ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

DICKEY – FULLER (ADF) UNIT ROOT TESTING: METHODS FOR THE LAG LENGTH SELECTION

By

Christodoulos C. Komninakidis

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece

June 2004

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΕΛΕΓΧΟΙ DICKEY – FULLER (ADF) ΓΙΑ ΜΟΝΑΔΙΑΙΑ ΡΙΖΑ: ΜΕΘΟΔΟΙ ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΤΗΣ ΤΑΞΗΣ ΤΗΣ ΑΥΤΟΠΑΛΙΝΔΡΟΜΗΣ ΕΞΙΣΩΣΗΣ ADF

Χριστόδουλος Χ. Κομνηνακίδης

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Ιούνιος 2004

ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

DICKEY-FULLER (ADF) UNIT ROOT TESTING: METHODS FOR THE LAG LENGTH SELECTION

Christodoulos C. Komninakidis

Approved by the Graduate Committee

E. Ioannidis Lecturer Thesis Supervisor	N. Fragkos Professor	I. Vrpntos Lecturer Members of the Commitee

Athens, November 2004

Michael Zazanis, Associate Professor
Director of the Graduate Program

DEDICATION

I consider that the most appropriate person to start my thank from is my brother, **Ioannis Komninakidis**. Dear brother, thank you for supporting me to finish my postgraduate studies, thank you for encouraging me in all difficult moments and commending me in the successes!

I would also thank my parents, **Chrysovergis Komninakidis** and **Efrosini Filippopoulou**, not only for their help, support and advice during my whole studies but also for their unique way to be always "close" to me!

Thank you very much and I share my happiness with you!

ACKNOWLEDGEMENTS

I would like to thank my supervisor **Evangelos Ioannidis** for his valuable help and advice during the development of this Thesis, for his contribution to the whole project, offering his own ideas, criticizing my ideas and making valuable comments on this Thesis. His role has been extended beyond the limits of this Thesis. He was always close to me, made me feel comfortable and helped me to build a scientific way of thinking.

For all your help and support: Thank you!

VITA

I was born in Athens on 24th December 1975

I was awarded the Degree of Mathematics from the Department of Mathematics of the University of Ioannina on 14th April 2000. During my studies I was attended and specialized in the subjects of the major of Statistics and Operational Researches and I was awarded the certificate of Statistics and Operational Researches.

In 2000 I was accepted by the International Full – Time M.Sc. Program of Statistics of the Department of Statistics of Athens University of Economics and Business. The present Thesis is the last requirement for the Degree of Master of Science in Statistics. My research interests are Time Series Analysis and Econometrics.

ABSTRACT

Christodoulos Komninakidis

DICKEY – FULLER (ADF) UNIT ROOT TESTING: METHODS FOR THE LAG LENGTH SELECTION

June 2004

In this thesis we discuss various methods for the selection of the number of lagged differences (lag length selection) that have to be included in an Augmented Dickey-Fuller autoregression in the context of testing the Hypothesis of a Unit root. The asymptotic properties of the test statistics under various lag length selection procedures proposed in the Literature are presented. Moreover, we discuss Monte Carlo simulation experiments from the Literature, in order to investigate the behavior of the various methods in finite samples. Finally, we conduct our own comparative Monte Carlo simulation study.

In the theoretical part, as well as in the simulation experiment, we focus particularly on a class of Modified Information Criteria (MICC) proposed by Ng and Perron (2001). These Criteria yield tests, which seem to be able to keep the nominal size much better than the Standard Information Criteria (AIC, BIC) do.

Our own Monte Carlo simulation analysis shows that the Modified Information Criteria have desirable size and power properties for the Dickey-Fuller unit root test statistics and in particular we (a) confirm the well known result that the size of the unit root test statistics is being significantly distorted when there is strong and negative serial correlation of the moving average type at frequency zero; this is less so with the Modified Akaike's Information Criterion (MAICC) which has the best performance, (b) we also show that the size and power properties of unit root test statistics are not significantly affected when there is serial correlation of the autoregressive type with any of the lag length selection procedures. A new result from our simulation is that (c) under autoregressive moving average error processes, when a strong moving average polynomial root at frequency zero is combined with a strong autoregressive polynomial root elsewhere, the size distortions of the unit root test statistics for all the lag length selection methods are reduced the stronger the autoregressive polynomial root.

ΠΕΡΙΛΗΨΗ

Χριστόδουλος Κομνηνακίδης

ΕΛΕΓΧΟΙ DICKEY – FULLER (ADF) ΓΙΑ ΜΟΝΑΔΙΑΙΑ ΡΙΖΑ: ΜΕΘΟΔΟΙ ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΤΗΣ ΤΑΞΗΣ ΤΗΣ ΑΥΤΟΠΑΛΙΝΔΡΟΜΗΣ ΕΞΙΣΩΣΗΣ ADF

Ιούνιος 2004

Στην παρούσα διατριβή αναλύουμε την επιλογή της τάξης της Αυτοπαλίνδρομης Εξίσωσης των Dickey – Fuller (ADF) στα πλαίσια του ελέγχου της Υπόθεσης της μοναδιαίας ρίζας. Παρουσιάζονται οι ασυμπτωτικές ιδιότητες των στατιστικών ελέγχων Dickey – Fuller όταν χρησιμοποιούν διάφορες μεθόδους επιλογής της τάξης της Αυτοπαλίνδρομης Εξίσωσης. Επιπλέον, παρουσιάζονται και αναλύονται Monte Carlo πειράματα προσομοίωσης που ήδη έχουν αναφερθεί στην Βιβλιογραφία με σκοπό την εξερεύνηση της συμπεριφοράς των παραπάνω στατιστικών ελέγχων σε πεπερασμένα δείγματα δεδομένων. Στο τέλος παρουσιάζουμε δικά μας συγκριτικά Monte Carlo πειράματα προσομοίωσης.

Στο θεωρητικό και εμπειρικό μέρος της διατριβής, δίνεται έμφαση σε μια οικογένεια τροποποιημένων κριτηρίων (MICC) για την επιλογή της τάξης της Αυτοπαλίνδρομης Εξίσωσης ADF που προτάθηκε από τους Ng και Perron (2001). Τα συγκεκριμένα κριτήρια φαίνεται να έχουν πόλυ καλύτερη απόδοση από την άποψη των Σφαλμάτων τύπου I και II από τα κλασσικά κριτήρια όπως το AIC και το BIC στα πλαίσια του ελέγχου υποθέσεων για μοναδιαία ρίζα.

Τα δικά μας Monte Carlo πειράματα προσομοίωσης δείχνουν ότι τα τροποποιημένα κριτήρια (MICC) οδηγούν τα στατιστικά τεστ των Dickey – Fuller σε επιθυμητές στατιστικές ιδιότητες και συγκεκριμένα (α) επιβεβαιώνουμε το γνωστό από τη βιβλιογραφία αποτέλεσμα ότι έχουμε σημαντικές αποκλίσεις του σφάλματος Τύπου I από την ονομαστική τιμή στην περίπτωση που υπάρχει ισχυρή αρνητική αυτοσυσχέτιση τύπου Κινητού Μέσου στην συχνότητα μηδέν. Το τροποποιημένο κριτήριο του Akaike (MAICC) παρουσιάζει τις μικρότερες αποκλίσεις και έχει την καλύτερη

απόδοση. Δείχνουμε ότι (β) όποιο κριτήριο και να εφαρμόσουμε δεν έχουμε σημαντικές αποκλίσεις του σφάλματος Τύπου Ι στην περίπτωση της ύπαρξης αυτοσυσχέτισης Αυτοπαλίνδρομου τύπου και τέλος (γ) στην περίπτωση αυτοσυσχέτισης τύπου Αυτοπαλίνδρομου Κινητού Μέσου, όταν συνδυάζεται μια ισχυρή ρίζα τύπου Κινητού Μέσου στην συχνότητα μηδέν με μια ισχυρή ρίζα Αυτοπαλίνδρομου τύπου σε οποιαδήποτε άλλη συχνότητα εκτός της μηδενικής τα Σφάλματα Τύπου Ι των στατιστικών τεστ των Dickey – Fuller μειώνονται ικανοποιητικά.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Many observed time series and especially economic and financial time series display a trend implying nonstationary behavior. Others appear to wander around as if they have no fixed population mean implying also nonstationary behavior. An enormous number of studies deals with the question whether e.g. time series exhibit linear time trend and in the case where the answer is positive what is the appropriate method for its elimination. One of the easiest ways to analyze such time series is to make these time series stationary. When a nonstationary series can be transformed to a stationary series by taking the first differences, the series is said to be "integrated of order 1" and is denoted by I(1). If the series needs to be differenced k times to be stationary, then the series is said to be I(k). The I(k) series ($k \neq 0$) is also called a "difference-stationary" process. Another important class are "trend-stationary" processes, meaning that they are stationary around a deterministic time trend and can be transformed to stationarity by regressing them on time.

The distinction between these two competing processes is emphasized in a highly influential study by Nelson and Plosser (1982). In this study, the authors clearly presented theoretical and empirical evidence that many important economic and financial time series are better characterized by difference-stationary processes. Since the extensive study by Nelson and Plosser (1982) evidence of the difference-stationary behavior in many time series was established and Nelson and Plosser's data set was used as example data set.

Dickey (1976) and Dickey and Fuller (1979, 1981) were the first who dealt with the problem of the discrimination between trend-stationary and difference-stationary processes. In particular, they approach it as the following hypothesis testing problem: Assume that the true process that generates the data of a time series under study satisfies an equation of the form:

$$y_t = d_t + \rho y_{t-1} + \varepsilon_t$$

where

$$\{\varepsilon_t\} \sim \text{i.i.d.}(0,\sigma^2) \tag{1}$$

where the error process $\{\varepsilon_t\}$ is assumed to be independent identically distributed with mean zero and constant variance σ^2 and d_t represents a set of deterministic components such as constant and/or time trend (linear, quadratic e.t.c.). Dickey (1976) and Dickey and Fuller (1979, 1981) confined themselves to a discussion of the cases of a zero intercept, a constant, and a linear time trend.

The time series defined in equation (1) is said to have an "autoregressive unit root" when $\rho = 1$. With $\rho = 1$, $\{\varepsilon_t\}$ is stationary and $\{\Delta y_t\}$ is stationary around the change in the deterministic part. In this case $\{y_t\}$ is said to be integrated of order 1 and is denoted by I(1). Stationary time series are said to be I(0). When $\{y_t\}$ is stationary and is thus I(0), and d_t is a linear time trend, then

$$y_t = d_t + \varepsilon_t$$

and the first-difference of $\{y_t\}$ is

$$\Delta y_t = \alpha + \varepsilon_t - \varepsilon_{t-1}$$

where $\alpha = d_t - d_{t-1}$ is constant. Thus $\{\Delta y_t\}$ has a moving average unit root. Moving average unit roots arise if a stationary time series differenced.

Tests using I(1) as the null hypothesis are based on the autoregressive unit root as null. Tests using I(0) as the null hypothesis are based on the unit moving average root as the null. Dickey (1976) and Dickey and Fuller (1979, 1981) considered test statistics for the autoregressive unit root since these are the ones most common. They actually proposed two test statistics, the Z_{DF} – test statistic and the τ_{DF} – test statistic, on the coefficient of y_{t-1} for the above testing purpose. These are the "standard Dickey-Fuller unit root test statistics". Dickey (1976) and Dickey and Fuller (1979, 1981) showed that ordinary asymptotic distribution theory breaks down in the case in which ρ is precisely equal to unity ($\rho=1$), and that the two proposed Dickey-Fuller unit root test statistics on the coefficient of y_{t-1} require to be compared with critical values taken from non-standard limiting distributions, the so called "Dickey-Fuller distributions", which differ depending on whether d_t in equation (1) is a constant or a time trend.

The assumption that the error process $\{\varepsilon_t\}$ in equation (1) is independent identically distributed is quite unrealistic for time series encountered in practice. Dickey and Fuller also considered the case of testing the unit root hypothesis that $\rho=1$ in more complicated models than the one described in equation (1), namely in models that are described as follows:

$$y_t = d_t + \rho y_{t-1} + u_t$$

where

$$u_t = n_1 u_{t-1} + n_2 u_{t-2} + \ldots\ldots + n_p u_{t-p} + \varepsilon_t \qquad (2)$$

with the characteristic that the error process $\{u_t\}$ follows an autoregressive process that has independent identically distributed innovations of bounded fourth moment, say AR(p) with p>1, the order of which is required to be finite and known. The test statistics concerning the unit root hypothesis under these models are the so called "Augmented Dickey-Fuller unit root test statistics" and their limiting distributions, the "Dickey-Fuller distributions", appeared to be the same as in the case where the error process is assumed to be independent identically distributed. The Augmented Dickey-Fuller unit root test statistics – which take into account possible serial correlation of the error process – are based on the estimation of the following "Augmented Dickey-Fuller autoregression":

$$y_t = d_t + \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots\ldots + \zeta_p \Delta y_{t-p} + \varepsilon_t \tag{3}$$

A statement that has to be emphasized is that the number of the additional lagged-difference terms of the dependent variable y_t that have to be included in the Augmented Dickey-Fuller autoregression depends on the true order of the error process. In fact, if we knew that the errors follow an autoregressive process with a known order p, then p lagged-difference terms of the dependent variable y_t should be included in the Augmented Dickey-Fuller autoregression. Naturally, a key shortcoming for applying the Dickey-Fuller as well as the Augmented Dickey-Fuller unit root test statistics in order to test the unit root hypothesis is that the practitioner must have knowledge about the order of the error process, knowledge that is not possible to get in practice since the true mechanism that generates the data is in most cases unknown.

Said and Dickey (1984) extended the Dickey-Fuller theory concerning the unit root hypothesis by allowing the error process in equation (2) to belong to the general class of ARMA(p, q) processes without needing to have any knowledge on the parameters p and q:

$$y_t = d_t + \rho y_{t-1} + u_t$$

where

$$u_t = \sum_{i=1}^{p} \varphi_i u_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{4}$$

where $\{\varepsilon_t\}$ is independent identically distributed innovations with zero mean, constant variance and bounded fourth moment. Assuming that $\{u_t\}$ is stationary and invertible with autoregressive and moving average polynomials that do not share common roots, $\{y_t\}$ evolves according to:

$$\Delta y_t = (\rho - 1) y_{t-1} + \sum_{i=1}^{\infty} \zeta_i u_{t-i} + \varepsilon_t \tag{5}$$

where the parameters $\{\zeta_i\}$ are appropriate functions of the parameters $\{\varphi_i$ and θ_j i=1, 2,…,p , j=1, 2,…..,q$\}$. The true order of the autoregression (5) is infinity when q>0. Because $\Delta y_t = u_t$ under the unit root hypothesis, (5) can be seen as an autoregression in Δy_t augmented by y_{t-1}, namely:

$$\Delta y_t = (\rho - 1) y_{t-1} + \sum_{i=1}^{\infty} \zeta_i \Delta y_{t-i} + \varepsilon_t \tag{6}$$

which is actually the "Augmented Dickey-Fuller autoregression" but with an infinite order. When the orders p and q are unknown, as it is the case in practice, Said and Dickey (1984) suggested approximating the infinite "Augmented Dickey-Fuller autoregression" by a truncated version whose order is a function of the number of observations T:

$$\Delta y_t = (\rho - 1) y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_t \tag{7}$$

Said and Dickey (1984) showed that when the truncation lag number k in equation (7) satisfies the conditions:

(A1) k is chosen as a function of T such that

$$\frac{k^3}{T} \longrightarrow 0 \text{ and } k \longrightarrow \infty \text{ as } T \longrightarrow \infty$$

(A2) There exist c>0 and r>0 such that $ck > T^{1/r}$

then the Augmented Dickey-Fuller unit root test statistics have the same asymptotic Dickey-Fuller distributions as in the case where the errors are independent identically distributed.

Ng and Perron (1995) also considered the conditions given by Said and Dickey (1984) and believed that are unnecessarily stringent in the sense that they can be a serious limitation for the application of the Augmented Dickey-

Fuller unit root test statistics. Ng and Perron (1995) proved the validity of Said and Dickey's results in the absence of the second condition (A2) making the Augmented Dickey-Fuller unit root tests theoretically more flexible. But, such conditions provided theoretical rather than practical guidelines for choosing the appropriate order for the Augmented Dickey-Fuller autoregression. So, a question that now arises and is of great interest for practical purposes is the following one:

How many lagged-difference terms of the dependent variable y_t should be included in the Augmented Dickey-Fuller autoregression in practice, provided that the conditions proposed by Said and Dickey are satisfied, before testing the unit root hypothesis?

If we add too many lagged-difference terms of the dependent variable y_t in the Augmented Dickey-Fuller autoregression the Augmented Dickey-Fuller unit root test statistics loose power because the estimate of the parameter on y_{t-1} becomes inaccurate. If we add too few the Augmented Dickey-Fuller unit root tests have eventually the wrong size (the probability of rejecting the null hypothesis of a unit root when it is true). Many studies, theories and Monte Carlo simulation experiments such that of Schwert (1989), Agiakloglou and Newbold (1991), Harris (1992) and Ng and Perron (1995, 2001) arisen concerning the issue of the lag length selection for the Augmented Dickey-Fuller autoregression so that the Augmented Dickey-Fuller unit root test statistics to have robust size and power properties. This is an active research area, and because the selection method of adjustment for serial correlation is not exact, several lag length selection methods (with different advantages and disadvantages) have been proposed. Some of the most popular and widely used selection methods for choosing the appropriate order for the Augmented Dickey-Fuller autoregression provided that they satisfy the conditions proposed by Said and Dickey (1984) are the following:

(a) Start with a large number k for the order of the Augmented Dickey-Fuller autoregression and test the significance of the largest lagged-difference term, using the ordinary t-test, referred to the usual critical value. If this is accepted re-estimate the Augmented Dickey-Fuller autoregression with the order reduced by 1 and repeat the test. Stop when the largest lag-differenced term is statistically significant.

(b) Estimate processes with every possible number of lagged-difference terms up to a maximum lag-differenced term, and select the order for the Augmented Dickey-Fuller autoregression that minimises a "Information Selection Criterion" such that of Akaike Information Criterion (AIC) or that of Schwarz Bayesian Information Criterion (BIC).

In this thesis, the asymptotic properties of the Augmented Dickey-Fuller unit root test statistics under the lag length selection methods of the type just mentioned for selecting the order for the Augmented Dickey-Fuller autoregression are going to be analysed, and Monte Carlo simulation experiments are going to be used to show their distinctive behavior in finite samples.

Monte Carlo simulation studies have repeatedly shown a strong association between the order of the Augmented Dickey-Fuller autoregression and the severity of size distortions and/or the extent of power loss of the Augmented Dickey-Fuller unit root test statistics. This association is more visible in the case where the underlying data generating process contains moving average error components.

Lag length selection methods such that of AIC and BIC tend to select values for the order of the Augmented Dickey-Fuller autoregression that are generally too small for the Augmented Dickey-Fuller unit root test statistics to have good size and power properties especially when the error process contains moving average components. The analysis shows that the lag length selection method based on the ordinary t –test for the significance on the last lagged-difference term has the ability to yield higher orders for the Augmented Dickey-Fuller autoregression than the AIC or BIC and reduce the size distortions of the Augmented Dickey-Fuller unit root test statistics. But, the lag length selection method based on the ordinary t –test tends to overparamererize data generating processes with errors that follow autoregressive processes in some cases, a result that leads to power losses. So, neither approach is fully satisfactory and an improved selection procedure for choosing the order for the Augmented Dickey-Fuller autoregression is needed.

Ng and Perron (2001) having firstly reconsidered issues related to standard Information Criteria such that of AIC and BIC for valid model comparisons they developed a class of Modified Information Criteria, denoted by MICC. These Modified Information Criteria have the key advantages that (a) they have a penalty factor that retains a stochastic term (b) take into account the fact that the bias in the sum of the autoregressive coefficients is highly dependent on the order of the Augmented Dickey-Fuller autoregression and (c) adapt to the type of the possible deterministic terms that might be included in the underlying data generating process. More precisely, the Modified Akaike Information Criterion, denoted by MAICC, is shown to lead to substantial size and power improvements for the Augmented Dickey-Fuller unit root test statistics when is used to select the order of the Augmented Dickey-Fuller autoregression compared to all the other lag length selection methods especially when the data generating process contains moving average error components.

In Chapter 2 up to Chapter 4 we are concerned with presenting the background of the univariate processes that contain a unit root. More specifically, the purpose of Chapter 2 is to present nonstationary univariate time series and how we can model them using trend-stationary or difference-stationary processes as well as to shortly compare them from the forecasting point of view in order to see their different implications in finite samples. Chapter 3 is devoted to nonstationary univariate processes that are characterized by difference-stationary processes and to unit root hypothesis testing problem in equation (1) and (2) under independent identically distributed and autoregressive error processes of finite and correctly specified order, respectively. Chapter 4 focuses to the extensions of the Augmented Dickey-Fuller unit root test statistics in the case in which the error process that described by equation (3) belongs to the general class of stationary and invertible ARMA processes the orders of which are assumed to be unknown as it is in practice.

Chapters 5 and 6 are the main focus of this thesis. The issue of choosing the appropriate order of the Augmented Dickey-Fuller autoregression using the various selection methods of the type mentioned above in order the Augmented Dickey-Fuller unit root test statistics to have robust and satisfactory size and power properties from the theoretical point of view is presented in Chapter 5. Chapter 6 is primarly devoted to the construction and extensive presentation of own as well as published Monte Carlo simulation experiments in order to examine the size and the power properties of the Augmented Dickey-Fuller unit root test statistics in finite samples when ordinary t – tests of various significance nominal levels and standard, Corrected and Modified Information Criteria selection methods are used for choosing the order for the Augmented Dickey-Fuller autoregression. Their majority is mainly focused on testing the unit root hypothesis in data generating processes of the form described in equations (1), (2) and (4) (with $d_t=0$) in the cases in which the errors follow a moving average and autoregressive process of order 1, denoted by MA(1) and AR(1), respectively, because these cases occur frequently in economic and financial data.

To provide a deep insight into the issue of testing unit root hypotheses in finite samples and choosing the appropriate order for the Augmented Dickey-Fuller autoregression in order the Augmented Dickey-Fuller unit root test statistics to have robust and satisfactory size and power properties we study the error processes from their spectral features perspective. More specifically, it is well known that the true spectrum of a moving average process of order 1 indicates a ‘hole’ at the frequency zero that is ‘stronger’ the closer the moving average polynomial root is to unity. It is also well known that the true spectrum of an autoregressive process of order 1 indicates a ‘peak’ at the frequency zero that is ‘stronger’ the closer the autoregressive polynomial root is to unity. In the first case where the errors are modelled by a MA(1) process, it has been noted that the Augmented Dickey-Fuller unit root test statistics seem unable to keep the nominal size at a permissible percentage rate when the underlying data generating process is indeed a unit root process and this inability increases the stronger the “hole” is at frequency zero. On the other hand, in the second case where the errors are modelled by an AR(1) process, the Augmented Dickey-Fuller unit root test statistics are able to keep the nominal size at a permissible percentage rate when the actual data generating process is a unit root process. The above statements lead to the conclusion that the inability of the Dickey-Fuller unit root test statistics to keep the nominal size at a permissible percentage rate when the actual data generating process is a unit root process is due to strong and negative serial correlation of the moving average type, i.e. a strong moving average polynomial root close to unity indicating a strong "hole" at frequency zero in the true spectrum of the error process. A question that is now arisen is the following:

What can we tell about the size and power performance of the unit root test statistics in the case where the actual data generating process that described in equations (1), (2) and (4) with moving average or autoregressive errors that are modeled by more complicated than a MA(1) or AR(1) processes?

or similarly

What can we tell about the size and power performance of the unit root test statistics in the case where the root/roots of the moving average and/or autoregressive polynomial of the error process are close to unity, indicating 'holes' and 'peaks' in the true spectrum of the error process, at frequencies different from zero as well as at various combinations of them including the frequency zero?

To answer these questions that are of great interest in the context of testing unit root hypotheses in finite samples we construct and fully present additional Monte Carlo simulation experiments that are combined with Figures in order for us to derive more robust and satisfactory averall conclusions.

According to many authors, there are many reasons to believe that economic and financial time series encountered in practice contain moving average components. Our Monte Carlo simulation experiments provide a useful tool for the broad application of the Dickey-Fuller and the Augmented Dickey-Fuller unit root test statistics.

CHAPTER 2

NONSTATIONARY UNIVARIATE PROCESSES

2.1 INTRODUCTION

Many observed time series seem to display nonstationary behavior. For economic time series, nonstationary behavior is often the most dominant characteristic. Some series grow in a secular way over long periods of time implying trending (nonstationary) behavior and others appear to wander around as if they have no fixed population mean implying nonstationarity in the mean.

Figures (A.1) through (A.4) in Appendix A represent (a) the total U.S. Industrial production index from 1990 up to 2003 after seasonal adjustment (b) the E.U. Exports at constant prices (in EUR billions) from 1990 up to 2003 after seasonal adjustment (c) the Unemployment Rate in Alaska (percent) from 1990 up to 2003 after seasonal adjustment and (d) the U.S. Consumers' Inflation Expectations (percent) from 1990 up to 2003 which clearly show such patterns.

Growth characteristics are especially evident in time series that represent aggregate economic behavior like gross domestic product and industrial production. Random wandering behavior is evident in many financial time series like interest rates and asset prices. Similar phenomena also arise in data from other sciences like communications and political sciences. Any attempt to explain or forecast series of this type requires a mechanism to be introduced to capture the nonstationary elements in the series, or the series have to be transformed in some way to achieve stationarity. Yet this is often much easier to say than it is to do in a satisfactory way. The problem is more cumbersome in the multivariate case, where several time series may have nonstationary characteristics and the interrelationships of these variables are the main object of study.

Nelson and Plosser (1982) have proposed to distinguish nonstationary time series of the type just mentioned from the fundamental nature of their underlying trend and, as far as this underlying trend is concerned, they mentioned that the tendency of economic time series to exhibit variation that increases in mean and dispersion in proportion to absolute level, motivates the transformation to natural logs and the assumption that trends are linear in the transformed data (Nelson and Plosser, 1982, p.141). Indeed, many economic and financial time series seem better characterized by a linear and mostly by an exponential time trend. In the latter case notice that if we take the natural log of the exponential time trend, the result is again a linear time trend. So, the linear specification of the deterministic trend is going to be used and its coefficients are assumed to be constant over the sample period. More specifically, the deterministic time trend, denoted hereafter by d_t, can be any of the following:

$$d_t = 0 \text{ (zero) },\ d_t = \alpha \text{ (constant) and } d_t = \alpha + \delta t \text{ (linear time trend)}$$

Some series should be best described as having a deterministic time trend meaning that are completely predictable (if we know the coefficient of time), while others should rather be described as having a stochastic trend meaning that increase systematically as t goes to infinity (sometimes systematically decreases) and there is no force to move toward their mean. The fundamental difference between the two is that the former is such that the time series is mean-reverting while the latter have no such inherent tendency. Nelson and Plosser (1982) also provided evidence that nonstationary time series of the type just mentioned are better characterized by stochastic trends. The statistical properties of nonstationary time series that are captured by stochastic trends are crucial in statistical and economic applications and will be of prime concern in this thesis.

This chapter follows Hamilton (1994, chapter 15, p.435-447) and proceeds as follows. Section 2.2 is devoted to modeling nonstationary univariate processes that indicate linear trending (nonstationary) behavior using trend-stationary and difference-stationary processes and to distinguishing them theoretically with section 2.3 being involved to compare them from the forecasting point of view in order to see their asymptotical implications as well as their implications in finite samples giving both theoretical and visual examples. In the end of this Chapter some remarks concerning these two competing processes are provided.

2.2 MODELS FOR NONSTATIONARY UNIVARIATE PROCESSES

It is widely known from the Wold's decomposition (Hamilton 1994, p.108-109) that an observed stationary time series can be written in the form:

$$y_t = \mu + \psi(L)\varepsilon_t \tag{2.2.1}$$

where $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ with $\psi_0 = 1$, $\{\varepsilon_t\}$ is a White Noise process with zero mean and σ^2 variance, denoted by $\{\varepsilon_t\} \sim WN(0, \sigma^2)$, μ is the unconditional mean of the series, L is the lag operator for which $Ly_t = y_{t-1}$, and the roots of the polynomial $\psi(z) = 1+\psi_1 z+ \psi_2 z^2+ \psi_3 z^3+\ldots\ldots$ are outside the complex unit circle meaning that:

$$\forall z \in C \quad \text{with } |z| \le 1 \text{ we have that } \psi(z) \ne 0$$

Finding the Wold representation in principle requires fitting an infinite number of parameters $\{\psi_1, \psi_2, \ldots..\}$ to the data. With finite number of observations this will never be possible. As a practical matter, we therefore need to make some additional assumptions about the nature of $\{\psi_1, \psi_2, \ldots..\}$. A typical assumption is that $\psi(L)$ can be expressed as a ratio of two finite-order polynomials:

$$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j = \frac{\theta(L)}{\varphi(L)} = \frac{1+\theta_1 L+\theta_2 L^2+\ldots\ldots+\theta_q L^q}{1-\varphi_1 L^1-\varphi_2 L^2-\ldots\ldots-\varphi_p L^p}$$

where $\varphi(L)$ and $\theta(L)$ both have all the roots outside the unit circle.

For the process of the form (2.2.1) it holds that:

$$E(y_t) = \mu \quad \text{and} \quad Var(y_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

The forecasting methods applied to this representation of stationary time series imply that if we try to forecast them farther into the future, the forecast is the unconditional mean of the series:

$$\lim_{s \to \infty} \hat{y}_{t+s/t} = \mu$$

for which we know that is independent of time.

The mean squared error (MSE) for this forecast is the unconditional variance of the series. So, the process of the form (2.2.1) is impossible to capture the trending behavior of a time series or the random fluctuation that implies nonstationarity in the mean.

The question that arises at this point is the following:

How can we model time series that indicate trending (nonstationary) behavior or random fluctuation that implies nonstationarity in the mean since the general class of the processes of the form (2.2.1) is not capable?

For the case of linear trending time series there are many approaches that can be considered appropriate to handle the situation. However, there are two approaches that are widely known and used in practice. The first widely known approach is to include a deterministic linear time trend instead of a mean in the stationary process of the form (2.2.1):

$$y_t = a + \delta t + \psi(L)\varepsilon_t \tag{2.2.2}$$

Such a series has the unconditional mean equal to:

$$E(y_t) = \alpha + \delta t$$

which is the deterministic trend of the series .

A process of the form (2.2.2) that exhibits a deterministic time trend is called **trend-stationary** meaning that is stationary around the deterministic time trend. A common method to get a stationary process from (2.2.2) is to substract the trend a+δt. Therefore, recalling the Box-Jenkins theory of stationary processes we can do the analogous statistical inference (Brockwell and Davis, 1996).

The second widely known approach is to consider a **difference-stationary** process, the so called **unit root process** with drift term, satisfying the equation:

$$y_t = y_{t-1} + \delta + \psi^*(L)\varepsilon_t$$

or

$$(1-L)y_t = \delta + \psi^*(L)\varepsilon_t \tag{2.2.3}$$

where δ denotes the mean of $(1-L)y_t$ and $\psi^*(L) = \dfrac{\theta^*(L)}{\varphi^*(L)}$ with $\varphi^*(L)$ being stationary and $\theta^*(L)$ being invertible .

Many scientists like Nelson and Plosser (1982) and D.N.Dejong, J.C.Nankervis, N.E.Savin and C.H.Whiteman (1992) included in their studies detailed discussions about the difference between these two widely known competing processes.

The issue now becomes to relate the processes of the form (2.2.2) and (2.2.3). To do this let us consider the case in which:

$$y_t = \alpha + \delta t + u_t \tag{2.2.4}$$

where u_t follows a zero mean ARMA(p , q) process :

$$(1-\phi_1 L-\phi_2 L^2-\ldots.-\phi_p L^p)u_t =$$

$$(1+\theta_1 L+\theta_2 L^2+\ldots\ldots+\theta_q L^q)\varepsilon_t$$

(2.2.5)

where the moving average operator $(1+\theta_1 L+\theta_2 L^2+\ldots.+\theta_q L^q)$ has all the roots outside the unit circle. The autoregressive operator in (2.2.5) can be viewed as:

$$(1-\phi_1 L-\phi_2 L^2-\ \ldots..-\phi_p L^p)=$$

$$(1-\lambda_1^{-1}L)(1-\lambda_2^{-1}L)\ldots\ldots.(1-\lambda_p^{-1}L)$$

If all the roots $\lambda_1,\lambda_2,\ldots\ldots.\lambda_p$ are outside the unit circle, (2.2.5) can be written as:

$$u_t=\frac{1+\theta_1 L+\theta_2 L^2+\ldots..+\theta_q L^q}{(1-\lambda_1^{-1}L)(1-\lambda_2^{-1}L)\ldots..(1-\lambda_p^{-1}L)}\varepsilon_t\equiv\psi(L)\varepsilon_t$$

where $\sum_{j=0}^{\infty}|\psi_j|<\infty$ and the roots of $\psi(z)=0$ outside the unit circle.

Thus, when $|\lambda_i^{-1}|<1$ for all the indices i, the process (2.2.4) would just be a special case of a trend stationary process (2.2.2).

Suppose now that $\lambda_1^{-1}=1$ and $|\lambda_i^{-1}|<1$ for i = 2, 3, 4,…., p.

Then (2.2.5) states that:

$$(1-L)(1-\lambda_2^{-1}L)\ldots\ldots(1-\lambda_p^{-1}L)u_t =$$

$$(1+\theta_1 L+\theta_2 L^2+\ldots\ldots+\vartheta_q L^q)\varepsilon_t$$

meaning that:

$$(1\text{-}L)u_t = \frac{1+\theta_1 L+\theta_2 L^2+.....+\theta_q L^q}{(1\text{-}\lambda_2^{-1}L)(1\text{-}\lambda_3^{-1}L).....(1\text{-}\lambda_p^{-1}L)}\varepsilon_t = \psi^*(L)\varepsilon_t$$

with $\sum_{j=0}^{\infty} |\psi_j^*| < \infty$ and roots of $\psi^*(z) = 0$ outside the unit circle.

Now if (2.2.4) is first-differenced, the result is:

$$(1\text{-}L)y_t = (1\text{-}L)\alpha + [\delta t - \delta(t-1)] + (1\text{-}L)u_t =$$
$$0 + \delta + \psi^*(L)\varepsilon_t = \delta + \psi^*(L)\varepsilon_t$$

which is of the form of the unit root process (2.2.3).

We can also notice that if we first-difference the trend-stationary process in order to achieve stationarity instead of substracting the linear time trend we get the following result:

$$y_t = \alpha + \delta t + \psi(L)\varepsilon_t \Rightarrow \Delta y_t = \delta + (1-L)\psi(L)\varepsilon_t$$

which implies that the time series y_t is stationary but with a non-invertible moving average polynomial . This result generally limits the time series since, for example, it can not be represented as an AR(∞) process.

The whole representation explains the term **"unit root"** process, that is, one of the roots of the autoregressive polynomial in (2.2.5) is unity and all the others are outside the unit circle. The process (2.2.3) is often called **integrated of order 1** and written as $\mathbf{y_t \sim I\ (1)}$, or equivalently, is called **autoregressive integrated moving average process** and written as $\mathbf{y_t}$ **~ARIMA(p, 1, q)**. Suppose now that the autoregressive polynomial in (2.2.5) has two roots equal to unity and all the others are outside the unit circle. Then it holds that:

$$(1-L)^2 y_t = \delta + \psi^{**}(L)\varepsilon_t$$

and y_t is said to be I (2) and written as $y_t \sim$ I (2), or equivalently, $y_t \sim$ARIMA(p, 2, q). In general y_t is said to be an autoregressive integrated of order d moving average process, wtitten as $y_t \sim$ ARIMA(p, d, q), if d - differences are necessary for achieving stationarity:

$$\varphi_p(L)(1-L)^d y_t = \theta_q(L)\varepsilon_t$$

The first parameter p refers to the number of autoregressive lags (not counting the unit roots), the second parameter d refers to the order of differencing in order to achieve stationarity and the third parameter q gives the number of the moving average lags. Of particular interest to us are the I (1) processes (e.g. ARIMA(p, 1, q)) .

2.3 SHORT COMPARISON OF TREND STATIONARY AND UNIT ROOT PROCESSES

Unit root and trend-stationary processes often imply very different predictions, so deciding which process to use is tremendously important for applied forecasters. Having modeled nonstationary univariate processes that indicate linear trending behavior using trend-stationary and unit root processes let us now compare them from the forecasting point of view in order to see their asymptotical implications as well as their implications in finite samples. In this comparison we follow Hamilton (1994) (Section 15, p.444-446).

a) <u>Comparison of Forecasts</u>

In order to forecast a trend-stationary process of the form (2.2.2), we have to simply add the deterministic component $\alpha+\delta t$ to the forecast of the stationary stochastic component:

$$\hat{y}_{t+s/t} = \alpha + \delta(t+s) + \psi_s \varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \ldots \tag{2.3.1}$$

As the forecast horizon s grows large, absolute summability of $\{\psi_j\}$ implies that this forecast converges in mean square (m.s.) to the time trend:

$$\mathrm{E}[\hat{y}_{t+s/t} - \alpha - \delta(t+s)]^2 \xrightarrow{m.s.} 0 \qquad \text{as} \quad s \to \infty$$

To forecast a unit root process of the form (2.2.3), recalling that Δy_t, where Δ is the differencing operator, is a stationary process, we use the same formula that used above:

$$\Delta\hat{y}_{t+s/t} = \delta + \psi^*_s \varepsilon_t + \psi^*_{s+1}\varepsilon_{t-1} + \psi^*_{s+2}\varepsilon_{t-2} + \ldots\ldots \tag{2.3.2}$$

The desired forecast is $\hat{y}_{t+s/t}$ and not $\Delta\hat{y}_{t+s/t}$. But, it holds that:

$$\begin{aligned} y_{t+s} &= (y_{t+s} - y_{t+s-1}) + (y_{t+s-1} - y_{t+s-2}) + \ldots\ldots + \\ &+ (y_{t+1} - y_t) + y_t = \Delta y_{t+s} + \Delta y_{t+s-1} + \ldots\ldots + \Delta y_{t+1} + y_t \end{aligned}$$

and

$$\begin{aligned} &\hat{y}_{t+s/t} = \Delta\hat{y}_{t+s/t} + \Delta\hat{y}_{t+s-1/t} + \ldots.. + \Delta\hat{y}_{t+1/t} + y_t = \\ &(\delta + \psi^*_s \varepsilon_t + \psi^*_{s+1}\varepsilon_{t-1} + \psi^*_{s+2}\varepsilon_{t-2} + \ldots\ldots.) + \\ &(\delta + \psi^*_{s-1}\varepsilon_t + \psi^*_s \varepsilon_{t-1} + \psi^*_{s+1}\varepsilon_{t-2} + \ldots\ldots) + \\ &(\delta + \psi^*_1 \varepsilon_t + \psi^*_2 \varepsilon_{t-1} + \psi^*_3 \varepsilon_{t-2} + \ldots\ldots) + y_t \end{aligned}$$

which results in:

$$\hat{y}_{t+s/t} = s\delta + y_t + (\psi^*_s + \psi^*_{s-1} + ... + \psi^*_1)\varepsilon_t +$$
$$(\psi^*_{s+1} + \psi^*_s + ... + \psi^*_2)\varepsilon_{t-1} + ...$$

(2.3.3)

As we can see from (2.3.1) and (2.3.3), the forecast $\hat{y}_{t+s/t}$ converges to a linear function of the forecast horizon s with slope δ. The main difference is in the intercept of the line. For trend-stationary processes, the forecast converges to a line where the intercept is the same regardless the value of y_t. By contrast, the intercept of the limiting forecast of a unit root process is continually changing with each new observation on y.

b) Comparison of Forecast Errors

For a trend-stationary process of the form (2.2.2) the s-period-ahead forecast error is given by the formula:

$$y_{t+s} - \hat{y}_{t+s/t} =$$
$$[\alpha + \delta(t+s) + \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + ... +$$
$$\psi_{s-1}\varepsilon_{t+1} + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + ...] -$$
$$-[\alpha + \delta(t+s) + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + ...] =$$
$$\varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + ... + \psi_{s-1}\varepsilon_{t+1}$$

(2.3.5)

The mean squared error (MSE) of this forecast is:

$$E\left(y_{t+s} - \hat{y}_{t+s/t}\right)^2 = \left(1 + \psi_1^2 + \psi_2^2 + ... + \psi_{s-1}^2\right)\sigma^2 \qquad (2.3.6)$$

The MSE increases with the forecasting horizon s, though as s becomes large, the added uncertainty from forecasting farther into the future becomes negligible:

$$\lim_{s \to \infty} E(y_{t+s} - \hat{y}_{t+s/t})^2 = (1+\psi_1^2+\psi_2^2+\ldots\ldots)\sigma^2 \tag{2.3.7}$$

which is the unconditional variance of the stationary component $\psi(L)\varepsilon_t$.
For the unit root process (2.2.3) the s-period-ahead forecast error is:

$$\begin{aligned}
&y_{t+s} - \hat{y}_{t+s/t} = (\Delta y_{t+s} + \Delta y_{t+s-1} + \ldots + \Delta y_{t+1} + y_t) - \\
&-(\Delta\hat{y}_{t+s/t} + \Delta\hat{y}_{t+s-1/t} + \ldots\ldots + \Delta\hat{y}_{t+1/t} + y_t) = \\
&= (\varepsilon_{t+s} + \psi_1^{*}\varepsilon_{t+s-1} + \ldots\ldots + \psi_{s-1}^{*}\varepsilon_{t+1}) + \\
&+ (\varepsilon_{t+s-1} + \psi_1^{*}\varepsilon_{t+s-2} + \ldots\ldots + \psi_{s-2}^{*}\varepsilon_{t+1}) + \ldots + \varepsilon_{t+1} = \\
&= \varepsilon_{t+s} + (1+\psi_1^{*})\varepsilon_{t+s-1} + (1+\psi_1^{*}+\psi_2^{*})\varepsilon_{t+s-2} + \ldots\ldots + \\
&+ (1+\psi_1^{*}+\psi_2^{*}+\psi_3^{*}+\ldots\ldots+\psi_{s-1}^{*})\varepsilon_{t+1}
\end{aligned}$$

with MSE

$$\begin{aligned}
&E(y_{t+s} - \hat{y}_{t+s/t})^2 = [1 + (1+\psi_1^{*})^2 + (1+\psi_1^{*}+\psi_2^{*})^2 + \ldots\ldots + \\
&(1+\psi_1^{*}+\psi_2^{*}+\ldots\ldots+\psi_{s-1}^{*})^2]\sigma^2
\end{aligned}$$

The MSE again increases with the length of the forecasting horizon s, though in contrast to the trend-stationary case, the MSE does not converge to any fixed value as s goes to infinity. Instead, it asymptotically approaches a linear function of s with slope $(1+\psi^{*}{}_1+\psi^{*}{}_2+\ldots\ldots)^2\sigma^2$.

To summarize, the forecast error of a trend-stationary process is given by $\psi(L)$ that is stationary. Hence, there is no influence of past or currect innovations on the mean of the dependent variable y_t, and all uncertainty about the forecast error of y_t is bounded in contrast to a unit root process that eventually grows linearly with the forecast horizon.

Let us now see the implications of these two competing processes at a finite sample size of observations. Assume that we have a finite sample size of T observations that were really generated by the following unit root process:

$$y_t = y_{t-1} + \varepsilon_t \qquad \text{true model} \qquad \text{(unit root)} \tag{2.3.8}$$

Consider trying to distinguish this from the following stationary process:

$$y_t = \varphi y_{t-1} + \varepsilon_t \;\; , |\, \varphi \,| < 1 \, , \qquad \text{false model} \qquad \text{(stationary)} \tag{2.3.9}$$

The s-period-ahead forecast of y_{t+s} assuming the process of the form (2.3.7) is:

$$\hat{y}_{t+s/t} = y_t \tag{2.3.10}$$

with MSE

$$\text{E}(y_{t+s} - \hat{y}_{t+s/t})^2 = s\sigma^2 \tag{2.3.11}$$

The corresponding forecast of y_{t+s} assuming the process of the form (2.3.8) is:

$$\hat{y}_{t+s/t} = \varphi^s y_t \tag{2.3.12}$$

with MSE

$$\text{E}(y_{t+s} - \hat{y}_{t+s/t})^2 = (1 + \varphi^2 + \varphi^4 + \ldots\ldots + \varphi^{2(s-1)})\sigma^2 \tag{2.3.13}$$

As we can notice, there exists a value of φ very close to unity such that the observable implications of the stationary process are arbitrarily close to those of the unit root process.

Below we show that the opposite is also true: there exists a value of θ very close to – 1 such that the observable implications of a unit root process are arbitrarily close to those of a stationary process. To see this consider the following process as the true process:

$$y_t = \varepsilon_t \qquad \text{true model} \qquad \text{(stationary)} \tag{2.3.14}$$

and consider trying to distinguish this from:

$$(1-L)y_t = (1+\theta L)\varepsilon_t \ , \ |\,\theta\,| < 1 \ , \quad \text{false model} \quad \text{(unit root)} \tag{2.3.15}$$

The s-period-ahead forecast of y_{t+s} assuming the process of the form (2.3.14) is:

$$\hat{y}_{t+s/t} = 0$$

with MSE

$$\mathrm{E}(y_{t+s} - \hat{y}_{t+s/t})^2 = \sigma^2$$

The s-period-ahead forecast of y_{t+s} assuming the process of the form (2.3.15) is:

$$\hat{y}_{t+s/t} = y_t + \vartheta\varepsilon_t = (\Delta y_t + \Delta y_{t-1} + \ldots\ldots + \Delta y_2 + y_1) + \vartheta\varepsilon_t =$$
$$[(\varepsilon_t + \theta\varepsilon_{t-1}) + (\varepsilon_{t-1} + \theta\varepsilon_{t-2}) + \ldots\ldots + (\varepsilon_2 + \theta\varepsilon_1) + \varepsilon_1] + \theta\varepsilon_t =$$
$$(1+\theta)(\varepsilon_t + \varepsilon_{t-1} + \ldots\ldots + \varepsilon_2 + \varepsilon_1$$

with MSE

$$\mathrm{E}(y_{t+s} - \hat{y}_{t+s/t})^2 = [1 + (s-1)(1+\theta)^2]\sigma^2$$

To emphasize the fact that for a given fixed sample size of observations T unit root and stationary processes have identical implications we give a visual example that shows that even when processes share the same realizations for the innovations, series that have unit roots and series that do not can behave similarly.

Example (Visual inspection)

Consider the following stochastic processes for the series y_t and z_t:

$$y_t = \begin{Bmatrix} y_{t-1} + \varepsilon_t \\ 0.9 y_{t-1} + \varepsilon_t \end{Bmatrix} \quad \text{and} \quad z_t = \begin{Bmatrix} 0.05 + z_{t-1} + \varepsilon_t \\ 0.05t + \varepsilon_t \end{Bmatrix}$$

with $\{\varepsilon_t\} \sim N(0, 0.1^2)$.

In each case the first processes correspond to a pure random walk and a random walk with drift, while the last correspond to a stationary AR(1) and a trend-stationary process. Figure (A.5) in Appendix A presents realizations of size 100 for each of these processes with $\{\varepsilon_t\}_{t=1}^{100}$ fixed. We labelled the resulting series as y_1, y_2 and z_1 and z_2 randomly. Can we tell which have unit roots? The answer is negative. So, visual inspection can not serve as a guide for discriminating series with or without unit roots.

2.4 REMARKS

The basic distinction between trend-stationary and unit root processes is that the former do, and the latter do not, tend to return to a fixed deterministic trend function. Since the non-trend component $\psi(L)\varepsilon_t$ in (2.2.2) is stationary with mean zero, the process is such that y_t tends to fluctuate about the fixed trend function $\alpha+\delta t$. In formulation (2.2.3), by contrast, the tendency is for y_t to grow at a rate δ from its current position, whatever might be. There is, expect in a limiting case, no tendency for y_t to return to any fixed trend path.

In addition, unit root and stationary processes have sharp asymptotic differences but for any fixed sample size there are unit root processes which behave like stationary processes and stationary processes which behave like unit root processes.

Thus, we must be very careful when testing whether a time series contains unit root or testing whether there exists a permanent effect on the level of the series. As far as this thesis is concerned, we will deal with unit root processes for analyzing the trending behavior as well as the nonstationarity in the mean of time series.

CHAPTER 3

UNIVARIATE UNIT ROOT PROCESSES

3.1 INTRODUCTION

Unit root processes have attracted a considerable amount of work in both statistical and economics literature. Indeed, the view that most economic time series are better characterized by unit root rather than trend-stationary processes has become prevalent. The work of Nelson and Plosser (1982) that found that most macroeconomic variables have a univariate time series structure with an autoregressive unit root has catalyzed the research with both empirical and theoretical dimensions. Nelson and Plosser's study was followed by a series of empirical analyses that confirmed their findings. Empirical applications of different methodologies generally confirmed the conclusion that most macroeconomic time series have an autoregressive unit root. These studies had many effects on statistical and economic theory. They seem to confirm previous analyses that had advanced the unit root processes for particular economic series, for example, consumption (Hall, 1978), velocity of money (Gould and Nelson, 1974) and stock prices (Samuelson, 1973).

In this chapter we will give the statistical inference concerning the univariate unit root processes. In particular, we are mainly interested in deciding whether to proceed with a unit root or a trend(or mean)-stationary process in order to describe the nonstationary behavior of a time series of the type mentioned in the previous chapter since we saw that they both lead to different implications.

This specific discrimination is actually a model selection problem. Fuller (1976) and Dickey-Fuller (1979) were the first who dealt with such discrimination. They actually approached it as a hypothesis testing problem and proposed two test statistics, the Z_{DF} – test statistic and the τ_{DF} – test statistic, commonly known as standard Dickey-Fuller (DF) test statistics and are based on a simple autoregression with or without a constant or a linear time trend. They are based on testing $\rho = 1$ in the equations:

$$y_t = \rho y_{t-1} + \varepsilon_t$$

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t$$

$$y_t = \alpha + \delta t + \rho y_{t-1} + \varepsilon_t$$

with the error process $\{\varepsilon_t\}$ being independent identically distributed, for determining whether y_t has a unit root so that y_t is a difference-stationary process.

Because ordinary asymptotic distribution theory breaks down in the case in which ρ is precisely equal to unity (ρ = 1), the two proposed Dickey-Fuller unit root test statistics on the coefficient y_{t-1} require to be compared with critical values taken from non-standard limiting distributions, the so called Dickey-Fuller distributions, which differ from depending on whether or not nuisance parameters such as constant or time trend are included in the estimated autoregression. But this can readily be done since Dickey and Fuller have provided the profession with the pertinent Tables.

The assumption that the error process in the above equations is independent identically distributed is quite unrealistic for time series encountered in practice. Dickey and Fuller (1979) also considered the case of testing the unit root hypothesis that $\rho = 1$ in more complicated than independent identically distributed error processes, namely under the assumption that the error process follow an autoregressive process, say AR(p) with p > 1, the order of which is required to be finite and correctly specified. When the error process $\{\varepsilon_t\}$ is correlated, there is a need to either change the estimation method (adopt another estimated equation) or modify the test statistics to obtain consistent estimators and test statistics. Dickey and Fuller (1979) use the first approach of changing the estimated equation, commonly known as the "Augmented Dickey-Fuller (ADF) autoregression". The Dickey-Fuller test statistics concerning the unit root hypothesis that ρ = 1 under the case of the correlated error processes are the so called "Augmented Dickey-

Fuller" unit root test statistics and their limiting distributions, the Dickey-Fuller distributions, are the same as in the case where the error process is assumed to be independent identically distributed not forgetting that are different from each other depending on whether or not nuisance parameters are included in the estimated Augmented Dickey-Fuller autoregression.

The difference between the simple autoregression and the Augmented Dickey-Fuller autoregression is that in the latter additional lagged-difference terms of the variable y_t have to be included in its right side in order to eliminate the serial correlation in the error process (see Introduction, equation 3). In addition, the appropriate number of the additional lagged-difference terms to be included in the right side of the Augmented Dickey-Fuller autoregression is depended on the true order of the error process.

Of course, there are many proposed test statistics concerning the unit root hypothesis that $\rho = 1$ for discriminating unit root and trend(or mean)-stationary processes in the literature . In this thesis we will deal with the Dickey-Fuller unit root test statistics since they are widely known from time series analysts and mostly they are easily computed.

The plan of this chapter is as follows. In section 3.2 we explain why the asymptotic distributions and the rates of convergence for the estimated coefficients of unit root processes differ from those for stationary processes when testing the unit root hypothesis. The above statements are illustrated using the first given simple autoregression with the error process being independent identically distributed. The presentation of the standard Dickey-Fuller unit root test statistics along with their asymptotic distributions in the case where different nuisance parameters such as constant and linear time trend are included in the simple autoregression when testing the unit root hypothesis is presented in section 3.3 with section 3.4 being involved with some remarks on the standard Dickey-Fuller unit root tests. Section 3.5 is concerned with testing the unit root hypothesis in the Augmented Dickey-Fuller autoregression under autoregressive error processes of finite and correctly specified orders with section 3.6 being involved with some remarks coming from the section 3.5.

This chapter is mainly based on the development of Hamilton (1994, chapter 17, p.486-543) and interested readers should refer to him for an extensive study on the theory that is going to be illustrated.

3.2 RATES OF CONVERGENCE

Consider the following process:

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{3.2.1}$$

with the error process $\{\varepsilon_t\}$ being independent identically normally distributed with zero mean and σ^2 variance, denoted by $\{\varepsilon_t\}\sim$ i.i.d.N(0 , σ^2) and $y_0 = 0$. Given T observations, the conditional maximum likelihood estimator of ρ is the ordinary least squares (OLS) estimator:

$$\hat{\rho}_T = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} \tag{3.2.2}$$

where $\sum$ denotes hereafter summation over the observations t=1 , 2 ,…., T. Mann and Wald (1943) first proved that if $|\rho|<1$, then:

$$\sqrt{T}(\hat{\rho}_T - \rho) \xrightarrow{L} \mathrm{N}[0,(1-\rho^2)] \Rightarrow \frac{\sqrt{T}(\hat{\rho}_T - \rho)}{\sqrt{(1-\rho^2)}} \xrightarrow{L} N(0,1) \tag{3.2.3}$$

(where $\xrightarrow{L}$ denotes convergence in distribution).

Suppose now that we would like to test the hypothesis that $\rho=1$. If (3.2.3) were also valid for the case when $\rho=1$, it would seem to claim that

$\sqrt{T}(\hat{\rho}_T - \rho)$ has zero variance, or that the distribution collapses to a point mass at zero:

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0$$

(where $\xrightarrow{p}$ denotes convergence in probability).

Rubin (1950) showed that $\hat{\rho}_T$ is a consistent estimator for all values of ρ. On the other hand, when ρ=1, White (1958) was unable to invert the limiting joint-moment generating function to obtain the limiting distribution. However, based on Donsker's theorem, he represented the limiting distribution of $T(\hat{\rho}_T - 1)/\sqrt{2}$ as that of the ratio of two integrals defined on the Wiener process. Later M.M.Rao (1961) obtained an expression of the limiting distribution of $T(\hat{\rho}_T - 1)/\sqrt{2}$ but it is not of an easily recognizable form. However, the standardization of $\hat{\rho}_T$ as $T(\hat{\rho}_T - \rho)/\sqrt{2}$ by White (1958) and M.M.Rao (1961) has been proved incorrect and the correct standardization of $\hat{\rho}_T$ is $T(\hat{\rho}_T - \rho)$ (Phillips 1987a). Fuller (1976) and Dickey and Fuller (1979) derived the asymptotic distribution of $\hat{\rho}_T$ under the assumption of the i.i.d. errors and assumption of $y_0 = 0$.

To get a better sense of why scaling by T is necessary when the true value of ρ is unity let us analyze apart the numerator and the denominator in the beneath equation which is actually the difference between the OLS estimator $\hat{\rho}_T$ and the true value ρ which in our case is assumed to be unity:

$$T(\hat{\rho}_T - 1) = \frac{\frac{1}{T}\Sigma y_t \varepsilon_{t-1}}{\frac{1}{T^2}\Sigma y_{t-1}^2} \tag{3.2.4}$$

We first work on the numerator. Since y_t is a unit root process, or equivalently, a driftless random walk under the assumption that $\rho=1$, it is normally distributed with a variance that grows at rate t:

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \ldots\ldots + \varepsilon_1 \sim N(0, t\sigma^2)$$

Notice also that:

$$y_t = y_{t-1} + \varepsilon_t \Leftrightarrow y_t^2 = (y_{t-1} + \varepsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\varepsilon_t + \varepsilon_t^2 \Leftrightarrow$$
$$\Leftrightarrow 2y_{t-1}\varepsilon_t = y_t^2 - y_{t-1}^2 - \varepsilon_t^2 \Leftrightarrow y_{t-1}\varepsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \varepsilon_t^2)$$

Taking the time series average, we obtain:

$$\frac{1}{T}\Sigma y_{t-1}\varepsilon_t = \frac{1}{2T}\Sigma(y_t^2 - y_{t-1}^2) - \frac{1}{2T}\Sigma\varepsilon_t^2 = \frac{1}{T}\frac{1}{2}(y_T^2 - y_0^2) - \frac{1}{2T}\Sigma\varepsilon_t^2 =$$
$$\frac{1}{T}\frac{1}{2}y_T^2 - \frac{1}{2T}\Sigma\varepsilon_t^2$$

Thus,

$$\frac{1}{T}\Sigma y_{t-1}\varepsilon_t = \frac{1}{T}\frac{1}{2}y_T^2 - \frac{1}{2T}\Sigma\varepsilon_t^2 \Leftrightarrow$$
$$\frac{1}{\sigma^2 T}\Sigma y_{t-1}\varepsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2 T}\Sigma\varepsilon_t^2 \qquad (3.2.5)$$

We know that $\Sigma\varepsilon_t^2$ is the sum of T i.i.d. random variables, each of mean zero and variance σ^2, and so, by the law of large numbers we have that:

$$\frac{1}{T}\Sigma\varepsilon_t^2 \xrightarrow{p} \sigma^2 \qquad (3.2.6)$$

Furthermore,

$$y_t \sim \mathrm{N}(0,\ \sigma^2 t) \Rightarrow y_T \sim \mathrm{N}(0,\ \sigma^2 T) \Rightarrow \frac{y_T}{\sigma\sqrt{T}} \sim \mathrm{N}(0,\ 1) \Rightarrow$$

$$\Rightarrow \frac{y_T^2}{(\sigma\sqrt{T})^2} \sim \chi_1^2 \tag{3.2.7}$$

where χ_1^2 is the chi-square distribution with one degree of freedom.

Using (3.2.6) and (3.2.7), it follows from (3.2.5) that:

$$\frac{1}{\sigma^2 T}\sum y_{t-1}\varepsilon_t \xrightarrow{L} \frac{1}{2}(\chi_1^2 - 1) \tag{3.2.8}$$

For the denominator we use the fact that $y_{t-1} \sim \mathrm{N}\ [0,\ \sigma^2(t\text{-}1)]$.

Then, $\mathrm{E}(y_{t-1}^2) = \sigma^2(t-1) \Rightarrow \sum \mathrm{E}(y_{t-1}^2) = \sigma^2 \sum (t-1) = \sigma^2 \frac{(T-1)T}{2}$

Now, dividing by T^2 we get:

$$\frac{1}{T^2}\sum \mathrm{E}(y_{t-1}^2) = \frac{\sigma^2}{2} - \frac{\sigma^2}{2T} \to \frac{\sigma^2}{2}$$

By doing these steps we have been able to establish that the denominator of $T(\hat{\rho}_T - 1)$ is a random variable with mean $\sigma^2/2$. We can not really say at this point whether

$$\frac{1}{T^2}\sum y_{t-1}^2$$

converges in probability to $\sigma^2/2$ or if it converges in distribution to a random variable with mean $\sigma^2/2$. Moreover, the asymptotic distribution of the $T(\hat{\rho}_T - 1)$ under the above considerations is not the usual Gaussian distribution but instead involves a χ_1^2 - variable in the numerator and a separated non-standard distribution in the denominator. The asymptotic distribution of $T(\hat{\rho}_T - 1)$ will be fully characterized in the next section.

Now consider again the OLS estimator $\hat{\rho}_T$ given by the form of (3.2.2). Another popular way to test the unit root hypothesis that $\rho = 1$ is based on the usual OLS t – test:

$$\frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{\hat{\rho}_T - 1}{\left\{s_T^2 / \sum y_{t-1}^2\right\}^{1/2}}$$

where $\hat{\sigma}_{\hat{\rho}_T}$ is the usual OLS standard error for the estimated coefficient $\hat{\rho}_T$ and

$$s_T^2 = \sum\left(y_t - \hat{\rho}_T y_{t-1}\right)^2 / (T-1)$$

Another but equivalent formulation of the usual OLS t – test is given by:

$$\frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{T^{-1}\sum y_{t-1}\varepsilon_t}{\left[T^{-2}\sum y_{t-1}^2\right]^{1/2}\left(s_T^2\right)^{1/2}}$$

and it is obvious that has terms that are equal to those included in the form of (3.2.4). So, repeating similarly the analysis for the numerator and the denominator of the above formulation of the usual OLS t – test we conclude that its distribution is not the usual normal distribution but instead involves a

χ_1^2 - variable in the numerator and a separated non-standard distribution in the denominator. The asymptotic distribution of the usual OLS t – test will be also fully characterized in the next section.

3.3 ASYMPTOTIC PROPERTIES OF THE DICKEY-FULLER (DF) STATISTICS

Before starting characterizing the asymptotic distribution of the $T(\hat{\rho}_T - 1)$ test statistic as well as that of the usual OLS t – test statistic concerning the unit root hypothesis that $\rho = 1$ under the process of the form (3.2.1) we will present some of its asymptotic properties when the true value of the coefficient ρ is unity in a form of a proposition without giving their proofs (Hamilton 1994, chapter 15, p.486).

<u>Proposition A</u>

Suppose that ξ_t follows a unit root process, or equivalently, a random walk without a drift term:

$$\xi_t = \xi_{t-1} + \varepsilon_t \tag{3.3.1}$$

where $\xi_0 = 0$ and $\{\varepsilon_t\}$~ i.i.d.$(0, \sigma^2)$. Then as $T \to \infty$:

a) $$T^{-\frac{1}{2}} \sum \varepsilon_t \xrightarrow{L} \sigma W(1) \tag{3.3.2}$$

b) $$T^{-1} \sum \xi_{t-1} \varepsilon_t \xrightarrow{L} \frac{1}{2} \sigma^2 \left\{ [W(1)]^2 - 1 \right\} \tag{3.3.3}$$

c) $$T^{-\frac{3}{2}}\sum t\varepsilon_t \xrightarrow{L} \sigma W(1) - \sigma\int_0^1 W(r)dr \qquad (3.3.4)$$

d) $$T^{-\frac{3}{2}}\sum \xi_{t-1} \xrightarrow{L} \sigma\int_0^1 W(r)dr \qquad (3.3.5)$$

e) $$T^{-2}\sum \xi_{t-1}^2 \xrightarrow{L} \sigma^2\int_0^1 [W(r)]^2 dr \qquad (3.3.6)$$

f) $$T^{-\frac{5}{2}}\sum t\xi_{t-1} \xrightarrow{L} \sigma\int_0^1 rW(r)dr \qquad (3.3.7)$$

The asymptotic distributions in the above proposition are all written in terms of functionals of the Wiener process denoted by W(r) for $0 \le r \le 1$. Although Fuller (1976) and Dickey and Fuller (1976,1979) did not provide the limiting distributions concerning the two test statistics considered in the previous section with functionals of the Wiener processes, they first derived their distributions and made their Tables. The functional forms of the Wiener processes are due to Phillips (1987a, 1988b).

Most of the studies about the asymptotic behavior of the OLS estimator $\hat{\rho}_T$ have been concerned with the following process:

$$y_t = \rho y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is assumed to be a sequence of independent normal random variables with zero mean and σ^2 variance, denoted by $\{\varepsilon_t\}$~i.i.d.N(0,σ^2), given that the data are believed to be generated by the process:

$$y_t = y_{t-1} + \varepsilon_t$$

It is also hard to believe that when the unit root hypothesis is rejected the AR(1) process without a constant or a deterministic time trend describes well most of the economic and financial time series which usually show some tendency to increase over time implying trending behavior as well as random fluctuation that implies nonstationarity in their mean. Hamilton (1994, p.487) summarizes the above considerations by dividing them up into the following four cases:

Case 1

Assume that the true process that generates the data is driven by the following unit root process which is often called driftless random walk:

$$y_t = y_{t-1} + \varepsilon_t \tag{3.3.8}$$

where $y_0 = 0$ and $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$ and that the estimated equation is given by the form:

$$y_t = \rho y_{t-1} + \varepsilon_t$$

Then, for the normalized least squares estimator, $\hat{\rho}_T$, we have that:

$$Z_{DF} = \mathrm{T}(\hat{\rho}_T - 1) = \frac{\frac{1}{T}\Sigma y_{t-1}\varepsilon_t}{\frac{1}{T^2}\Sigma y_{t-1}^2} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\int_0^1 [W(r)]^2 dr} \tag{3.3.9}$$

and for the studentized least squares estimator, $\hat{\rho}_T$, we have that:

$$\tau_{DF} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\left\{\int_0^1 [W(r)]^2 dr\right\}^{\frac{1}{2}}} \tag{3.3.10}$$

where $\hat{\sigma}^2_{\hat{\rho}_T} = \dfrac{s_T^2}{\Sigma y_{t-1}^2}$ and $s_T^2 = \dfrac{1}{T-1}\Sigma(y_t - \hat{\rho}_T y_{t-1})^2$

The result (3.3.9) allows the point estimate $\hat{\rho}_T$ to be used by itself to test the null hypothesis of unit root, that is $\rho = 1$, without needing to calculate its standard error. The asymptotic density[1] of the Z_{DF} – test statistic under this case is displayed in Figure (A.6) in Appendix A and clearly shows that the distribution involved is non-normal and asymmetric, with a fat left tail.
This is the Z_{DF} Dickey-Fuller distribution.

The τ_{DF} – test statistic taken from (3.3.10) is actually the ordinary least squares (OLS) t – test statistic. Although the τ_{DF} – test statistic is calculated in the usual way, it does not have a limiting Gaussian distribution when the true process is characterized by the unit root hypothesis of $\rho = 1$ as Figure (A.7) in Appendix A shows. More specifically, is shifted to left of the standard normal density. **This is the τ_{DF} Dickey-Fuller distribution**.

---

[1] This density is actually a kernel estimate of the density of $\hat{\rho}_T$ on the basis of 10,000 replications of a Gaussian random walk $y_t = y_{t-1} + \varepsilon_t$, t = 0 , 1 , ….., 1000 , $y_t = 0$ for t < 0. The kernel involved is the standard normal density, and bandwidth $h = cs1000^{-1/5}$, where s is the sample standard error and c = 1. The scale factor c has been chosen by experimenting with various values. The value c = 1 is about the smallest one for which the kernel estimate remains a smooth curve, for smaller values of c the kernel becomes wobbly. The same holds true of the density of the studentized $\hat{\rho}_T$ coefficient. The densities that will be considered in the other cases have been derived with analogous way.

Case 2

Assume that the true process that generates the data is driven by the following unit root process which is often called driftless random walk:

$$y_t = y_{t-1} + \varepsilon_t$$

where $y_0 = 0$ and $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$ and that the estimated equation is given by the form:

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t$$

Then, looking at the least squares estimators

$$\begin{bmatrix} T^{\frac{1}{2}}(\hat{\alpha}_T - 0) \\ T(\hat{\rho}_T - 1) \end{bmatrix} = \begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum y_{t-1} \\ T^{-\frac{3}{2}}\sum y_{t-1} & T^{-2}\sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum \varepsilon_t \\ T^{-1}\sum y_{t-1}\varepsilon_t \end{bmatrix}$$

for the normalized least squares estimator, $\hat{\rho}_T$, we have that:

$$Z_{DF} = T(\hat{\rho}_T - 1) \xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int_0^1 W(r)dr}{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2} \qquad (3.3.11)$$

and for the studentized least squares estimator, $\hat{\rho}_T$, we have that:

$$\tau_{DF} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\} - W(1)\int_0^1 W(r)dr}{\left\{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2\right\}^{\frac{1}{2}}} \tag{3.3.12}$$

where $$\hat{\sigma}^2_{\hat{\rho}_T} = s_T^2 [0 \quad 1] \begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y^2_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and $$s_T^2 = \frac{1}{T-2} \Sigma (y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2$$

Again result (3.3.11) allows the point estimate $\hat{\rho}_T$ to be used by itself to test the null hypothesis of unit root existness, that is $\rho = 1$, without needing to calculate its standard error.

Moreover, the asymptotic distribution of the estimate $\hat{\rho}_T$ is not the same as the asymptotic distribution in (3.3.9) when a constant term is included in the estimated equation. This conclusion also holds for the τ_{DF} – test statistic.

The density of the Z_{DF} – test statistic under this case is displayed in Figure (A.8) in Appendix A and compared to Figure (A.6) we can see that is farther left of zero and has a fatter left tail. Figure (A.9) in the same Appendix shows the density of the τ_{DF} – test statistic under this case compared to the standard normal density. We can see that the density of the τ_{DF} – test statistic is shifted even more to the left of the standard normal density than in Figure (A.7).

Statistical tables for the distribution of Z_{DF} – test statistic and of τ_{DF} – test statistic under various sample sizes for this case are reported in Tables (B.1) and (B.2) in Appendix B, respectively.

Case 3

Assume that the true process that generates the data is driven by the following unit root process with drift term which is often called random walk with a drift:

$$y_t = \alpha + y_{t-1} + \varepsilon_t$$

where $y_0 = 0$, $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$, $\alpha \neq 0$ and that the estimated equation is given by the form:

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t$$

Then, looking at the least squares estimators:

$$\begin{bmatrix} T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha) \\ T^{\frac{3}{2}}(\hat{\rho}_T - 1) \end{bmatrix} = \begin{bmatrix} 1 & T^{-2}\sum y_{t-1} \\ T^{-2}\sum y_{t-1} & T^{-3}\sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum \varepsilon_t \\ T^{-\frac{3}{2}}\sum y_{t-1}\varepsilon_t \end{bmatrix}$$

it holds that :

$$\begin{bmatrix} T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha) \\ T^{\frac{3}{2}}(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} \mathrm{N}\left(0, \sigma^2 Q^{-1}\right) \qquad (3.3.13)$$

where $$Q = \begin{bmatrix} 1 & \frac{\alpha}{2} \\ \frac{\alpha}{2} & \frac{\alpha^2}{3} \end{bmatrix}$$

As can be noticed, both estimated coefficients are asymptotically normally distributed. So, any t-test involving any coefficients from this regression can be compared with the usual t-tables.

Case 4

Assume that the true process that generates the data is driven by the following unit root process with drift term which is often called random walk with a drift:

$$y_t = \alpha + y_{t-1} + \varepsilon_t$$

where $y_0 = 0$, $\{\varepsilon_t\}$~ i.i.d.(0, σ^2) and that the estimated equation is given by the form:

$$y_t = \alpha + \rho y_{t-1} + \delta t + \varepsilon_t \qquad (3.3.14)$$

Then, if $\alpha \neq 0$, we can rewrite the equation (3.3.14) as:

$$y_t = (1-\rho)\alpha + \rho\left[y_{t-1} - \alpha(t-1)\right] + (\delta + \rho\alpha)t + \varepsilon_t = \alpha^* + \rho^* g_{t-1} + \delta^* t + \varepsilon_t \qquad (3.3.15)$$

where $\alpha^* = (1-\rho)a$, $\rho^* = \rho$, $\delta^* = \delta + \rho\alpha$, and $g_t = y_t - \alpha t$

The maintained hypothesis is that:

$$\alpha = \alpha_0 \,, \quad \rho = 1 \,, \quad \delta = 0$$

which in the transformed system would mean that:

$$\alpha^{*}=0\ ,\quad \rho^{*}=1\ ,\quad \delta^{*}=\alpha_{0}$$

Therefore, looking at the least squares estimators

$$\begin{bmatrix} T^{\frac{1}{2}}\left(\hat{\alpha}_{T}^{*}-0\right) \\ T\left(\hat{\rho}_{T}^{*}-1\right) \\ T^{\frac{3}{2}}\left(\hat{\delta}_{T}^{*}-\alpha_{0}\right) \end{bmatrix} = \begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum g_{t-1} & T^{-2}\sum t \\ T^{-\frac{3}{2}}\sum g_{t-1} & T^{-2}\sum g_{t-1}^{2} & T^{-\frac{5}{2}}\sum t g_{t-1} \\ T^{-2}\sum t & T^{-\frac{3}{2}}\sum t g_{t-1} & T^{-3}\sum t^{2} \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum \varepsilon_{t} \\ T^{-1}\sum g_{t-1}\varepsilon_{t} \\ T^{-\frac{3}{2}}\sum t\varepsilon_{t} \end{bmatrix}$$

we have that :

$$\begin{bmatrix} T^{\frac{1}{2}}\left(\hat{\alpha}_{T}^{*}-0\right) \\ T\left(\hat{\rho}_{T}^{*}-1\right) \\ T^{\frac{3}{2}}\left(\hat{\delta}_{T}^{*}-\alpha_{0}\right) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} * \begin{bmatrix} 1 & \int_{0}^{1} W(r)dr & \frac{1}{2} \\ \int_{0}^{1} W(r)dr & \int_{0}^{1} [W(r)]^{2} dr & \int_{0}^{1} rW(r)dr \\ \frac{1}{2} & \int_{0}^{1} rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} *$$

$$* \begin{bmatrix} W(1) \\ \frac{1}{2}\left\{[W(1)]^{2}-1\right\} \\ W(1)-\int_{0}^{1} W(r)dr \end{bmatrix} = K \qquad (3.3.16)$$

As can be noticed, $\hat{\rho}_T^*$ which is the OLS estimate of ρ based on (3.3.15) is identical to $\hat{\rho}_T$, the OLS estimate of ρ based on (3.3.14).

Thus, the asymptotic distribution of the $Z_{DF} = T(\hat{\rho}_T - 1)$ test statistic is given by the middle row of (3.3.16). As we can also see, this distribution does not depend on either σ or α, no matter whether or not the true value of α is zero.

For the studentized least squares estimator, $\hat{\rho}_T$, we have that:

$$\tau_{DF} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{T(\hat{\rho}_T - 1)}{\left(T^2 \hat{\sigma}^2_{\hat{\rho}_T}\right)^{\frac{1}{2}}}$$

The asymptotic distribution of the denominator, $T^2 \hat{\sigma}^2_{\hat{\rho}_T}$, is given by:

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 W(r)dr & \frac{1}{2} \\ \int_0^1 W(r)dr & \int_0^1 [W(r)]^2 dr & \int_0^1 rW(r)dr \\ \frac{1}{2} & \int_0^1 rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \equiv V$$

whereas the asymptotic distribution of the numerator, $T(\hat{\rho}_T - 1)$, is given by the middle row of K. So, we get the limiting distribution of the τ_{DF} – test statistic by diving the above two limiting distributions V and K. Again, this distribution does not depend on α or σ.

The densities of Z_{DF} and τ_{DF} – test statistics under this case (the latter compared with the standard normal density) are displayed in Figures (A.10) and (A.11) in Appendix A. Again, these densities are farther to the left and heavier-tailed compared to the corresponding densities in Figures (A.6) through (A.9).

Statistical tables for the distribution of Z_{DF} – test statistic and of τ_{DF} – test statistic under various sample sizes for this case are reported in Tables (B.1) and (B.2) in Appendix B, respectively.

3.4 REMARKS ON THE DF – STATISTICS

As we noticed from the previous results, the asymptotic distributions of the standard Dickey-Fuller unit root test statistics under the null hypothesis of a unit root depend on whether or not a constant term or a deterministic time trend is included in the estimated equation as well as on the form of the assumed data generating process. Another important comment on the standard Dickey-Fuller unit root test statistics is that they have been derived to distinguish between unit root processes, or equivalently, random walks and stationary AR(1) processes around a constant or linear time trend with the latter being rather simple processes for characterizing time series encountered in practice when the unit root hypothesis is rejected.

Note also that the Dickey-Fuller unit root test statistics are actually t- and F- test statistics with critical values simulated by Dickey and Fuller.

Readers who are interested in getting further insight into the proofs of Dickey-Fuller unit root test statistics should referred to Hamilton (1994, chapter 17, p.475-543).

3.5 ASYMPTOTIC PROPERTIES OF THE AUGMENTED DICKEY-FULLER (ADF) STATISTICS

The assumption that the error process in the equation:

$$y_t = d_t + \rho y_{t-1} + \varepsilon_t$$

where $d_t = 0$, $d_t = \alpha$ (constant) and $d_t = \alpha + \delta t$ (linear time trend), is independent identically distributed is quite unrealistic for time series encountered in practice. When the error process $\{\varepsilon_t\}$ is correlated, there is a need to either change the estimation method (adopt another estimated equation) or modify the test statistics to obtain consistent estimators and test statistics.

Dickey and Fuller (1979) and later Said and Dickey (1984) used the first approach of changing the estimated equations. In particular, Dickey and Fuller (1979) considered the case of testing the unit root hypothesis that $\rho = 1$ in the above equation under the assumption that the error process $\{\varepsilon_t\}$ follow an autoregressive process, say AR(p) with $p > 1$, the order of which is required to be finite and correctly specified and Said and Dickey (1984) under the assumption that the error process $\{\varepsilon_t\}$ follow an autoregressive moving average process, say ARMA(p,q), the order of which is not required to be finite and correctly specified given appropriate conditions for it. Having in mind the above equation, the Dickey-Fuller's assumption that the error process follows an AR(p) with $p>1$ process covers the case in which Δy_t is a stationary AR(p) process under the unit root hypothesis against the alternative that y_t is an AR(p+1) process. Before starting deriving the unit root tests under the Dickey-Fuller's assumption for the error process, let us see an alternative representation for the autoregressive processes.

Suppose that the true process that generates the data satisfies the equation:

$$\left(1-\varphi_1 L-\varphi_2 L^2-.....-\varphi_p L^p\right)y_t=\varepsilon_t$$

(3.5.1)

where $\{\varepsilon_t\}$~ i.i.d.(0, σ^2) with finite bounded fourth moment. We will write the autoregression (3.5.1) in a different form. To do so, we define:

$$\rho=\varphi_1+\varphi_2+.....+\varphi_p \tag{3.5.2}$$

$$\zeta_j=-\left(\varphi_{j+1}+\varphi_{j+2}+.....+\varphi_p\right) \quad \text{for} \quad j=1,2,3,......,p-1 \tag{3.5.3}$$

So, we have:

$$\begin{aligned}
&(1-\rho L)-\left(\zeta_1 L+\zeta_2 L^2+......+\zeta_{p-1}L^{p-1}\right)(1-L)=\\
&1-\rho L-\zeta_1 L+\zeta_1 L^2-\zeta_2 L^2+\zeta_2 L^3-.......-\zeta_{p-1}L^{p-1}+\zeta_{p-1}L^p=\\
&1-(\rho+\zeta_1)L-(\zeta_2-\zeta_1)L^2-(\zeta_3-\zeta_2)L^3-.....-\left(\zeta_{p-1}-\zeta_{p-2}\right)L^{p-1}-\left(-\zeta_{p-1}\right)L^p=\\
&1-[(\varphi_1+\varphi_2+....+\varphi_p)-(\phi_2+\varphi_3+....+\varphi_p)]L-\\
&-[-(\varphi_3+\varphi_4+....+\varphi_p)+(\varphi_2+\varphi_3+.....+\varphi_p)]L^2-.......-\\
&-[-\varphi_p+(\varphi_{p-1}+\varphi_p)]L^{p-1}-\varphi_p L^p=\\
&1-\varphi_1 L-\varphi_2 L^2-.........-\varphi_{p-1}L^{p-1}-\varphi_p L^p
\end{aligned}$$

(3.5.4)

Thus, the autoregression (3.5.1) can be written as:

$$\left[(1-\rho L)-\left(\zeta_1 L+\zeta_2 L^2+......+\zeta_{p-1}L^{p-1}\right)(1-L)\right]y_t=\varepsilon_t \tag{3.5.5}$$

or

$$y_t=\rho y_{t-1}+\zeta_1\Delta y_{t-1}+\zeta_2\Delta y_{t-2}+........+\zeta_{p-1}\Delta y_{t-p+1}+\varepsilon_t \tag{3.5.6}$$

One of the advantages of writing the autoregression (3.5.1) in the form of (3.5.6) is that under the hypothesis of a unit root only one of the regressors in (3.5.6), namely y_{t-1}, is I(1), whereas all the other regressors $\left(\Delta y_{t-1}, \Delta y_{t-2}, \ldots\ldots, \Delta y_{t-p+1}\right)$ are stationary. Since no knowledge of any population parameters is needed to write the process in this form it is convenient to estimate the parameters by direct application of the ordinary least squares (OLS) procedure.

Let us now define the true process that will generate the data in this section as well as the appropriate estimated equation for the construction of the tests concerning the unit root hypothesis.

Assume that the true process that generates the data has the following form:

$$y_t = d_t + \rho y_{t-1} + u_t \qquad (3.5.7)$$

$$u_t = n_1 u_{t-1} + n_2 u_{t-2} + \ldots . n_p u_{t-p} + \varepsilon_t$$

where the error process u_t follows an AR(p) process that has independent identically distributed innovations of finite bounded fourth moment. We are interested in testing the null hypothesis of a unit root that $\rho = 1$ in the equation of the form (3.5.7) which implies that the difference series Δy_t follows a stationary AR(p) process against the alternative that $\rho < 1$ which implies that the series y_t follows an AR(p+1) process.

The equation of the form (3.5.7) is equivalent to the equation of the form:

$$y_t = d_t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots\ldots . + \varphi_{p+1} y_{t-p-1} + \varepsilon_t$$

where

$$\varphi_1 = \rho + n_1$$

$$\varphi_i = n_i - \rho n_{i-1} \qquad \text{for } i = 2,3,\ldots.,p$$

$$\varphi_{p+1} = -\rho \varphi_p$$

with the latter being also equivalent to the equation of the form:

$$y_t = d_t + \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots\dots + \zeta_p \Delta y_{t-p} + \varepsilon_t \qquad (3.5.8)$$

with the parameters ρ and ζ_i being equal to those specified in (3.5.2) and (3.5.3) respectively.

The last defined equation (3.5.8) which includes lagged changes in the dependent variable y_t is the so called **Augmented Dickey-Fuller autoregression** and the test statistics concerning the unit root hypothesis in the equation defined in (3.5.7) are the so called **Augmented Dickey-Fuller unit root test statistics**.

We are ready now to generalize the results of the previous section under the four different cases when the estimated equation is written now in the form of (3.5.8). Before starting doing that, we will give a second proposition taken by Hamilton (1994, p.505) which its results are useful for constructing the limiting distributions of the Augmented Dickey-Fuller unit root test statistics. Again, its proofs are not given as it is not in the scope of this thesis.

Proposition B

Let $u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ and $\{\varepsilon_t\} \sim$ i.i.d$(0, \sigma^2)$ with finite fourth moment.

Define:

$$\gamma_j \equiv \mathrm{E}\left(u_t u_{t-j}\right) = \sigma^2 \sum_{s=0}^{\infty} \psi_s \psi_{s+j} \qquad \text{for } j = 0, 1, 2, \dots\dots$$

$$\lambda = \sigma \sum_{j=0}^{\infty} \psi_j = \sigma\psi(1)$$

$$\xi_t \equiv u_1 + u_2 + \dots\dots + u_t \qquad \text{for } t = 1, 2, \dots, T$$

with $\xi_0 = 0$. Then, as $T \to \infty$:

a) $$T^{-\frac{1}{2}} \sum u_t \xrightarrow{L} \lambda W(1)$$

b) $$T^{-\frac{1}{2}} \sum u_{t-j}\varepsilon_t \xrightarrow{L} N\left(0, \sigma^2 \gamma_0\right)$$

c) $$T^{-1} \sum u_t u_{t-j} \xrightarrow{p} \gamma_j \quad \text{for } j = 1, 2, \ldots$$

d) $$T^{-1} \sum \xi_{t-1}\varepsilon_t \xrightarrow{L} \left(\frac{1}{2}\right)\sigma\lambda\left\{[W(1)]^2 - 1\right\}$$

e) $$T^{-1} \sum \xi_{t-1}u_{t-j} \xrightarrow{L} \left(\frac{1}{2}\right)\left\{\lambda^2 [W(1)]^2 - \gamma_0\right\} \quad \text{for } j = 0$$

$$T^{-1} \sum \xi_{t-1}u_{t-j} \xrightarrow{L} \left(\frac{1}{2}\right)\left\{\lambda^2 [W(1)]^2 - \gamma_0\right\} + \gamma_0 + \gamma_1 + \ldots\ldots + \gamma_{j-1}$$

for $j = 1, 2, \ldots.$

f) $$T^{-\frac{3}{2}} \sum \xi_{t-1} \xrightarrow{L} \lambda \int_0^1 W(r)\,dr$$

g) $$T^{-\frac{3}{2}} \sum t u_{t-j} \xrightarrow{L} \lambda\left\{W(1) - \int_0^1 W(r)\,dr\right\} \qquad \text{for } j = 0, 1, 2, \ldots\ldots$$

h) $$T^{-2} \sum \xi_{t-1}^2 \xrightarrow{L} \lambda^2 \int_0^1 [W(r)]^2\,dr$$

i) $\quad T^{-\frac{5}{2}}\sum t\xi_{t-1} \xrightarrow{L} \lambda\int_0^1 rW(r)dr$

j) $\quad T^{-3}\sum t\xi_{t-1}^2 \xrightarrow{L} \lambda^2\int_0^1 r[W(r)]^2 dr$

k) $\quad T^{-(\nu+1)}\sum t^{\nu} \rightarrow \frac{1}{\nu+1}$ for $\nu = 0 , 1 , 2 , \ldots\ldots$

We are now in the position to generalize the results of the previous section under the four cases when the estimated equation is in the form of (3.5.8).

Case 1

Assume that the true process that generates the data satisfies the equation of the form (3.5.8) with $d_t = 0$ and $\rho = 1$:

$$y_t = \zeta_1\Delta y_{t-1} + \zeta_2\Delta y_{t-2} + \ldots\ldots + \zeta_p\Delta y_{t-p} + y_{t-1} + \varepsilon_t$$

and that the estimated equation is given by the following form:

$$y_t = \zeta_1\Delta y_{t-1} + \zeta_2\Delta y_{t-2} + \ldots\ldots + \zeta_p\Delta y_{t-p} + \rho y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$~ i.i.d.$(0, \sigma^2)$ with finite and bounded fourth moment and the roots of

$$\left(1-\zeta_1 z-\zeta_2 z^2-\ldots\ldots-\zeta_p z^p\right)=0$$

are outside the unit circle. Let $\hat{\zeta}_T = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots\ldots, \hat{\zeta}_{p,T}\right)'$ be the $p \times 1$ vector of estimated OLS coefficients on the lagged changes in y, and let ζ be the corresponding true value. Then, for the normalized least squares estimator, $\hat{\rho}_T$, we have that:

$$Z_{DF} = \frac{\mathrm{T}\left(\hat{\rho}_T - 1\right)}{1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \ldots\ldots - \hat{\zeta}_{p,T}} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\int_0^1 [W(r)]^2 dr} \tag{3.5.9}$$

and for the studentized least squares estimator, $\hat{\rho}_T$, we have that:

$$\tau_{DF} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\left\{\int_0^1 [W(r)]^2 dr\right\}^{\frac{1}{2}}} \tag{3.5.10}$$

where

$$\hat{\sigma}^2_{\hat{\rho}_T} = s_T^2 e'_{p+1} \left(\sum_{t=p+1}^{T} x_t x_t'\right)^{-1} e_{p+1}$$

in which

$$x_t = \left(\Delta y_{t-1}, \Delta y_{t-2}, \ldots\ldots, \Delta y_{t-p}, y_{t-1}\right)'$$

and e_{p+1} denotes a $(p+1) \times 1$ vector with unity in the last position and zeros elsewhere and

$$s_T^2 = (T - p - 1)^{-1} \sum_{t = p+1}^{T} \left(y_t - \hat{\zeta}_{1,T}\Delta y_{t-1} - \ldots.. - \hat{\zeta}_{p,T}\Delta y_{t-p} - \hat{\rho}_T y_{t-1}\right)^2$$

Case 2

Assume that the true process that generates the data satisfies the equation of the form (3.5.8) with $d_t = 0$ and $\rho = 1$:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots\ldots.. + \zeta_p \Delta y_{t-p} + y_{t-1} + \varepsilon_t$$

and that the estimated equation is given by the following form:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots.. + \zeta_p \Delta y_{t-p} + \alpha + \rho y_{t-1} + \varepsilon_t \qquad (3.5.11)$$

where $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$ with finite and bounded fourth moment and the roots of

$$\left(1 - \zeta_1 z - \zeta_2 z^2 - \ldots\ldots - \zeta_p z^p\right) = 0$$

are outside the unit circle. Let $\hat{\zeta}_T = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots\ldots., \hat{\zeta}_{p,T}\right)'$ be the $p \times 1$ vector of estimated OLS coefficients on the lagged changes in y, and let ζ be the corresponding true value. Then, for the normalized least squares estimator, $\hat{\rho}_T$, we have that:

$$Z_{DF} = \frac{\mathrm{T}\left(\hat{\rho}_T - 1\right)}{1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \ldots\ldots - \hat{\zeta}_{p,T}} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\} - W(1)\int_0^1 W(r)dr}{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2}$$

(3.5.12)

and for the studentized least squares estimator, $\hat{\rho}_T$, we have that:

$$\tau_{DF} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{L} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\} - W(1)\int_0^1 W(r)dr}{\left\{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2\right\}^{\frac{1}{2}}} \qquad (3.5.13)$$

where

$$\hat{\sigma}^2_{\hat{\rho}_T} = s_T^2 e'_{p+2}\left(\sum_{t=p+2}^{T} x_t x_t'\right)^{-1} e_{p+2}$$

in which

$$x_t = \left(\Delta y_{t-1}, \Delta y_{t-2}, \ldots\ldots, \Delta y_{t-p}, 1, y_{t-1}\right)'$$

and e_{p+2} denotes a $(p+2)\times 1$ vector with unity in the last position and zeros elsewhere and

$$s_T^2 = [T-(p+2)]^{-1}\sum_{t=p+1}^{T}\left(y_t - \hat{\zeta}_{1,T}\Delta y_{t-1} - \ldots\ldots - \hat{\zeta}_{p,T}\Delta y_{t-p} - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}\right)^2$$

<u>Case 3</u>

Assume that the true process that generates the data satisfies the equation of the form (3.5.8) with an included constant ($d_t = \alpha$) and $\rho = 1$:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots\ldots + \zeta_p \Delta y_{t-p} + \alpha + y_{t-1} + \varepsilon_t$$

and that the estimated equation is given by the following form:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots\ldots + \zeta_p \Delta y_{t-p} + \alpha + \rho y_{t-1} + \varepsilon_t$$

(3.5.14)

where $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$ with finite bounded fourth moment and the roots of

$$\left(1 - \zeta_1 z - \zeta_2 z^2 - \ldots\ldots - \zeta_p z^p\right) = 0$$

are outside the unit circle.

The estimated equation of the form (3.5.14) is equivalent to the following transformed specification:

$$y_t = \zeta_1 \omega_{t-1} + \zeta_2 \omega_{t-2} + \ldots\ldots + \zeta_p \omega_{t-p} + w + \rho y_{t-1} + \varepsilon_t \qquad (3.5.15)$$

where $\omega_t = \Delta y_t - w$ and $w = \dfrac{\alpha}{1 - \zeta_1 - \zeta_2 - \ldots\ldots - \zeta_p}$

Let $\gamma_j = \mathrm{E}\left(\omega_t \omega_{t-j}\right)$ and $\hat{\zeta}_T = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots\ldots\ldots \hat{\zeta}_{p,T}\right)^{/}$ be the $p \times 1$ vector of estimated OLS coefficients of $\left(\omega_{t-1}, \omega_{t-2}, \ldots\ldots, \omega_{t-p}\right)$ which are identical to the coefficients of $\left(\Delta y_{t-1}, \Delta y_{t-2}, \ldots\ldots\ldots \Delta y_{t-p}\right)$ in the original estimated equation. Now, if $\rho = 1$ and $\alpha \neq 0$, we have that:

$$\begin{bmatrix} \mathrm{T}^{\frac{1}{2}}\left(\hat{\zeta}_T - \zeta\right) \\ \mathrm{T}^{\frac{1}{2}}\left(\hat{\mu}_T - \mu\right) \\ \mathrm{T}^{\frac{3}{2}}\left(\hat{\rho}_T - 1\right) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} V & 0 & 0 \\ 0^{/} & 1 & \frac{\mu}{2} \\ 0^{/} & \frac{\mu}{2} & \frac{\mu^2}{3} \end{bmatrix}^{-1} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \qquad (3.5.16)$$

where

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} V & 0 & 0 \\ 0' & 1 & \frac{\mu}{2} \\ 0' & \frac{\mu}{2} & \frac{\mu^2}{3} \end{bmatrix} \right) \quad \text{and} \quad V = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdot\ \cdot\ \cdot & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdot\ \cdot\ \cdot & \gamma_{p-3} \\ \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \gamma_{p-2} & \gamma_{p-3} & \cdot\ \cdot\ \cdot & \gamma_0 \end{bmatrix}$$

As we can see from (3.5.16), $\hat{\rho}_T$ converges at rate $T^{3/2}$ to a Gaussian variable and all the other estimated coefficients converge at rate $T^{1/2}$ to Gaussian variables. So, any tests involving the above coefficients of the estimated equation of the form (3.5.14) can be compared with the usual Tables.

Case 4

Assume that the true process that generates the data satisfies the equation of the form (3.5.8) with an included constant ($d_t = \alpha$) and $\rho = 1$:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots\ldots + \zeta_p \Delta y_{t-p} + \alpha + y_{t-1} + \varepsilon_t$$

and that the estimated equation is given by the following form:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots\ldots + \zeta_p \Delta y_{t-p} + \alpha + \rho y_{t-1} + \delta t + \varepsilon_t \qquad (3.5.17)$$

where $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$ with finite bounded fourth moment and the roots of

$$\left(1 - \zeta_1 z - \zeta_2 z^2 - \ldots\ldots - \zeta_p z^p\right) = 0$$

are outside the unit circle. The estimated equation of the form (3.5.17) is equivalent to the following transformed specification:

$$y_t = \zeta_1 \omega_{t-1} + \zeta_2 \omega_{t-2} + \ldots\ldots + \zeta_p \omega_{t-p} + w^* + \rho m_{t-1} + \delta^* t + \varepsilon_t \qquad (3.5.18)$$

where $\omega_t = \Delta y_t - w$ and $w = \dfrac{\alpha}{1-\zeta_1-\zeta_2-\ldots..-\zeta_p}$ and $w^* = (1-\rho)w$ and $m_{t-1} = y_{t-1} - \mu(t-1)$ and $\delta^* = \delta + \rho\mu$.

Let $\gamma_j = \mathrm{E}\left(\omega_t \omega_{t-j}\right)$ and $\hat{\zeta}_T = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots\ldots.., \hat{\zeta}_{p,T}\right)'$ be the $p \times 1$ vector of estimated OLS coefficients of $\left(\omega_{t-1}, \omega_{t-2}, \ldots\ldots, \omega_{t-p}\right)$ which are identical to the coefficients of $\left(\Delta y_{t-1}, \Delta y_{t-2}, \ldots\ldots\ldots \Delta y_{t-p}\right)$ in the original estimated equation. Now, if $\rho = 1$ and $\delta = 0$, and $\lambda = \dfrac{\sigma}{1-\zeta_1-\zeta_2-\ldots\ldots-\zeta_p}$ we have that:

$$\begin{bmatrix} T^{\frac{1}{2}}\left(\hat{\zeta}_T - \zeta\right) \\ T^{\frac{1}{2}} w^* \\ T\left(\hat{\rho}_T - 1\right) \\ T^{\frac{3}{2}}\left(\hat{\delta}^*_T - \delta^*\right) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} V & 0 & 0 & 0 \\ 0' & 1 & \lambda \int_0^1 W(r)dr & \frac{1}{2} \\ 0' & \lambda \int_0^1 W(r)dr & \lambda^2 \int_0^1 [W(r)]^2 dr & \lambda \int_0^1 rW(r)dr \\ 0' & \frac{1}{2} & \lambda \int_0^1 rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} l_1 \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{[W(1)]^2 - 1\right\} \\ \sigma\left\{W(1) - \int_0^1 W(r)dr\right\} \end{bmatrix}$$

(3.5.19)

where $l_1 \sim \mathrm{N}\left(0, \sigma^2 V\right)$ and V has been defined in the previous case .

From (3.5.19) we can deduce that:

$$T^{\frac{1}{2}}\left(\hat{\zeta}_T - \zeta\right) \xrightarrow{L} N\left(0, \sigma^2 V^{-1}\right) \tag{3.5.20}$$

For the normalized least squares estimator, $\hat{\rho}_T$, we have that:

$$Z_{DF} = \frac{T\left(\hat{\rho}_T - 1\right)}{1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \dots\dots - \hat{\zeta}_{p,T}} \xrightarrow{L}$$

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 W(r)dr & \frac{1}{2} \\ \int_0^1 W(r)dr & \int_0^1 [W(r)]^2 dr & \int_0^1 rW(r)dr \\ \frac{1}{2} & \int_0^1 rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \frac{1}{2}\left\{[W(1)]^2 - 1\right\} \\ W(1) - \int_0^1 W(r)dr \end{bmatrix} = B \tag{3.5.21}$$

and for the studentized least squares estimator, $\hat{\rho}_T$, we have that:

$$\tau_{DF} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{L} \frac{B}{\sqrt{Q}}$$

where

$$Q = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 W(r)dr & \frac{1}{2} \\ \int_0^1 W(r)dr & \int_0^1 [W(r)]^2 dr & \int_0^1 rW(r)dr \\ \frac{1}{2} & \int_0^1 rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

3.6 REMARKS ON THE ADF – STATISTICS

The assumption that the error process in the simple autoregression is uncorrelated (i.i.d. or White Noise) is too restrictive for the real world. A major reason is that it is hard to believe that when the standard Dickey-Fuller test statistics reject the unit root hypothesis the stationary AR(1) process with or without a constant or a linear time trend describes well most of the economic and financial time series in practice. Another important reason is that if we allow for the error process to be serially correlated, for example to be an AR(p) with p>1 process, the standard Dickey-Fuller unit root test statistics are not any more valid. Dickey and Fuller (1979) approached the latter more realistic case by modifying the simple autoregression and got valid unit root test statistics, the so called Augmented Dickey-Fuller unit root test statistics. The latter case is actually covers the case in which the difference series follows a stationary AR(p) process around or not a constant or a linear time trend under the hypothesis of a unit root whereas the series follows an AR(p+1) process under the alternative.

As can also be noticed, the asymptotic distributions of the Augmented Dickey-Fuller unit root test statistics are identical to those of standard Dickey-Fuller unit root test statistics. The reason behind this result is that in the regression of I(1) variable on I(0) and I(1) variables, the asymptotic distribution of the coefficient of I(1) and I(0) variables are independent (G.S.Maddala and In-Moo Kim 2002).

Furthermore, the unit root test statistics proposed by Dickey-Fuller are based on finite-order autoregressive error processes, the orders of which assumed to be known. This assumption rarely holds in applied time series analysis and for this reason suggestions in selecting the unknown but finite p are needed. The last fact motivated Said and Dickey (1984) to extend firstly the Augmented Dickey-Fuller unit root tests in the general class of ARMA(p,q) error processes covering actually the case in which, under the hypothesis of a unit root the series of the first-differences are of the general ARMA(p,q) form with p and q unknown and secondly to show that there is no need to have exact knowledge on the parameters p and q if some specified conditions hold. These issues are of great importance and are going to be fully analyzed in the next chapter.

CHAPTER 4

UNIT ROOT TESTING UNDER ARMA ERROR PROCESSES

4.1 INTRODUCTION

The standard Dickey-Fuller (DF) unit root testing procedures assume independent identically distributed errors and the Augmented Dickey-Fuller (ADF) unit root testing procedures take into account the possibility of the serial correlation of errors. In particular, they allow for the serial correlation to belong to the wide class of autoregressive (AR) processes and actually cover the case in which, under the hypothesis of a unit root the differenced-data (Δy_t) follow a stationary AR(p) process while under the alternative the data themselves (y_t) follow an AR(p+1) process. A key shortcoming for applying the Augmented Dickey-Fuller unit root tests is that the practitioner must have prior knowledge about the order of the error process, knowledge that is not possible to get in practice since the true mechanism that generates the data is in most cases unknown. In fact, if we knew that the error process follows an AR process with a known order p, then the order of the Augmented Dickey-Fuller autoregression should be set to the value of p before applying the unit root testing procedures.

The Augmented Dickey-Fuller (ADF) unit root testing procedures were later extended by Said and Dickey (1984) to allow for the serial correlation to belong to the general class of autoregressive moving average (ARMA) error processes and cover the case in which, under the hypothesis of a unit root, the series of the first-differences are of the general stationary ARMA(p,q) form with p and q unknown while under the alternative the data themselves follow an ARMA(p+1,q) process. More specifically, in the case where the error process includes moving average components, they showed that the Augmented Dickey-Fuller autoregression is still valid for testing the unit root null if the number of lagged terms of Δy_t introduced as regressors increases with the sample size at a controlled rate $T^{1/3}$. Essentially, the moving average error terms are being approximated by including enough autoregressive terms.

In this Chapter, we present Said and Dickey's results. It is possible to approximate an ARIMA(p, 1, q) process by a truncated autoregression (which is actually the Augmented Dickey-Fuller autoregression with an infinite order) whose order is a function of the number of observations T. Using the least squares procedure to estimate the coefficients in this autoregressive approximation produces test statistics whose limiting distributions are the same as those tabulated in Chapter 3 and proposed by Dickey and Fuller (1976, 1979). Thus, it is possible to test the null hypothesis of a unit root without knowing p or q.

This Chapter is organized as follows. In section 4.2, we present the Said and Dickey's key results for testing the hypothesis of a unit root in data whose innovations exhibit all sorts of dynamics best represented by some ARMA(p,q) process. Then, in section 4.3 we present Said and Dickey's modeling approach to this specific and of great practical interest unit root hypothesis testing issue. Section 4.4 is concerned with an extensive discussion on the Said and Dickey's key findings. In the end of this Chapter we give an Appendix that is involved with some key mathematical proofs on Said and Dickey's modeling approach in order their development to be more understandable.

4.2 SAID AND DICKEY'S KEY RESULTS

Said and Dickey (1984) considered the following process with the serial correlation in errors described by:

$$y_t = \rho y_{t-1} + u_t \qquad \text{for } t = 1,2,3,\ldots\ldots$$

and

$$u_t = \sum_{i=1}^{p} \phi_i u_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad \text{for } t = \ldots.,-2,-1,0,1,2,\ldots. \tag{4.2.1}$$

and it is assumed that $\{u_t\}$ is a stationary and invertible ARMA(p, q) process with p and q unknown, $y_0 = 0$ and $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$ with finite fourth moment. In the case where the true mechanism that generates the data is assumed to satisfy the general process of the form (4.2.1) it holds that, if $\rho < 1$ then $\{y_t\} \sim$ ARMA(p+1,q) whereas if $\rho = 1$ then $\{y_t\} \sim$ ARIMA (p,1,q). The null hypothesis to be tested is that of $\rho = 1$. Notice that, under the null hypothesis that $\rho = 1$ the process of the form (4.2.1) can be written in the form:

$$\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{\infty} \zeta_i u_{t-i} + \varepsilon_t \tag{4.2.2}$$

where the coefficients ζ_i $(i = 1,2,3,\ldots., \infty)$ are appropriate functions of the parameters $\{\phi_i, \theta_j, i = 1,2,..,p, j = 1,2,\ldots,q\}$.

Under the null hypothesis that $\rho = 1$ it also holds that:

$$u_t = y_t - y_{t-1} = \Delta y_t$$

Said and Dickey (1984) proceeded with estimating the coefficients in (4.2.2) under the unit root hypothesis by regressing the data of the first-differences Δy_t on y_{t-1}, Δy_{t-1}, …… , Δy_{t-k} where k is an appropriately chosen integer since there is no knowledge about p and q and it is impossible to estimate an infinite number of parameters from a finite sample of observations T. More specifically, Said and Dickey (1984) have shown that to account for the time-dependence of the general ARMA(p,q) form in the errors under the unit root null, one can augment the regression of the form (4.2.2) with, say, k lagged terms of Δy_t:

$$\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk} \qquad (4.2.3)$$

The purpose of the number of the lagged terms of Δy_{t-i} included in the above equation is to ensure that the innovations $\left\{\varepsilon_{tk},\ t = 1,2,\ldots,T\right\}$ will behave approximately as independent identically distributed. Then, they proved that in order to get consistent estimates of the coefficients and test statistics it is necessary to let k be a function of the number of the observations T. They actually gave the following two conditions for the above purpose:

(A1) $\frac{k^3}{T} \to 0$ and $k \to \infty$ as $T \to \infty$

(A2) there exists $c > 0$ and $r > 0$ such that $ck > T^{1/r}$, meaning that k is bounded below by a positive multiple of $T^{1/r}$ for some $r > 0$.

In this case, the limiting distribution of the τ_{DF} –test statistic for the coefficient on the lagged dependent variable y_{t-1} considered in Chapter 3 and is the basis for testing the unit root hypothesis have the same τ_{DF}-Dickey-Fuller distribution as when the errors are independent identically distributed.

The last result extends naturally to the inclusion of deterministic components in (4.2.1). In that case the Wiener (Brownian Motion) process is replaced by its detrended counterpart.

This second condition is actually a lower bound condition that restricts k to be at least a polynomial rate in T and its role is to force k to be large enough that the truncated autoregression of the form (4.2.3) is a good approximation to the autoregression of the form of (4.2.2). We will see that (A2) enters the analysis only when considering the stationary lagged terms Δy_{t-i}.

In practice, Said and Dickey (1984) showed that it would be sufficient to increase the number of lagged terms on Δy_t as a function of the sample size, at a rate no faster than $T^{1/3}$. This does not suggest, however, what should be the appropriate value of k in a given application. This is no trivial matter. Choosing too small a k will distort the size (significance level) of the test statistics. Choosing too big a k will reduce the degrees of freedom and the power of the tests, which will too often lead falsely to the non-rejection of the unit root hypothesis. So, selection methods are needed for choosing the appropriate lag number k to be included in the autoregression of the form (4.2.3) provided that these selection methods must satisfy the Said and Dickey's conditions. These issues are going to be fully analyzed in the next sections.

4.3 SAID AND DICKEY'S MAIN STEPS IN THE PROOFS

Said and Dickey (1984) first considered the case where the errors belong to the simple stationary and invertible ARMA(1,1) process with normal innovations and then extended to the more general case where the errors belong to the stationary and invertible ARMA(p,q) processes of unknown order and with independent identically distributed innovations.

We will present the Said and Dickey's approach to the already mentioned unit root hypothesis testing problem in the case where the errors follow a stationary and invertible ARMA(1,1) process and then the extension to the general case where the errors follow a stationary and invertible ARMA(p,q) process of unknown p and q case comes up naturally with analogous statements.

Assume that the true mechanism that generates the data is given by the following equation with the error process described by:

$$y_t = \rho y_{t-1} + u_t \qquad \text{for } t = 1, 2, 3, \ldots\ldots\ldots$$

and

$$u_t = \phi u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \qquad \text{for } t = \ldots\ldots, -2, -1, 0, 1,2, \ldots.$$

$$(4.3.1)$$

and it is assumed that $|\ \varphi\ | < 1$, $|\ \theta\ | < 1$, $y_0 = 0$ and $\{\varepsilon_t\} \sim$ i.i.d.N$(0, \sigma^2)$ with finite fourth moment. If $\rho < 1$ then $\{y_t\} \sim$ ARMA(2,1) whereas if $\rho = 1$ then $\{y_t\} \sim$ ARIMA (1, 1, 1). The desired unit root hypothesis that $\rho = 1$ and will be tested. Under the null hypothesis that $\rho = 1$, the process of the form (4.3.1) can be written in the form:

$$y_t - y_{t-1} = (\rho - 1)y_{t-1} + (\phi + \theta)\left(u_{t-1} - \theta u_{t-2} + \theta^2 u_{t-3} - \ldots\ldots\right) + \varepsilon_t$$

or

$$\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{\infty} \zeta_i u_{t-i} + \varepsilon_t \tag{4.3.2}$$

where $\zeta_0 = \rho - 1$ and $\zeta_i = (\varphi+\theta)(-\theta)^{i-1}$, $i > 0$.

Under the null hypothesis that $\rho = 1$ it also holds that:

$$u_t = y_t - y_{t-1} = \Delta y_t$$

Said and Dickey (1984) proceeded with estimating the coefficients in (4.3.2) under the unit root hypothesis by regressing the data of the first-differences Δy_t on y_{t-1}, Δy_{t-1} ,…., Δy_{t-k} where k is an appropriately chosen integer since it is impossible to estimate an infinite number of parameters from a finite sample of observations T.

Letting $\zeta_0 = \rho - 1$ and $\zeta_i = (\varphi+\theta)(-\theta)^{i-1}$, $i > 0$, be the coefficients in (4.3.2) and $X_t = (u_{t-1}, u_{t-2}, \ldots\ldots, u_{t-k})'$, $U_t = \left(y_{t-1}, X_t'\right)$ and $\zeta' = (\zeta_0, \zeta_1, \ldots\ldots, \zeta_k)$, the truncated version of (4.3.2) under the unit root hypothesis is:

$$\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk} \tag{4.3.3}$$

Notice that the truncated innovations, $\left\{\varepsilon_{tk},\ t = 1,2,\ldots.,T\right\}$, are not independent identically distributed random variables for any k. In fact $\varepsilon_{tk} = \Delta y_t - U_t' \zeta$. If we define $\hat{\zeta} = \left(\hat{\zeta}_{0,T}, \hat{\zeta}_{1,T}, \ldots\ldots\ldots \hat{\zeta}_{k,T}\right)$ to be the vector of the estimated coefficients in the truncated autoregression of the form (4.3.3), the OLS deviation of the estimated coefficient vector $\hat{\zeta}$ from the true-valued coefficient vector ζ is given by the formula:

$$\left(\hat{\zeta} - \zeta\right) = \left(\Sigma U_t U_t'\right)^{-1} \left(\Sigma U_t \varepsilon_{tk}\right) \tag{4.3.4}$$

where $\sum$ denotes hereafter summation over the observations t = k+1, …., T.

Consulting the rates of convergence for this case in order to get statistical meaningful test statistics and limiting distributions, we define the beneath $(k+1)\times(k+1)$ diagonal matrix:

$$D_T = diag\left\{(T-k)^{-1}, (T-k)^{-1/2}, (T-k)^{-1/2}, \ldots\ldots, (T-k)^{-1/2}\right\} \tag{4.3.5}$$

We also let $D_T(1)$ be the first diagonal element of the diagonal matrix D_T and $D_T(k)$ be the lower right $(k \times k)$ block of D_T.

Pre-multiplying equation (4.3.4) by D_T^{-1} we have that:

$$D_T^{-1}\left(\hat{\zeta}-\zeta\right) = \left(D_T R_0 D_T\right)^{-1}\left(D_T \Sigma U_t \varepsilon_{tk}\right) \tag{4.3.6}$$

where

$$R_0 = \Sigma U_t U_t^{/} = \begin{bmatrix} \Sigma y_{t-1}^2 & \Sigma y_{t-1} X_t^{/} \\ \Sigma y_{t-1} X_t & \Sigma X_t X_t^{/} \end{bmatrix} \tag{4.3.7}$$

and

$$D_T R_0 D_T = \begin{bmatrix} (T-k)^{-2} \Sigma y_{t-1}^2 & (T-k)^{-3/2} \Sigma y_{t-1} X_t^{/} \\ (T-k)^{-3/2} \Sigma y_{t-1} X_t & (T-k)^{-1} M_k \end{bmatrix}$$

where

$$M_k = \Sigma X_t X_t^{/}$$

We also let $R_0(1)$ be the first diagonal element of the matrix R_0. Using also the beneath definitions of: $\gamma_u(i-j) = \mathrm{E}\left(u_{t-i} u_{t-j}\right)$, which denote the autocovariance function of the stationary series u_t, $\Gamma_{ij} = \gamma_u(i-j)$, where Γ is a $k \times k$ matrix, $\gamma^{/} = \{\gamma_u(1), \gamma_u(2), \ldots\ldots, \gamma_u(k)\}$, $W_t = \varepsilon_1 + \varepsilon_2 + \ldots\ldots + \varepsilon_t$ and the block diagonal matrix $R = diag\left\{(1-\phi)^{-2}(1+\theta)^2(T-k)^{-2}\sum W_{t-1}^2, \Gamma\right\}$, we are in the position to present some useful findings of Said and Dickey (1984) that are involved with the notion of || . || to signify the usual Euclidean norm. These are going to be presented in terms of five steps in order to be simplified.

Let us reconsider the formula (4.3.6) in order to prove firstly the consistency of the least squares estimated coefficient vector $\hat{\zeta}$:

Step 1: Asymptotic equivalence of $\mathbf{D_T R_0 D_T}$ and R

$$D_T^{-1}(\hat{\zeta} - \zeta) = (D_T R_0 D_T)^{-1}(D_T \Sigma U_t \varepsilon_{tk}) = \left(D_T^{-1} R_0^{-1} D_T^{-1}\right)(D_T \Sigma U_t \varepsilon_{tk}) =$$
$$\left(D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1} + R^{-1}\right)(D_T \Sigma U_t \varepsilon_{tk}) =$$
$$\left(D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1}\right)\left(D_T \sum U_t \varepsilon_{tk}\right) + \left(R^{-1} D_T \sum U_t \varepsilon_{tk}\right)$$

Taking now the usual Euclidean norm of both sides above we obtain:

$$\| D_T^{-1}(\hat{\zeta} - \zeta) \| = \| (D_T R_0 D_T)^{-1}(D_T \Sigma U_t \varepsilon_{tk}) \| = \| \left(D_T^{-1} R_0^{-1} D_T^{-1}\right)(D_T \Sigma U_t \varepsilon_{tk}) \| =$$
$$\| \left(D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1} + R^{-1}\right)(D_T \Sigma U_t \varepsilon_{tk}) \| =$$
$$\| \left(D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1}\right)\left(D_T \sum U_t \varepsilon_{tk}\right) + \left(R^{-1} D_T \sum U_t \varepsilon_{tk}\right) \| \leq$$
$$\| \left(D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1}\right) \| \| \left(D_T \sum U_t \varepsilon_{tk}\right) \| + \| R^{-1} \| \| \left(D_T \sum U_t \varepsilon_{tk}\right) \|$$

Said and Dickey (1984, p.601-603) proved that:

$$k^{1/2} \| \left(D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1}\right) \| = o_p(1) \quad \text{and} \quad \| R^{-1} \| = O_p(1)$$

under the condition $\frac{k^3}{T} \longrightarrow 0$ as $T \longrightarrow \infty$, where $o_p(1)$ and $O_p(1)$ denote convergence to zero and boundness in probability, respectively, as their next Theorem and Lemma claim:

Theorem 1

Under the assumptions of the process (4.2.1) with $\rho = 1$ and $T^{-\frac{1}{3}} k \rightarrow 0$ (notice that this is the first part of the condition (A1)), it holds that $k^{\frac{1}{2}} \| D_T^{-1} R_0^{-1} D_T^{-1} - R^{-1} \| = o_p(1)$.

Lemma 1

Under the assumptions of the process (4.2.1) with $\rho = 1$, $\|R^{-1}\| = O_p(1)$.

Step 2: Consistency of $\hat{\zeta} = \left(\hat{\zeta}_{0,T}, \hat{\zeta}_{1,T}, \ldots\ldots \zeta_{k,T}\right)$

It now remains to consider the term $\left(D_T \Sigma U_t \varepsilon_{tk}\right)$, which is actually the second term of (4.3.6), in order to find its order. Said and Dickey (1984) gave the following Lemma that is concerned with it:

Lemma 2

Under the conditions (A1) and (A2), as $T \rightarrow \infty$:

a) $\| D_T \Sigma U_t \left(\varepsilon_{tk} - \varepsilon_t\right) \| = O_p(T^{-1})$

b) $\| D_T \Sigma U_t \varepsilon_t \| = O_p(k^{\frac{1}{2}})$

c) $\| D_T \sum U_t \varepsilon_{tk} \| = O_p(k^{\frac{1}{2}})$

<u>Proof</u>: See Appendix of Chapter 4

The above mathematical proof is actually based on the second part of the condition (A1) as well as on the condition (A2). Through Lemmas 1, 2 and Theorem 1, Said and Dickey (1984) proved the consistency of the least squares estimated coefficient vector $\hat{\zeta}$ as the next theorem claims:

<u>Theorem 2</u>

Under the assumptions of the process (4.3.1) with $\rho = 1$, it holds that $\| \hat{\zeta} - \zeta \| \xrightarrow{p} 0$ under the conditions (A1) and (A2).

<u>Proof</u>: See Appendix of Chapter 4

To summarize, Said and Dickey (1984) proved that in order to get consistent estimates of the coefficients and test statistics it is necessary to let k be a function of the number of the observations T. Let us now consider the rates of convergence of $\hat{\zeta}_{0,T}$ and $\hat{\zeta}(k) = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{k,T}\right)$ under the two imposed conditions (A1) and (A2).

Step 3: The vector $\hat{\zeta}(k) = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{k,T}\right)$ is $\sqrt{T}$ - consistent

The rate at which the vector $\hat{\zeta}(k) = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{k,T}\right)$ converges can be found by writing:

$$
\begin{aligned}
&(T-k)^{1/2}(\hat{\zeta}(k)-\zeta(k)) = (D_T(k)M_k D_T(k))^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_{tk} = \\
&R^{*-1}(T-k)^{-1/2}\sum X_t\varepsilon_{tk} = (R^{*-1}-\Gamma^{-1}+\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk} = \\
&(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk} + \Gamma^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_{tk} = \\
&(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk} + \Gamma^{-1}(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t+\varepsilon_t) = \\
&(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk} + \Gamma^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_t + \\
&+\Gamma^{-1}(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t)
\end{aligned}
$$

where R^{*-1} denotes $(D_T(k)M_k D_T(k))^{-1}$.

Taking now the usual Euclidean norm of both sides above we obtain:

$$
\begin{aligned}
&\| (T-k)^{1/2}(\hat{\zeta}(k)-\zeta(k)) \| = \\
&\| (R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk} + \Gamma^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_t + \\
&\Gamma^{-1}(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t) \| \le \\
&\| (R^{*-1}-\Gamma^{-1}) \| \| (T-k)^{-1/2}\sum X_t\varepsilon_{tk} \| + \\
&\| \Gamma^{-1} \| \| (T-k)^{-1/2}\sum X_t\varepsilon_t \| + \| \Gamma^{-1} \| \| (T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t) \| \\
&= O_p(1) + O_p(k^{1/2}) + O_p(T^{-1}) \Leftrightarrow \\
&\| \hat{\zeta}(k)-\zeta(k) \| = O_p(T^{-1/2}) + O_p(k^{1/2}T^{-1/2}) + O_p(T^{-3/2}) = O_p(T^{-1/2}) \Rightarrow
\end{aligned}
$$

$$\sqrt{T}(\hat{\zeta}(k)-\zeta(k)) = O_p(1) \quad \text{for } i=1,2,3,\ldots,T.$$

So, the estimated coefficient vector $\hat{\zeta}(k) = \left(\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{k,T}\right)$ is $\sqrt{T}$ – consistent under both conditions (A1) and (A2).

Step 4: The coefficient $\hat{\zeta}_{0,T}$ is T – consistent

As far as the rate of convergence of the estimated coefficient $\hat{\zeta}_{0,T}$ is concerned, it holds that:

$$(T-k)(\hat{\zeta}_{0,T}-\zeta_0)=(D_T(1)R_0(1)D_T(1))^{-1}(T-k)^{-1}\Sigma y_{t-1}\varepsilon_{tk}=$$
$$R^{**-1}(T-k)^{-1}\Sigma y_{t-1}\varepsilon_{tk}=(R^{**-1}-R^{-1}+R^{-1})(T-k)^{-1}\Sigma y_{t-1}\varepsilon_{tk}=$$
$$(R^{**-1}-R^{-1})(T-k)^{-1}\Sigma y_{t-1}\varepsilon_{tk}+R^{-1}(T-k)^{-1}\Sigma y_{t-1}\varepsilon_{tk}=$$
$$(R^{**-1}-R^{-1})(T-k)^{-1}\Sigma y_{t-1}\varepsilon_{tk}+R^{-1}(T-k)^{-1}\Sigma y_{t-1}(\varepsilon_{tk}-\varepsilon_t+\varepsilon_t)=$$
$$(R^{**-1}-R^{-1})(T-k)^{-1}\sum y_{t-1}\varepsilon_{tk}+R^{-1}(T-k)^{-1}\sum y_{t-1}\varepsilon_t+$$
$$R^{-1}(T-k)^{-1}\sum y_{t-1}(\varepsilon_{tk}-\varepsilon_t)$$

where R^{**-1} denotes $(D_T(1)R_0(1)D_T(1))^{-1}$.
Taking now the usual Euclidean norm of both sides above we obtain:

$$\| (T-k)(\hat{\zeta}_{0,T}-\zeta_0)\| =$$
$$\| (R^{**-1}-R^{-1})(T-k)^{-1}\sum y_{t-1}\varepsilon_{tk}+R^{-1}(T-k)^{-1}\sum y_{t-1}\varepsilon_t+$$
$$R^{-1}(T-k)^{-1}\sum y_{t-1}(\varepsilon_{tk}-\varepsilon_t)\| \leq$$
$$\| (R^{**-1}-R^{-1})\| \| (T-k)^{-1}\sum y_{t-1}\varepsilon_{tk}\| +\| R^{-1}\| \| (T-k)^{-1}\sum y_{t-1}\varepsilon_t\| +$$
$$\| R^{-1}\| \| (T-k)^{-1}\sum y_{t-1}(\varepsilon_{tk}-\varepsilon_t)\|$$
$$=O_p(1)+O_p(k^{1/2})+O_p(1)\Leftrightarrow$$
$$\| \hat{\zeta}_{0,T}-\zeta_0\| =O_p(T^{-1})+O_p(k^{1/2}T^{-1})+O_p(T^{-1})=O_p(T^{-1})\Rightarrow$$

$$T(\hat{\zeta}_{0,T}-\zeta_0)=O_p(1)$$

meaning that $\hat{\zeta}_{0,T}$ is T-consistent under the condition (A1), whether or not the lower bound condition (A2) is satisfied.

Step 5: Asymptotic convergence of $\hat{\zeta}_{0,T}-\zeta_0$ to the DF- distribution

Their next step was to derive the limiting distributions of the Dickey-Fuller unit root test statistics. The first element of $D_T^{-1}\left(\hat{\zeta}-\zeta\right)$ is:

$$(T-k)(\hat{\rho}_T-1)=\left[(T-k)^{-2}\sum y_{t-1}^2\right]^{-1}(T-k)^{-1}\sum y_{t-1}\varepsilon_{tk}$$

Using the beneath lemma:

Lemma 3

For $T\to\infty$:

a) $\dfrac{1}{(T-k)^2}\sum y_{t-1}^2=\dfrac{(1+\theta)^2}{(1-\phi)^2}\sum W_{t-1}^2+O_p(T^{-\frac{1}{2}})$

b) $\dfrac{1}{(T-k)}\sum y_{t-1}\varepsilon_{tk}=\dfrac{1}{(T-k)}\sum y_{t-1}\varepsilon_t+O_p(T^{-1})$

c) $\dfrac{1}{(T-k)}\sum y_{t-1}\varepsilon_t=\dfrac{1}{(T-k)}\dfrac{(1+\theta)}{(1+\phi)}\sum W_{t-1}\varepsilon_t+O_p(T^{-\frac{1}{2}})$

along with the definition of Dickey-Fuller (1979) for the random variables (Γ,ξ) where $\Gamma=\sum_{i=1}^{\infty}\omega_i^2Z_i^2$ and $\xi=\left(\sum_{i=1}^{\infty}\omega_iZ_i^2\right)-\frac{1}{2}$ where $Z_i \sim$ i.i.d.N (0,1) and $\omega_i=(-1)^{i+1}\left[\dfrac{2}{(2i-1)\pi}\right]$ and their proof that:

$$\left(\sigma^2T^{-2}\sum W_{t-1}^2,\sigma^2T^{-1}\sum W_{t-1}\varepsilon_t\right)\xrightarrow{L}(\Gamma,\xi)$$

they showed that the distribution of $\Gamma^{-1}\xi$ is the same as the limiting distribution of:

$$(1-\phi)^{-1}(1+\theta)(T-k)(\hat{\rho}_T-1)$$

The limiting distribution of the above Z_{DF} - test statistic involves the unknown parameters φ and θ and for this reason can not be used directly for unit root testing. However, these unknown parameters can be consistently estimated and there exists a transformation of the Z_{DF} – test statistic which eliminates these parameters asymptotically (Xiao and Phillips (1998)) but such considerations are not in the scope of this Thesis. In contrast, the limiting distribution of the τ_{DF} –test statistic associated with $\hat{\rho}_T$ does not depend on any unknown parameters as the next theorem claims:

<u>Theorem 3</u>

Under the assumptions of Theorem 1, define:

$$\tau_{DF} = \left(C_{11}\hat{\sigma}^2\right)^{-\frac{1}{2}}\left(\hat{\rho}_T - 1\right)$$

where C_{11} is the upper left-hand corner element of R_0^{-1} , $\hat{\rho}_T - 1 = \hat{\zeta}_{0,T} - \zeta_0$ and $\hat{\sigma}^2$ is the error mean square from regression (4.3.3). Then the limiting distribution of τ_{DF}-test statistic is equal to the distribution of $\Gamma^{-1/2}\xi$ which is equivalent to the distribution of the form (3.3.10).

So, the limiting distribution of the studentized τ_{DF} – test statistic of the coefficient on the lagged dependent variable y_{t-1} for the case considered in this section has the same τ_{DF} - Dickey-Fuller distribution as when the errors are independent identically distributed.

Said and Dickey (1984) proved that the above results(Theorems and Lemmas) are still valid to the higher order ARMA error processes of unknown order and for this reason they generalized them immediately.

4.4 RELAXING SAID AND DICKEY'S CONDITIONS: THE NG AND PERRON'S EXTENSIONS

We saw that Said and Dickey (1984) imposed the following conditions in order to guaranty the consistency of the least squares estimated coefficient vector $\hat{\zeta} = \left(\hat{\zeta}_{0,T}, \hat{\zeta}_{1,T}, \ldots\ldots\zeta_{k,T}\right)$:

(A1) k is chosen as a function of T such that:

$$\frac{k^3}{T} \to 0 \quad \text{and} \quad k \to \infty \text{ as } T \to \infty$$

(A2) there exists c > 0 and r > 0 such that:

$$ck > T^{\frac{1}{r}}$$

meaning that k is bounded below by a positive multiple of $T^{\frac{1}{r}}$ for some strictly positive r. This second is actually a lower bound condition that restricts k to grow at least at a polynomial rate in T and its role is to force k to be large enough that the truncated Augmented Dickey-Fuller autoregression of the form (4.3.3) is a good approximation to the true process of the form (4.3.2) and enters the analysis only as a condition to ensure the $\sqrt{T}$-consistency of the estimated coefficient vector $\hat{\zeta}(k) = \left(\hat{\zeta}_{1,T},\ldots\ldots,\hat{\zeta}_{k,T}\right)$ (see Step 3 and its mathematical proof in Appendix in the end of the Chapter).

Condition (A1) has also been used by Berk (1974) who dealt with approximating a stationary process $\{\omega_t\}$ by fitting a truncated autoregression of order, say, k:

$$\omega_t = \sum_{i=1}^{k} d_i \omega_{t-i} + \varepsilon_{tk}$$

In particular, he proved that under the condition (A1) and

(A3) k satisfies $k^{\frac{1}{2}} \sum_{i=k+1}^{\infty} |d_i| \to 0$ and $k \to \infty$ as $T \to \infty$

the estimates involved in the fitted truncated autoregression are consistent. The last condition (A3) is satisfied for any stationary and invertible ARMA process as long as $k \to \infty$ as $T \to \infty$, irrespective of the rate at which k increases. Condition (A3) is more relaxed than the condition (A2) as it allows k to grow at slower rates compared to (A2) which restricts k to grow at least at a polynomial rate. In addition, Berk (1974) and later Lewis and Reinsel (1985) in a related work of approximating a stationary process by fitting a truncated autoregression of order, say, k used the condition (A1) along with the condition below:

$$\text{(A4) k satisfies} \quad T^{\frac{1}{2}} \sum_{i=k+1}^{\infty} |d_i| \to 0 \text{ as } k \to \infty \text{ and } T \to \infty$$

to prove $\sqrt{T}-$consistency of the estimated coefficients involved in the fitted truncated autoregression.

The conditions (A3) and (A4) proposed by Berk (1974) and Lewis and Reinsel (1985) have direct application to the case where $\omega_t=\Delta y_t$ and $d_i=\zeta_i$ if the process $\{y_t\}$ defined in (4.2.1) and analyzed by Said and Dickey (1984) is assumed to be stationary.

Ng and Perron (1995) considered the validity of all Said and Dickey's results in the absence of the lower bound condition (A2). The importance of such a result will be mentioned later in this Chapter and is going to be analyzed extensively in Chapter 5. Their following lemma is involved with the above statement:

Lemma 4

Suppose that the DGP for $\{y_t\}$ is given by (4.2.2) where $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$ with bounded fourth moment and $\{u_t\}$ is a stationary and invertible process with autoregressive and moving average polynomials that do not share common roots. Let also the limiting distribution of the studentized $\tau_{DF}-$test statistic be obtained from the truncated autoregression of the form (4.2.3) with k chosen such as to satisfy only (A1).

Then, (a) the asymptotic distribution of the $\tau_{DF}-$test statistic continues to be given by (3.3.10) without (A2) and (b) the estimated coefficient vector $\hat{\zeta}(k)=\left(\hat{\zeta}_{1,T},.......\hat{\zeta}_{k,T}\right)$ is not in general $\sqrt{T}-$consistent for the true-valued coefficient vector ζ if (A2) or (A4) does not hold. In that case, if there exists $\lambda\in(0,1)$ and C > 0 such that $|\zeta_j|\leq C\lambda^j$, then it follows $\lambda^{-k}(\hat{\zeta}_{i,T}-\zeta_i)=O_p(1)$ for i = 1, 2, …. , k.

This lemma of Ng and Perron (1995) states that the consistency of the estimated coefficients on the stationary regressors in the Augmented Dickey-Fuller autoregression of the form (4.2.3) is not violated if (A1) alone holds. It also states that there is no change in the structure of the limiting distribution of the studentized $\tau_{DF}-$test statistic for testing the unit root hypothesis that $\rho=1$ if (A1) alone holds. Although $\sqrt{T}-$consistency of the estimated coefficients on the stationary regressors in the Augmented Dickey-Fuller autoregression is not assured without the condition (A2), $\hat{\zeta}_{0,T}$ is still T-consistent (remember that we proved the T – consistency of $\hat{\zeta}_{0,T}$ without the lower-bound condition (A2) in step 4). The importance of the last lemma comes from the fact that allows k to be chosen to have a logarithmic rate, which we will see that is of special interest, in contrast to (A2) or (A4) which rule it out. On the other hand, all these statements do not indicate that (A2) condition is not important. Therefore, choices of k that satisfy conditions (A2) and (A4) will yield coefficient estimates on the stationary regressors in the Augmented Dickey-Fuller autoregression that achieve $\sqrt{T}-$ consistency and for this reason can be expected to lead to unit root tests that have better finite-sample properties than choices of k that do not satisfy (A2).

Let us now see why the $\sqrt{T}-$consistency of the estimated coefficients on the stationary regressors in the Augmented Dickey-Fuller autoregression is not assured without the condition (A2). To do this, we have to re-establish the order of the term $\| D_T\sum U_t(\varepsilon_{tk}-\varepsilon_t)\|=\| D_T\sum X_t(\varepsilon_{tk}-\varepsilon_t)\|$.

It holds that:

$$E\| D_T \sum_{t=k+1}^{T} X_t(\varepsilon_{tk}-\varepsilon_t)\| = E\left\{(T-k)^{-1}\sum_{j=1}^{k}\left[\sum_{t=k+1}^{T} u_{t-j}(\varepsilon_{tk}-\varepsilon_t\right]^2\right\}$$

But,

$$E\left\{(T-k)^{-1}\sum_{j=1}^{k}\left[\sum_{t=k+1}^{T-k} u_{t-j}(\varepsilon_{tk}-\varepsilon_t)\right]^2\right\} =$$

$$E\left\{\sum_{j=1}^{k}\left[(T-k)^{-1/2}\sum_{t=k+1}^{T-k} u_{t-j}(\varepsilon_{tk}-\varepsilon_t)\right]^2\right\} =$$

$$\sum_{j=1}^{k} E\left[(T-k)^{-1/2}\sum_{t=k+1}^{T-k} u_{t-j}(\varepsilon_{tk}-\varepsilon_t\right]^2 \le$$

$$ck(T-k)\sum_{i=k+1}^{\infty}\zeta_i^2 \le C_1 k\lambda^{2k}/(1-\lambda^2)\longrightarrow 0$$

provided that $\lambda \in (0,1)$ satisfies the condition $|\zeta_j| \le C\lambda^j$ for some constant C_1 different from c and $k \longrightarrow \infty$, which is assured under the condition (A1). So, using the above derivations we get:

$$\| D_T \sum_{t=k+1}^{T} X_t(\varepsilon_{tk}-\varepsilon_t)\| = o_p(1)$$

under the absence of the lower-bound condition (A2).

The rate at which the estimated coefficient vector $\hat{\zeta}(k) = \left(\hat{\zeta}_{1,T},\hat{\zeta}_{2,T},\ldots\ldots,\hat{\zeta}_{k,T}\right)$ converges under the absence of the lower bound condition (A2) can be found by writing:

$$(T-k)^{1/2}(\hat{\zeta}(k)-\zeta(k))=(D_T(k)M_kD_T(k))^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_{tk}=$$
$$R^{*-1}(T-k)^{-1/2}\sum X_t\varepsilon_{tk}=(R^{*-1}-\Gamma^{-1}+\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk}=$$
$$(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk}+\Gamma^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_{tk}=$$
$$(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk}+\Gamma^{-1}(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t+\varepsilon_t)=$$
$$(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk}+\Gamma^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_t+$$
$$+\Gamma^{-1}(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t)$$

Taking the usual Euclidean norm of both sides above we obtain:

$$\|(T-k)^{1/2}(\hat{\zeta}(k)-\zeta(k))\|=$$
$$\|(R^{*-1}-\Gamma^{-1})(T-k)^{-1/2}\sum X_t\varepsilon_{tk}+\Gamma^{-1}(T-k)^{-1/2}\sum X_t\varepsilon_t+$$
$$\Gamma^{-1}(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t)\|\leq$$
$$\|(R^{*-1}-\Gamma^{-1})\|\|(T-k)^{-1/2}\sum X_t\varepsilon_{tk}\|+$$
$$\|\Gamma^{-1}\|\|(T-k)^{-1/2}\sum X_t\varepsilon_t\|+\|\Gamma^{-1}\|\|(T-k)^{-1/2}\sum X_t(\varepsilon_{tk}-\varepsilon_t)\|$$
$$=O_p(1)+O_p(k^{1/2})+o_p(1)=O_p(k^{1/2})\Rightarrow$$
$$\|(T-k)^{1/2}(\hat{\zeta}(k)-\zeta(k))\|=O_p(k^{1/2})$$

If condition (A2) holds, we saw that: $\sqrt{T}(\hat{\zeta}_{i,k}-\zeta_i)=O_p(1)$ for i=1,2,3,….,T.

If condition (A2) does not hold, then:

$$\|\lambda^{-k}(\hat{\zeta}(k)-\zeta(k))\|=O_p(k^{1/2}) \quad or \quad \lambda^{-k}(\hat{\zeta}(k)-\zeta(k))=O_p(1)$$

for i=1,2,3,….,T since $k^{1/2}\lambda^k \longrightarrow 0$ as $k\longrightarrow\infty$.

Having concerned with the appropriate conditions that the degree of augmentation, or the lag order k, must satisfied in order to get statistical meaningful results, a direct application of them provide theoretical rather than practical guidance for choosing it. So, selection methods are needed for choosing the appropriate lag number k to be included in the truncated autoregression of the form (4.2.3) provided that these selection methods, whatever will be, must satisfy the conditions just mentioned.

Appendix of Chapter 4: Mathematical proofs

Lemma 2 (p. 70-71)

As $T \rightarrow \infty$:

a) $\| D_T \Sigma U_t \left(\varepsilon_{tk} - \varepsilon_t\right) \| = O_p(T^{-1})$

b) $\| D_T \Sigma U_t \varepsilon_t \| = O_p(k^{\frac{1}{2}})$

c) $\| D_T \Sigma U_t \varepsilon_{tk} \| = O_p(k^{\frac{1}{2}})$

Proof

a) It holds that:

$$E \| D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk} - \varepsilon_t) \| =$$

$$(T-k)^{-2} E\left[\sum_{t=k+1}^{T} y_{t-1}(\varepsilon_{tk} - \varepsilon_t)\right]^2 + E\left\{(T-k)^{-1} \sum_{j=1}^{k} \left[\sum_{t=k+1}^{T} u_{t-j}(\varepsilon_{tk} - \varepsilon_t\right]^2\right\}$$

For the first term we have the following: Since $\{u_t\}$ is a stationary invertible ARMA process, there exists a sequence of real numbers $\{\zeta_j\}$, a number $0 < \lambda < 1$ and a number M such that $|\zeta_j| < M\lambda^j$ (Fuller, 1976, Theorem 2.7.2). Taking the above considerations into account we have that:

$$(T-k)^{-2}E\left[\sum_{t=k+1}^{T} y_{t-1}(\varepsilon_{tk}-\varepsilon_t)\right]^2 = (T-k)^{-2}E\left[\sum_{t=k+1}^{T} y_{t-1}\sum_{i=k+1}^{\infty}\zeta_i u_{t-i}\right]^2 =$$

$$\sum_{i=k+1}^{\infty}\sum_{j=k+1}^{\infty}\zeta_i\zeta_j\left[(T-k)^{-2}\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}E(y_{t-1}u_{t-i}y_{s-1}u_{s-j})\right] \le$$

$$C_1\sum_{i=k+1}^{\infty}\sum_{j=k+1}^{\infty}\zeta_i\zeta_j \le C_1\sum_{i=k+1}^{\infty}\zeta_i\sum_{j=k+1}^{\infty}\zeta_j \le$$

$$C_1\sum_{i=k+1}^{\infty}\lambda^i\sum_{j=k+1}^{\infty}\lambda^j = C_1\left(\sum_{i=k+1}^{\infty}\lambda^i\right)^2 = C_1\frac{\lambda^{2k}}{(1-\lambda)^2} \to 0 \ \textbf{for}\ \ \mathbf{k}\to\infty \ \ \textbf{as}\ \ T\to\infty$$

where C_1 is a constant number.

For the second term, it holds that:

$$E\left\{(T-k)^{-1}\sum_{j=1}^{k}\left[\sum_{t=k+1}^{T-k}u_{t-j}(\varepsilon_{tk}-\varepsilon_t)\right]^2\right\} =$$

$$E\left\{\sum_{j=1}^{k}\left[(T-k)^{-1/2}\sum_{t=k+1}^{T-k}u_{t-j}(\varepsilon_{tk}-\varepsilon_t)\right]^2\right\} =$$

$$\sum_{j=1}^{k}E\left[(T-k)^{-1/2}\sum_{t=k+1}^{T-k}u_{t-j}(\varepsilon_{tk}-\varepsilon_t\right]^2 \le$$

$$ck(T-k)\sum_{i=k+1}^{\infty}\zeta_i^2 = O_p(T^{-2})$$

under the lower bound condition $ck>T^{1/r}$ for some $c>0$ and $r>0$ and the direct application of assumption (iv) of Berk's (1974) Theorem 2, p.495.

Finally, using the above derivations we get:

$$\| D_T \sum_{t=k+1}^{T} U_t(\varepsilon_{tk} - \varepsilon_t) \| = O_p(T^{-1})$$

b) It holds that:

$$E \| D_T \sum_{t=k+1}^{T} U_t \varepsilon_t \|^2 =$$

$$(T-k)^{-2} E\left(\sum_{t=k+1}^{T} y_{t-1}\varepsilon_t \right)^2 + (T-k)^{-1} \sum_{j=1}^{k} E\left(\sum_{t=1}^{T-k} u_{t-j}\varepsilon_t \right)^2$$

But,

$$(T-k)^{-1} \sum_{j=1}^{k} E\left(\sum_{t=1}^{T-k} u_{t-j}\varepsilon_t \right)^2 = k\gamma_u(0)\sigma^2 = O_p(k)$$

using the independence of u_{t-j} and ε_t for j > 0 and the direct application of Berk's (1974) Lemma 1, p.492.

On the other hand, using the same facts as above, it holds that:

$$E\left(\sum_{t=k+1}^{T} y_{t-1}\varepsilon_t \right)^2 = \sigma^2 \sum_{t=k+1}^{T} E\left(y_{t-1}^2 \right) = O_p(T^2)$$

(see Hamilton, 1994, p.497). So,

$$E \| D_T \sum_{t=k+1}^{T} U_t \varepsilon_t \|^2 = O_p(k) + O_p(T^2) = O_p(k) \Rightarrow$$

$$\| D_T \sum_{t=k+1}^{T} U_t \varepsilon_t \| = O_p(k^{1/2})$$

c) Combining parts (a) and (b) we have that:

$$\| D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} \| = \| D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk} - \varepsilon_t + \varepsilon_t) \| =$$

$$\| D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk} - \varepsilon_t) + D_T \sum_{t=k+1}^{T} U_t \varepsilon_t \| \leq$$

$$\| D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk} - \varepsilon_t) \| + \| D_T \sum_{t=k+1}^{T} U_t \varepsilon_t \| =$$

$$O_p(T^{-1}) + O_p(k^{1/2}) = O_p(k^{1/2})$$

Proof of Theorem 2 (p.71)

Under the assumptions of the process (4.3.1) with $\rho = 1$, it holds that $\| \hat{\zeta} - \zeta \| \xrightarrow{p} 0$ under the conditions (A1) and (A2).

Proof

Let us denote $K = D_T R_0 D_T$. We then write:

$$D_T^{-1}(\hat{\zeta} - \zeta) = (D_T R_0 D_T)^{-1} D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} = K^{-1} D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} =$$

$$(K^{-1} - R^{-1} + R^{-1}) D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} =$$

$$(K^{-1} - R^{-1}) D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} + R^{-1} D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} =$$

$$(K^{-1} - R^{-1}) D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} + R^{-1} D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk} - \varepsilon_t + \varepsilon_t) =$$

$$(K^{-1} - R^{-1}) D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} + R^{-1} D_T \sum_{t=k+1}^{T} U_t \varepsilon_t + R^{-1} D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk} - \varepsilon_t)$$

Taking norms we get:

$$\| D_T^{-1}(\hat{\zeta}-\zeta)\| \le$$

$$\| K^{-1}-R^{-1}\| \| D_T \sum_{t=k+1}^{T} U_t \varepsilon_{tk} \| + \| R^{-1}\| \| D_T \sum_{t=k+1}^{T} U_t \varepsilon_t \| +$$

$$\| R^{-1}\| \| D_T \sum_{t=k+1}^{T} U_t (\varepsilon_{tk}-\varepsilon_t) \| =$$

$$o_p(1)+O_p(1)O_p(k^{1/2})+O_p(1)O_p(T^{-1}) =$$

$$o_p(1)+O_p(k^{1/2})+O_p(T^{-1}) = O_p(k^{1/2})$$

Finally,

$\| D_T^{-1}(\hat{\zeta}-\zeta)\| = O_p(k^{1/2}) \Rightarrow \| \hat{\zeta}-\zeta \| = o_p(1)$ since $\| D_T^{-1}\| = O_p(T\text{-}k)$ under the condition of $\frac{k^3}{T} \to 0$ as $T \to \infty$.

CHAPTER 5

LAG LENGTH SELECTION METHODS FOR THE ADF TESTS

5.1 INTRODUCTION

Model selection is a crucial issue in the science of statistics. Model selection refers to the procedure that selects hypothesized models which describe in a best way an experiment under study. The usual statistical theory involves the selection of a class of models using various procedures including applications of significance tests or maximization of specified criteria that work in a satisfactory way under certain conditions. Model selection in our case refers to the procedure that selects the appropriate lag number to be included in the Augmented Dickey-Fuller autoregression as well as in the Augmented Dickey-Fuller τ_{DF} – test statistic in order to obtain desirable statistical properties for the resulting testing procedure. In this Chapter we will present the theory concerning the choice of the lag number to be included in the Augmented Dickey-Fuller autoregression and its implications to the Augmented Dickey-Fuller τ_{DF} – test statistic when it tests unit root hypotheses using selection criteria of the type just mentioned. We will notice in the foregoing sections that the theory is developed for giving satisfactory results asymptotically, so Monte Carlo simulation experiments are needed for examining the results of the various selection criteria in finite samples. The latter will be the main issue of the next Chapter.

The data generating process that is considered and used in this Chapter has the same form considered by Said and Dickey (1984):

$$y_t = \rho y_{t-1} + u_t$$

and

$$u_t \sim \text{ARMA(p, q)}$$

with the autoregressive and moving average parameters involved in the stationary ARMA(p, q) error process denoted by $\{\phi_i, \theta_j, i = 1,2,..,p, j = 1,2,...,q\}$, respectively, and the unit root hypothesis that $\rho = 1$ is going to be tested through the studentized unit root τ_{DF} – test statistic using the Augmented Dickey-Fuller autoregression of the form:

$$\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk} \tag{5.1.1}$$

for various values of the parameters p and q.

5.2 THE SIMPLEST METHOD

The simplest way to choose the lag number k to be included in the Augmented Dickey-Fuller autoregression of the form (5.1.1) before applying the Augmented Dickey-Fuller unit root τ_{DF} – test is that of being independent of the number of observations T meaning, for example, that we can allow it to take any value we like. Ng and Perron (1995) fixed k to belong in the range [1, 10] and constructed Monte Carlo simulation experiments to investigate the performance of the unit root τ_{DF} – test statistic in the context of its size and the power. These are going to be illustrated extensively in the next Chapter.

5.3 THE DETERMINISTIC METHOD

Schwert (1989) suggested choosing the lag number k to be included in the Augmented Dickey-Fuller autoregression before applying the Augmented Dickey-Fuller unit root τ_{DF} – test as a deterministic function of the number of observations T. In particular, Schwert (1989) defined:

$$k = \text{int}\left\{ c\left(\frac{T}{100}\right)^{\frac{1}{d}} \right\}$$

where 'int' denotes the integer before the first decimal point, for arbitrary given constants c and d. More specifically, values of c=4 and 12 and d=4 were used for k to be calculated for different sample size choices. Schwert (1989) also constructed extensive Monte Carlo simulation experiments to investigate the performance of the unit root τ_{DF} – test statistic in the context of its size and the power. These are also going to be illustrated extensively in the next Chapter.

The practical problem that appears from using this specified deterministic method for selecting the lag number k is that one is faced with a given sample size in practice and this fact leads to the calculation of the right c and d in the deterministic function of k unless c and d happen to be chosen correctly.

Furthermore, deterministic methods like this proposed by Schwert (1989) as well as the simplest one for choosing the lag number k to be included in the Augmented Dickey-Fuller autoregression are not good practical guidelines for asymptotic inference. Indeed, the value of k that ensures an exact size close to the nominal size and also produces high power is strongly dependent on the actual data generating process and its error process, that is, the values of the moving average and autoregressive parameters, information that these two criteria do not take into account and for this reason have to be avoided as a matter of practice. However, as we will notice in the next Chapter, studying the results of these two selection methods mentioned above, a significant problem on the statistical properties of the unit root τ_{DF} – test statistic starts to appear.

5.4 SEQUENTIAL METHODS

The "sequential tests", as Ng and Perron (1995) claim, belong to a "general-to-specific" modeling strategy and used for choosing, for example, between a model with, say, (k-h)-regressors and a model with k-regressors. In the case where the model is the Augmented Dickey-Fuller autoregression of the form (5.1.1), the discrimination of the Augmented Dickey-Fuller autoregression of order, say, k-h and the Augmented Dickey-Fuller autoregression of order k can be done as follows:

Let $\hat{\zeta}(k-h,k)$ denote the estimated vector of the coefficients $(\hat{\zeta}_{k-h+1,T},......,\hat{\zeta}_{k,T})$ obtained by applying the least squares procedure to the Augmented Dickey-Fuller autoregression of the form (5.1.1), with:

$$\hat{\sigma}_k^2 = \sum_{t=k+1}^{T} \hat{\varepsilon}_{tk}^2$$

and let:

$$R_k = \sum_{t=k+1}^{T} (y_{t-1}, \Delta y_{t-1}, ...\Delta y_{t-k})^{/} (y_{t-1}, \Delta y_{t-1}, ...\Delta y_{t-k})$$

Let also $R_{0,k}^{-1}(h)$ be the lower-right $(h \times h)$ block of the matrix R_k^{-1}.

Giving the above definitions, the discrimination of the Augmented Dickey-Fuller autoregression with (k-h) and k-lagged terms can be done by considering the hypothesis that the coefficients on the last h-lagged terms are jointly equal to zero and can be tested through the Wald test-statistic that defined as follows:

$$J(k-h,k) = \hat{\zeta}(k-h,k)^{/}\,(R_{0,k}^{-1}(h))^{-1}\,\hat{\zeta}(k-h,k)\,/\,\hat{\sigma}_k^2$$

(5.4.1)

Ng and Perron (1995) gave the beneath Lemma that is concerned with the limiting distribution of the Wald, J(k-h, k), test statistic defined above and is useful in establishing later in this section the limiting distribution of the studentized τ_{DF} – test statistic when testing the null hypothesis of the unit root.

Lemma 4

Let $\{y_t\}$ be generated by (4.2.1) and suppose that the conditions (A1) and (A4) hold. Let $\hat{\zeta}(k) = (\hat{\zeta}_{1,T},\ldots,\hat{\zeta}_{k,T})$ be the estimated coefficient vector on the stationary variables included in the Augmented Dickey-Fuller autoregression of the form (5.1.1) and let the Wald, J(k-h, k), test-statistic be as defined in (5.4.1). Then, the Wald, J(k-h, k) test-statistic is asymptotically chi-squared distributed with h-degrees of freedom.

The Wald test statistic that has been described above leads to the standard t-test statistic if we replace h with unity, h=1.

Simply speaking, in the case of the Augmented Dickey-Fuller autoregression of order, say, k, the standard t-test deals with the significance of the coefficient on the last k-th lagged term and takes the form of:

$$t_{\hat{\zeta}_k} = \sqrt{T}\,\hat{\zeta}_k \left[\hat{\sigma}_k^2 T R_0^{-1}(1)\right]^{-\frac{1}{2}} \qquad (5.4.2)$$

where $R_0^{-1}(1)$ is the first diagonal element of the matrix R_0^{-1} defined in (4.3.7), and is actually the square root of the Wald, J(k-1, k), test statistic.

The standard t-test chooses a value of $\hat{k}$ if $t_{\hat{\zeta}_k}$ is significant at some pre-specified level α in an estimated Augmented Dickey-Fuller autoregression of the form (5.1.1) of order $\hat{k}$, whereas the t-tests $t_{\hat{\zeta}_k}$ are insignificant in the estimated autoregressions of order k for all k in the range $(\hat{k}, k_{\max}]$.

Having illustrated the Wald as well as the standard t – test statistics of the form (5.4.1) and (5.4.2) respectively, let us make some comments on them. The Wald, J(k-h, k), test statistic requires $\hat{k}$ to be chosen from a set of possible values {0, 1, 2, …. , k_{max}} where k_{max} has to be selected a priori in a way to satisfy the condition (A1). Ng and Perron (2001) set, for example, $k_{\max} = \text{int}\left[10\left(\frac{T}{100}\right)^{1/4}\right]$ but mentioned that other values are also valid. Ng and Perron (1995) gave the following definition that is concerned with this specific Wald test statistic.

Definition 1

The general-to-specific modeling strategy chooses $\hat{k}$ to be either a) h+1 if, at a prespecified level α, J(k-h, k) is the first test in the sequence J(i-h, i) where {i = k_{max}-1, ….. , 1} which is significantly different from zero b) zero if J(i-h, i) is not significantly different from zero for all {i = k_{max}-1, ….. , 1}.

Letting h being equal to unity in the above definition we get a special case of the general-to-specific modeling strategy. This is the standard t-test concerning the significance of the coefficient on the last lagged-term. Furthermore, the Wald, J(k-h, k), test statistic is chi-square distributed with h-degrees of freedom under the conditions (A2) and (A4) meaning that k has to increase at some polynomial rate or at a rate that (A2) or (A4) is satisfied. In fact, Ng and Perron (1995) stated that the truncated lag number selected by general-to-specific modeling strategy will be grown at the same rate as k_{max} as the next Lemma states:

Lemma 6

If the order $\hat{k}$ of the Augmented Dickey-Fuller autoregression is selected by means of general-to-specific modeling strategy and its maximum order k_{max} increases at a rate such that (A1) and (A4) are satisfied, then its order $\hat{k}$ increases at the same rate as k_{max}.

According to this Lemma, if k_{max} is selected to increase at a polynomial rate then $\hat{k}$ selected due to the general-to-specific strategy will also increase at a polynomial rate. This implies that conditions (A1) or (A4) can be satisfied with judicious choice of k_{max}, thereby ensuring that the results of the Lemma 4 hold.

Now we are in the position to determine the asymptotic distribution of the unit root τ_{DF} –test statistic when k chosen due to the general-to-specific modeling strategy.

Theorem 4

If the maximum order of the Augmanted Dickey-Fuller autoregression k_{max} satisfies conditions (A1) and (A4) and its order $\hat{k}$ is chosen according to the general-to-specific modeling strategy, then the limiting distribution of the studentized τ_{DF} –test statistic when testing the hypothesis of the unit root is of the form of (3.3.10).

5.5 INFORMATION CRITERIA (IC) METHODS

Time series analysis involves both model identification and parameters estimation. The identification problem is the more difficult part. Once the functional form of the model is specified, estimating the parameters of the model is usually straightforward.

To identify the model that best represents the data of a time series, it is necessary to be clear about the purpose of the model. Here, we are concerned with the ″predictive accuracy″ of a model. A model is chosen by studying the data values in a fit set. It is then used to predict future values of the time series. The model can be forced to fit the data increasingly well by increasing its order. However, as it is well known, the fact that the fit set errors are small is no guarantee that the prediction set errors will be.

Many of the terms in a complex model may simply be accounting for noise in the data. Such overfitted models may predict future values of the time series quite poorly. Even when the true order of the data generating process is known, a lower order may give better predictions. Thus, to arrive at a model that represents only the main features of the series, selection criteria that balance model fit and model complexity must be used.

In the context of a stationary process that is approximated by fitting a truncated autoregression of order, say, k Hannan and Deistler (1988) considered the global minimization of the beneath objective function of the general form:

$$IC(k) = \ln(\tilde{\sigma}_k^2) + k\frac{C_T}{T} \qquad (5.5.1)$$

for selecting its appropriate order in order to minimize its prediction error, where C_T is a sequence that satisfies $C_T > 0$ and $\frac{C_T}{T} \to 0$ as $T \to \infty$ and $\tilde{\sigma}_k^2$ is the estimated variance computed from the fitted truncated autoregression with, say, k regressors and a sample size of T observations using least squares procedure.

Two of the most known criteria are that of Akaike, named Akaike Information Criterion and denoted as AIC and that of Schwarz, named Schwarz or Bayesian Information Criterion and denoted as BIC. The use of AIC has originally been motivated by the wish to maximize the value of a quantity which called ″predictive accuracy″ of a model and the use of BIC is motivated by the Bayesian idea that one should choose a model with the largest ″posterior probability″. The AIC is obtained as a special case of

(5.5.1) with $C_T = 2$ and the BIC as a special case of (5.5.1) with $C_T = \ln T$. Of course, there are many other criteria that actually differ only in C_T, the weight applied to overfitting, but all use k as the penalty to overfitting.

Ng and Perron (1995) summarized a result of Hannan and Deistler (1988) that is relevant to the context of the unit root testing issue, as we will notice in the subsequent analysis in this section, and is concerned with a basic property that the order of the fitted truncated autoregression satisfies when the latter is fitted to approximate a stationary process. Specifically, it is proportional to a specific function of the number of observations T as the next Lemma claims:

Lemma 7

Let $\{\omega_t\}$ belong to the general class of the stationary and invertible ARMA processes with finite fourth moment and let us assume that we fit a truncated autoregression of order, say, k:

$$\omega_t = \sum_{i=1}^{k} d_i \omega_{t-i} + \varepsilon_{tk} \tag{5.5.2}$$

and define:

$$\tilde{\sigma}_k^2 = \frac{1}{T-k} \sum_{t=k+1}^{T} \tilde{\varepsilon}_{tk}^2$$

where $\tilde{\varepsilon}_{tk}$ are the residuals obtained from the fitted truncated autoregression of the form (5.5.2) by applying the least squares procedure. Let C_T be a function of T such that $C_T > 0$ and $\frac{C_T}{T} \to 0$ as $T \to \infty$ and

$$\tilde{k}_T = \arg \min_{k \le k_{max}} \left[\ln(\tilde{\sigma}_k^2) + k \frac{C_T}{T} \right]$$

provided that k_{max} satisfies the condition (A1). Then, $\lim_{T \to \infty} \frac{\tilde{k}_T}{b \ln T} = 1$, for some constant b.

The result that the global minimization of the objective function of the form (5.5.1) chooses a value of k that is proportional to lnT in a univariate stationary and invertible ARMA process is due to Shibata (1980) and Hannan and Deistler (1988) provided a unified asymptotic framework to show that the feature of lnT proportionality is generic to Information Criteria based rules applied, in particular, to the general class of the stationary and invertible ARMA processes. The result of the above Lemma 7 is useful in studying the properties of the truncated lag number k within the context of an Augmented Dickey-Fuller autoregression of the form (5.1.1) derived for an autoregressive integrated moving average ARIMA (p, 1, q) process. Indeed, the following Lemma shows that the result of Lemma 7 extends to this latter case:

Lemma 8

Let $\{y_t\}$ satisfy the following process:

$$y_t = \rho y_{t-1} + u_t \qquad \text{for } t = 1,2,3,\ldots\ldots$$

and

$$u_t = \sum_{i=1}^{p} \phi_i u_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad \text{for } t = \ldots,-2,-1,0,1,2,\ldots$$

and define:

$$\hat{\sigma}_k{}^2 = (T-k)^{-1} \sum_{t=k+1}^{T} \hat{\varepsilon}_{tk}^2$$

where $\hat{\varepsilon}_{tk}$ are the least squares residuals obtained from the Augmented Dickey-Fuller autoregression of the form (5.1.1). Define also:

$$\tilde{\sigma}_k{}^2 = (T-k)^{-1} \sum_{t=k+1}^{T} \tilde{\varepsilon}_{tk}^2$$

where $\tilde{\varepsilon}_{tk}$ are the least squares residuals obtained from the fitted truncated autoregression of the form (5.5.2) with $\omega_t = \Delta y_t$. Then:

$$\tilde{\sigma}_k^2 = \hat{\sigma}_k^2 + o_p(T^{-1/2})$$

uniformly in k, provided that k satisfies the condition (A1).

Lemma 8 implies that the difference between the residual sum of squares from the Augmented Dickey-Fuller autoregression of the form (5.1.1) and the restricted one of the form (5.5.2) with $\omega_t = \Delta y_t$ is $o_p(T^{-1/2})$ uniformly in k. Therefore, the Information Criteria and the corresponding values of k that minimize such criteria are asymptotically the same in both cases. Thus, the AIC and BIC, when applied to the Augmented Dickey-Fuller autoregression denined in (5.1.1), also select truncation lags proportional to lnT under the null hypothesis of a unit root that $\rho = 1$. Notice also that the lnT proportionality satisfies the condition (A1) that $\frac{k^3}{T} \to 0$ and $k \to \infty$ as $T \to \infty$ in contrast to the conditions (A2) or (A4) which rule it out since they require for k to grow at least at a polynomial rate. As far as the limiting distribution of the unit root τ_{DF} –test statistic is concerned, is summarized in the following Theorem of Ng and Perron (1995):

Theorem 5

Let $\{y_t\}$ satisfy the assumptions of the general process of the form (4.2.1). Under the null hypothesis of a unit root, if the order of the Augmented

Dickey-Fuller autoregression selected using an Information Criterion in the class of IC(k) as defined in (5.5.1) such that to satisfy the condition (A1), then the unit root τ_{DF} –test statistic has the limiting distribution defined by (3.3.10).

5.6 CORRECTED INFORMATION CRITERIA (ICC) METHODS

Let $\{\omega_t\}$ belong to the general class of the stationary and invertible ARMA processes with finite fourth moment and let also be approximated by fitting a truncated autoregression of order, say, k:

$$\omega_t = \sum_{i=1}^{k} d_i \omega_{t-i} + \varepsilon_{tk} \tag{5.6.1}$$

Ng and Perron (2000) restudied issues related to the order of such an above autoregression selected using the global minimization of the objective function of the form (5.5.1). In particular, they studied the sensitivity of the estimated order to (a) whether the effective number of observations is held fixed when estimating models of different order, (b) whether the estimate of the variance is adjusted for the degrees of freedom, and (c) how the penalty factor for overfitting is defined in relation to the total sample size. Their Monte Carlo simulation experiments showed that the lag length order selected using the global minimization of the objective function of the form (5.5.1) is sensitive to these parameters in finite samples. Furthermore, theoretical considerations revealed that the global minimization of the beneath Corrected objective function:

$$ICC(k) = \ln(\tilde{\sigma}_k^2) + k\frac{C_T}{T - k_{\max}} \tag{5.6.2}$$

where $\tilde{\sigma}_k^2 = \frac{1}{T-k_{\max}} \sum_{t=k+1}^{T} \tilde{\varepsilon}_{tk}^2$ denotes the estimated variance computed from the fitted truncated autoregression of the form (5.6.1) and $C_T > 0$ and $\frac{C_T}{T} \to 0$ as $T \to \infty$, provided that k_{max} satisfies the condition (A1), that hold the effective sample size fixed across models to be compared is more appropriate than the objective function of the form (5.5.1) in the sense that when it is used for selecting the order of a truncated autoregression when the latter is fitted to approximate a stationary process gives more precise estimates for the coefficients involved in the same fitted truncated autoregression[2].

Related issues concerning firstly the choice of the order of the Augmented Dickey-Fuller autoregression derived for an autoregressive integrated moving average, ARIMA(p,1,q), process through the Corrected Information Criteria and secondly the unit root hypothesis testing issue through the unit root τ_{DF} - test statistic, follow the theory presented in section 5.5.

Simulation experiments concerning the limiting behavior of the unit root τ_{DF} - test statistic under this truncation lag selection will be presented and analyzed fully in the next Chapter.

[2] Interested readers should refer to Ng and Perron (2000) for an extensive study on this particular formulation of the Corrected Information Criteria stated.

5.7 MODIFIED INFORMATION CRITERIA (MICC) METHODS

The AIC and BIC or AICC and BICC considered in the previous sections belong to the class of Information based rules where the chosen value of k to be included in the Augmented Dickey-Fuller autoregression of the form (5.1.1) is either:

$$k_{IC} = \arg \min_{k \in [0,\ldots\ldots,k_{\max}]} IC(k) \quad \text{or} \quad k_{ICC} = \arg \min_{k \in [0,\ldots\ldots,k_{\max}]} ICC(k)$$

where

$$IC(k) = \ln(\hat{\sigma}_k^2) + k\frac{C_T}{T} \qquad \text{or} \qquad ICC(k) = \ln(\hat{\sigma}_k^2) + k\frac{C_T}{T - k_{\max}}$$

where $\hat{\sigma}_k^2 = (T-k)^{-1} \sum_{t=k+1}^{T} \hat{\varepsilon}_{tk}^2$ or $\hat{\sigma}_k^2 = (T-k_{\max})^{-1} \sum_{t=k+1}^{T} \hat{\varepsilon}_{tk}^2$

and $\frac{C_T}{T} \to 0$ as $T \to \infty$, and $C_T>0$.

The various criteria differ in C_T, the weight applied to overfitting, but all use k as the penalty to overfitting.

Ng and Perron (2001) argued that, with integrated data, this penalty may be a poor approximation to the cost of underfitting. So, on the basis of the Corrected Information Criteria, they suggested a class of Modified Information Criteria, denoted hereafter by MICC, that takes better account of the cost of overfitting. Following their lines, when the Modified Information Criteria are used as selection method for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression of the form (5.1.1) lead to substantial size and power improvements for the unit root τ_{DF} – test statistic.

We illustrate these considerations starting by the first argument that, with integrated data, the penalty k may be a poor approximation to the cost of underfitting and for this reason an alternative penalty for integrated data must be considered. This issue will be illustrated by the derivation of AICC when dealing with integrated data.

Derivation of AIC with an alternative penalty for integrated data

Suppose that the data generated by a finite order AR(k_0) with normal errors and a unit root meaning for example that the data generating process is of the form (5.1.1) with $k = k_0$, $\zeta_0 = 0$ and $\varepsilon_{tk} \sim$ i.i.d.N(0,1).

For notation we let $\zeta^0(k) = (0, \zeta_1, \ldots\ldots, \zeta_k)'$, $\hat{\zeta}(k) = (\hat{\zeta}_{0,T}, \hat{\zeta}_{1,T}, \ldots\ldots \hat{\zeta}_{k,T})'$, $\zeta^0_{-0}(k) = (\zeta_1, \ldots\ldots. \zeta_k)'$ and $\hat{\zeta}_{-0}(k) = (\hat{\zeta}_{1,T}, \ldots\ldots. \hat{\zeta}_{k,T})'$.

The goal is to select the lag length k between 0 and some upper bound k_{max} that satisfies the condition (A1). In these considerations for the lag length k condition (A1) for the upper bound k_{max} is rather strong in the sense that $\frac{k_{\max}}{T} \to 0$ and $k_{\max} \to \infty$ as $T \to \infty$ also works well.

Let $f\left(\Delta y / \zeta^0(k)\right)$ be the likelihood function of the data $\left(\Delta y_{k_{\max}+1}, \ldots\ldots, \Delta y_T\right)$ conditional on the initial observations $\left(y_0, \ldots\ldots, y_{k_{\max}}\right)$ to ensure that each competing model is evaluated with the same number of effective observations, namely $T - k_{max}$.

The Kullback – Leibler distance between the true probability distribution and the estimated one is:

$$Q = E[\ln(f(\Delta y / d^0(k))) - \ln(f(\Delta y / \hat{d}(k)))]$$

with sample analogue:

$$\breve{Q} = \left(T - k_{\max}\right)^{-1} \sum_{t = k_{\max}+1}^{T} \ln[f(\Delta y_t / \zeta^0(k))] - \left(T - k_{\max}\right)^{-1} \sum_{t = k_{\max}+1}^{T} \ln[f(\Delta y_t / \hat{\zeta}(k))]$$

as discussed in Gourieroux-Monfort (1995). Akaike's suggestion was to find a Q^* such that $\lim_{T \to \infty} E[T(Q - Q^*)] = 0$ so that Q^* is unbiased for Q to order T^{-1}.

Let $U_t = \left(y_{t-1}, X_t\right)$ with $X_t = \left(\Delta y_{t-1}, \ldots\ldots, \Delta y_{t-k}\right)$ and

$$\Phi_T(k) = \left(\frac{1}{\hat{\sigma}_k^2}\right)\left[\hat{\zeta}(k) - \zeta^0(k)\right]^{/}\left[\sum_{t = k_{\max}+1}^{T} U_t U_t^{/}\right]\left[\hat{\zeta}(k) - \zeta^0(k)\right]$$

where

$$\hat{\sigma}_k^2 = \left(T - k_{\max}\right)^{-1} \sum_{t = k+1}^{T} \hat{\varepsilon}_{tk}^2$$

Using Taylor expansions, we have that:

$$TQ = \frac{\Phi_T(k)}{2} + o_p(1) \quad \text{and} \quad T\breve{Q} = -\frac{\Phi_T(k)}{2} + o_p(1)$$

Since $T(Q - \breve{Q}) = \Phi_T(k) + o_p(1)$, $\lim_{T \to \infty} E[T(Q - Q^*)] = 0$ for $Q^* = \breve{Q} + \Phi_T(k)$ and the remainder term is uniformly integrable.

Let us now study $\Phi_T(k)$ in the context of integrated data.

Given the asymptotic block diagonality of the matrix $D_T[\sum_{t = k_{\max}+1}^{T} U_t U_t^{/}]D_T$ with $D_T = diag\left\{\left(T - k_{\max}\right)^{-1}, \left(T - k_{\max}\right)^{-1/2}, \ldots\ldots\ldots, \left(T - k_{\max}\right)^{-1/2}\right\}$ we have that:

$$\Phi_T(k) = \left(\frac{1}{\hat{\sigma}_k^2}\right)\hat{\zeta}_{0,T}^2 \sum_{t=k_{\max}+1}^{T} y_{t-1}^2 +$$

$$+\left(\frac{1}{\hat{\sigma}_k^2}\right)\left[\hat{\zeta}_{-0}(k)-\zeta_{-0}^0(k)\right]^{/}\left[\sum_{t=k_{\max}+1}^{T} X_t X_t^{/}\right]\left[\hat{\zeta}_{-0}(k)-\zeta_{-0}^0(k)\right]+o_p(1)=$$

$$=\left(\frac{1}{\hat{\sigma}_k^2}\right)\hat{\zeta}_0^2 \sum_{t=k_{\max}+1}^{T} y_{t-1}^2+\chi_k^2+o_p(1)$$

(5.7.1)

where χ_k^2 is a chi-squared distributed random variable with k degrees of freedom and is asymptotically independent of the first term.

Hence, a Q^* that will satisfy $\lim_{T\to\infty} E[T(Q-Q^*)]=0$ is:

$$Q^* = (T-k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \ln\left[f\left(\Delta y_t/\zeta^0(k)\right)\right]-$$

$$-(T-k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \ln\left[f\left(\Delta y_t/\hat{\zeta}(k)\right)\right]+\left(\frac{1}{\hat{\sigma}_k^2}\right)\hat{\zeta}_{0,T}^2 \sum_{t=k_{\max}+1}^{T} y_{t-1}^2+k$$

(5.7.2)

Under normality, the second term is proportional to $-\left[\left(T-k_{\max}\right)/2\right]\ln\left(\hat{\sigma}_k^2\right)$.

Since the first term is common to all models and does not depend on k, minimizing Q^* with respect to k is equivalent to minimizing:

$$MAICC=\ln\left(\hat{\sigma}_k^2\right)+\frac{2\left(\tau_T(k)+k\right)}{T-k_{\max}} \qquad (5.7.3)$$

where

$$\tau_T(k) = \left(\frac{1}{\hat{\sigma}_k^2}\right)\hat{\zeta}_{0,T}^2 \sum_{t=k_{\max}+1}^{T} y_{t-1}^2$$

and

$$\hat{\sigma}_k^2 = \left(T - k_{\max}\right)^{-1} \sum_{t=k_{\max}+1}^{T} \hat{\varepsilon}_{tk}^2$$

The important step is the relation given by (5.7.1) in the sense that the remainder term is $o_p(1)$ uniformly in k provided that conditions (A1) and (A4) hold. As a natural generalization of the MAICC is the Modified Information Criteria (MICC) which select k as:

$$k_{MICC} = \arg\min_{k} MICC(k)$$

where

$$MICC(k) = \ln\left(\hat{\sigma}_k^2\right) + \frac{C_T[\tau_T(k) + k]}{T - k_{\max}} \tag{5.7.4}$$

with $C_T > 0$ and $\frac{C_T}{T} \to 0$ as $T \to \infty$.

The MAICC is taken from (5.7.4) by setting $C_T = 2$ and the MBICC by setting $C_T = \ln(T - k_{max})$.

Let us now make some comments on the above derivation of the Modified Information Criteria (MICC) to see its differences from the Corrected Information Criteria that we presented in the previous section. A first notable element is the imposition of the null hypothesis $\zeta_0 = 0$ in the derivation of the penalty term. Ng and Perron (2001) mentioned that the idea of imposing the null hypothesis in model selection when the purpose is testing hypotheses appears new and may have implications beyond the unit root theory. A second element is that the penalty function is now stochastic.

As $T \to \infty$ it holds that:

$$\left(\frac{1}{\hat{\sigma}_k^2}\right)\hat{\zeta}_{0,T}^2 \sum_{t=k_{\max}+1}^{T} y_{t-1}^2 \xrightarrow{L} \frac{\left(\frac{\sigma^2}{\sigma_\varepsilon^2}\right)\left[[W(1)]^2-1\right]^2}{2\int_0^1 W(r)^2\,dr} \equiv \tau(W)$$

where W(r) denotes the standard Brownian Motion process. Strictly speaking, we could use the mean of τ(W), which is independent of k, instead of $\tau_T(k)$, to construct Q^* and the objective function would then reduce to standard AICC. The central theme for retaining $\tau_T(k)$ is that, unless T and k are very large, it varies substantially with k. Furthermore, the imposition of the null hypothesis $\zeta_0 = 0$ in model selection allows us to avoid using asymptotic expected values to approximate the penalty factor. So, we hope that $\tau_T(k)$ will better capture the relevant cost of selecting different orders in finite samples.

5.8 REMARKS ON THE LAG LENGTH SELECTION METHODS

The results of this Chapter can be summarized as follows. An Information Criterion will choose values of the lag number to be included in the Augmented Dickey-Fuller autoregression that are proportional to lnT, a rate ruled out by the condition (A2). But the lag number selected using the J(m, r) statistic to test for the significance of lagged terms will increase at the same rate as the prespecified k_{max}, itself increasing at a polynomial rate. Because a logarithmic rate of increase is slow compared to a polynomial rate, an Information Criterion will choose values of the lag number that are generally much smaller than those chosen by a general-to-specific t – test, for example. Although the log proportionality rule might fail the lower-bound condition (A2), the limiting distribution of the unit root τ_{DF} – test statistic is unaffected. In such a case the estimates of the coefficients on the stationary regressors Δy_{t-i} in the Augmented Dickey-Fuller autoregression will be consistent at a rate slower than $\sqrt{T}$ for some data generating processes.

CHAPTER 6

FINITE SAMPLE BEHAVIOR OF THE LAG LENGTH SELECTION METHODS: A MONTE CARLO INVESTIGATION

6.1 INTRODUCTION

The Dickey-Fuller (DF) and the Augmented Dickey-Fuller (ADF) unit root tests have been developed for testing the null hypothesis of a unit root against the alternative of stationarity. While the presence or absence of a unit root has important implications, many remain skeptical about the conclusions drawn from such tests. This concern is justifiable, as these specific tests generally suffer from two problems. First, in the case where the errors are serially correlated of the autoregressive type with an autoregressive polynomial root close to but less than unity, these tests have low power. Secondly, in the case where the errors follow a moving average process with a moving average polynomial root close to but less than unity, these tests suffer from severe size distortions. The consequence are over-rejections of the unit root hypothesis. In practice, while few economic time series are found to have serial correlation of the autoregressive type with an autoregressive polynomial root that is close to unity, many do exhibit a large moving average polynomial root. It is therefore desirable to have powerful unit root tests that are robust to size and power distortions.

The implementation of the Augmented Dickey-Fuller unit root tests necessitates the selection of an autoregressive truncation lag, say, k. This is required in the Augmented Dickey-Fuller autoregression used to form the Augmented Dickey-Fuller unit root τ_{DF} – test statistic.

As we will notice in this Chapter, Monte Carlo simulation experiments repeatedly show a strong association between k, the number of the lag that have to be included in the Augmented Dickey-Fuller autoregression, and the severity of size distortions and/or the extent of power loss. The problem is that while a small k is adequate for finite order autoregressive error processes and autoregressive moving average error processes with small moving average polynomial roots, a large k is generally necessary for error processes with a moving average polynomial root that is close to unity. For the latter class of error processes, lag length selection methods that presented in the previous Chapter 5 are going to be used for examining extensively the size and power properties of the τ_{DF} – test statistic when testing the hypothesis of the unit root in finite samples.

More specifically, we assume the following data generating processes:

$$y_t = \rho y_{t-1} + u_t$$

where

$$u_t \sim AR(p), MA(q) \text{ and } ARMA(p,q)$$

in order to test the null hypothesis of the unit root, that is $\rho = 1$, for various values of the parameters p and q and examine the effects of the autoregressive and moving average polynomial roots in combination with the lag length selection methods that presented in Chapter 5 on the size and power properties of the unit root τ_{DF} – test statistic.

The first half of this Chapter is mainly focused on presententing Monte Carlo simulation experiments from the Literature. Their majority is concentrated on the cases in which the errors follow a moving average and autoregressive process of order 1, because these cases occur frequently in time series considered in practice.

In the second half of this Chapter we construct and present an own Monte Carlo simulation study that is based on error processes which go beyond the standard AR(1) and MA(1) usually considered in the Literature.

The latter class of error processes are also analysed from their spectral features perspective in order for us to derive useful statistical information about the importance of knowing the distance of the autoregressive and/or moving average polynomial root/roots of an error process from the bounds of the unit circle in the context of testing unit root hypotheses in finite samples.

6.2 THE SIMPLEST METHOD: NG AND PERRON (1995)

In section 5.2 we saw that the simplest way to choose the lag length order of the Augmented Dickey-Fuller autoregression before applying the τ_{DF} – test statistic in order to test the null hypothesis of the unit root is that of being independent of the number of observations T, meaning, for example, that we can allow it to take any value we like.

Ng and Perron (1995) considered the following data generating processes:

MA case	AR case
$y_t = \rho y_{t-1} + u_t$	$y_t = \rho y_{t-1} + u_t$
where	where
$u_t \sim MA(1)$	$u_t \sim AR(1), AR(2), AR(3), AR(4)$
(MA parameter: θ)	(AR parameters: φ_i , i=1,2,3,4)
(MA case)	(AR case)

for numerous parameterizations of the error parameters θ and φ_i in order to test the hypothesis of the unit root, that is ρ=1, through the unit root τ_{DF} – test statistic. Fixing the order of the Augmented Dickey-Fuller autoregression to belong in the range [1, 10], they constructed Monte Carlo simulation experiments to investigate the performance of the unit root τ_{DF} – test statistic in the context of its size and the power in finite samples.

The results they reported are based on 5,000 simulations for different values of the error parameters θ and φ_i (i=1, 2, 3, 4). For each parameterization, the selected values of the lag number k and the corresponding values of the unit root τ_{DF} – test statistic are recorded. For a given sample size T, different values for k_{max} were examined for condition (A1) to be satisfied.

Tables (B.3) and (B.4) in Appendix B report their results for the moving average and autoregressive case using the experimental sample size of T=100 observations with k_{max} = 10.

For the moving average case, looking at the Table (B.3) we notice that when the moving average error parameter θ is equal to – 0.8, fixing k to be 4 yields an exact size of 28.3% instead of 5% with power 59% (instead of 100%). Furthermore, size distortions are much smaller as k becomes larger. This is not true for the power which gets worse as k grows. Size distortions are much smaller when the moving average error parameter θ is positive, but the unit root τ_{DF} – test statistic seems to be oversized when k is odd and undersized when k is even.

For the autoregressive case, looking at the Table (B.4) we notice that the exact size of the unit root τ_{DF} – test statistic for all choices of the lag number k is approximately close to the nominal size of 5% provided that k is at least equal to the true order. We also notice that an overparameterized model is associated with lower power, meaning that if the lag number k is further increased, the size of the unit root τ_{DF} – test statistic is approximately in the right level but power can drop dramatically.

Thus, we conclude that while a large chosen lag number k reduces the size distortions in both moving average and autoregressive case, it generally yields lower power.

Ng and Perron (1995) also constructed analogous Monte Carlo simulation experiments for the sample size choices of T=200 and T=500 observations (not presented in this thesis). The result, as they mentioned in their study, was that the power of the unit root τ_{DF} – test statistic increased for every value of k in both moving average and autoregressive case.

As far as the size of the unit root τ_{DF} – test is concerned, the results for the autoregressive case were qualitatively the same as when T=100. For the positive moving average error parameters, there was a zig-zag pattern of size distortions between odd and even choices of the lag number k and this existed even when T=500. **For the negative moving average error parameters, size distortions increased with T for any given value of the lag number k.**

The latter statement has to be emphasized. For example, they mentioned that when θ= - 0.8 and k=3 the exact size of the unit root τ_{DF} – test statistic increased from 43.4% to 59.8% as T increased from 100 to 500.

6.3 THE DETERMINISTIC METHOD: SCHWERT (1989)

Schwert (1989) first presented Monte Carlo evidence to point out the size distortion problems of the commonly Dickey-Fuller and the Augmented Dickey-Fuller unit root tests. He argued that the distribution of the Dickey-Fuller and the Augmented Dickey-Fuller unit root tests is far different from the distribution reported by Dickey-Fuller in finite samples if the underlying data generating process contains moving average error components.

To prove the above statements he considered the data generating process below:

$$y_t = \rho y_{t-1} + u_t$$

where

$$u_t \sim MA(1)$$

(MA parameter: θ)

(MA case)

and tested the null hypothesis of the unit root, that is ρ=1, using the experimental sample size choices of T=25, 50, 100, 250, 500 and 1000 observations. Each experiment was simulated 10,000 times in order for the sampling distribution of the unit root τ_{DF} – test statistic to be created. The moving average error parameter θ was set equal to -0.8, -0.5, 0, 0.5 and 0.8, respectively.

Defining:

$$k = \text{int}\left\{ c\left(\frac{T}{100}\right)^{\frac{1}{d}} \right\}$$

where 'int' denotes the integer before the first decimal point, and for given arbitrary constants c and d we can notice that the above choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression is a deterministic function of the number of observations T. Values of c=4 and 12 and d=4 were used for the lag number k to be calculated for the different sample size choices.

Schwert's results[3] (not reported in this thesis) showed that the size of the unit root τ_{DF} – test statistic had significant distortions from the specified nominal sizes of 1% and 5% and that it got significantly better with c=12 the closer the moving average error parameter θ was to -1. He also noted that the exact size of the unit root τ_{DF} – test statistic depends on the choice of the lag number k.

As we have already mentioned in section 5.3, the practical problem that appears when using this specific deterministic method for selecting the lag order of the Augmented Dickey-Fuller autoregression before the application of the unit root τ_{DF} – test statistic is that one is faced with a given sample size in practice and this fact leads to the right calculation of c and d in the deterministic function of the lag number k unless c and d happen to be chosen correctly. So, deterministic methods like this proposed by Schwert (1989) as well as the simplest one mentioned by Ng and Perron (1995) for choosing the lag order of the Augmented Dickey-Fuller autoregression are not good practical guidelines for statistical inference in finite samples.

However, studying the results of these two lag length selection methods, a significant problem when testing the unit root hypothesis through the τ_{DF} – test – statistic seems to appear when the actual data generating process is a unit root process with errors that follow a MA(1) process with a moving average error parameter θ that is close to -1, meaning that the moving average polynomial of the MA(1) error process has a root close to unity. That is the over-rejections of the unit root hypothesis.

[3] Interested readers should refer to Schwert (1989) for an extensive study on his Monte Carlo simulation experiment Tables.

6.4 THE GENERAL-TO-SPECIFIC AND THE INFORMATION CRITERIA (IC) METHODS: NG AND PERRON (1995)

Since we have presented the theory of the general-to-specific modeling strategy along with the Information Criteria for selecting the appropriate lag number k to be included in the Augmented Dickey-Fuller autoregression before testing the null hypothesis of the unit root through the τ_{DF} - test statistic we continue with the implications of these theoretical results to the τ_{DF} - test statistic in finite samples as we done in the case of the simplest and deterministic selection methods.

Ng and Perron (1995) constructed Monte Carlo simulation experiments to investigate the performance of the unit root τ_{DF} - test statistic in the context of its size and the power by allowing the lag number k in the Augmented Dickey-Fuller autoregression to take the value that is based on two Information Criteria, the AIC and BIC, and on the standard t-tests that are based on the significance on the last lag at the nominal levels of 5% and 10%, denoted by $t_{sig}(5)$ and $t_{sig}(10)$, respectively, according to the general-to-specific modeling strategy.

They considered the following data generating processes:

$y_t = \rho y_{t-1} + u_t$	$y_t = \rho y_{t-1} + u_t$
where	where
$u_t \sim MA(1)$	$u_t \sim AR(1), AR(2), AR(3), AR(4)$
(MA parameter: θ)	(AR parameters: φ_i , i=1,2,3,4)
(MA case)	(AR case)

for numerous parameterizations of the error parameters θ and φ_i in order to test the hypothesis of the unit root, that is ρ=1, through the unit root τ_{DF} – test statistic. The results they reported are based on 5,000 simulations for different values of the error parameters θ and φ_i (i=1, 2, 3, 4). For each parameterization, the selected values of the lag number k and the corresponding values of the unit root τ_{DF} – test statistic are recorded. For a given sample size T, different values for k_{max} were examined for condition (A1) to be satisfied. The experimental sample sizes were that of T = 100, 200 and 500 observations. The focus was for T = 100 with k_{max} =10.

Tables (B.5) and (B.6) give the frequency counts of the selected lag number k according to the lag length selection methods just mentioned in the moving average and autoregressive case for T = 100 observations, respectively. A quick look at these two Tables leads us to the conclusion that Information Criteria such that of AIC and BIC choose values of k that are consistently less than 3 for both moving average and autoregressive case. Furthermore, in the moving average case (Table (B.5)), the values of the lag number k selected by the general-to-specific modeling strategy t - tests are in the range [2, k_{max} =10] with some mass concentrated at k = 1.

On the other hand, for the autoregressive case (Table (B.6)), the lag number k selected according to the same lag length selection methods seems pretty well given that the data generating processes considered are of order no higher than 4. Information Criteria selection methods yield very parsimonious models when the data generating process is a unit root process with MA(1) errors no matter what is the value of the moving average error parameter θ, or equivalently, no matter what is the distance of the moving average polynomial root of the error process from unity. The cost of parsimony will be judged in terms of the size and the power of the unit root τ_{DF} - test statistic.

The results of Tables (B.7) and (B.8) are concerned with the size and the power of the unit root τ_{DF} - test statistic in the moving average and autoregressive case for T = 100 observations, respectively.

Turning first to the moving average case (Table (B.7)), we can notice that for the positive values of the moving average error parameter θ, the size of the unit root τ_{DF} - test statistic is similar to all methods of selecting the lag number k and can be said that is fairly close to the nominal size of 5%.

From the frequency counts point of view (Table (B.5)), when θ = - 0.8 the standard 10% t-test picks k to be 5 or smaller approximately 43% of the time, whereas the AIC picks k to be in the same range twice as often. Although such variations in the choice of the lag number k appear to yield small size differences, power is slightly higher the more parsimonious the model.

A result that must be emphasized is that of large size distortions when the moving average parameter θ of the error process is negative and close to the value of –1, or equivalently, when the moving average polynomial root of the error process is close to unity. We noticed the same phenomenon when we referred to the simplest and deterministic method of selecting the lag number k to be included in the Augmented Dickey-Fuller autoregression.

Now, for the autoregressive case (Table (B.8)), all methods of selecting the lag number k produce an exact size which is close to the nominal size of 5%. The standard 10% t-test ($t_{sig}(10)$) tends to have lower power, however.

Looking again at the frequency counts (Table (B.6)), the standard 10% t-test selects the lag number k to be greater than 4 more than 40% of the time when the actual data generating process is a unit root process with AR(4) errors indicating overparameterization. Thus, underparameterization is associated with larger size distortions and overparameterization with power loss in the case of T = 100 observations.

The size of the unit root τ_{DF} - test statistic for the moving average case with T = 200 observations is reported in Table (B.9). We notice that the size distortions persist as T increases. But in cases for which size distortions is not a issue, as in the autoregressive case, the discrepancies in power across selection procedures vanish almost completely when the sample size grows.

Table (B.10) shows the size and the power of the unit root τ_{DF} - test statistic for the autoregressive case at T = 200 observations. Compared to the results obtained with T = 100 observations, power is higher throughout, and the differences in power across selection methods are smaller.

So, standard t-tests have generally an advantage over Information Criteria in the sense that they produce more accurate size without much loss of power to the τ_{DF} - test statistic not forgetting that tend to oveparameterize unit root processes with underlying autoregressive error processes in some cases.

To summarize, an overly parsimonious estimated autoregression may have large size distortions and an overparameterized one may have too low power. The size problem is more severe than the power loss in the sense that differences in power across the selection methods diminish as sample size increases but the size distortions persist even for large sample sizes for some methods of selecting the lag number k. At this point of view, standard t – tests for the significance on the last lag, and especially the $t_{sig}(10)$, will have an advantage over the Information Criteria based rules such that of AIC or BIC because the former produce more robust properties across unit root processes of different underlying error processes.

6.5 THE COMPARATIVE PERFORMANCE OF THE MODIFIED INFORMATION CRITERIA METHODS: OUR OWN MONTE CARLO INVESTIGATION

In the previous sections we presented Monte Carlo simulation experiments in order to examine the size and power performance of the τ_{DF} - test statistic when testing the unit root hypothesis in finite samples using the various methods mentioned there for selecting the lag number k to be included in the Augmented Dickey-Fuller autoregression. We similarly proceed here and present Monte Carlo simulation experiments for examining the size and power performance of the τ_{DF} - test statistic when testing the unit root null in finite samples using now the Corrected (AICC, BICC) and the Modified (MAICC, MBICC) Information Criteria as lag length selection methods. We also make comparisons among all the lag length selection methods. Ng and Perron (2001) have investigated the performance of Modified Information Criteria (MICC) in the context of Generalized Least Squares (GLS) estimation for the parameters of the Augmented Dickey-Fuller autoregression. In this section we corfirm their findings in the context of Ordinary Least Squares (OLS) estimation for the parameters of the Augmented Dickey-Fuller autoregression.

Simulation studies reported up to now have come up with the conclusion that the standard t –test based on the significance on the last lag at the nominal size of 10% ($t_{sig}(10)$) works better than the standard Information Criteria (AIC, BIC) in the sense that when it chooses the lag number k to be included in the Augmented Dickey-Fuller autoregression it produces smaller size distortions to the τ_{DF} - test statistic with no great loss of power when testing the unit root hypothesis in finite samples.

For the above testing purposes and comparisons we constructed Monte Carlo simulation experiments using the experimental sample sizes of T = 100, 250 and 500 observations like those of Ng and Perron (1995, 2001). For each sample size, different values of k_{max} are used for condition (A1) to be satisfied. The selection of k_{max} is actually based again on the form:

$$k_{max} = \text{int}\left[10\left(\frac{T}{100}\right)^{1/4}\right]$$

which proposed and used by Ng and Perron (1995, 2001).

The results we report are based on 5,000 simulations. The data generating processes that are considered in this section are of the same form mentioned in section 6.2 for comparison purposes:

$$y_t = \rho y_{t-1} + u_t$$

where

$$u_t \sim MA(1)$$

(MA parameter: θ)

(MA case)

$$y_t = \rho y_{t-1} + u_t$$

where

$$u_t \sim AR(1), AR(2), AR(3), AR(4)$$

(AR parameters: φ_i , i=1,2,3,4)

(AR case)

For each parameterization of the error parameters θ and φ_i (i=1, 2, 3, 4), the selected values of the lag number k and the corresponding values of the unit root τ_{DF} – test statistic are recorded.

We first examine the number of times that k = i (i = 1, 2, …, 10) is being selected by the Corrected and Modified Information Criteria in 5,000 simulations using the sample size of T = 100 observations. We confirm that when the error process contains a strong moving average polynomial root close to unity Corrected Criteria tend to select higher k than the non – Corrected. Modified Criteria select even higher values.

Tables (B.11) and (B.12) represent frequency counts for the selected lag number k. A first sign from looking at both Tables is that the Corrected (AICC and BICC) and the Modified (MAICC and MBICC) Information Criteria select values of k that are less than 4 in both moving average and autoregressive case. If we now make a first comparison among all the lag length selection methods in the context of the frequency counts in the moving average case we notice that the Corrected and the Modified Information Criteria tend to select values for k that are generally larger. For example, when $\theta = -0.8$ we saw that the AIC picks k to be 4 or greater almost rarely in contrast now to the AICC that picks k to be 4 or greater approximately 62% and the MAICC approximately 86% of the time, percentages that are by far greater than the percentages we took even from the standard t – test for the significance on the last lag at the nominal size of 10%. The same conclusions hold if we compare the BIC with the BICC and MBICC but with the difference not being so dramatic. Analogous remarks are also derived when considering the other values of the moving average error parameter θ, leading us to the conclusion that the AICC and even more the MAICC is more flexible than the standard Information Criteria and the standard t - tests according to the general-to-specific modeling strategy.

Furthermore, the k's selected from the Corrected and Modified Information Criteria are concentrated at small numbers if someone thinks that the true order of such an estimated autoregression that tries to approximate a unit root process that contains moving average error terms is infinite. We will judge this parsimony in terms of the size and power of the unit root τ_{DF} – test statistic soon enough in this section.

Let us now make the comparison in the context of the frequency counts in the autoregressive case. The k's selected from all the lag length selection methods considered up to now seem appropriate given that the data generating processes are of order no higher than 4, not forgetting that the $t_{sig}(10)$-test tends to overparameterize unit root processes that have autoregressive error terms in some cases.

We continue our Monte Carco simulation analysis by investigating the size implications of the unit root τ_{DF} – test statistic using the Corrected and Modified Information Criteria as selection methods for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression.

Turning first to the moving average case (Tables (B.13) through (B.15)), we notice that for the positive values of the moving average error parameter θ and for all the experimental sample sizes the size of the unit root τ_{DF} – test statistic is almost in the same level and fairly close to the nominal size of 5%. This is a conclusion that holds for all the lag length selection methods we have seen up to now. The large size distortions of the unit root τ_{DF} – test statistic when the moving average error parameter θ is large and negative in the context of using the standard Information Criteria (AIC, BIC) and the standard t – tests ($t_{sig}(5)$, $t_{sig}(10)$) for significance on the last lag as lag length selection methods are well documented in the previous Monte Carlo simulation analysis.

The use of the Corrected and Modified Information Criteria for the selection of the lag number k to be included in the Augmented Dickey-Fuller autoregression improve by far the situation. For example, when θ = - 0.8 and T = 100 the size of the unit root τ_{DF} – test statistic under the use of AICC is 22.9% instead of 5%, and under the use of MAICC is 8.3% instead of 5%, indicating enormous improvements. Analogous conclusions derived when looking at the BICC and MBICC in the same Tables. Such improvements on the size of the unit root τ_{DF} – test statistic are also noticed at the experimental sample sizes of T = 250 and 500 observations, respectively.

A result that has to be emphasized at this point is that in the moving average case and when the moving average parameter θ of the error process is negative and close to the value of –1, meaning that the moving average polynomial of the error process has a root close to unity, the

MAICC has the best performance in the context of size of the unit root τ_{DF} – test statistic.

For the autoregressive case, Tables (B.16) through (B.18) indicate that the Corrected and Modified Information Criteria produce an exact size close to the desired nominal size, a fact that is also true for the standard ones as well as for the standard t –tests for the significance on the last lag.

In addition, in the autoregressive case, for the first two unit root processes where the errors follow an AR(1) process with different specified autoregressive error parameters we know that there exists an autoregressive polynomial root for each case that is not so close to unity. This notice made us to investigate the size and power performance of the unit root τ_{DF} – test statistic when a unit root process has autoregressive errors that indicate a root that is close to the bounds of the unit circle. So, additional Monte Carlo simulation experiments were constructed for studying in depth the performance of the unit root τ_{DF} – test statistic in the case were the errors follow an AR(1) process focusing mostly to the autoregressive error parameters φ = 0.8 and –0.8 which indicate an autoregressive polynomial root closer to the bounds of the unit circle compared to the root that derived under the autoregressive error parameters φ= 0.6 and –0.6. Tables (B.19) through (B.21) indicate that for all the values of the autoregressive error parameter φ, including the parameters φ = 0.8 and –0.8, and for all the experimental sample sizes the size of the unit root τ_{DF} – test statistic is close to the nominal size of 5%, meaning that the size of the τ_{DF} – test statistic is not affected or distorted whatever is the distance of the autoregressive polynomial root of the error process from the bounds of the unit circle.

The power performance of the unit root τ_{DF} – test statistic using the Corrected and Modified Information Criteria as lag length selection methods is going now to be examined.

Turning first to moving average case (Tables (B.22) and (B.23)), we notice that for positive values of the moving average error parameter θ where the size is around 5% the power is slightly higher the more parsimonious the model. It is well known that the BIC or the BICC criterion imposes a heavy penalty for overparameretization. Thus, for moving average error processes with a positive moving average parameter θ, the BIC and BICC tend to yield higher power for a given sample size. For the negative values of the moving average error parameter θ there is no need to focus on the power of the unit root τ_{DF} – test statistic since the size of the test can not be considered in a constant level for the power to be evaluated but we have calculated it.

In the autoregressive case (Tables (B.16), (B.17), (B.19) and (B.20)), the power of the unit root τ_{DF} – test statistic is in the same range for all the lag length selection methods with the AIC being slightly higher and any differences across them vanish almost completely when T increases.

A general feature of our results is that the Modified Information Criteria (MICC) and especially the Modified Akaike Information Criterion (MAICC) takes the advantage of the standard t –test on the significance on the last lag at the nominal size of 10% (t_{sig}(10)) according to the general – to – specific modeling strategy since the presented Monte Carlo simulation analysis showed that they produce more accurate size and power properties for the unit root τ_{DF} - test statistic especially when the data generating processes are unit root processes with moving average error terms.

6.6 SOME INTERESTING CASES IN THE CONTEXT OF TESTING UNIT ROOT HYPOTHESES: OUR OWN MONTE CARLO INVESTIGATION

It is well known that the true spectrum of a moving average process of order 1 indicates a 'hole' at the frequency zero that is 'stronger' the closer the moving average polynomial root is to unity. It is also well known that the true spectrum of an autoregressive process of order 1 indicates a 'peak' at the frequency zero that is 'stronger' the closer the autoregressive polynomial root is to unity.

From the Monte Carlo simulation analysis reported up to now we noticed the inability of the τ_{DF} – test statistic to keep the nominal size when the actual data generating process is of the form:

$$y_t = y_{t-1} + u_t$$

where

$$u_t \sim MA(1)$$

(MA parameter: θ)

(MA case)

and more specifically when the error process has a moving average polynomial root that is close to unity. Having in mind the fundamental characteristic ("hole") of the true spectrum of a moving average process that has a moving average polynomial root close to unity, this inability seems to be a result of the occurrence of the 'hole' at the frequency zero and increases the 'stronger' the 'hole' is at the frequency zero.

From the same Monte Carlo simulation analysis we also noticed the ability of the τ_{DF} – test statistic to keep the nominal size when an actual data generating process is of the form:

$$y_t = y_{t-1} + u_t$$

where

$$u_t \sim AR(1), AR(2), AR(3), AR(4)$$

(AR parameters: φ_i , i=1, 2, 3, 4)

(AR case)

Having in mind the fundamental characteristic ("peak") of the true spectrum of an autoregressive process that has an autoregressive polynomial root close to unity, it seems that the occurrence of a "mild" or "strong" "peak" at the frequency zero in the true spectrum of the autoregressive error process of order 1 does not affect the size performance of the unit root τ_{DF} – test statistic. A question that arises at this point is the following:

What can we tell about the size performance of the unit root τ_{DF} – test statistic in the case where the actual data generating process is a unit root process with moving average or autoregressive errors that are modeled by more complicated than a MA(1) or AR(1) processes?

or similarly

What can we tell about the size performance of the unit root τ_{DF} – test statistic in the case where the root/roots of the moving average and/or autoregressive polynomial of the error process are close to unity, indicating 'holes' and 'peaks' in the true spectrum of the error process, at different from zero frequencies as well as at various combinations of them including the frequency zero?

Figures (A.12) through (A.34) illustrate some of these interesting cases which are taken into consideration in Tables (B.24) through (B.26).

To answer the above questions that are of great interest, analogous to the previous Monte Carlo simulation experiments were constructed focusing only to the experimental sample size of T = 100 observations.

When the error process contains moving average polynomial roots at frequencies far from zero then the size of the unit root τ_{DF} – test statistic is acceptable with any of the lag length selection criteria

Turning first to the case where the data generating process is a unit root process that has errors that follow moving average processes more complicated than the MA(1) process and with characteristics like those indicated in Figures (A.12) through (A.19) we notice from Table (B.24) that the size of the unit root τ_{DF} – test statistic is fairly close to the nominal size of 5% when the root/roots of the moving average polynomial of the error process is/are close to the unit circle but at a frequency or frequencies away from zero.

When the error process contains a strong moving average polynomial root at frequency zero the size of the unit root τ_{DF} – test statistic has the best performance under the use of MAICC

The problem starts to appear again when the moving average polynomial of the error process has at least one root close to unity and at the frequency zero no matter at what frequencies are its other roots and how close or not are to the bounds of the unit circle.

For example, when we consider as a data generating process a unit root process with errors that follow a moving average process of order 3 with specified parameters θ_1 = - 0.99, θ_2 = 0.98, θ_3 = - 0.97 (case 4 in Table (B.24)) in order its polynomial to have two roots equal to the value of 0.99 at the frequencies zero and 0.5π, respectively, (see Figure (A.15)) we notice that the size of the unit root τ_{DF} – test statistic using the standard and the Corrected Information Criteria as well as the standard t – tests for the significance on the last lag according to the general-to-specific modeling strategy for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression is 39% (using the AIC), 45.1% (using the BIC), 27% (using the $t_{sig}(10)$), 29% (using the $t_{sig}(5)$), 33.8% (using the AICC) and 48.4% (using the BICC) instead of 5% indicating enormous size distortions.

On the other hand, the size of the unit root τ_{DF} – test statistic using the Modified Information Criteria for the choice of the lag number to be included in the Augmented Dickey-Fuller autoregression is 19.85% (using the MAICC) and 23.5% (using the MBICC) which are by far better than the results obtained using the Corrected Information Criteria noting also here significant size distortions.

When the error process is purely autoregressive the size performance of the unit root τ_{DF} – test statistic is not influenced with any of the lag length

<u>selection criteria no matter how close the autoregressive polynomial roots are to unit circle and at what frequencies</u>

Turning now to the case where the data generating process is a unit root process with errors that follow autoregressive processes more complicated than the AR(1) process and with characteristics like those indicated in Figures (A.20) through (A.26) we notice from Table (B.25) that the size of the unit root τ_{DF} – test statistic is close to the nominal size of 5% for all the lag length selection methods when the root/roots of the autoregressive polynomial of the error process is/are close to unity at frequency zero, at frequencies far from zero or at any combination of them whatever is the selection method for choosing the lag number k to be included in the Augmented Dickey-Fuller autoregression.

<u>Under autoregressive moving average error processes, when a strong moving average polynomial root at frequency zero is combined with a strong autoregressive polynomial root elsewhere, the size distortions of the unit root τ_{DF} – test statistic for all the lag length selection criteria are reduced the stronger the autoregressive polynomial root</u>

We also checked the size performance of the unit root τ_{DF} – test statistic when the actual data generating process is a unit root process with errors that follow ARMA(p, q) processes and with characteristics like those indicated in Figures (A.27) through (A.34).

More specifically, we investigated the case where the autoregressive polynomial of the error process has root/roots close to the bounds of the unit circle at various frequencies including the frequency zero indicating ‘peaks’ in the true spectrum of the autoregressive moving average error process and the case where the moving average polynomial of the same error process has root/roots close to the bounds of the unit circle at various frequency choices including the frequency zero, indicating ‘holes’ in the same true spectrum of the autoregressive moving average error process.

The first case that is considered in Table (B.26) examines the size performance of the τ_{DF} – test statistic when is applied to test the unit root hypothesis in the Augmented Dickey-Fuller autoregression that is trying to approximate a unit root process (of T=100 observations) with errors that follow an ARMA(2, 1) process of specified autoregressive and moving average parameters φ_1= - 0.31, φ_2= - 0.25 and θ= - 0.99 in order to have the following characteristics: (a) the autoregressive polynomial root is equal to the value of 0.5 at frequency 0.6π indicating a mild "peak" at this frequency in the true spectrum of the error process and (b) the moving average polynomial root is equal to the value of 0.99 at frequency zero indicating a strong "hole" at this frequency in the true spectrum of the same error process (see Figure (A.27)). We notice that the size of the unit root τ_{DF} – test statistic using the standard t – tests ($t_{sig}(5)$, $t_{sig}(10)$) for the significance on the last lag for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression is 23.94% (using the $t_{sig}(5)$) and 21.4% (using the $t_{sig}(10)$) instead of 5% indicating enormous size distortions. The same conclusions hold for the size of the unit root τ_{DF} – test statistic when using the standard as well as the Corrected Information Criteria (AIC, BIC, AICC, BICC) for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression. Furthermore, the size of the unit root τ_{DF} – test statistic when using the Modified Information Criteria (MAICC, MBICC) for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression is 14.9% (using the MAICC) and 15.02% (using the MBICC) instead of 5% which are by far better than the results obtained when using the previous lag length selection methods noting also here significant size distortions.

The second, third and fourth case that is considered in Table (B.26) examine the size performance of the τ_{DF} – test statistic when the actual data generating process is a unit root process (of T=100 observations) with errors that follow an ARMA(2, 1) process of specified autoregressive and moving average parameters (φ_1= - 0.43, φ_2= - 0.49 and θ= - 0.99), (φ_1= - 0.56, φ_2= - 0.81 and θ= - 0.99), (φ_1= - 0.61, φ_2= - 0.98 and θ= - 0.99), respectively, in order to have the following characteristics: (a) the autoregressive polynomial roots are equal to the values of 0.7, 0.9 and 0.99 at frequency 0.6π indicating stronger "peaks" than the first case in this frequency in the true spectrum of the error process and (b) the moving average polynomial root is equal to the value of 0.99 at frequency zero indicating a strong "hole" at this frequency in the true spectrum of the same error process (see Figures (A.28) through (A.30)). We notice that the size distortions of the unit root τ_{DF} – test statistic using the standard t – tests ($t_{sig}(5)$, $t_{sig}(10)$) for the significance on the last lag for the choice of the lag number to be included in the Augmented Dickey-Fuller autoregression is getting smaller the stronger the "peak" is in the true spectrum of the error process at frequency 0.6π. The same conclusions hold for the size of the unit root τ_{DF} – test statistic when using the standard, the Corrected and the Modified Information Criteria for the choice of the lag number to be included in the Augmented Dickey-Fuller autoregression, with the latter ones, and especially the MAICC, giving the more acceptable sizes for the unit root τ_{DF} – test statistic.

The inability of the τ_{DF} – test statistic to keep the nominal size when the actual data generating process is a unit root process with errors that follow a moving average process with the characteristic that the moving average polynomial has a root close to unity at frequency zero, indicating a strong "hole" at frequency zero in the true spectrum of the error process, is well documented in the previous Monte Carlo simulation experiments. Looking at the latter four cases we notice that the combination of a strong "hole" at frequency zero and a strong "peak" at any frequency choices different from zero in the true spectrum of an error process makes this inability to get decreased the stronger the "peak" is in the true spectrum of the error process.

The next two cases in Table (B.26) (cases 5 and 6) make the above statement more readable. More specifically, the fifth case examines the size performance of the τ_{DF} – test statistic when the actual data generating process is a unit root process (of T=100 observations) with errors that follow an ARMA(4, 1) process of specified autoregressive and moving average parameters φ_1= - 0.63, φ_2= - 0.84, φ_3= - 0.17, φ_4= - 0.40 and θ= - 0.99 in order to have the following characteristics: (a) the autoregressive polynomial roots are equal to the values of 0.7 at frequency 0.4π and 0.9 at frequency

0.7π indicating a mild "peak" at the frequency 0.4π and a stronger one at the frequency 0.7π in the true spectrum of the error process and (b) the moving average polynomial root is equal to the value of 0.99 at frequency zero indicating a strong "hole" at this frequency in the true spectrum of the same error process (see Figure (A.31)). We notice that the size of the unit root τ_{DF} – test statistic using the standard t – tests ($t_{sig}(5)$, $t_{sig}(10)$) for the significance on the last lag for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression is 10.08% (using the $t_{sig}(5)$) and 8.72% (using the $t_{sig}(10)$) instead of 5% indicating significant size distortions. The same conclusions hold for the size of the unit root τ_{DF} – test statistic when using the standard as well as the Corrected Information Criteria (AIC, BIC, AICC, BICC) for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression. Furthermore, the size of the unit root τ_{DF} – test statistic using the Modified Information Criteria (MAICC, MBICC) for the choice of the lag number k to be included in the Augmented Dickey-Fuller autoregression is 6.64% (using the MAICC) and 8.37% (using the MBICC) instead of 5% which are by far better than the results obtained using the previous lag length selection methods noting also here size distortions. The sixth case that is considered in Table (B.26) is the same with the fifth one with the difference that the autoregressive polynomial roots are closer to the value of unity indicating stronger "peaks" in the true spectrum of the error process ARMA(4, 1) (see Figure (A.32)). We also notice in this case that the size of the unit root τ_{DF} – test statistic is fairly close to the nominal size of 5%, meaning that the combination of a strong "hole" at frequency zero and strong "peaks" at any frequency choices in the true spectrum of an error process does not produce significant size distortions to the τ_{DF} – test statistic when testing the unit root hypothesis in finite samples.

The case before the last one that is considered in Table (B.26) examine the size performance of the τ_{DF} – test statistic when the actual data generating process is a unit root process (of T=100 observations) with errors that follow an ARMA(2, 2) process of specified autoregressive and moving average parameters $\varphi_1 = -1.16$, $\varphi_2 = -0.98$ and $\theta_1 = -1.16$, $\theta_2 = 0.98$, respectively, in order to have the following characteristics: (a) the autoregressive polynomial root is equal to the value of 0.99 at frequency 0.3π indicating a strong "peak" at this frequency in the true spectrum of the error process and (b) the moving average polynomial root is equal to the value of 0.99 at frequency 0.7π, indicating a strong "hole" at this frequency in the true spectrum of the same error process (see Figure (A.33)). We notice in this case that the size of the unit root τ_{DF} – test statistic is not distorted, meaning that a strong "hole" at a frequency different from zero or a combination of a strong "hole" at a frequency different from zero and strong "peaks" at any frequency choices in the true spectrum of an error process does not produce significant size distortions to the τ_{DF} – test statistic when testing the unit root hypothesis in finite samples. The last case that is considered in Table (B.26) and is illustrated in Figure (A.34) leads to the same above conclusions.

6.7 SUMMARY

In this Chapter we analyzed issues related to the selection of the truncation lag number k that has to be included in the Augmented Dickey-Fuller autoregression before the application of the unit root τ_{DF} – test statistic proposed by Dickey and Fuller (1979) and Said and Dickey (1984). We have mainly focused on the implications of the lower-bound condition (A2) imposed by Said and Dickey (1984) on the Dickey-Fuller distribution. Lag length selection procedures that do not satisfy this condition tend to select truncation lag numbers that are too small for some parameter values of an error process. Standard Information Criteria based rules such as AIC and BIC fit into this category.

A general feature of the presented Monte Carlo simulation analysis is that an overly parsimonious model can have large size distortions, but an overparameterized model may have low power. But the size problem is more severe than the power loss in the sense that discrepancies in power across lag length selection procedures diminish as T increases, but size distortions persist even for large sample sizes for some methods of selecting the lag number that has to be included in the Augmented Dickey-Fuller autoregression. In this regard, the standard t – test at the nominal size of 10%, $t_{sig}(10)$, for the significance on the last lag according to the general-to-specific scheme will have an advantage over standard Information selection methods such as the AIC and BIC, because the former does not produce significant size and power distortions to the τ_{DF} – test statistic when it tests the unit root hypothesis in finite samples not forgetting also that tend to overparameterize in some cases. Furthermore, the Monte Carlo simulation study has repeatedly shown a strong association between the lag number k and the severity of size distortions and/or power loss. The problem is that while a small k is adequate for finite order autoregressive error processes and autoregressive moving average error processes with small moving average polynomial roots, a large k is generally necessary for noise functions with a moving average root that is close to unity and especially when the true spectrums of these error processes indicate this specific polynomial root at frequency zero.

Ng and Perron (2001) provided an improved procedure for choosing the lag number k that has to be included in the Augmented Dickey-Fuller autoregression before the application of the unit root τ_{DF} – test statistic. They argued that the penalty k assigned to overfitting in standard Information Criteria such as the AIC and BIC underestimates the cost of a low order model when the unit root process contains errors that have moving average terms that indicate a moving average polynomial root that is close to unity and especially when the true spectrum of these errors indicate this specific polynomial root at frequency zero. For the latter case, they suggest a class of Modified Information Criteria (MICC) that takes better account of the cost of underfitting. In particular, the Modified Akaike's Information Criterion (MAICC) is shown to lead to substantial size and power improvements over standard Information Criteria and standard t – tests on the significance on the last lag according to general-to-specific modeling scheme. The key distinction between the Modified and standard Information Criteria is that the former include a penalty factor that is sample dependent and take account of the fact that the bias in the estimate of the sum of the autoregressive parameters is highly dependent on the lag number k that has to be included in the Augmented Dickey-Fuller autoregression.

Our own Monte Carlo simulation analysis shows that the Modified Information Criteria and especially the Modified Akaike Information Criterion (MAICC) improve by far the size and power properties of the unit root τ_{DF} – test statistic especially in the case in which there are errors that are modeled by a MA(1) process with a moving average polynomial root close to unity.

To provide a deep insight into the issue of testing unit root hypotheses in finite samples and choosing the appropriate order for the Augmented Dickey-Fuller autoregression in order the Augmented Dickey-Fuller unit root τ_{DF} – test statistic to have robust and satisfactory size and power properties we study various error processes from their spectral perspective. Our own Monte Carlo simulation analysis shows that (a) the unit root τ_{DF} – test statistic is being significantly distorted when there is strong and negative serial correlation of the moving average type at frequency zero with the MAICC having the best performance (b) the size and power properties of unit root test τ_{DF} – test statistic are not significantly affected when there is serial correlation of the autoregressive type with any of the lag length selection procedures and (c) under autoregressive moving average error processes, when a strong moving average polynomial root at frequency zero is combined with a strong autoregressive polynomial root elsewhere, the size distortions of the unit root τ_{DF} – test statistic for all the lag length selection methods are reduced the stronger the autoregressive polynomial root.

CHAPTER 7

CONCLUSIONS

An enormous amount of analytical literature has been focused on testing unit roots in time series. When the true process that generates the data is a random walk, Fuller (1976) and Dickey-Fuller (1979, 1981) derived limiting distributions of unit root test statistics, the so called Z_{DF} and τ_{DF} Dickey-Fuller test statistics, when the estimated model is (a) the true model, (b) a random walk with shift in mean and (c) random walk with shift in mean and a linear time trend. The distribution of the Dickey-Fuller unit root test statistics relied on the error process being white noise, and so these test statistics are not valid if the error process is serially correlated. Such a restriction is a considerable drawback in applying these tests to time series in practice.

Dickey-Fuller (1979, 1981) and Said and Dickey (1984) considered the latter case and derived the limiting distribution of the unit root test statistics when the time series is an autoregressive integrated of order 1 and autoregressive integrated of order 1 moving average process, denoted by ARIMA(p,1,0) and ARIMA(p,1,q), respectively, with known p and q. These are the so called Augmented Dickey-Fuller test statistics. In some circumstances p and q may be unknown, in which Said and Dickey (1984) demonstrated that an ARIMA(p,1,q) process can be approximated by an ARIMA(k_T,1,0) process, the so called Augmented Dickey-Fuller autoregression, where k_T is a function of the sample size T and has to be chosen so that to satisfy some certain assumptions and conditions. This is a model selection problem. The usual statistical theory involves the selection of a model using various procedures including applications of significance tests such that of standard t – tests or maximization of specified criteria that work in a satisfactory way under certain conditions such that of AIC and BIC. These selection procedures are developed for giving satisfactory results asymptotically, so Monte Carlo simulation analysis is needed for examining their performance in finite samples.

The Monte Carlo simulation experiments presented in the literature show that when there are errors that are modeled by a moving average process of order 1, denoted by MA(1), with a moving average polynomial root close to unity, a high order Augmented autoregression is necessary for the unit root test statistics to have good size and power properties, but selection criteria of the type mentioned above tend to select the truncation lag k_T that is very small.

Ng and Perron (2001) considered a class of Modified Information Criteria (MICC) with a penalty factor that is sample dependent. It takes into account the fact that the bias in the sum of the autoregressive parameters is highly dependent on k_T. These Criteria yield tests which seem to be able to keep the nominal size much better than the Standard Information Criteria (AIC, BIC) do.

Our own Monte Carlo simulation analysis shows that the Modified Information Criteria and especially the Modified Akaike Information Criterion (MAICC) improve by far the size and power properties of the unit root test statistics especially in the case where there are errors that are modeled by a MA(1) process with a moving average polynomial root close to unity.

To provide a deep insight into the issue of testing unit root hypotheses in finite samples and choosing the appropriate order for the Augmented Dickey-Fuller autoregression in order the Augmented Dickey-Fuller unit root test statistics to have robust and satisfactory size and power properties we study various error processes from their spectral perspective. Our own Monte Carlo simulation analysis shows that (a) the unit root τ_{DF} – test statistic is being significantly distorted when there is strong and negative serial correlation of the moving average type at frequency zero with the MAICC having the best performance (b) the size and power properties of unit root τ_{DF} – test statistic are not significantly affected when there is serial correlation of the autoregressive type with any of the lag length selection procedures. A new result of our own simulation study is that (c) under autoregressive moving average error processes, when a strong moving average polynomial root at frequency zero is combined with a strong autoregressive polynomial root elsewhere, the size distortions of the unit root τ_{DF} – test statistic for all the lag length selection methods are reduced the stronger the autoregressive polynomial root.

APPENDIX A

FIGURES

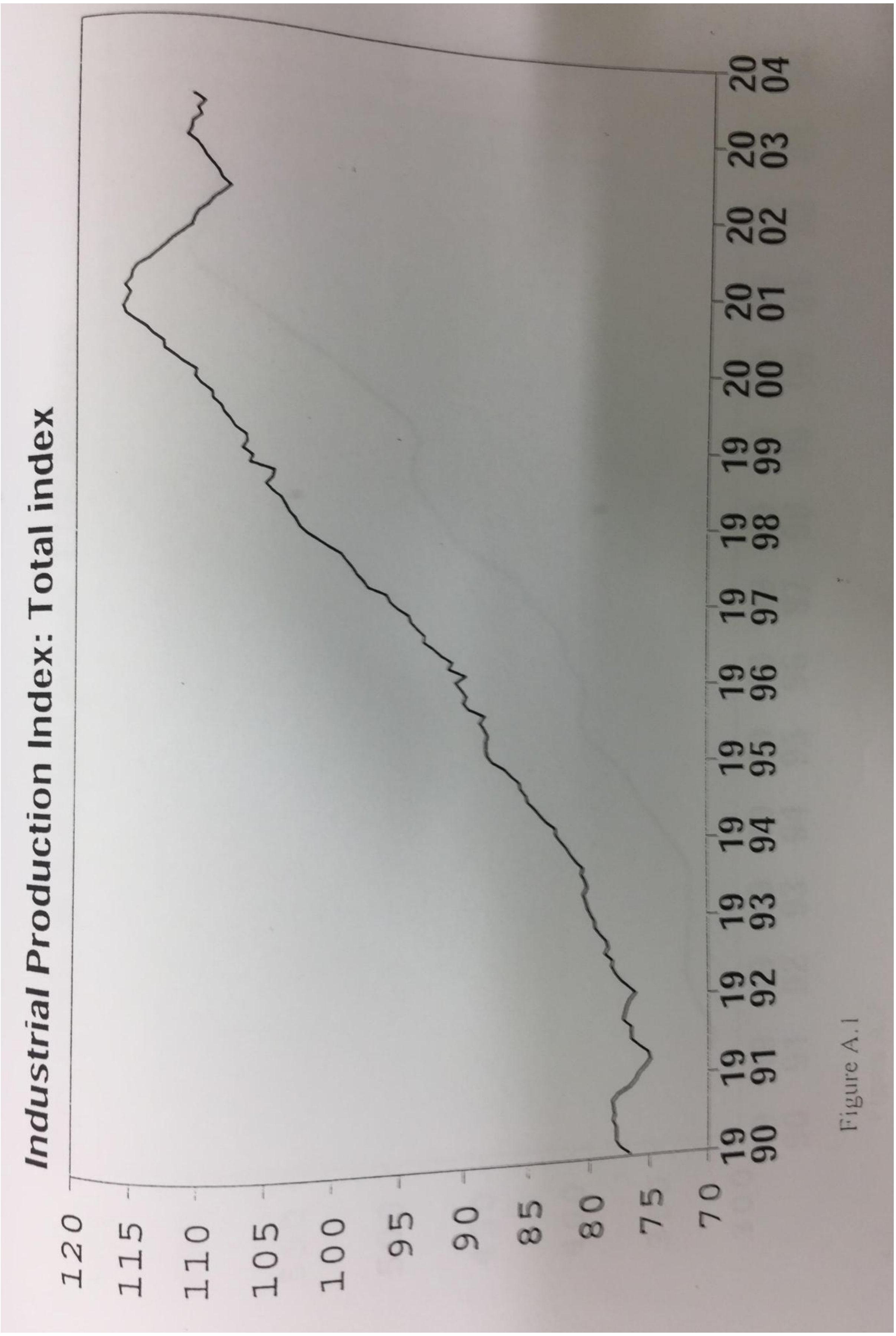

Industrial Production Index: Total index
120
115
110
105
100
95
90
85
80
75
70
19 90
19 91
19 92
19 93
19 94
19 95
19 96
19 97
19 98
19 99
20 00
20 01
20 02
20 03
20 04

Figure A.1

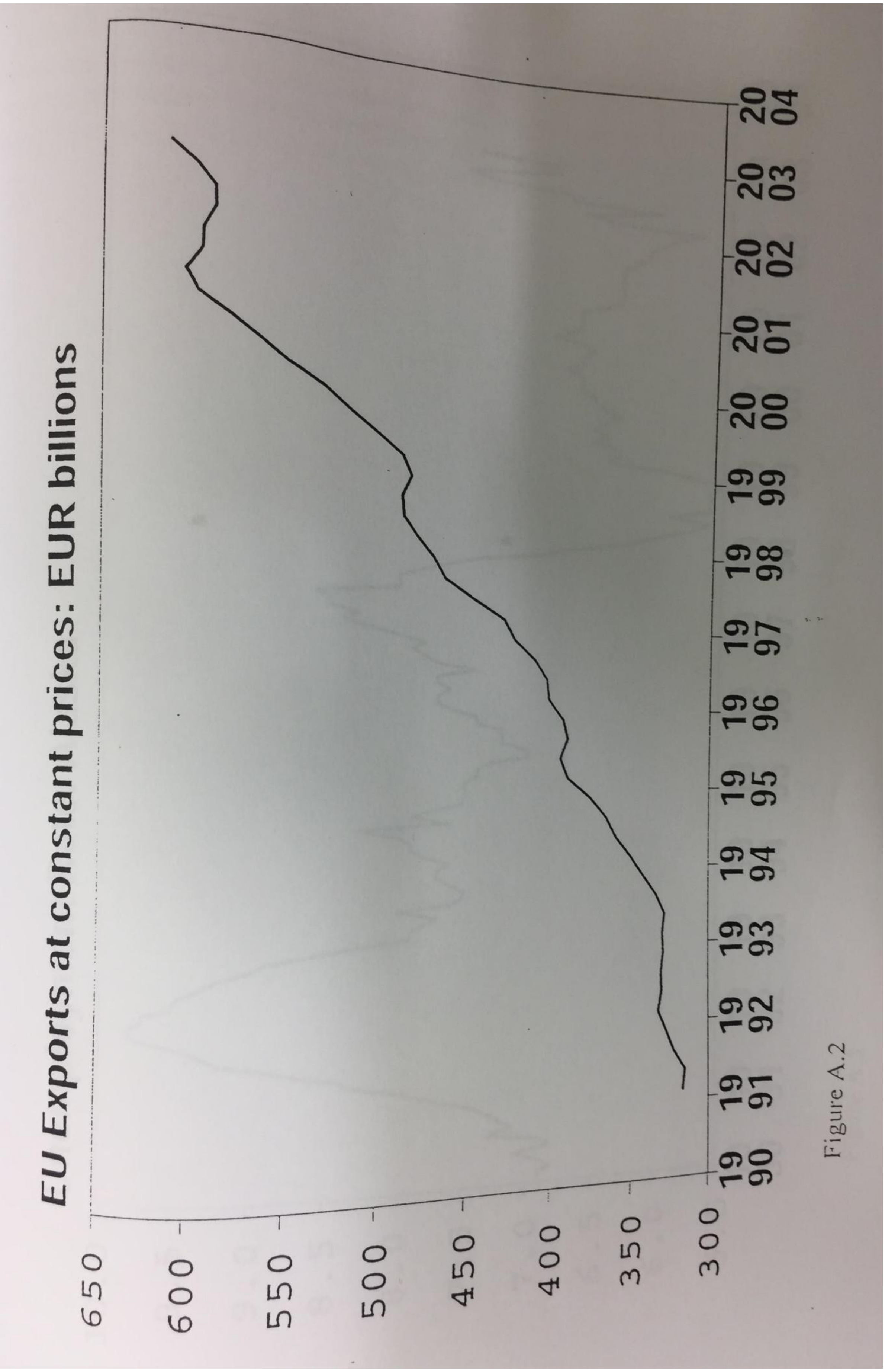

EU Exports at constant prices: EUR billions
650
600
550
500
450
400
350
300
19 90
19 91
19 92
19 93
19 94
19 95
19 96
19 97
19 98
19 99
20 00
20 01
20 02
20 03
20 04

Figure A.2

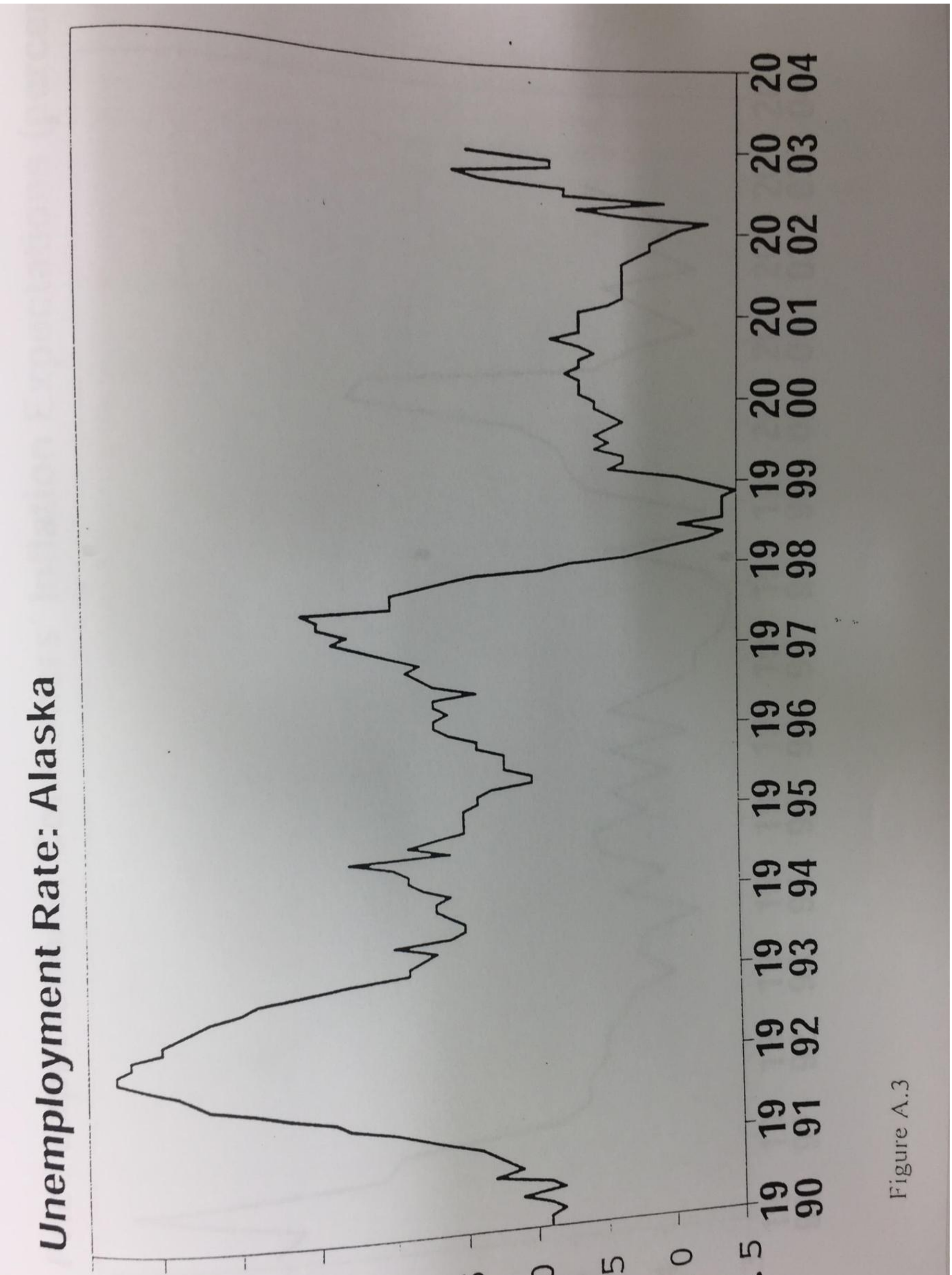
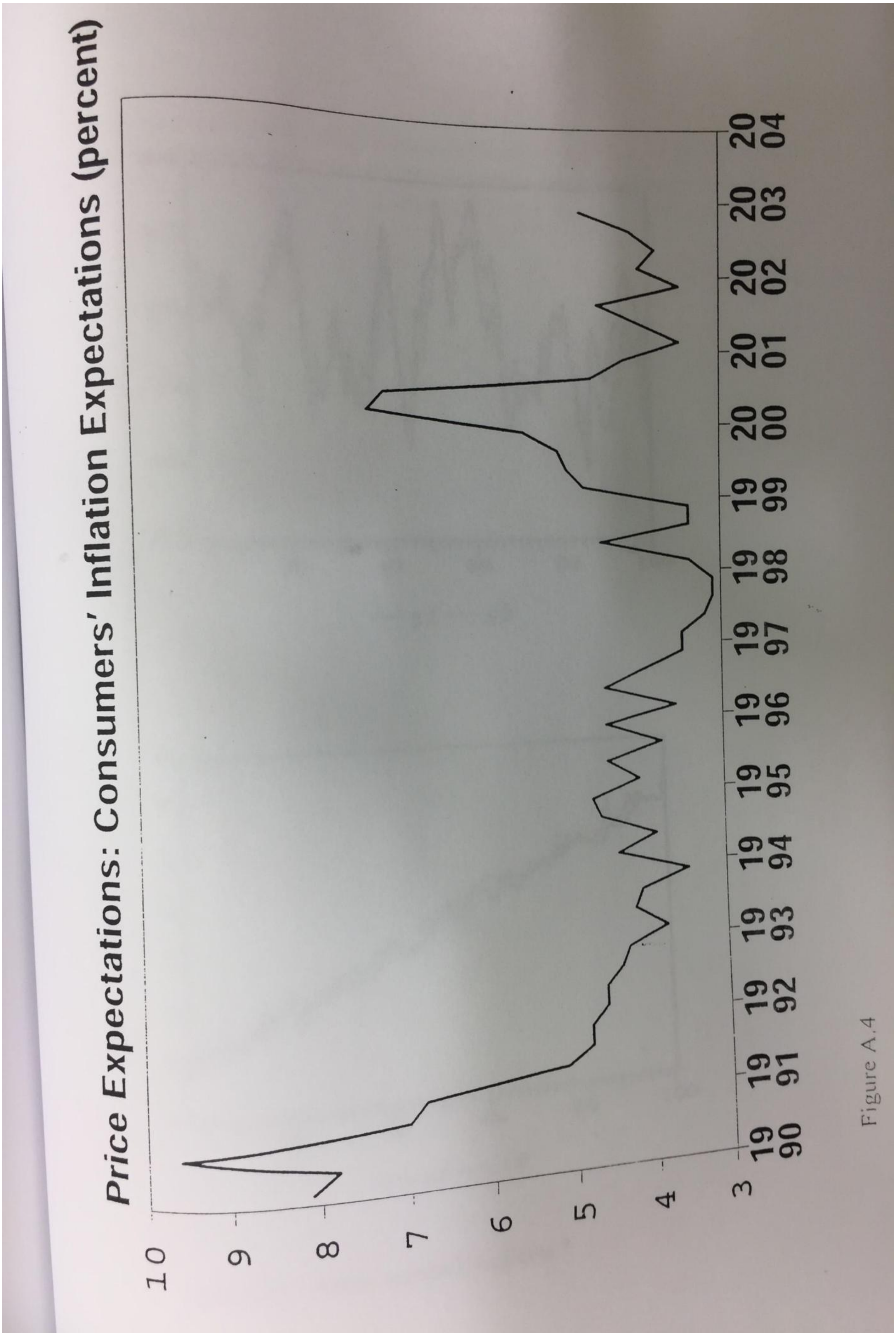

Unemployment Rate: Alaska
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
19 90
19 91
19 92
19 93
19 94
19 95
19 96
19 97
19 98
19 99
20 00
20 01
20 02
20 03
20 04

Figure A.3


Price Expectations: Consumers' Inflation Expectations (percent)
10
9
8
7
6
5
4
3
19 90
19 91
19 92
19 93
19 94
19 95
19 96
19 97
19 98
19 99
20 00
20 01
20 02
20 03
20 04
Figure A.4

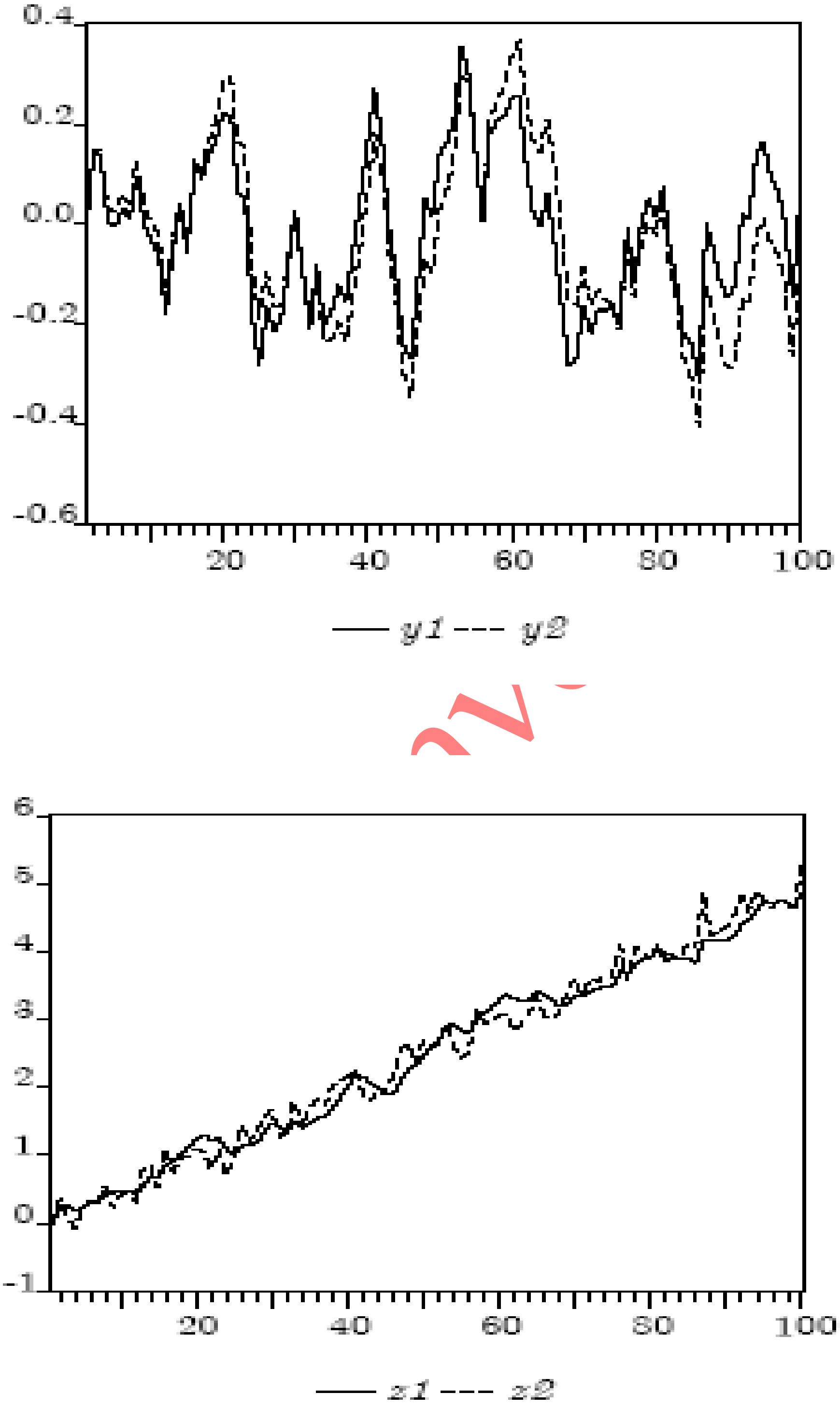


Figure A.5 – Where are the Unit Roots ?

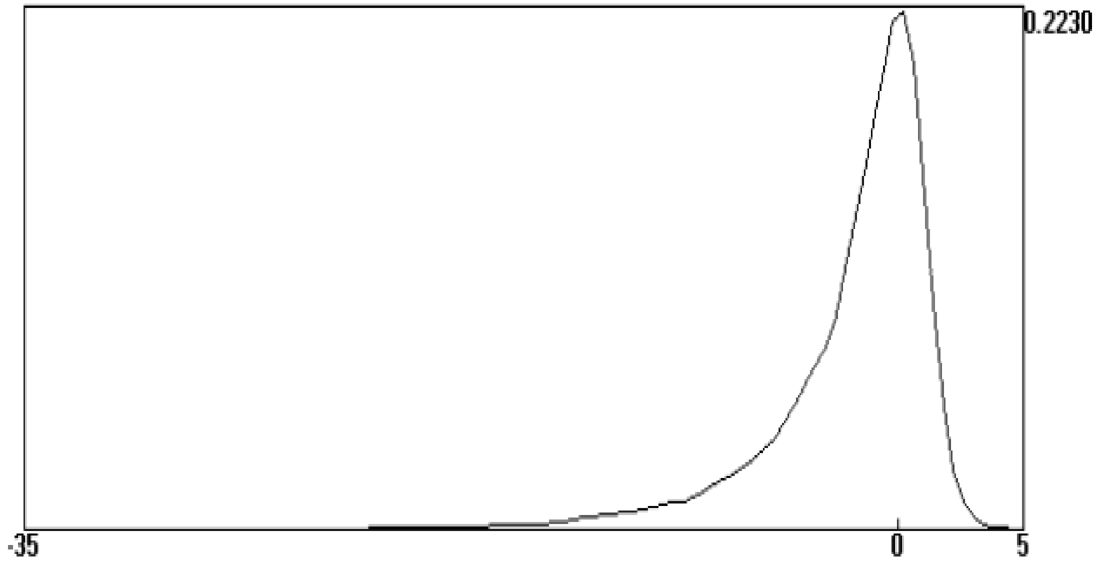


Figure A.6 – Density of the unit root Z_{DF} – test statistic under case 1

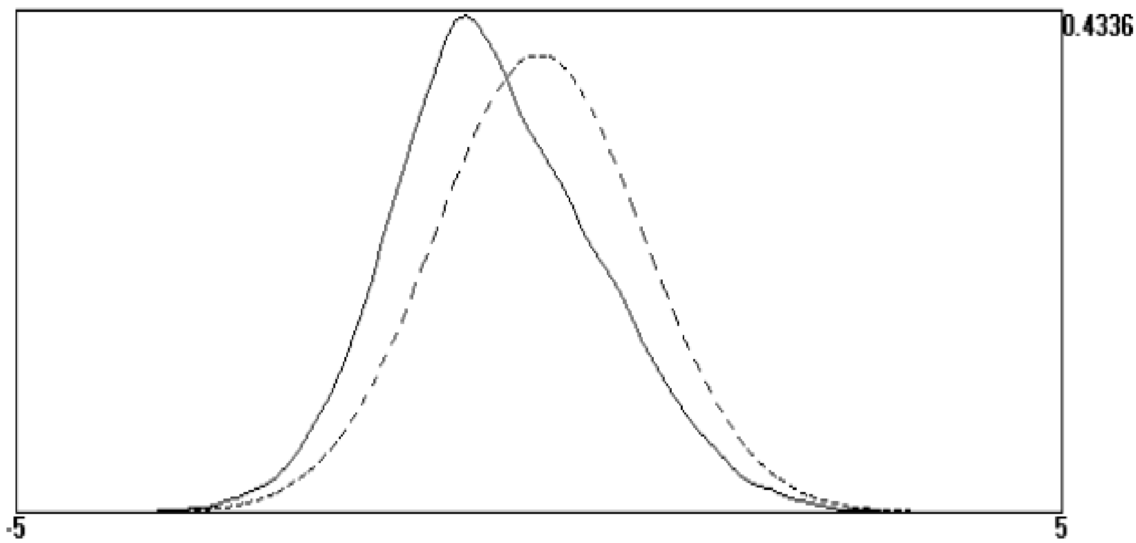


Figure A.7 – Density of the unit root τ_{DF} – test statistic compared with the standard Normal density (dashed curve) under case 1

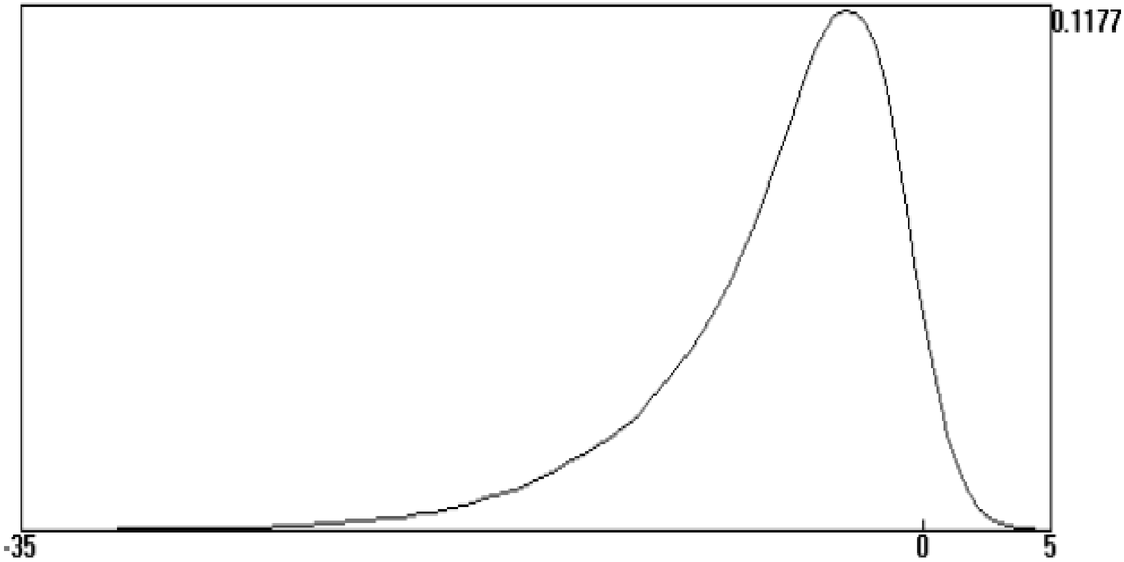


Figure A.8 – Density of the unit root Z_{DF} – test statistic under case 2

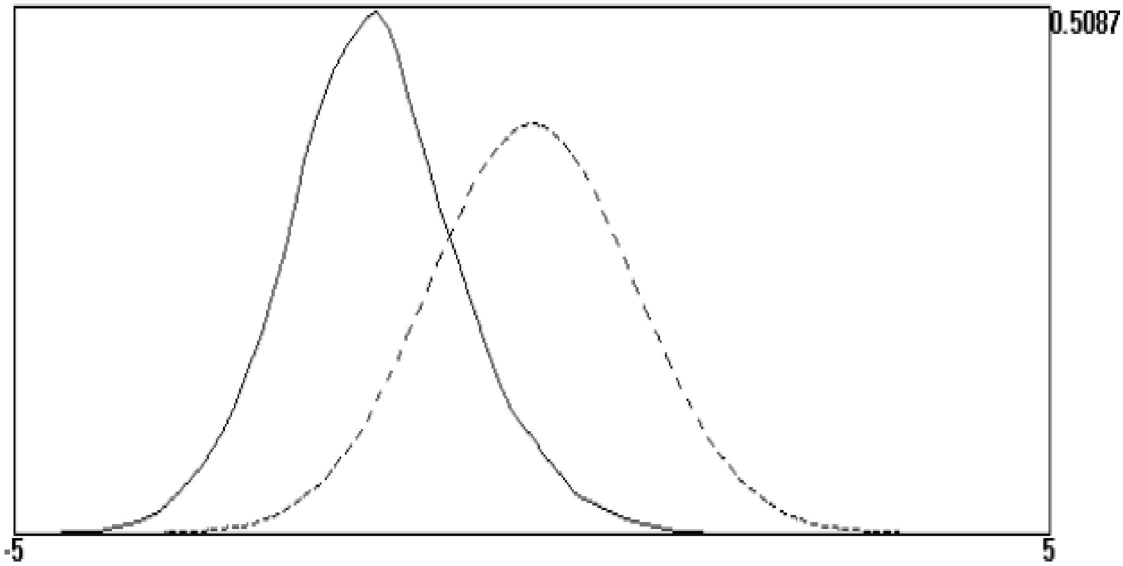


Figure A.9 – Density of the unit root τ_{DF} – test statistic compared with the standard Normal density (dashed curve) under case 2

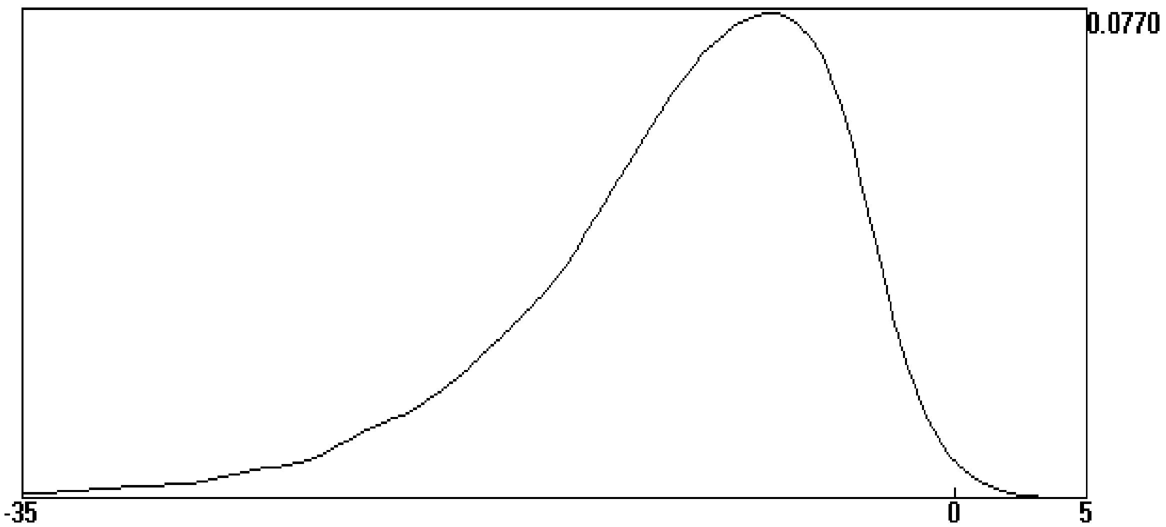


Figure A.10 – Density of the unit root Z_{DF} – test statistic under case 4

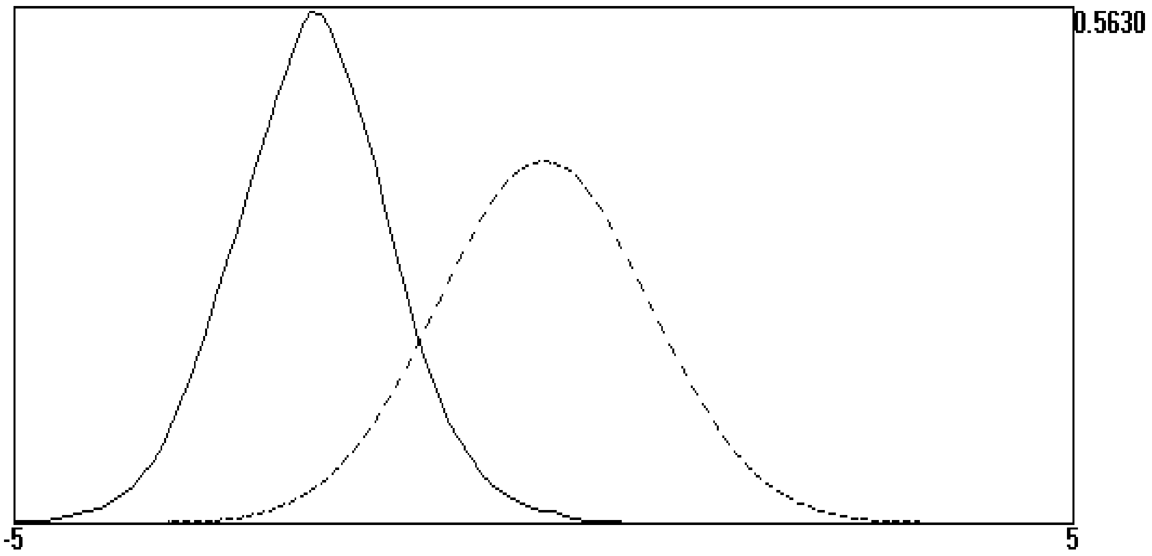
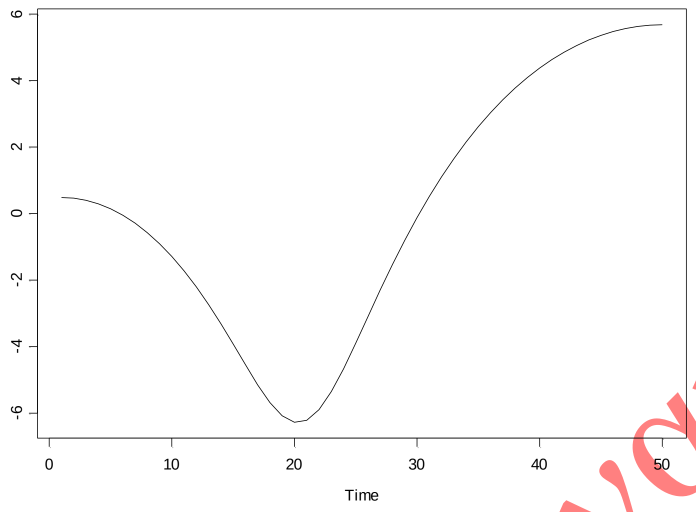


Figure A.11 – Density of the unit root τ_{DF} – test statistic compared with the standard Normal density (dashed curve) under case 4

TRUE SPECTRUMS OF MOVING AVERAGE PROCESSES WITH SPECIFIED PARAMETERS AND ROOTS
T=100 OBSERVATIONS



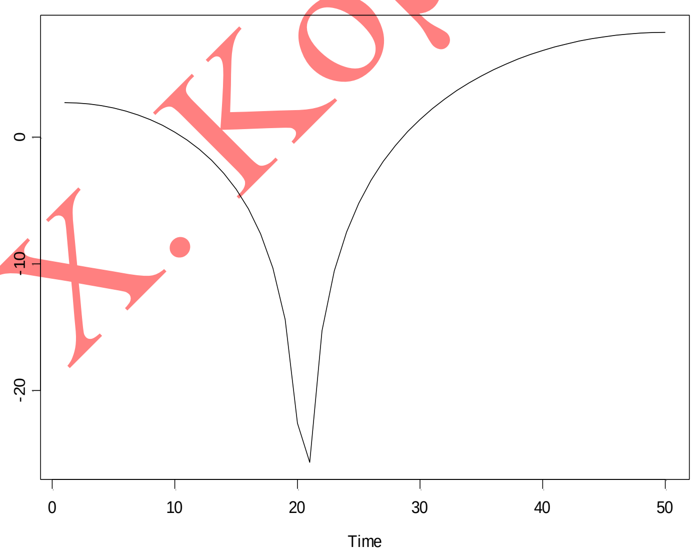
True spectrum of the process u_t where $u_t \sim MA(2)$

with MA parameters:
θ_1 = - 0.43
θ_2 = 0.49

and real MA root:
0.7 at frequency 0.4π

(case 1 in Table B.24)

Figure A.12



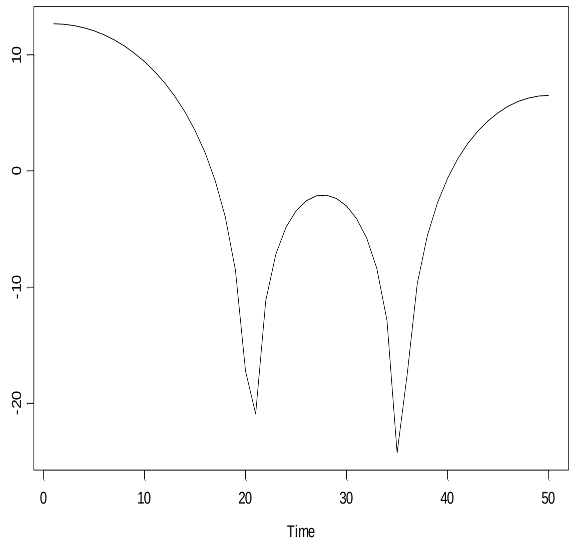
True spectrum of the process u_t where $u_t \sim MA(2)$

with MA parameters:
θ_1 = - 0.61
θ_2 = 0.98

and real MA root:
0.99 at frequency 0.4π

(case 2 in Table B.24)

Figure A.13



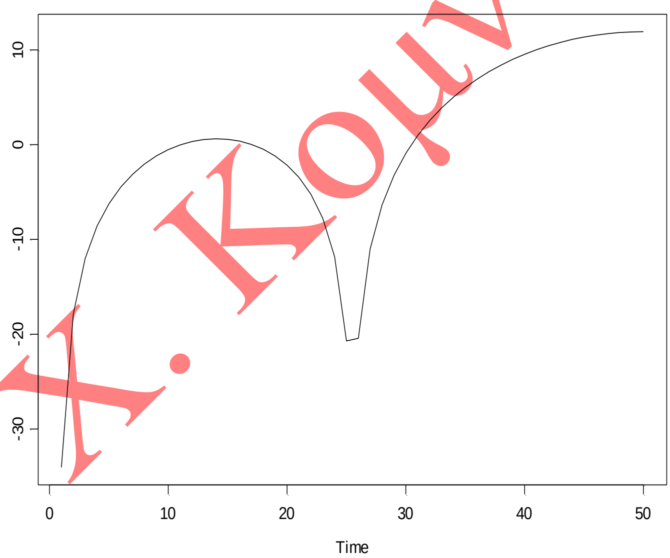
True spectrum of the process u_t where $u_t \sim MA(4)$

with MA parameters:
$\theta_1 = 0.55$
$\theta_2 = 1.25$
$\theta_3 = 0.54$
$\theta_4 = 0.96$

and real MA roots:
0.99 at frequency 0.4π
0.99 at frequency 0.7π

(case 3 in Table B.24)

Figure A.14



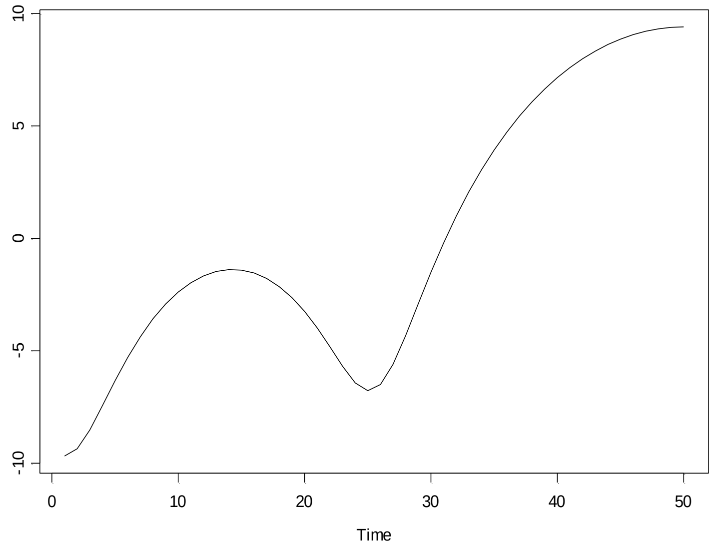
True spectrum of the process u_t where $u_t \sim MA(3)$

with MA parameters:
$\theta_1 = -0.99$
$\theta_2 = 0.98$
$\theta_3 = -0.97$

and real MA roots:
0.99 at frequency 0
0.99 at frequency 0.5π

(case 4 in Table B.24)

Figure A.15



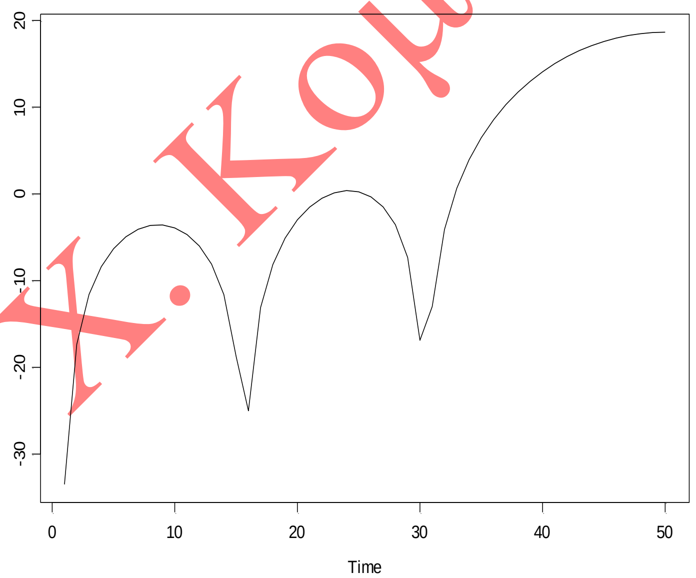
<u>True spectrum</u> of the process u_t where $u_t \sim MA(3)$

with MA parameters:
$\theta_1 = -0.8$
$\theta_2 = 0.64$
$\theta_3 = -0.51$

and real MA roots:
0.8 at frequency 0
0.8 at frequency 0.5π

(case 5 in Table B.24)

Figure A.16



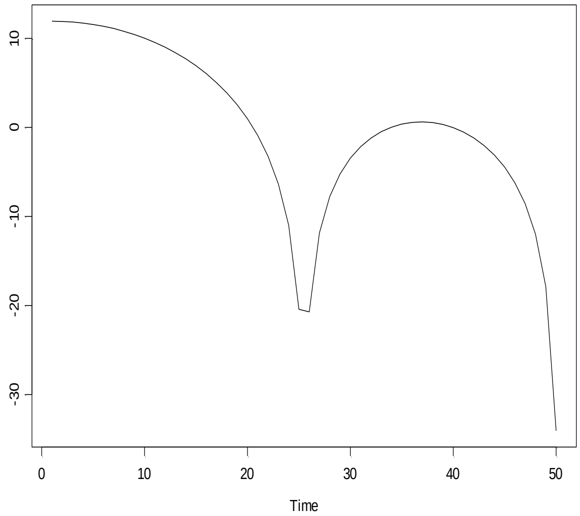
<u>True spectrum</u> of the process u_t where $u_t \sim MA(5)$

with MA parameters:
$\theta_1 = -1.54$
$\theta_2 = 1.79$
$\theta_3 = -1.78$
$\theta_4 = 1.50$
$\theta_5 = -0.95$

and real MA roots:
0.99 at frequency 0
0.99 at frequency 0.3π
0.99 at frequency 0.6π

(case 6 in Table B.24)

Figure A.17



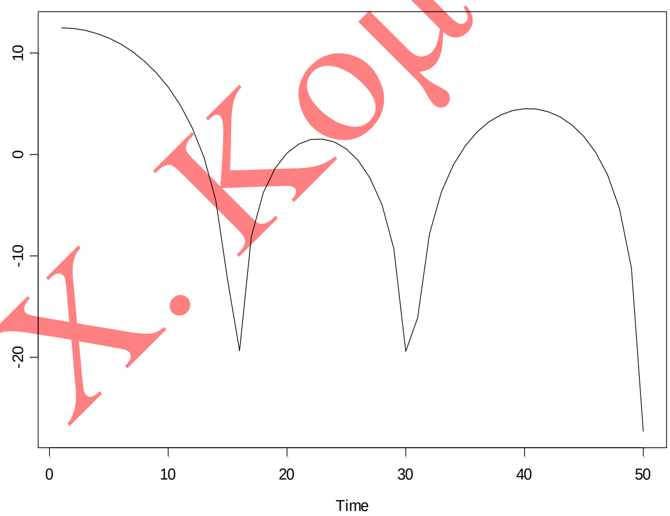
True spectrum of the process u_t where $u_t \sim MA(3)$

with MA parameters:
$\theta_1 = 0.99$
$\theta_2 = 0.98$
$\theta_3 = 0.97$

and real MA roots:
0.99 at frequency π
0.99 at frequency 0.5π

(case 7 in Table B.24)

Figure A.18



True spectrum of the process u_t where $u_t \sim MA(5)$

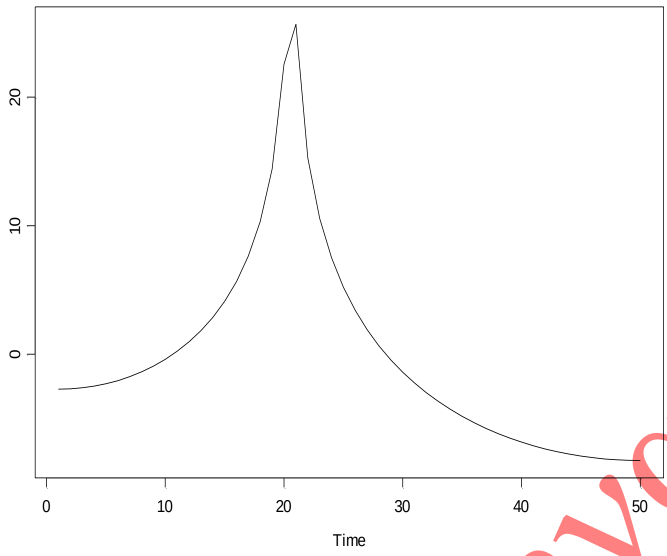
with MA parameters:
$\theta_1 = 0.44$
$\theta_2 = 0.70$
$\theta_3 = 0.69$
$\theta_4 = 0.43$
$\theta_5 = 0.95$

and real MA roots:
0.99 at frequency π
0.99 at frequency 0.3π
0.99 at frequency 0.6π

(case 8 in Table B.24)

Figure A.19

TRUE SPECTRUMS OF AUTOREGRESSIVE PROCESSES WITH SPECIFIED PARAMETERS AND ROOTS T=100 OBSERVATIONS



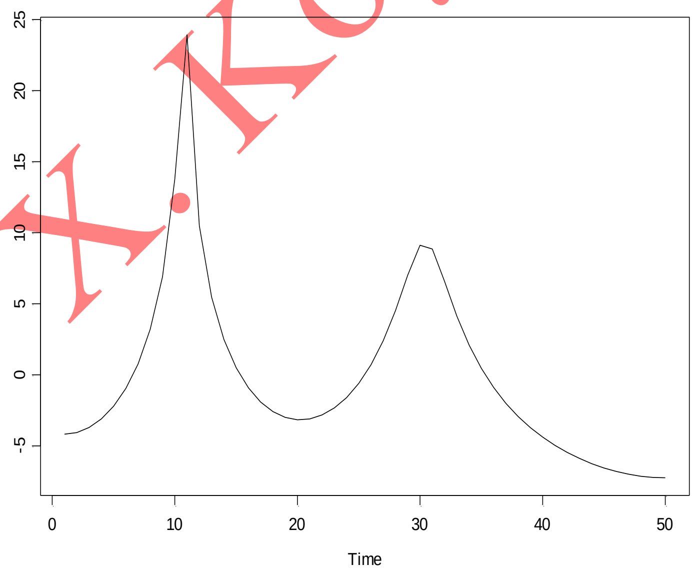
True spectrum of the process u_t where $u_t \sim AR(2)$

with AR parameters: $\varphi_1 = 0.61$ and $\varphi_2 = - 0.98$

and real AR root: 0.99 at frequency 0.4π

(case 1 in Table B.25)

Figure A.20



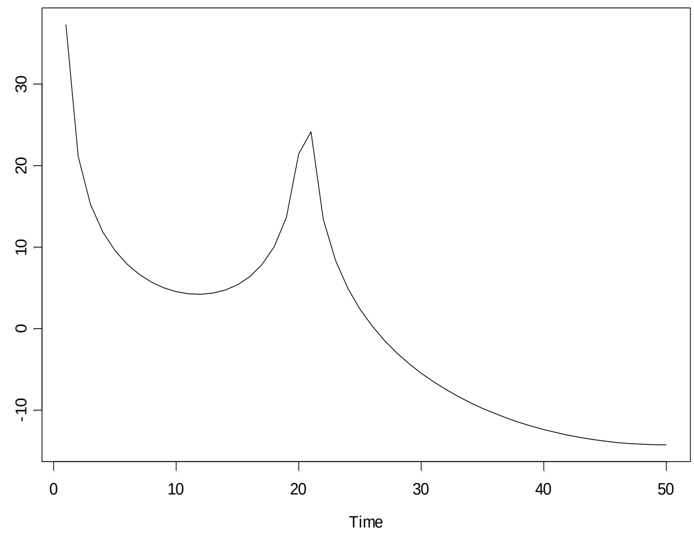
True spectrum of the process u_t where $u_t \sim AR(6)$

with AR parameters:
$\varphi_1 = 0.40$, $\varphi_2 = - 0.38$
$\varphi_3 = 0.34$, $\varphi_4 = - 0.45$
$\varphi_5 = - 0.39$, $\varphi_6 = - 0.13$

and real AR roots:
0.99 at frequency 0.2π
0.90 at frequency 0.6π
0.40 at frequency 0.8π

(case 2 in Table B.25)

Figure A.21



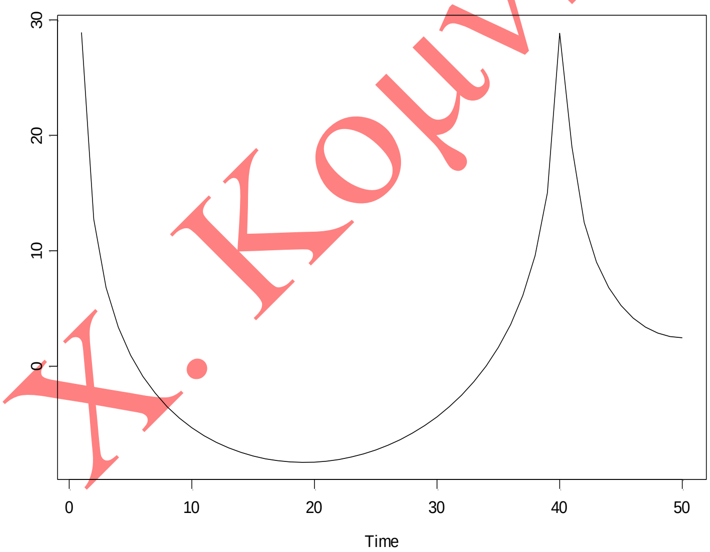
<u>True spectrum</u> of the process u_t where $u_t \sim AR(3)$

with AR parameters:
$\varphi_1 = 1.60$, $\varphi_2 = -1.59$
$\varphi_3 = 0.97$

and real AR roots:
0.99 at frequency 0
0.99 at frequency 0.4π

(case 3 in Table B.25)

Figure A.22



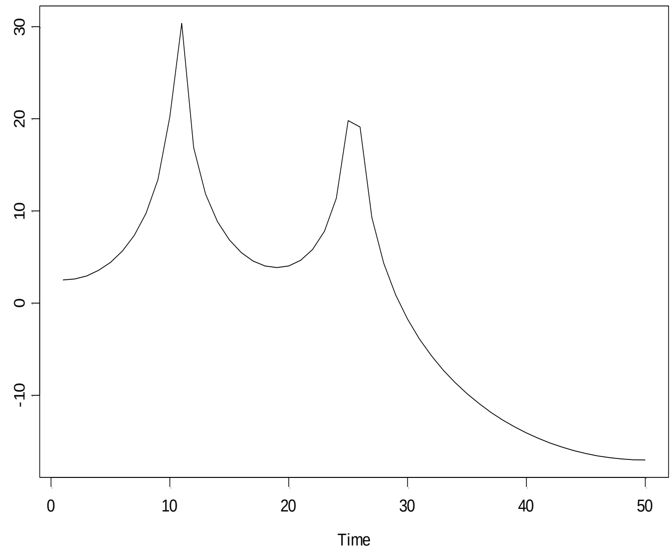
<u>True spectrum</u> of the process u_t where $u_t \sim AR(3)$

with AR parameters:
$\varphi_1 = -0.61$
$\varphi_2 = -0.61$
$\varphi_3 = 0.97$

and real AR roots:
0.99 at frequency 0
0.99 at frequency 0.8π

(case 4 in Table B.25)

Figure A.23



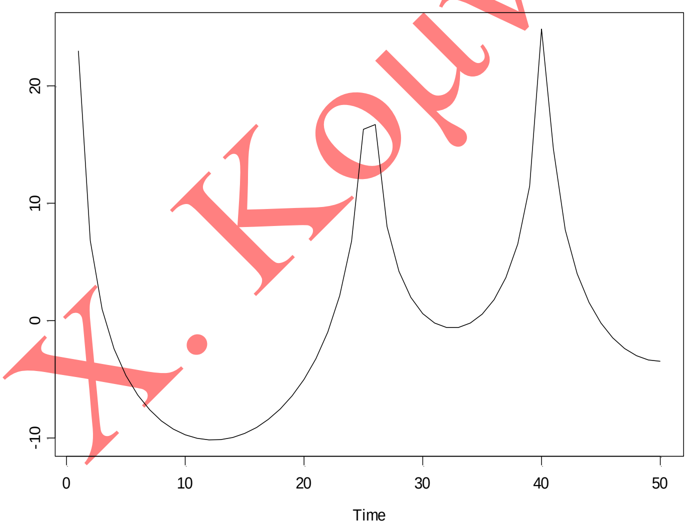
True spectrum of the process u_t where $u_t \sim AR(4)$

with AR parameters:
$\varphi_1 = 1.60$
$\varphi_2 = -1.96$
$\varphi_3 = 1.57$
$\varphi_4 = -0.96$

and real AR roots:
0.99 at frequency 0.2π
0.99 at frequency 0.5π

(case 5 in Table B.25)

Figure A.24



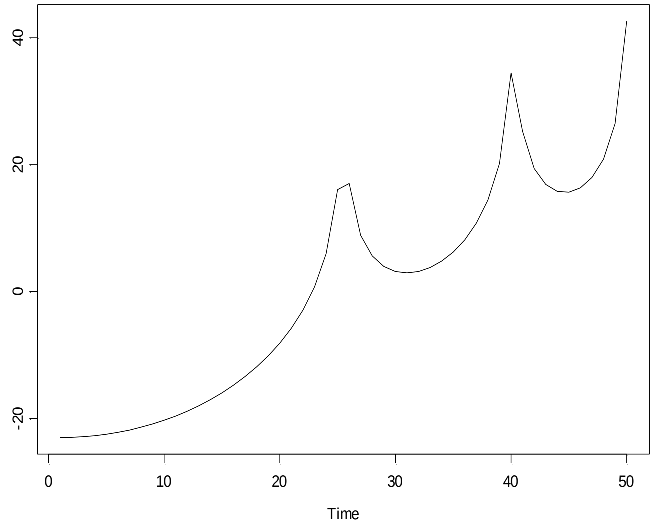
True spectrum of the process u_t where $u_t \sim AR(5)$

with AR parameters:
$\varphi_1 = -0.61$
$\varphi_2 = -0.37$
$\varphi_3 = 0.37$
$\varphi_4 = 0.59$
$\varphi_5 = 0.95$

and real AR roots:
0.99 at frequency 0
0.99 at frequency 0.5π
0.99 at frequency 0.8π

(case 6 in Table B.25)

Figure A.25



<u>True spectrum</u> of the process u_t where $u_t \sim AR(5)$

with AR parameters:

$\varphi_1 = -2.59$
$\varphi_2 = -3.55$
$\varphi_3 = -3.51$
$\varphi_4 = -2.51$
$\varphi_5 = -0.95$

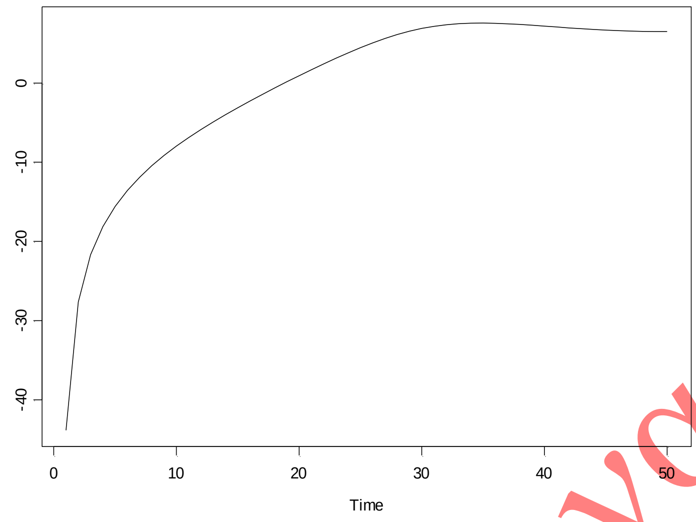
and real AR roots:

0.99 at frequency π
0.99 at frequency 0.5π
0.99 at frequency 0.8π

(case 7 in Table B.25)

Figure A.26

TRUE SPECTRUMS OF AUTOREGRESSIVE MOVING AVERAGE PROCESSES WITH SPECIFIED PARAMETERS AND ROOTS T=100 OBSERVATIONS

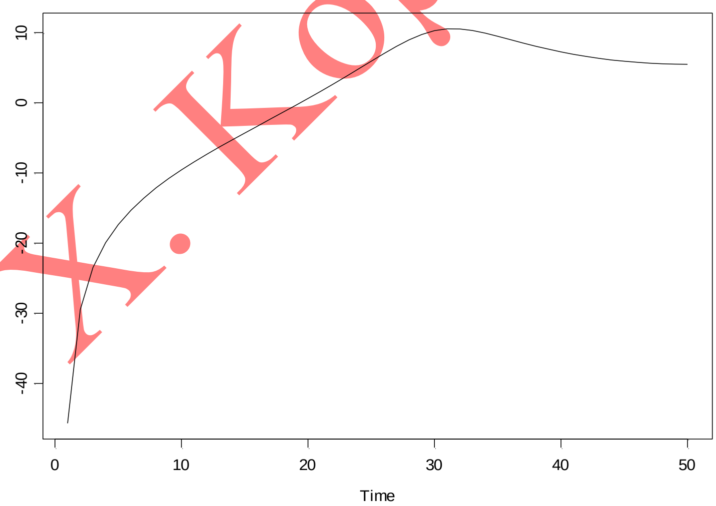


True spectrum of the process u_t where $u_t \sim ARMA(2,1)$

MA parameter:
θ = - 0.99
real MA root:
0.99 at frequency 0
AR parameters:
φ_1 = - 0.31
φ_2 = - 0.25
real AR root:
0.5 at frequency 0.6π

(case 1 in Table B.26)

Figure A.27

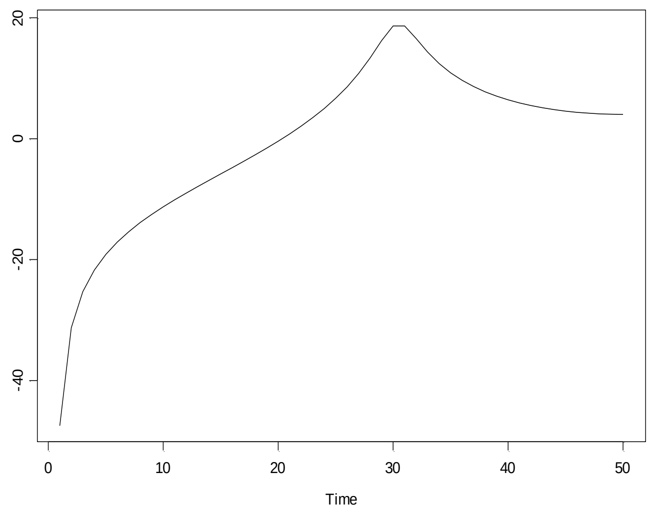


True spectrum of the process u_t where $u_t \sim ARMA(2,1)$

MA parameter:
θ = - 0.99
real MA root:
0.99 at frequency 0
AR parameters:
φ_1 = - 0.43
φ_2 = - 0.49
real AR root:
0.7 at frequency 0.6π

(case 2 in Table B.26)

Figure A.28

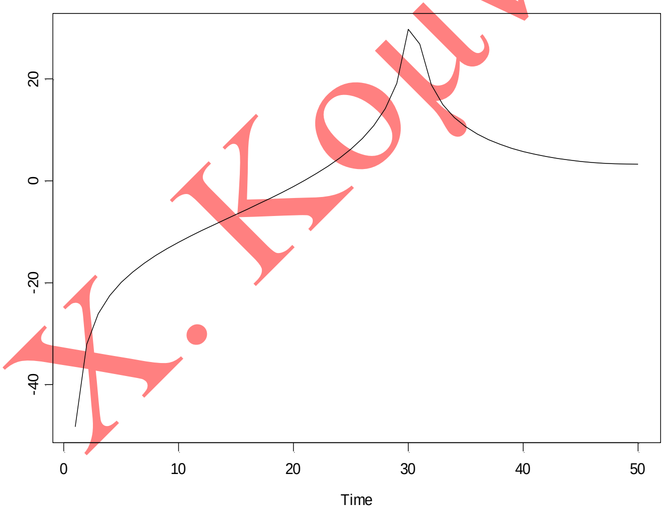


True spectrum of the process u_t where $u_t \sim$ ARMA(2,1)

MA parameter:
θ = - 0.99
real MA root:
0.99 at frequency 0
AR parameters:
φ_1 = - 0.56
φ_2 = - 0.81
real AR root:
0.9 at frequency 0.6π

(case 3 in Table B.26)

Figure A.29



True spectrum of the process u_t where $u_t \sim$ ARMA(2,1)

MA parameter:
θ = - 0.99
real MA root:
0.99 at frequency 0
AR parameters:
φ_1 = - 0.61
φ_2 = - 0.98
real AR root:
0.99 at frequency 0.6π

(case 4 in Table B.26)

Figure A.30

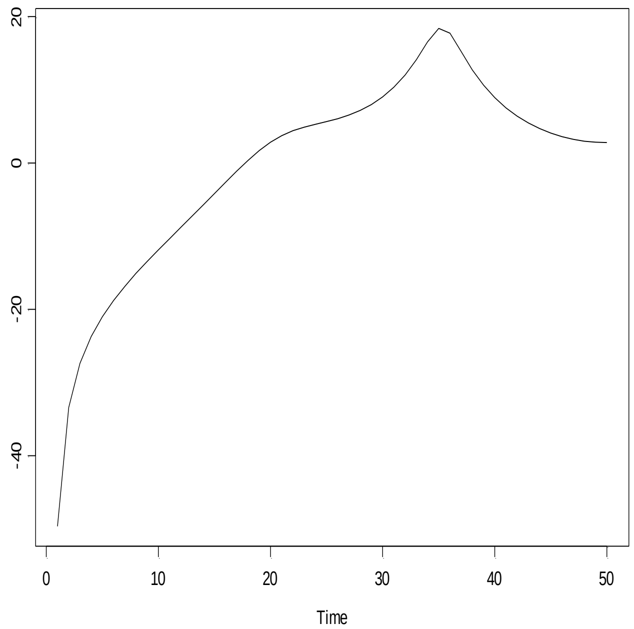


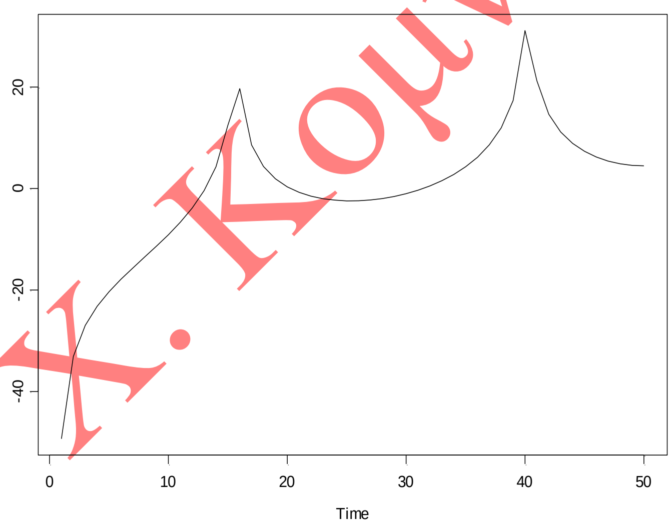
Figure A.31

<u>**True spectrum**</u> **of the process u_t where $u_t \sim ARMA(4,1)$**

MA parameter:
$\theta = -0.99$
real MA root:
0.99 at frequency 0
AR parameters:
$\varphi_1 = -0.63$
$\varphi_2 = -0.84$
$\varphi_3 = -0.17$
$\varphi_4 = -0.40$
real AR roots:
0.7 at frequency 0.4π

0.9 at frequency 0.7π

(case 5 in Table B.26)

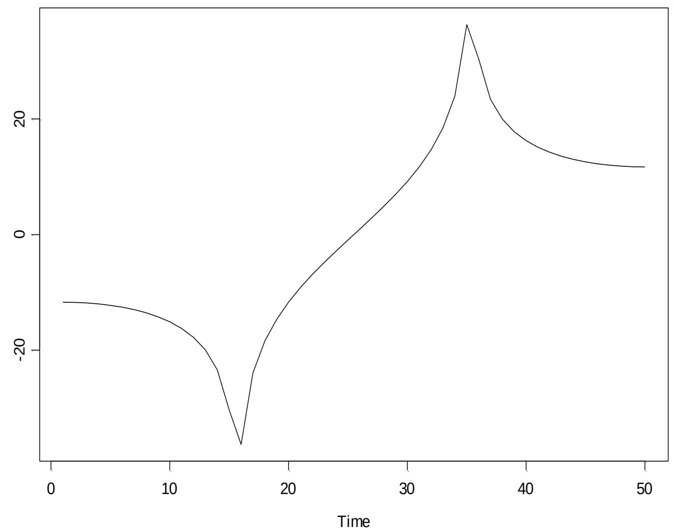


<u>**True spectrum**</u> **of the process u_t where $u_t \sim ARMA(4,1)$**

MA parameter:
$\theta = -0.99$
real MA root:
0.99 at frequency 0
AR parameters:
$\varphi_1 = -0.44$
$\varphi_2 = -0.10$
$\varphi_3 = -0.43$
$\varphi_4 = -0.96$
real AR roots:
0.99 at frequency 0.3π
0.99 at frequency 0.8π

Figure A.32

(case 6 in Table B.26)

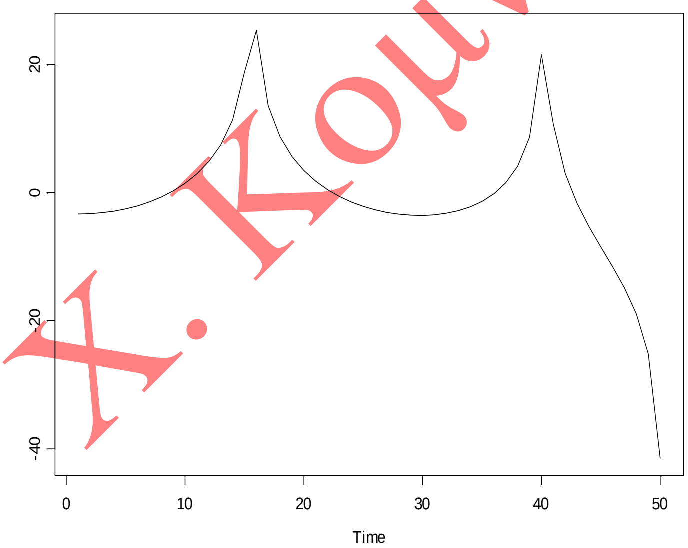


True spectrum of the process u_t where $u_t \sim$ ARMA(2,2)

MA parameters:
θ_1 = - 1.16
θ_2 = 0.98
real MA root:
0.99 at frequency 0.3π
AR parameters:
φ_1 = - 1.16
φ_2 = - 0.98
real AR root:
0.99 at frequency 0.7π

(case 7 in Table B.26)

Figure A.33



True spectrum of the process u_t where $u_t \sim$ ARMA(4,1)

MA parameter:
θ = - 0.99
real MA root:
0.99 at frequency π
AR parameters:
φ_1 = - 0.44
φ_2 = - 0.10
φ_3 = - 0.43
φ_4 = - 0.96
real AR roots:
0.99 at frequency 0.3π
0.99 at frequency 0.8π

(case 8 in Table B.26)

Figure A.34

APPENDIX B

STATISTICAL TABLES

Table B.1 Critical values of the unit root Z_{DF} – test statistic for testing $H_0 : \rho = 1$ vs. $H_1 : \rho < 1$

Sample size T	Case 1				Case 2				Case 4			
	Probability of a smaller value				Probability of a smaller value				Probability of a smaller value			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-11.9	-9.3	-7.3	-5.3	-17.2	-14.6	-12.5	-10.2	-22.5	-19.9	-17.9	-15.6
50	-12.9	-9.9	-7.7	-5.5	-18.9	-15.7	-13.3	-10.7	-25.7	-22.4	-19.8	-16.8
100	-13.3	-10.2	-7.9	-5.6	-19.8	-16.3	-13.7	-11.0	-27.4	-23.6	-20.7	-17.5
250	-13.6	-10.3	-8.0	-5.7	-20.3	-16.6	-14.0	-11.2	-28.4	-24.4	-21.3	-18.0
500	-13.7	-10.4	-8.0	-5.7	-20.5	-16.8	-14.0	-11.2	-28.9	-24.8	-21.5	-18.1
∞	-13.8	-10.5	-8.1	-5.7	-20.7	-16.9	-14.1	-11.3	-29.5	-25.1	-21.8	-18.3

Source : Fuller (1976) , Table (8.5.1)

Table B.2	Critical values of the unit root τ_{DF} – test statistic for testing $H_0 : \rho = 1$ vs. $H_1 : \rho < 1$											
Sample size T	Case 1				Case 2				Case 4			
	Probability of a smaller value				Probability of a smaller value				Probability of a smaller value			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-2.66	-2.26	-1.95	-1.60	-3.75	-3.33	-3.00	-2.63	-4.38	-3.95	-3.60	-3.24
50	-2.62	-2.25	-1.95	-1.61	-3.58	-3.22	-2.93	-2.60	-4.15	-3.80	-3.50	-3.18
100	-2.60	-2.24	-1.95	-1.61	-3.51	-3.17	-2.89	-2.58	-4.04	-3.73	-3.45	-3.15
250	-2.58	-2.23	-1.95	-1.62	-3.46	-3.14	-2.88	-2.57	-3.99	-3.69	-3.43	-3.13
∞	-2.58	-2.23	-1.95	-1.62	-3.43	-3.12	-2.86	-2.57	-3.96	-3.66	-3.41	-3.12

Source : Fuller (1976) , Table (8.5.2)

Table B.3		Size and Power of the unit root τ_{DF} – test statistic , Moving Average Case , T=100 Specified nominal size $\alpha = 0.05$ (5,000 simulations)									
ρ	**θ**	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**k=6**	**k=7**	**k=8**	**k=9**	**k=10**
1.0	0.8	0.123	0.031	0.075	0.037	0.062	0.041	0.062	0.041	0.051	0.041
1.0	0.5	0.103	0.047	0.065	0.052	0.057	0.054	0.053	0.053	0.051	0.048
1.0	0.3	0.073	0.051	0.055	0.056	0.053	0.047	0.050	0.048	0.045	0.046
1.0	0	0.049	0.048	0.046	0.044	0.046	0.044	0.045	0.044	0.044	0.041
1.0	-0.3	0.091	0.062	0.057	0.052	0.052	0.048	0.049	0.048	0.048	0.045
1.0	-0.5	0.214	0.099	0.068	0.055	0.051	0.051	0.052	0.050	0.049	0.048
1.0	-0.8	0.880	0.640	0.434	0.283	0.200	0.132	0.110	0.082	0.074	0.060
0.95	0.8	0.307	0.044	0.176	0.059	0.129	0.064	0.106	0.067	0.085	0.064
0.95	0.5	0.237	0.074	0.130	0.089	0.103	0.086	0.086	0.079	0.077	0.072
0.95	0.3	0.162	0.088	0.101	0.092	0.092	0.085	0.087	0.081	0.075	0.069
0.95	0	0.117	0.111	0.108	0.102	0.099	0.093	0.083	0.082	0.080	0.068
0.95	-0.3	0.219	0.139	0.122	0.108	0.100	0.091	0.094	0.089	0.087	0.081
0.95	-0.5	0.477	0.246	0.157	0.115	0.104	0.087	0.085	0.081	0.078	0.069
0.95	-0.8	0.997	0.941	0.782	0.584	0.444	0.329	0.255	0.203	0.164	0.130
0.85	0.8	0.788	0.201	0.524	0.212	0.379	0.203	0.278	0.166	0.207	0.149
0.85	0.5	0.708	0.316	0.425	0.297	0.308	0.252	0.239	0.203	0.191	0.169
0.85	0.3	0.598	0.393	0.395	0.334	0.306	0.267	0.241	0.214	0.196	0.171
0.85	0	0.510	0.436	0.399	0.343	0.316	0.271	0.251	0.218	0.193	0.166
0.85	-0.3	0.746	0.540	0.452	0.376	0.344	0.287	0.266	0.233	0.209	0.176
0.85	-0.5	0.961	0.779	0.614	0.482	0.423	0.353	0.315	0.275	0.242	0.208
0.85	-0.8	1.000	1.000	0.996	0.956	0.886	0.759	0.671	0.565	0.483	0.399

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta \varepsilon_{t-1}$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995) , Table (1)

Table B.4

Size and Power of the unit root τ_{DF} – test statistic , Autoregressive Case, T=100
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	φ_1	φ_2	φ_3	φ_4	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
1.0	0.6	0.0	0.0	0.0	0.058	0.055	0.057	0.053	0.053	0.053	0.058	0.055	0.055	0.052
1.0	-0.6	0.0	0.0	0.0	0.054	0.054	0.054	0.052	0.052	0.047	0.045	0.043	0.044	0.043
1.0	0.4	0.2	0.0	0.0	0.033	0.052	0.052	0.051	0.050	0.052	0.050	0.048	0.047	0.046
1.0	0.3	0.3	0.25	0.14	0.062	0.034	0.038	0.051	0.054	0.052	0.050	0.047	0.048	0.052
0.95	0.6	0.0	0.0	0.0	0.391	0.346	0.320	0.281	0.260	0.225	0.210	0.191	0.178	0.154
0.95	-0.6	0.0	0.0	0.0	0.078	0.075	0.073	0.068	0.070	0.061	0.061	0.058	0.056	0.055
0.95	0.4	0.2	0.0	0.0	0.125	0.354	0.328	0.288	0.275	0.237	0.225	0.204	0.194	0.165
0.95	0.3	0.3	0.25	0.14	0.137	0.650	0.865	0.903	0.876	0.814	0.763	0.675	0.608	0.522
0.85	0.6	0.0	0.0	0.0	0.976	0.938	0.883	0.805	0.720	0.626	0.560	0.479	0.435	0.357
0.85	-0.6	0.0	0.0	0.0	0.252	0.224	0.207	0.188	0.171	0.159	0.154	0.134	0.129	0.122
0.85	0.4	0.2	0.0	0.0	0.818	0.933	0.889	0.799	0.718	0.626	0.560	0.479	0.435	0.357
0.85	0.3	0.3	0.25	0.14	0.688	0.664	0.991	0.993	0.984	0.960	0.908	0.827	0.748	0.647

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source: Ng & Perron (1995), Table (2)

Table B.5	Frequency Count of the Selected Lag Lengths k , Moving Average Case , T = 100 (5,000 simulations)									
Method	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
					θ=0.8					
$t_{sig}(10)$	0.01	0.028	0.115	0.131	0.149	0.113	0.135	0.104	0.121	0.102
$t_{sig}(5)$	0.013	0.103	0.210	0.157	0.147	0.092	0.097	0.063	0.065	0.053
AIC	0.016	0.149	0.325	0.219	0.150	0.069	0.039	0.020	0.009	0.005
BIC	0.151	0.381	0.326	0.100	0.035	0.004	0.001	0.001	0.000	0.000
					θ=0.3					
$t_{sig}(10)$	0.304	0.081	0.067	0.055	0.059	0.062	0.074	0.079	0.086	0.086
$t_{sig}(5)$	0.468	0.071	0.047	0.040	0.038	0.039	0.043	0.031	0.049	0.039
AIC	0.676	0.131	0.037	0.013	0.007	0.002	0.002	0.001	0.000	0.000
BIC	0.611	0.037	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
					θ= - 0.5					
$t_{sig}(10)$	0.220	0.189	0.075	0.061	0.058	0.068	0.069	0.077	0.077	0.083
$t_{sig}(5)$	0.387	0.212	0.054	0.041	0.039	0.041	0.041	0.041	0.041	0.039
AIC	0.484	0.331	0.073	0.025	0.011	0.007	0.003	0.001	0.000	0.000
BIC	0.621	0.169	0.016	0.003	0.000	0.000	0.000	0.000	0.000	0.000
					θ= - 0.8					
$t_{sig}(10)$	0.074	0.113	0.109	0.116	0.076	0.097	0.075	0.092	0.080	0.085
$t_{sig}(5)$	0.123	0.162	0.121	0.100	0.061	0.065	0.045	0.047	0.045	0.041
AIC	0.197	0.225	0.146	0.091	0.033	0.019	0.008	0.003	0.002	0.001
BIC	0.264	0.172	0.056	0.019	0.002	0.000	0.000	0.000	0.000	0.000

DGP : $y_t = y_{t-1} + u_t$, $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ **Regression :** $\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995) , Table (4)

Table B.6	Frequency Count of the Selected Lag Lengths k, Autoregressive Case, T = 100 (5,000 simulations)									
Method	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
φ_1=0.6, φ_2=0.0, φ_3=0.0, φ_4=0.0										
$t_{sig}(10)$	0.407	0.045	0.047	0.051	0.057	0.064	0.071	0.078	0.091	0.089
$t_{sig}(5)$	0.644	0.037	0.033	0.033	0.037	0.042	0.040	0.043	0.047	0.044
AIC	0.878	0.075	0.027	0.010	0.004	0.002	0.001	0.001	0.001	0.000
BIC	0.976	0.020	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
φ_1= - 0.6, φ_2=0.0, φ_3=0.0, φ_4=0.0										
$t_{sig}(10)$	0.407	0.049	0.051	0.050	0.060	0.069	0.068	0.076	0.081	0.090
$t_{sig}(5)$	0.652	0.036	0.037	0.035	0.041	0.040	0.036	0.040	0.040	0.044
AIC	0.866	0.086	0.027	0.011	0.006	0.003	0.000	0.000	0.000	0.000
BIC	0.978	0.018	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
φ_1=0.4, φ_2=0.2, φ_3=0.0, φ_4=0.0										
$t_{sig}(10)$	0.200	0.242	0.055	0.057	0.060	0.058	0.073	0.070	0.092	0.089
$t_{sig}(5)$	0.391	0.285	0.038	0.039	0.038	0.038	0.043	0.035	0.042	0.044
AIC	0.455	0.443	0.057	0.021	0.008	0.003	0.002	0.001	0.001	0.000
BIC	0.680	0.277	0.008	0.001	0.000	0.000	0.000	0.000	0.000	0.000
φ_1=0.3, φ_2=0.3, φ_3=0.25, φ_4=0.14										
$t_{sig}(10)$	0.008	0.095	0.299	0.149	0.059	0.063	0.067	0.078	0.084	0.092
$t_{sig}(5)$	0.023	0.192	0.388	0.137	0.034	0.037	0.042	0.047	0.043	0.046
AIC	0.026	0.196	0.494	0.204	0.035	0.016	0.007	0.003	0.001	0.001
BIC	0.082	0.369	0.419	0.076	0.005	0.002	0.000	0.000	0.000	0.000

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995), Table (5)

Table B.7	Size and Power of the unit root τ_{DF} – test statistic Moving Average Case $T = 100$, $k_{max} = 10$ Specified nominal size $\alpha = 0.05$ (5,000 simulations)				
ρ	θ	$t_{sig}(10)$	$t_{sig}(5)$	AIC	BIC
1.0	0.8	0.069	0.073	0.068	0.071
1.0	0.5	0.083	0.087	0.082	0.088
1.0	0.3	0.075	0.077	0.070	0.069
1.0	0	0.063	0.059	0.052	0.046
1.0	- 0.3	0.097	0.126	0.127	0.174
1.0	- 0.5	0.116	0.158	0.167	0.244
1.0	- 0.8	0.304	0.424	0.561	0.733
0.95	0.8	0.136	0.151	0.146	0.158
0.95	0.5	0.164	0.184	0.170	0.196
0.95	0.3	0.158	0.162	0.152	0.144
0.95	0	0.153	0.151	0.140	0.126
0.95	- 0.3	0.228	0.292	0.294	0.393
0.95	- 0.5	0.254	0.336	0.377	0.510
0.95	- 0.8	0.534	0.704	0.877	0.963
0.85	0.8	0.347	0.387	0.405	0.451
0.85	0.5	0.445	0.510	0.520	0.586
0.85	0.3	0.465	0.513	0.536	0.505
0.85	0	0.486	0.540	0.580	0.575
0.85	- 0.3	0.555	0.682	0.758	0.859
0.85	- 0.5	0.627	0.753	0.860	0.936
0.85	- 0.8	0.825	0.908	0.996	1.000

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta \varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995) , Table (6)

Table B.8	Size and Power of the unit root τ_{DF} – test statistic Autoregressive case $T = 100$, $k_{max} = 10$ Specified nominal size $\alpha = 0.05$ (5,000 simulations)							
ρ	φ_1	φ_2	φ_3	φ_4	$t_{sig}(10)$	$t_{sig}(5)$	AIC	BIC
1.0	0.6	0.0	0.0	0.0	0.078	0.075	0.066	0.060
1.0	- 0.6	0.0	0.0	0.0	0.068	0.066	0.066	0.060
1.0	0.4	0.2	0.0	0.0	0.066	0.062	0.055	0.047
1.0	0.3	0.3	0.25	0.24	0.066	0.062	0.058	0.052
0.95	0.6	0.0	0.0	0.0	0.371	0.399	0.404	0.394
0.95	- 0.6	0.0	0.0	0.0	0.101	0.099	0.087	0.080
0.95	0.4	0.2	0.0	0.0	0.346	0.336	0.338	0.267
0.95	0.3	0.3	0.25	0.24	0.822	0.840	0.886	0.837
0.85	0.6	0.0	0.0	0.0	0.782	0.870	0.960	0.972
0.85	- 0.6	0.0	0.0	0.0	0.269	0.274	0.268	0.256
0.85	0.4	0.2	0.0	0.0	0.763	0.824	0.899	0.867
0.85	0.3	0.3	0.25	0.24	0.901	0.937	0.976	0.947

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995) , Table (7)

Table B.9	Size of the unit root τ_{DF} – test statistic Moving Average Case $T = 200$, $k_{max} = 12$ Specified nominal size $\alpha = 0.05$ (5,000 simulations)				
ρ	**θ**	**$t_{sig}(10)$**	**$t_{sig}(5)$**	**AIC**	**BIC**
1.0	0.8	0.056	0.060	0.059	0.063
1.0	0.5	0.061	0.064	0.056	0.064
1.0	0.3	0.061	0.064	0.061	0.066
1.0	0	0.064	0.066	0.059	0.057
1.0	- 0.3	0.067	0.076	0.076	0.102
1.0	- 0.5	0.085	0.110	0.121	0.168
1.0	- 0.8	0.177	0.250	0.366	0.557

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995) , Table (6)

Table B.10

Size and Power of the unit root τ_{DF} – test statistic, Autoregressive Case, T = 200, k_{max} = 12
Specified nominal size α = 0.05
(5,000 simulations)

ρ	φ_1	φ_2	φ_3	φ_4	$t_{sig}(10)$	$t_{sig}(5)$	AIC	BIC
1.0	0.6	0.0	0.0	0.0	0.063	0.060	0.057	0.054
1.0	- 0.6	0.0	0.0	0.0	0.063	0.064	0.058	0.056
1.0	0.4	0.2	0.0	0.0	0.062	0.061	0.056	0.048
1.0	0.3	0.3	0.25	0.14	0.076	0.072	0.070	0.059
0.95	0.6	0.0	0.0	0.0	0.738	0.815	0.897	0.908
0.95	- 0.6	0.0	0.0	0.0	0.166	0.168	0.160	0.153
0.95	0.4	0.2	0.0	0.0	0.712	0.709	0.837	0.784
0.95	0.3	0.3	0.25	0.14	0.979	0.988	1.000	0.998
0.85	0.6	0.0	0.0	0.0	0.955	0.974	1.000	1.000
0.85	- 0.6	0.0	0.0	0.0	0.608	0.603	0.738	0.749
0.85	0.4	0.2	0.0	0.0	0.954	0.975	1.000	1.000
0.85	0.3	0.3	0.25	0.14	0.994	0.997	1.000	1.000

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Source : Ng & Perron (1995) , Table (9)

Table B.11	Frequency Count of Selected Lag Lengths k, Moving Average Case, T = 100 (5,000 simulations)									
Method	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
					θ=0.8					
AICC	0.0016	0.051	0.1682	0.2122	0.1816	0.1362	0.0956	0.065	0.0492	0.0394
BICC	0.0704	0.3332	0.34	0.16	0.0682	0.0206	0.0056	0.0016	0.0002	0.0002
MAICC	0.0028	0.0666	0.1704	0.2364	0.157	0.1462	0.0736	0.00684	0.0408	0.0378
MBICC	0.08	0.396	0.2566	0.1816	0.0456	0.0296	0.006	0.0028	0.0008	0.001
					θ=0.5					
AICC	0.2112	0.3844	0.1802	0.085	0.0528	0.0282	0.0234	0.013	0.0112	0.0106
BICC	0.5894	0.3402	0.0578	0.0102	0.0016	0.0004	0	0.0002	0.0002	0
MAICC	0.2068	0.4134	0.1602	0.0954	0.0482	0.028	0.018	0.0138	0.0074	0.0088
MBICC	0.511	0.4234	0.0426	0.0166	0.0022	0.0026	0.001	0.0004	0	0.0002
					θ= 0.3					
AICC	0.5892	0.2012	0.0776	0.0424	0.0292	0.02	0.011	0.0116	0.0102	0.0076
BICC	0.894	0.0924	0.0096	0.0032	0.0006	0.0002	0	0	0	0
MAICC	0.5824	0.1958	0.0938	0.0432	0.0262	0.0188	0.0122	0.0114	0.0076	0.0086
MBICC	0.8188	0.1518	0.0168	0.0082	0.0018	0.0018	0.0006	0	0	0.0002
					θ= - 0.3					
AICC	0.613	0.1948	0.0652	0.0458	0.0224	0.0184	0.012	0.0122	0.007	0.0092
BICC	0.8974	0.0892	0.0104	0.0028	0.0002	0	0	0	0	0
MAICC	0.5786	0.2206	0.072	0.0484	0.0224	0.0194	0.0106	0.013	0.0068	0.0082
MBICC	0.817	0.139	0.029	0.0078	0.004	0.0014	0.0006	0.0006	0.0004	0.0002
					θ= - 0.5					
AICC	0.2702	0.3732	0.1586	0.0756	0.0426	0.0286	0.0174	0.0142	0.0086	0.011
BICC	0.6298	0.3106	0.0482	0.0102	0.0012	0	0	0	0	0
MAICC	0.191	0.393	0.1796	0.0938	0.0496	0.0322	0.0206	0.0168	0.0118	0.0116
MBICC	0.4208	0.4006	0.1134	0.0416	0.0124	0.006	0	0	0.0006	0.0006
					θ= - 0.8					
AICC	0.0666	0.1422	0.1756	0.1984	0.1392	0.1084	0.0592	0.0488	0.0314	0.0302
BICC	0.2846	0.3566	0.21	0.0984	0.0324	0.0126	0.0044	0.0008	0.0002	0
MAICC	0.0016	0.0332	0.1054	0.1966	0.1688	0.1622	0.0998	0.0938	0.0676	0.071
MBICC	0.0092	0.1316	0.2174	0.2334	0.1468	0.1106	0.0532	0.044	0.0272	0.0266

DGP : $y_t = y_{t-1} + u_t \;,\; u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ **Regression :** $\Delta y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.12	Frequency Count of Selected Lag Lengths k , Autoregressive Case , T = 100 (5,000 simulations)									
Method	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**k=6**	**k=7**	**k=8**	**k=9**	**k=10**
	φ_1=0.6 , φ_2=0.0 , φ_3=0.0 , φ_4=0.0									
AICC	0.723	0.116	0.056	0.041	0.015	0.016	0.007	0.0011	0.008	0.007
BICC	0.952	0.038	0.007	0.002	0	0.001	0	0	0	0
MAICC	0.702	0.127	0.062	0.041	0.02	0.02	0.009	0.008	0.007	0.007
MBICC	0.917	0.061	0.011	0.008	0.002	0	0	0	0.001	0
	φ_1= - 0.6 , φ_2=0.0 , φ_3=0.0 , φ_4=0.0									
AICC	0.732	0.101	0.05	0.036	0.027	0.014	0.011	0.01	0.01	0.009
BICC	0.953	0.039	0.003	0.005	0	0	0	0	0	0
MAICC	0.727	0.101	0.05	0.042	0.022	0.016	0.013	0.012	0.01	0.007
MBICC	0.923	0.051	0.012	0.009	0.001	0.003	0	0	0	0.0010
	φ_1=0.4 , φ_2=0.2 , φ_3=0.0 , φ_4=0.0									
AICC	0.274	0.456	0.109	0.063	0.031	0.017	0.013	0.013	0.012	0.012
BICC	0.621	0.348	0.023	0.005	0.001	0.002	0	0	0	0
MAICC	0.313	0.431	0.098	0.063	0.033	0.017	0.013	0.012	0.011	0.009
MBICC	0.674	0.289	0.028	0.006	0.001	0.001	0	0	0	0.001
	φ_1=0.3 , φ_2=0.3 , φ_3=0.25 , φ_4=0.14									
AICC	0.705	0.127	0.063	0.034	0.021	0.019	0.012	0.008	0.003	0.008
BICC	0.964	0.027	0.007	0.002	0	0	0	0	0	0
MAICC	0.712	0.113	0.069	0.037	0.02	0.017	0.012	0.008	0.006	0.006
MBICC	0.914	0.063	0.017	0.003	0.001	0.001	0.001	0	0	0

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.13	Size of the unit root τ_{DF} – test statistic Moving Average Case, $T = 100$ $k_{max} = 10$, Specified nominal size $\alpha = 0.05$ (5,000 simulations)				
ρ	**θ**	**AICC**	**BICC**	**MAICC**	**MBICC**
1.0	0.8	0.0586	0.0638	0.0276	0.0216
1.0	0.5	0.0612	0.0646	0.0316	0.0382
1.0	0.3	0.0598	0.0642	0.0364	0.0442
1.0	- 0.3	0.0664	0.0754	0.0402	0.0504
1.0	- 0.5	0.1022	0.1442	0.055	0.0688
1.0	- 0.8	0.229	0.4078	0.083	0.0964

DGP : $y_t = \rho y_{t-1} + u_t \;,\; u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.14	Size of the unit root τ_{DF} – test statistic Moving Average Case , T = 250 $k_{max} = 12$, Specified nominal size $\alpha = 0.05$ (5,000 simulations)				
ρ	**θ**	**AICC**	**BICC**	**MAICC**	**MBICC**
1.0	0.8	0.0502	0.0526	0.0336	0.024
1.0	0.5	0.0498	0.0534	0.0322	0.0326
1.0	0.3	0.0474	0.056	0.0354	0.0418
1.0	- 0.3	0.0624	0.079	0.045	0.0574
1.0	- 0.5	0.0756	0.1162	0.0466	0.0648
1.0	- 0.8	0.1678	0.3282	0.0778	0.1054

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.15	Size of the unit root τ_{DF} – test statistic Moving Average Case, $T = 500$ $k_{max} = 14$, Specified nominal size $\alpha = 0.05$ (5,000 simulations)				
p	**θ**	**AICC**	**BICC**	**MAICC**	**MBICC**
1.0	0.8	0.052	0.0516	0.044	0.0356
1.0	0.5	0.0572	0.0544	0.0506	0.0428
1.0	0.3	0.0504	0.0588	0.044	0.049
1.0	- 0.3	0.058	0.071	0.05	0.0588
1.0	- 0.5	0.0608	0.092	0.0514	0.0676
1.0	- 0.8	0.1134	0.2358	0.0787	0.1154

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.16

Size and Power of the unit root τ_{DF} – test statistic , Autoregressive Case , T = 100 , k_{max} = 10
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	φ_1	φ_2	φ_3	φ_4	AICC	BICC	MAICC	MBICC
1.0	0.6	0.0	0.0	0.0	0.064	0.06	0.038	0.049
1.0	- 0.6	0.0	0.0	0.0	0.035	0.036	0.032	0.034
1.0	0.4	0.2	0.0	0.0	0.052	0.043	0.036	0.039
1.0	0.3	0.3	0.25	0.14	0.062	0.047	0.039	0.037
0.95	0.6	0.0	0.0	0.0	0.33	0.324	0.256	0.278
0.95	- 0.6	0.0	0.0	0.0	0.331	0.325	0.254	0.281
0.95	0.4	0.2	0.0	0.0	0.281	0.244	0.196	0.173
0.95	0.3	0.3	0.25	0.14	0.13	0.135	0.097	0.088
0.85	0.6	0.0	0.0	0.0	0.805	0.84	0.647	0.733
0.85	- 0.6	0.0	0.0	0.0	0.916	0.957	0.735	0.812
0.85	0.4	0.2	0.0	0.0	0.736	0.706	0.576	0.603
0.85	0.3	0.3	0.25	0.14	0.099	0.121	0.075	0.076

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.17 — **Size and Power of the unit root τ_{DF} – test statistic , Autoregressive Case , $T = 250$, $k_{max} = 12$**
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	φ_1	φ_2	φ_3	φ_4	AICC	BICC	MAICC	MBICC
1.0	0.6	0.0	0.0	0.0	0.057	0.059	0.048	0.056
1.0	- 0.6	0.0	0.0	0.0	0.068	0.065	0.054	0.063
1.0	0.4	0.2	0.0	0.0	0.058	0.055	0.044	0.042
1.0	0.3	0.3	0.25	0.14	0.043	0.038	0.036	0.035
0.95	0.6	0.0	0.0	0.0	0.833	0.855	0.771	0.832
0.95	- 0.6	0.0	0.0	0.0	0.869	0.881	0.785	0.854
0.95	0.4	0.2	0.0	0.0	0.798	0.779	0.73	0.602
0.95	0.3	0.3	0.25	0.14	0.22	0.207	0.208	0.185
0.85	0.6	0.0	0.0	0.0	0.997	1.000	0.985	0.993
0.85	- 0.6	0.0	0.0	0.0	0.999	1.000	0.989	0.994
0.85	0.4	0.2	0.0	0.0	0.996	0.999	0.988	0.995
0.85	0.3	0.3	0.25	0.14	0.217	0.212	0.206	0.185

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.18

Size of the unit root τ_{DF} – test statistic , Autoregressive Case , T = 500 , k_{max} = 14
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	φ1	φ2	φ3	φ4	AICC	BICC	MAICC	MBICC
1.0	0.6	0.0	0.0	0.0	0.055	0.051	0.047	0.05
1.0	- 0.6	0.0	0.0	0.0	0.05	0.046	0.044	0.044
1.0	0.4	0.2	0.0	0.0	0.053	0.052	0.048	0.042
1.0	0.3	0.3	0.25	0.14	0.054	0.051	0.047	0.041

DGP : $y_t = \rho y_{t-1} + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \varepsilon_t$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.19	Size and Power of the unit root τ_{DF} – test statistic Autoregressive Case , T = 100 , k_{max} = 10 Specified nominal size α = 0.05 (5,000 simulations)				
ρ	φ	AICC	BICC	MAICC	MBICC
1.0	0.8	0.05	0.04	0.038	0.039
1.0	0.5	0.054	0.047	0.036	0.038
1.0	0.3	0.066	0.061	0.045	0.051
1.0	- 0.3	0.065	0.058	0.037	0.042
1.0	- 0.5	0.049	0.047	0.035	0.037
1.0	- 0.8	0.061	0.052	0.037	0.041
0.95	0.8	0.285	0.278	0.209	0.24
0.95	0.5	0.317	0.313	0.248	0.278
0.95	0.3	0.345	0.343	0.263	0.297
0.95	- 0.3	0.355	0.355	0.279	0.305
0.95	- 0.5	0.353	0.353	0.272	0.302
0.95	- 0.8	0.346	0.353	0.241	0.283
0.85	0.8	0.649	0.693	0.544	0.594
0.85	0.5	0.817	0.868	0.67	0.754
0.85	0.3	0.878	0.921	0.711	0.81
0.85	- 0.3	0.897	0.958	0.734	0.805
0.85	- 0.5	0.917	0.958	0.751	0.815
0.85	- 0.8	0.925	0.961	0.704	0.757

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varphi u_{t-1} + \varepsilon_t$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.20	Size and Power of the unit root τ_{DF} – test statistic Autoregressive Case , T = 250 , k_{max} = 12 Specified nominal size α = 0.05 (5,000 simulations)				
ρ	**φ**	**AICC**	**BICC**	**MAICC**	**MBICC**
1.0	0.8	0.052	0.05	0.042	0.046
1.0	0.5	0.066	0.062	0.052	0.056
1.0	0.3	0.051	0.05	0.044	0.049
1.0	- 0.3	0.053	0.052	0.044	0.047
1.0	- 0.5	0.058	0.054	0.043	0.048
1.0	- 0.8	0.041	0.041	0.035	0.037
0.95	0.8	0.76	0.791	0.707	0.768
0.95	0.5	0.851	0.869	0.797	0.844
0.95	0.3	0.84	0.863	0.777	0.831
0.95	- 0.3	0.882	0.903	0.828	0.879
0.95	- 0.5	0.88	0.894	0.799	0.86
0.95	- 0.8	0.881	0.903	0.812	0.868
0.85	0.8	0.99	0.998	0.952	0.988
0.85	0.5	0.998	1.000	0.987	0.994
0.85	0.3	1.000	1.000	0.992	0.997
0.85	- 0.3	0.998	1.000	0.981	0.99
0.85	- 0.5	1.000	1.000	0.995	0.997
0.85	- 0.8	0.998	1.000	0.984	0.987

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varphi u_{t-1} + \varepsilon_t$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.21	Size of the unit root τ_{DF} – test statistic Autoregressive Case , T = 500 , k_{max} = 14 Specified nominal size α = 0.05 (5,000 simulations)				
ρ	φ	AICC	BICC	MAICC	MBICC
1.0	0.8	0.048	0.048	0.043	0.047
1.0	0.5	0.047	0.051	0.042	0.047
1.0	0.3	0.053	0.051	0.049	0.051
1.0	- 0.3	0.047	0.042	0.042	0.041
1.0	- 0.5	0.052	0.048	0.043	0.045
1.0	- 0.8	0.044	0.037	0.036	0.036

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varphi u_{t-1} + \varepsilon_t$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.22	Size and Power of the unit root τ_{DF} – test statistic Moving Average Case, T = 100, $k_{max} = 10$ Specified nominal size $\alpha = 0.05$ (5,000 simulations)				
ρ	**θ**	**AICC**	**BICC**	**MAICC**	**MBICC**
1.0	0.8	0.0586	0.0638	0.0276	0.0216
1.0	0.5	0.0612	0.0646	0.0316	0.0382
1.0	0.3	0.0598	0.0642	0.0364	0.0442
1.0	- 0.3	0.0664	0.0754	0.0402	0.0504
1.0	- 0.5	0.1022	0.1442	0.055	0.0688
1.0	- 0.8	0.229	0.4078	0.083	0.0964
0.95	0.8	0.272	0.321	0.159	0.165
0.95	0.5	0.351	0.398	0.248	0.261
0.95	0.3	0.353	0.392	0.274	0.314
0.95	- 0.3	0.398	0.451	0.303	0.347
0.95	- 0.5	0.491	0.644	0.322	0.374
0.95	- 0.8	0.782	0.951	0.382	0.424
0.85	0.8	0.728	0.79	0.525	0.598
0.85	0.5	0.813	0.88	0.655	0.734
0.85	0.3	0.893	0.945	0.735	0.806
0.85	- 0.3	0.914	0.973	0.71	0.79
0.85	- 0.5	0.93	0.984	0.684	0.741
0.85	- 0.8	0.982	1.000	0.834	0.846

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.23	Size and Power of the unit root τ_{DF} – test statistic Moving Average Case , T = 250 , k_{max} = 12 Specified nominal size α = 0.05 (5,000 simulations)				
ρ	θ	AICC	BICC	MAICC	MBICC
1.0	0.8	0.0502	0.0526	0.0336	0.024
1.0	0.5	0.0498	0.0534	0.0322	0.0326
1.0	0.3	0.0474	0.056	0.0354	0.0418
1.0	- 0.3	0.0624	0.079	0.045	0.0574
1.0	- 0.5	0.0756	0.1162	0.0466	0.0648
1.0	- 0.8	0.1678	0.3282	0.0778	0.1054
0.95	0.8	0.749	0.788	0.661	0.674
0.95	0.5	0.83	0.848	0.758	0.784
0.95	0.3	0.848	0.896	0.765	0.839
0.95	- 0.3	0.88	0.943	0.802	0.867
0.95	- 0.5	0.897	0.972	0.799	0.869
0.95	- 0.8	0.964	1.000	0.875	0.887
0.85	0.8	0.993	0.999	0.978	0.986
0.85	0.5	0.998	1.000	0.991	0.995
0.85	0.3	1.000	1.000	0.989	0.991
0.85	- 0.3	1.000	1.000	0.996	0.997
0.85	- 0.5	0.999	1.000	0.997	0.998
0.85	- 0.8	1.000	1.000	1.000	1.000

DGP : $y_t = \rho y_{t-1} + u_t \quad , \quad u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

Regression : $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.24

Size of the unit root τ_{DF} – test statistic, Moving Average Case, T = 100, k_{max} = 10
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	θ_1	θ_2	θ_3	θ_4	θ_5	$t_{sig}(5)$	$t_{sig}(10)$	AIC	BIC	AICC	BICC	MAICC	MBICC
1.0	-0.43	0.49	0	0	0	0.059	0.055	0.069	0.068	0.057	0.054	0.029	0.026
1.0	-0.61	0.98	0	0	0	0.064	0.057	0.052	0.069	0.0615	0.0835	0.0305	0.0287
1.0	0.55	1.25	0.54	0.96	0	0.043	0.046	0.029	0.038	0.053	0.086	0.032	0.0305
1.0	-0.99	0.98	-0.97	0	0	0.29	0.27	0.39	0.451	0.338	0.484	0.1985	0.235
1.0	-0.8	0.64	-0.51	0	0	0.124	0.099	0.118	0.206	0.1355	0.225	0.068	0.102
1.0	-1.54	1.79	-1.78	1.50	-0.95	0.164	0.155	0.157	0.245	0.143	0.245	0.0975	0.124
1.0	0.99	0.98	0.97	0	0	0.0738	0.068	0.0642	0.084	0.0622	0.0826	0.0394	0.0289
1.0	0.44	0.70	0.69	0.43	0.95	0.0668	0.0674	0.0666	0.0722	0.0628	0.0654	0.0252	0.0122

DGP : $y_t = \rho y_{t-1} + u_t$, $u_t \sim MA(q)$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.25

Size of the unit root τ_{DF} – test statistic, Autoregressive Case, $T = 100$, $k_{max} = 10$
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	$t_{sig}(5)$	$t_{sig}(10)$	AIC	BIC	AICC	BICC	MAICC	MBICC
1.0	0.61	-0.98	0	0	0	0	0.049	0.048	0.054	0.052	0.063	0.054	0.042	0.044
1.0	0.40	-0.38	0.34	-0.45	-0.39	-0.13	0.061	0.055	0.046	0.045	0.055	0.0652	0.0424	0.0448
1.0	1.60	-1.59	0.97	0	0	0	0.065	0.069	0.065	0.053	0.051	0.044	0.041	0.039
1.0	-0.61	0.61	0.97	0	0	0	0.075	0.074	0.07	0.069	0.062	0.046	0.039	0.044
1.0	1.60	-1.96	1.57	-0.96	0	0	0.061	0.055	0.041	0.04	0.055	0.056	0.044	0.045
1.0	-0.61	-0.37	0.37	0.59	0.95	0	0.062	0.06	0.063	0.054	0.057	0.054	0.046	0.051
1.0	-2.59	-3.55	-3.51	-2.51	-0.95	0	0.0338	0.0346	0.0386	0.0338	0.0352	0.0328	0.0222	0.0214

DGP : $y_t = \rho y_{t-1} + u_t$, $u_t \sim AR(p)$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

Table B.26

Size of the unit root τ_{DF} – test statistic
Autoregressive Moving Average Case , T = 100 , k_{max} = 10
Specified nominal size $\alpha = 0.05$
(5,000 simulations)

ρ	φ_1	φ_2	φ_3	φ_4	θ_1	θ_2	$t_{sig}(5)$	$t_{sig}(10)$	AIC	BIC	AICC	BICC	MAICC	MBICC
1.0	-0.31	-0.25	0	0	-0.99	0	0.2394	0.214	0.2698	0.4438	0.2816	0.455	0.149	0.1502
1.0	-0.43	-0.49	0	0	-0.99	0	0.2232	0.169	0.2014	0.3326	0.2178	0.346	0.1152	0.1156
1.0	-0.56	-0.81	0	0	-0.99	0	0.111	0.092	0.1104	0.1802	0.1244	0.1996	0.073	0.0732
1.0	-0.61	-0.98	0	0	-0.99	0	0.0354	0.0302	0.036	0.0618	0.036	0.062	0.022	0.021
1.0	-0.63	-0.84	-0.17	-0.40	-0.99	0	0.1008	0.0872	0.1052	0.1646	0.1041	0.1696	0.0664	0.0837
1.0	-0.44	-0.10	-0.43	-0.96	-0.99	0	0.045	0.0402	0.046	0.066	0.052	0.076	0.059	0.0652
1.0	-1.16	-0.98	0	0	-1.16	0.98	0.0506	0.0496	0.055	0.0624	0.053	0.02	0.022	0.0235
1.0	-0.44	-0.10	-0.43	-0.96	-0.99	0	0.0503	0.046	0.051	0.0738	0.0454	0.0724	0.053	0.0582

DGP : $y_t = \rho y_{t-1} + u_t \; , \; u_t \sim ARMA(p,q)$ **Regression :** $y_t = \zeta_0 y_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta y_{t-i} + \varepsilon_{tk}$

REFERENCES

Agiakloglou, C. and Newbold, P. (1992). Empirical evidence on Dickey-Fuller type tests, *Journal of Time Series Analysis*, 13, 471-483

Ahn, S.K. (1993). Some tests for unit roots in autoregressive integrated moving average models with deterministic trends, *Biometrika*, 80, 855-868

Ahtola, J. and Tiao, G.C. (1984). Parameter inference for nearly nonstationary first order autoregressive model, *Biometrika*, 71, 263-273

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models, *Biometrika*, 60, 255-265

Andrews, D.W.K. (1989). Power in econometric applications, *Econometrica*, 57, 1059-1090

Berk, K.N. (1974). Consistent autoregressive spectral estimates, *Annals of Statistics*, 2, 489-502

Blough, S.R. (1992). The relationship between power and level for generic unit root tests in finite samples, *Journal of Applied Econometrics*, 7, 295-308

Brockwell, P.J. and Davis, R.A. (1987). *Time Series: Theory and Methods.* Springer Series In Statistics, New York

Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting.* Springer Series In Statistics, New York

Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes, *Annals of Statistics*, 15, 367-401

Chan, N.H. and Wei, C.Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics*, 15, 1050-1063

Chan, N.H. (1988). The parameter inference for nearly nonstationary time series, *Journal of the American Statistical Association*, 83, 857-862

DeJong, D.N., Nankervis, J.C., Savin, N.E. and Whiteman, C.H. (1992). Integration versus trend stationary in time series, *Econometrica*, 60, 423-433

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

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

Eastwood, B.J. and Gallant, R.A. (1991). Adaptive rules for seminonparametric estimators that achieve asymptotic normality, *Econometric Theory*, 7, 307-340

Elder, J. and Kennedy, P.E. (2001). F versus t tests for unit roots, *Economics Bulletin*, 3, 1-6

Elder, J. and Kennedy, P.E. (2001). Testing for unit roots : What should students be taught? *Journal of Economic Education*, 137-146

Elliot, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root, *Econometrica*, 64, 813-836

Evans, G.B.A. and Savin, N.E. (1981). Testing for unit roots : 1, *Econometrica*, 49, 753-779

Evans, G.B.A. and Savin, N.E. (1984). Testing for unit roots : 2, *Econometrica*, 52, 1241-1270

Evans, G.B.A. and Savin, N.E. (1981). The calculation of the limiting distribution of the least squares estimator of a parameter in a random walk model, *Annals of Statistics*, 9, 1114-1118

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York, Wiley

Geweke, J. and Meese, R. (1981). Estimating regression models of finite but unknown order, *International Economic Review*, 22, 55-70

Gourieroux, C. and Monfort, A. (1995). *Statistical and Econometric Models.* Cambridge University Press, Cambridge, United Kingdom

Hall, A. (1989). Testing for a unit root in the presence of moving average errors, *Biometrika*, 76, 49-56

Hall, A. (1994). Testing for a unit root in time series with pretest data based model selection, *Journal of Business and Economic Statistics*, 12 , 461-470

Hall, W.J., Wijsman, R.A. and Chosh, J.K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis, *Annals of Mathematical Statistics*, 36, 575-614

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey

Hannan, E.J. and McDougall, A.J. (1988). Regression procedures for ARMA Estimation, *Journal of the American Statistical Association*, 83, 490-498

Hannan, E.J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. New York, Wiley

Hannan, E.J., McDougall, A.J. and Poskitt, D.S. (1989). Recursive estimation of autoregressions, *Journal of the Royal Statistical Society*, 51, 217-233

Harris, R.I.D. (1992). Testing for unit roots using the augmented Dickey-Fuller Test, *Economic Letters*, 38, 381-386

Kodde, D.A., Palm, F.C. and Pfann, G.A. (1990). Asymptotic least squares estimation efficiency considerations and applications, *Journal of Applied Econometrics*, 5, 229-243

Koehler, A.B. and Murphree, E.S. (1988). A comparison of the Akaike and Schwarz criteria for selecting model order, *Applied Statistics*, 37, 187-195

Lopez, J.H. (1997). The power of the ADF test, *Economics Letters*, 57, 5-10

Maddala, G.S. and Kim, I.M. (1998). *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge, United Kingdom

Nabeya, S.N. and Tanaka, K. (1990). A general approach to the limiting distribution for estimators in time series regression with nonstable autoregressive errors, *Econometrica*, 58, 145-163

Nabeya, S.N. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative, *Annals of Statistics*, 16, 218-235

Nelson, C.R. and Plosser, C. (1982). Trends and random walks in macroeconomic time series, *Journal of Monetary Economics*, 10, 139-162

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association*, 90, 268-281

Ng, S. and Perron, P. (2000). A note on the selection of time series models, *Manuscript*, Department of Economics, Boston College, Boston

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power, *Econometrica*, 69, 1519-1554

Oksendal, B. (1998). *Stochastic Differential Equations.* Springer Series In Statistics, New York

Perron, P. and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, *The Review of Economic Studies*, 63, 435-463

Perron, P. and Ng, S. (1998). An autoregressive spectral density at frequency zero for nonstationary tests, *Econometric Theory*, 14, 560-603

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

Perron, P. (1991). A continuous time approximation to the unstable first order autoregressive process: The case without an intercept, *Econometrica*, 59, 211-236

Perron, P. and Vogelsang, T.J. (1992). Nonstationarity and level shifts with application to purchasing power parity, *Journal of Business and Economic Statistics*, 10, 301-320

Perron, P. and Rodriguez, G. (1998). GLS detrending, efficient unit root tests and structural change, *Manuscript*, Boston University, Boston

Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series Regression, *Biometrika*, 75, 335-346

Phillips, P.C.B. (1987). Time series regression with a unit root, *Econometrica*, 55, 277-301

Phillips, P.C.B. and Park, J.Y. (1988). Asymptotic equivalence of ordinary least squares in regression with integrated regressors, *Journal of the*

American Statistical Association, 83, 111-115

Phillips, P.C.B. and Solo, V. (1992). Asymptotics for linear processes, *Annals of Statistics*, 20, 971-1001

Potscher, B.M. (1989). Model selection under nonstationary: Autoregressive models and stochastic linear regression models, *Annals of Statistics*, 17, 1257-1274

Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association*, 89, 1420-1437

Ross, S. (1996). *Stochastic Processes*. New York, Wiley

Said, S.E. and Dickey, D.A. (1985). Hypothesis testing in ARIMA (p,1,q) Models, *Journal of the American Statistical Association*, 80, 369-374

Said, S.E. and Fuller, D.A. (1984). Testing for unit roots in autoregressive moving average models of unknown order, *Biometrika*, 71, 599-607

Said, S.E. (1991). Unit roots test for time series data with a linear time trend, *Journal of Econometrics*, 47, 285-303

Schwert, G.W. (1989). Tests for unit roots: A Monte Carlo investigation, *Journal of Business and Economic Statistics*, 7, 147-160

Saikkonen, P. and Luukkonen, R. (1993). Testing for a moving average unit root in autoregressive integrated moving average models, *Journal of the American Statistical Association*, 88, 596-601

Sawa, T. (1978). Information criteria for discriminating among alternative regression models, *Econometrica*, 46, 1273-1291

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

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Annals of Statistics*, 8, 147- 164

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz, *Journal of the Royal Statistical Society*, 41, 276-278

Tanaka, K. (1996). *Time Series Analysis*. New York, Wiley

Taylor, C.C. (1987). Akaike's information criterion and the histogram. *Biometrika*, 74, 636-639

Tsay, R.S. (1984). Order selection in nonstationary autoregressive models, *Annals of Statistics*, 12, 1425-1433

Tyostheim, D. and Auestad, B.H. (1994). Nonparametric identification of nonlinear time series: Selecting significant lags, *Journal of the American Statistical Association*, 89, 1410-1419

Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-PLUS*. Springer Series In Statistics, New York

Xiao, Z. and Phillips, P.C.B. (1998). An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy, *Econometrics Journal*, 1, 27-43

Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-PLUS*. Springer Series In Statistics, New York