30$ and for $N(0,1)$ distributed residuals, as the parameter value decreases, the bootstrap bias takes the negatively large values. On the other hand, for $LOGN(0,1)$ distributed residuals, at the parameter values for which powers of the tests differ from zero, the bootstrap bias takes the negative values. Since df-AR Test is a lower-tail test and in this type of test, the null hypothesis is rejected if the test statistic is smaller than the critical value, that is the test statistic is more negative than the critical value, the bootstrap method tries to take a value negatively large to reject the null hypothesis. This is the situation which is wanted. That is, there is a positive correlation between the bootstrap bias taking the value negatively large and the power of the test. Hence, the bootstrap bias values are better for the situation for $N(0,1)$ distributed residuals.

The same thing may be said for Figure 4.8. Moreover, this graph shows that as the sample size increases, the negatively large values of the bootstrap bias are obtained without the parameter value being far away from the unity.

For the positive values of c, the bootstrap bias takes the positively large values, and as the sample size increases, the magnitude of these values increase too much. Hence, for $c = 10$, the value of bootstrap bias is not included in these figures.

To show the effect of the distribution of the residuals on the distribution of the test statistics, Figure 4.9 and Figure 4.10 are given. Both these graphs are obtained for $n = 250$ and the parameter value $\phi = 1$. These figures show that the sufficient

bootstrap test statistics and the original test statistics are influenced badly when $LOGN(0,1)$ distributed residuals are used. While the distribution of the sufficient bootstrap test statistics is normal for $N(0,1)$ distributed residuals, the distribution of the sufficient bootstrap test statistics tend to be piled up on a determined value for $LOGN(0,1)$ distributed residuals. Also, while the distribution of the original test statistics seems normal for $N(0,1)$ distributed residuals, the distribution of the original test statistics is skewed to the left for $LOGN(0,1)$ distributed residuals.

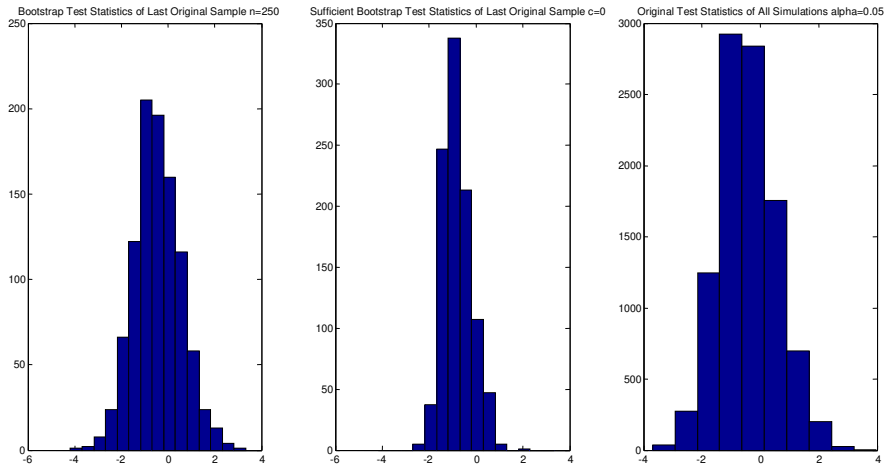

Figure 4.9 The distribution of 1,000 bootstrap test statistics of the last original sample, the distribution of 1,000 sufficient bootstrap test statistics of the last original sample, and the distribution of 10,000 original test statistics for i.i.d. $N(0,1)$ distributed residuals, respectively ($n = 250, \phi = 1$)

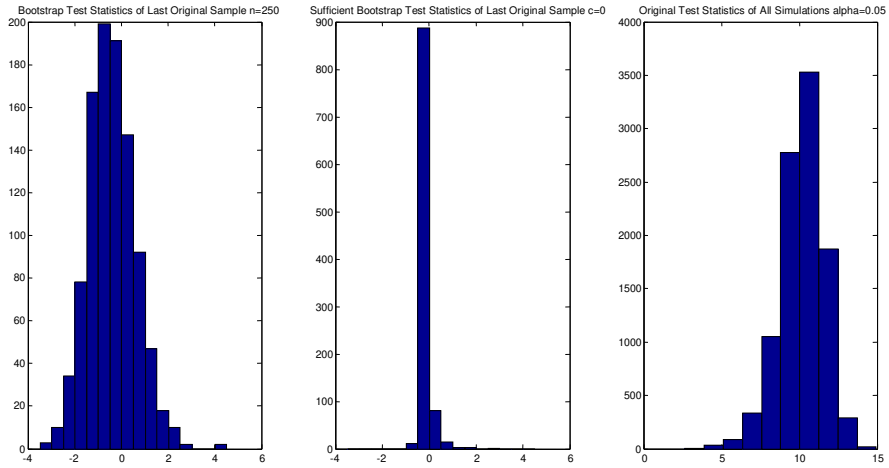

Figure 4.10 The distribution of 1,000 bootstrap test statistics of the last original sample, the distribution of 1,000 sufficient bootstrap test statistics of the last original sample, and the distribution of 10,000 original test statistics for i.i.d. $LOGN(0,1)$ distributed residuals, respectively $(n = 250, \phi = 1)$

4.2 Dependent Residuals

In this section, the residuals for the model (3.8.4) are supposed to be dependent on each other by the relation given with (4.2.1). Also, to show the effect of the magnitude of dependency on the power of the test, two different parameter values for dependency considered: $\beta = 0.2$ for "weak dependency" and $\beta = 0.8$ for "strong dependency" using the formulas

$$a_t = \beta a_{t-1} + u_t \qquad (4.2.1)$$

where u_t's are i.i.d. $N(0,1)$ distributed residuals.

To evaluate powers of the tests, the parameter values used in section 4.1 are also used for this study. However, the ones for $N(0,1)$are used for the weak dependency, while the parameters of $LOGN(0,1)$ distributed residuals are preferred for the strong dependency case. These values are listed in Table 4.4.

Table 4.4 The values of c and ϕ used for the simulation study for df-AR unit root test with weakly dependent a_t and i.i.d. $N(0,1)$ distributed u_t residuals, and strongly dependent a_t and i.i.d. $N(0,1)$ distributed u_t residuals.

n		VALUES FOR $\beta = 0.2$ WEAKLY DEPENDENT a_t RESIDUALS												
30	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.33	1.03	1.00	0.97	0.93	0.83	0.67	0.50	0.33				
50	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.20	1.02	1.00	0.98	0.96	0.90	0.80	0.70	0.60				
100	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.10	1.01	1.00	0.99	0.98	0.95	0.90	0.85	0.80				
250	c	10	1	0	-1	-2	-5	-10	-15	-20				
	ϕ	1.04	1.004	1.00	0.996	0.992	0.98	0.96	0.94	0.92				
n		VALUES FOR $\beta = 0.8$ STRONGLY DEPENDENT a_t RESIDUALS												
30	c	10	1	0	-1	-2	-5	-10	-15	-20	-24			
	ϕ	1.33	1.03	1.00	0.97	0.93	0.83	0.67	0.50	0.33	0.20			
50	c	10	1	0	-1	-2	-5	-10	-15	-20	-30	-40		
	ϕ	1.20	1.02	1.00	0.98	0.96	0.90	0.80	0.70	0.60	0.40	0.20		
100	c	10	1	0	-1	-2	-5	-10	-15	-20	-30	-40	-50	
	ϕ	1.10	1.01	1.00	0.99	0.98	0.95	0.90	0.85	0.80	0.70	0.60	0.50	
250	c	10	1	0	-1	-2	-5	-10	-15	-20	-30	-40	-50	-100
	ϕ	1.04	1.004	1.000	0.996	0.992	0.98	0.96	0.94	0.92	0.88	0.84	0.80	0.60

The algorithm for dependent residuals is the same with the algorithm for independent residuals. However, please note that in Step1, a random sample of n residuals from $N(0,1)$ distributionis drawn for u_t's,and the residuals a_t are obtained by the Equation (4.2.1).

Table 4.5 Empirical rejection probabilities of df-AR unit root test with weakly dependent a_t and i.i.d. $N(0,1)$ distributed u_t residuals, and strongly dependent a_t and i.i.d. $N(0,1)$ distributed u_t residuals (**M:** method, ϕ**:** parameter, **A:** asymptotic method, **B:** bootstrap method, **S:** sufficient bootstrap method)

n	M	$\beta = 0.2$ WEAKLY DEPENDENT a_t RESIDUALS												
30	ϕ	1.33	1.03	1.00	0.97	0.93	0.83	0.67	0.50	0.33				
	A	0.00	0.01	0.02	0.03	0.05	0.15	0.53	0.86	0.98				
	B	0.00	0.01	0.02	0.03	0.04	0.13	0.47	0.82	0.97				
	S	0.00	0.02	0.03	0.04	0.06	0.15	0.44	0.77	0.95				
50	ϕ	1.20	1.02	1.00	0.98	0.96	0.90	0.80	0.70	0.60				
	A	0.00	0.01	0.02	0.03	0.04	0.14	0.48	0.83	0.97				
	B	0.00	0.01	0.01	0.02	0.04	0.12	0.44	0.80	0.96				
	S	0.00	0.02	0.03	0.04	0.06	0.16	0.43	0.75	0.94				
100	ϕ	1.10	1.01	1.00	0.99	0.98	0.95	0.90	0.85	0.80				
	A	0.00	0.01	0.02	0.03	0.04	0.12	0.44	0.79	0.96				
	B	0.00	0.01	0.01	0.02	0.04	0.11	0.42	0.78	0.96				
	S	0.00	0.02	0.03	0.04	0.06	0.16	0.42	0.73	0.93				
250	ϕ	1.04	1.004	1.00	0.996	0.992	0.98	0.96	0.94	0.92				
	A	0.00	0.01	0.02	0.03	0.04	0.12	0.43	0.78	0.95				
	B	0.00	0.01	0.02	0.02	0.04	0.12	0.42	0.78	0.95				
	S	0.00	0.02	0.03	0.05	0.07	0.16	0.43	0.73	0.92				
n	M	$\beta = 0.8$ STRONGLY DEPENDENT a_t RESIDUALS												
30	ϕ	1.33	1.03	1.00	0.97	0.93	0.83	0.67	0.50	0.33	0.20			
	A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.10	0.20			
	B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.08	0.17			
	S	0.00	0.00	0.01	0.01	0.01	0.02	0.04	0.07	0.12	0.19			
50	ϕ	1.20	1.02	1.00	0.98	0.96	0.90	0.80	0.70	0.60	0.40	0.20		
	A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.17	0.48		
	B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.15	0.44		
	S	0.00	0.01	0.01	0.01	0.02	0.02	0.04	0.05	0.08	0.20	0.43		
100	ϕ	1.10	1.01	1.00	0.99	0.98	0.95	0.90	0.85	0.80	0.70	0.60	0.50	
	A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.18	0.42	
	B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.17	0.41	
	S	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.13	0.25	0.42	
250	ϕ	1.04	1.004	1.000	0.996	0.992	0.98	0.96	0.94	0.92	0.88	0.84	0.80	0.60
	A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.18	0.98
	B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.18	0.98
	S	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.06	0.10	0.18	0.29	0.92

Table 4.5 shows that for weakly dependent residuals with $\beta = 0.2$, all the methods give smaller significance levels under $\phi = 1$. The power values reach their maximum at $\phi = 0.92$ for $n = 250$ and $\phi = 0.33$ for $n = 30$. Also it is shown that as the sample size increases three methods, *the asymptotic*, *the bootstrap*, and *the sufficient bootstrap* give nearly the same power values. However, as $\phi > 1$, the power of the test goes to zero as in i.i.d. distributed residuals. Figure 4.11 and Figure 4.12 show the effect of the sample size on powers of the tests.

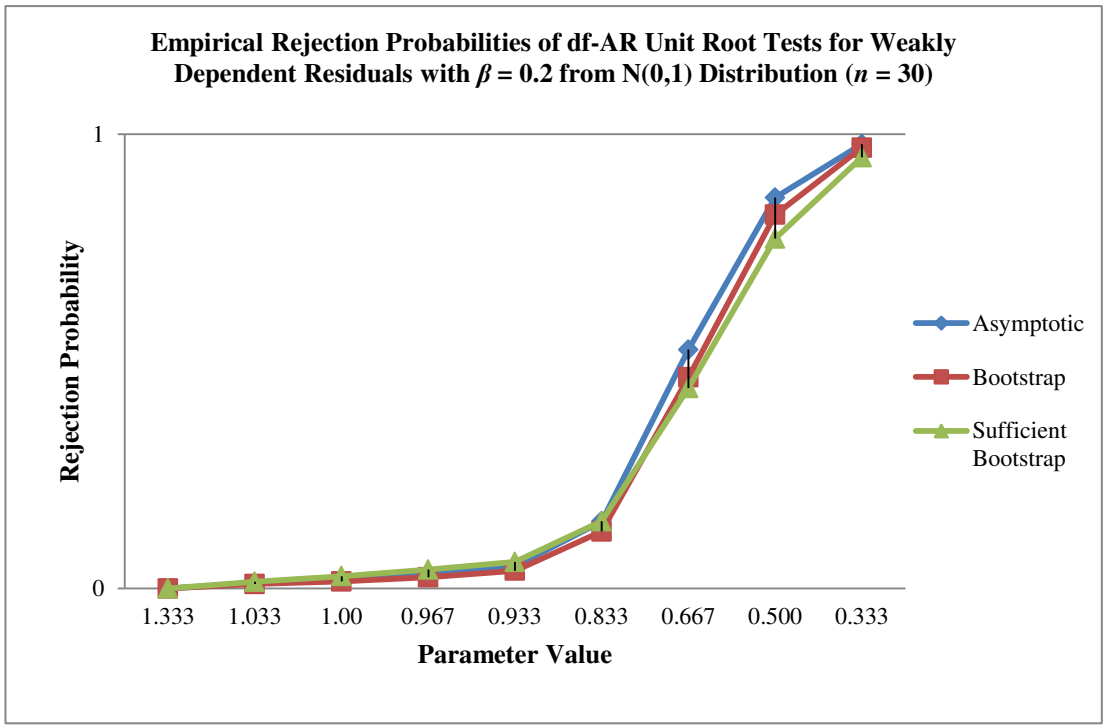

Figure 4.11 Empirical rejection probabilities of df-AR unit root tests for weakly dependent residuals with $\beta = 0.2$ from $N(0,1)$distribution $(n = 30)$

Figure 4.13 and Figure 4.14 show that for weakly dependent residuals with $\beta = 0.2$, in near-unit root processes the sufficient bootstrap method is more successful. Two different methods, the asymptotic and the bootstrap, show nearly the same patterns.

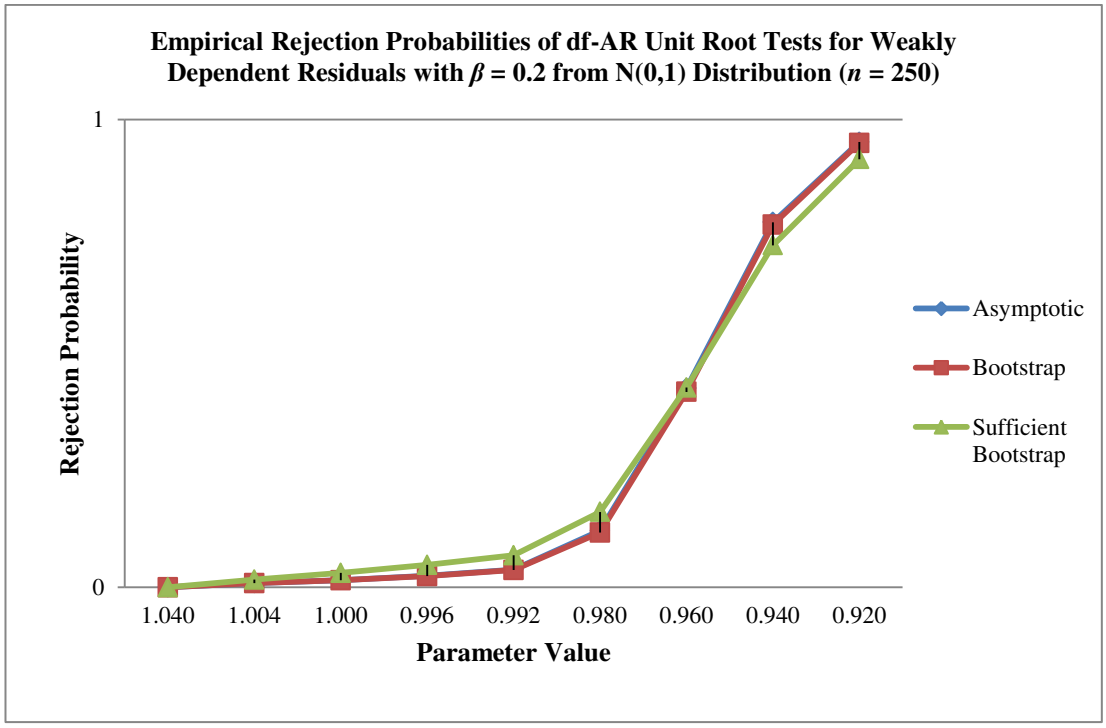

Figure 4.12 Empirical rejection probabilities of df-AR unit root tests for weakly dependent residuals with $\beta = 0.2$ from $N(0,1)$ distribution $(n = 250)$

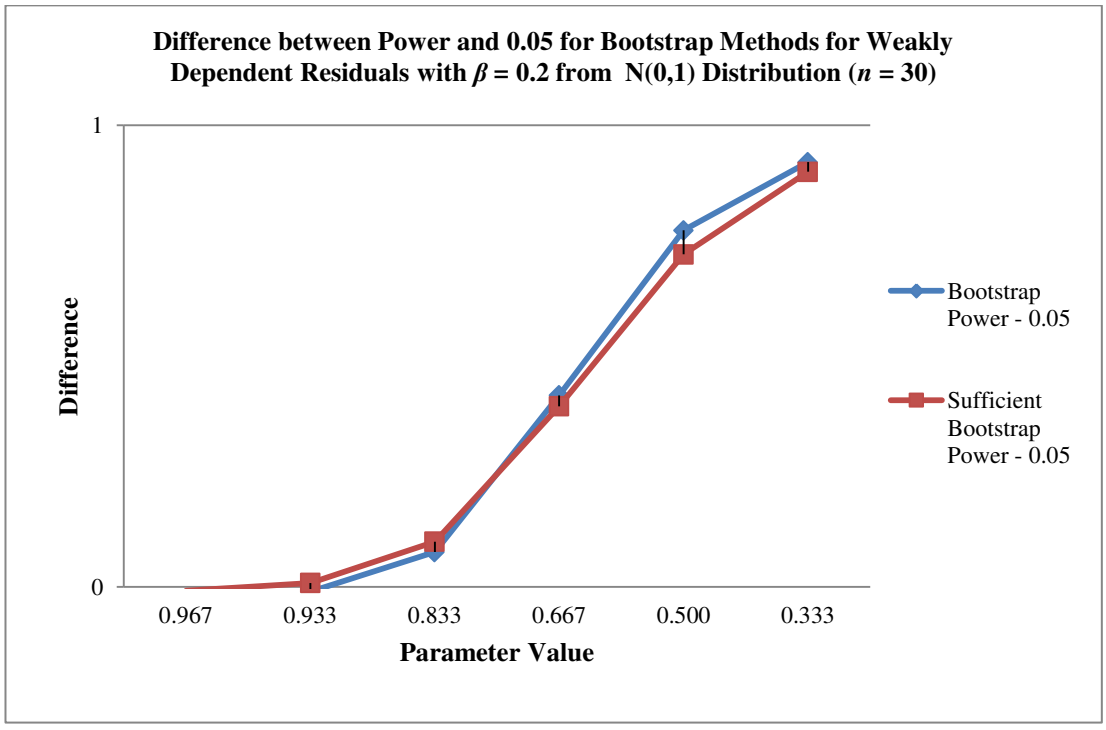

Figure 4.13 Difference between power and 0.05 for bootstrap methods for weakly dependent residuals with $\beta = 0.2$ from $N(0,1)$ distribution $(n = 30)$

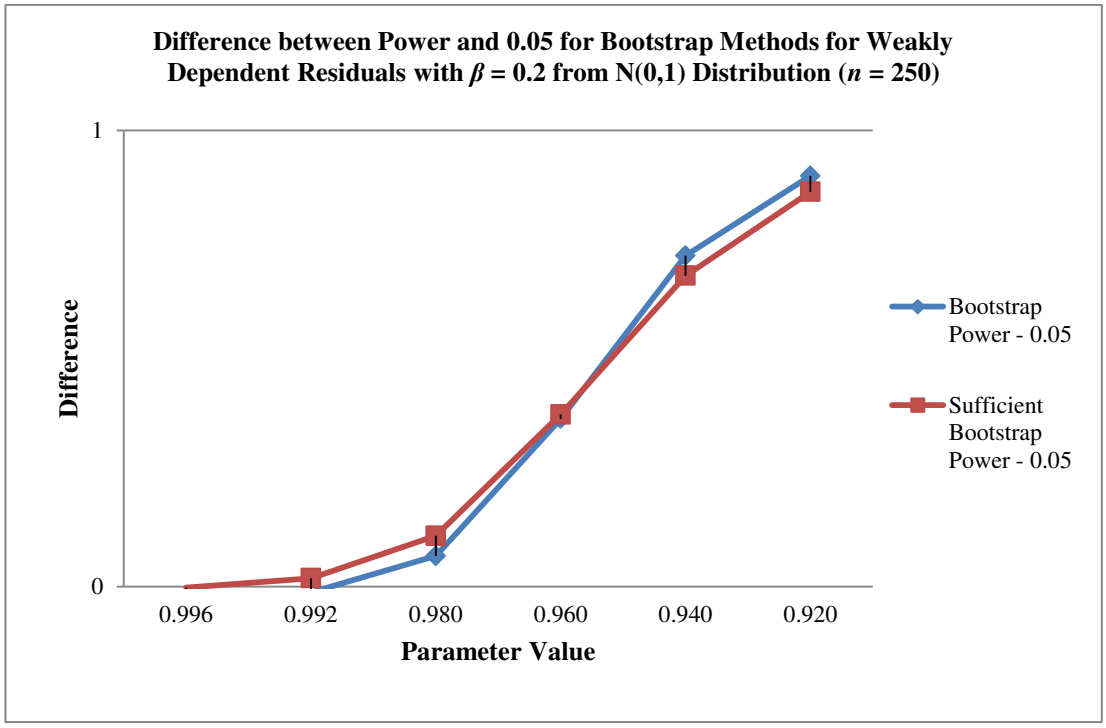

Figure 4.14 Difference between power and 0.05 for bootstrap methods for weakly dependent residuals with $\beta = 0.2$ from $N(0,1)$ distribution $(n = 250)$

Table 4.5 shows that for strongly dependent residuals with $\beta = 0.8$, all df-AR tests behave badly under $\phi = 1$ even for the sample size of 250. For all sample sizes, when $\phi = 1$, the exact level of the test is 0. This situation does not change until the parameter value being far away from the unity except for the method of sufficient bootstrap. It is surprising that the sufficient bootstrap method gives more satisfactory results than other methods in near-unit root processes. For instance, when the sample size $n = 250$ and $\phi = 0.92$; the sufficient bootstrap method has a power equals to 0.06 while the other methods have a power equals to zero. For the same sample size, when $\phi = 0.88$; the sufficient bootstrap method has a power equals to 0.10 while the other methods have a power equals to 0.01. Although, at the limit value of the parameters, the other methods have power values which are bigger than the values obtained by the sufficient bootstrap method, it can be said that the sufficient bootstrap method is more successful. Also it is seen from Table 4.5 that in the situation of strongly dependent residuals, all tests can't even get close to the power of one except from the sample size $n = 250$. For the sample sizes $n = 50$ and $n = 100$, the highest power value is approximately 0.50. Moreover, for

both strongly dependent and weakly dependent residuals, the power of the asymptotic and the bootstrap tests increase very fast after a specific parameter value, while for the sufficient bootstrap method the power increases slowly. Besides, for all sample sizes two methods, *the asymptotic* and *the bootstrap*, give the same power values, whereas *the sufficient bootstrap* gives more satisfactory results compared with the others. However, as $\phi > 1$, the power of all tests is zero. Figure 4.15 and Figure 4.16 show the effect of the sample size on powers of the tests.

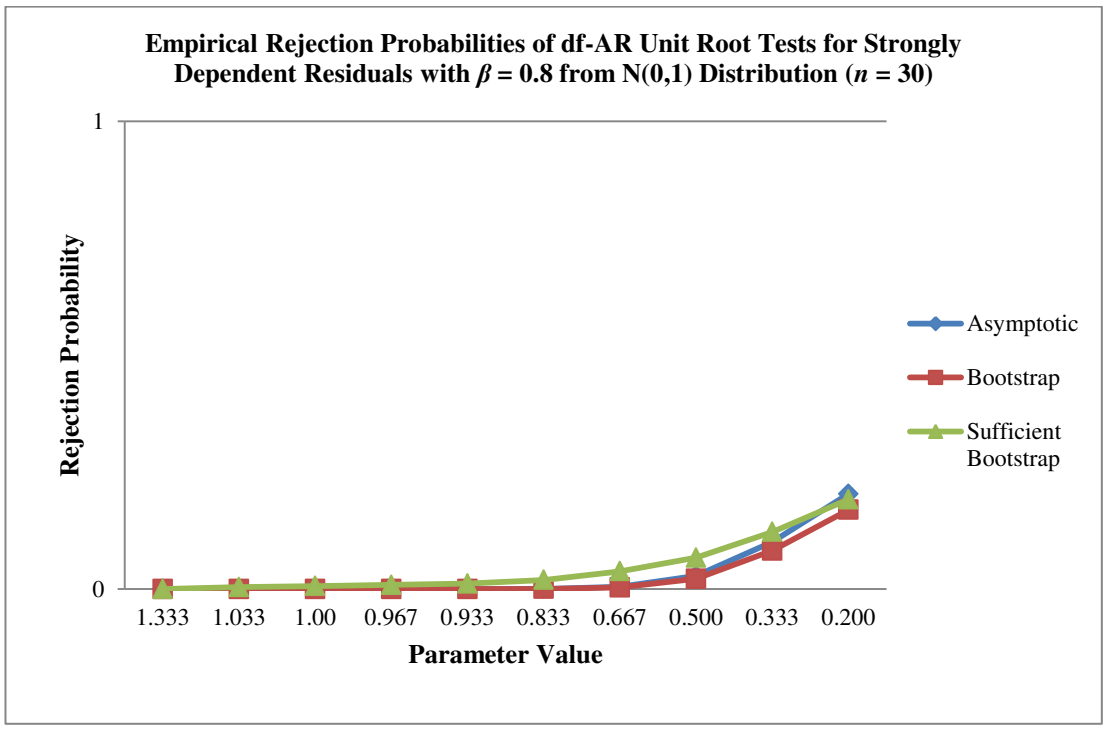

Figure 4.15 Empirical rejection probabilities of df-AR unit root tests for strongly dependent residuals with $\beta = 0.8$ from $N(0,1)$ distribution $(n = 30)$

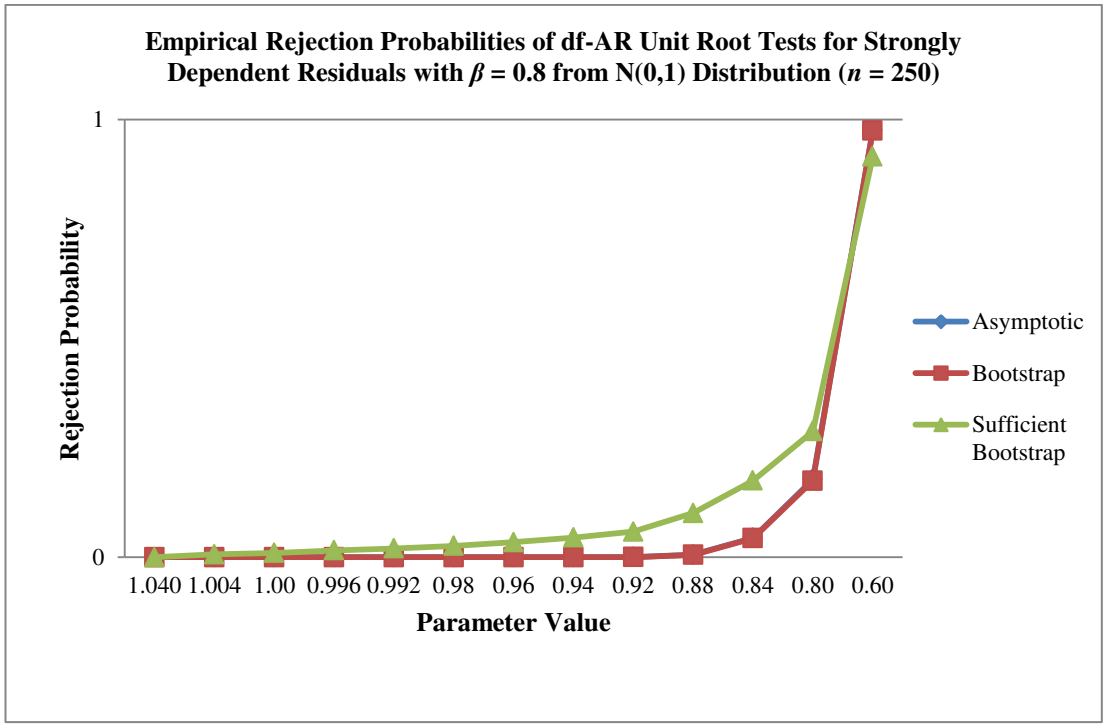

Figure 4.16 Empirical rejection probabilities of df-AR unit root tests for strongly dependent residuals with $\beta = 0.8$ from $N(0,1)$ distribution $(n = 250)$

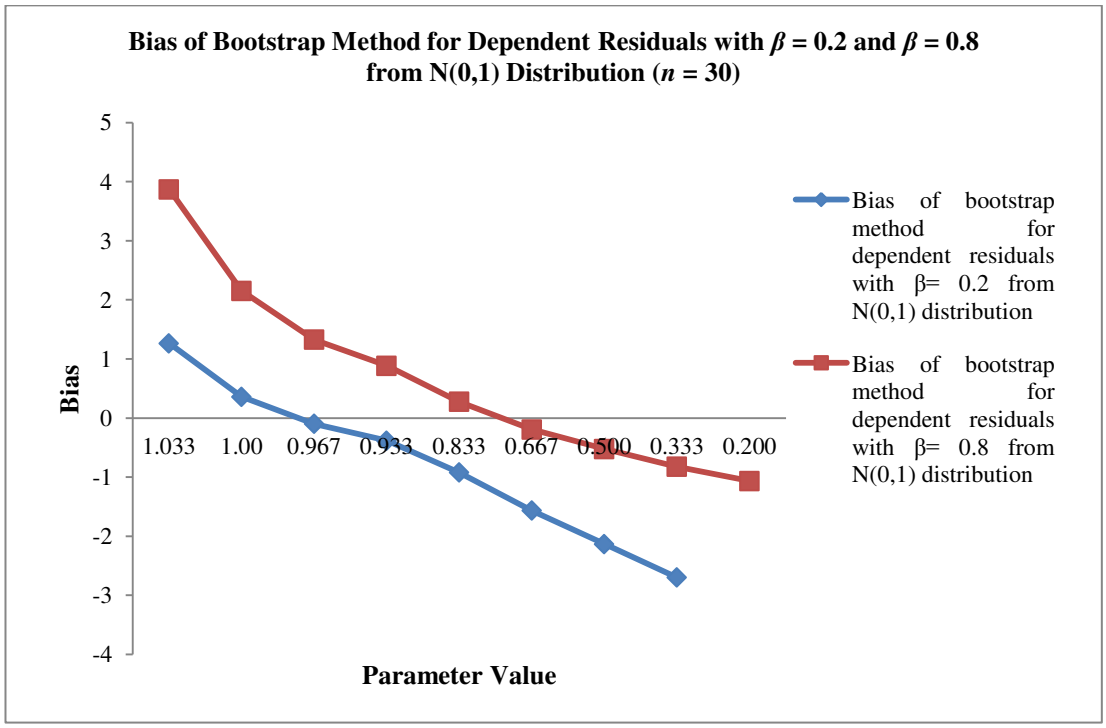

Figure 4.17 Bias of bootstrap method for dependent residuals with $\beta = 0.2$ and $\beta = 0.8$ $(n = 30)$

The patterns for the bias of bootstrap are the same for i.i.d. residual case. See Figures 4.17 and 4.18.

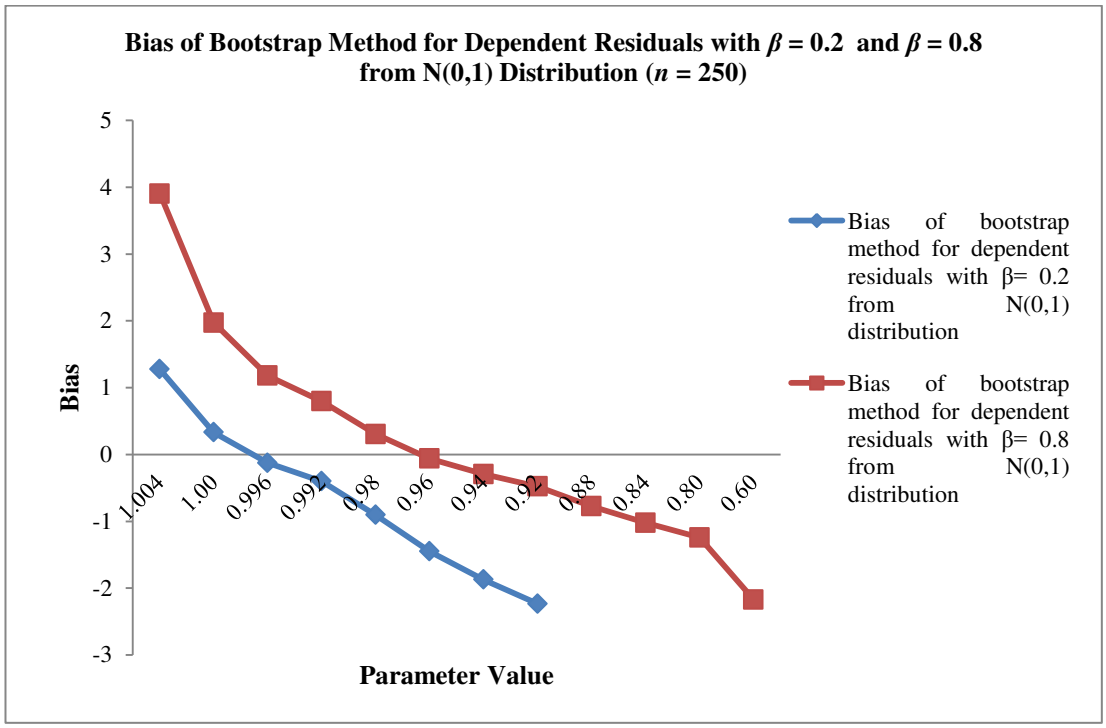

Figure 4.18 Bias of bootstrap method for dependent residuals with $\beta = 0.2$ and $\beta = 0.8$ $(n = 250)$

To show the effect of the magnitude of dependency on the distribution of the test statistics, Figure 4.19 and Figure 4.20 are given. Both these graphs are obtained by using the sample size $n = 250$ and the parameter value $\phi = 1$. These figures show that the sufficient bootstrap test statistics and the original test statistics are influenced badly in the situation of strongly dependent residuals. While the distribution of the sufficient bootstrap test statistics is normal for weakly dependent residuals, the distribution of the sufficient bootstrap test statistics tend to be piled up on several determined value for strongly dependent residuals. Also, while the distribution of the original test statistics is normal for weakly dependent residuals, the distribution of the original test statistics is skewed to the right for strongly dependent residuals.

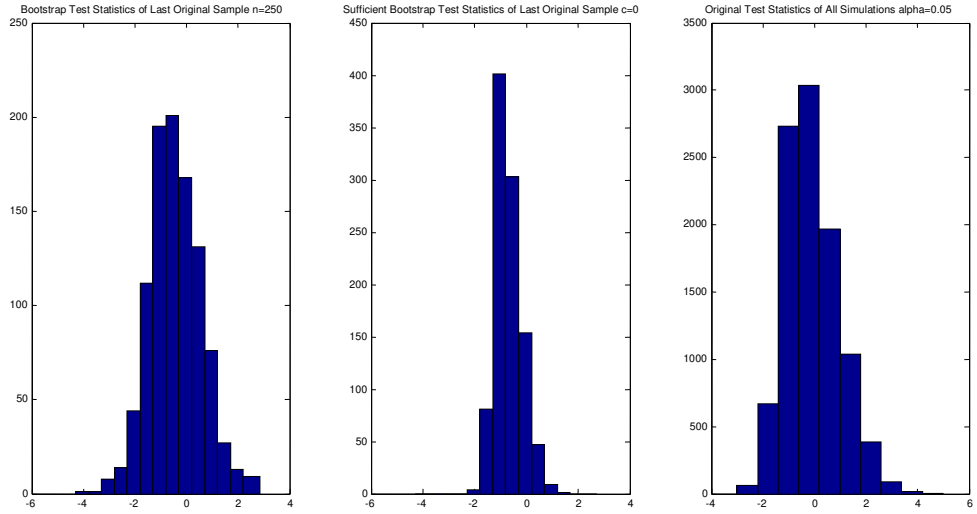

Figure 4.19 The distribution of 1,000 bootstrap test statistics of the last original sample, the distribution of 1,000 sufficient bootstrap test statistics of the last original sample, and the distribution of 10,000 original test statistics for weakly dependent residuals with $\beta = 0.2$ from $N(0,1)$ distribution, respectively $(n = 250, \phi = 1)$

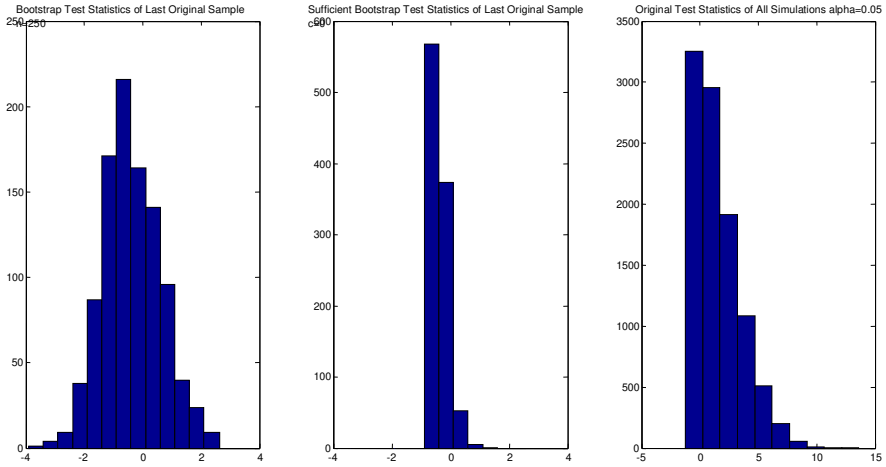

Figure 4.20 The distribution of 1,000 bootstrap test statistics of the last original sample, the distribution of 1,000 sufficient bootstrap test statistics of the last original sample, and the distribution of 10,000 original test statistics for strongly dependent residuals with $\beta = 0.8$ from $N(0,1)$ distribution, respectively $(n = 250, \phi = 1)$

CHAPTER FIVE
CONCLUSIONS

In this thesis, a group of simulation study is conducted to show the performance of asymptotic, bootstrap, and sufficient bootstrap df-AR Unit Root tests. For their powers on the unit root tests, two different situations are taken into consideration. Firstly, the residuals are supposed to be independent from each other. Secondly, the residuals are supposed to be dependent on each other. Considering these different situations, obtained results can be summarized as follows:

1. When the residuals are independent from each other and follow the $N(0,1)$ distribution, all three methods give nearly the same satisfactory results especially in the situation of the large sample size.

2. When the residuals are independent from each other and follow the $LOGN(0,1)$ distribution, asymptotic and bootstrap methods give nearly the same unsatisfactory results. The unit power is obtained for a smaller parameter value compared with $N(0,1)$ distributed residuals. For $LOGN(0,1)$ distributed residuals the power of the test increases very fast after a specific parameter value as the parameter value decreases, while for $N(0,1)$ distributed residuals the power of the test increases slowly.

3. The bootstrap bias values for $N(0,1)$ distributed independent residuals are smaller than the bootstrap bias values for $LOGN(0,1)$ distributed independent residuals for fixed values of n and ϕ.

4. The sufficient bootstrap method which is resulted in nearly 35% decrease in the sample size gives satisfactory results compared with the other methods especially when the residuals are dependent on each other. Only for $LOGN(0,1)$ distributed independent residuals, the sufficient bootstrap does not show good performance.

5. While the distribution of the sufficient bootstrap test statistics is normal for $N(0,1)$ distributed residuals, the distribution of the sufficient bootstrap test statistics tend to be piled up on a determined value for $LOGN(0,1)$ distributed residuals. Also, while the distribution of the original test statistics is normal for $N(0,1)$ distributed residuals, the distribution of the original test statistics is skewed to the left for $LOGN(0,1)$ distributed residuals.

6. When the residuals are dependent on each other, the sufficient bootstrap method gives more satisfactory results than other methods in near-unit root processes.

7. When the residuals are dependent on each other, as the quantity of dependency increases from $\beta = 0.2$ to $\beta = 0.8$, the power of the test decreases too much. The power close to one is obtained only for the sample size $n = 250$. Unlike the strong dependency, weak dependency does not affect the power of the test significantly.

8. The bootstrap bias values for weakly dependent residuals are smaller than the bootstrap bias values for strongly dependent residuals while all the other parameters, *n* and ϕ, are the same.

9. While the distribution of the sufficient bootstrap test statistics is normal for weakly dependent residuals, the distribution of the sufficient bootstrap test statistics tend to be piled up on several determined value for strongly dependent residuals. Also, while the distribution of the original test statistics is normal for weakly dependent residuals, the distribution of the original test statistics is skewed to the right for strongly dependent residuals.

10. For both independent residuals and dependent residuals, when $\phi > 1$, the power of the test is zero. For the positive values of *c*, the bootstrap bias takes the positively large values, and as the sample size increases, the magnitude of these values increase too much.

11. The bootstrap method and the sufficient bootstrap method show nearly the same pattern as regards the speed of their powers being move away from the nominal level of the test.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of Second International Symposium on Information Theory* (Eds. B. N.Petrov and F. Csaki), 267-281, Akadamiai Kiado, Budapest.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19* (6), 716-723.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics, 30* (*A*), 9-14.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrica, 66* (2), 237-242.

Andrews, D.W.K. (1999). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Cowles Foundation discussion paper, No. 1230* (Cowles Foundation for Research in Economics, Yale University).

Andrews, D.W.K., & Buchinsky, M. (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica, 68* (1), 23-51.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrica, 74* (3), 457-468.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association, 83* (403), 687-697.

Boik, R. J. (2006). *Lecture notes: Statistics 550*. Montana State University, Bozeman.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics, 16* (4), 1709-1722.

Bose, A. (1990). Bootstrap in moving average models. *Annals of the Institute of Statistical Mathematics, 42* (4), 753-768.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B. 26* (2), 211-252.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis forecasting and control* (2nd ed.). San Francisco: Holden-Day.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd ed.). New York: Springer-Verlag.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli, 3* (2), 123-148.

Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *Annals of Statistics, 26* (1), 48-83.

Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics, 14*, 1171-1179.

Chan, N.H., & Wei, C.Z. (1988). Limiting distribution of least squares estimates of unstable autoregressive processes. *Annals of Statistics, 16*, 367-401.

Chang, Y., & Park, J. Y. (2000). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis, 24* (4), 379-400.

Choi, E., & Hall, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *Journal of the Royal Statistical Society, Series B. 62* (2), 461-477.

Cornish, E.A., & Fisher, R.A. (1937). Moments and cumulants in the specification of distributions. *International Statistical Review* (5), 307-322.

Dickey, D. A. (1976). *Estimation and hypothesis testing for nonstationary time series*. Ph.D. Thesis, Iowa State University.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74* (366), 427-431.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49* (4), 1057-1072.

Dickey, D. A., & Pantula, S. G. (1987). Determining the order of differencing in AR processes. *Journal of Business and Economic Statistics, 5* (4), 455-461.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics, 7* (1), 1-26.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82* (397), 171-185.

Enders, W. (Ed.).(1948). *Applied econometric time series.* Iowa: Wiley.

Feller, W. (1957). *An introduction to probability theory and its applications, volume I*. (3rd ed.). New York: John Wiley & Sons, Inc.

Ferretti, N., & Romo, J. (1996). Unit root bootstrap tests for AR(1) models. *Biometrica, 83* (4), 849-860.

Fuller, W. A. (Ed.). (1996). *Introduction to statistical time series* (2nd ed.). New York: Wiley.

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2* (2), 111-120.

Hall, P. (1985). Resampling a coverage process. *Stochastic Process Applications, 19*, 259-269.

Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics, 14* (4), 1431-1452.

Hall, P. (Ed.). (1992). *The bootstrap and edgeworth expansion*. New York : Springer.

Hall, P., Horowitz, J.L., & Jing, B.-Y. (1995). On blocking rules on the bootstrap with dependent data. *Biometrica*, *82* (3), 561-574.

Härdle, W., Horowitz , J. L., & Kreiss, J. -P. (2002). *Bootstrap methods for time series*. Retrieved February 10, 2012, from, http://www.econstor.eu/bitstream/10419/62726/1.

Horowitz, J.L. (2001). *The bootstrap, handbook of econometrics*. *Volume 5,* Elsevier Science B.V.

Kreiss, J.-P. (1992). *Bootstrap procedures for AR(∞) processes*. *Lecture Notes in Economics and Mathematical Systems, 376* (Springer-Verlag, Heidelberg).

Kreiss, J.-P., & Paparoditis, E. (2011). Bootstrap methods for dependent data: a review. *Journal of the Korean Statistical Society, 40*, 357-378.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics, 17* (3), 1217-1241.

Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics, 27* (1), 386-404.

Paparoditis, E., & Politis, D. N. (2005). Bootstrapping unit root tests for autoregressive time series. *Journal of the American Statistical Association, 100* (470), 545-553.

Perron, P. (1989). The great crash, the oil-price shock, and the unit root hypothesis. *Econometrica*, Econometric Society, *57* (6), 1361-1401.

Perron, P., & Vogelsang, T., J. (1992).Testing for a unit root in a time series with a changing mean: corrections and extensions. *Journal of Business & Economic Statistics*, American Statistical Association, *10* (4), 467-470.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrica, 75* (2), 335-346.

Politis, D.N. ,& Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics, 22* (4), 2031-2050.

Said, S. E., & Dickey, D. A. (1984).Testing for unit roots in autoregressive-moving average of unknown order. *Biometrica, 71* (3), 599-607.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrica, 63* (1)*,* 117-126.

Singh, S., & Sedory, S. A. (2011). Sufficient bootstrapping. *Computational Statistics and Data Analysis, 55* (4)*,* 1629-1637.

Slutzky, E. (1927). The summation of random causes as the source of cyclic processes, *Econometrica, 5* (2), 105-146 (1937). Translated from the earlier paper

of the same title in *Problems of Economic Conditions* (Ed. Conjuncture Institute Moscow).

Swensen, A. R. (2003). A note on the power of bootstrap unit root tests. *Econometric Theory, 19* (1), 32-48.

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis, in *Advanced Seminar on Spectral Analysis* (Ed. B. Harris), 25-46, John Wiley, New York.

Wei, W. (2006). *Time series analysis-univariate and multivariate methods* (2nd ed.). The United States of America: Pearson Education, Inc.

Wikipedia (a). Dickey-Fuller test. Retrieved March 10, 2013 from http://en.wikipedia.org/wiki/Dickey%E2%80%93Fuller_test.

Wikipedia (b). Dickey-Fuller test. Retrieved March 10, 2013 from http://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test.

Wikipedia (c). Martingale. Retrieved April 15, 2013 from http://en.wikipedia.org/wiki/Martingale_(probability_theory).

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society, London*, *226* (*A*), 267-298.

APPENDICES

MATLAB R2011b programme codes for the conducted simulation study

```
function[asy_sign_level,boot_sign_level,suff_boot_sign_level,Bootstrapbias,mean_l
ength_for_suffboot]=DF_boot_formula_thesis(n,S,B,c,alpha)
% residuals from N(0,1) distribution
bias_for_bootstrap=[];
TotalH0=0;TotalH1=0;count=0;count_suff=0;
TS=[];
suff_mean_length=[];
a=1+(c/n)
for s=1:S;
disp('simulation='),disp(s)
Zt=randn(n,1);
%Generation of AR(1) series
X_vec=filter(1,[1 -a],Zt);
X=X_vec;
%Dickey-Fuller Unit Root Test results for the original sample
[h,pValue,stat,coeff,reg] = adftest(X,'lags',0,'alpha',0.05,'model','AR');
zt_new = getfield(reg,'res');
zt_new=[X(1);zt_new];
zt_new=zt_new-mean(zt_new);
if h==0;
 TotalH0=TotalH0+1;
else
 TotalH1=TotalH1+1;
end;
%Test Statistic for the original sample
TS_original=stat;
TS=[TS;TS_original];
%Bootstrapping the residuals
```

```
zt_star=zt_new(ceil(rand(length(zt_new),B)*length(zt_new)));
%Bootstrapping the series
X_star=filter(1,[1 -1],zt_star);
X_star(1,:)=X(1);
TS_boot=[];
TS_boot_suff=[];
for i=1:B
[h,pValue,stat_boot] = adftest(X_star(:,i),'lags',0,'alpha',0.05,'model','AR');
TS_boot=[TS_boot;stat_boot];
end
meanTestStat_boot=mean(TS_boot);
bias_for_bootstrap=[bias_for_bootstrap;TS_original-meanTestStat_boot];
%percentile for (100*alpha)%(the bootstrap critical value)
Bootstrap_critical_value(s)=prctile(TS_boot,[(alpha)*100]);
%for the sufficient bootstrap
length_suff=[];
for i=1:B
zt_star_suff=zt_new(unique(ceil(rand(length(zt_new),1)*length(zt_new))));
%Bootstrapping the series
X_star_suff=filter(1,[1 -1],zt_star_suff);
X_star_suff=[X(1); X_star_suff];
length_suff=[length_suff;length(X_star_suff)];
[h,pValue,stat_boot_suff] = adftest(X_star_suff,'lags',0,'alpha',0.05,'model','AR');
TS_boot_suff=[TS_boot_suff;stat_boot_suff];
end
suff_mean_length=[suff_mean_length;mean(length_suff)];
%percentile for (100*alpha)%(the sufficient bootstrap critical value)
Bootstrap_critical_value_suff(s)=prctile(TS_boot_suff,[(alpha)*100]);
end;
for s=1:S
if TS(s,1)<mean(Bootstrap_critical_value);
count=count+1;
```

```
end
end
for s=1:S
if TS(s,1)<mean(Bootstrap_critical_value_suff)
count_suff=count_suff+1;
end
end
Bootstrapbias=mean(bias_for_bootstrap);
mean_length_for_suffboot=mean(suff_mean_length);
asy_sign_level=TotalH1/S;%if a=1 then equals to alpha, otherwise equals to power
boot_sign_level=count/S;
suff_boot_sign_level=count_suff/S;
minimum=min(min(TS_boot),min(TS_boot_suff));
maximum=max(max(TS_boot),max(TS_boot_suff));
binwidth = 0.5;
bins=minimum:binwidth:maximum;
subplot(1,3,1); hist(TS_boot,bins)
title('Bootstrap Test Statistics of LastOriginalSample n=30')
subplot(1,3,2); hist(TS_boot_suff,bins)
title('SufficientBootstrap Test Statistics of LastOriginalSample c=-15')
subplot(1,3,3); hist(TS)
title('Original Test Statistics of All Simulations alpha=0.05')
end
```

(Please note that the values of n, and c have to be changed for the first two figures.)