Order Number 9321543

Testing for stationarity, trend stationarity and unit root

Guo, Shengyi, Ph.D.

Southern Methodist University, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

TESTING FOR STATIONARITY, TREND STATIONARITY AND UNIT ROOT

A Thesis Presented to the Graduate Faculty of

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Economics

by

Shengyi Guo

(B.S., Hebei University, 1984)
(M.A., Southern Methodist University, 1989)

May 22, 1993

TESTING FOR STATIONARITY, TREND STATIONARITY AND UNIT ROOT

Approved by:

HERMAN J. BIERENS

ESFANDIAR MAASOUMI

THOMAS B. FOMBY

HENRY L. GRAY

ACKNOWLEDGEMENT

Over the years of my graduate training, I have accumulated a large amount of intellectual debt. I hope these words can repay some of it.

I am indebted to the members of my dissertation committee, Professors Herman Bierens, Essie Maasoumi, Tom Fomby and Buddy Gray, for guidance and encouragement. Professor Esfandiar Maasoumi has been very helpful by putting my work in the context of traditional model specification testing using variable addition method, by inviting me to present my work in the econometrics workshop, by criticizing my work and suggesting improvement, and in many other innumerable ways.

Most of all, I am grateful to my advisor, Professor Herman Bierens. Before his coming to SMU, I had been learning the theory of unit root testing through working for Professors Balke and Fomby, and through reading a series of papers by Peter C. B. Phillips. Through his lectures, Professor Bierens made many difficult concepts, such as convergence of random elements in metric space, the Skorohod metric, and the Functional Central Limit Theorem, meaningful. His unselfish and stimulating contribution to my thinking was instrumental in the writing and completion of the dissertation.

I am grateful to the faculty and staff members of the Department of Economics. I have been very lucky to be able to benefit both intellectually and financially from working on several summer projects as research assistant. For this I am grateful to Professors Kathy Hayes, Dan Slottje and Joseph Hirschberg (1990), Nathan Balke and Tom Fomby (1989) and Tom Fomby (1988). I want to thank Professor Joseph Hirschberg for his invaluable advice on computer algorithms and for many hours of help on SAS and FORTRAN, as well as some other nonhuman languages. Professor Nathan Balke taught my first graduate macroeconomics course and has patiently discussed any of my research ideas with me. Professor Tom Fomby introduced me to the beauty of the Box-Jenkins methodology. My skill in RATS programming has grown from zero by working for Professors Balke and Fomby and discussing problems with them. For all of this, I am grateful.

Ms. Pam Lange of the English Department at SMU, in addition to creating an enjoyable environment for graduate students in the ESL class, kindly helped to polish the dissertation to the present form. I am grateful for this.

Since the summer of 1991, I have been working as a consultant and programmer at the Research Department of the Federal Reserve Bank of Dallas, which allows sufficient time for me to finish the dissertation. The excellent research facility there speeded up the completion of my work.

I am grateful to the United Nations Fund for Population Activities (UNFPA) for the generous fellowship (1986-1987) through the State Education Commission of the People's Republic of China, which made it possible for me to come to study in this country.

Needless to say, the humble views expressed here are only my own and should not be interpreted as those of the Federal Reserve Bank of Dallas, or the UNFPA, or the staff members of these institutions.

Last but not the least, I am grateful to my wife, Qing, for encouragement and for taking care of the household duties when it seemed that I would be working on my dissertation forever.

Guo, Shengyi

B.S., Hebei University, 1984
M.A., Southern Methodist University, 1989

Testing for Stationarity, Trend Stationarity and Unit Root

Advisor: Herman J. Bierens,

Robert H. and Nancy Dedman Trustee Professor of Economics

Doctor of Philosophy degree conferred May 22, 1993

Thesis completed March 1, 1993

Testing for the existence of unit root is important because different asymptotic theory of inference applies depending on whether the series has a unit root or not. In the dissertation, several new tests for stationarity and trend stationarity, as well as a new test for unit root, are developed. The new tests developed in the dissertation all have some properties superior to the existing tests.

Tests for parameter stability also have power against the unit root hypothesis. We examine the power of the CUSUM test and the parameter fluctuation test. The case of a break trend and the issues concerning the selection of the break point are also discussed. We propose a new adaptive estimator of the long run variance that improves the asymptotic power of the tests. Our tests take stationarity as the null hypothesis.

We also propose a test for trend stationarity against the unit root that does not require the estimation of the long run variance. Our test takes stationarity as the null hypothesis, and it is consistent at the rate of the sample size. The asymptotic distribution of the test statistic under the null is standard Cauchy.

In another chapter, we introduce a new transformation of the Dickey-Fuller statistic that does not require the estimate of the long run variance. Our new test is still consistent. Finally, all of the new tests are applied to the extended Nelson-Plosser macroeconomic time series.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS xi

LIST OF TABLES xii

Chapter

1. INTRODUCTION 1

1.1. Introduction and Motivation 1

1.2. Objective and Methodology 2

1.3. Outline of the Dissertation 2

1.4. Notational Conventions 4

2. LITERATURE REVIEW 5

2.1. Introduction 5

2.2. Review of Probability Theory 7

2.2.1. Some Introductory Probability Theory 7

2.2.2. Brownian Motion and Functional Central Limit Theorems 14

2.3. Representation of Deterministic Trend 18

2.3.1. Polynomial Time Trend 19

2.3.2. Orthogonal Time Polynomials 20

2.3.3. Chebishev Polynomials 23

2.4. Asymptotic Distribution of the Unit Root Tests 23

2.5. Review of Unit Root Tests 26

2.5.1. Phillips and Perron's Extension of the Dickey-Fuller Test 27

2.5.2. Test for Unit Root by Variable Addition 32

2.5.3. Higher Order Autocorrelation and Unit Root 34

2.5.4. Nonlinear Trend Stationarity and Unit Root 35

2.5.5. Discussion 35

3. SPURIOUS REGRESSION, PARAMETER INSTABILITY AND UNIT ROOT 37

3.1. Introduction 37

3.2. Spurious Regression, Parameter Instability and Unit Root 38

3.2.1. Introduction 38

3.2.2. The CUSUM Test Using OLS Residuals 40

3.2.3. The Fluctuation of the Recursive Estimate of Trend Coefficients 42

3.2.4. Distinguishing Deterministic Trend with Breaks from Stochastic Trend 46

3.3. An Adaptive Estimator of the Long Run Variance 47

3.3.1. Introduction 47

3.3.2. An Adaptive Estimator of the Long Run Variance 48

3.3.3. Monte Carlo Simulation 54

3.4. Application of the Adaptive Estimator to Testing for Stationarity 54

3.4.1. The χ^2 Test of Park 55

3.4.2. The CUSUM Test 56

3.5. Summary and Conclusion 56

Appendix A 58

Appendix B 64

4. CAUCHY TEST FOR TREND STATIONARITY AGAINST THE UNIT ROOT MODEL 86

4.1. Introduction 86

4.2. Stationarity and Unit Root Around a General Deterministic Trend 88

4.2.1. The Model 88

4.2.2. The Asymptotic Theory 89

4.2.3. Test Statistic 91

4.2.4. Monte Carlo Simulation 93

4.3. Tests Using Orthogonal Polynomials 95

4.3.1. The Cauchy Test 96

4.3.2. A t-test 96

4.3.3. Monte Carlo Simulation 97

4.4. Further Discussion and Other Tests for Stationarity 98

4.5. Summary and Conclusion 100

Appendix 101

5. A TRANSFORMATION OF THE DICKEY-FULLER TEST WITHOUT ESTIMATING THE LONG RUN VARIANCE 113

5.1. Introduction 113

5.2. The Phillips and Phillips-Perron Transformation of the Dickey-Fuller Statistic 114

5.2.1. The Test 114

5.2.2. Simulation Results 116

5.3. An Alternative Transformation of the Dickey-Fuller Statistic 116

5.3.1. The Test 116

5.3.2. Simulation Results 117

5.4. Transformation of Q with Improved Power 119

5.4.1. The Test 119

5.4.2. Simulation Results 120

5.5. Summary and Conclusion 121

6. TREND, STATIONARITY AND UNIT ROOT IN MACROECONOMIC TIME SERIES 132

6.1. Introduction 132

6.2. Description of the Data 132

6.3. Testing for Linear Trend Stationarity and Unit Root 133

6.4. Breaking Trend Analysis 137

6.5. Summary and Conclusion 138

7. SUMMARY AND CONCLUSION 148

REFERENCES 151

LIST OF ILLUSTRATIONS

Figure Page

6.1. The Extended Nelson-Plosser Economic Time Series 134

LIST OF TABLES

Table Page

3.2.1. Simulated Asymptotic Distribution of the CUSUM Test 66

3.2.2. Monte Carlo Simulation on the CUSUM Test with An Intercept 67

3.2.3. Monte Carlo Simulation of the CUSUM Test with a Linear Trend 69

3.2.4. Simulated Asymptotic Distribution of the Fluctuation Test 71

3.2.5. Simulated Asymptotic Distribution of the CUSUM Test under the Presence of a one-time Shift in the Mean 72

3.2.6. Simulated Asymptotic Distribution of the CUSUM Test under the Presence of a one-time Shift in both the Intercept and the Slope 74

3.3.1. The Estimates of the Long Run Variance 76

3.3.2. Simulated Values of $\hat{\theta}_1$ and $\hat{\rho}$ 79

3.4.1. Monte Carlo Simulation on the $\chi^2(1)$ Test 80

3.4.2. Monte Carlo Simulation on the $\chi^2(1)$ Test Using the Adaptive Estimator 81

3.4.3. Monte Carlo Simulation on the CUSUM Test Using the Adaptive Estimator 82

3.4.4. Monte Carlo Simulation on the CUSUM Test with a Linear Trend Using the Adaptive Estimator 84

4.2.1. Monte Carlo Simulation on the Cauchy Test in Theorem 4.2.5 106

4.2.2. Values of the Quasi-nuisance Parameters in Theorem 4.2.5 107

4.3.1. Monte Carlo Simulation on the Cauchy Test Using Orthonormal Polynomials ... 108

4.3.2. Monte Carlo Simulation on the Multiplier $(1+T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^2)$ ... 109

4.3.3. Monte Carlo Simulation on the Quality of Approximation of the Asymptotic Distribution in Theorem 2.3.1 ... 110

4.3.4. Monte Carlo Simulation on the t_4 Test in Equation 4.3.2 ... 112

5.2.1. Monte Carlo Simulation on the Newey-West Estimator ... 122

5.2.2. Monte Carlo Simulation on the Z Test ... 123

5.3.1. Simulated Critical Values ... 126

5.3.2. Monte Carlo Simulation on the Q Test ... 128

5.3.3. The Asymptotics for the Power of the Z and Q Tests ... 129

5.4.1. The Empirical Distribution of $|\hat{\rho}|^{\sqrt{T}}$ ... 130

5.4.2. Monte Carlo Simulation on the Improved Q Test ... 131

6.3.1. Application of the Traditional Tests for Unit Root ... 140

6.3.2. Application of the CUSUM, Fluctuation Tests for Stationarity and the Modified Dickey-Fuller Test for Unit Root ... 142

6.3.3. Application of the Cauchy Tests ... 144

6.4.1. Results on Breaking Trend and Unit Root ... 145

6.4.2. Optimal Break Fraction ... 147

This work is dedicated to my daughter

Rose

for all the joy and work she has brought to me.

CHAPTER 1

INTRODUCTION

1.1 Introduction and Motivation

Ever since the publication of the influential study of Nelson and Plosser (1982) on the unit root behavior of economic time series, there has been intense interest in econometric modeling involving unit root processes.

The unit root behavior is both theoretically relevant and empirically important. A unit root in certain economic time series is implied by some intertemporal models of economic behavior: for example, marginal utility of consumption, cf. Hall (1978); stock prices, cf. Samuelson (1973); foreign exchange rate, cf. Meese and Singleton (1983). Interestingly, the absence of a unit root in linear combinations of certain economic time series (cointegration) is also suggested by economic theory: for example, the spot price and forward price of a financial asset, cf. Hakkio and Rush (1989); stock price and dividend, cf. Campbell and Shiller (1987); labor income and consumption, cf. Campbell (1987).

The dissertation is concerned with testing for unit root and stationarity in time series. Strictly speaking, the "stationarity" of a stochastic process requires that the underlying probability measure does not change over time. But unless otherwise stated in this dissertation, the "stationarity" of a time series only requires that it satisfies certain Functional Central Limit Theorems, cf. Theorem 2.2.4.

The asymptotic theory of inference invariably involves the use of various Central Limit Theorems (CLT's). Different CLT's are used depending on whether the stochastic process has a unit root or not. In the case of stationarity, classical CLT's are applicable. But in the case of a unit root, Functional Central Limit Theorems (FCLT's) need to be used.

The existing tests for unit root have several drawbacks. First of all, most of the existing tests take the unit root as the null model and stationarity as the alternative model. This is fine for testing economic theory that implies unit root behavior. As a test for economic theory that implies stationarity (e. g. cointegration), it puts the theory at a severe disadvantage. Secondly, the asymptotic theory requires the estimation of the long run variance, which may not behave well in finite samples. Thirdly, the existing test for stationarity is consistent at a rate slower than the sample size, which affects the asymptotic power of the tests.

1.2 Objective and Methodology

The purpose of this dissertation is to develop several new tests for stationarity and unit root that have superior finite and large sample properties over the existing tests.

We are going to use the theory of weak convergence on metric spaces, cf. Billingsley (1964). Using the theory of quadratic forms of normal random variables, Dickey and Fuller (1979) established the finite sample and asymptotic distributions of their tests for unit root under a restrictive I.I.D. normality assumption. As shown by Phillips (1987), the theory of weak convergence can be applied to more general stochastic processes, for example, mixing process.

1.3 Outline of the Dissertation

Chapter 1 is an introduction.

In Chapter 2, we briefly review the basic probability theory, introducing the concept of a mixing process, and the Functional Central Limit Theorems (FCLT's). Then we use the FCLT's to derive the asymptotic distribution of the Dickey-Fuller class of tests for unit root with general deterministic trend. With this characterization of the asymptotic distribution of the tests for unit root, we then review and contrast the existing tests for unit root.

There has been a large literature on testing for structural stability of regression coefficients. The existing tests can be classified into three categories. The first is the CUSUM tests using recursive residuals. The second is based on the fluctuation of parameter estimates over subsamples. The third assumes that the coefficients evolve as a random walk over time; it then tests the variance of innovations to the random walk being zero. Using results on spurious regression, we show that tests for parameter stability also have power against the unit root hypothesis, but an estimate of the long run variance is needed to implement the test. The conventional method of estimating the long run variance impairs the power of the test. Therefore, we propose a new adaptive estimator of the long run variance that enhances the asymptotic power of the tests. Our tests take stationarity as the null hypothesis. Monte Carlo simulation is conducted to study the finite sample properties of the tests. This makes up Chapter 3.

Using results on spurious regression and generalizing the Cauchy test of Bierens (1991a), we propose a test for trend stationarity against the unit root that does not require the estimation of the long run variance. Our test takes stationarity as the null hypothesis, and the unit root as the alternative model. It is consistent at the rate of the sample size. So it should be more powerful asymptotically than the existing tests for stationarity. Since the asymptotic distribution of the test statistic under the null hypothesis is standard Cauchy (Student t distribution with one degree of freedom), tabulation of critical values by simulation is not necessary. The asymptotic distribution under the null hypothesis does not depend on the alternative model against which the null is tested. Therefore our new test avoids most of the problems with the traditional tests. Monte Carlo simulation is conducted to study the finite sample properties of the tests. This makes up Chapter 4.

The Phillips-Perron transformation of the Dickey-Fuller test statistics for unit root requires an estimate of the long run variance. Because the long run variance is the sum of an infinite number of autocovariances, its estimate has to be truncated in finite samples. This truncation can make the finite sample estimate unreliable. In Chapter 3, we propose a new transformation of the Dickey-Fuller statistic that does not require the estimate of the long run variance. Yet our new test is still consistent. The finite sample properties of the test are investigated with Monte Carlo simulation.

In Chapter 6, all the tests developed in the dissertation are applied to the Nelson and Plosser (1982) macroeconomic time series extended up to 1988. In Chapter 7, we summarize and conclude.

1.4 Notational Conventions

We will follow the notational conventions in the literature, cf. Phillips (1987) and Phillips and Perron (1988). In particular, note the following. $\mathbb{R}^k$ stands for k dimensional Euclidean space. W(r) stands for standard Brownian motion on the unit interval [0, 1]. $\Rightarrow$ denotes weak convergence. $\xrightarrow{p}$ denotes convergence in probability. $\xrightarrow{a.s.}$ denotes convergence almost surely. All integrals are taken over the interval [0, 1], unless otherwise specified; integrals with respect to the Lebesgue measure such as $\int_0^1 rW(r)dr$ and $\int_0^1 f(r)dr$, are often written as $\int rW$ and $\int f$. The index for summation is t, and the range of summation is from 1 to T. All limits are taken for $T \to \infty$. Upper case letters, such as $\{Y_t\}$, denote a stochastic process; lower case letters, such as $\{y_t\}$, denote a realization of $\{Y_t\}$. A functional is understood to be a function of a function.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Box and Jenkins (1976) systematized time series model identification using sample autocorrelations and partial autocorrelations. Their method is intuitive and easy to follow. Their prescription for detecting nonstationarity is,

> failure of the estimated autocorrelation function to die out rapidly might logically suggest that we should treat the underlying stochastic process as nonstationary [...], but possibly as stationary in [first difference], or in some higher difference (p. 175).

They further emphasized that,

> it is the failure of the estimated autocorrelation function to die out rapidly that suggests nonstationarity. It need not happen that the estimated correlations are extremely high even at low lags (p. 175).

The concept of "die out rapidly" is very vague; its unwise use can lead to incorrect treatment of the series of interest, i. e. differencing the series while it is already stationary, or assuming stationarity while it needs to be differenced. Because the asymptotic theory of inference for stationary process and nonstationary process is very different, the incorrect treatment of the time series of interest can lead to serious problems.

For the simple AR(1) model,

(2.1.1) $$X_t = \rho X_{t-1} + u_t \qquad u_t \sim \text{IID}(0, \sigma^2)$$

the least squares estimate of ρ is,

(2.1.2) $$\hat{\rho} = \Sigma_{t=2}^{T} X_t X_{t-1} / \Sigma_{t=2}^{T} X_{t-1}^2.$$

We have the following asymptotic results on $\hat{\rho}$ depending on the true autoregressive coefficient ρ.

(i) If $|\rho| < 1$, then X_t is stationary. For fixed x_0, $\sqrt{T}(\hat{\rho}-\rho)$ is asymptotically normal, cf. Mann and Wald (1943).

(ii) If $\rho = 1$, then X_t has a unit root. The asymptotic distribution of $T(\hat{\rho}-1)$ is non-normal and can be expressed as a functional of a Brownian motion, cf. White (1958), Dickey and Fuller (1979), Phillips (1987).

(iii) If $|\rho| > 1$, then X_t is explosive. For $x_0=0$, $|\rho|^T(\hat{\rho}-\rho)$ is asymptotically Cauchy, cf. White (1958).

Therefore the asymptotic distribution theory is very different depending on whether ρ is inside, on or outside the unit circle.

All parameter estimators are functions of the data. Therefore to study the asymptotic behavior of $\hat{\rho}$ given in (2.1.2), first we need to study the asymptotic behavior of X_t. Let X_t be generated by,

(2.1.3) $$X_t = \rho X_{t-1} + u_t, \qquad \rho = 1$$

then

(2.1.4) $$X_t = \Sigma_{i=1}^{t} u_i + X_0.$$

Given that X_0 is a random variable not related to t, the asymptotic properties of X_t is the same as that of $\Sigma_{i=1}^{t} u_i$. In what follows, we assume $X_0 = 0$ for simplicity.

Define the quantity,

$$X_{[Tr]} = \Sigma_{t=1}^{[Tr]} u_t, \; r \in [0, 1]$$

where $[\cdot]$ here means the integral truncation operator. For fixed T and r, $X_{[Tr]}$ is a random variable; for fixed T, $X_{[Tr]}$ is a random function in r. So in order to develop an asymptotic theory for $X_{[Tr]}$, we need to study random functions.

In order to understand the literature on unit root, some knowledge of FCLT's is necessary. In Section 2.2, we first review the basic probability theory on metric spaces, and introduce the FCLT's and the concept of a mixing process. In Section 2.3, we discuss the representation of deterministic trend. In Section 2.4, we derive the asymptotic distribution of the Dickey-Fuller class of tests for unit root under general conditions. In Section 2.5, we review the literature on unit root tests.

2.2 Review of Probability Theory

In the following, we present some basic probability theory on random elements of metric spaces; then the theory will be specialized to establish the FCLT's which is the building block of asymptotic theory for unit root processes.

2.2.1 Some Introductory Probability Theory

Let $(\Omega,\mathcal{F},P)$ be a probability space, and $(\Psi,\mathcal{G})$ be a measurable space. We are going to study mappings from Ω to Ψ.

Definition 2.2.1

A mapping $X: \Omega \mapsto \Psi$ is $\mathcal{F}|\mathcal{G}$ measurable if

$\{\omega \in \Omega \mid x(\omega) \in G\} = F \in \mathcal{F}, \forall G \in \mathcal{G}.$

If X is $\mathcal{F}|\mathcal{G}$ measurable, then X is called a random element on $(\Omega,\mathcal{F},P)$.

The set G is called the image of the set F under the mapping X, denoted as $G = X(F)$, and F is called the inverse image of G under the relation X, denoted as $F = X^{-1}(G)$. If $(\Psi, \mathcal{G})$ is implicit, we simply say X is $\mathcal{F}$-measurable, or measurable. Note that the definition of a measurable mapping does not require the existence of a probability measure, but for the measurable mapping to have an interpretation as a random element, a probability measure is necessary.

If we are to study the convergence of a series of measurable mappings from Ω to Ψ, we need a measure of the distance between two elements of the space Ψ. Therefore Ψ is endowed with a metric ρ, so that (Ψ,ρ) forms a metric space. With the metric ρ, open and closed sets on (Ψ,ρ) are well defined. Crucial to our study of a random element is the minimal Borel field containing all open sets in the metric space. We denote this Borel field as $\mathfrak{B}$. Its elements are called Borel sets, denoted as B and the boundary of B is denoted as ∂B. Throughout, we take $\mathfrak{G} = \mathfrak{B}$[1]. We see that the metric ρ is related to random elements in two ways: first ρ affects the measurability of each mapping through $\mathfrak{B}$; second ρ affects the convergence of random elements through the definition of a distance. A weaker metric allows much easier convergence, cf. Bierens (1992b, Chapter 9). Since the Borel field $\mathfrak{B}$ is larger for a weaker metric, a weaker metric makes Borel measurability more difficult, cf. Pollard (1984, Chapter 4).

To emphasize the domain space, the range space, or both, of a random element, the random element can be called a random element on the probability space $(\Omega,\mathfrak{F},P)$, a random element of the measure [metric] space $(\Psi,\mathfrak{G})$ [(Ψ,ρ)], or a random element from the probability space $(\Omega,\mathfrak{F},P)$ to the measure space $(\Psi,\mathfrak{G})$. Here are some examples of metric spaces and random elements.

(i) If $\Psi = \mathbb{R}$, then X is a random variable; if $\Psi = \mathbb{R}^k$, then X is called a random vector,

$$\rho(x, y) = [\Sigma_{i=1}^{k}(x_i-y_i)^2]^{1/2} \tag{2.2.1}$$

is the Euclidean distance which is the basis for least squares estimation.

(ii) If Ψ is a space of functions on $\Theta \subset \mathbb{R}^k$, then X is called a random function. Two spaces of

[1]For other choices of $\mathfrak{G}$, see Pollard (1984). Note that a metric ρ is not necessary for the definition of a random element as long as the Borel field $\mathfrak{G}$ is well defined.

functions on $\Theta = [0, 1]$ we shall use later on are the space of continuous functions, denoted as C[0,1], and the space of functions with discontinuities of the first kind at most at countably many points, denoted as D[0,1]. The metric,

(2.2.2) $$\rho(x, y) = \sup_{s \in \Theta} | x(s)-y(s) |$$

is the uniform metric. The uniform metric is suitable for most studies. But for some studies, the uniform metric is too conservative. Another metric on C that is going to be used in Chapter 4 is,

$$\rho(x,y) = \int_{\Theta} [x(r)-y(r)]^2 dr.$$

(iii) For $\Psi = D[0,1]$, another more involved metric is often used. Let Λ be the space of strictly increasing continuous real functions $\lambda(s)$ on [0, 1] with $\lambda(0) = 0$ and $\lambda(1) = 1$, $\Lambda \subset C[0,1]$. $\rho(x, y)$ is defined as the infimum of those ϵ for which there is $\lambda \in \Lambda$ such that,

(2.2.3a) $$\sup_{s,t \in [0,1], s<t} | \ln[(\lambda(t)-\lambda(s))/(t-s)] | \leq \epsilon$$

and

(2.2.3b) $$\sup_{s \in [0,1]} | x(s)-y(\lambda(s)) | \leq \epsilon.$$

This is the famous Skorohod metric, denoted as ρ^0, cf. Bierens (1992b, Chapter 9). Think of $\lambda(s)$ as a monotonic transformation of time [0, 1] keeping the end points unchanged, then the two inequalities say that as long as the slope of $\lambda(s)$ is close to one, that is, as long as the time scale is not distorted too much by the transformation, the two functions stay close on different time scales. Note that, fixing $\lambda(s) = s$, the "Skorohod" metric is reduced to the uniform metric, so the uniform metric dominates the Skorohod metric. We see that the time transformation involved in the Skorohod norm is designed to "discount" the importance of discontinuities in the elements of D. For the definition of another metric on D, see Billingsley (1968, p. 113).

In general, the distance between x and y, the elements of (Ψ,ρ), is the "length" of the "difference" between x and y. So the metric space is often endowed with scalar multiplication, vector addition and a dot product, the metric space is turned into a normed linear dot product space (or vector space), cf. Dhrymes (1989, p. 64). The relationship among them is,

$$\rho(x, y) = \|x\text{-}y\|^2 = \mathrm{dot}(x\text{-}y, x\text{-}y).$$

For example, the metric spaces we will use, $\mathbb{R}^k$, C[0, 1] and D[0, 1] are all normed linear dot product spaces. For this reason, the Skorohod (uniform or sup) metric is also called Skorohod (uniform or sup) norm.

Given the metric space (Ψ,ρ), various modes of convergence can be defined. Let X_t and X be a series of random elements. Then X_t is said to converge to X in probability, denoted as $X_t \xrightarrow{P} X$ or plim $X_t = X$, if for any $\epsilon > 0$,

$$\lim_{t\to\infty} P(\omega \in \Omega \mid \rho(x_t(\omega), x(\omega)) < \epsilon) = 1.$$

X_t is said to converge to X almost surely, denoted as $X_t \xrightarrow{a.s.} X$, if,

$$P(\omega \in \Omega \mid \lim_{t\to\infty} x_t(\omega) = x(\omega)) = 1.$$

Almost sure(ly) is equivalently called almost everywhere (a. e.) or almost certainly (a. c.). Note that the limit in plim is a regular algebraic limit in the Euclidean space $\mathbb{R}$, while the limit in a. s. convergence is a limit in the metric space $(\Psi,\mathfrak{B})$.

In order to define the convergence in distribution, we need a concept on (Ψ,ρ) similar to the cumulative distribution function in $\mathbb{R}$. To this end, define a real function $\mu: \mathfrak{B} \mapsto \mathbb{R}$ for a given random element X,

$$\mu(B) = P(X^{-1}(B)), \ \forall\ B \in \mathfrak{B}.$$

It can be shown that μ is a probability measure on $(\Psi,\mathfrak{B})$. It is called the measure induced by the random element X. Then the measure μ will serve in a similar role to the cumulative distribution function.

Definition 2.2.2 A series of probability measures μ_t on $(\Psi,\mathfrak{B})$ is said to converge to μ, denoted as $\mu_t \rightarrow \mu$, if $\lim_{t\rightarrow\infty}\mu_t(B) = \mu(B)$, $\forall$ B $\in \mathfrak{B}$ with $\mu(\partial B) = 0$.

The condition $\mu(\partial B) = 0$ is similar to the restriction that the convergence of distribution functions only holds at the continuity points of the limiting distribution function. Note that the limit in the definition is the regular limit in Euclidean space $\mathbb{R}$. Now we can define the convergence in distribution of random elements.

Definition 2.2.3 Let $\{X_t\}_{t=1}^{\infty}$ and X be random elements on $(\Omega,\mathfrak{F},P)$, and $\{\mu_t\}_{t=1}^{\infty}$ and μ be the corresponding induced probability measures on $(\Psi,\mathfrak{B})$; then X_t is said to converge weakly to X, denoted as $X_t \Rightarrow X$, if $\mu_t \rightarrow \mu$.

As mentioned earlier, the OLS estimator involving unit root processes is a function of the unit root process, so given the asymptotic distribution of X_t, we also need to know the asymptotic distribution of functions of X_t. To this end, we have the following theorem:

Theorem 2.2.1 (Continuous Mapping Theorem, CMT)

Let $\{X_t\}$ and X be random elements on $(\Omega,\mathfrak{F},P)$, μ be the probability measure induced by X, and ϕ be a Borel measurable mapping from the metric space (Ψ,ρ) to the metric space (Ψ^*,ρ^*) such that ϕ is continuous on a Borel set $\Psi_0 \subset \Psi$, with $\mu(\Psi_0) = 1$. Then $\phi(X_t) \Rightarrow \phi(X)$.

Proof: Billingsley (1968, Theorem 5.1, p.30).

Therefore weak convergence is invariant under continuous transformations. Two measures of the dependence between two Borel fields are defined as follows:

Definition 2.2.4 Let $\mathfrak{G}$ and $\mathfrak{H}$ be Borel fields in $\mathfrak{F}$, $\mathfrak{G} \subset \mathfrak{F}$, $\mathfrak{H} \subset \mathfrak{F}$, define

$$\phi(\mathfrak{G}, \mathfrak{H}) = \sup_{\{G \in \mathfrak{G}, H \in \mathfrak{H}: P(G)>0\}} |P(H|G) - P(H)|$$

$$\alpha(\mathcal{G}, \mathcal{H}) = \sup_{\{G \in \mathcal{G}, H \in \mathcal{H}\}} |P(G \cap H) - P(G)P(H)|$$

where $P(H|G) = P(H \cap G)/P(G)$

They are equal to zero if all the events in $\mathcal{G}$ and $\mathcal{H}$ are independent. To study the dependence between random elements, we need the following concept.

Definition 2.2.5 Let X be a random element on $(\Omega, \mathcal{F}, P)$; then the Borel field generated by X, denoted as $\mathcal{F}(X)$, is the minimal Borel field that contains all sets of the form $(\omega \in \Omega \mid x(\omega) \in B) \in \mathcal{F}$, $\forall\, B \in \mathcal{B}$.

Note that $\mathcal{F}(X) \subset \mathcal{F}$, actually all sets of the form $\{X^{-1}(B) \mid B \in \mathcal{B}\}$ make up a Borel field, so

$$\mathcal{F}(X) = \{(\omega \in \Omega \mid X(\omega) \in B) \mid B \in \mathcal{B}\} = \{X^{-1}(B) \mid B \in \mathcal{B}\}.$$

That is, the inverse images of the Borel sets in $\mathcal{B}$ form a Borel field in Ω, cf. Dhrymes (1989, Proposition 4, p. 12) for the general case or Bierens (1992b, Theorem 1.1.1) for the case of $\Psi = \mathbb{R}^k$.

Definition 2.2.6 For a series of random elements $\{X_t\}$ on $(\Omega, \mathcal{F}, P)$, let

$$\mathcal{F}^t_{-\infty} = \mathcal{F}(\cdots, X_{t-1}, X_t);$$

$$\mathcal{F}^{\infty}_t = \mathcal{F}(X_t, X_{t+1}, \cdots).$$

$\mathcal{F}^t_{-\infty}$ is the Borel field generated by the history of X_t prior to (and including) time t, $\mathcal{F}^{\infty}_t$ is the Borel field generated by the future of X_t after (and including) time t. However the union of countable Borel fields $\bigcup_{i=0}^{\infty}\mathcal{F}(X_{t-i}, \cdots, X_{t-1}, X_t)$ is, in general, not a Borel field. So Bierens (1992b) has defined $\mathcal{F}^t_{-\infty}$ as the minimal Borel field containing $\bigcup_{i=0}^{\infty}\mathcal{F}(X_{t-i}, \cdots, X_{t-1}, X_t)$. $\mathcal{F}^{\infty}_t$ is defined similarly.

If $\mathcal{G}$ and $\mathcal{H}$ are the Borel fields generated by two random elements, then $\phi(\mathcal{G},\mathcal{H})$ and $\alpha(\mathcal{G},\mathcal{H})$ measure the dependence between the two random elements. Two measures of the serial dependence, the mixing coefficients, of a series of random elements $\{X_t\}$ are defined as follows:

Definition 2.2.7 Let

$\phi(m) = \sup_t \phi(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+m}^{\infty})$,

$\alpha(m) = \sup_t \alpha(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+m}^{\infty})$.

If $\lim_{m\to\infty}\phi(m) = 0$, then X_t is called ϕ-mixing (uniform mixing);

If $\lim_{m\to\infty}\alpha(m) = 0$, then X_t is called α-mixing (strong mixing).

They both measure the dependence among random elements at least m periods apart. Since $\phi(m) \geq \alpha(m)$, ϕ-mixing implies α-mixing. The rate at which $\phi(m)$ or $\alpha(m)$ approaches zero defines their sizes. If $\phi(m) = O(m^{-\lambda})$, for any $\lambda > p$, then $\phi(m)$ is said of size p. Similarly if $\alpha(m) = O(m^{-\lambda})$ for any $\lambda > p$, then $\alpha(m)$ is said of size p. A stochastic process X_t is ϕ- (α-) mixing of size p if it is ϕ- (α-) mixing and $\phi(m)$ ($\alpha(m)$) is of size p[2]. The mixing coefficients are related to the well-known autocovariance of a time series through the following theorem:

Theorem 2.2.2 Let $Eu_t = Eu_{t+m} = 0$.

(i) If $Eu_t^2 < \infty$, $Eu_{t+m}^2 < \infty$ and u_t is ϕ-mixing then

$$|Eu_t u_{t+m}| \leq 2\phi(m)^{1/2}(Eu_t^2)^{1/2}(Eu_{t+m}^2)^{1/2}.$$

(ii) If $E|u_t^\beta| < \infty$, $E|u_{t+m}|^\beta < \infty$, $\beta > 2$, and u_t is α-mixing, then

$$|Eu_t u_{t+m}| \leq 2(2^{1/2}+1)\alpha(m)^{(\beta-2)/(2\beta)}(Eu_t^2)^{1/2}(E|u_{t+m}|^\beta)^{1/\beta}.$$

Proof: White (1984), Corollary 6.16, p. 148.

[2]Some authors have preferred to call it mixing of size -p.

2.2.2 Brownian Motion and Functional Central Limit Theorems

The asymptotic theory for unit root processes often involves functionals of a Brownian motion on the unit interval. With the theory of weak convergence on metric spaces, the Functional Central Limit Theorem can be proved. First we define the Brownian motion.

Definition 2.2.8 Let W be a random function from $(\Omega,\mathcal{F},P)$ to (C,ρ), where ρ is the uniform metric, such that,

(i) $W(0) = 0$;

(ii) $W(t)-W(s) \sim N(0, (t-s)), \forall\, t, s \in [0, 1], \quad s < t$;

(iii) $W(t_1)-W(s_1)$ and $W(t_2)-W(s_2)$ are independent, $\quad \forall\, t_i, s_i \in [0, 1], i = 1, 2,$

$s_1 < t_1 < s_2 < t_2$.

Then W(t) is called a standard Brownian motion (or a Wiener process).

Therefore by definition, the standard Brownian motion has independent normally distributed increments, with variance increasing proportionally with time. It is standardized to be zero at time zero. Note in particular that $W(1) \sim N(0, 1)$. W has a continuous path: $W \in C[0, 1]$. The probability measure on C induced by the Brownian motion is called the Wiener measure.

As an exercise on Brownian motion, we present the following theorem which is useful for calculating the covariance between functionals of Brownian motion.

Theorem 2.2.3 For any deterministic continuous functions f and g on [0, 1], for any $\tau \in [0, 1]$,

(i) $E(\int_0^1 f dW \int_0^\tau g dW) = \int_0^\tau fg dr$;

(ii) $E(\int_0^1 fW \int_0^\tau gW) = \int_0^\tau f(r)[\int_0^r g(s)s ds + \int_r^\tau g(s)r ds]dr + \int_\tau^1 f(r)[\int_0^\tau g(s)s ds]dr$.

Proof: Note that Brownian Motion W has independent increments,

$$E[dW(r)dW(s)] = \begin{cases} 0 & \text{if } r \neq s \\ dr & \text{if } r = s \end{cases}$$

Therefore

$$E[W(r)W(s)] = \min(r, s);$$

$$\text{(i) } E(\int_0^1 f dW \int_0^1 g dW) = \int_0^1 f(r)\Big[\int_0^1 g(s)E[dW(r)dW(s)]\Big] = \int_0^1 f(r)g(r)dr;$$

$$\text{(ii) } E(\int_0^1 fW \int_0^1 gW) = \int_0^1 \int_0^1 f(r)g(s)E[W(r)W(s)]dsdr$$

$$= \int_0^1 f(r)\Big[\int_0^r g(s)E[W(r)W(s)]ds + \int_r^1 g(s)E[W(r)W(s)]ds\Big]dr$$

$$+ \int_r^1 f(r)[\int_0^1 g(s)E[W(r)W(s)]dsdr$$

$$= \int_0^1 f(r)[\int_0^r g(s)sds + r\int_r^1 g(s)ds]dr + \int_r^1 f(r)[\int_0^1 g(s)sds]dr. \qquad \square$$

With the theory of weak convergence of random elements, we now study the asymptotic behavior of the partial sum,

$$\text{(2.2.4)} \qquad X_{[Tr]} = \Sigma_{t=1}^{[Tr]} u_t, \; r \in [T^{-1}, 1]$$

$$= 0, \; r \in [0, T^{-1})$$

where for fixed t, u_t: $\Omega \mapsto \mathbb{R}$ is random variable on $(\Omega, \mathcal{F}, P)$. For fixed T and r, $X_{[Tr]}$ is either the sum of random variables or zero, so is a random variable itself. But for fixed T, $X_{[Tr]}$ is a random function on [0, 1]. Furthermore, for fixed T and fixed $\omega \in \Omega$, $X_t(\omega)$ is right continuous, with finite left limits, and the number of discontinuities is T, so $X_{[Tr]}(\omega) \in D[0,1]$. First we impose some restrictions on u_t.

Assumption 2.2.1 Let $\{u_t\}_{t=1}^{\infty}$ be a series of random variables on $(\Omega,\mathcal{F},P)$

(i) $E(u_t) = 0, \forall\ t$;

(ii) $\sup_t E|u_t|^\beta < \infty$ for some $\beta \geq 2$;

(iii) $\sigma^2 = \lim_{T\to\infty} E[(1/\sqrt{T})\Sigma_{t=1}^{T} u_t]^2,\ 0 < \sigma^2 < \infty$;

(iv) either $\beta \geq 2$ and u_t is ϕ-mixing with mixing coefficients $\phi(m)$ satisfying

$\Sigma_{m=1}^{\infty}\phi(m)^{1-1/\beta} < \infty$;

or $\beta > 2$ and u_t is α-mixing with mixing coefficients $\alpha(m)$ satisfying

$\Sigma_{m=1}^{\infty}\alpha(m)^{1-2/\beta} < \infty$.

Condition(ii) controls the degree of heterogeneity or limits the probability of outlier occurrence. Conditions (iii) and (iv) control the serial dependence in the series. As noted by Phillips (1987), there is a tradeoff between the probability of outliers (represented by β) and the degree of serial dependence (represented by $\alpha(m)$ or $\phi(m)$) as indicated by the inequality in (iv). σ^2 in (iii) is called the long run variance of u_t. The following theorem says that the standardized partial sum converges to a Brownian motion.

Theorem 2.2.4 (Functional Central Limit Theorem, FCLT)

Let $X_{[Tr]}$ be defined as in (2.2.4), and let u_t satisfy Assumption (2.2.1). Then

$$\frac{1}{\sqrt{T}\sigma} X_{[Tr]} \Rightarrow W(r), \qquad r \in [0, 1].$$

Proof: Herrndorf (1984).

The functional central limit theorem is also called the invariance principle. Note that $X_{[Tr]} \in D[0, 1]$, but $W \in C[0, 1]$. Applying the CMT to the FCLT, we have, Theorem 2.2.5 Let f be a continuous real function on $\mathbb{R}$; let ΔX_t satisfy Assumption (2.2.1).

Then

(i) $T^{-1}\Sigma_{t=1}^{T}f(T^{-1/2}X_t/\sigma) \Rightarrow \int_0^1 f(W)$;

(ii) $T^{-1}\Sigma_{t=1}^{T}f(t/T)(T^{-1/2}X_t/\sigma) \Rightarrow \int_0^1 fW$.

Proof: (i) Bierens (1992b, Lemma. 9.2.1); (ii) Bierens (1991a, p. 4).

The continuous function for CMT in (i) of the above theorem is the composite function,

$$\phi(x) = \int_0^1 f[x(t)],$$

and that in (ii) is the composite function,

$$\phi(x) = \int_0^1 f(t)x(t).$$

The asymptotic distribution of the OLS estimate of the first order autoregressive coefficient with a unit root also involves the asymptotic distribution of $\Sigma_{t=2}^{T}X_{t-1}u_t$; for this we have the following theorem.

Theorem 2.2.6 Let $\Delta X_t = u_t$ satisfy Assumption (2.2.1). Then

(i) $T^{-1}\Sigma_{t=2}^{T}X_{t-1}u_t/\sigma \Rightarrow \int_0^1 WdW + (1/2)(1-\sigma_u^2/\sigma^2) \equiv (1/2)[W(1)^2-\sigma_u^2/\sigma^2]$.

(ii) If f is a continuous real function on $\mathbb{R}$, then

$T^{-1/2}\Sigma_{t=1}^{[T\tau]}f(t/T)u_t/\sigma \Rightarrow \int_0^\tau fdW, \qquad \tau \in [0, 1]$

where $\sigma_u^2 = \lim_{T\to\infty}T^{-1}\Sigma_{t=1}^{T}Eu_t^2$.

Proof: (i) Phillips (1987, Theorem 3.1(b)); (ii) Bierens (1991a, Lemma 1).

Note that the integral in (i) is a stochastic integral. The equality is derived using Ito's Lemma by which

$$dW^2 = 2(WdW+dt),$$

so

$$\int WdW = (1/2)(\int dW^2 - \int dt) = (1/2)[W(1)^2-1].$$

The mapping $\phi(X) = \int_0^1 X(r)dX(r)$, $X \in C[0, 1]$, is not continuous, so we cannot apply the CMT to FCLT, cf. Phillips (1987). In contrast, the mapping $\phi(X) = \int_0^1 f(r)dX(r)$, $X \in C[0, 1]$, is continuous; therefore result (ii) is a direct application of the FCLT and CMT.

In particular, note that if

$$f(t/T) = e^{([T\tau]/T-t/T)c}, \qquad c \in \mathbb{R},$$

then the asymptotic distribution $\int_0^\tau e^{(\tau-r)c}dW$ in Theorem (2.2.6)(ii) is an Ornstein-Uhlenbeck process. This stochastic process has been used extensively in the study of near-unit root processes, cf. Phillips (1987), Chan and Wei (1987).

It is easy to see that, if function f in Theorems 2.2.5(ii) and 2.2.6(ii) is discontinuous at a point in (0, 1) but with finite left and right limits, then the theorems still hold but with piecewise integral on the right hand side. This result is required to do breaking trend analysis.

2.3 Representation of Deterministic Trend

Given T observations of a time series, call it y^T, then it forms a point in the Euclidean space $\mathbb{R}^T$. y^T can be written as a linear combination of the basis of $\mathbb{R}^T$. The most common basis is formed by the column vectors e_i, $i = 1, 2, \cdots, T$, where the jth element of e_i is given by the indicator function $I(i=j)$. This set of basis vectors $\{e_1, e_2, \cdots, e_T\}$ defines the conventional coordinate system in Euclidean space $\mathbb{R}^T$. Another basis is formed by the set of column vectors τ_i, $i = 1, 2, \cdots, T$, where the jth element of τ_i is j^{i-1}. This set of basis vectors $\{1, \tau_2, \cdots, \tau_T\}$ forms the basis for polynomial time regression. The difference between statistics and linear algebra on vector space is that in statistics a point in $\mathbb{R}^T$ can be expressed as the sum of a point in an Euclidean space of much lower dimension (k) and a random error. If k is not substantially lower than T, then the statistical description of the data is called overfitting. There are many orthogonal bases in $\mathbb{R}^T$, but here we discuss only two of them. The first is orthogonal polynomials in time for polynomial time trend. The second is the Chebishev polynomials for a much weaker time trend, cf. Bierens (1992a).

2.3.1 Polynomial Time Trend

Polynomials in time have been the simplest and most popular representation of deterministic trend in time series. At least the linear trend can be rationalized by a constant increment. The following regression, often called detrending, is usually estimated,

(2.3.1) $$Y_t = b_0 + tb_1 + t^2b_2 + \cdots + t^kb_k + e_t \equiv f(t)b + e_t,$$

where $f(t) = (1, t, \cdots, t^k)$, $b = (b_0, b_1, \cdots, b_k)'$. The OLS estimate of b is,

$$\hat{b} = [\Sigma f(t)'f(t)]^{-1}[\Sigma f(t)'Y_t].$$

Let $D_T = \text{diag}(1, T, \cdots, T^k)$, and then it is trivial that,

(2.3.2) $$(1/T)\ D_T^{-1}[\Sigma f(t)'f(t)]D_T^{-1} \to \int f'f.$$

If e_t satisfies Assumption (2.2.1), then by Theorem 2.2.6 (ii),

(2.3.3) $$(1/\sqrt{T})\ D_T^{-1}[\Sigma f(t)'e_t] \Rightarrow \sigma \int f dW.$$

Therefore,

(2.3.4) $$\sqrt{T}D_T(\hat{b}-b) \Rightarrow \sigma[\int f'f]^{-1}[\int f dW].$$

Note that the k^{th} order polynomial in time can also be written as,

(2.3.1)′ $$Y_t = f(t/T)b^* + e_t,$$

where $b^* = (b_0, b_1/T, \cdots, b_k/T^k)'$. Then (2.3.2)-(2.3.4) are respectively, (2.3.2)′ $$(1/T)[\Sigma f(t/T)'f(t/T)] \to \int f'f,$$

(2.3.3)′ $$(1/\sqrt{T})[\Sigma f(t/T)'e_t] \Rightarrow \sigma \int f dW,$$

(2.3.4)′ $$\sqrt{T}(\hat{b}^*-b^*) \Rightarrow \sigma[\int f'f]^{-1}[\int f dW].$$

Therefore if the polynomials in time are expressed in the form of f(t/T), the asymptotic theory is much easier to handle, since we do not have to bother with the D_T matrix. For this reason, from now on we will adopt this representation of the deterministic trend, and continue calling the function [a+(t/T)b] a linear trend.

By Theorem 2.2.3, the asymptotic covariance of $\sqrt{T}(\hat{b}^*-b^*)$ is $\sigma^2[\int f'f]^{-1}$. In the variable addition approach to testing for stationarity against unit root, we need to orthogonalize the covariance matrix of the parameter estimates. There are two equivalent ways of doing this. The first is to orthogonalize the covariance $[\int f'f]^{-1}$, the other is to orthogonalize the regressors. It turns out that for the representation and proof of the asymptotic theory, it is much easier to orthogonalize the regressors.

2.3.2 Orthogonal Time Polynomials

The orthogonal time polynomials can be constructed as in Hamming (1986, Section 26.6). They are defined as follows. Let

(2.3.5a) $$p_{0,T}(t) = 1$$

(2.3.5b) $$p_{1,T}(t) = t - \frac{T+1}{2T}$$

(2.3.5c) $$p_{2,T}(t) = (t - \alpha_2)p_{1,T}(t) - \beta_2 p_{0,T}(t)$$

(2.3.5d) $$p_{3,T}(t) = (t - \alpha_3)p_{2,T}(t) - \beta_3 p_{1,T}(t)$$

where $\alpha_i = \dfrac{\Sigma(t/T)p_{i-1,T}^2(t/T)}{\Sigma p_{i-1,T}^2(t/T)}$, $\beta_i = \dfrac{\Sigma(t/T)p_{i-1,T}(t/T)p_{i-2,T}(t/T)}{\Sigma p_{i-2,T}^2(t/T)}$, for i = 2, 3.

The polynomials $p_{i,T}(t/T)$, i= 0, 1, 2 and 3, as constructed above, are exactly orthogonal for t = 1, 2, $\cdots$, T. It can be shown that, when $T \to \infty$,

$$\alpha_2 \to 1/2, \qquad \beta_2 \to 1/12,$$

$$\alpha_3 \to 1/2, \qquad \beta_3 \to 1/15.$$

Then as $T \to \infty$, for $t = [Tr]$, $r \in [0, 1]$,

(2.3.6a) $$p_{0,T}(t/T) \to q_0(r) = 1$$

(2.3.6b) $$p_{1,T}(t/T) \to q_1(r) = r-1/2$$

(2.3.6c) $$p_{2,T}(t/T) \to q_2(r) = (r-1/2)^2 - 1/12$$

(2.3.6d) $$p_{3,T}(t/T) \to q_3(r) = (r-1/2)^3 - (3/20)(r-1/2)$$

which are a system of orthogonal functions on the interval [0, 1]. The orthonormal polynomials are constructed as,

(2.3.7) $$p_{i,T}^*(t/T) = p_{i,T}(t/T)/\sqrt{\Sigma p_{i,T}^2(t/T)}, \qquad i = 0, 1, 2, 3.$$

Note that $p_{i,T}^*(t/T)$, i = 0, 1, 2, 3, are a class of orthonormal elements in $\mathbb{R}^T$. Now we run the orthonormal regression,

(2.3.8) $$y_t = \sum_{i=0}^{3} b_i p_{i,T}^*(t/T) + e_t.$$

Denote the coefficient vector as b, and we have the following theorem:

Theorem 2.3.1

For the orthonormal polynomial regression (2.3.8),

(i) if $e_t = u_t$ satisfy Assumption 2.2.1, then

$$(\hat{b}-b) \Rightarrow \sigma(\int dW,\ \int q_1^* dW,\ \int q_2^* dW,\ \int q_3^* dW)' \sim N(0, \sigma^2 I_4);$$

(ii) if $\Delta e_t = u_t$ satisfy Assumption 2.2.1, then

$$T^{-1}(\hat{b}-b) \Rightarrow \sigma(\int W,\ \int q_1^* W,\ \int q_2^* W,\ \int q_3^* W)'$$

where $q_i^*(r) = \dfrac{q_i(r)}{\sqrt{\int q_i^2(r)dr}}$, $i = 0, 1, 2, 3.$

Proof: Since the regressors are orthonormal,

$$(\hat{b}-b) = X'e = (\Sigma p_{0,T}^*(t/T)e_t,\ \Sigma p_{1,T}^*(t/T)e_t,\ \Sigma p_{2,T}^*(t/T)e_t,\ \Sigma p_{3,T}^*(t/T)e_t)'.$$

By the definition of the Riemann integral,

$$\Sigma_{t=1}^T p_{i,T}^2(t/T)\,\frac{1}{T} \to \int_0^1 q_i^2, \qquad i = 0, 1, 2, 3.$$

(i) Under H_0, $e_t = u_t$, by Theorem 2.2.6(ii),

$$T^{-1/2}\Sigma p_{i,T}(t/T)u_t \Rightarrow \int q_i dW, \qquad i = 0, 1, 2, 3.$$

Therefore,

$$\Sigma p_{i,T}^*(t/T)e_t = \frac{\Sigma p_{i,T}(t/T)u_t}{\sqrt{\Sigma p_{i,T}(t/T)^2}}$$

$$= \frac{T^{-1/2}\Sigma p_{i,T}(t/T)u_t}{\sqrt{T^{-1}\Sigma p_{i,T}(t/T)^2}}$$

$$\Rightarrow \frac{\int q_i dW}{\int q_i^2} = \int q_i^* dW, \quad i = 0, 1, 2, 3.$$

It is easy to show that the above weak convergence for $i = 0, 1, 2, 3$ also holds jointly. So

$$(\hat{b}-b) \Rightarrow \sigma(\int dW,\ \int q_1^* dW,\ \int q_2^* dW,\ \int q_3^* dW)'.$$

Furthermore, using Theorem 2.2.3(i) and the orthonormality of q_i^*, it is trivial that,

$$\sigma(\int dW,\ \int q_1^* dW,\ \int q_2^* dW,\ \int q_3^* dW)' \sim N(0, \sigma^2 I_4).$$

(ii) Under H_1, $e_t = \Sigma_{j=1}^t u_j$, using Theorem 2.2.5(ii), it is immediate that

$$T^{-1}\Sigma p^*_{i,T}(t/T)e_t \Rightarrow \sigma \int q^*_i W, \qquad i = 0, 1, 2, 3.$$

It is easy to show that the above weak convergence for i = 0, 1, 2, 3 also holds jointly. □

Note that $q^*_i(r)$, i = 0, 1, 2, 3, form a system of orthonormal functions in C[0, 1]. The coefficient estimate $\hat{b}$ in (i) has an orthogonal covariance matrix; therefore it is very easy to construct χ^2, Cauchy and t statistics.

2.3.3 Chebishev Polynomials

Bierens (1992a) has used linear combinations of Chebishev polynomials to represent any nonlinear trend. They are defined as follows. For t = 1, 2, ⋯, T, let,

$$p_{0,T}(t) = 1,$$

$$p_{k,T}(t) = \sqrt{2}\cos[k\pi(t-1/2)/T], \qquad k = 1, 2, \cdots, T-1.$$

It is easy to show that the polynomials $P_{k,T}(t)$, k = 0, 1, ⋯, T-1, are orthogonal for fixed T and t = 1, 2, ⋯, T, cf. Bierens (1992a). It is trivial that, with k fixed as $T \to \infty$,

$$p_{k,T}(t) \to \sqrt{2}\cos(kr\pi),$$

where r is defined through t = [Tr]. For k = 0, 1, 2, ⋯, the limit functions, $\sqrt{2}\cos(kr\pi)$, are a class of orthonormal functions in C[0, 1].

2.4 Asymptotic Distributions of the Unit Root Tests

As an application of the FCLT and the CMT, we derive the asymptotic distribution of the unit root test with a general deterministic time trend. The results of Phillips (1987) without an intercept, Phillips and Perron (1988) with an intercept or a linear time trend, and Ouliaris, Park and Phillips (1988) with polynomial time trend are all specializations of our general theorem.

Theorem 2.4.1 Let ΔX_t satisfy Assumption 2.2.1, f(r) be a continuous function on [0, 1], $\hat{\rho}$ be the OLS estimate of ρ in the following regression equation,

(2.4.1) $$X_t = f(t/T)\theta + \rho X_{t-1} + e_t, \qquad t = 2, 3, \cdots, T.$$

Let $t_{\rho=1}$ be the t-statistic for $\rho = 1$. Then

(2.4.2) $$T(\hat{\rho}-1) \Rightarrow [\int_0^1 W^{*2}]^{-1}[\int_0^1 WdW + \lambda - \eta]$$

$$\equiv [\int_0^1 W^{*2}]^{-1}[\int_0^1 W^{*}dW + \lambda];$$

$$t_{\rho=1} \Rightarrow (\sigma_u/\sigma)[\int_0^1 WdW + \lambda - \eta]/[\int_0^1 W^{*2}]^{1/2}$$

$$\equiv (\sigma_u/\sigma)[\int_0^1 W^{*}dW + \lambda]/[\int_0^1 W^{*2}]^{1/2}$$

where $\lambda = (1/2)(1-\sigma_u^2/\sigma^2)$, $\eta = [\int_0^1 Wf(r)dr][\int_0^1 f(r)'f(r)dr]^{-1}[\int_0^1 f(r)'dW]$, $W^*(r)$ is the projection residual of Brownian motion on f(r).

Proof: It is trivial that

$$\hat{\rho} - \rho = [\Sigma_{t=2}^T x_{t-1}^{*2}]^{-1}[\Sigma_{t=2}^T x_{t-1}^{*} u_t]$$

where X_{t-1}^* is the linear projection residual of X_{t-1} on f(t/T). But

$$X_{t-1}^* = X_{t-1} - f(t/T)[\Sigma f(t/T)'f(t/T)]^{-1}[\Sigma f(t/T)'X_{t-1}]$$

so X_{t-1}^* is a continuous function of X_{t-1} and $\Sigma f(t/T)'X_{t-1}$. Since

$$T^{-3/2}\Sigma f(t/T)'X_{t-1} \Rightarrow \sigma \int_0^1 f(r)W$$

$$T^{-1}[\Sigma f(t/T)'f(t/T)] \rightarrow \int_0^1 f(r)'f(r)dr$$

$$T^{-1}\Sigma f(t/T) \rightarrow \int_0^1 f(r)dr$$

then

(2.4.3) $$T^{-1/2}\sigma^{-1}X^*_{[Tr]} \Rightarrow W(r) - f(r)[\int_0^1 f(r)'f(r)dr]^{-1}\int_0^1 f(r)'W \equiv W^*(r).$$

By Theorem 2.2.4

$$T^{-2}\sigma^{-2}\Sigma X^{*2}_{t-1} \Rightarrow \int_0^1 W^{*2}.$$

Similarly,

$$T^{-1}\sigma^{-2}\Sigma X^*_{t-1}u_t = T^{-1}\sigma^{-2}\Sigma X_{t-1}u_t$$

$$- T^{-1/2}\sigma^{-2}[T^{-3/2}\Sigma X_{t-1}f(t/T)][T^{-1}\Sigma f(t/T)'f(t/T)]^{-1}[\Sigma f(t/T)'u_t]$$

(2.4.4) $$\Rightarrow \int_0^1 W dW + \lambda - [\int_0^1 W f(r)][\int_0^1 f(r)'f(r)dr]^{-1}[\int_0^1 f(r)'dW].$$

Therefore,

$$T(\hat{\rho}-1) \Rightarrow [\int_0^1 W^{*2}]^{-1}\Big[\int_0^1 W dW + \lambda - [\int_0^1 W f(r)][\int_0^1 f(r)'f(r)dr]^{-1}[\int_0^1 f(r)'dW]\Big]$$

$$\equiv [\int_0^1 W^{*2}]^{-1}[\int_0^1 W^* dW + \lambda].$$

The proof for the t-statistic is similar. □

Let X^*_t be the linear projection residuals of X_t on $f(t/T)$; if we had estimated a first order autoregression using the detrended series X^*_t,

(2.4.5) $$X^*_t = \tilde{\rho}X^*_{t-1} + e_t$$

then we would have had,

(2.4.6) $$T(\tilde{\rho}-1) \Rightarrow [\int_0^1 W^{*2}]^{-1}[\int_0^1 W^* dW^* + \lambda]$$

which is the same as (2.4.2), since $\int W^* dW = \int W^* dW^*$. This procedure of detrending first and then doing first order autoregression has been used by Stock and Watson (1988) and Bierens (1991b).

Note that no matter what deterministic trend $f(t/T)$ is included in regression equation

(2.4.1), under the unit root hypothesis ΔX_t is assumed to be as in Assumption 2.2.1, that is, unit root without drift. If the unit root process has a drift,

(2.4.7) $$X_t = \mu + X_{t-1} + u_t, \qquad \mu \neq 0$$

then as a data generating process, Equation 2.4.7 specifies that,

$$f(t/T) \equiv 1,\ \theta = \mu,\ e_t = u_t,\ \rho = 1$$

but (2.4.7) can be rewritten as,

(2.4.7)′ $$X_t = X_0 + \mu t + \Sigma_{j=1}^{t} u_j.$$

Therefore the deterministic component is a linear trend (with random intercept X_0). It is crucial for 2.4.3 that X_{t-1}^{*} does not have any deterministic trend; therefore adequate detrending is a prerequisite for valid unit root test, cf. Schmidt (1988), Campbell and Perron (1990). As a regression equation, Equation 2.4.7 should account for the deterministic linear trend.

It is trivial that both $\hat{\rho}$ and $\tilde{\rho}$ are invariant to the value of θ in Equation 2.4.7. So the test in the theorem is also a test for the unit root process,

$$X_t = f(t/T)\theta + \Sigma_{j=1}^{t} u_j$$

against the trend stationary process,

$$X_t = f(t/T)\theta + u_t.$$

2.5 Review of the Existing Unit Root Tests

If Box and Jenkins made the time series modeling into an industry, then Dickey and Fuller made unit root testing into a profession. Dickey (1976), Fuller (1976) and Dickey and Fuller (1979) statistically formalized the notation of a unit root, and proposed several statistical tests for unit root. In the jargon of Box-Jenkins, the null model is that y_t has a unit root, i. e. y_t is ARIMA(0,1,0), and the alternative model is that y_t is stationary, i. e. y_t is ARIMA(1,0,0). The two test statistics Dickey and Fuller proposed are the normalized bias $T(\hat{\rho} - 1)$ and the t statistic for $\rho = 1$, call it τ. The asymptotic distribution is derived under the assumption of u_t being IID normal, which is very restrictive. As Dickey and Fuller (1979) have shown, the asymptotic distributions of the test statistics are non-normal but they can be expressed as functionals of a Brownian motion. Thus they have to be calculated by simulation. The critical values can be found in Fuller (1976).

To test for the null of ARIMA(p-1,1,0) against the alternative of ARIMA(p,0,0), the regression equation needs to be augmented with (p-1) lagged differences in y_t. The augmented Dickey-Fuller (ADF) τ statistic still follows the same asymptotic distribution as in the simple IID case. To test the null of ARIMA(p-1,1,q) against the alternative of ARIMA(p,0,q), the mixed ARMA process needs to be approximated by an infinite order AR process, so the number of augmenting lagged differences needs to increase with the number of observations at a controlled rate. The corresponding ADF τ statistic has the same asymptotic distribution as in the simple IID case, cf. Said and Dickey (1984).

The results of the ADF test may be contradictory with different numbers of lagged differences. The power of the test may also be substantially reduced if a large number of augmented terms are included. Therefore there is a need for an alternative testing strategy that is invariant for a general stationary process.

Our review of the literature on unit root tests is aimed at providing a basis for developing new tests, rather than aimed at providing a guideline for empirical application. For the latter purpose, see the excellent reviews by Diebold and Rudebusch (1990), Jolado, Jenkins and Sosvilla-Rivero (1990), and Campbell and Perron (1991).

2.5.1 Phillips and Perron's Extension of the Dickey-Fuller Test

If u_t is a general stationary (or mixing) process, then the asymptotic distribution of the Dickey-Fuller test depend on the unknown correlation structure of u_t, cf. Theorem 2.4.1.

Specifically it depends on the unknown long run variance of u_t which in turn depends on the sum of autocovariances of u_t at all lags. Phillips (1987), and Phillips and Perron (1988) nonparametrically corrected the Dickey-Fuller statistics for the unknown correlation structure of u_t. Thus their test for unit root can be conducted using the Dickey-Fuller statistics as if u_t were IID normal. The nonparametric correction uses results in statistics on estimating the spectral density at frequency zero. More importantly, Phillips and Perron used the theory of weak convergence in the proof of the asymptotic results, substantially relaxing the unduely restrictive assumptions of Dickey and Fuller. Through these and later works of Phillips and coauthors, the method of the theory of weak convergence became popular in theoretical econometrics. Bierens (1992b) provides a step by step treatment of the theory of weak convergence.

The asymptotic distributions of the Phillips and Phillips and Perron test are simplifications of those in Theorem 2.4.1.

2.5.1.A Examples

(i) For the Phillips (1987) test, the regression equation is,

$$X_t = \rho X_{t-1} + e_t.$$

The asymptotic distributions are as in Theorem 2.4.1 with $W^* = W$ and $\eta = 0$.

(ii) For the Phillips and Perron (1988) test, if the regression equation includes only an intercept,

$$X_t = a + \rho X_{t-1} + e_t$$

then

$f(r) = 1$, $\quad W^* = W - \int W$ is the demeaned Brownian motion, $\quad \eta = W(1)\int W$.

(iii) For the Phillips and Perron (1988) test, if the regression equation also includes a linear trend,

$$X_t = a + b(t/T) + \rho X_{t-1} + e_t$$

then

$$f(r) = (1\ r), \qquad W^* = W - (4\text{-}6r)\int W - (12\text{-}6r)\int rW$$

$$\eta = 12[\int rW\text{-}(1/2)\int W][(1/2)W(1)\text{-}\int W]+W(1)\int W.$$

W^* above is called the detrended Brownian motion.

(iv) For the Ouliaris, Park and Phillips (1988) test, the regression equation includes a time polynomial of order k,

$$X_t = b_0 + b_1(t/T) + \cdots + b_k(t/T)^k + \rho X_{t-1} + e_t$$

then

$$f(r) = (1\ r \cdots r^k), \qquad W^* \text{ and } \eta \text{ are as given in the theorem.}$$

For this class of tests, the nuisance parameter σ^2 in λ is consistently estimated by one of the kernel methods, among which the Newey and West (1987) estimator is very popular. For detailed treatment of the estimation of σ^2, see Chapter 3.

(v) For the Perron (1989) test, the regression equation includes a break in the intercept,

$$X_t = b_0 + c_0 D_t(\tau) + b_1(t/T) + e_t$$

where $D_t(\tau) = 0$ if $t < [T\tau]$, and 1 if $t \geq [T\tau]$. The function f in Theorem 2.4.1 is given by $f(r) = (1, I(r \geq \tau), r)$. The cases of a break in the slope or in both the intercept and the slope can be treated similarly.

2.5.1.B Limitations of the Traditional Approach to Unit Root Testing

(i) The asymptotic distribution is nonstandard.

The asymptotic distribution involves functionals of Brownian motion, and therefore is nonstandard. The asymptotic distribution has to be calculated by simulation. Given the available computing technology, this is of course not a big problem.

(ii) The asymptotic distribution depends on the alternative model.

Interestingly, the asymptotic distribution of the test statistics depends on the alternative model against which the unit root model is tested, as is shown in the previous subsection. The following models have been investigated as alternative models to the unit root model and the corresponding asymptotic distributions have been calculated by simulation.

(a) (i) Stationarity with zero mean, (ii) stationarity with nonzero mean, and (iii) linear trend stationarity; Fuller (1976) has three sets of critical values.

(b) Stationarity around time polynomial; Ouliaris, Park and Phillips (1990) have four sets of critical values for a (2^{nd}, 3^{rd}, 4^{th}, 5^{th}) order polynomial of time.

(c) Stationarity around breaking trend; Perron (1989) has twenty seven sets of critical values for the cases of a single break in the intercept, a break in the slope of the linear trend, and a break in both the intercept and the slope of the linear trend for nine locations of the break.

(d) More flexible nonlinear trend; Bierens (1992a) has 80 sets of critical values for the case of trend being approximated by Chebishev polynomials of order up to 40.

(iii) Unit root is taken as the null model.

The unit root model has been taken as the null model. As a test for economic theory which suggests the presence of unit root behavior against alternative theories, it puts alternative theories at an extreme disadvantage. As an extension to the unit root testing, the test for cointegration also takes no-cointegration as the null model. As a test for some economic theory which suggests cointegration among economic time series, the test takes as the null hypothesis that the theory is false. This is counter-intuitive, since before starting out testing the theory, the investigator must believe that there is some truth in the theory, otherwise he would not bother testing it with real data. Of course, given the problems with the power of any test, there is some conceptual problem with taking any model as the null model. This is the problem with the classical approach to inference.

(iv) The power of the tests is low.

Simulations show that the Dickey-Fuller test and the Phillips-Perron test have low power to distinguish the stationary processes with large AR components from a unit root process, and have substantial size distortion for the unit root processes with large MA components, cf. Schwert (1989).

(v) The estimate of the long run variance is needed.

The long run variance is usually estimated using results on estimation of spectral density at the origin with various window choices, the most popular of which are the Bartlett window, the Parzen window, and the Quadratic window, cf. Priestley (1981). Some of the methods guarantee nonnegativity of the estimate. For econometric applications, see Newey and West (1987), and Andrews (1991). Since the long run variance depends on the sum of autocovariances at all lags, given limited data, some truncation has to be made. The accumulation of the higher order autocovariances that are truncated may still be substantially large for some stationary processes. Although the estimate is heteroskedasticity and autocorrelation consistent, the convergence to the probability limit can be slow, creating problems in finite samples.

2.5.2 Test for Unit Root by Variable Addition

2.5.2A The Variable Addition Approach and the χ^2 Test of Park

Variable addition has been used as a model specification test. Specifically the residual from the regression model is regressed on some variables ("the added variables") in the information set, and then the significance of the coefficients on the added variables is tested. Significance of the added variables indicates model misspecification, cf. Pagan (1984). However insignificance of the coefficients on the added variables does not rule out other forms of model misspecification.

Due to the difference in the persistence of shocks to stationary and nonstationary processes, the cross moments between the time series and the time trend have different orders of magnitude under stationarity and unit root, cf. Theorems 2.2.5 and 2.2.6. If the time series of interest is regressed on some superfluous (irrelevant) trending regressors, for example a time trend:

(2.5.1) $$X_t = at + e_t,$$

the OLS estimate of parameter a is,

(2.5.2) $$\hat{a} = \Sigma X_t t / \Sigma t^2.$$

If X_t satisfies Assumption 2.2.1, then $\hat{a}$ is $T^{3/2}$-consistent for $a = 0$; if ΔX_t satisfies Assumption 2.2.1, then $\hat{a}$ is $T^{1/2}$-consistent for $a = 0$. So the OLS estimate of the superfluous coefficients behaves differently under stationarity and a unit root. Park (1990) proposes to use the properly standardized F-test for the insignificance of the superfluous coefficients as a test for unit root. When properly adjusted for nuisance parameters as Phillips and Perron (1987) did, the resulting F-test has an asymptotic χ^2 distribution under the null of stationarity.

Let Y_t be generated by the following model, (2.5.3) $$Y_t = b_0 + b_1(t/T) + \cdots + b_k(t/T)^k + X_t.$$

If we run the following regression,

(2.5.4) $$Y_t = b_0 + b_1(t/T) + \cdots + b_k(t/T)^k + b_{k+1}(t/T)^{k+1} + \cdots + b_{k+q}(t/T)^{k+q} + e_t$$

then the regressors $((t/T)^{k+1}, \cdots, (t/T)^{k+q})$ are superfluous. Let $\hat{b}$ be the OLS estimate of the parameters, F be the regression F statistic or the Wald statistic for the insignificance of $(b_{k+1}, \cdots, b_{k+q})'$. Under the unit root hypothesis, ΔX_t satisfies Assumption 2.2.1,

$$(1/\sqrt{T})(\hat{b}_{k+1}, \cdots, \hat{b}_{k+q})' \Rightarrow \sigma M^{*-1} \int g^*(r)'W$$

(2.5.5) $$(1/T)qF \Rightarrow (\int W^{*2} - \int W^{**2}) / \int W^{**2},$$

where

$g^*(r)$ is the residual of a linear projection of $g(r) = (r^{k+1}, r^{k+2}, \cdots, r^{k+q})$ on $f(r)$,

$f(r) = (1, r, \cdots, r^k)$,

W^* is the Brownian motion detrended with $f(r)$,

W^{**} is the Brownian motion detrended with $(f(r)\ g(r))$,

$M^* = \int_0^1 g^{*\prime} g^*$.

If X_t is stationary, then

$$\sqrt{T}(\hat{b}_{k+1}, \cdots, \hat{b}_{k+q})' \Rightarrow \sigma M^{*-1} \int g^*(r)'dW \sim N(0, \sigma^2 M^{*-1}),$$

(2.5.6) $$qF = (\hat{b}_{k+1}, \cdots, \hat{b}_{k+q})[\hat{\sigma}_e^{-2} \Sigma g^*(t/T)'g^*(t/T)](b_{k+1}, \cdots, b_{k+q})' \Rightarrow (\sigma^2/\sigma_u^2)\chi_q^2,$$

where $\hat{\sigma}_e^2$ is the OLS estimate of the variance of e_t.

Taking the stationarity as the null hypothesis, the test statistic is defined as

$$P_T = (\hat{\sigma}_e^2/\hat{\sigma}^2)qF.$$

Then under stationarity, $P_T \Rightarrow \chi_q^2$. While under the unit root, $\hat{\sigma}^2 = O_p(m_T)$ where m_T is the lag truncation parameter in the long run variance estimation; therefore, $(m_T/T)P_T = O_p(1)$. Thus the P_T test is consistent at rate (T/m_T).

The variable addition approach of Park can be summarized as follows: (i) the asymptotic distribution under stationarity is well known; (ii) stationarity can be taken as the null model; (iii) the specification of deterministic trend can be very flexible; but (iv) the construction of the test requires spectral estimation of the long run variance and the asymptotic power is affected.

2.5.2B The Variable Addition Approach and the Cauchy Test of Bierens (1991a)

The OLS estimate $\hat{a}$ of the superfluous time trend in (2.5.1) is asymptotically normal with variance proportional to the long run variance, but the rate of convergence to the asymptotic distribution is different under stationarity and unit root. If another quantity can be constructed that also has an asymptotic normal distribution with a variance that depends on the long run variance proportionally, then the quotient of the two asymptotic normal quantities, after orthogonalization and standardization, is asymptotically standard Cauchy. This is the approach adopted by Bierens (1991a). Bierens' approach shares the first three features of Park's approach. But the advantage of the Bierens' approach is that it does not require the estimation of the long run variance.

2.5.3 Higher Order Autocorrelation and Unit Root

If the time series has a unit root, then the autocorrelation function is equal to one at all lags. Hasza (1980) derived the asymptotic distribution of the sample autocorrelation at finite lags for finite order ARIMA process. Bierens (1991b) extended the results of Hasza to the case where the first difference of the series is a mixing process. Let the regression equation be,

$$X_t = \rho X_{t-p} + u_t.$$

For fixed and finite p, $T(\hat{\rho}-1)$ is similar to the Dickey-Fuller test statistic for unit root, but its asymptotic distribution depends on the unknown truncated long run variance when u_t is not a martingale difference series. In order to get rid of the nuisance parameter, the lag length p in the regression is allowed to increase with the sample size T at a controlled rate, i. e. $p = o(T^{1/3})$. Bierens used the results to construct a test for unit root which is asymptotically more powerful than the Phillips and Perron tests. The asymptotic distribution is the same as that of the Dickey-Fuller test. Both the case of an intercept and the case of a linear trend are treated.

In the context of the variance ratio test for random walks, Richardson and Stock (1989) derived asymptotic theory when the lag length increases proportionally to the sample size, $p = O(T)$. However they noted that in this case, their test statistic is inconsistent. Separately Bierens has indicated that it is possible to construct a consistent unit root test with $p = O(T)$.[3]

2.5.4 Nonlinear Trend Stationarity and Unit Root

In Bierens (1992a), the nonlinear trend is approximated by a set of Chebishev polynomials introduced in Section 2.3. First the series is detrended by linear time trend, and then it is detrended by a weak nonlinear trend which is approximated by a linear combination of a set of Chebishev polynomials. Finally the test statistic is constructed using the detrended series.

2.5.5 Discussion

Theorem 2.4.1 states that the asymptotic distribution for unit root test can be conveniently characterized as a family of distributions. In the case of testing for unit root against stationarity with a zero mean, the asymptotic distributions are functionals of Brownian

[3]Private communication.

motion. In the case of stationarity with a nonzero mean or the case of trend stationarity, the asymptotic distributions can be obtained simply by replacing the Brownian motion with the demeaned or detrended Brownian motion.

For general mixing processes, all the distributions in unit root testing rely on asymptotic theory. Each different test relies on a different aspect of the asymptotic theory. The finite sample properties of each test depend on the quality of approximation of that aspect of the asymptotic theory that the test relies on. A good test should have small or no size distortion, good power, and be easy to calculate and convenient to use.

CHAPTER 3

SPURIOUS REGRESSION, PARAMETER INSTABILITY AND UNIT ROOT

3.1 Introduction

Regression estimates of parameters involving stationary processes are consistent under certain conditions. Regression estimates of parameters involving unit root processes are spurious except in the case of cointegration, cf. Phillips (1986). Under certain conditions, the spurious estimates of the parameters converge to nondegenerate random variables, rather than to fixed constants. Therefore, tests for parameter instability also have power against the unit root model[4].

There has been a large literature on structural change, partly because the economic systems are under the attack of shocks of various nature, e. g. depression, war, surges in oil prices, and changes in weather. Prominent tests for parameter instability include the classical CUSUM test of Brown, Durbin and Evans (1975), the recent parameter fluctuation test of Ploberger, Kramer and Kontrus (1989, hence PKK), and the random walk test of Nabeya and Tanaka (1988). The CUSUM test relies on the residuals of the regression, the fluctuation test relies on a sequence of parameter estimates over the subsamples, while the random walk test relies on the Lagrangian Multiplier principle. Kwiatkowski, Phillips, Schmidt and Shin (1992) have developed a test for stationarity similar to the random walk test for parameter stability.

In this chapter, we develop tests for stationarity using the CUSUM test and the parameter fluctuation test. Note that the power of our test relies on the fact that under the alternative of unit root, the parameter estimates behave as if the true parameters were unstable.

[4] I thank Nathan Balke for this observation.

Therefore, our test does not distinguish between true parameter instability and unit root, e. g. Perron (1988, 1989). In order to distinguish breaking trend from unit root, the breaking trend should be included in the regression as part of the maintained assumption. The proofs for the theorems in section 3.2 are straightforward applications of the theorems in Chapter 2, and thus they are omitted.

3.2 Spurious Regression, Parameter Instability and Unit Root

3.2.1 Introduction

The null model is that the series Y_t is stationary,

(3.2.1a) $$H_0\colon Y_t = u_t,\ u_t \sim \text{Assumption (2.2.1)}$$

and the alternative model is that Y_t has a unit root,

(3.2.1b) $$H_1\colon \Delta Y_t \sim \text{Assumption (2.2.1)}.$$

From the invariance principle, the quasi-sample mean $\bar{Y}(\tau) = [T\tau]^{-1}\Sigma_{t=1}^{[T\tau]}Y_t$ behaves differently under H_0 and H_1,

(3.2.2a) $$\text{under } H_0,\ T^{1/2}\bar{Y}(\tau) \Rightarrow \sigma\tau^{-1}W(\tau),$$

(3.2.2b) $$\text{under } H_1,\ T^{-1/2}\bar{Y}(\tau) \Rightarrow \sigma\tau^{-1}\textstyle\int_0^\tau W.$$

Therefore under stationarity, $T^{-1/2}\bar{Y}(\tau)$ approaches zero, while under H_1 it converges to a nondegenerate normal random variable. Define

(3.2.3) $$S(\tau) = \frac{[T\tau]}{\sigma\sqrt{T}}[\bar{Y}(\tau) - \bar{Y}(1)].$$

Then we have Theorem 3.2.1.

Theorem 3.2.1

(i) Under H_0, $S(\tau) \Rightarrow W(\tau) - \tau W(1) \equiv W^0(\tau)$, the Brownian Bridge;

(ii) under H_1, $S(\tau)/T \Rightarrow \int_0^\tau W - \tau \int_0^1 W$.

Note that $S(\tau)$ is invariant to a linear transformation of Y_t, so we could have included an intercept in equation (3.2.1a). The nuisance parameter σ^2 needs to be estimated consistently. One choice is the Newey-West estimate $\hat{\sigma}^2$ defined as,

$$\hat{\sigma}^2 = (1/T)\Sigma_{j=-(m_T-1)}^{m_T-1} k(j/m_T)\Sigma_{t=|j|+1}^{T} u_t u_{t-|j|} \quad (3.2.4)$$

where $k(j/m_T) = j/m_T$, $m_T \to \infty$ as $T \to \infty$, but $m_T = o(T^{1/2})$. Note that the controlled rate at which m_T approaches infinity can be greater than the original Newey-West rate, cf. Bierens (1992b). Denote the resulting test statistic as $\hat{S}(\tau)$. We have,

Theorem 3.2.2

(i) Under H_0, $\hat{S}(\tau) \Rightarrow W(\tau) - \tau W(1) \equiv W^0(\tau)$, the Brownian Bridge;

(ii) under H_1, $\hat{S}(\tau)/(T/m_T) \Rightarrow \int_0^\tau W - \tau \int_0^1 W$.

After regressing Y_t on an intercept, the statistic $\sup_{\tau \in [0,1]} |\hat{S}(\tau)|$ is the CUSUM test using OLS residuals. The statistic is also the fluctuation test of the regression parameter, since the two tests are the same with an intercept as the only regressor. As a test statistic, other functions of $\hat{S}(\tau)$ are also admissible. For example, $(\sup_{\tau \in [0,1]} \hat{S}(\tau) - \inf_{\tau \in [0,1]} \hat{S}(\tau))$, cf. Lo (1991). Note that due to the estimation of σ^2, the $\hat{S}(\tau)$ test is consistent at the rate (T/m_T), instead of at the rate T.

Park (1990) and Bierens (1991a) explored the variable addition approach to test for stationarity. The variable addition approach recognizes the superfluous nature of the variable addition (the true value of the parameters being zeros), but the parameter estimates on the maintained trends cannot be used in constructing the test. Here the approach using the parameter stability test only recognizes the stability of the parameter, be it zero or not. The use of fewer (or no) superfluous regressors may improve the power of the test. Traditional tests for unit root only use the full sample result, while the parameter stability tests uses the information in the fluctuation of the underlying process within the sample. Therefore the parameter stability test may have superior power.

3.2.2 The CUSUM Test Using OLS Residuals

3.2.2A. The Test Statistic

The most interesting competing models for most macroeconomic time series are trend stationarity versus unit root with drift. In this section, we construct a test for stationarity with a general deterministic time trend. Let the models of interest be,

(3.2.5) H_0: $Y_t = f(t/T)b + u_t,\ u_t \sim$ Assumption (2.2.1)

(3.2.6) H_1: $Y_t = f(t/T)b + X_t,\ \Delta X_t \sim$ Assumption (2.2.1).

Let e_t be the OLS residuals of the following regression equation,

$$y_t = f(t/T)\hat{b} + e_t.$$

Define

(3.2.7) $$\xi(\tau) = T^{-1}\Sigma_{t=1}^{[T\tau]}e_t.$$

Then we have

Theorem 3.2.3

(i) Under H_0, $T^{1/2}\xi(\tau) \Rightarrow \sigma\left[W(\tau) - [\int_0^\tau f][\int_0^1 f'f]^{-1}[\int_0^1 f'dW]\right]$;

(ii) under H_1, $T^{-1/2}\xi(\tau) \Rightarrow \sigma\left[\int_0^\tau W - [\int_0^\tau f][\int_0^1 f'f]^{-1}[\int_0^1 f'W]\right]$.

Note that if f(t/T) = 1 (an intercept), this theorem collapses to Theorem 3.2.1. A test for stationarity can be based on functionals of $(\sqrt{T}\xi(\tau)/\hat{\sigma})$, e. g. the CUSUM type of test[5]

$$CU = \max_{r \in [\tau, 1-\tau]} |(\sqrt{T}\xi(r)/\hat{\sigma}|,$$

or the range over standard deviation (R/S) test,

$$RS = \left[\max_{r \in [\tau, 1-\tau]} \sqrt{T}\xi(r)/\hat{\sigma} - \min_{r \in [\tau, 1-\tau]} \sqrt{T}\xi(r)/\hat{\sigma}\right],$$

where τ is called the trimming parameter. We will study the properties of both tests.

3.2.2B. Simulated Critical Values

Table 3.2.1 is the simulated asymptotic distribution for the cases of no trend and a linear trend respectively. The simulated null distribution under linear trend stationarity lies to the left of that under nonzero mean stationarity. Formally, the null distribution under nonzero mean stochastically dominates (FSD) that under linear trend in the first degree. Specifically, for the case of an intercept only in the regression and no trimming ($\tau = 0$),

$$P(CU \leq 1.207) = 0.90,\ P(CU \leq 1.333) = 0.95,\ P(CU \leq 1.605) = 0.99,$$

$$P(RS \leq 1.563) = 0.90,\ P(RS \leq 1.697) = 0.95,\ P(RS \leq 1.960) = 0.99;$$

for the case of a linear trend and no trimming,

$$P(CU \leq 0.810) = 0.90,\ P(CU \leq 0.881) = 0.95,\ P(CU \leq 1.019) = 0.99,$$

$$P(RS \leq 1.455) = 0.90,\ P(RS \leq 1.572) = 0.95,\ P(RS \leq 1.828) = 0.99.$$

3.2.2C. Simulation Results

The simulation results on the empirical size and power of the CU and RS tests with an intercept are collected in Tables 3.2.2. For the case of white noise ($\rho = 0$ and $\theta = 0$), the sizes of both tests are close to the nominal size. But as the value of ρ increases, there is more and

[5]Stock (1992) uses a class of transformations of the CUSUM statistic in designing a Bayesian rule for deciding whether a series is I(0) or I(1). The CUSUM statistic, as a test for stationarity, is mentioned but not pursued. The CUSUM test also serves as a vehicle for experimenting with the new long run variance estimator to be introduced in the next section.

more overrejection if the Newey-West estimator is used, but there is not enough rejection if the QS estimator (with prewhitening\recoloring) is used. The reason is that the QS estimator with prewhitening\recoloring is more precise; the long run variance estimate used in the denominator is large if ρ is large; therefore, the test statistics are small in absolute value. Since the Newey-West estimator has a window width/truncation parameter, the long run variance is underestimated when ρ is large; therefore, the test statistics are large in absolute value. Because of this underestimation, the power of the tests is fairly good using the Newey-West estimator, but very poor using the QS estimator.

Table 3.2.3 contains the simulation results for the case of a linear trend. The results are similar to the case of an intercept; the above discussion still applies here.

3.2.3 The Fluctuation of the Recursive Estimate of Trend Coefficients

3.2.3A. The Test Statistic

Let $b(\tau)$ be the OLS estimate of b using only the first $[T\tau]$ observations, then

$$b(\tau) - b = [\Sigma_{t=1}^{[T\tau]} f'(t/T)f(t/T)]^{-1}[\Sigma_{t=1}^{[T\tau]} f'(t/T)e_t] \tag{3.2.8}$$

where $e_t = u_t$ under H_0 and $e_t = X_t$ under H_1. We have,

Theorem 3.2.4

(i) Under H_0, $\sqrt{T}(b(\tau) - b) \Rightarrow \sigma[\int_0^\tau f'f]^{-1}[\int_0^\tau f'dW]$;

(ii) Under H_1, $(1/\sqrt{T})(b(\tau) - b) \Rightarrow \sigma[\int_0^\tau f'f]^{-1}[\int_0^\tau f'W]$.

So under H_0, $b(\tau)$ is consistent for b. As in PKK (1989), our test statistic will be based on functionals of the following quantity,

$$\hat{S}(\tau) = \frac{[T\tau]}{T\hat{\sigma}} [\Sigma_{t=1}^{T} f'(t/T)f(t/T)]^{1/2}[b(\tau) - b(1)],\ \tau \in (0, 1).$$

Its asymptotic distribution is given below.

Theorem 3.2.5

(i) Under H_0, $\hat{S}(\tau) \Rightarrow \tau\,[\int_0^1 f'f]^{1/2}\Big[[\int_0^\tau f'f]^{-1}[\int_0^\tau f'dW] - [\int_0^1 f'f]^{-1}[\int_0^1 f'dW]\Big]$;

(ii) Under H_1, $\hat{S}(\tau)/(T/m_T) \Rightarrow \tau\,[\int_0^1 f'f]^{1/2}\Big[[\int_0^\tau f'f]^{-1}[\int_0^\tau f'W] - [\int_0^1 f'f]^{-1}[\int_0^1 f'W]\Big]$.

For a linear trend, f(r) = (1, r). If the square root of a matrix is taken as its Choleski decomposition, then the asymptotic distribution under H_0 is

$$12\begin{bmatrix} 1 & 0 \\ 1/2 & 1/\sqrt{12} \end{bmatrix}\left\{\begin{bmatrix} (1/3)W(\tau) - (2\tau)^{-1}\int_0^\tau r dW \\ -(2\tau)^{-1}W(\tau) + \tau^{-2}\int_0^\tau r dW \end{bmatrix} - \tau\begin{bmatrix} (1/3)W(1) - (1/2)\int_0^1 r dW \\ -(1/2)W(1) + \int_0^1 r dW \end{bmatrix}\right\}$$

(3.2.9) $\equiv 12A[B(\tau)-\tau B(1)]$

Remarks

(i) If f(t/T) = 1 includes only an intercept, then Theorem 3.2.5 collapses to Theorem 3.2.1.

(ii) Theorem 3.2.5 (i) extends the results in PKK (1989) to the case of trending data. They assume that,

(3.2.10) f(t/T) is either stochastic or deterministic

(3.2.11a) $\frac{1}{T}\Sigma_{t=1}^{T} f'(t/T)f(t/T) \rightarrow \int_0^1 f'f \equiv R$

(3.2.11b) $\frac{1}{[T\tau]}\Sigma_{t=1}^{[T\tau]} f'(t/T)f(t/T) \rightarrow R$

(3.2.12) $\frac{1}{\sqrt{T}}\Sigma_{t=1}^{[T\tau]} f(t/T)u_t \Rightarrow \sigma R^{1/2}W(\tau)$

So in our notation

(3.2.13) $$[\int_0^\tau f'f][\int_0^1 f'f]^{-1} = \tau I_{k+1}.$$

Then under H_0,

(3.2.14) $$\frac{1}{\sqrt{T}}[\Sigma_{t=1}^{[T\tau]} f'(t/T)f(t/T)]\ (b(\tau) - b(1)) \Rightarrow \sigma\ R^{1/2}\left[W(\tau) - \tau W(1)\right].$$

After standardization using (3.2.11) and (3.2.13),

(3.2.15) $$\frac{[T\tau]}{T}[\Sigma_{t=1}^{T} f'(t/T)f(t/T)]^{1/2}\ (b(\tau) - b(1)) \Rightarrow \sigma\ W^0(\tau)$$

cf. PKK (1989), equations (11) and (12). Therefore the asymptotic distribution does not depend on f, the regressors in the regression.

For Theorem 3.2.5, conditions (3.2.11b) and (3.2.13) are no longer true, since they also depend on the curvature of the function f. Therefore simplification of the asymptotic distribution to a Brownian Bridge as in (3.2.15) is no longer possible. Results in Theorem 3.2.5 typify the way in which the asymptotic distribution depends on the regressors in tests for unit root.

Note that $\hat{S}(\tau)$ is a random function in $D^k[0, 1]$. There are many ways to transform the random function in $D^k[0, 1]$ into a random variable. The norm in $\mathbb{R}^k$ transforms the random function in $D^k[0, 1]$ into random function in $D[0, 1]$. Many transformations can turn the random function in $D[0, 1]$ into a random variable. For example, for the first step, the Max norm and the Euclidean norm can be used; for the second step, the supremum, the integral over (0, 1) and the range can be used. PKK (1989) used the Max norm $\|S\|_\infty$ along with the supremum. We will examine the use of the Max norm $\|S\|_\infty$ and the Euclidean norm $\|S\|$ along with the supremum. The supremum and the range of $\|\hat{S}(\tau)\|$ over (0, 1) are the same since the function is positive (due to taking the norm) and the minimum is 0 (at $\tau = 1$). Our test statistics are denoted as

$$FL_1 = \sup_{r \in [\tau, 1-\tau]} \| \hat{S}(r) \|_\infty,$$

$$FL_2 = \sup_{r \in [\tau, 1-\tau]} \| \hat{S}(r) \|,$$

where τ is the trimming parameter.

Our fluctuation tests are related to the "direct" tests for changing trend of Chu and White (1992). Their tests are based on the same idea of parameter fluctuation over subsamples. While they test for the fluctuation in the intercept and the slope separately, we test for the overall fluctuation of the intercept and the slope as a whole. Stacking their tests for fluctuation in the intercept and the slope together into vector $\tilde{S}$, then $\tilde{S}$ can be defined similarly to $\hat{S}$, but with the matrix Choleski decomposition replaced by an appropriate diagonal scaling matrix.

3.2.3B. Simulated Critical Values

The asymptotic distribution under H_0 with a linear trend in Theorem 3.2.5 is calculated by simulation, and reported in Table 3.2.4. The critical values for no-trimming are substantially larger than those for 15% trimming. The reason for this can easily be seen from the asymptotic distribution (3.2.8). Take a realization $\hat{D}(\tau)$ from the asymptotic distribution. At the beginning of the sample, τ is small, hence τ^{-2} is very large, and the (2,1) element of $B(\tau)$, hence that of $B(\tau)-\tau B(1)$, is very large. Because the Choleski factor A is lower triangular, the very large (2,1) element of $B(\tau)$ makes the (2,1) element of $\hat{D}(\tau)$ very large. The very large (2,1) element of $\hat{D}(\tau)$ makes the Max norm and the Euclidean norm very large and approximately equal. But with some trimming (15% for example), τ^{-2} is no longer so large, therefore, the Max norm and the Euclidean norm are not large. But the above argument still makes their values close. This is reflected in the tabulation of critical values in Table 3.2.2. Specifically, for the case of a linear trend and no trimming,

$$P(FL_1 \leq 64.056) = 0.90, \; P(FL_1 \leq 75.772) = 0.95, \; P(FL_1 \leq 99.607) = 0.99,$$

$$P(FL_2 \leq 64.056) = 0.90,\ P(FL_2 \leq 75.773) = 0.95,\ P(FL_2 \leq 99.607) = 0.99;$$

for the case of a linear trend and 15% trimming ($\tau = 0.15$),

$$P(FL_1 \leq 4.618) = 0.90,\ P(FL_1 \leq 5.204) = 0.95,\ P(FL_1 \leq 6.382) = 0.99,$$

$$P(FL_2 \leq 4.930) = 0.90,\ P(FL_2 \leq 5.527) = 0.95,\ P(FL_2 \leq 6.741) = 0.99.$$

3.2.4 Distinguishing Deterministic Trend with Breaks from Stochastic Trend

As mentioned before, our proposed tests do not have any power to distinguish a break in the trend coefficients from a true unit root. If there is a break in the trend, the test statistics tend to have a larger value, thus resulting in the rejection of the null model. However, if the correct breaking trend is included in the regression and the series is stationary with a breaking trend, then the test statistics should have small values, thus not resulting in the rejection of the null. If an incorrect breaking trend is included in the regression, the test statistics should still be large. In reality, the location of the break is never known in advance. Most likely it is chosen by a data-based selection procedure, for example by a plot of the data. Based on the above analysis, the selection procedure for the break point can be based on a grid search over the break fraction (η), for example, for $\eta = 0.1, 0.2, \cdots, 0.9$. The value of η at which the CUSUM test statistic is the smallest is chosen as the true break fraction.

Once the break point is found, the time trend with break can be formed and the CUSUM test can be done as usual. Similarly to the Perron approach, the asymptotic distribution is incorrect if the endogeneity of the breaking point is not taken into account.

Table 3.2.5 tabulates the null distribution of the CUSUM test under the presence of a one-time shift in the mean. Table 3.2.6 tabulates the null distribution of the CUSUM test under the presence of a one-time shift in the intercept and the slope (the locations of the shifts for the mean and slope are the same). Specifically, for the case of an intercept only and no trimming, the minimum CUSUM statistic, denoted as mCU, has the following critical values,

$$P(mCU \leq 0.717) = 0.90,\ P(mCU \leq 0.776) = 0.95,\ P(mCU \leq 0.913) = 0.99;$$

for the case of a linear trend and trimming,

$$P(mCU \leq 0.510) = 0.90,\ P(mCU \leq 0.539) = 0.95,\ P(mCU \leq 0.606) = 0.99.$$

3.3 An Adaptive Estimator of the Long Run Variance

3.3.1 Introduction

The asymptotic distributions of our test for stationarity as well as that of Park (1990) and Bierens (1991a) all depend on the long run variance proportionally. In order for the test to be free of nuisance parameters, an estimate of the long run variance is needed. The long run variance is that of the series itself under stationarity, and that of its first difference under unit root. Therefore consistent estimation of the long run variance under both unit root and stationarity calls for a switch in the random variable used in the estimation. However such a switch is not necessary in the traditional test for unit root when the unit root model is taken as the null model, cf. Phillips and Perron (1988). With the stationarity as the null model, the estimate of the long run variance often uses a detrended version of the series in estimation, which is consistent under stationarity. But under the alternative of unit root, this detrended series still has a unit root, the estimate of the long run variance using this detrended series explodes with the sample size, so when this estimate appears in the denominator of the test statistic, the test is at best consistent at a rate less than the sample size.

In this section, we develop an adaptive consistent long run variance estimator. The consistency is proved for the class of kernels used in Andrews (1991). The series used in constructing our estimator is a weighted average of the series itself and its first difference, with the weight being a random variable. Our estimator is adaptive in that, under stationarity, the weight on the first difference approaches zero, and thus the level part dominates; while under unit root, the weight on the level approaches zero, and thus the first difference part dominates. Our estimator is consistent under both stationarity and unit root. Therefore, when the unit root is taken as the alternative model, the test that employs our long run variance estimate can be consistent at the rate of the sample size, providing more power in large samples.

3.3.2. An Adaptive Estimator of the Long Run Variance

To facilitate the proof of the theoretical results, we assume in this section that u_t satisfies assumption A in the appendix.

Let Y_t be the series of interest; $V_t(\theta)$ be a transformation of Y_t with parameter θ; $\hat{\theta}$ be an estimate of θ. For simplicity, we abstract from any deterministic trend in Y_t. We are interested in distinguishing two characterizations of Y_t: the null model of stationarity, cf equation (3.2.1a), and the alternative model of unit root, cf. equation (3.2.1b). As an example of the dependence of asymptotic distributions on σ^2, we take a look at Theorem (3.2.1). Note that the long run variance σ^2 involved is that of u_t. Therefore, in order to consistently estimate σ^2, the random variable used in the estimation should be Y_t under (3.2.1a) and ΔY_t under (3.2.1b). It seems that consistent estimation of σ^2 requires knowledge of (3.2.1a) and (3.2.1b). But distinguishing the two characterizations of Y_t is exactly the purpose of the exercise. The random variable used in the estimation of σ^2 can be a function of Y_t such that asymptotically it is close to Y_t under stationarity and close to ΔY_t under the unit root. For this purpose, define the adaptive estimate of u_t

$$V_t(\hat{\theta}_1) = (1-\hat{\theta}_1)Y_t + \hat{\theta}_1 \Delta Y_t = Y_t - \hat{\theta}_1 Y_{t-1} \qquad (3.3.1)$$

which is a weighted average of Y_t and ΔY_t. The random weight $\hat{\theta}_1$ satisfies the following,

Assumption 3.3.1

(i) Under H_0, $\hat{\theta}_1 = o_p(T^{-1/2})$;

(ii) Under H_1, $1-\hat{\theta}_1 = o_p(T^{-3/2})$.

For the ease of exposition, define the true value of θ_1 to be $\theta_{10} = 0$ under H_0, $\theta_{10} = 1$ under H_1. Then Assumption (3.3.1) simply requires that $\hat{\theta}_1$ be consistent for θ_{10} at a certain rate. In large samples, $V_t(\hat{\theta}_1)$ always extracts the stationary part of Y_t. Define,

(3.3.2) $$J_T = \Sigma_{j=-(T-1)}^{T-1}\Gamma(j)$$

where $\Gamma(j) = E(V_t V_{t+j})$. Then $\sigma^2 = \lim_{T\to\infty} J_T$. Given u_t (= $V_t(\theta_0)$), the kernel estimate of σ^2 is,

(3.3.3) $$J_T(\theta_0) = \Sigma_{j=-(T-1)}^{T-1} k(j/m_T)\hat{\Gamma}(j,\theta_0)$$

where m_T is the bandwidth parameter, or what is commonly known as the lag truncation parameter,

$k(j/m_T)$ is the weighting function for the sample autocorrelation function at lag j,

$$\hat{\Gamma}(j,\theta_0) = \tfrac{1}{T}\Sigma_{t=1}^{T-j} V_t(\theta_0)V_{t+j}(\theta_0) \text{ for } j \geq 0, \text{ and } \hat{\Gamma}(j) = \hat{\Gamma}(-j) \text{ for } j < 0.$$

The difference between J_T and $J_T(\theta_0)$ are the weighting function $k(j/m_T)$ and the error involved in approximating the expectation by the sample quantity. The class of kernels k_1 are as defined in Andrews (1991),

$$k_1 = \Big\{ k(\cdot): \mathbb{R} \mapsto [-1,1] \mid k(0) = 1,\ k(x)=k(-x),\ \forall x \in \mathbb{R},\ \textstyle\int_{-\infty}^{\infty}|k(x)|dx < \infty,$$
$$k(x) \text{ is continuous at } x = 0, \text{ and at all but a finite number of other points} \Big\}$$

For example, for the Bartlett estimator, $k(x) = 1 - |x|$, $|x| \leq 1$; 0, otherwise. If u_t is not observable, then a model has to be estimated. With an estimate of θ, say $\hat{\theta}$, we get an estimate of u_t, say $V_t(\hat{\theta})$. Define,

(3.3.4) $$J_T(\hat{\theta}) = \Sigma_{j=-(T-1)}^{T-1} k(j/m_T)\hat{\Gamma}(j,\hat{\theta})$$

The difference between $J_T(\hat{\theta})$ and $J_T(\theta_0)$ is the estimation error in $\hat{\theta}$. The adaptive estimate (3.3.1) can be viewed in this way. If the null model for Y_t is trend stationarity,

(3.3.5) $$H_0: Y_t = \mu + \beta(t/T) + u_t \equiv x_t\theta_2 + u_t$$

and the alternative model is unit root with drift

(3.3.6) $$H_1: Y_t = Y_0 + \mu(t/T) + \Sigma_{j=1}^{t} u_j \equiv x_t\theta_2 + \Sigma_{j=1}^{t} u_j$$

then in order to construct an appropriate unit root test, Y_t has to be detrended. Note that now we are using x_t to denote the deterministic trend, instead of a realization of a stochastic process. The deterministic trend x_t in the above equation has the same order of magnitude as a stationary variable. Denote the detrending parameters as θ_2, and its estimate as $\hat{\theta}_2$.

Assumption 3.3.2

(3.3.7) $$\text{(i) } \sqrt{T}(\hat{\theta}_2 - \theta_{20}) = O_p(1) \text{ under } H_0,$$

(3.3.8) $$\text{(ii) } (1/\sqrt{T})(\hat{\theta}_2 - \theta_{20}) = O_p(1) \text{ under } H_1.$$

For the OLS estimate of θ_2, Assumption (3.3.2) is guaranteed by an argument similar to Theorem 3.2.3. Similarly to (3.3.1), we define,

(3.3.9) $$V_t(\theta) = (Y_t - x_t\theta_2) - \theta_1(Y_{t-1} - x_{t-1}\theta_2)$$

When $\hat{\theta}$ is consistent for θ, under H_0 for example, then (3.3.4) is consistent for σ^2, cf. Newey and West (1987), Andrews (1991), among others. However when $\hat{\theta}_2$ is not consistent for θ_2, under H_1 for example, (3.3.4) might still be consistent for σ^2 if $\hat{\theta}_1$ is "sufficiently" consistent for θ_1. Proof of the consistency of estimate for σ^2 in the literature has been using a Taylor series expansion approach to bound the difference between $\hat{\Gamma}(j,\hat{\theta})$ and $\hat{\Gamma}(j,\theta_0)$. From (3.3.10),

$$(3.3.10)\qquad \frac{\partial V_t(\theta)}{\partial\theta} = -\begin{bmatrix} y_{t-1} - x_{t-1}\theta_2 \\ (x_t - x_{t-1}\theta_1)' \end{bmatrix} \quad \overset{\theta=\theta_0}{\underset{H_1}{=}} \quad \begin{bmatrix} -\Sigma_{j=1}^{t-1} u_j \\ -(x_t - x_{t-1})' \end{bmatrix}$$

$$(3.3.11)\qquad \frac{\partial^2 V_t(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} 0 & x_{t-1} \\ x_{t-1}' & 0 \end{bmatrix}.$$

The two-term Taylor series expansion for $V_t(\theta)$ is,

$$(3.3.12)\qquad V_t(\theta) = V_t(\theta_0) + [\partial V_t(\theta_0)/\partial\theta](\theta-\theta_0) + (1/2)(\theta-\theta_0)'[\partial^2 V_t(\bar{\theta})/\partial\theta\partial\theta'](\theta-\theta_0)$$

where $\bar{\theta} = \theta_0 + \alpha(\theta-\theta_0)$, $\alpha \in [0, 1]$. So although the first element of the first order derivative, $\partial V_t(\theta)/\partial\theta$, does not satisfy the uniform bounding condition of Assumption B(ii) of Andrews (1991) or Theorem 2(ii) of Newey and West (1987) in the Taylor series expansion, the strong consistency of $\hat{\theta}_1$ in Assumption (3.3.1)(ii) compensates for this unboundedness. The second order derivative, $\partial^2 V_t(\theta)/\partial\theta\partial\theta'$, does not depend on θ; it satisfies the uniform bounding condition, of Assumption C(ii) of Andrews (1991). Its main diagonal elements are equal to zeros, so the second order term in the Taylor series expansion involves only the cross product term, $(\hat{\theta}_1 - \theta_{10})(\hat{\theta}_2 - \theta_{20})$; here the strong consistency in $\hat{\theta}_1$ again compensates for the inconsistency of $\hat{\theta}_2$.

The above analysis suggests that (3.3.12) provide a consistent estimate of σ^2, which is indeed the case, as proved in the following theorem.

Theorem 3.3.1 Let u_t satisfy Assumption A in Appendix A, $\hat{\theta}_1$ satisfy Assumption (3.3.1), and $\hat{\theta}_2$ satisfy Assumption (3.3.2), $k(\cdot) \in k_1$, $m_T \to \infty$.

(i) If $m_T^2/T \to 0$, then $J_T(\hat{\theta}) - J_T = o_p(1)$, and $J_T(\hat{\theta}) - J_T(\theta_0) = o_p(1)$.

(ii) If $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which k_q, $|f^{(q)}| < \infty$, then $\sqrt{T/m_T}(J_T(\hat{\theta}) - J_T) = O_p(1)$, and $\sqrt{T/m_T}(J_T(\hat{\theta}) - J_T(\theta_0)) = o_p(1)$, where k_q, $f^{(q)}$ are defined in the appendix.

Proof: see Appendix A.

Therefore, consistent estimation of σ^2 requires only that $m_T = o(\sqrt{T})$, but the rate of consistency is unknown. The condition in (ii) may be stronger or weaker than that in (i) depending on the value of q. For the Bartlett estimator, q = 1, cf. Andrews (1991). From (ii), the rate of consistency of the kernel estimator is $\sqrt{T/m_T} = T^{q/(2q+1)}$, the limit of which when q $\to \infty$ is $\sqrt{T}$, the familiar result in nonparametric statistical estimation. The larger the value of q, the slower m_T grows, and the higher the rate of consistency[6].

Assumption (3.3.1) suggests that $\hat{\theta}_1$ be constructed as a random variable bounded to the interval [0, 1] which approaches the correct bound asymptotically. The easiest way to bound a random variable to [0, 1] is the exponential function and the probability measure. First we discuss the constructs using the exponential function. Let $A_T(y)$ be a function of the data such that,

(3.3.13a) $$\text{under } H_0,\ A_T \xrightarrow{p} +\infty$$

(3.3.13b) $$\text{under } H_1,\ A_T \xrightarrow{p} 0$$

then let $\hat{\theta}_1 = \exp(-|A_T|^{\nu})$ where ν controls the rate of convergence of $\hat{\theta}_1$ ($\nu>0$). A_T can easily be constructed from the traditional test for unit root. Let B_T be the traditional test statistic. If it is a consistent test, then under our H_0, $B_T = O_p(T^{\epsilon})$, $\epsilon > 0$; under our H_1, $B_T = O_p(1)$. It is trivial that $A_T = T^{-\eta}B_T$, $\eta < \epsilon$, satisfies (3.3.13). For example, for $B_T = T(1-\hat{\rho})$, where $\hat{\rho}$ is the

[6]The difference in the growth rate of m_T required for consistency of $J_T(\hat{\theta})$ in the literature largely is the result of using different probabilistic inequalities in proving the consistency. For example, there are two uses for $m_T = o(T^{1/4})$ in the proof of Newey and West (1987). The first is for $J_T(\theta_0)$-$EJ_T(\theta_0)$ for which they used a bound on $Var(\Sigma_{t=j+1}^{T} u_t u_{t-j}) \le T(j+1)D^*$, while Andrews (1991) uses a sharper bound on $var(J_T(\theta_0))$ due to Hannan (1970), and Hansen (1992) uses yet another sharp inequality. The second is for $J_T(\hat{\theta})$-$J_T(\theta_0)$, they use some bounds on $V_t(\theta)\partial V_t(\theta)/\partial\theta$, while Bierens (1992b) uses a bound on $\frac{1}{T}\Sigma_{t=1}^{T-j} V_t(\theta)\partial V_t(\theta)/\partial\theta$. For Theorem (3.3.1)(i), we use the same bound as Newey and West, but for (ii), the bound on $\frac{1}{T}\Sigma_{t=1}^{T-j} V_t(\theta)\partial V_t(\theta)/\partial\theta$ due to Bierens (1992b) is used.

first order autocorrelation coefficient of y_t, it is well known that $\epsilon = 1$.

It is desirable that B_T, and so A_T, does not depend on the scale of y_t. Since otherwise, rescaling y_t would change the stationarity/unit root behavior of y_t. This rules out $[1/(T^{-2}\Sigma y_t^2)]$ as a choice for B_T, the inverse of which is the basis for the Kwiatkowski, Phillips and Schmidt (1990) test for stationarity.

Conditional on a first stage statistic B_T, we can infer the probability that the series y_t is stationary (I(0)), $P(I(0)|B_T)$, and the probability that it has a unit root (I(1)), $P(I(1)|B_T)$, cf. Stock (1992) and Elliott and Stock (1992). Let $\hat{\theta}_1 = P(I(1))|B_T)$, then $\hat{\theta}_1 \xrightarrow{P} 0$ under H_0, $\hat{\theta}_1 \xrightarrow{P} 1$ under H_1. With this estimate of θ_1, testing for unit root/stationarity is seen as an iterative process. But the rate of convergence of $\hat{\theta}_1$ is difficult to establish; therefore, we do not further pursue this choice of $\hat{\theta}_1$ here.

The kernel estimators of the long run variance approximate the unknown autocorrelations in the series by an increasingly higher order moving average process (justifying the truncation). When the serial correlations are high, for a near unit root process for example, the truncation may introduce substantial error into the estimation. But on the other hand, if the serial correlation is low, for a series close to white noise for example, such truncation may work well. To improve the quality of estimation, Andrews and Monahan (1992) propose a prewhitening/recoloring procedure, which consists of four steps,

(i) estimate parameters θ_1 and θ_2, and construct $V_t(\hat{\theta}_1,\hat{\theta}_2)$ as before;

(ii) prewhitening — fit a simple AR(p) model to $V_t(\hat{\theta}_1,\hat{\theta}_2)$, typically $p = 1$, save the the AR coefficient estimates (say $\hat{\theta}_3$) and the residuals $V_t^*(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3) = V_t(\hat{\theta}_1,\hat{\theta}_2)-\Sigma_{i=1}^{p} V_t(\hat{\theta}_1,\hat{\theta}_2)\hat{\theta}_{3i}$;

(iii) apply the kernel estimation using V_t^*, denote the estimate as $J_T^*(\hat{\theta})$;

(iv) recoloring — the final estimate of the long run variance of $V_t(\theta_0)$ is $J_T^*(\hat{\theta})(1-\Sigma_{i=1}^{p}\hat{\theta}_{3i})^{-2}$.

The residuals after the prewhitening are expected to be close to white noise; therefore, the kernel estimate is expected to do better. If the coefficient estimates in the prewhitening are consistent for some points in the parameter space, the recoloring is consistent for the inverse transformation. It is easy to show that Theorem (3.3.1) holds for the prewhitening/recoloring kernel estimate.

3.3.3 Monte Carlo Simulation

Tables 3.3.1 and 3.3.2 contains the simulation results for the data generating process

$$(1\text{-}\rho B)Y_t = (1\text{-}\theta B)u_t.$$

Indeed, for all the choices of the truncation lag m_T, our adaptive estimate of σ^2 is very close to the Newey-West estimate for $\rho = (0, 0.8)$, $\rho = 0.9$ ($\theta = 0, 0.5$), and $\theta = 0.95$ ($\theta = 0.0$). This is because for these cases, the adaptive weight $\hat{\theta}_1$ is very close to the asymptotic value of zero, cf. Table 3.3.2. Because of the superconsistency of $\hat{\rho}$ for the unit root process, the adaptive weight $\hat{\theta}_1$ is very close to the asymptotic value of one, cf. Table 3.3.2. The level part in (3.3.1) receives a negligible weight, so the first difference part dominates. While the Newey-West estimate explodes because it is estimating an infinite quantity, our adaptive estimate remains finite because it is consistent for σ^2 which is the long run variance of the first difference of Y_t. The difference in the behavior of $\hat{\sigma}^2$ for the unit root case is all that matters for the power of the test for stationarity. For the near unit root process ($\rho = 0.95$), the adaptive estimator gives substantial weight to the first difference part ($\hat{\theta}_1 = (0.335, 0.811)$ for $\theta = (0, -0.5)$), which pulls down the estimate of σ^2. Note that for the near-unit root process, the Newey-West estimate is already a underestimate of σ^2, but our adaptive estimator is a further underestimate.

3.4 Application of the Adaptive Estimator to Testing for Stationarity

With the new consistent long run variance estimator, we can construct some new unit root tests with improved power, at least in large samples. In this section, we examine the power improvement for Park's (1990) χ^2 test, the CUSUM and fluctuation test introduced earlier in the chapter.

3.4.1 Park's (1990) χ^2 Test

3.4.1A The Test Statistic

For the ease of proof, we present our result for the case of orthogonal polynomial regression introduced in Chapter 2.

Let the null hypothesis be linear trend stationarity, and the alternative be unit root with drift, cf. equations (3.2.5) and (3.2.6) with $f(t/T)b = b_0 + b_1(t/T)$. Define

(3.4.1) $$N_1 = \hat{b}_2/\hat{\sigma}$$

where $\hat{b}$ is the regression coefficient in the orthogonal regression 2.3.8. Then from Theorem 2.3.1, it is trivial that

Theorem 3.4.1

(i) Under H_0, $N_1 \Rightarrow N(0, 1)$;

(ii) under H_1, $N_1/T \Rightarrow \int q_2^*(r)W(r)dr \sim N(0, s^2)$

where s^2 can be calculated from Theorem 2.2.3.

With multiple superfluous regressors, we can construct a χ^2 test. Define

(3.4.2) $$N_2 = (\hat{b}_2^2 + \hat{b}_3^2)/\hat{\sigma}^2.$$

Then trivially from Theorem 2.3.1, we have

Theorem 3.4.2

(i) Under H_0, $N_2 \Rightarrow \chi^2(2)$;

(ii) under H_1, $N_2/T^2 \Rightarrow [\int q_2^*(r)W(r)dr]^2 + [\int q_3^*(r)W(r)dr]^2$

The N_2 test is similar to the χ^2 test of Park (1990), but with improved power.

3.4.1.B Simulation Results

Tables (3.4.1) and (3.4.2) contains the results of simulation with the χ^2 test with one degree of freedom (one superfluous regressor). The size distortion is high for near unit root processes using both estimates of the long run variance. As a comparison, if we used the true value of the long run variance σ^2, the size and power comparison would be much better. Therefore, the size over-distortion is due to the underestimate of the long run variance of both the original and the adaptive Newey-West estimator. The power of the test using the adaptive estimator is much better. Improving the power while keeping the size distortion at the same level as before is the sole objective of designing the adaptive estimator.

3.4.2 The CUSUM Test

Table 3.4.3 contains simulation results on the CUSUM test for stationarity using the new adaptive estimator, Table 3.4.4 is the result with a linear trend. Comparing with the results in Section 3.2, we see that indeed the power of the test has been improved substantially.

3.5 Summary and Conclusion

Using results on spurious regression, we find that tests for parameter stability also have power against unit root process. The asymptotic distributions of the CUSUM test using OLS residuals and of the parameter fluctuation test in subsamples are derived under the null of trend stationarity. Simulation results show that the tests have reasonable empirical sizes and reasonable power.

Many inference procedures in time series analysis relies on asymptotic theory, the asymptotic distribution involved often is proportional to (the square root of) the long run variance of the underlying series. To get a nuisance parameter free distribution, a consistent estimate of the long run variance is needed. The problem with the existing long run variance estimators, e. g. the Newey-West estimator, is that they are inconsistent under interesting alternative hypothesis, therefore the power of the test may be affected. For example, if the null hypothesis is stationarity and the alternative is unit root, the Newey-West estimator is consistent under the null, but explodes under the alternative. If the test statistic is constructed as some sample quantity divided by this long run variance estimator, the power of the test is weakened.

In this chapter, we propose an adaptive estimator of the long run variance. The series used in the Newey-West type of estimator is a weighted average of the series of interest under the null and under the alternative, with the weight being a random variable. Our estimator is consistent under both the null and the alternative.

Finally we apply the adaptive estimator of the long run variance to the CUSUM test, the parameter fluctuation test, and Park's χ^2 test. Indeed using the adaptive long run variance estimator significantly improves the power of the tests.

Appendix A. Proof of Theorem 3.3.1

To simplify the proof, we adopt the following assumptions from Andrews (1991). Due to the nonstationary nature of the trending regressors x_t, the sample quantities are used instead of the expectation as in Andrews (1991). Let θ_0 be the true value of the parameter θ, and Θ be a small convex neighborhood of θ_0.

Assumption A

(i) u_t is mean zero, fourth order stationary random variable;

(ii) $\sum_j |\Gamma(j)| < \infty$;

(iii) $\sum_{j,m,n} \kappa(0,j,m,n) < \infty$,

where $\kappa(t,t+j,t+m,t+n) = E(u_t u_{t+j} u_{t+m} u_{t+n})$

Assumption B

(i) $\sqrt{T}(\hat{\theta}-\theta_0) = O_p(1)$;

(ii) $(\hat{\theta}_1-\theta_{10}) = O_p(T^{-3/2})$ and $(\hat{\theta}_2-\theta_{20}) = O_p(\sqrt{T})$;

(iii) $\frac{1}{T}\Sigma_t \sup_{\theta \in \Theta} V_t^2(\theta) < \infty$;

(iv) $\frac{1}{T}\Sigma_t \sup_{\theta \in \Theta} \left\| \frac{\partial V_t(\theta)}{\partial \theta} \right\|^2 < \infty$

(v) $\frac{1}{T}\Sigma_t \sup_{\theta \in \Theta} \left(\frac{\partial V_t(\theta)}{\partial \theta} \sqrt{T}(\hat{\theta}-\theta_0) \right)^2 < \infty$

The proof is similar to that for Theorem 1 of Andrews (1991). It consists of some Taylor series expansions and inequalities involving V_t, $\partial V_t/\partial\theta$, and $\partial^2 V_t/\partial\theta\partial\theta'$. For ease of notation, we drop the subscript T from m_T and J_T.

Under the conditions in (i) and (ii) of the theorem, Proposition 1 of Andrews (1991) holds under both H_0 and H_1. For (i), $J(\theta_0)$-$J = o_p(1)$; for (ii), $\sqrt{T/m}\,[J(\theta_0)\text{-}J] = O_p(1)$. So to complete the proof, we need to provide bounds for $[J(\hat{\theta})-J(\theta_0)]$. For this, we follow the steps in Andrews (1991) which relies on Taylor series expansions of $J(\theta)$ at θ_0.

For (i), we need the following mean value expansion for J,

$$\frac{\sqrt{T}}{m}[J(\hat{\theta}) - J(\theta_0)] = \frac{1}{m}\frac{\partial J(\bar{\theta})}{\partial \theta}[\sqrt{T}(\hat{\theta} - \theta_0)]$$

(3A.1)
$$= \frac{1}{m}\Sigma_j\, k(j/m_T)\left[\frac{\partial \hat{\Gamma}(j,\bar{\theta})}{\partial \theta}\right][\sqrt{T}(\hat{\theta} - \theta_0)]$$

where $\bar{\theta} = \theta_0 + \alpha_1(\theta\text{-}\theta_0)$, $\alpha_1 \in [0,\ 1]$.

$$\left|\frac{\sqrt{T}}{m}[J(\hat{\theta}) - J(\theta_0)]\right| \leq \frac{1}{m}\Sigma_j\, |k(j/m)|\left|\frac{\partial \hat{\Gamma}(j,\bar{\theta})}{\partial \theta}[\sqrt{T}(\hat{\theta} - \theta_0)]\right|$$

But $\frac{1}{m}\Sigma_j\, |k(j/S_T)| \rightarrow \int |k(x)|dx < \infty$, if it can be shown that

(3A.2)
$$\left|\frac{\partial \hat{\Gamma}(j,\bar{\theta})}{\partial \theta}[\sqrt{T}(\hat{\theta} - \theta_0)]\right| = O_p(1)$$

then

(3A.3)
$$\frac{\sqrt{T}}{m}[J(\hat{\theta}) - J(\theta_0)] = O_p(1).$$

For (ii), we will use the following Taylor series expansion for J,

$$\sqrt{\frac{T}{m}}[J(\hat{\theta}) - J(\theta_0)] = \frac{1}{\sqrt{m}}\frac{\partial J(\theta_0)}{\partial \theta}[\sqrt{T}(\hat{\theta} - \theta_0)] + \frac{1}{2}[\sqrt{T}(\hat{\theta} - \theta_0)]'\frac{1}{\sqrt{Tm}}\frac{\partial^2 J(\bar{\theta})}{\partial \theta \partial \theta'}[\sqrt{T}(\hat{\theta} - \theta_0)]$$

(3A.4)
$$\equiv L_{1T}[\sqrt{T}(\hat{\theta} - \theta_0)] + \frac{1}{2}[\sqrt{T}(\hat{\theta} - \theta_0)]'L_{2T}[\sqrt{T}(\hat{\theta} - \theta_0)]$$

where $\bar{\theta} = \theta_0 + \alpha_2(\theta\text{-}\theta_0)$, $\alpha_2 \in [0,\ 1]$. If it can be shown that

(3A.5)
$$L_{1T}[\sqrt{T}(\hat{\theta} - \theta_0)] = o_p(1),$$

and

(3A.6)
$$[\sqrt{T}(\hat{\theta} - \theta_0)]'L_{2T}[\sqrt{T}(\hat{\theta} - \theta_0)] = o_p(1)$$

then

(3A.7) $$\sqrt{\tfrac{T}{m}}[J(\hat{\theta}) - J(\theta_0)] = o_p(1).$$

Note that

$$\frac{\partial^2 V_t(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} 0 & x_{t-1} \\ x'_{t-1} & 0 \end{bmatrix}$$

so

$$[\sqrt{T}(\hat{\theta} - \theta_0)]'L_{2T}[\sqrt{T}(\hat{\theta} - \theta_0)] = \frac{1}{\sqrt{Tm}} \Sigma_j k(j/m)\tfrac{1}{T}\Sigma_t\, x_{t-1}[T(\hat{\theta}_2-\theta_{20})(\hat{\theta}_1-\theta_{10})]$$

but from Assumption (3.2.1) and (3.2.2),

(3A.8) $$T(\hat{\theta}_2-\theta_{20})(\hat{\theta}_1-\theta_{10}) = O_p(1) \text{ under both } H_0 \text{ and } H_1$$

therefore

$$\left|[\sqrt{T}(\hat{\theta} - \theta_0)]'L_{2T}[\sqrt{T}(\hat{\theta} - \theta_0)]\right|$$

$$\leq \sqrt{\tfrac{m}{T}}\, \tfrac{1}{m}\Sigma_j|k(j/m)|\, \left\|\tfrac{1}{T}\Sigma_t\, x_{t-1}\right\| O_p(1)$$

$$= O_p(\sqrt{m/T})$$

(3A.6) is proved. Note that (3A.8) is the only result that depends on the extremely strong rate of convergence of $\hat{\theta}$ under H_1.

The rest of the proof consists of establishing (3A.2) and (3A.5) under both H_0 and H_1. First we need some results on Taylor series expansion,

(3A.9a) $$V_t(\theta) = V_t(\theta_0) + \frac{\partial V_t(\theta_0)}{\partial\theta}(\theta-\theta_0) + \tfrac{1}{2}(\theta-\theta_0)'\frac{\partial^2 V_t(\bar{\theta})}{\partial\theta\partial\theta'}(\theta-\theta_0)$$

(3A.9b) $$\frac{\partial V_t(\theta)}{\partial\theta} = \frac{\partial V_t(\theta_0)}{\partial\theta} + \frac{\partial^2 V_t(\check{\theta})}{\partial\theta\partial\theta'}(\theta - \theta_0)$$

where $\bar{\theta} = \theta_0 + \alpha_3(\theta-\theta_0)$, $\alpha_3 \in [0, 1]$, $\check{\theta} = \theta_0 + \alpha_4(\theta-\theta_0)$, $\alpha_4 \in [0, 1]$.

Under H_0,

$$\frac{\partial V_t(\theta_0)}{\partial \theta} = - \begin{bmatrix} u_t \\ x_t' \end{bmatrix}$$

Assumptions B(i)(iii)(iv) hold.

$$\sup_{j \geq 1} \left\| \frac{\partial \hat{\Gamma}(j,\bar{\theta})}{\partial \theta} \right\| = \sup_{j \geq 1} \left\| \frac{1}{T}\Sigma_t \left(V_t(\bar{\theta}) \frac{\partial V_{t+j}(\bar{\theta})}{\partial \theta} + V_{t+j}(\bar{\theta}) \frac{\partial V_t(\bar{\theta})}{\partial \theta} \right) \right\|$$

$$\leq 2\left[\frac{1}{T}\Sigma_t \sup_{\theta \in \Theta} V_t^2(\theta)\right]^{1/2} \left[\frac{1}{T}\Sigma_t \sup_{\theta \in \Theta} \left\| \frac{\partial V_t(\theta)}{\partial \theta} \right\|^2\right]^{1/2}$$

$$= O_p(1) \text{ by B(iii)(iv)}$$

(3A.2) holds by B(i).

$$L_{1T} = - \frac{1}{\sqrt{m}} \frac{1}{T} \Sigma_j \, k(j/m) \Sigma_t (u_t (u_{t+j}, x_{t+j}) + u_{t+j}(u_t, x_t))$$

$$= - \frac{1}{\sqrt{m}} \frac{1}{T} \Sigma_j \, k(j/m) \Sigma_t (2u_t u_{t+j}, u_t x_{t+j} + u_{t+j} x_t)$$

By (A.13) of Andrews (1991),

$$\frac{1}{T} \Sigma_j \, k(j/m) \Sigma_t u_t u_{t+j} = O_p(1)$$

Similar to (A.15) of Andrews, it is easy to show that,

$$E\left(\frac{1}{\sqrt{m}} \frac{1}{T} \Sigma_j \, k(j/m) \Sigma_t u_t x_{t+j}\right)^2 = O_p(m/T)$$

$$E\left(\frac{1}{\sqrt{m}} \frac{1}{T} \Sigma_j \, k(j/m) \Sigma_t u_{t+j} x_t\right)^2 = O_p(m/T)$$

so $L_{1T} = o_p(1)$, and (3A.5) holds by B(i). The proof under H_0 is complete.

Under H_1,

$$\frac{\partial V_t(\theta_0)}{\partial\theta} = \begin{bmatrix} -\Sigma_{j=1}^{t-1}u_t \\ -(x_t - x_{t-1})' \end{bmatrix} = \begin{bmatrix} -\Sigma_{j=1}^{t-1}u_t \\ 0 \\ -1/T \end{bmatrix}$$

so $\partial V_t(\theta)/\partial\theta$ is unbounded in Θ. However under Assumption B(ii),

$$\frac{\partial V_t(\theta_0)}{\partial\theta}(\hat\theta-\theta_0) = -\frac{1}{\sqrt{T}}\Sigma_{j=1}^{t-1}u_t\sqrt{T}(\hat\theta_1-\theta_{10}) - \frac{1}{T}(\hat\theta_{2,2}-\theta_{2,2,0}) = O_p(1/\sqrt{T}).$$

From (3A.9b),

$$\begin{aligned}\frac{\partial V_t(\theta)}{\partial\theta}(\hat\theta-\theta_0) &= \frac{\partial V_t(\theta_0)}{\partial\theta}(\hat\theta-\theta_0) + (\theta-\theta_0)'\frac{\partial^2 V_t(\bar\theta)}{\partial\theta\partial\theta'}(\hat\theta-\theta_0)\\ &= O_p(1/\sqrt{T})\end{aligned}$$

Using the Taylor series expansion for $V_t(\theta)$, (3A.9a), and the above results, it is trivial that B(iii) and B(v) hold, but B(i) (iv) do not. Note that for the above results to hold, it is not necessary that $\sqrt{T}(\hat\theta_1-\theta_{10}) = O_p(1/T)$. They are valid as long as $\sqrt{T}(\hat\theta_1-\theta_{10}) = O_p(1/\sqrt{T})$. However,

$$\sup_{j\geq 1}\left|\frac{\partial\hat\Gamma(j,\bar\theta)}{\partial\theta}\sqrt{T}(\hat\theta-\theta_0)\right|$$

$$= \sup_{j\geq 1}\left|\frac{1}{T}\Sigma_t\left(V_t(\bar\theta)\frac{\partial V_{t+j}(\bar\theta)}{\partial\theta}\sqrt{T}(\hat\theta-\theta_0) + V_{t+j}(\bar\theta)\frac{\partial V_t(\bar\theta)}{\partial\theta}\sqrt{T}(\hat\theta-\theta_0)\right)\right|$$

$$\leq 2\left[\frac{1}{T}\Sigma_t\sup_{\theta\in\Theta}V_t^2(\theta)\right]^{1/2}\left[\frac{1}{T}\Sigma_t\sup_{\theta\in\Theta}\left(\frac{\partial V_t(\theta)}{\partial\theta}\sqrt{T}(\hat\theta-\theta_0)\right)^2\right]^{1/2} = O_p(1)$$

(3A.2) is proved.

$$L_{1T}[\sqrt{T}(\hat\theta - \theta_0)] = -\frac{1}{\sqrt{m}}\frac{1}{T}\Sigma_j k(j/m)\Sigma_t u_t\left[\Sigma_{k=1}^{t+j-1}u_k\sqrt{T}(\hat\theta_1-\theta_{10}) + (1/\sqrt{T})(\hat\theta_{2,2}-\theta_{2,2,0})\right]$$

(3A.10) $$-\frac{1}{\sqrt{m}}\frac{1}{T}\Sigma_j k(j/m)\Sigma_t u_{t+j}\left[\Sigma_{k=1}^{t-1}u_k\sqrt{T}(\hat\theta_1-\theta_{10}) + (1/\sqrt{T})(\hat\theta_{2,2}-\theta_{2,2,0})\right]$$

It is shown in Bierens (1992b) (cf. equations (9.5.8) and (9.5.10)) that,

$$\frac{1}{T}\Sigma_t u_t\Sigma_{k=1}^{t+j-1}u_k = O_p(\sqrt{j}).$$

From this, we have,

$$\frac{1}{\sqrt{m}} \Sigma_j\ k(j/m)\frac{1}{T}\Sigma_t u_t \Sigma_{k=1}^{t+j-1} u_k = O_p(m).$$

From Assumption B(ii),

$$\frac{1}{\sqrt{m}} \Sigma_j\ k(j/m)\frac{1}{T}\Sigma_t u_t \Sigma_{k=1}^{t+j-1} u_k \sqrt{T}(\hat{\theta}_1-\theta_{10}) = O_p(m/T) \tag{3A.11}$$

Similar to (A.15) of Andrews (1991), it is easy to show that,

$$E\Big(\frac{1}{\sqrt{m}}\frac{1}{T}\Sigma_j\ k(j/m)\Sigma_t u_t\Big)^2 = O_p(m/T)$$

so

$$\frac{1}{\sqrt{m}}\frac{1}{T}\Sigma_j\ k(j/m)\Sigma_t u_t(1/\sqrt{T})(\hat{\theta}_{2,2}-\theta_{2,2,0}) = o_p(1)$$

A similar result holds for the second term in (3A.10). So (3A.5) holds. Note that it is not necessary for (3A.5) that $\sqrt{T}(\hat{\theta}_1-\theta_{10}) = O_p(1/T)$; it holds as long as $\sqrt{T}(\hat{\theta}_1-\theta_{10}) = o_p(1/m)$. The proof under H_1 is complete. □

Appendix B. Andrews Optimal Quadratic Spectral Estimator of long run variance

The materials covered here are from Andrews (1991), and are included only for completeness and for illustrative purposes.

The quantities k_q and $f^{(q)}$ in Theorem 3.2.1 measure the smoothness of the kernel $k(\cdot)$ and the spectral density function $f(\lambda)$ of u_t at zero, cf Andrews (1991),

$$k_q = \lim_{x \to 0} \frac{1-k(x)}{|x|^q} \qquad \text{for } q \in [0, \infty)$$

$$f^{(q)} = \frac{1}{2\pi} \Sigma_j |j|^q \Gamma(j) \qquad \text{for } q \in [0, \infty)$$

Examples of kernels in k_1 and their q values in Theorem (3.2.1)(ii) are as follows:

(i) Truncated $\quad k(x) = \begin{cases} 1 \text{ for } |x| \leq 1 \\ 0 \text{ otherwise} \end{cases}$

(ii) Bartlett $\quad k(x) = \begin{cases} 1-|x| \text{ for } |x| \leq 1 \\ 0 \text{ otherwise} \end{cases} \qquad q = 1$

(iii) Parzen $\quad k(x) = \begin{cases} 1-6x^2+6|x|^3 \text{ for } 0 \leq |x| \leq 1/2 \\ 2(1-|x|)^3 \text{ for } 1/2 \leq |x| \leq 1 \\ 0 \text{ otherwise} \end{cases} \qquad q = 2$

(iv) Tukey-Hanning $\quad k(x) = \begin{cases} (1+\cos(\pi x))/2 \text{ for } |x|<1 \\ 0 \text{ otherwise} \end{cases} \qquad q = 2$

(v) Quadratic Spectral $\quad k(x) = \frac{3}{(6\pi x/5)^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right) \qquad q = 2$

Theorem 2 of Andrews (1991) proves that for any given sequence of bandwidth parameters $\{m_T\}$, the QS kernel is preferred to other kernels in the asymptotic truncated MSE. Define

$$\alpha(q) = (f^{(q)}/f(0))^2$$

$$m_T^* = \left(qk_q^2\alpha(q)T/\int k^2(x)dx\right)^{1/(2q+1)}$$

then Andrews (1991), in Corollary 1, shows that the sequence m_T^* is preferred to other sequences $\{m_T\}$ in the asymptotic truncated MSE. For the familiar kernels,

Bartlett	$m_T^* = 1.1447(\alpha(1)T)^{1/3}$
Parzen	$m_T^* = 2.6614(\alpha(2)T)^{1/5}$
Tukey-Hanning	$m_T^* = 1.7462(\alpha(2)T)^{1/5}$
QS	$m_T^* = 1.3221(\alpha(2)T)^{1/5}$

But m_T^* involves unknown parameters $\alpha(q)$, which can be estimated by assuming and consistently estimating an AR(1) model for $V_t(\theta)$,

$$\hat{\alpha}(2) = 4\hat{\rho}^2/(1-\hat{\rho})^4$$

$$\hat{\alpha}(1) = 4\hat{\rho}^2/[(1-\hat{\rho})^2(1+\hat{\rho})^2]$$

$$\hat{m}_T = \left(qk_q^2\hat{\alpha}(q)T/\int k^2(x)dx\right)^{1/(2q+1)}$$

Theorem 3 of Andrews (1991) establishes the consistency and optimality of the automatic bandwidth estimator $\hat{m}_T$. In the simulation, we use the QS estimator with an automatic bandwidth estimator with $q = 2$.

Table 3.2.1 Simulated Asymptotic Distribution of the CUSUM Test

$\alpha = P(Q \leq q)$

	$\tau = 0$				$\tau = 0.15$			
	range		sup		range		sup	
α*100	k=0	k=1	k=0	k=1	k=0	k=1	k=0	k=1
1	0.703	0.660	0.417	0.351	0.618	0.601	0.392	0.328
5	0.807	0.760	0.493	0.408	0.713	0.706	0.471	0.387
10	0.872	0.818	0.545	0.440	0.778	0.765	0.526	0.420
15	0.922	0.864	0.586	0.465	0.823	0.809	0.568	0.444
20	0.963	0.905	0.620	0.487	0.865	0.850	0.604	0.467
30	1.037	0.969	0.681	0.525	0.937	0.917	0.668	0.506
40	1.105	1.030	0.740	0.561	1.002	0.980	0.732	0.542
50	1.172	1.091	0.802	0.597	1.069	1.044	0.795	0.580
60	1.241	1.155	0.869	0.635	1.142	1.112	0.865	0.618
70	1.321	1.228	0.948	0.678	1.222	1.191	0.945	0.664
80	1.418	1.319	1.052	0.732	1.322	1.286	1.049	0.720
90	1.563	1.455	1.207	0.810	1.476	1.429	1.207	0.801
95	1.697	1.572	1.333	0.881	1.600	1.550	1.333	0.873
99	1.960	1.828	1.605	1.019	1.850	1.813	1.605	1.011

The simulation is based on 10000 replications of samples of 500 observations. k is the maintained polynomial time trend. RANGE is $\left[\sup_{r \in [\tau,1-\tau]} D(r) - \inf_{r \in [\tau,1-\tau]} D(r)\right]$, SUP is $\sup_{r \in [\tau,1-\tau]} |D(r)|$, where D(r) is the asymptotic distribution in Theorem 3.2.5(i).

Table 3.2.2 Monte Carlo Simulation on the CUSUM Test

with An Intercept

A. The SUP Test ($\tau = 0$)

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	5.6	9.8	51.5	63.5	79.7	85.9	93.8	97.2	99.3	99.7
	4	5.1	10.2	30.3	42.4	55.6	68.1	78.7	86.2	97.9	98.8
	8	4.1	9.3	14.7	22.4	30.8	41.3	54.8	64.8	91.2	94.7
	12	3.5	9.4	8.9	15.5	20.1	29.7	36.8	49.9	82.0	88.8
	QS	5.3	9.4	1.4	4.7	0.8	2.3	0.1	0.4	0.2	0.5
	∞	5.5	9.5	2.0	4.4	1.6	3.5	1.0	2.2	100.0	100.0
0.5	2	0.1	0.3	39.1	51.8	74.7	82.6	92.0	96.4	99.7	99.9
	4	0.2	0.8	25.2	35.8	53.1	65.5	78.3	85.9	98.1	99.1
	8	0.9	2.2	11.9	21.8	28.1	42.0	53.8	64.4	91.9	95.2
	12	1.0	4.3	7.2	13.6	16.1	26.6	38.5	49.7	82.2	89.0
	QS	0.2	0.9	29.1	39.9	50.0	61.8	67.0	77.9	70.9	80.5
	∞	8.2	14.5	2.4	4.7	1.5	2.9	0.3	0.9	100.0	100.0
-0.5	2	7.2	13.3	54.5	65.8	79.1	86.5	94.6	96.9	99.7	99.9
	4	5.3	10.6	30.5	42.3	54.4	67.0	79.6	86.8	97.1	98.5
	8	3.8	8.9	15.3	23.6	31.2	40.4	51.9	63.1	89.6	94.0
	12	3.8	8.0	10.1	17.1	18.8	28.7	36.2	47.7	80.1	86.6
	QS	0.8	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
	∞	4.0	8.6	2.6	5.0	1.9	4.4	0.3	1.8	100.0	100.0

B. The RANGE Test ($\tau = 0$)

θ	m	ρ 0.00 5%	10%	0.80 5%	10%	0.90 5%	10%	0.95 5%	10%	1.00 5%	10%
0.0	2	3.4	8.9	66.1	78.0	91.7	96.6	98.4	99.3	99.9	100.0
	4	3.4	8.6	34.7	47.7	69.5	79.9	88.7	94.1	99.1	99.2
	8	2.6	7.5	11.3	21.6	34.2	47.8	59.4	72.0	93.9	96.0
	12	1.7	6.3	4.4	10.8	13.5	28.0	31.8	50.1	82.6	90.4
	QS	4.1	9.0	0.3	1.0	0.2	0.3	0.0	0.0	0.0	0.0
	∞	3.5	9.4	1.2	2.4	0.5	1.8	0.2	0.5	100.0	100.0
0.5	2	0.0	0.3	53.0	65.2	89.3	94.3	98.7	99.9	100.0	100.0
	4	0.3	0.7	29.9	44.7	66.9	79.4	87.8	93.6	99.1	99.5
	8	0.7	2.8	11.4	21.4	29.9	46.2	61.1	75.1	94.1	96.6
	12	1.1	4.4	5.6	12.6	11.4	25.0	33.5	51.6	84.1	91.1
	QS	0.4	0.6	36.8	49.1	61.1	77.2	77.5	85.8	56.3	68.9
	∞	13.7	23.2	2.0	3.6	0.6	1.7	0.0	0.2	100.0	100.0
-0.5	2	7.0	13.9	71.6	82.2	91.0	95.6	98.4	99.3	99.8	99.9
	4	4.3	10.5	41.4	54.2	68.4	78.2	90.4	95.2	99.1	99.6
	8	2.7	7.6	13.2	26.5	31.1	44.6	61.6	71.9	93.0	96.3
	12	1.7	5.8	5.7	12.9	13.2	24.9	32.7	50.1	79.8	87.8
	QS	0.4	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	∞	3.5	7.3	1.0	3.5	0.3	1.0	0.0	0.4	100.0	100.0

The simulation is based on 1000 replications of the following DGP,

$$(1\text{-}\rho B)\, y_t = (1\text{-}\theta B)u_t, \text{ where } u_t \sim \text{NIID}(0, 1), t = 1, 2, \cdots, T, T = 300$$

m is the number of autocorrelations used in the Newey-West estimate of long run variance. The column for m = ∞ uses the true σ^2; it is only meant as a comparison.

Table 3.2.3 Monte Carlo Simulation on the CUSUM Test

with a Linear Trend

A. The SUP Test ($\tau = 0$)

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	3.1	7.8	66.1	76.0	91.2	96.4	97.7	99.0	100.0	100.0
	4	3.2	7.5	35.3	49.4	67.1	80.1	87.9	92.0	97.0	98.7
	8	2.8	7.2	14.0	23.8	35.9	48.5	60.6	73.0	88.4	93.2
	12	2.3	6.8	6.3	15.5	18.9	31.2	34.9	52.6	71.2	82.8
	QS	3.4	8.0	0.6	2.2	0.1	0.3	0.2	0.2	0.0	0.1
	∞	3.9	7.9	1.4	3.2	0.3	1.0	0.0	0.1	100.0	100.0
0.5	2	0.2	0.2	51.8	63.9	88.8	94.5	97.8	99.3	99.8	99.9
	4	0.4	1.4	31.8	44.0	68.1	79.1	86.8	94.0	97.0	99.1
	8	1.3	3.9	13.9	24.8	34.8	50.0	60.7	71.8	86.8	91.6
	12	2.3	6.9	7.6	16.5	15.9	30.7	34.8	52.3	71.9	81.1
	QS	0.5	1.5	36.9	49.3	65.7	76.7	77.7	86.8	82.6	91.3
	∞	12.3	22.7	1.1	4.3	0.4	1.1	0.0	0.2	100.0	100.0
-0.5	2	7.8	15.0	69.4	78.4	90.4	94.6	98.4	99.3	99.8	99.9
	4	5.1	11.9	41.3	54.3	67.7	78.6	88.3	93.5	97.8	99.0
	8	4.4	8.9	17.4	29.2	34.2	48.4	59.2	72.8	87.9	92.3
	12	2.5	8.9	8.1	17.4	15.5	29.0	36.1	52.0	70.1	82.4
	QS	0.5	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	∞	3.8	6.8	1.5	4.3	0.1	0.7	0.3	0.5	100.0	100.0

B. The Range Test ($\tau = 0$)

θ	m	ρ = 0.00 5%	0.00 10%	0.80 5%	0.80 10%	0.90 5%	0.90 10%	0.95 5%	0.95 10%	1.00 5%	1.00 10%
0.0	2	3.5	8.2	67.8	79.0	93.1	97.3	98.0	99.4	100.0	100.0
	4	3.1	8.2	35.8	50.4	70.3	80.2	88.9	93.8	97.7	99.4
	8	2.3	8.2	13.2	22.5	33.8	48.1	58.6	71.7	88.7	92.8
	12	2.5	7.4	5.7	13.0	16.8	28.7	33.7	48.8	71.6	82.1
	QS	3.8	7.9	0.4	2.1	0.1	0.4	0.0	0.3	0.0	0.1
	∞	3.3	8.4	1.1	2.3	0.2	1.0	0.0	0.2	100.0	100.0
0.5	2	0.2	1.2	53.7	66.9	90.8	94.9	98.4	99.5	99.8	99.9
	4	0.5	1.3	32.6	44.9	70.2	80.9	88.8	94.3	98.5	99.2
	8	1.3	5.2	14.0	23.8	33.7	48.0	60.0	72.2	87.4	92.6
	12	2.9	6.7	8.3	15.7	15.7	29.2	35.9	51.3	70.7	81.8
	QS	0.6	1.3	37.5	50.6	67.1	77.6	79.8	86.8	84.8	92.1
	∞	14.5	23.8	1.6	3.6	0.4	0.8	0.0	0.2	100.0	100.0
-0.5	2	8.2	15.5	70.6	81.0	91.3	96.0	98.8	99.4	99.9	99.9
	4	6.0	12.6	41.8	54.9	69.1	79.3	90.8	95.1	98.3	99.3
	8	4.1	9.5	15.7	27.2	35.0	46.8	61.2	72.4	88.2	93.1
	12	3.0	8.5	7.8	15.3	16.0	28.3	34.8	50.3	70.4	81.6
	QS	0.6	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	∞	4.0	7.6	1.4	2.9	0.2	0.6	0.3	0.6	100.0	100.0

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho B)\, y_t = (1-\theta B)u_t, \text{ where } u_t \sim \text{NID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300$$

m is the number of autocorrelations used in the Newey-West estimate of long run variance. The row for m = ∞ uses the true σ^2; it is only meant as a comparison. The row QS uses the QS estimate of Andrews (1991); see appendix B for details.

Table 3.2.4 Simulated Asymptotic Distribution of the Fluctuation Test

$\alpha = P(Q \leq q)$

	$\tau = 0$		$\tau = 0.15$	
α*100	Max-norm	Euclidean norm	Max-norm	Euclidean norm
1	9.425	9.431	1.467	1.613
5	12.741	12.741	1.763	1.946
10	15.336	15.337	1.969	2.173
15	17.548	17.548	2.113	2.343
20	19.393	19.393	2.248	2.486
30	23.261	23.261	2.473	2.737
40	26.931	26.932	2.709	2.981
50	31.234	31.234	2.944	3.233
60	36.422	36.423	3.222	3.509
70	42.393	42.393	3.551	3.839
80	50.961	50.961	3.965	4.240
90	64.056	64.056	4.618	4.930
95	75.772	75.773	5.204	5.527
99	99.607	99.607	6.382	6.741

The simulation is based on 10000 replications of samples of 500 observations. SUP is $\sup_{r \in [\tau, 1-\tau]} \| D(r) \|$, where D(r) is the asymptotic distribution for k = 1 in Theorem 3.2.4(i), $\|D\|$ is the Max norm $\|D\|_\infty$ for the column Max-norm, and it is the Euclidean norm for the column Euclidean norm.

Table 3.2.5 The Simulated Distribution of the CUSUM Test

under the Presence of a one-time Shift in the Mean

A. (no trimming)

	η (break fraction)									
α*100	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	mCU
1	0.398	0.378	0.382	0.373	0.373	0.373	0.381	0.388	0.395	0.333
10	0.519	0.498	0.489	0.484	0.480	0.480	0.483	0.495	0.513	0.409
20	0.588	0.559	0.544	0.538	0.536	0.537	0.543	0.556	0.582	0.449
30	0.645	0.615	0.593	0.584	0.581	0.581	0.592	0.609	0.640	0.480
40	0.703	0.665	0.640	0.628	0.624	0.625	0.638	0.658	0.694	0.507
50	0.761	0.719	0.690	0.672	0.665	0.669	0.683	0.709	0.753	0.535
60	0.827	0.780	0.741	0.717	0.709	0.718	0.736	0.768	0.813	0.568
70	0.902	0.847	0.803	0.773	0.760	0.774	0.799	0.840	0.891	0.603
80	0.994	0.933	0.882	0.842	0.827	0.841	0.874	0.929	0.991	0.648
90	1.146	1.071	1.002	0.947	0.926	0.947	1.001	1.068	1.139	0.717
95	1.271	1.197	1.113	1.047	1.007	1.034	1.103	1.189	1.266	0.776
99	1.522	1.440	1.344	1.235	1.192	1.229	1.337	1.430	1.531	0.913

B. (trimming = 0.15)

	η (break fraction)									
α*100	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	mCU
1	0.383	0.356	0.351	0.342	0.351	0.349	0.350	0.356	0.376	0.305
10	0.508	0.476	0.465	0.458	0.456	0.455	0.456	0.475	0.499	0.382
20	0.580	0.543	0.524	0.516	0.514	0.514	0.519	0.537	0.570	0.422
30	0.637	0.603	0.574	0.562	0.561	0.561	0.573	0.594	0.632	0.453
40	0.698	0.655	0.625	0.610	0.604	0.607	0.621	0.648	0.688	0.483
50	0.757	0.712	0.675	0.657	0.648	0.652	0.669	0.700	0.748	0.513
60	0.822	0.775	0.731	0.705	0.695	0.703	0.723	0.762	0.810	0.545
70	0.900	0.845	0.795	0.761	0.748	0.761	0.789	0.836	0.888	0.584
80	0.993	0.933	0.880	0.834	0.816	0.833	0.867	0.926	0.989	0.629
90	1.144	1.069	1.002	0.942	0.920	0.941	0.994	1.067	1.139	0.700
95	1.271	1.194	1.110	1.044	1.002	1.031	1.101	1.188	1.265	0.763
99	1.522	1.441	1.336	1.234	1.183	1.226	1.337	1.431	1.531	0.899

The simulation is based on 10000 replications of samples of 500 observations. The column under "no break" is for the case of a constant intercept. It is taken from Table 3.2.1 (the SUP k=0 column). η is the break fraction. "mCU" is for the minimum of the CUSUM test across the grid values of η.

Table 3.2.6 The Simulated Distribution of the CUSUM Test

under the Presence of a one-time Shift in Both the Intercept and the Slope

A. (no trimming)

	η (break fraction)									
α*100	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	mCU
1	0.333	0.318	0.308	0.303	0.305	0.304	0.307	0.315	0.333	0.276
10	0.419	0.393	0.379	0.371	0.368	0.370	0.378	0.395	0.418	0.330
20	0.463	0.433	0.413	0.403	0.402	0.404	0.415	0.437	0.462	0.356
30	0.498	0.467	0.442	0.430	0.426	0.431	0.447	0.469	0.498	0.375
40	0.531	0.499	0.470	0.454	0.450	0.455	0.474	0.500	0.530	0.392
50	0.564	0.531	0.499	0.479	0.473	0.479	0.499	0.531	0.563	0.408
60	0.601	0.566	0.530	0.506	0.498	0.506	0.530	0.564	0.601	0.426
70	0.641	0.600	0.565	0.536	0.527	0.536	0.565	0.603	0.642	0.447
80	0.694	0.649	0.609	0.575	0.562	0.574	0.609	0.651	0.691	0.473
90	0.767	0.721	0.676	0.632	0.613	0.633	0.674	0.724	0.764	0.510
95	0.836	0.786	0.737	0.682	0.659	0.684	0.733	0.786	0.830	0.539
99	0.972	0.918	0.859	0.777	0.745	0.785	0.842	0.907	0.956	0.606

B. (trimming = 0.15)

α*100	η (break fraction) 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	mCU
1	0.320	0.305	0.293	0.284	0.287	0.287	0.291	0.302	0.317	0.262
10	0.407	0.380	0.361	0.350	0.347	0.351	0.361	0.383	0.405	0.310
20	0.451	0.422	0.397	0.384	0.380	0.383	0.398	0.422	0.450	0.337
30	0.488	0.455	0.427	0.410	0.405	0.410	0.427	0.455	0.486	0.356
40	0.519	0.486	0.455	0.434	0.429	0.436	0.457	0.487	0.519	0.374
50	0.553	0.521	0.483	0.460	0.452	0.460	0.484	0.518	0.552	0.391
60	0.590	0.554	0.515	0.486	0.477	0.486	0.515	0.552	0.591	0.409
70	0.632	0.592	0.552	0.517	0.505	0.517	0.551	0.592	0.634	0.429
80	0.686	0.640	0.595	0.556	0.539	0.555	0.597	0.640	0.683	0.456
90	0.762	0.712	0.664	0.617	0.592	0.614	0.660	0.714	0.760	0.493
95	0.831	0.776	0.724	0.665	0.637	0.668	0.721	0.776	0.824	0.526
99	0.966	0.907	0.848	0.763	0.732	0.775	0.838	0.903	0.949	0.592

For explanation, see the note for Table 3.2.5.

Table 3.3.1 The Estimates of the Long Run Variance

A. m = 2

ρ	θ=0			θ=0.5			θ=-0.5		
	σ^2	NW	Adap	σ^2	NW	Adap	σ^2	NW	Adap
1.00	1	146.22	1.98	0.25	37.39	10.69	2.25	325.42	2.15
0.95	400	25.81	10.13	100	6.87	6.84	900	56.58	4.28
0.90	100	13.43	12.94	25	3.86	3.84	225	29.82	13.93
0.80	25	6.60	6.58	6.25	2.12	2.11	56.25	14.57	14.45
0.00	1	0.99	0.99	0.25	0.59	0.59	2.25	1.88	1.88

B. m = 4

ρ	θ=0			θ=0.5			θ=-0.5		
	σ^2	NW	Adap	σ^2	NW	Adap	σ^2	NW	Adap
1.00	1	240.78	2.62	0.25	61.26	17.13	2.25	536.17	2.45
0.95	400	41.21	15.76	100	10.65	10.61	900	90.63	5.88
0.90	100	20.72	19.93	25	5.64	5.62	225	46.29	21.06
0.80	25	9.49	9.47	6.25	2.78	2.77	56.25	21.28	21.11
0.00	1	0.98	0.98	0.25	0.45	0.45	2.25	2.00	1.99

C. m = 8

ρ	θ=0			θ=0.5			θ=-0.5		
	σ^2	NW	Adap	σ^2	NW	Adap	σ^2	NW	Adap
1.00	1	423.49	3.75	0.25	107.38	29.37	2.25	943.32	2.83
0.95	400	68.47	25.32	100	17.32	17.25	900	150.78	8.36
0.90	100	32.37	31.08	25	8.50	8.47	225	72.52	31.90
0.80	25	13.26	13.22	6.25	3.64	3.63	56.25	30.10	29.84
0.00	1	0.97	0.96	0.25	0.36	0.36	2.25	2.06	2.05

D. m = 12

ρ	$\theta=0$			$\theta=0.5$			$\theta=-0.5$		
	σ^2	NW	Adap	σ^2	NW	Adap	σ^2	NW	Adap
1.00	1	597.83	4.71	0.25	151.39	40.74	2.25	1331.71	3.11
0.95	400	91.60	33.02	100	22.94	22.85	900	201.69	10.19
0.90	100	41.01	39.32	25	10.60	10.56	225	91.83	39.44
0.80	25	15.39	15.35	6.25	4.14	4.12	56.25	35.24	34.93
0.00	1	0.96	0.95	0.25	0.32	0.32	2.25	2.06	2.06

E. m = 16

ρ	$\theta=0$			$\theta=0.5$			$\theta=-0.5$		
	σ^2	NW	Adap	σ^2	NW	Adap	σ^2	NW	Adap
1.00	1	764.01	5.52	0.25	193.35	51.28	2.25	1701.83	3.34
0.95	400	111.20	39.21	100	27.67	27.56	900	244.74	11.54
0.90	100	47.43	45.41	25	12.16	12.11	225	106.04	44.68
0.80	25	16.63	16.58	6.25	4.42	4.41	56.25	38.35	38.00
0.00	1	0.94	0.94	0.25	0.30	0.30	2.25	2.06	2.05

Note: The simulation is based on 1000 replications of the following data generating process,

$$(1-\rho B)\, y_t = (1-\theta B) u_t, \qquad u_t \sim \text{NIID}(0, 1), \; T = 300.$$

σ^2 is the true long run variance; NW is the Newey-West estimator ($a_T = 0$); Adap is the adaptive Newey-West estimator in section 3.3; m is the truncation lag, $[T^{1/4}] = 4$. The numbers presented in the tables are sample averages across 1000 replications.

F. Optimal Quadratic Spectral Estimator

With Andrews Automatic Bandwidth Estimator

ρ	$\theta=0$			$\theta=0.5$			$\theta=-0.5$		
	σ^2	NW	Adap	σ^2	NW	Adap	σ^2	NW	Adap
1.00	1	+	5.55	0.25	485.42	29.57	2.25	+	5.18
0.95	400	367.02	57.54	100	14.34	14.34	900	2550.46	20.41
0.90	100	94.32	85.20	25	6.09	6.09	225	576.14	131.89
0.80	25	24.10	24.11	6.25	2.60	2.60	56.25	125.17	123.44
0.00	1	1.00	1.00	0.25	0.47	0.47	2.25	2.99	2.99

Note: (+) $\geq$ 366720.70. QS is the Optimal Quadratic Spectral estimator with Automatic bandwidth estimator and prewhitening/recoloring; Adap is the adaptive QS estimator. The numbers presented in the tables are sample averages across 1000 replications.

Table 3.3.2 Simulated Values of $\hat{\theta}_1$ and $\hat{\rho}$

ρ	θ=0		θ=0.5		θ=-0.5	
	$\hat{\theta}_1$	$\hat{\rho}$	$\hat{\theta}_1$	$\hat{\rho}$	$\hat{\theta}_1$	$\hat{\rho}$
1.00	0.92	0.98	0.36	0.92	0.98	0.99
0.95	0.37	0.94	0.00	0.75	0.81	0.96
0.90	0.01	0.89	0.00	0.60	0.32	0.94
0.80	0.00	0.79	0.00	0.38	0.00	0.88
0.00	0.00	-0.00	0.00	-0.40	0.00	0.39

Note: $\hat{\theta}_1$ is the weight for the first difference in estimating $\hat{u}_t$, cf. equation (3.3.1), while $\hat{\rho}$ is the first order autoregressive coefficient of y_t. The numbers presented in the tables are sample averages across 1000 replications. For the design of the simulation, see note for Table 3.3.1.

Table 3.4.1 Monte Carlo Simulation on the $\chi^2(1)$ test

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	4.6	10.8	30.8	39.2	46.0	53.4	58.3	64.9	81.2	84.1
	4	4.5	11.8	21.1	30.7	37.1	44.7	50.0	57.5	76.7	80.3
	8	5.9	12.9	14.7	22.5	28.4	35.8	40.8	47.7	68.8	74.3
	12	7.4	14.2	12.7	20.0	24.5	31.4	36.3	42.8	65.3	69.4
	QS	4.6	10.6	6.3	11.2	11.6	16.6	15.5	21.5	34.2	41.9
	∞	4.0	9.1	4.0	8.5	3.4	7.6	1.7	5.2	96.1	96.7
0.5	2	0.3	1.9	26.5	36.1	45.5	54.1	59.9	65.2	81.4	84.1
	4	1.4	3.5	21.0	29.1	37.8	45.3	52.5	59.1	76.7	80.7
	8	2.4	5.2	16.6	23.0	27.6	36.6	43.9	49.7	69.3	74.2
	12	3.7	7.8	14.2	21.7	23.5	32.2	38.1	46.2	64.0	69.9
	QS	1.1	3.2	22.3	30.7	37.4	44.7	49.0	55.4	70.0	75.0
	∞	5.6	9.8	4.4	9.9	2.5	6.3	2.1	5.5	96.8	97.6
-0.5	2	7.3	13.7	32.5	40.9	47.1	52.7	63.1	68.8	78.5	82.4
	4	6.8	12.8	23.3	32.1	37.1	45.1	54.0	60.5	73.7	77.4
	8	6.3	13.2	16.5	25.1	26.8	35.6	44.0	51.5	68.4	72.0
	12	7.3	14.5	14.8	21.4	22.9	30.7	37.3	45.5	64.0	68.8
	QS	2.0	5.5	1.2	3.3	1.2	3.0	4.6	7.2	15.8	20.5
	∞	4.3	9.4	4.2	8.7	2.8	6.5	2.6	5.3	96.5	96.8

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho B)\, y_t = (1-\theta B) u_t, \text{ where } u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300.$$

"m" is the number of autocorrelations used in the Newey-West estimate of long run variance.

The column for m = ∞ uses the true σ^2; it is only meant as a comparison.

Table 3.4.2 Monte Carlo Simulation on the $\chi^2(1)$ test

Using the Adaptive Estimator

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	4.6	10.9	30.9	39.3	46.2	53.6	65.6	71.9	91.6	93.0
	4	4.6	12.0	21.3	30.7	37.6	44.9	57.9	64.5	90.4	91.4
	8	5.9	13.0	14.7	22.5	29.1	36.3	51.2	56.2	88.5	90.1
	12	7.5	14.2	12.8	20.1	24.9	31.6	47.0	53.1	87.0	89.1
	QS	4.6	10.5	6.2	11.4	11.8	16.9	36.2	41.9	85.1	87.5
	∞	4.0	9.1	4.0	8.5	3.4	7.6	1.7	5.2	96.1	96.7
0.5	2	0.3	1.9	26.5	36.1	45.5	54.2	59.9	65.2	81.7	84.3
	4	1.4	3.5	21.0	29.2	37.9	45.3	52.5	59.2	77.0	81.2
	8	2.5	5.2	16.8	23.0	27.7	36.7	43.9	50.0	69.4	74.7
	12	3.7	7.5	14.2	21.7	23.6	32.2	38.5	46.2	64.5	70.0
	QS	1.1	3.2	22.3	30.8	37.1	44.8	49.1	55.2	70.6	75.7
	∞	5.6	9.8	4.4	9.9	2.5	6.3	2.1	5.5	96.8	97.6
-0.5	2	7.4	13.8	32.5	41.0	56.2	62.5	84.0	86.0	96.1	96.5
	4	6.9	12.8	23.4	32.1	49.2	55.5	81.2	84.1	95.7	96.2
	8	6.4	13.5	16.7	25.0	40.4	48.2	77.1	81.1	94.8	96.0
	12	7.3	14.5	15.0	21.5	36.1	44.1	74.7	78.4	94.5	95.7
	QS	1.9	5.5	1.4	3.2	14.1	18.8	66.5	71.6	93.2	94.7
	∞	4.3	9.4	4.2	8.7	2.8	6.5	2.6	5.3	96.5	96.8

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho B)\, y_t = (1-\theta B)u_t, \text{ where } u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300$$

m is the number of autocorrelations used in the Adaptive Newey-West estimate of long run variance. The column for m = ∞ uses the true σ^2; it is only meant as a comparison.

Table 3.4.3. Monte Carlo Simulation on the CUSUM Test

Using the Adaptive Estimator

A. The SUP Test ($\tau = 0$)

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	5.7	9.8	51.6	63.8	80.0	86.4	96.2	98.0	99.9	100.0
	4	5.3	10.1	30.5	42.7	56.3	68.9	88.5	92.4	99.8	99.8
	8	4.2	9.4	14.8	22.7	32.1	42.3	76.1	81.6	99.2	99.5
	12	3.6	9.5	8.9	15.6	21.4	30.5	67.5	75.0	99.0	99.2
	QS	5.3	9.2	1.5	4.5	2.6	4.0	53.5	57.5	98.2	98.4
	∞	5.5	9.5	2.0	4.4	1.6	3.5	1.0	2.2	100.0	100.0
0.5	2	0.1	0.3	39.2	51.9	74.9	83.0	92.2	96.4	99.7	99.9
	4	0.2	0.9	25.4	36.1	53.6	65.5	78.5	86.0	98.3	99.2
	8	0.8	2.5	12.0	22.1	28.3	42.5	54.1	64.5	93.3	95.7
	12	0.9	4.6	7.1	13.8	16.2	26.6	38.7	49.8	85.5	90.5
	QS	0.2	0.9	28.8	40.1	50.2	61.6	67.2	77.9	91.6	94.9
	∞	8.2	14.5	2.4	4.7	1.5	2.9	0.3	0.9	100.0	100.0
-0.5	2	7.2	13.2	54.9	66.0	90.2	93.1	99.5	99.6	100.0	100.0
	4	5.3	10.6	30.9	42.6	78.2	84.6	98.5	98.9	99.9	99.9
	8	3.9	9.2	15.4	23.9	64.0	71.1	96.2	97.4	99.9	99.9
	12	3.8	8.1	10.1	17.2	56.5	63.6	95.3	96.1	99.9	99.9
	QS	0.8	2.1	0.0	0.0	34.4	38.1	90.7	92.1	99.9	99.9
	∞	4.0	8.6	2.6	5.0	1.9	4.4	0.3	1.8	100.0	100.0

B. The RANGE Test ($\tau = 0$)

θ	m	ρ 0.00 5%	10%	0.80 5%	10%	0.90 5%	10%	0.95 5%	10%	1.00 5%	10%
0.0	2	3.5	9.0	66.2	78.2	92.0	96.9	99.0	99.4	100.0	100.0
	4	3.4	8.9	34.7	48.0	70.6	80.2	93.1	96.4	100.0	100.0
	8	2.9	7.9	11.5	21.8	35.7	49.3	80.0	86.5	99.3	99.6
	12	1.7	6.3	4.4	10.8	16.1	29.9	67.6	76.7	99.1	99.3
	QS	4.1	8.6	0.3	1.0	1.7	2.2	53.6	55.6	97.8	98.2
	∞	3.5	9.4	1.2	2.4	0.5	1.8	0.2	0.5	100.0	100.0
0.5	2	0.0	0.3	53.3	65.7	89.4	94.3	98.8	99.9	100.0	100.0
	4	0.3	0.7	30.1	45.3	67.3	79.5	87.8	93.8	99.2	99.5
	8	0.7	2.9	11.7	21.3	30.2	46.5	61.6	75.1	94.6	96.9
	12	1.1	4.5	5.8	13.1	11.8	25.2	34.0	52.0	86.5	92.0
	QS	0.3	0.7	36.8	48.8	61.1	76.8	77.2	85.9	93.9	96.7
	∞	13.7	23.2	2.0	3.6	0.6	1.7	0.0	0.2	100.0	100.0
-0.5	2	7.0	13.9	72.4	82.5	95.5	97.5	99.7	99.7	100.0	100.0
	4	4.4	10.8	42.1	54.6	84.9	90.3	99.3	99.5	100.0	100.0
	8	2.8	7.7	13.8	27.1	68.6	76.5	98.1	98.7	100.0	100.0
	12	2.1	5.8	5.6	13.4	60.2	66.7	97.3	97.9	99.9	100.0
	QS	0.4	1.3	0.0	0.0	34.0	37.1	91.6	92.3	99.9	99.9
	∞	3.5	7.3	1.0	3.5	0.3	1.0	0.0	0.4	100.0	100.0

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho B)\, y_t = (1-\theta B)u_t, \text{ where } u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300$$

m is the number of autocorrelations used in the Newey-West estimate of long run variance. The column for m = ∞ uses the true σ^2; it is only meant as a comparison.

Table 3.4.4 Monte Carlo Simulation on the CUSUM Test

with a Linear Trend Using the Adaptive Estimator

A. The SUP Test ($\tau = 0$)

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	3.1	7.9	66.2	76.4	91.6	96.4	98.2	99.1	100.0	100.0
	4	3.3	7.7	35.5	49.5	67.9	80.6	91.4	94.2	99.0	99.4
	8	2.9	7.2	14.1	24.3	37.2	49.4	74.4	81.4	96.9	98.3
	12	2.4	6.9	6.3	15.7	20.8	32.5	59.5	70.3	95.4	96.2
	QS	3.6	7.9	0.6	2.0	0.8	1.1	41.8	44.4	90.2	91.6
	∞	3.9	7.9	1.4	3.2	0.3	1.0	0.0	0.1	100.0	100.0
0.5	2	0.2	0.2	51.9	64.3	88.9	94.7	97.9	99.3	99.8	99.9
	4	0.4	1.5	32.0	44.2	68.2	79.6	86.9	94.2	97.1	99.1
	8	1.3	3.9	13.9	25.3	35.0	50.3	61.1	71.9	87.1	92.1
	12	2.6	7.4	7.7	16.7	16.2	31.3	35.5	52.9	72.4	81.8
	QS	0.5	1.4	36.9	49.4	65.7	76.7	78.0	86.6	88.1	93.4
	∞	12.3	22.7	1.1	4.3	0.4	1.1	0.0	0.2	100.0	100.0
-0.5	2	7.9	15.1	69.9	78.7	93.7	96.6	99.5	99.8	99.9	100.0
	4	5.3	11.9	41.6	54.8	81.8	87.8	98.8	99.2	99.7	99.2
	8	4.4	9.2	17.7	29.6	64.1	72.1	96.3	97.9	97.6	99.7
	12	2.6	9.0	8.4	17.6	54.2	62.3	94.7	96.3	99.3	99.5
	QS	0.5	1.6	0.0	0.0	27.5	30.1	85.0	85.8	99.0	99.0
	∞	3.8	6.8	1.5	4.3	0.1	0.7	0.3	0.5	100.0	100.0

B. The Range Test (τ = 0)

		ρ									
		0.00		0.80		0.90		0.95		1.00	
θ	m	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	2	3.5	8.1	67.9	79.2	93.1	97.5	98.5	99.4	100.0	100.0
	4	3.1	8.3	36.2	50.7	70.7	80.9	91.1	95.6	99.2	99.7
	8	2.4	8.4	13.4	22.7	35.7	49.3	74.3	81.6	97.1	98.2
	12	2.5	7.6	5.9	13.2	18.1	30.1	60.5	68.1	94.9	96.3
	QS	3.6	7.9	0.4	2.0	0.8	1.4	41.5	44.3	90.3	91.0
	∞	3.3	8.4	1.1	2.3	0.2	1.0	0.0	0.2	100.0	100.0
0.5	2	0.2	0.2	54.0	66.9	90.8	95.0	98.5	99.5	99.9	99.9
	4	0.5	1.3	32.9	45.0	70.5	80.9	89.1	94.4	98.5	99.2
	8	1.5	5.0	14.5	23.9	34.0	48.5	60.3	72.2	88.0	92.7
	12	2.9	6.8	8.4	15.9	16.0	29.7	36.2	51.6	71.7	82.6
	QS	0.6	1.2	37.7	50.8	66.8	77.6	79.4	87.1	89.2	94.3
	∞	14.5	23.8	1.6	3.6	0.4	0.8	0.0	0.2	100.0	100.0
-0.5	2	8.2	15.8	71.1	81.4	94.3	97.4	99.8	99.9	100.0	100.0
	4	6.2	12.7	42.1	55.1	82.6	88.9	98.9	99.2	99.7	99.7
	8	4.2	9.6	16.1	27.6	64.0	71.7	96.7	97.8	99.6	99.7
	12	3.0	9.0	7.9	15.9	55.1	62.5	94.9	96.1	99.5	99.6
	QS	0.6	1.3	0.0	0.0	27.3	30.0	84.8	85.7	98.9	99.0
	∞	4.0	7.6	1.4	2.9	0.2	0.6	0.3	0.6	100.0	100.0

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho B)\, y_t = (1-\theta B)u_t, \text{ where } u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300$$

m is the number of autocorrelations used in the Newey-West estimate of long run variance. The row for m = ∞ uses the true σ^2; it is only meant as a comparison. The row QS uses the QS estimate of Andrews (1991); see appendix B for details.

CHAPTER 4

CAUCHY TEST FOR TREND STATIONARITY AGAINST THE UNIT ROOT MODEL

4.1 Introduction

The traditional approach to unit root testing can be characterized as follows. First, the null hypothesis is taken as the unit root model. Second, the asymptotic distribution of the test statistics under the null is a function of Brownian motion and tabulation of critical values by simulation is needed. Third, the asymptotic distribution of the test statistic under the null depends on the specification of the trend component under the alternative, i.e. whether a nonzero mean and/or linear trend are allowed, cf. Dickey and Fuller (1979, 1981), Phillips (1987), Phillips and Perron (1988); the order of polynomial trend, cf. Ouliaris, Park and Phillips (1988); and whether breaks in the trend are allowed, cf. Perron (1989).

As an extension of the unit root literature, the test for cointegration also takes the unit root as the null model. Therefore, as a test for an economic theory which suggests cointegration among economic variables, it takes as the null that the theory is false.

Recently Bierens (1991a) has proposed a new approach to the unit root testing which takes issue with the first two features of the traditional unit root testing. It takes stationarity as the null model and unit root as the alternative, with the asymptotic distribution of the test statistic under the null being standard Cauchy. Therefore tabulation of critical values by simulation is unnecessary. However Bierens' Cauchy test does not have any power to distinguish trend stationarity from unit root with drift, which are the two primary competing models for most historical macroeconomic time series.

The approach taken by Bierens implicitly uses the variable addition method proposed by Park (1990). First, the time series of interest is regressed on some trend component and some superfluous regressors, then the coefficient estimates on the superfluous regressors are used to construct the test statistic. The nuisance parameter appears in the asymptotic distribution proportionally. If another asymptotic normal random variable can be constructed that also depends on the nuisance parameter proportionally, then the quotient of the two random variables is asymptotically free from the nuisance parameter. If the two random variables are asymptotically independent with equal variance, their quotient is asymptotically standard Cauchy. This is the approach taken by Bierens (1991a).

The test statistic proposed by Park (1990) is the adjusted F (or Wald) statistic on the significance of the superfluous regressors. Under stationarity, it is distributed as Chi-square with the degree of freedom being equal to the number of superfluous regressors. However, an estimate of the long run variance is needed which can cause some problem in finite samples, because its convergence to the asymptotic distribution may be slow. This is also part of the problem with the Phillips-Perron test. Park's test allows very general deterministic trend under both the null and the alternative.

This chapter is based upon and is a generalization of Bierens' work. We will construct several tests for stationarity against unit root. It takes the variable addition approach of Park and proposes a Cauchy test for stationarity with a maintained deterministic time trend. We form the two asymptotic normal quantities in two ways. First, we fit some deterministic trend and superfluous trend; the coefficient on the superfluous trend forms one asymptotic normal quantity, the other asymptotic normal quantity is the partial sum of detrended series using the correct trend specification. Second, an orthogonal polynomial regression is estimated; the two asymptotic normal quantities are just the coefficient estimates on superfluous trends. The asymptotic distribution of our test statistics is standard Cauchy and no estimate of the long run variance is needed.

The asymptotic theory and a test statistic using the first approach are developed in Section 4.2 The test statistic using orthogonal polynomial regression is developed in Section 4.3. We give further discussion of the empirical results in Section 4.4. Section 4.5 summarizes and concludes this chapter. The proof of all theoretical results is collected in the appendix to the chapter.

4.2 Stationarity and Unit Root Around a General Deterministic Trend

4.2.1 The Model

The two competing models for Y_t are stationarity,

$$Y_t = u_t \tag{4.2.1}$$

and unit root

$$\Delta Y_t = u_t. \tag{4.2.2a}$$

The unit root model can also be written as,

$$Y_t = Y_0 + \Sigma_{j=1}^{t} u_j \tag{4.2.2b}$$

By the invariance principle (Theorem 2.2.4),

$$\Sigma_{j=1}^{t} u_j \simeq t^{1/2}\sigma W(1), \tag{4.2.3}$$

so the stochastic trend is asymptotically equivalent to a square root trend in probability. In a regression of y_t on an intercept and $t^{1/2}$, the coefficient on the time trend $t^{1/2}$ approaches zero under the null, but approaches a nondegenerate normal distribution under the unit root. It turns out, however, that all trending regressors, deterministic or stochastic, have the ability to distinguish stationarity from unit root. Let the null hypothesis be that Y_t is trend stationary, (4.2.4) $$H_0\text{: } Y_t = f(t/T)b + u_t,\ u_t \sim \text{Assumption 2.2.1},$$

and the alternative be that Y_t is generated by a unit root process around a deterministic trend,

(4.2.5) $$H_1\text{: } Y_t = f(t/T)b + X_t,\ \Delta X_t \sim \text{Assumption (2.2.1).}$$

We will run the following auxiliary regression,

(4.2.6) $$y_t = f(t/T)b + g(t/T)\beta + e_t$$

where f(t/T) is the maintained deterministic trend and g(t/T) represents the superfluous trend. If there are multiple superfluous regressors, then $g(\cdot)$ is a vector function. In this section, we take $g(\cdot)$ as a scalar function. In Section 4.3, we study the case of multiple superfluous regressors. The OLS estimate of $\hat{\beta}$ is, noting that $\beta = 0$ under both H_0 and H_1,

(4.2.7) $$\hat{\beta} = [\Sigma g^*(t/T)'g^*(t/T)]^{-1}[\Sigma g^*(t/T)'e_t]$$

where $g^*(t/T)$ is the linear projection residual of g(t/T) on f(t/T). In Bierens (1991a), f(t) = 1 under stationarity; and f(t) does not exist under the unit root. So Bierens' results are restrictive in that no deterministic trend is allowed under either the null or the alternative.

4.2.2 The Asymptotic Theory

The asymptotic distribution of $\hat{\beta}$ is given in Theorem 4.2.1.

Theorem 4.2.1 (asymptotic distribution of $\hat{\beta}$)

(i) Under H_0, $\sqrt{T}\hat{\beta} \Rightarrow N(0, \sigma^2 s_1^2)$;

(ii) under H_1, $(1/\sqrt{T})\hat{\beta} \Rightarrow N(0, \sigma^2 s_2^2)$;

where s_1^2 and s_2^2 are given in the Appendix.

So the rate of convergence of $\hat{\beta}$ is different under the null and the alternative. Specifically under H_0, $\hat{\beta}$ approaches zero at rate $\sqrt{T}$, because OLS is consistent despite the serial correlation in u_t.

$\sqrt{T}\hat{\beta}$ has a nondegenerate distribution. Under H_1, $(1/\sqrt{T})\hat{\beta}$ approaches a nondegenerate distribution, $\sqrt{T}\hat{\beta}$ is divergent. A test based on $\sqrt{T}\hat{\beta}$, would be consistent if the nuisance parameter σ^2 can be estimated. Actually the test proposed by Park (1990) is just the corrected F test for ($\beta = 0$) which is asymptotically distributed as χ_1^2. Note that $\beta = 0$ under both H_0 and H_1. A test can not be based on the coefficient of the maintained time trend (b) since b is unknown.

The asymptotic distributions in Theorem 4.2.1 are Gaussian with nuisance parameter σ^2. We follow Bierens (1991a) to construct a Cauchy test. Therefore we need to define another statistic ξ which behaves similarly to $\sqrt{T}\hat{\beta}$, i.e. having a limiting normal distribution under H_0 and divergent under H_1. If ξ were defined as in Bierens,

$$\xi(\tau) = [T\tau]^{-1}\Sigma_{t=1}^{[T\tau]}(y_t-\bar{y}),$$

then since the trend component is present under both the null and the alternative, $\xi(\tau)$ behaves similarly asymptotically. So $\xi(\tau)$ does not have any power to distinguish the two models. However, if the series Y_t is detrended first, then the residuals should behave differently under the null and the alternative. Because under H_0 the residual should be stationary, while under H_1 it contains a stochastic trend. Accordingly we define $\xi(\tau)$ as follows,

$$\xi(\tau) = [T\tau]^{-1}\Sigma_{t=1}^{[T\tau]}y_t^*$$

where y_t^* is the linear projection residual of y_t on $f(t/T)$. The joint asymptotic distribution of $\hat{\beta}$ and ξ is given below.

Theorem 4.2.2

(i) Under H_0, $\sqrt{T}(\hat{\beta}, \xi(\tau))' \Rightarrow N(0, \sigma^2 Q_1(\tau))$;

(ii) under H_1, $(1/\sqrt{T})(\hat{\beta}, \xi(\tau))' \Rightarrow N(0, \sigma^2 Q_2(\tau))$;

where Q_1 and Q_2 are given in the Appendix.

Now we orthogonalize the two normal random variables to construct a Cauchy random variable. Define

$$(\gamma_{1T}(\tau), \gamma_{2T}(\tau))' = L^{-1}\sqrt{T}(\hat{\beta}, \xi(\tau))' \sim N(0, \sigma^2 I_2)$$

where the matrix L is such that $Q_1 = LL'$. For example, the Cholesky factor of Q_1 can be chosen as the matrix L. However as noted in the Appendix, the need for matrix decomposition can be avoided in certain circumstances. The asymptotic distributions of γ_{1T} and γ_{2T} are given in the following theorem.

Theorem 4.2.3

(i) Under H_0, $(\gamma_{1T}(\tau), \gamma_{2T}(\tau))' \sim N(0, \sigma^2 I_2)$;

(ii) under H_1, $T^{-1}(\gamma_{1T}(\tau), \gamma_{2T}(\tau))' \Rightarrow \sigma(\zeta_1, \zeta_2)'$;

where (ζ_1, ζ_2) is multivariate normal.

4.2.3 The Test

Now we define the test statistic. First let

$$\xi^* = \frac{\gamma_{2T}}{1+T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^{-2}}$$

where $\hat{\sigma}_\Delta^{-2}$ is a consistent estimate of the variance of Δy_t^*. Note that the denominator of ξ^* is invariant to a linear transformation of y_t. Its limiting distribution is given in the next theorem.

Theorem 4.2.4

(i) Under H_0, $(\gamma_{1T}(\tau), \xi^*)' \Rightarrow N(0, I_2)$;

(ii) under H_1, $\xi_T^* \Rightarrow \sigma_u^2 /(\sigma\zeta_2)$.

Define the statistic,

(4.2.8) $$C_1 = \frac{\gamma_{1T}}{\xi^*} = \frac{\gamma_{1T}}{\gamma_{2T}}\left[1+T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^{-2}\right]$$

which is invariant to a linear transformation of y_t. Since ξ^* converges to a nondegenerate distribution under both the null and the alternative, the asymptotic behavior of C_1 is dominated by that of γ_{1T}. The major result is presented in the following theorem.

Theorem 4.2.5

(i) Under H_0, $C_1 \Rightarrow \text{Cauchy}(0, 1)$;

(ii) under H_1, $T^{-1}C_1 \Rightarrow (\sigma^2/\sigma_u^2)(\zeta_1\zeta_2)$.

Note that the Cauchy distribution is the Student t distribution with one degree of freedom. The Cauchy test is free of nuisance parameters. Specifically its asymptotic distribution under the null does not depend on the long run variance, so in contrast to the Phillips-Perron test and the Park χ^2 test, our test does not need the estimate of the long run variance. It is strongly consistent against the unit root model. The asymptotic distribution of the test statistic does not depend on the use of a specific set of superfluous regressors in general. Of course, the value of the test statistic is going to be different through the matrices Q_1 and L for different $g(\cdot)$ functions. The difference in the rate of convergence of the test statistic, essentially resulting from Theorems (4.2.2)(ii) and (4.2.3)(ii), does not depend on any nuisance parameter, so the power of the test in large samples is not related to any nuisance parameter.

However there is some arbitrariness in the construction of ξ^* and C_1. For example let ϵ be any positive number, then the new statistic

$$\frac{\gamma_{1T}}{\gamma_{2T}}(1+\epsilon^2T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^{-2}) \Rightarrow \text{Cauchy}(0, 1) \text{ under } H_0.$$

But for large ϵ, the term $(1+\epsilon^2T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^{-2})$, although it asymptotically equals to 1 under H_0, can be very large in finite samples. For small ϵ, the term $(1+\epsilon^2T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^{-2})$, although it asymptotically diverges under H_1, can be very small in finite samples. So even though the test is asymptotically invariant to a linear transformation of Y_t, in finite samples it can be very dependent on the way the test is constructed. Some simulation on the magnitude of $(1+T^{-1}\gamma_{2T}^{2}\hat{\sigma}_{\Delta}^{-2})$ is collected in Table 4.3.2 for the test in Section 4.3.

Although the nuisance parameters asymptotically vanish under the null, they may affect the power of the test in finite samples as they appear in the asymptotic distribution under the alternative, cf. Theorem 4.2.5 (ii). For the moment, let

$$(\gamma_{1T}^{*}, \gamma_{2T}^{*})' = \sigma^{-1}(\gamma_{1T}, \gamma_{2T})' \Rightarrow N(0, I_2)$$

then

$$C_1 = \frac{\gamma_{1T}}{\xi^{*}} = \frac{\gamma_{1T}^{*}}{\gamma_{2T}^{*}}\left[1+(T^{-1}\gamma_{2T}^{*2})\frac{\sigma^2}{\hat{\sigma}_{\Delta}^{2}}\right].$$

If $(\sigma^2/\sigma_{\Delta}^2)$ is large, then there may be substantial size distortion. If $(\sigma^2/\sigma_{\Delta}^2)$ is small, then the power of the test in finite samples is low. This is confirmed by the calculation in Bierens (1991a, Table 3). For the ARMA(1,1) process $u_t = \phi u_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$, $\epsilon_t \sim \text{IID}(0, 1)$,

(4.2.9a) $$\frac{\sigma^2}{\sigma_{\Delta}^2} = \left[\frac{1-\theta}{1-\phi}\right]^2 \left[\frac{1-\phi^2}{2[1-\phi^2+(\theta-\phi)(1-\phi)(1+\theta)]}\right]$$

(4.2.9b) $$\frac{\sigma^2}{\sigma_{u}^2} = \left[\frac{1-\theta}{1-\phi}\right]^2 \left[\frac{1-\phi^2}{1-\phi^2+(\theta-\phi)^2}\right]$$

They are going to be used in the discussion of the simulation results.

4.2.4 Monte Carlo Simulation

The data generating process is the same as in Bierens (1991a),

$$(1-\phi B)(1-\rho B)\, Y_t = (1-\theta B)\, \epsilon_t, \qquad \epsilon_t \sim \text{NIID}(0, 1) \qquad t = 1, 2, \cdots, T$$

$\phi = \{0, 0.8, 0.9, 0.95\}$, $\rho = \{0, 0.8, 0.9, 0.95, 0.99, 1\}$, and $\theta = \{0, 0.5\}$, $T = \{100, 300, 500\}$. The case of $T = 500$ is intended to capture some asymptotic properties of the test. In the formal notation of equations (4.2.4) and (4.2.5),

$$\text{if } \rho = 1, \qquad u_t = \Delta Y_t, \qquad (1\text{-}\phi B)u_t = (1\text{-}\theta B)\epsilon_t$$

$$\text{if } \rho < 1, \qquad u_t = Y_t, \qquad (1\text{-}\phi B)(1\text{-}\rho B)u_t = (1\text{-}\theta B)\epsilon_t.$$

Note that the test is invariant to the coefficient b on the maintained trend in equations (4.2.4) and (4.2.5), so it is set to zero in the simulation for simplicity. In the simulation, we set $f(t) = (1\ t)$ to capture a linear trend, and $g(t) = t^2$ for the spurious regressor, and $\tau = 0.5$ for the construction of $\xi(\tau)$.

Table 4.2.1 presents the results of the Monte Carlo simulation. The empirical size is close to the nominal size except for the near-unit root case ($\rho = 0.95, 0.99$), or when the stationary process has high persistence (ϕ and ρ are high). However the test has almost no power against the random walk model, which is not surprising. But the power against general ARIMA models seems satisfactory. The test has some power to distinguish stationary processes with moderate persistence ($\rho = 0.8$, $\phi = 0.8$) from a genuine unit root ($\rho = 1$, $\phi = 0.8$). Note that the size and power results of our test with a maintained linear trend are very close to those of Bierens (1991a) with no trend. The results with $g(t) = t^3$ and $g(t) = t^5$ are similar, as suggested by the theoretical argument, and are not reported.

The quasi-nuisance parameters in Equation 4.2.9 are numerically listed in Table 4.2.2. This explains the low power against the random walk models and the substantial size distortion for near-unit root models. For a random walk model, σ^2/σ_u^2 is one, which makes the multiplier (the term in the bracket) in Equation 4.2.8 small, resulting in not enough rejection of the null. For near-unit root models (with both ρ and ϕ large), σ^2/σ_Δ^2 is huge, which makes the multiplier in Equation 4.2.8 substantially larger than one, resulting in excess rejections of the null. As Bierens (1991a) has pointed out, the power of the test is good against unit root alternatives with strong persistence, but its power is low against alternatives with weak persistence.

Applying the filer (1-0.5B) to the series always reduces the quasi-nuisance parameters, and this also translates into size and power. For stationary processes with low serial correlation, the reduction in $(\sigma^2/\sigma^2_{\Delta u})$ is small, and the size is not affected much by the transformation. For stationary processes with high persistence, the reduction is much larger. Although the relative change is small, the size distortion is smaller after applying the filter. For an alternative model with high persistence, applying the filter substantially reduces the nuisance parameter (σ^2/σ^2_u). Since the nuisance parameter remains large, the power of the test is not affected. For an alternative model with low serial correlation, applying the filter substantially reduces the nuisance parameter.

The discussion above leads to the following recommendations: (i) if the series of interest has high persistence and the first difference still has high persistence, then apply the filter (1-0.5B); (ii) if the series itself has high persistence but the first difference of the series does not have high persistence (near-unit root vs random walk), do the test both with and without the filter, and be careful in interpreting the results; (iii) in other cases, whether one applies the filter or not does not affect the results.

4.3 Tests Using Orthogonal Polynomials

The key to constructing the Cauchy test is the derivation of two random variables that are asymptotically normally distributed. The quality of the approximation of asymptotic distribution in finite samples depends on the various FCLT's. In the construction of ξ in the last section, only a fraction (τ) of the observations are used. This approximation can possibly be improved by using more observations in the construction of the partial sum. However the partial sum of the OLS residuals quickly approaches zero (if an intercept is included in f(t)). Therefore we must construct the normal quantity in other ways.

If the maintained trend $f(\cdot)$ and the superfluous trend $g(\cdot)$ are orthogonal, then from Equation 4.2.7, (4.3.1) $$\hat{\beta} = [\Sigma g(t/T)'g(t/T)]^{-1}[\Sigma g(t/T)'e_t].$$

Therefore specification of the maintained trend is not necessary, as long as the superfluous trend is known to be orthogonal to the possibly unknown true trend. If it were not for the quantity $\hat{\sigma}_\Delta^{-2}$ in the definition of ξ^* which makes the test statistic invariant to a linear transformation, then the true trend $f(\cdot)$ would be irrelevant. Bierens (1992b) uses a system of orthogonal Chebishev polynomials to approximate the unknown trend $f(\cdot)$. When the approximation is close enough, the variable addition approach can be used.

We can include multiple superfluous regressors in the auxiliary regression, and orthogonalize (asymptotically) the coefficients on the superfluous regressors in the construction of the Cauchy test. In this section, we represent the maintained trend by a set of orthogonal polynomials; additional multiple orthogonal polynomials are used as superfluous regressors. The test is constructed by using the coefficients on the superfluous regressors. The maintained trend is assumed to be a linear time trend.

4.3.1 The Cauchy Test

Let the orthogonal polynomials be as given in Section 2.3.2. The asymptotic distribution of the OLS estimate of the orthonormal regression, cf. Equation 2.3.8,

$$y_t = \Sigma_{i=0}^{3} b_i p_{i,T}^*(t/T) + e_t,$$

is normal with the covariance matrix being the identity matrix scaled by σ^2, cf. Theorem 2.3.1. Let $\gamma_{1T} = \hat{b}_2$, $\gamma_{2T} = \hat{b}_3$ as in Section 4.2. Then the Cauchy test can be constructed the same as in Section 4.2. The resulting Cauchy test statistic is denoted as C_2.

4.3.2 A t-test

Since the coefficients of all the superfluous regressors are zero, if more superfluous regressors are included in the regression, then the power of the test may be improved.

Continuing the discussion in the last subsection, if k (≥ 2) superfluous orthonormal polynomials are included in the regression. The asymptotic IID normal coefficients estimates $(\hat{b}_2, \hat{b}_3, \cdots, \hat{b}_{k+1})'$ can easily be used to construct a t-test.

$$\text{Under } H_0, \qquad \frac{\hat{b}_2}{\sqrt{\Sigma_{i=3}^{k+1}\hat{b}_i^2}} \sim t_{(k-1)}.$$

A consistent t test can be constructed as in Equation 4.2.8, for example,

$$C_2^* = \frac{\hat{b}_2}{\sqrt{\Sigma_{i=3}^{k+1}\hat{b}_i^2}}\left[1+T^{-1}[\Sigma_{i=3}^{k+1}\hat{b}_i^2]\hat{\sigma}_\Delta^{-2}\right]. \tag{4.3.2}$$

It is trivial to show that

$$\text{under } H_0, \qquad C_2^* \Rightarrow t_{(k-1)},$$

$$\text{under } H_1, \qquad T^{-1}C_2^* \Rightarrow \text{nondegenerate distribution.}$$

Note that Bierens (1992b) has used a similar "approximate" t-test, where the maintained trend $f(\cdot)$ is assumed to be unknown and is approximated by a set of Chebishev polynomials.

4.3.3 Monte Carlo Simulation

The results of Monte Carlo simulation on C_2 are presented in Table 4.3.1 The results in general are the same as those for the earlier tests; namely the size is close to the nominal size but the power against the random walk model is low. One advantage of the test is that it treats the near unit root model as stationary (at the expense of also treating the random walk model as stationary), in contrast to the Dickey-Fuller test that treats near unit root model as unit root process.

Table 4.3.2 collects simulation results on the magnitude of the multiplier $(1+T^{-1}\gamma_{2T}^2\hat{\sigma}_\Delta^{-2})$. The pattern in the sample mean of the multiplier roughly corresponds to the pattern of empirical size and power in Table 4.3.1. Simulation shows that γ_{1T} and γ_{2T} are approximately distributed as normal in finite samples with equal variance; therefore, the size distortion and the lack of power of the test is caused by the ill behavior of the multiplier. Table 4.3.3 contains some statistics on the quality of approximation for the asymptotic distribution of $\hat{b}$ in Theorem 2.3.1(i). By Theorem 2.3.1 with σ^2=1, all the entries in Table 4.3.3a corresponding to $\rho \neq 1$ should be zero, and all the entries in Table 4.3.3b corresponding to $\rho \neq 1$ should be one. Results in the table indicate that the asymptotic approximation is rather poor in finite samples (T = 100 here).

Table 4.3.4 contains the simulation result on the t_4 test of Equation 4.3.2. Indeed the test is more powerful than the Cauchy test, since the additional superfluous regressors provide more information on the spurious nature of the trend under the unit root.

Given the low power of the test against the random walk model, the finding of a unit root should be regarded as strong evidence for unit root, just as the rejection of unit root by the Dickey-Fuller test should be regarded as strong evidence for stationarity.

4.4 Further Discussion and Other Tests for Stationarity

Unit root tests are designed to distinguish different characterizations of trend in a stochastic process, including stochastic trend (random walk) and deterministic trend (e. g. time polynomial, polynomial with breaks, or seasonality). In order to test for the presence (unit root) or the absence (stationarity) of stochastic trend, an adequate specification of the deterministic trend is necessary. Overspecification of the deterministic trend does not affect the test asymptotically. However, underspecification has serious consequences. Since all the unit root tests rely on the notion that the variance of a unit root process increases linearly while the variance of a stationary process remains constant, if there is any trend remaining in the series, the test would tend to interpret it as evidence of nonstationarity, that is, a unit root in our context. Therefore, a nonrejection of the unit root model by the traditional test does not necessarily imply the existence of a unit root; similarly the rejection of the null of stationarity by our test does not necessarily imply the existence of a unit root. This is just like the existing many classes of model specification tests. Since the test can have power against a class of alternatives, the rejection of the null model only implies that the null model is not an adequate characterization of the data; it does not necessarily imply the correctness of a specific alternative model.

Since a process with seasonal unit root (complex unit root) has the same degree of nonstationarity as a unit root process, our test also has power against the seasonal unit root model. This reinforces our earlier point that the rejection of the null model does not necessarily imply the validity of a specific alternative model.

Kahn and Ogaki (1991) propose a test for stationarity based on the asymptotic distribution theory for their unit root test (Kahn and Ogaki, 1990). As for the χ^2 test of Park (1990), the need to estimate the long run variance makes the test consistent at rate T^δ ($0 < \delta < 1$). While our test is consistent at rate T, its power against the random walk model in finite samples is low. Note that the LM test for stationarity in Kwiatkowski, Phillips, Schmidt and Shin (1992) is also consistent at rate T^δ.

The asymptotic distribution of our test statistic does not depend on the specification of the maintained deterministic trend under both the null and the alternative. Suppose the deterministic trend is a linear trend with a break in the intercept and/or slope and the break point is known. Then the Cauchy test for stationarity around the breaking trend depends on the break point, but its asymptotic distribution is still Cauchy. Such is not the case with the traditional test for unit root. The traditional test for unit root around a breaking trend is a function of the break point; its asymptotic distribution is a random function of the break point, cf. Perron (1989). In order to make the asymptotic distribution of the Perron-type test free of the break point, the break point has been assumed to be deterministic but unknown. Since the break point is unknown, it needs to be estimated. To account for the estimation of the break point, the modified Perron-type test statistic is taken as a continuous functional of the Perron-type test in certain regions of the parameter space for the break point. By the continuous mapping theorem, the asymptotic distribution of the modified Perron-type test is a functional of the asymptotic distribution of the Perron-type test. It is free of nuisance parameter — the break point. It is not clear how our test can be conducted for a deterministic but unknown break point.

4.5 Summary and Conclusion

Both the famous Dickey-Fuller test and the Phillips-Perron test for unit root take the unit root model as the null; the asymptotic distributions of the test statistics are nonstandard, are dependent on nuisance parameters, and are also dependent on the alternative model. Bierens (1991a) constructed a test which takes the stationarity as the null model and the unit root with drift as the alternative model. The asymptotic distribution is standard Cauchy.

Using the variable addition approach of Park (1990), we propose some generalizations of the Bierens Cauchy test with maintained linear trend under both the null and the alternative. Monte Carlo simulation indicates that the test has reasonable power to distinguish a stationary process with strong persistence against a unit root process. Like the Dickey-Fuller test, it does not have any power to distinguish the near-unit root process from a random walk.

The framework used in this chapter can also be used to test for stationarity around a stochastic trend, i.e. cointegration, as in Engle and Granger (1987). But the power of the test is likely to be very low, so we do not pursue it further here.

Appendix

Proof of Theorem 4.2.1

Note that

$$T\hat{\beta} = [T^{-1}\Sigma g^*(t/T)'g^*(t/T)]^{-1}[\Sigma g^*(t/T)'e_t].$$

From the definition of the Riemann integral,

$$T^{-1}\Sigma g^*(t/T)'g^*(t/T) \rightarrow \int g^{*\prime}g^*.$$

Under H_0, $e_t = u_t$,

$$(1/\sqrt{T})\Sigma g^*(t/T)'e_t \Rightarrow \sigma \int g^{*\prime}dW \qquad \text{by Theorem 2.2.6 (ii)}$$

$$\sqrt{T}\hat{\beta} \Rightarrow \sigma[\int g^{*\prime}g^*]^{-1}[\int g^{*\prime}dW] \sim N(0, s_1^2), \qquad \text{by Theorem 2.2.6 (ii)}$$

where $s_1^2 = \sigma^2[\int g^{*\prime}g^*]^{-1}$ by Theorem 2.2.3 (i).

Under H_1, $e_t = \Sigma_j u_j$,

$$T^{-3/2}\Sigma g^*(t/T)'e_t \Rightarrow \sigma \int g^{*\prime}W, \qquad \text{by Theorem 2.2.5(ii)}$$

$$(1/\sqrt{T})\hat{\beta} \Rightarrow \sigma[\int g^{*\prime}g^*]^{-1}[\int g^{*\prime}W],$$

and s_2^2 can be calculated using Theorem 2.2.3(ii). □

Proof of Theorem 4.2.2

Note that

$$\sqrt{T}\xi = (T/[T\tau])\ T^{-1/2}\Sigma_{t=1}^{[T\tau]}y_t^*.$$

(i) Under H_0,

$$T^{-1/2}\Sigma_{t=1}^{[T\tau]}y_t^* \Rightarrow \sigma\tilde{W}(\tau),$$

where

$$\tilde{W}(\tau) \equiv W(\tau) - [\int_0^\tau f][\int_0^1 f'f]^{-1}[\int_0^1 f'dW] \sim N(0, s_3^2),$$

$$s_3^2 = \tau - [\int_0^\tau f][\int_0^1 f'f]^{-1}[\int_0^\tau f'].$$

Note that $\tilde{W}(r)$ is different from the detrended Brownian motion $W^*(r)$ in Theorem 2.4.1, because $W^*(r)$ is the limiting distribution of standardized partial sums under the assumption of a unit root, while $\tilde{W}(r)$ here is the limiting distribution under stationarity. From the above results,

$$T^{1/2}\xi \Rightarrow \tau^{-1}\sigma\tilde{W}(\tau) \sim N(0, \tau^{-2}\sigma^2 s_3^2).$$

The convergence of $\hat{\beta}$ in Theorem 4.2.1 (i) and ξ here is also joint, and their covariance is, from Theorem 2.2.3,

$$s_{13} = \tau^{-1}\sigma^2[\int g^{*\prime}g^*]^{-1}\{[\int_0^\tau g^{*\prime}] - [\int_0^\tau g^{*\prime}f][\int_0^1 f'f]^{-1}[\int_0^\tau f']\}.$$

Define Q_1 as

$$Q_1 = \begin{bmatrix} s_1^2 & s_{13} \\ s_{13} & s_3^2 \end{bmatrix}.$$

Then the result under review follows immediately.

(ii) Under H_1,

$$T^{-3/2}\Sigma_{t=1}^{[T\tau]} y_t^* \Rightarrow \sigma\int_0^\tau W^*(r). \qquad \text{by equation (2.4.3) and CMT}$$

So,

$$(1/\sqrt{T})\xi \Rightarrow \tau^{-1}\sigma\int_0^\tau W^*(r).$$

Note that if the maintained trend is a linear trend, $f(r) = (1\ r)$, the superfluous trend is a quadratic trend, $g(r) = (r^2)$, and the partial sum is the half-way sum, $\tau = 1/2$, then the covariance s_{13} is equal to zero. □

Proof of Theorem 4.2.3

(i) obvious from the definition of γ_{1T}, γ_{2T} and L.

(ii) follows from Theorems 4.2.1 (ii) and 4.2.2 (ii),

$$T^{-1}(\gamma_{1T}, \gamma_{2T})' \Rightarrow \sigma L^{-1}([\int g^{*\prime}g^{*}]^{-1}[\int g^{*\prime}dW]', \tau^{-1}\sigma\int_0^\tau W^{*}(r))'$$

$$\stackrel{\text{def}}{=} \sigma(\zeta_1, \zeta_2)'. \qquad \square$$

Proof of Theorem 4.2.4

First we study the asymptotic behavior of $\hat{\sigma}^2_\Delta$,

$$Y_t^* = Y_t - T^{-1}[f(t/T)][T^{-1}\Sigma f(t/T)'f(t/T)]^{-1}[\Sigma f(t/T)'Y_t]$$

$$= e_t - T^{-1}[f(t/T)][T^{-1}\Sigma f(t/T)'f(t/T)]^{-1}[\Sigma f(t/T)'e_t],$$

$$\Delta Y_t^* = \Delta e_t - T^{-1}[f(t/T) - f((t-1)/T)][T^{-1}\Sigma f(t/T)'f(t/T)]^{-1}[\Sigma f(t/T)'e_t]$$

$$= \Delta e_t - T^{-1}[f^{(1)}(t/T)(1/T)+O(T^{-2})][(\int f'f)^{-1}+o(1)][\Sigma f(t/T)'e_t],$$

where $f^{(1)}(r) = df(r)/dr$ is the first order derivative.

(i) Under H_0, $e_t = u_t$.

$$\Delta Y_t^* = \Delta u_t -$$

$$T^{-1/2}[f^{(1)}(t/T)(1/T)+O(T^{-2})][(\int f'f)^{-1}+o(1)][T^{-1/2}\Sigma f(t/T)'u_t]$$

$$= \Delta u_t - T^{-3/2}[f^{(1)}(t/T)(\int f'f)^{-1}\sigma W(1) + o_p(1)]$$

$$= \Delta u_t - O_p(T^{-3/2}),$$

Note that

$$\hat{\sigma}^2_\Delta = T^{-1}\Sigma_{t=2}^T(\Delta y_t^*)^2$$

$$= T^{-1}\Sigma_{t=2}^{T}(\Delta u_t)^2 + O_p(T^{-3}) - 2(T^{-1}\Sigma_{t=2}^{T}\Delta u_t)O_p(T^{-3/2}).$$

By the Law of Large Numbers (LLN),

$$T^{-1}\Sigma_{t=2}^{T}\Delta u_t \xrightarrow{p} 0,\ T^{-1}\Sigma_{t=2}^{T}(\Delta u_t)^2 \xrightarrow{p} \sigma_{\Delta u}^2.$$

We have,

$$\hat{\sigma}_{\Delta}^2 \xrightarrow{p} \sigma_{\Delta u}^2.$$

From Theorem 4.2.3 (i),

$$T^{-1}\gamma_{2T}^2 \xrightarrow{p} 0.$$

Therefore $\xi^* - \gamma_{2T} \xrightarrow{p} 0$, and the desired result follows again from Theorem 4.2.3 (i).

(ii) Under H_1, $e_t = \Sigma_j u_j$. Note that

$$\Delta Y_t^* = u_t - O_p(T^{-1/2}),$$

$$\hat{\sigma}_{\Delta}^2 = T^{-1}\Sigma_{t=2}^{T}u_t^2$$

$$= T^{-1}\Sigma(\Delta u_t)^2 + O_p(T^{-1}) - 2(T^{-1}\Sigma_{t=2}^{T}u_t\, O_p(T^{-1/2}),$$

$$T^{-1}\Sigma_{t=2}^{T}u_t \xrightarrow{p} 0,\ T^{-1}\Sigma_{t=2}^{T}u_t^2 \xrightarrow{p} \sigma_u^2 \text{ by LLN.}$$

We have,

$$\hat{\sigma}_{\Delta}^2 \xrightarrow{p} \sigma_u^2.$$

The desired result follows by multiplying the numerator and denominator of ξ^* by T^{-1}, and using Theorem 4.2.3 (ii),

$$\xi^* = \frac{T^{-1}\gamma_{2T}}{T^{-1}+T^{-2}\gamma_{2T}^2\hat{\sigma}_{\Delta}^{-2}} \Rightarrow \frac{\sigma_u^2}{\sigma\zeta_2}. \qquad \square$$

Proof of Theorem 4.2.5

(i) The result is immediate from Theorem 4.2.4 (i).

(ii) From Theorems 4.2.3 (ii) and 4.2.4 (ii), γ_{1T} diverges at rate T, and ξ^* converges to a nondegenerate distribution; therefore, their ratio C_1 diverges at rate T,

$$T^{-1}C_1 = T^{-1}\gamma_{1T}/\xi^* \Rightarrow (\sigma^2/\sigma_u^2)(\zeta_1\zeta_2). \qquad \square$$

Table 4.2.1 Monte Carlo Simulation on the Cauchy Test in Theorem 4.2.5

ϕ θ	T	ρ 0.0		0.8		0.9		0.95		0.99		1.0	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0 0.0	100	5.0	10.5	4.9	10.2	7.0	14.6	6.1	15.2	8.1	19.6	9.7	21.8
	300	5.5	9.9	5.1	10.8	5.8	11.1	5.5	11.8	9.7	31.8	14.1	38.1
	500	5.4	9.4	5.7	11.8	6.1	11.2	6.1	12.4	11.4	36.4	21.1	51.2
0.0 0.5	100	3.6	8.1	6.4	11.6	5.4	11.9	7.3	15.0	7.9	14.3	8.6	16.9
	300	5.5	10.5	5.0	9.6	5.4	10.0	5.1	10.8	7.4	16.5	9.5	18.7
	500	4.5	9.2	4.9	9.5	5.1	10.3	6.0	11.7	6.7	14.6	8.5	20.2
0.8 0.0	100			5.5	17.7	10.1	36.7	24.0	56.5	37.4	67.3	39.8	67.1
	300			4.9	10.3	6.6	23.1	19.6	49.2	59.4	81.9	65.3	83.8
	500			6.7	13.0	6.9	19.1	14.3	39.8	65.2	82.3	75.7	88.9
0.8 0.5	100			5.0	10.7	7.0	21.4	9.1	35.2	17.4	50.7	18.7	52.4
	300			5.2	9.8	5.2	13.5	8.5	32.4	42.0	69.0	52.8	74.5
	500			5.7	11.8	4.8	10.5	6.7	25.5	51.8	74.5	63.0	81.3
0.9 0.0	100					26.6	57.8	43.5	72.1	58.8	78.8	59.3	80.0
	300					18.8	48.9	46.1	71.8	79.7	89.4	81.3	90.2
	500					11.6	37.5	38.2	64.7	81.3	90.4	86.8	92.4
0.9 0.5	100					13.9	42.5	28.4	63.0	44.3	71.7	45.4	73.5
	300					10.3	32.2	33.2	61.9	71.1	84.8	76.9	87.5
	500					6.7	21.8	25.3	53.2	76.6	88.3	83.4	91.6
0.95 0.0	100							61.1	80.8	70.4	85.1	72.7	86.0
	300							68.9	83.5	86.4	92.3	90.0	95.4
	500							65.5	80.9	89.2	94.6	91.8	95.5
0.95 0.5	100							46.5	71.1	61.6	80.1	64.9	82.9
	300							58.4	79.0	82.6	92.1	87.7	93.4
	500							56.6	76.5	87.5	93.3	90.3	95.8

Note: The numbers in each cell are the percentage of rejections in 1000 replications. The Monte Carlo simulation results are based on 1000 replications of the following Data Generating Process:

$$(1-\phi B)(1-\rho B)\, Y_t = (1-\theta B)\epsilon_t, \qquad \epsilon_t \sim \text{NIID}(0, 1)$$

where the pseudo-random numbers $\{\epsilon_t\}$ are generated using the random number generator RAN of RATS. The simulation is executed using (mainframe) RATS Version 3.11.

Table 4.2.2 Values of the Quasi-Nuisance Parameters in Theorem 4.2.5

ϕ	σ^2/σ_u^2		$\sigma^2/\sigma_{\Delta u}^2$	
	θ=0	θ=0.5	θ=0	θ=0.5
0.00	1.00	0.20	0.50	0.07
0.50	3.00	1.00	3.00	0.50
0.80	9.00	5.00	22.50	4.17
0.90	19.00	13.57	95.00	18.27
0.95	39.00	32.50	390.00	76.47
0.99	199.00	191.35	9950.00	1982.07

Note: For ARMA model, the quasi-nuisance parameter is defined by Equation 4.2.9.

Table 4.3.1 Monte Carlo Simulation on the Cauchy Test

Using Orthogonal Polynomials

ϕ θ	T	ρ = 0.0 5%	0.0 10%	0.8 5%	0.8 10%	0.9 5%	0.9 10%	0.95 5%	0.95 10%	0.99 5%	0.99 10%	1.0 5%	1.0 10%
0.0 0.0	100	5.0	10.2	5.9	11.9	5.0	12.4	6.2	12.4	7.7	20.2	8.3	18.7
	300	5.3	11.2	4.5	9.3	4.2	10.1	4.4	10.4	8.9	29.5	14.4	38.7
	500	5.1	10.8	6.3	10.8	6.1	11.7	5.1	10.2	12.6	35.7	22.4	52.2
0.0 0.5	100	5.1	10.4	4.9	10.4	4.8	9.3	4.7	11.7	6.7	14.6	8.8	17.5
	300	5.4	11.7	4.6	9.6	4.2	9.7	5.7	10.4	7.3	13.9	7.8	16.9
	500	5.6	10.0	4.0	8.4	5.0	8.8	5.1	10.5	6.7	13.2	7.5	18.5
0.8 0.0	100			6.3	17.0	10.8	37.3	29.3	59.3	42.3	70.4	45.8	70.6
	300			4.7	10.6	5.5	21.7	22.7	48.8	64.1	82.0	70.2	84.6
	500			5.4	10.9	5.6	16.9	15.3	40.2	67.9	84.3	77.3	89.4
0.8 0.5	100			4.4	10.4	7.7	23.0	9.8	35.4	21.8	53.0	24.3	55.6
	300			5.2	11.1	5.8	12.6	7.9	32.5	45.9	71.1	53.9	73.9
	500			5.6	10.2	4.9	11.6	6.1	24.5	54.4	74.9	65.3	82.8
0.9 0.0	100					31.4	58.6	49.7	75.1	66.3	82.3	66.6	83.3
	300					18.8	46.8	48.1	70.8	81.1	90.4	82.4	91.9
	500					13.2	37.0	39.6	65.6	81.8	91.1	88.6	94.3
0.9 0.5	100					16.3	43.5	36.1	65.1	52.4	73.4	54.2	75.1
	300					9.4	32.3	32.8	60.2	71.8	85.7	79.2	88.5
	500					7.0	23.2	26.7	53.3	75.4	88.2	83.9	92.3
0.95 0.0	100							67.5	84.1	76.5	88.4	77.6	88.7
	300							69.3	83.8	88.6	93.7	91.1	95.7
	500							65.5	81.3	90.4	94.8	93.1	96.1
0.95 0.5	100							49.9	74.5	69.4	84.9	71.9	84.7
	300							60.5	79.0	86.0	92.9	88.6	94.5
	500							56.6	77.6	86.9	93.2	91.7	96.1

Note: See note for Table 4.2.1.

Table 4.3.2 Monte Carlo Simulation

on the Multiplier $(1+T^{-1}\gamma_{2T}^{2}\hat{\sigma}_{\Delta}^{-2})$

ϕ / θ		ρ 0.0	0.8	0.9	0.95	0.99	1.0
0.0	mean	1.0	1.1	1.4	1.8	2.2	2.2
0.0	var	0.0	0.0	0.4	1.4	3.0	2.8
0.0	mean	1.0	1.0	1.1	1.1	1.2	1.2
0.5	var	0.0	0.0	0.0	0.0	0.1	0.1
0.8	mean		3.1	5.8	8.9	12.4	11.9
0.0	var		9.2	41.1	108.5	257.7	193.9
0.8	mean		1.8	3.1	4.8	6.6	6.6
0.5	var		1.3	8.0	25.1	55.4	56.3
0.9	mean			10.4	15.8	22.7	23.9
0.0	var			176.2	409.9	855.8	950.8
0.9	mean			6.7	9.7	14.4	13.3
0.5	var			63.4	156.6	351.1	283.3
0.95	mean				24.5	36.3	36.0
0.0	var				997.4	2373.9	2528.8
0.95	mean				16.7	23.2	24.6
0.5	var				486.2	911.3	975.5

Note: "mean" and "var" are the sample mean and sample variance, respectively, across 1000 replications of samples of 100 observations. For simulation setup, see note for Table 4.2.1.

Table 4.3.3 Monte Carlo Simulation
on the Quality of Approximation of
the Asymptotic Distribution in Theorem 2.3.1

A. Sample Mean

ϕ		ρ					
θ		0.0	0.8	0.9	0.95	0.99	1.0
0.0	b_1	0.01	-0.02	0.00	0.26	- 0.97	2.87
0.0	b_2	0.01	0.17	-0.19	- 0.49	0.06	- 0.89
	b_3	-0.06	0.16	-0.10	0.17	0.69	- 0.65
0.0	b_1	-0.02	0.04	-0.19	0.17	- 0.20	- 0.84
0.5	b_2	-0.01	0.11	0.03	0.05	- 0.45	0.52
	b_3	0.01	0.03	-0.28	- 0.19	- 0.03	- 0.06
0.8	b_1		-0.29	-0.07	4.48	10.04	- 9.54
0.0	b_2		-0.30	-0.11	1.71	5.61	- 1.46
	b_3		0.94	1.40	- 2.50	2.59	3.33
0.8	b_1		-0.13	-0.62	- 0.41	-14.50	- 2.33
0.5	b_2		-0.16	0.67	- 0.22	0.20	- 2.64
	b_3		-0.47	-0.50	0.36	- 1.42	- 2.07
0.9	b_1			-0.71	- 1.04	3.30	53.06
0.0	b_2			2.43	- 1.62	- 2.43	- 1.84
	b_3			-0.23	- 4.53	- 2.36	0.83
0.9	b_1			1.16	0.22	- 5.23	1.41
0.5	b_2			-1.37	0.69	- 1.85	- 3.85
	b_3			-0.76	0.39	2.21	2.22
0.95	b_1				-22.73	5.42	55.88
0.0	b_2				- 3.54	5.55	0.65
	b_3				- 4.25	10.79	- 1.89
0.95	b_1				- 5.15	- 9.66	-22.63
0.5	b_2				- 2.77	4.83	9.43
	b_3				- 1.21	1.72	5.90

B. Sample Variance

ϕ		ρ					
θ		0.0	0.8	0.9	0.95	0.99	1.0
0.0	b_1	1.01	23.97	89.80	329.77	+	+
0.0	b_2	0.98	20.68	68.76	186.69	645.53	969.59
	b_3	1.04	19.19	55.23	114.75	229.89	257.14
0.0	b_1	0.24	5.88	23.66	81.10	830.41	+
0.5	b_2	0.27	5.51	18.48	52.48	168.20	249.28
	b_3	0.29	4.99	14.07	32.26	56.25	60.84
0.8	b_1		592.74	+	+	+	+
0.0	b_2		511.80	+	+	+	+
	b_3		435.57	+	+	+	+
0.8	b_1		129.37	598.78	+	+	+
0.5	b_2		113.78	454.52	+	+	+
	b_3		103.93	333.86	680.64	+	+
0.9	b_1			+	+	+	+
0.0	b_2			+	+	+	+
	b_3			+	+	+	+
0.9	b_1			+	+	+	+
0.5	b_2			+	+	+	+
	b_3			981.49	+	+	+
0.95	b_1				+	+	+
0.0	b_2				+	+	+
	b_3				+	+	+
0.95	b_1				+	+	+
0.5	b_2				+	+	+
	b_3				+	+	+

Note: (+) $\geq$ 1000.

Table 4.3.4 Monte Carlo Simulation on the t_4 Test in Equation 4.3.2

ϕ		ρ											
θ	T	0.0		0.8		0.9		0.95		0.99		1.0	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100	5.5	9.4	11.7	20.0	19.3	30.7	33.5	46.5	47.7	58.9	49.7	59.5
0.0	300	5.7	10.9	7.3	14.5	9.7	20.2	22.6	34.7	58.1	65.7	65.3	72.9
	500	4.4	8.6	7.6	14.5	6.4	13.1	15.4	25.0	63.1	71.4	72.3	78.8
0.0	100	4.0	7.8	8.5	15.8	10.9	19.4	23.4	32.5	30.6	42.7	33.2	44.1
0.5	300	4.5	9.3	6.8	12.9	7.0	13.1	11.3	19.9	34.6	44.2	42.0	51.8
	500	5.1	9.0	4.5	10.9	6.5	13.1	9.6	16.0	31.2	43.6	47.5	57.2
0.8	100			46.5	57.4	67.8	74.7	82.1	85.4	85.6	89.4	86.8	89.0
0.0	300			25.1	37.2	53.4	62.9	74.9	79.4	91.7	93.6	93.5	95.5
	500			17.0	27.3	43.6	53.5	67.7	75.1	94.7	96.3	95.4	97.0
0.8	100			24.9	35.6	53.8	63.4	65.9	73.2	75.9	81.1	77.8	83.4
0.5	300			12.6	23.9	32.4	44.7	62.8	69.7	85.2	88.8	88.7	90.6
	500			9.8	19.3	24.3	37.6	53.4	62.4	89.2	91.7	92.3	93.7
0.9	100					81.8	85.8	90.5	92.4	92.7	94.7	94.0	95.2
0.0	300					75.4	80.3	89.0	91.5	95.6	96.4	97.6	98.4
	500					68.3	73.7	83.5	87.5	96.7	97.1	97.5	98.2
0.9	100					71.6	76.3	84.8	87.5	87.7	90.2	89.2	91.5
0.5	300					59.7	67.5	81.9	86.1	94.3	95.3	95.9	96.8
	500					50.5	59.2	79.5	84.5	96.6	97.6	97.0	97.5
0.95	100							94.2	95.2	96.1	97.3	96.8	97.8
0.0	300							93.8	94.9	97.4	98.0	98.9	99.2
	500							93.8	95.1	98.4	98.4	98.7	99.0
0.95	100							88.4	90.8	94.4	95.8	93.6	95.4
0.5	300							92.2	93.9	97.6	98.2	98.1	98.3
	500							91.6	93.6	97.6	98.6	98.9	99.1

CHAPTER 5

A TRANSFORMATION OF THE DICKEY-FULLER TEST WITHOUT ESTIMATING THE LONG RUN VARIANCE

5.1. Introduction

Dickey and Fuller (1979, 1981) proposed the first class of statistical tests for unit root, but the form of the tests depends on the short run dynamics of the series. Phillips (1987), and Phillips and Perron (1988) nonparametrically corrected the Dickey-Fuller tests such that the resulting test statistics are invariant to the short run dynamics, yet the asymptotic distributions of the tests are the same as those of the original Dickey-Fuller tests.

But the nonparametric correction requires the estimate of the long run variance of the series, which often needs prudent choice of lag truncation or bandwidth parameters. Such an estimate often behaves very badly in finite samples. Different choices of the truncation lag could lead to contradictory outcome of the tests, making inference difficult.

In this chapter, we show that there is a simple transformation of the Phillips-Perron tests that is invariant to the long run variance parameter; therefore, its estimate is not necessary. The asymptotic distributions of the new test statistics are tabulated by simulation. The finite sample properties of the new tests are investigated by simulation, and comparison is made with the Phillips (1987) and Phillips and Perron (1988) tests. All proofs in Sections 5.2.2 and 5.2.3 can be derived from those of Phillips (1987) and Phillips and Perron (1988), or from Bierens (1992b), and therefore are omitted.

5.2 The Phillips and Phillips-Perron Transformation of the Dickey-Fuller Statistic

5.2.1 The Test

Given a time series Y_t, assuming that its deterministic component can be represented by f(t/T)b, then we have,

(5.2.1) $$(1-\rho L)(Y_t - f(t/T)b) = u_t,$$

where u_t is assumed to satisfy Assumption 2.2.1. For $|\rho| < 1$, Y_t is stationary around the deterministic trend f(t/T)b. Recall the regression equation in Section 2.4,

(5.2.2) $$X_t = f(t/T)b + \rho X_{t-1} + u_t.$$

Let $\hat{\rho}$ be the OLS estimate of ρ. The asymptotic distribution of $T(\hat{\rho}-1)$ under unit root is given by Theorem 2.4.1, which is partly reproduced as follows.

Theorem 5.2.1

(i) Under H_0, $T(\hat{\rho}-1) \Rightarrow [\int W^{*}dW + \lambda] / \int W^{*2}$;

(ii) under H_1, $(\hat{\rho}-1) \xrightarrow{p} -(1-\rho^{*}) < 0$;

where λ, σ^2, σ_u^2 and W^{*} are as defined in Theorem 2.4.1, $\rho^{*} = E(Y_t^{*}Y_{t-1}^{*})/E(Y_t^{*2})$,

$Y_t^{*} = Y_t - f(t/T)b$.

The asymptotic distribution under H_0 involves λ which depends on the unknown cumulated short run dynamics through the long run variance σ^2. Note that the first part of the distribution,

$$\int W^{*}dW / \int W^{*2}$$

is free of nuisance parameters and is tabulated in Fuller (1976). Phillips (1987) and Phillips and

Perron (1988) corrected the distributions for the second part, $\lambda/\int W^{*2}$, nonparametrically. Their test statistics are,

(5.2.3) $$Z = T(\hat{\rho}-1) - \tfrac{1}{2}(\hat{\sigma}^2-\hat{\sigma}_u^2)/T^{-2}\Sigma_{t=2}^{T}\hat{y}_{t-1}^{*2}$$

where $\hat{y}_t^*$ is an estimate of Y_t^*, usually by detrending using OLS, and $\hat{\sigma}^2$ and $\hat{\sigma}_u^2$ are consistent estimates of σ^2 and σ_u^2, respectively. Under H_1, the detrending by OLS is consistent. Consistent estimation of σ_u^2 is easy, but consistent estimation of σ^2 is more involved. One popular consistent estimator is the Newey and West (1987) estimator, also known as the Bartlett estimator. If σ_u^2 and σ^2 are estimated using the OLS residuals, as recommended by Phillips (1987), then the estimates are also consistent under the alternative. The asymptotic distribution of the Z test are summarized below.

Theorem 5.2.2

(i) Under H_0, $Z \Rightarrow \int W^* dW/\int W^{*2}$;

(ii) under H_1, $Z/T \xrightarrow{P} -(1-\rho^*) - \tfrac{1}{2}(\sigma_*^2-\sigma_{u*}^2)/\sigma_{y*}^2 < 0$;

where $\sigma_*^2 = \lim T^{-1}E(\Sigma_{t=1}^{T}u_{t*})^2$, $\sigma_{u*}^2 = \lim T^{-1}\Sigma_{t=1}^{T}E(u_{t*}^2)$, $u_{t*} = Y_t^* - \rho^* Y_{t-1}^*$,

$\sigma_{y*}^2 = \lim T^{-1}\Sigma_{t=1}^{T}Ey_{t*}^2$.

Therefore the Z test is consistent. But in finite samples, $\hat{\sigma}^2$ could behave badly relative to its asymptotic distribution. Table 5.2.1 presents simulation results for the Newey-West (1987) estimator for some simple ARMA models. In all cases except $\rho = 0$ and $\rho = 0.8$, $\theta = 0.5$, $\hat{\sigma}$ is a substantial underestimate of σ^2. Because for some of the ARMA models the autocovariance does not die out quickly enough, the lag truncation (m) in the Newey-West (1987) estimator cuts off too much autocovariance at high lags. If σ^2 is underestimated, then the correction term in the Phillips-Perron type of test is underestimated, so algebraically Z/T is overestimated, resulting in lower power.

5.2.2 Simulation Results

For comparison with the tests to be developed later in this chapter, we study the finite sample properties of the Phillips (1987) and Phillips and Perron (1988) tests with Monte Carlo simulation. The results are reported in Table 5.2.2. The simulation evidence agrees broadly with the evidence in the literature. cf. Phillips and Perron (1988) and Schwert (1989). Namely, the power of the Z test is fairly good, and the size of the test is close to the nominal size except for the case of $\theta = 0.5$. When $\theta = 0.5$, the moving average polynomial partially neutralizes the unit root, making the series appear stationary. Therefore, the null of unit root is rejected more often than the nominal size suggests.

5.3 An Alternative Transformation of the Dickey-Fuller Statistic

5.3.1 The Test

It turns out that there is an alternative transformation of the Dickey-Fuller statistic $T(\hat{\rho}\text{-}1)$ that does not need the estimate $\hat{\sigma}^2$. Since $\int WdW = (1/2)(W^2(1)\text{-}1)$, cf. Phillips (1987) or Bierens (1992b), rewrite the asymptotic distribution of Theorem 5.2.1 (i) as,

$$[(1/2)W^2(1)\text{-}\eta]/\int W^2 - (1/2)\sigma_u^2/[\sigma^2 \int W^2].$$

Then under H_0 the second part can be consistently estimated by $(1/2)\hat{\sigma}_u^2/T^{-2}\Sigma_{t=2}^T \hat{y}_{t-1}^{*2}$. Therefore define the new class of test statistics as,

$$Q = T(\hat{\rho}\text{-}1)+\tfrac{1}{2}\hat{\sigma}_u^2/T^{-2}\Sigma_{t=2}^T \hat{y}_{t-1}^{*2}. \qquad (5.3.1)$$

Its asymptotic distribution is given in Theorem 5.3.1.

Theorem 5.3.1

(i) Under H_0, $Q \Rightarrow \left[\tfrac{1}{2}W^2(1) - \eta\right]/\int W^{*2}(r)dr \equiv D$;

(ii) under H_1, $Q/T \xrightarrow{p} -(1\text{-}\rho^*) + (1/2)\sigma_{u*}^2/\sigma_{y*}^2$.

Therefore our new test is also consistent at rate T. But it does not need the estimate of the long run variance σ^2. Under H_1, it approaches negative infinity. Following the convention in the literature, for the cases of nointercept, only an intercept, and a linear trend, the asymptotic distributions under H_0 are denoted as D, D_μ and D_τ respectively; the Phillips and Phillips and Perron tests are denoted as Z, Z_μ and Z_τ respectively; and the new tests are denoted as Q, Q_μ and Q_τ respectively. The asymptotic distributions under H_0 are calculated by simulation, and reported in Tables 5.3.1. In particular, the (1%, 5%, 10%) critical values are, (0.00, 0.01, 0.05) for Q, (-4.39, -2.73, -1.95) for Q_μ and (-4.97, -3.06, -2.18) for Q_τ.

The fact that Z/T asymptotically is more negative than Q/T under H_1 does not necessarily mean that Z is more powerful, because the entire null distribution for Z lies to the left of that for Q. A more negative test statistic for Z has to be compared against a more negative critical value. Asymptotically, $(Q - Z)/T \xrightarrow{p} (1/2)(\sigma_*^2/\sigma_{y*}^2)$ under H_1. Thus the difference depends on the variance ratio which has nothing to do with the first order autocorrelation coefficient of Y_t. The larger is the variance ratio $(\sigma_*^2/\sigma_{y*}^2)$, the more powerful is the Z test. But the larger σ^2 is, the more difficult it is to estimate σ^2.

5.3.2 Simulation Results

For $\rho = 1$, the Q test is sensitive to starting value while the Phillips Z test is not. For example, separate simulation shows that, with $\theta = 0$ and a starting value of zero, the Q test rejects the null of unit root 7.2% (at 5%) and 12.0% (at 10%); with $\theta = 0$ and a random starting value from N(0, 20), i. e. a 20-step random walk, the Q test rejects the null 17.7% (at 5%) and 20.7% (at 10%). The Phillips Z test is robust to the starting value. It rejects 5% (at 5%) and 10.2% (at 10%) with a starting value zero, and rejects 4.5% (at 5%) and 9.3% (at 10 %) respectively. The reason may be that for the Q test, the critical values are confined to a very small interval; a small disturbance in the starting value may result in a large change in the test statistic relative to the critical values. For the Z test, such is not the case. Table 5.3.2

reports the finite sample performance of our new tests Q in comparison with the Z class of tests.

For our tests, the size is close to the nominal size except for $\theta = 0.5$, but the Q_μ and Q_τ tests do not have any power for $\rho \geq 0.9$. This can be explained by asymptotics under the alternative of stationarity. Table 5.3.3 presents some results on the causes of weak power. It is true that asymptotically both the Z test and the Q test approach negative infinity. But depending on the autocorrelation structure of the time series under study, it may take a huge number of observations for the asymptotics to work.

For example, for $(\rho, \theta) = (0.9, 0)$, $\text{plim}(Q_\mu/T) = -0.005$, $T \cdot \text{plim}(Q_\mu/T) = -1.5$ for $T = 300$ which roughly equals the 14% critical value. In order to reject the null at 5%, it is required that $T \geq 547$. For $(\rho, \theta) = (0.9, -0.5)$, $\text{plim}(Q_\mu/T) = -0.0016$, $T \cdot \text{plim}(Q_\mu/T) = -0.47$ for $T = 300$. In order to reject the null at 5%, it is required that $T \geq 1755$. The numbers are even more dramatic for $\rho = 0.95$ and Q_τ. See Tables 5.3.3b and 5.3.3c.

The existence of (low) power in the simulation reported in Tables 5.3.2b. and 5.3.2c is the result of (fortunate) underestimation of $\hat{\rho}$ and $\hat{\sigma}^2_{y*}$. In simulation with 1000 replications, for $(\rho, \theta) = (0.95, -0.5)$, the mean of $\hat{\rho}_\mu$ is 0.964, and $\hat{\rho}_\mu$ is smaller than its probability limit ρ_* (= 0.97) in 70% of the replications. For the same simulation, the mean of $\hat{\sigma}^2_{y\mu}$ is 19.632, and $\hat{\sigma}^2_{y\mu}$ is smaller than its probability limit of $\sigma^2_{y\mu}$ (= 22.56) in 71 % of the replications. But the net effect of the two sources of the underestimation is the underestimation of the test statistic, which results in the existence of (low) power for the Q_μ test.

In contrast to the Q_μ test, the result in Table 5.2.2b for the Phillips-Perron test Z_μ is much better. Therefore it seems that the additional correction term in Phillips and Perron test helps to differentiate the alternative from the null. The reason that the Phillips and Perron test performs better is that the additional correction term drives the plim(Z/T) further away from zero. The test statistic, which asymptotically is the product of plim(Z/T) and T, is much further away from zero than the Q test.

5.4 Transformations of Q with Improved Power

5.4.1 The Test

The problem with the Q test in Section 5.3 is that although Q diverges to negative infinity asymptotically, the probability limit of Q/T is too small in absolute value. If we can find a random variable A_T such that,

(5.4.1a) $$A_T - 1 = o_p(1) \text{ under } H_0,$$

(5.4.1b) $$A_T = o_p(1) \text{ and } A_T > 0 \text{ a. s. under } H_1,$$

then Q/A_T is also a consistent test. The new test Q/A_T has the same asymptotic distribution as Q under the null. However the new test Q/A_T is more powerful since it diverges to negative infinity at a faster rate. One choice of A_T can be derived from the following theorem.

Theorem 5.4.1 Let $\hat{\rho}$ be as defined in Section 5.2, then

(i) under H_0, $T \cdot \ln(|\hat{\rho}|)$ has the same asymptotic distribution as $T(\hat{\rho}-1)$;

(ii) under H_1, $T \cdot \ln(|\hat{\rho}|) \xrightarrow{P} -\infty$.

Proof: (i) $T \cdot \ln(|\hat{\rho}|)$ can be rewritten as,

(5.4.2) $$\begin{aligned} T \cdot \ln(|\hat{\rho}|) &= T \cdot [\ln(|\hat{\rho}|)/\ln(\hat{\rho})] \ln(1+\hat{\rho}-1) \\ &= T(\hat{\rho}-1) \cdot [\ln(|\hat{\rho}|)/\ln(\hat{\rho})][\ln(1+\hat{\rho}-1)/(\hat{\rho}-1)]. \end{aligned}$$

Since $\hat{\rho} \xrightarrow{P} 1$, $P(\hat{\rho}>0) \rightarrow 1$, that is, $P(\hat{\rho} = |\hat{\rho}|) \rightarrow 1$. So

(5.4.3) $$\ln(|\hat{\rho}|)/\ln(\hat{\rho}) \xrightarrow{P} 1.$$

Since $\hat{\rho}-1 = O_p(T^{-1})$,

(5.4.4) $$\ln(1+\hat{\rho}-1)/(\hat{\rho}-1) \xrightarrow{P} 1.$$

The desired result follows immediately from results (5.4.2)-(5.4.4).

(ii) By Theorem 5.2.1 (ii),

$$\hat{\rho} \xrightarrow{P} \rho^* < 1,$$

so

$$|\hat{\rho}| \xrightarrow{P} |\rho^*| < 1, \qquad \ln(|\hat{\rho}|) \xrightarrow{P} \ln(|\rho^*|) < 0.$$

The desired result follows trivially. □

The reason that the absolute value of $\hat{\rho}$ is taken is that $\hat{\rho}$ may be negative for some data generating processes, for example, when the series of interest Y_t is an AR(1) process with a negative autoregressive parameter. It is easy to see from the above theorem that,

(5.4.5a) $$|\hat{\rho}|^T \Rightarrow e^D \text{ under } H_0,$$

(5.4.5b) $$|\hat{\rho}|^T \xrightarrow{P} 0 \text{ under } H_1.$$

Let $A_T = |\hat{\rho}|^{T^\eta}$, $\eta \in (0,1)$. Then $\ln(A_T) = T^{\eta-1}[T \cdot \ln(|\hat{\rho}|)]$. It is trivial that,

(5.4.6a) $$\ln(A_T) = O_p(T^{\eta-1}) \text{ under } H_0,$$

(5.4.6b) $$\ln(A_T) = O_p(T^\eta) \text{ and } \ln(A_T) \xrightarrow{P} -\infty.$$

Therefore 5.4.1a and 5.4.1b hold. The choice of η controls the trade-off between the size distortion under the null and the improvement in the power under the alternative. The smaller η is, the faster does $\ln(A_T)$ in 5.4.6a converge to 0, the faster does A_T converge to 1 under H_0. The larger η is, the faster the $\ln(A_T)$ in 5.4.6b diverges to negative infinity, the faster A_T converges to zero.

5.4.2 Simulation Results

In the simulation, we set $\eta = 0.5$. For the case of the regression equation having an intercept, Table 5.4.1 summarizes the empirical distribution of $|\hat{\rho}|^{\sqrt{T}}$ under H_0 whose probability limit is 1. Tables 5.4.2 contains results on the size and power of the improved Q test. We see that the power of the test indeed improves dramatically, but at the cost of slightly larger size distortion. However the trade-off between improved power and slightly larger size distortion does seem to be a favorable one.

5.5 Summary and Conclusion

The class of tests for unit root proposed by Phillips and coauthors requires the estimation of the long run variance, the estimate of which may not behave well in finite samples. In this chapter, we propose another transformation of the Dickey-Fuller test statistic that does not require the estimation of the long run variance. The critical values of the new tests are calculated by simulation. Simulation shows that the size of the new test is close to the nominal size, but the power of the new test is substantially lower than that of the Phillips and Perron test due to the way the new test is constructed. In summary, the new transformation of the Dickey-Fuller test statistics is not very successful. Finally a transformation of the Q test is suggested, with the new test a significant improvement on the Q test.

Table 5.2.1. Monte Carlo Simulation on the Newey-West Estimator

θ	m	ρ			
		0.00	0.80	0.90	0.95
0.0	2	0.99	6.60	13.43	25.81
	8	0.97	13.26	32.37	68.47
	16	0.94	16.63	47.43	111.20
	∞	1.00	25.00	100.00	400.00
0.5	2	0.59	2.12	3.86	6.87
	8	0.36	3.64	8.50	17.32
	16	0.30	4.42	12.16	27.67
	∞	0.25	6.25	25.00	100.00
-0.5	2	1.88	14.57	29.82	57.58
	8	2.06	30.10	72.52	150.78
	16	2.06	38.35	106.04	244.74
	∞	2.25	56.25	225.00	900.00

Note: The Newey-West estimate of σ^2 is calculated as,

$$\hat{\sigma}^2 = \Sigma_{j=-(m-1)}^{m-1} k(j/m)\, \Sigma_{t=j+1}^{T} y_t y_{t-j}.$$

For consistency of $\hat{\sigma}^2$, it is required that, $m \to \infty$ and $m = o(\sqrt{T})$, cf. Andrews (1991), Theorem 1(a) and Bierens (1992b). The row for $m = \infty$ is the true value of σ^2, being included only for comparison. The simulation is based on 1000 replications of the data generating process,

$$(1-\rho L)\, y_t = (1-\theta L)u_t,\ u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 320,$$

with starting value zero. The first 20 values of y_t are discarded. The numbers in the table are sample averages across 1000 replications.

Table 5.2.2. Monte Carlo Simulation on Z Test

A. The Z Test

θ	m	ρ = 0.0 5%	0.0 10%	0.8 5%	0.8 10%	0.9 5%	0.9 10%	0.95 5%	0.95 10%	1.0 5%	1.0 10%
0.0	0	100.0	100.0	100.0	100.0	100.0	100.0	97.1	100.0	4.9	10.4
	2	100.0	100.0	100.0	100.0	100.0	100.0	96.5	99.9	5.1	10.3
	4	100.0	100.0	100.0	100.0	100.0	100.0	96.4	99.9	5.1	10.3
	8	100.0	100.0	100.0	100.0	100.0	100.0	95.6	99.8	5.4	11.1
	16	100.0	100.0	100.0	100.0	100.0	100.0	95.3	99.5	6.0	11.4
0.5	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	25.6	34.1
	4	100.0	100.0	100.0	100.0	100.0	100.0	96.5	99.9	22.8	31.0
	8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	24.2	31.6
	16	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	30.0	36.4
-0.5	2	100.0	100.0	100.0	100.0	100.0	100.0	91.4	98.4	2.8	7.4
	4	100.0	100.0	100.0	100.0	100.0	100.0	91.8	98.6	3.7	7.8
	8	100.0	100.0	100.0	100.0	100.0	100.0	90.6	98.5	3.9	7.9
	16	100.0	100.0	100.0	100.0	100.0	100.0	86.7	98.0	3.6	7.3

B. The Z_μ Test

θ	m	ρ = 0.0		0.8		0.9		0.95		1.0	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	0	100.0	100.0	100.0	100.0	100.0	100.0	75.1	88.5	6.2	10.9
	2	100.0	100.0	100.0	100.0	100.0	100.0	75.1	89.4	6.1	10.9
	4	100.0	100.0	100.0	100.0	99.9	100.0	75.7	89.1	6.4	10.9
	8	100.0	100.0	100.0	100.0	99.9	100.0	75.7	89.4	6.4	12.0
	16	100.0	100.0	100.0	100.0	99.9	100.0	77.4	90.4	7.0	12.5
0.5	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	38.7	49.1
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	37.7	75.0
	8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	40.8	48.9
	16	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	50.2	58.0
-0.5	2	100.0	100.0	100.0	100.0	99.3	100.0	59.0	78.6	2.6	6.5
	4	100.0	100.0	100.0	100.0	99.4	100.0	61.6	81.4	3.1	7.5
	8	100.0	100.0	100.0	100.0	98.9	100.0	59.7	80.8	3.0	7.8
	16	100.0	100.0	100.0	100.0	97.0	99.6	50.5	75.1	2.6	8.0

C. The Z_τ Test

θ	m	ρ = 0.0 5%	0.0 10%	0.8 5%	0.8 10%	0.9 5%	0.9 10%	0.95 5%	0.95 10%	1.0 5%	1.0 10%
0.0	0	100.0	100.0	100.0	100.0	97.9	99.7	41.5	63.6	5.7	10.8
	2	100.0	100.0	100.0	100.0	97.8	99.7	43.7	65.5	6.6	11.1
	4	100.0	100.0	100.0	100.0	97.8	99.7	46.5	65.8	6.7	12.0
	8	100.0	100.0	100.0	100.0	97.7	99.7	49.4	66.2	6.6	13.0
	16	100.0	100.0	100.0	100.0	98.6	99.7	49.6	68.2	7.3	13.3
0.5	2	100.0	100.0	100.0	100.0	100.0	100.0	99.8	99.9	61.5	70.5
	4	100.0	100.0	100.0	100.0	100.0	100.0	99.8	99.9	61.6	69.9
	8	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	67.9	74.6
	16	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	78.5	83.8
-0.5	2	100.0	100.0	100.0	100.0	83.5	95.7	24.4	41.5	2.1	4.5
	4	100.0	100.0	100.0	100.0	86.7	96.3	26.9	45.7	2.7	5.4
	8	100.0	100.0	100.0	100.0	82.0	94.3	24.7	44.7	2.5	5.3
	16	100.0	100.0	99.7	100.0	61.0	86.8	12.6	33.2	1.1	4.0

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho L)\, y_t = (1-\theta L)u_t,\ u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300.$$

The long run variance estimator is the Newey-West (1987) estimator using OLS residuals.

Table 5.3.1. Simulated Critical Values

A. The Distribution D

$\alpha = P(D \leq q)$

$\alpha*100$	q	$\alpha*100$	q
0 (min)	0.000	100 (max)	8.048
1	0.001	99	4.226
2	0.002	98	3.595
3	0.005	97	3.250
4	0.008	96	2.986
5	0.012	95	2.296
6	0.017	94	2.651
7	0.025	93	2.532
8	0.032	92	2.423
9	0.041	91	2.315
10	0.050	90	2.217
15	0.106	85	1.853
20	0.187	80	1.618
30	0.368	70	1.271
40	0.575	60	1.011
50	0.787	50	0.787

B. The Distribution D_μ

$\alpha = P(D_\mu \leq q)$

$\alpha*100$	q	$\alpha*100$	q
0 (min)	-8.908	100 (max)	7.999
1	-4.394	99	4.438
2	-3.670	98	3.655
3	-3.247	97	3.267
4	-2.938	96	2.919
5	-2.733	95	2.700
6	-2.552	94	2.529
7	-2.376	93	2.357
8	-2.204	92	2.209
9	-2.068	91	2.086
10	-1.950	90	1.989
15	-1.470	85	1.527
20	-1.122	80	1.185
30	-0.623	70	0.661
40	-0.255	60	0.270
50	0.001	50	0.001

C. The Distribution D_T

$$\alpha = P(D_T \leq q)$$

$\alpha*100$	q	$\alpha*100$	q
0 (min)	-12.078	100 (max)	10.247
1	- 4.973	99	4.813
2	- 4.091	98	4.086
3	- 3.668	97	3.646
4	- 3.370	96	3.313
5	- 3.055	95	2.987
6	- 2.816	94	2.755
7	- 2.627	93	2.546
8	- 2.454	92	2.376
9	- 2.310	91	2.215
10	- 2.179	90	2.086
15	- 1.553	85	1.560
20	- 1.161	80	1.162
30	- 0.574	70	0.631
40	- 0.204	60	0.251
50	0.010	50	0.010

The simulation is based on 10000 replications of samples of 1000 observations.

Table 5.3.2. Monte Carlo Simulation on the Q Test

A. The Q Test

θ	ρ									
	0.00		0.80		0.90		0.95		1.00	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100.0	100.0	99.8	99.8	93.9	94.2	73.5	75.5	7.7	12.6
0.5	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	25.9	27.1
-0.5	100.0	100.0	95.6	95.9	77.8	78.2	58.4	61.1	5.1	10.5

B. The Q_μ Test

θ	ρ									
	0.00		0.80		0.90		0.95		1.00	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100.0	100.0	97.2	99.2	25.1	41.3	5.0	9.7	5.6	11.8
0.5	100.0	100.0	100.0	100.0	100.0	100.0	96.7	98.2	20.5	29.5
-0.5	100.0	100.0	26.6	48.4	2.8	7.9	2.0	5.1	5.1	11.5

C. The Q_τ Test

θ	ρ									
	0.00		0.80		0.90		0.95		1.00	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100.0	100.0	96.7	98.9	25.0	44.0	5.8	12.9	5.8	10.0
0.5	100.0	100.0	100.0	100.0	100.0	100.0	97.6	98.6	49.8	61.2
-0.5	100.0	100.0	21.8	46.3	2.2	8.2	1.7	4.0	5.2	10.2

Table 5.3.3. The Asymptotics for the Power of the Z and Q Tests

A. - plim(Z/T) and - plim(Q/T)

θ	ρ							
	0.00		0.80		0.90		0.95	
	Z	Q	Z	Q	Z	Q	Z	Q
0.0	1.000	0.500	0.200	0.020	0.100	0.005	0.050	0.001
0.5	0.504	0.180	0.068	0.006	0.033	0.002	0.016	0.000
-0.5	1.176	0.980	1.080	0.180	1.005	0.069	0.756	0.002

Note: The numbers in the table are the probability limit of Q/T and Z/T. The minus sign is omitted for simplicity.

B. Minimum Number of Observations at which

the Z_μ and Q_μ test Are Significant at 5 %

θ	ρ							
	0.00		0.80		0.90		0.95	
	Z_μ	Q_μ	Z_μ	Q_μ	Z_μ	Q_μ	Z_μ	Q_μ
0.0	15	6	72	137	143	547	285	2187
0.5	13	3	14	16	15	40	19	122
-0.5	29	16	208	435	438	1755	899	7052

The numbers in the table are calculated as [5% critical value/Plim]+1, where plim is from the previous table.

C. Minimum Number of Observations at which

the Z_τ and Q_τ test Are Significant at 5 %

θ	ρ							
	0.00		0.80		0.90		0.95	
	Z_μ	Q_μ	Z_μ	Q_μ	Z_μ	Q_μ	Z_μ	Q_μ
0.0	22	7	108	153	215	612	430	2445
0.5	19	4	20	17	22	45	29	136
-0.5	43	17	314	486	661	1962	1355	7884

The numbers in the table are calculated as [5% critical value/Plim]+1, where plim is from the previous table.

Table 5.4.1 The Empirical Distribution of $|\hat{\rho}|^{\sqrt{T}}$

prob	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
value	0.30	0.54	0.63	0.68	0.73	0.77	0.82	0.85	0.90	0.95	1.06

Note: The simulation is based on 1000 replications of the following data generating process,

$$Y_t = Y_{t-1} + u_t,\ u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,$$

with $Y_0 = 0$, T = 300.

Table 5.4.2. Monte Carlo Simulation on the Improved Q Test

A. The Q Test

θ	ρ									
	0.00		0.80		0.90		0.95		1.00	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100.0	100.0	99.8	99.8	93.9	93.9	73.0	73.9	7.5	11.8
0.5	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	25.7	26.4
-0.5	100.0	100.0	95.6	95.6	77.3	77.8	58.0	59.7	4.9	9.9

B. The Q_μ Test

θ	ρ									
	0.00		0.80		0.90		0.95		1.00	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100.0	100.0	99.8	99.8	86.4	89.3	34.2	42.9	13.4	18.7
0.5	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	48.9	52.8
-0.5	100.0	100.0	92.3	93.4	32.1	43.6	10.6	15.0	10.6	15.9

C. The Q_τ Test

θ	ρ									
	0.00		0.80		0.90		0.95		1.00	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0.0	100.0	100.0	99.9	99.9	88.2	90.9	42.2	50.2	16.9	23.8
0.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	80.7	82.7
-0.5	100.0	100.0	92.4	94.4	34.9	46.6	12.2	17.7	13.0	19.3

The simulation is based on 1000 replications of the following DGP,

$$(1-\rho L)\, y_t = (1-\theta L)u_t,\ u_t \sim \text{NIID}(0, 1),\ t = 1, 2, \cdots, T,\ T = 300.$$

η in Equation 5.4.6 is set at $\eta = 0.5$.

CHAPTER 6

TREND, STATIONARITY AND UNIT ROOTS IN MACROECONOMIC TIME SERIES

6.1 Introduction

Nelson and Plosser (1982) applied the Dickey-Fuller τ test to 14 annual U.S. macroeconomic time series; the null model of unit root was rejected only for the unemployment series. Loosely speaking, the DF τ test "finds" unit root in all macroeconomic time series except the unemployment rate. That finding has been interpreted as evidence supporting the real business cycle theory which argues that the permanent technological shocks to the economy are the driving force of the business cycles. However, statistically it can be true that the null model of unit root was not rejected just because the test does not have enough power to distinguish the null model from the alternative model. The Nelson-Plosser data set serves as an ideal laboratory for contrasting different unit root tests. Recently Schotman and van Dijk (1991) extended the Nelson-Plosser data set up to 1988 in studying the test for unit root in a Bayesian framework.

In this chapter, we apply the new tests developed in the previous chapters to the extended Nelson-Plosser data set and the Post-War quarterly real GDP series.

6.2 Description of the Data

The original Nelson-Plosser data set contains 14 historical annual macroeconomic time series. All series end at 1970 with starting dates from 1860 to 1909. The 14 time series are, (1) real GNP, (2) nominal GNP, (3) real per capita GNP, (4) industrial production, (5) employment, (6) unemploy-ment rate, (7) GNP deflator, (8) consumer price index, (9) nominal wages, (10) real wages, (11) money stock, (12) velocity, (13) bond yield, and (14) common stock prices. The extended data set ends at 1988. We also analyze the Post-War quarterly real GDP (in 1987 dollars) which is from 1947:1 to 1992:3. All of the series except the bond yield are in logarithms.

All of the series are plotted in Figure 6.1. We see that all the series (except unemployment rate) display strong trends (either linear or quadratic), which makes unit root with drift and trend stationarity the interesting competing hypotheses. For this reason, a linear trend is included as part of the null hypothesis for all the series except unemployment. Only an intercept is included for the unemployment series. To isolate the effect of the new tests from that of the new observations, both the original Nelson-Plosser data set and the extended data set will be used.

6.3 Testing for Linear Trend Stationarity and Unit Root

The results of the traditional unit root tests on the extended Nelson-Plosser series are summarized in Table 6.3.1. As in many other studies, the results can be summarized as follows: (i) there is strong evidence against unit roots in Industrial Production (IP) and Unemployment Rate (UNEMP); (ii) there is some evidence against unit roots in Real GNP (RGNP), Per Capita Real GNP (PCRGNP), Employment (EMP); (iii) there is no evidence against unit roots in the rest of the series. The three groups of series above will be called the Groups (i), (ii) and (iii) series respectively in the following discussion.

The results using the CUSUM test, the Fluctuation test and the modified DF test are reported in Table 6.3.2. The results from the CUSUM and Fluctuation tests can be summarized as follows: (i) there is strong evidence against linear trend stationarity for the Group (iii) series above; (ii) there is no evidence against linear trend stationarity for the Groups (i) and (ii) series above. The traditional tests for unit roots and our CUSUM tests for stationarity paint similar pictures in terms of evidence for and against stationarity (or unit root). Note that the evidence against linear trend stationarity can be interpreted either as evidence for a breaking trend or as evidence for unit root.

However, the modified DF test does not reject the null of unit root for any series, reflecting the weak power of the test.

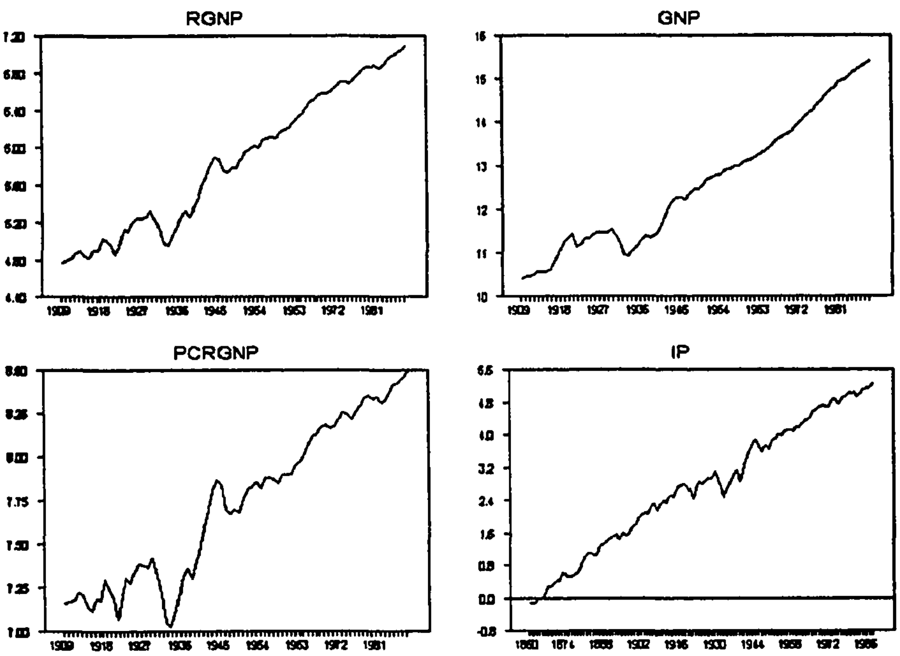

Figure 6.1 The Extended Nelson-Plosser Economic Time Series

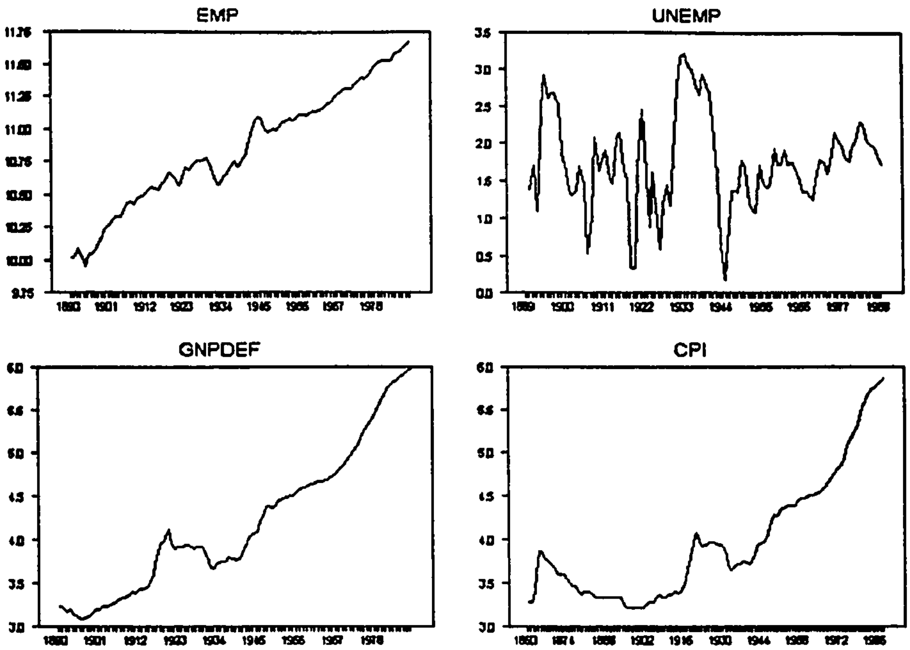

Figure 6.1 Continued

According to our Cauchy tests in Table 6.3.3, the null model of linear trend stationarity is strongly rejected for nominal GNP (GNP), Consumer Prices (CPI) and Velocity (VEL), and marginally rejected for GNP Deflator (GNPDEF) and Stock Prices (SP500). After applying the MA filter (1-0.5B), the stationarity of GNPDEF is no longer rejected. But the evidence against stationarity remains strong for GNP, CPI and VEL, and that for SP500 still remains marginal.

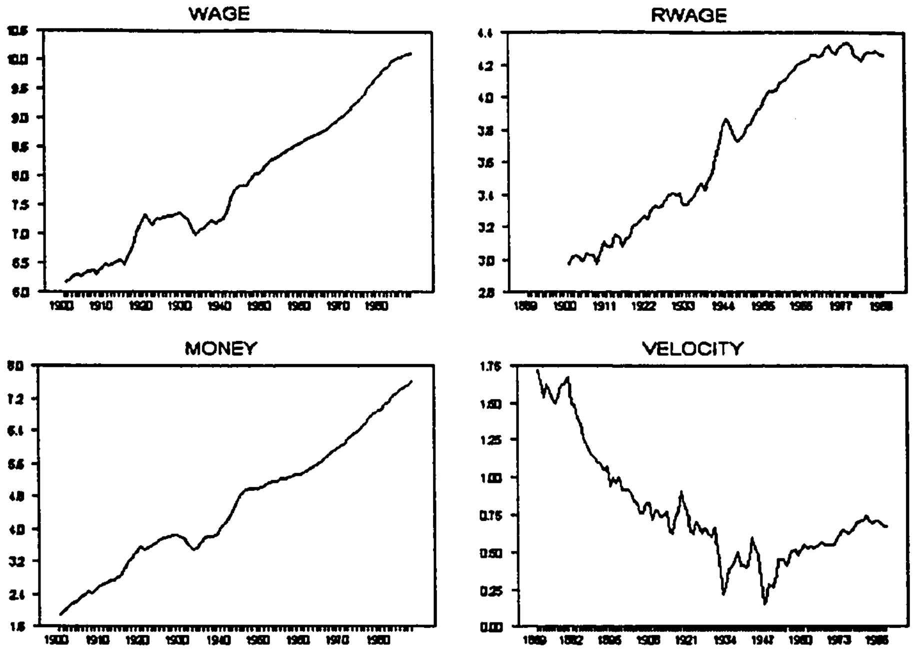

Figure 6.1 Continued

Summarizing the results of the tests so far, there is strong evidence against linear trend stationarity (there is no evidence against the unit root) for Group (iii) series, but there is no evidence against linear trend (there is some or strong evidence against the unit root) for the rest of the series.

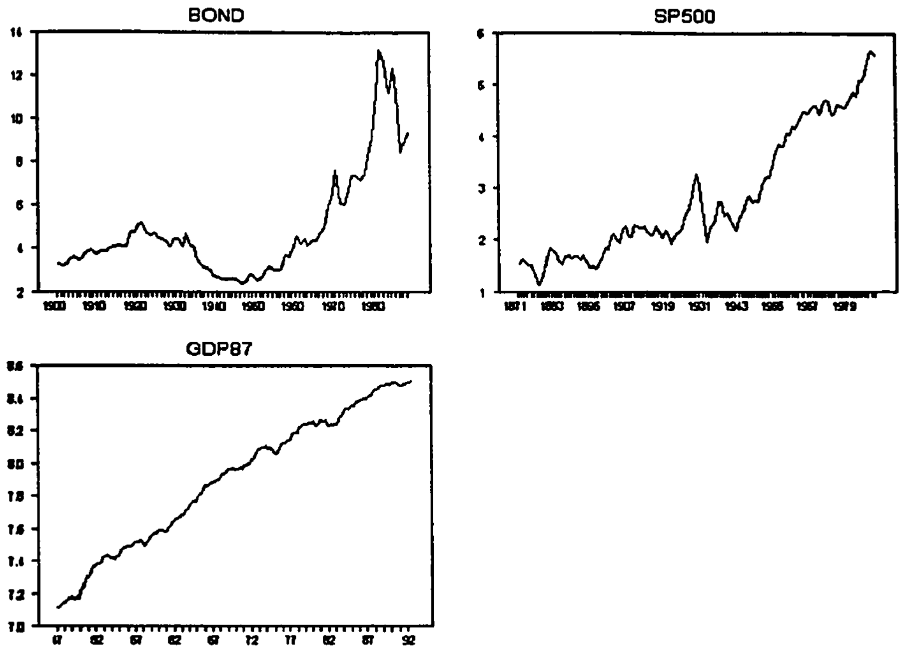

Figure 6.1 Continued

6.4. Breaking Trend Analysis

The evidence against linear trend stationarity may be due to either a break in the linear trend, a higher order trend, or a unit root. If the rejection is due to a break in the trend, then after specifying the correct break in the trend, there should be no evidence against stationarity left in the series. If the rejection is due to a unit root, then evidence against stationarity should remain no matter what kind of break is specified. The reason is that a unit root process is equivalent to a stationary process with a break in each time period. Using the methodology in Chapter 3, we test for stationarity with breaks in linear trend. The break fraction is assumed to be 10%, 20%,···, 90% of the sample, since these are the break fractions for which critical values are available.

The results for the break trend analysis are presented in Table 6.4.1. Except for the extended CPI series, for the series for which the linear trend stationarity is rejected in the last section, there exists at least one breaking trend such that the stationarity is not rejected at the 5% significance level. The location of the break for which the CUSUM statistic is the smallest is called the "optimal" break point. Because of the endogeneity of the search for the break point, the critical values need to be adjusted. The correct critical values are calculated in Chapter 3 and are given in the last column of Tables 3.2.5 and 3.2.6. The minimum CUSUM statistic (mCU) in the last column of Table 6.4.1 should be compared with this set of critical values. Except for the extended CPI series, there is no evidence against breaking trend. Note that the long run variance estimate used in the CUSUM test is from the Adaptive QS estimator which improves the power of the test; therefore, the evidence here for breaking trend is indeed very strong.

For comparison with previous work on breaking trend, we list the "optimal" break fraction in Table 6.4.2. The "optimal" break fraction is in broad agreement with that in Perron (1989) that the Great Depression is the major cause of break.

6.5. Summary and Conclusion

Most of the traditional tests for unit root do not reject the unit root hypothesis for most of the Nelson-Plosser series. Various tests for stationarity that are developed in earlier chapters reveal very strong evidence against linear trend stationarity for most of the series. However, when a break in the linear trend is specified, the evidence for unit root, or against stationarity, almost no longer exists. In this sense, our new CUSUM test is in agreement with the Perron approach. Note that the empirical results in this chapter show that the modified Dickey-Fuller test introduced in Chapter 5 has very little power.

The evidence for breaking trend in this chapter, as well as that in the earlier literature, warns against the common practice of testing for unit root against a linear trend and, if unit root is not rejected, continuing to test for cointegration. Care should be taken when deciding how best to model a macroeconomic time series. It might be that the distribution theory with a breaking trend provides a better approximation to the finite sample distribution than that under the unit root.

Table 6.3.1 Application of Traditional Tests for Unit Root

to the Extended Nelson-Plosser Data Set

A. The Dickey-Fuller Tests

Variable	τ_τ k=2	k=4	k=8	ρ_τ k=2	k=4	k=8
RGNP	3.35*	2.88	2.74	25.54**	22.27**	56.20***
GNP	1.74	1.25	2.20	7.40	4.40	19.42**
PCRGNP	3.44**	3.02	2.91	26.77**	25.16**	96.36***
IP	3.33*	3.22*	2.85	23.80**	26.63**	31.33***
EMP	2.87	3.38*	3.65**	19.41*	24.95**	157.93***
UNEMP	3.41**	3.45***	3.39***	26.56***	39.48***	190.09***
GNPDEF	1.63	1.45	1.61	6.65	6.16	14.30
CPI	0.59	1.77	1.49	1.54	3.78	5.38
WAGE	2.12	1.91	2.26	9.69	9.86	33.32***
RWAGE	1.45	1.09	0.86	7.32	5.78	6.30
MONEY	2.51	2.64	3.00	16.28	22.91**	763.82***
VEL	1.32	1.19	1.36	3.99	2.74	2.92
BOND	1.37	1.81	0.31	4.28	8.88	1.01
SP500	1.92	1.63	1.16	8.54	6.89	4.50
QRGDP	2.36	1.98	2.21	11.98	9.26	10.40

B. The Phillips-Perron Tests

Variable	$Z(t_\tau)$ k=4	k=7	k=10	$Z(\alpha_\tau)$ k=4	k=7	k=10
RGNP	2.76	2.58	2.38	13.90	11.97	9.93
GNP	1.54	1.54	1.56	5.16	5.17	5.30
PCRGNP	2.85	2.67	2.48	14.53	12.59	10.49
IP	3.34*	3.16*	3.08	21.02*	18.67*	17.55
EMP	2.76	2.61	2.50	14.87	13.25	12.09
UNEMP	3.84***	3.74***	3.61***	26.48***	25.05***	23.09***
GNPDEF	1.39	1.48	1.53	4.28	4.80	5.10
CPI	0.30	0.37	0.39	0.73	0.93	0.97
WAGE	1.83	1.88	1.91	6.78	7.15	7.37
RWAGE	1.33	1.25	1.26	5.52	5.04	5.11
MONEY	1.94	2.07	2.08	8.66	9.73	9.78
VEL	1.56	1.44	1.39	4.17	3.43	3.15
BOND	1.57	1.46	1.39	5.18	4.56	4.12
SP500	1.90	1.68	1.70	7.77	6.20	6.33
QRGDP	1.80	1.80	1.73	6.97	7.00	6.54

Note: (1) The data are the annual macroeconomic time series used in Nelson and Plosser (1982), and extended up to 1988 by Schotman and van Dijk (1991). The last series is the post-War quarterly real GNP from Citibase.

(2) All the series used in the unit root test are in logarithms except the bond yield (BOND).

(3) τ and α are the Augmented Dickey-Fuller t and T($\hat{\rho}$-1) statistics with a linear trend (no trend for UNEMP). Z(t) and Z(α) are the corresponding Phillips-Perron corrected t and T($\hat{\rho}$-1) statistics. The negative signs on the values of the test statistics are dropped for the ease of presentation.

(4) The asymptotic critical values at 1%, 5% and 10% significance levels for the test statistics are, respectively,

for τ_μ and Z(t_μ), -3.43, -2.86, -2.57, for τ_τ and Z(t_τ), -3.96, -3.41, -3.12,

for α_μ and Z(α_μ), -20.7, -14.1, -11.3 for α_τ and Z(α_τ), -29.5, -21.8, -18.3.

(5) The corrections for Z(t) and Z(α) are calculated using the Newey-West estimator with window width k.

(6) (*, **, ***) indicate being significant at (10%, 5%, 1%) respectively.

Table 6.3.2 Application of the CUSUM and the Fluctuation Test for Stationarity and the Modified Dickey-Fuller Test (Q) for Unit Root

A. Nelson-Plosser Data Set

Variable	CUSUM (τ=0.15) CU	RS	Fluctuation (τ=0.15) FL_1	Q
RGNP	0.73	1.33	3.21	-0.75
GNP	2.81***	4.46***	9.33***	0.19
PCRGNP	0.55	1.04	2.47	-0.99
IP	0.54	0.95	1.78	-1.49
EMP	0.67	1.30	2.90	-0.66
UNEMP	0.50	0.72	1.12	-2.11
GNPDEF	2.32***	4.43***	8.56***	-1.08
CPI	4.96***	9.26***	19.81***	0.98
WAGE	2.55***	4.35***	7.72***	-0.01
RWAGE	0.71	1.34	3.29	-1.60
MONEY	2.09***	3.49***	7.06***	-0.04
VEL	4.72***	9.14***	15.79***	0.14
BOND	3.40***	6.68***	10.24***	7.86
SP500	2.48***	3.80***	7.22***	-0.03

B. Extended Nelson-Plosser Data Set

Variable	CUSUM (τ=0.15) CU	RS	Fluctuation (τ=0.15) FL_1	Q
RGNP	0.44	0.75	2.13	-1.30
GNP	4.53***	7.54***	14.28***	0.34
PCRGNP	0.35	0.66	1.97	-1.60
IP	0.48	0.92	1.94	-1.54
EMP	0.51	0.92	2.70	-0.52
UNEMP	0.39	0.68	1.42	-2.69
GNPDEF	5.38***	10.10***	17.60***	0.30
CPI	6.88***	13.23***	16.09***	2.17
WAGE	4.07***	7.00***	12.83***	0.00
RWAGE	3.52***	6.65***	9.13***	2.49
MONEY	2.95***	5.03***	7.48***	1.16
VEL	8.01***	15.62***	31.04***	-0.68
BOND	5.33***	9.81***	18.33***	0.19
SP500	4.27***	7.12***	15.42***	0.26
QRGDP	8.48***	15.72***	34.35***	0.51

Note: The long run variance is estimated using the Adaptive Andrews QS estimator. (*, **, ***) indicate being significant at (10%, 5%, 1%) respectively.

Table 6.3.3 Application of The Cauchy Tests With Linear Trend

to the Extended Nelson-Plosser Data Set

	Without MA filter		With MA filter (1-0.5B)	
Variable	C_1	C_2	C_1	C_2
RGNP	- 1.46	- 1.32	- 0.92	- 0.67
GNP	- 6.28	181.98***	- 5.45	-13.33**
PCRGNP	- 1.64	- 1.26	- 1.21	- 0.69
IP	- 1.41	- 1.70	- 0.89	- 1.36
EMP	- 1.91	- 0.47	0.76	- 0.14
UNEMP	0.06	- 0.11	0.01	- 0.24
GNPDEF	6.93*	7.43*	4.37	3.74
CPI	53.68**	20.66**	17.22**	79.27***
WAGE	- 5.59	5.59	- 5.15	5.65
RWAGE	1.37	1.62	0.98	0.89
MONEY	1.90	2.13	2.54	1.30
VEL	8.65*	46.70**	6.70	406.85***
BOND	4.78	4.86	3.26	2.89
SP500	4.37	7.67*	2.65	6.95*
QRGDP	25.27**	21.82**	61.52**	17.63**

Note: C_1 and C_2 are Cauchy tests; C_1 is from Theorem 5.2.5 based on residuals; C_2 is based on orthonormal polynomial regression. The critical values for a two-sided test at (1%, 5%, 10%) significance levels are (63.66, 12.71, 6.31), respectively. (*, **, ***) indicate being significant at (10%, 5%, 1%) respectively.

Table 6.4.1 Results on Breaking Trend and Unit Root

A. Nelson-Plosser Data Set

Variable	η (break fraction)									mCU
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
RGNP	0.46	0.39	0.23	0.28	0.31	0.30	0.41	0.50	0.48	0.23
GNP	2.81*	2.32*	0.31	0.27	0.42	0.36	1.24*	2.09*	2.40*	0.27
PCRGNP	0.35	0.30	0.27	0.29	0.29	0.32	0.49	0.55	0.53	0.27
IP	0.49	0.43	0.40	0.38	0.42	0.32	0.47	0.53	0.55	0.32
EMP	0.34	0.30	0.28	0.27	0.32	0.64	0.75*	0.64	0.58	0.27
UNEMP	0.42	0.43	0.41	0.52	0.53	0.67	0.37	0.43	0.48	0.37
GNPDEF	3.06*	2.86*	2.71*	2.17*	0.34	1.44*	1.81*	1.54*	1.96*	0.34
CPI	1.72*	0.47	0.33	0.36	0.36	2.73*	3.57*	4.12*	4.58*	0.33
WAGE	2.83*	2.62*	2.46*	0.25	0.39	1.91*	1.00*	1.84*	2.29*	0.25
RWAGE	0.36	0.31	0.27	0.29	0.29	0.35	0.43	0.51	0.56	0.27
MONEY	2.00*	1.75*	1.62*	1.50*	0.35	1.99*	2.83*	2.36*	2.33*	0.35
VEL	1.19*	0.46	0.40	0.36	0.31	0.42	0.43	0.55	0.51	0.31
BOND	2.97*	3.04*	3.48*	2.50*	0.37	0.37	2.03*	2.55*	2.69*	0.37
SP500	2.12*	1.01	0.90*	0.44	0.38	0.31	0.23	0.29	0.46	0.23
	5% critical values									
no trend	1.27	1.19	1.11	1.04	1.00	1.03	1.10	1.19	1.27	0.76
trend	0.83	0.78	0.72	0.67	0.64	0.67	0.72	0.78	0.82	0.53

B. Extended Nelson-Plosser Data Set

Variable	η (break fraction) 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	mCU
RGNP	0.31	0.26	0.34	0.31	0.29	0.34	0.33	0.40	0.45	0.26
GNP	4.11*	0.49	0.24	0.40	0.85*	1.76*	2.24*	2.45*	2.98*	0.24
PCRGNP	0.28	0.29	0.35	0.29	0.32	0.37	0.36	0.33	0.35	0.28
IP	0.41	0.39	0.38	0.36	0.36	0.44	0.53	0.55	0.55	0.36
EMP	0.37	0.35	0.31	0.28	0.49	0.42	0.38	0.39	0.40	0.28
UNEMP	0.48	0.36	0.41	0.62	0.51	0.37	0.42	0.48	0.47	0.36
GNPDEF	5.30*	5.10*	3.45*	0.41	1.45*	1.87*	1.64*	1.75*	0.74	0.41
CPI	5.40*	4.37*	3.47*	3.31*	2.17*	3.45*	3.34*	4.24*	5.89*	2.17*
WAGE	3.85*	4.06*	0.36	0.26	1.14*	1.50*	2.03	2.32*	2.28*	0.26
RWAGE	3.45*	3.54*	3.24*	2.40*	0.31	0.39	0.47	0.45	0.49	0.31
MONEY	3.73*	2.85*	2.74*	0.30	2.11*	2.46*	2.21*	2.10*	2.36*	0.30
VEL	2.63*	1.07*	0.65	0.46	0.33	0.39	0.51	0.51	1.43*	0.33
BOND	4.84*	3.76*	0.88*	0.37	0.24	0.25	0.24	0.32	0.80	0.24
SP500	2.75*	1.09*	0.52	0.39	0.28	0.28	0.27	0.46	1.24*	0.27
QRGDP	7.70*	7.26*	5.43*	0.49	1.77*	1.53*	1.65*	3.61*	7.02*	0.49
	5% critical values									
no trend	1.27	1.19	1.11	1.04	1.00	1.03	1.10	1.19	1.27	0.76
trend	0.83	0.78	0.72	0.67	0.64	0.67	0.72	0.78	0.82	0.53

Note: A star (*) in this table indictes significance at 5%. However some of the tests may be significant at a lower significance level.

Table 6.4.2 Optimal Break Fraction

Variable	Nelson-Plosser Data Set CUSUM	Nelson-Plosser Data Set Perron	Extended Data Set CUSUM
RGNP	0.3	0.33	0.2
GNP	0.4	0.33	0.3
PCRGNP	0.3	0.33	0.1
IP	0.6	0.63	0.4
EMP	0.4	0.49	0.4
GNPDEF	0.5	0.49	0.4
CPI	0.3	0.63	0.5
WAGE	0.4	0.41	0.4
RWAGE	0.3	0.41	0.5
MONEY	0.5	0.49	0.4
VEL	0.5	0.59	0.5
BOND	0.5	0.41	0.5
SP500	0.7	0.59	0.7
QRGDP	0.4	NA	NA

Note: The columns "CUSUM" and "Perron" are the optimal break fractions determined by our mCU test and Perron (1989) respectively.

CHAPTER 7

SUMMARY AND CONCLUSION

The purpose of this dissertation has been to develop several new tests for stationarity and a new test for unit root. The tests developed in the previous chapters all have some superior properties either in finite samples or in large samples.

A unit root process is equivalent to a stochastic trend. Therefore, tests for parameter stability should have good power against unit root processes. In Chapter 3, for the case of trending regressors, we derived the asymptotic distributions of the CUSUM test using OLS residuals and of the parameter fluctuation test over subsamples. Simulation shows that these two tests have good power against the unit root alternative. We also further developed the CUSUM test when a break in the trend is allowed under the null hypothesis, and discussed the issue of the "optimal" selection of the break point.

Beginning with Phillips (1987), in the context of test for unit root and stationarity, the short run dynamics have been corrected by nonparametric methods. The nonparametric estimation of the nuisance parameter (i. e. the long run variance) can crucially affect the finite sample properties of the tests. The asymptotic distributions of the tests for stationarity are often proportional to the square root of the long run variance. If, as suggested by Phillips (1987) and Phillips and Perron (1988), the detrended series is used in the estimation of the long run variance, the estimate diverges to infinity under unit root. Since the estimate of the long run variance appears in the denominator of the tests statistics, the power of the tests is affected. In Chapter 3, we propose an adaptive estimator of the long run variance that approaches a finite constant under both the null and the alternative. Therefore, the tests using our adaptive estimator should have superior power. Simulation shows that indeed the power of the tests improves when the adaptive estimator is used.

The finite sample properties of a consistent estimator of the long run variance can be very poor. Therefore, when the estimate of the long run variance is used in constructing a test, the outcome of the test may be dependent on the way the long run variance is estimated. In Chapter 4, we generalize the Cauchy test of Bierens (1991a) to the case of a general deterministic trend, possibly with breaks; several Cauchy tests were proposed. The Cauchy tests avoid having to estimate the nuisance parameter. Although the test is constructed differently depending on the nature of the deterministic trend, the asymptotic distribution under the null of stationarity is always standard Cauchy. Simulation shows that the tests have weak power against pure random walk alternatives, but good power against general unit root alternatives.

In Chapter 5, we propose another transformation of the Dickey-Fuller test for unit root. Our proposed transformation is different from that of Phillips (1987) and Phillips and Perron (1988) in that ours does not need the estimate of the long run variance. Given the problems with most of the long run variance estimators in finite samples, our test can potentially behave better in finite samples. We also discussed ways to enhance the power of the test.

In Chapter 6, all the new tests developed in the previous chapters were applied to the extended Nelson-Plosser data set. While the CUSUM test and the Fluctuation test reject linear trend stationarity for most of the time series, there is no evidence against stationarity around a linear trend with an one-time break (except for the Consumer Price Index series). This lends more support to the argument of Perron (1989) for a breaking trend characterization of most of the historical macroeconomic time series.

The poor behavior of the existing long run variance estimators in finite samples originates from the strong short run dependence (autocorrelation) in the time series. It seems more fruitful to model both the long run and the short run dynamics jointly. Since all the distribution theories involving unit root processes are asymptotic, future research should pay more attention to the quality of approximation to the distributions in finite samples.

As an extension of the unit root tests, the concept of cointegration, introduced by Engle and Granger (1987), has been an active area of econometric research. While the modeling of multiple time series using cointegration method is well developed, especially since the publication of a series of papers by Johansen and coauthors (1988, 1990, 1991), introducing the method of cointegration to panel data analysis is a challenging and fruitful area for future research.

REFERENCES

Andrews, Donald W. K. 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

______ and J. Christopher Monahan 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 50, 953-966.

Bierens, Herman J. 1991a, Testing stationarity against the unit root hypothesis, manuscript, Free University and Southern Methodist University (November).

______ 1991b, Higher order sample autocorrelation and the unit root hypothesis, forthcoming in Journal of Econometrics.

______ 1992a, The unit root hypothesis versus nonlinear trend stationarity, manuscript, Southern Methodist University (February).

______ 1992b, Topics in Advanced Econometrics. Combridge University Press (forthcoming).

______ and Shengyi Guo 1992, Testing for stationarity and trend stationarity against the unit root hypothesis, Econometric Reviews 12, 1-32.

Billingsley, Patrick 1964, Convergence of Probability Measures. John Wiley.

Box, George E. P. and Gwilym M. Jenkins 1976, Time Series Analysis: Forecasting and Control. Holden-Day.

Brown, R. L., Durbin, J. J. and J. M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, Journal of the Royal Statistical Society B, 37, 149-192.

Campbell, John Y. 1987, Does saving anticipate declining labor income? an alternative test of the Permanent Income Hypothesis, Econometrica 55, 1249-1273.

______ and Pierre Perron 1991, Pitfalls and opportunities: what macroeconomists should know about unit roots. In NBER Macroeconomics Annual, ed. Oliver J. Blanchard and Stanley Fischer, 141-201. MIT Press.

______ and Robert J. Shiller 1987, Cointegration and tests of present value models, Journal of Political Economy 95, 1062-1088.

Chan, N. H. and C. Z. Wei 1987, Asymptotic inference for nearly nonstationary AR(1) processes, The Annals of Statistics 15, 1050-1063.

Chu, Chia-Shang J. and Halbert White 1992, A Direct Test for Changing Trend, Journal of

Business and Economic Statistics 10, 289-299.

Dhrymes, Phoebus J. 1989, Topics in Advanced Econometrics: Probability Foundations. Springer-Verlag.

Dickey, David A. and Wayne A. Fuller 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427-431.

______ and Wayne A. Fuller 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057-1072.

Diebold, Francis X. and Marc Nerlove 1990, Unit roots in economic time series: a selective survey, 3-69, in Thomas B. Fomby and George F. Rhodes, Jr. ed. Advances in Econometrics, Volume 8. JAI Press.

Dolado, Juan J., Jenkins, Tim and Simon Sosvilla-Rivero 1990, Cointegration and Unit Roots, Journal of Economic Surveys 4, 249-273.

Elliott, Graham and James H. Stock 1992, Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown, NBER Technical Working Paper 122.

Engle, Robert F. and Clive W. J. Granger 1987, Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica 55, 251-276.

Hall, R. E. 1978, Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, Journal of Political Economy 86, 971-987.

Hakkio, Craig S. and Mark Rush 1989, Market Efficiency and Cointegration: An Application to the Sterling and Deutschemark Exchange Markets, Journal of International Money and Finance 8, 75-88.

Hansen, Bruce E. 1992, Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes, Econometrica 60, 967-972.

Hamming, R. W. 1986, Numerical Methods for Scientists and Engineers. Dover Publications.

Hannan, E. J. 1970, Multiple Time Series. John Wiley and Sons.

Hasza, D. P. 1980, The Asymptotic Distribution of the Sample Autocorrelations for an Integrated ARMA Process, Journal of the American Statistical Association 75, 349-352.

Herrndorf, N. 1984, A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables, Annals of Probability 12, 141-153.

Johansen, Soren 1988, Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamics and Control 12, 231-254.

______ 1991, Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica 59, 1551-1580.

______ and Katarina Juselius 1990, Maximum Likelihood Estimation and Inference on Cointegration — with Applications to the Demand for Money, Oxford Bulletin of

Economics and Statistics 52, 169-210.

Kahn, James A. and Masao Ogaki 1990, A Chi-square Test for a Unit Root, Economics Letters 34, 37-42.

Kahn, James A. and Masao Ogaki 1991, A Consistent Test for the Null of Stationarity Against the Alternative of a Unit Root, Rochester Center for Economic Research Working Paper 304.

Kwiatkowski, Denis, Phillips, Peter C. B., Schmidt, Peter and Yongcheol Shin 1992, Testing the Null Hypothesis of Stationarity against the Alternative of A Unit Root, Journal of Econometrics 54, 159-178.

Lo, Andrews W. 1991, Long-Term Memory in Stock Market Prices, Econometrica 59, 1279-1313.

Meese , R. A. and K. J. Singleton 1983, On Unit Roots and the Empirical Modeling of Exchange Rates, International Economic Review 24, 1029-1035.

Mann, H. B. and A. Wald 1943, On the Statistical Treatment of Linear Stochastic Difference Equations, Econometrica 11, 172-200.

Nabeya, Seiji and Katsuto Tanaka 1988, Asymptotic Theory of a Test for the Constancy of Regression Coefficients against the Random Walk Alternative, The Annals of Statistics 16, 218-235.

Nelson, Charles R. and Charles I. Plosser 1982, Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics 10, 139-162.

Newey, Whiteney K. and Kenneth D. West 1987, A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703-708.

Ouliaris, Sam, Park, Joon Y. and Peter C. B. Phillips 1988, Testing for a Unit Root in the Presence of a Maintained Trend. In Advances in Econometrics and Modeling, ed. Baldev Raj, 7-28. Kluwer Academic Publishers.

Pagan, Adrian R. 1984, Model Evaluation by Variable Addition. In Econometrics and Quantitative Economics, ed. David F. Hendry and Kenneth F. Wallis, 103-133. Oxford: Blackwell.

Park, Joon Y. 1990, Testing for Unit Root and Cointegration by Variable Addition. In Advances in Econometrics 8, ed. Thomas B. Fomby and George F. Rhodes, Jr., 107-133. JAI Press.

Perron, Pierre 1988, Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach, Journal of Economic Dynamics and Control 12, 297-332.

______ 1989, The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, Econometrica 57, 1361-1401.

Phillips, Peter C. B. 1987, Time Series Regression with a Unit Root, Econometrica 55, 277-301.

______ and Pierre Perron 1988, Testing for Unit Root in Time Series Regression, Biometrika

71, 599-607.

Ploberger, W. and W. Kramer, 1992, The CUSUM Test with OLS Residuals, Econometrica 60, 271-285.

______, Kramer, W. and K. Kontrus, 1989, A New Test for Structural Stability in the Linear Regression Model, Journal of Econometrics 40, 307-318.

Pollard, David 1984, Convergence of Stochastic Processes. Springer-Verlag.

Richardson, Mathew and James H. Stock 1989, Drawing Inferences from Statistics Based on Multiyear Asset Returns, Journal of Financial Economics 25, 323-348.

Said, E. S. and David A. Dickey 1984, Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order, Biometrika 71, 599-607.

Samuelson, P. A. 1973, Proof That Properly Discounted Present Values of Assets Vibrate Randomly, Bell Journal of Economics and Management Science 4, 369-374.

Schmidt, Peter 1990, Dickey-Fuller Tests with Drift. In Advances in Econometrics 8, ed. Thomas B. Fomby and George F. Rhodes, Jr., 161-200. JAI Press.

Schotman, Peter C. and Herman F. van Dijk 1991, On Bayesian Routes to Unit Roots, Journal of Applied Econometrics 6, 387-401.

Schwert, G. W. 1989, Tests for Unit Roots: A Monte Carlo Investigation, Journal of Business and Economic Statistics 7, 147-159.

Stock, James H. 1992, Deciding between I(1) and I(0). NBER Technical Working Paper 121.

White, Halbert 1984, Asymptotic Theory for Econometricians. Academic Press.

White, J. S. 1958, The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case, Annals of Mathematical Statistics 29, 1188-1197.