T .s to certify that the

dissertation entitled

Analysis of Economic Time Series with Long Memory

presented by

Hyung Seung Lee

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Economics

Peter Schmidt
Major professor

Date July 27, 1995

ANALYSIS OF ECONOMIC TIME SERIES WITH LONG MEMORY

By

Hyung Seung Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1995

ABSTRACT

ANALYSIS OF ECONOMIC TIME SERIES WITH LONG MEMORY

By

Hyung Seung Lee

This dissertation focuses on economic time series that follow a general ARFIMA(p,d,q) process with $0< d <1$, which is intermediate between short memory ($d =0$) and unit root ($d =1$). Chapter 2 considers the unit root test proposed by Kwiatkowski, Phillips, Schmidt and Shin (KPSS, 1992) against I(d) alternatives. We show that the KPSS unit root test is consistent against stationary long memory processes ($d <1/2$) but is not consistent against nonstationary long memory processes ($d >1/2$). Therefore, the KPSS test only can distinguish short memory processes ($d=0$), stationary long memory processes and nonstationary processes. Simulation results are provided to support our asymptotic findings.

Chapter 3 considers the non-parametric estimation of the differencing parameter in the ARFIMA(p,d,q) process using the Adjusted Minimum Distance Estimator (AMDE) of Chung and Schmidt (1995). We compute the asymptotic bias of the AMDE and the MDEs that occur if we ignore short-run dynamics and estimate the (0,d,0) model. Our computational results for the ARFIMA(1,d,0) and

ARFIMA(0,d,1) models show that the asymptotic bias is larger when the short-run dynamics are stronger and when the number of ignored low-order autocorrelations is smaller.

Chapter 4 considers the estimation of the cointegrating coefficient in the case of fractional cointegration. We derive the asymptotic distribution of the OLS estimator under fairly strong assumptions and find that its order in probability is T^{d-1}, for $-1/2 < d < 3/2$ except $d = 1/2$. Also we derive the asymptotic distribution of OLS in differences and find that it is not consistent unless the error and the regressor are uncorrelated. We provide simulation results that support our asymptotic findings.

ACKNOWLEDGMENTS

This dissertation can be finished by guidance, support and encouragement from many people around me. Especially, the wonderful guidance from Professor Peter Schmidt makes it possible for me to complete the dissertation smoothly and quickly. His careful comments and advice, which are essential to improve this dissertation from the beginning to the end, are highly appreciated. I also wish to thank my other committee members, Professor Richard Baillie, Professor Ching-Fan Chung and Professor Jeffrey Wooldridge for their useful comments.

The financial support from Korean government allowed me to start this program and I thank everyone who involved it. I am grateful to many people who are important to me during the whole process and would like to name some of them who are special. The aid from Dr. Jaeyoun Hwang, Dr. Dongin Lee and Dr. Kyungso Im is valuable in studying econometrics. The friendship with Inkyo Cheong, Youngmin Kwon and Hailong Qian has made the sailing enjoyable.

Lastly, I express thanks to encouragement and support from my grandma, parents, parents-in-law, aunts, brothers and sister, brother and sister-in-law, and Maeng's family. The love from my wife, Jeongeun, and two daughters, Jimin and Jiwon is always with me and I am indebted to them forever.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Classical methods of time series analysis assume stationarity, so that the series fluctuates around its mean level (or a trend) without changes in its autocovariance structure over time. Stationary series are often assumed to follow an ARMA process, which implies that the series has short memory in the sense that its autocorrelations and impulse response weights decay at a geometric rate. Models which can deal with time series that are more persistent than a short memory process have focused primarily on the existence of a unit root process.

However, the classification of time series into either unit root or stationary ARMA processes is too extreme and restrictive. Between these two types of processes, a long memory process can be considered to cover intermediate cases which are not well fit by either short memory or unit root models. In the data, when the sample autocorrelations do not decay quickly for long lags and yet the low order autocorrelations are not close to unity, we can suspect a long memory process. That is, a long memory process displays autocorrelations that are too small at low orders for a unit root, but too persistent at long lags for a stationary ARMA process.

There are many cases of long memory models in the physical sciences. Data with hyperbolically decaying autocorrelations and impulse response weights were firstly observed by Hurst (1951, 1956) and Mandelbrot and Wallis (1968) in hydrology and climatology. In economics, many financial data, such as forward premiums, interest rate differentials and inflation rates, have recently been found to display long memory characteristics. Baillie (1995) provides a survey of the application of long memory models in economics.

There are several possible definitions of the concept of long memory. McLeod and Hipel (1978) defined a stochastic process to be long memory if the autocorrelation function is not summable; that is, with ρ_j = j-th autocorrelation,

(1) $$\lim_{T\to\infty} \sum_{j=-T}^{T} \left|\rho_j\right| \neq \text{finite}.$$

A more specific definition of long memory is that the process has autocorrelations that decline hyperbolically at large lags:

(2) $$\gamma_k \sim \lambda k^{-\alpha},\ (\lambda > 0,\ 0 < \alpha < 1),\ \text{ as } k \to \infty.$$

(Here $a_k \sim b_k$ means $a_k / b_k \to 1$ as $k \to \infty$.) In (1) above, λ can be a constant, or more generally it can be a function of k that is slowly varying at infinity (i.e., $\lambda(ck) / \lambda(k) \to 1$, as $k \to \infty$, for any $c > 0$). Long memory can be defined equivalently in terms of the behavior of the spectral density as one approaches the zero frequency. A long memory process has infinite spectral density at zero frequency, as does a unit root process; however, unlike a unit root process, the spectral density at zero of the first difference of a long memory process vanishes.

Rosenblatt (1956) has defined long memory based on the dependence between two points of a process. Mandelbrot and Van Ness (1968) and Mandelbrot (1970) formalized Hurst's empirical findings and defined fractional Gaussian noise which is designed to account for the long run behavior of a long memory time series. Granger and Joyeux (1980) and Hosking (1981) proposed an alternative long memory process, the fractionally integrated process. Geweke and Porter-Hudak (1983) proved that the definition of fractional Gaussian noise by Mandelbrot and Van Ness and the fractionally integrated process are equivalent.

We will consider in detail the fractionally integrated process of Granger (1980), Granger and Joyeux (1980) and Hosking (1981). A time series $\{y_t\}$ is a fractionally integrated process of order d, I(d), if

(3) $$(1-L)^d y_t = \varepsilon_t,$$

where L is the lag operator, d is the differencing parameter and $\{\varepsilon_t\}$ is a white noise process with zero mean and finite variance σ_ε^2. For any d > -1, y_t is invertible (Odaki (1993)) and $(1-L)^d$ can be expressed via the binomial expansion:

(4) $$(1-L)^d = \sum_{j=0}^{\infty}(-1)^j \binom{d}{j} = \sum_{j=0}^{\infty}\pi_j L^j = F(-d,1;1;L),$$

where for j = 0, 1, 2, ...,

(5A) $$\binom{d}{j} = \frac{d(d-1)\cdots(d-j+1)}{j!},$$

(5B) $$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<k\le j}\frac{k-1-d}{k},$$

(5C) $$\Gamma(x) = \text{gamma function} = \begin{cases} \int_0^\infty t^{x-1}e^{-t}dt, & x>0, \\ \infty, & x=0, \\ \dfrac{\Gamma(1+x)}{x}, & x<0, \end{cases}$$

(5D) F(a,b;c;z) = hypergeometric function

$$= 1 + \frac{ab}{c \cdot 1} z + \frac{a(a+1)b(b+1)}{c(c+1)\cdot 1 \cdot 2} z^2 + \cdots$$

$$= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)\Gamma(j+1)} z^j .$$

Therefore, the infinite AR representation is the following:

(6) $$y_t = \sum_{j=1}^{\infty} \phi_j y_{t-j} + \varepsilon_t, \text{ where } \phi_j = -\pi_j.$$

y_t is a stationary process for $d < 1/2$. Its infinite MA representation can be expressed as follows:

(7) $$y_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j},$$

where

(8) $$\theta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \; j = 0, 1, 2, \ldots .$$

Its variance-covariance structure is as follows:

(9) $$\sigma_y^2 = \frac{\sigma_\varepsilon^2 \Gamma(1-2d)}{\Gamma^2(1-d)},$$

(10) $$\rho_j = \frac{\Gamma(j+d)\Gamma(1-d)}{\Gamma(d)\Gamma(j-d+1)} = \prod_{k=1}^{j} \frac{k-1+d}{k-d}, \; j = 1, 2, 3, \ldots .$$

Thus for $-1 < d < 1/2$, y_t is a stationary and invertible fractionally integrated process. Since θ_j represents the impact of ε_{t-j} on y_t, the cumulative impulse response ($=\theta(1)$) which is the total effect of a unit innovation can be obtained by summation of θ_j:

(11) $$\theta(1) = \sum_{j=0}^{\infty} \theta_j = \lim_{N\to\infty} \sum_{j=0}^{N} \theta_j = \lim_{N\to\infty} \frac{N^d}{\Gamma(1+d)} = \begin{cases} 0, & d < 0 \\ \infty, & d > 0. \end{cases}$$

For more details, see Sowell (1990).

The AR weights, MA weights and autocorrelations all decay hyperbolically, though at different hyperbolic rates:

(12A) $\pi_j \sim \frac{1}{\Gamma(-d)} j^{-d-1}$ as $j \to \infty$,

(12B) $\theta_j \sim \frac{1}{\Gamma(d)} j^{d-1}$ as $j \to \infty$,

(12C) $\rho_j \sim \frac{\Gamma(1-d)}{\Gamma(d)} j^{2d-1}$ as $j \to \infty$.

This is in contrast to the case of a stationary ARMA process, for which the autocorrelations decrease rapidly (at an exponential rate rather than at the hyperbolic rate for an I(d) process). Since θ_j is close to zero for large j as long as $d < 1$, an I(d) process with $1/2 \leq d < 1$ is still mean-reverting, even though it is not stationary. Baillie (1995) showed this result using the cumulative impulse response functions of the first difference of an I(d) process. For further details see Chung (1994b).

The spectral density at zero frequency is another measure of persistence in a time series. The spectral density of an I(d) process is: for $-\pi \leq \omega \leq \pi$,

(13) $$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left|1 - e^{-i\omega}\right|^{-2d} = \frac{\sigma_\varepsilon^2}{2\pi}\left|2\sin(\omega/2)\right|^{-2d}.$$

The spectral density at $\omega = 0$ is infinite for $d > 0$, finite for $d = 0$, and zero for $d < 0$. More specifically, because $\sin\omega \sim \omega$ as $\omega \to 0$, $f(0) \sim (\sigma^2/2\pi)\omega^{-2d}$ as $\omega \to 0$. Thus

(14) $$f(0) = \begin{cases} 0, & d < 0, \\ \infty, & d > 0. \end{cases}$$

$$(15) \qquad f'(0) = \begin{cases} 0, & d \le -1/2, \\ -\infty, & -1/2 < d < 0, \\ \infty, & d > 0. \end{cases}$$

Therefore, the differencing parameter d is not identified by the level or derivative of its spectral density at zero frequency (Sowell, 1992b).

An I(d) process can be extended to cover more general economic time series models when ε_t in (3) is allowed to follow a general stationary ARMA process. A time series $\{y_t\}$ is an autoregressive fractionally integrated moving average process of order p, d, q, or ARFIMA(p,d,q), if it satisfies:

$$(16) \qquad (1-L)^d y_t = \varepsilon_t = \Phi(L)^{-1}\Theta(L)u_t \quad [\text{so } \Phi(L)\varepsilon_t = \Theta(L)u_t];$$

$$(17) \qquad \Phi(L) = 1+\phi_1 L+\phi_2 L^2+\cdots+\phi_p L^p \text{ and } \Theta(L) = 1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q;$$

where all the roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle, Φ and Θ have no common roots, and $\{u_t\}$ is white noise. For $-1< d <1/2$ the ARFIMA(p,d,q) process is stationary and invertible. For the general ARFIMA(p,d,q) process, Sowell (1992a) and Chung (1994a) show how to compute the autocovariance and autocorrelation functions. In a stationary and invertible ARFIMA(p,d,q) model the autocorrelations decay at the same hyperbolic rate as in the corresponding I(d) process; the rate of decay is independent of the ARMA parameters. Specifically, if ρ_k^*, k = 0, 1, ..., represent the autocorrelations of the ARFIMA(p,d,q) process,

$$(18) \qquad \rho_k^* \sim Ck^{2d-1}, \text{ as } k\to\infty.$$

Here $C\neq 0$ is a constant that depends on the ARMA parameters.

Since the ARFIMA process is an I(d) process with ARMA error or an ARMA process with an I(d) error, its spectral density function is:

$$(19)\qquad f^*(\omega) = \frac{\left|\Theta(e^{-i\omega})\right|^2}{\left|\Phi(e^{-i\omega})\right|^2} f(\omega)$$

$$= \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left|\Theta(e^{-i\omega})\right|^2}{\left|\Phi(e^{-i\omega})\right|^2} \left|1 - e^{-i\omega}\right|^{-2d}$$

$$= \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left|\Theta(e^{-i\omega})\right|^2}{\left|\Phi(e^{-i\omega})\right|^2} \left[2\left|1 - \cos(\omega)\right|\right]^{-2d}, \quad -\pi < \omega < \pi.$$

At zero frequency the spectral density is similar to the expression for the I(d) case:

$$(20)\qquad f^*(0) \sim \frac{\sigma_\varepsilon^2}{2\pi} \left[\frac{\Theta(1)}{\Phi(1)}\right]^2 \omega^{-2d} \quad \text{as } \omega \to 0 .$$

The asymptotic distribution for many statistics based on data generated by an I(d) process will be established using a functional central limit theorem involving the fractional Brownian motion of Mandelbrot and Van Ness (1968):

$$(21)\qquad W_d(r) = \frac{1}{\Gamma(d+1)} \int_0^r (r-s)^d dW(s), \; r \in [0, 1],$$

where W(s) is the standard Brownian motion. To state this functional central limit theorem, suppose that v_t is an I(d) process and V_t is its cumulation:

$$(22)\qquad V_t = \sum_{j=1}^{t} v_j .$$

Thus $(1-L)^d v_t = u_t$ where u_t is short memory. We follow Lee and Schmidt (1995) in assuming the following (Assumption A):

(A1) v_t is I(d) with $|d| < 1/2$.

(A2) u_t is iid N(0, σ^2_u).

This assumption is somewhat stronger than others have made, and probably stronger than necessary; see Sowell (1990), Lo (1991) and Hosking (1984).

Define the variance of the partial sum process as in Sowell (1990):

$$\sigma_T^2 = \mathrm{Var}(V_T) = \mathrm{Var}(\sum_{j=1}^{T} v_j). \tag{23}$$

Then when v_t follows Assumption A, Sowell (1990) shows that:

$$\sigma_T^2 = \sigma_u^2 \frac{\Gamma(1-2d)}{\Gamma(1+2d)\Gamma(1+d)\Gamma(1-d)}\left[\frac{\Gamma(1+d+T)}{\Gamma(T-d)} - \frac{\Gamma(1+d)}{\Gamma(-d)}\right] \tag{24}$$

and, as $T \to \infty$,

$$\frac{\sigma_T^2}{T^{1+2d}} \to \sigma_u^2 \frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)} \equiv \omega_d^2. \tag{25}$$

Furthermore, under Assumption A and using results of Davydov(1970), Sowell (1990, p.498) shows the following invariance principle, for $r \in [0,1]$:

$$\frac{V_{[rT]}}{\sigma_T} \Rightarrow W_d(r), \tag{26}$$

or, equivalently,

$$\frac{V_{[rT]}}{T^{d+1/2}} \Rightarrow \omega_d W_d(r). \tag{27}$$

The discussion above has focused on the case of a stationary long-memory process, and specifically on the I(d) process with $|d|<1/2$. However, we will be interested primarily in positive values of d, because an I(d) process with $d<0$ is anti-persistent, which is not of empirical relevance. In some cases it may be useful, theoretically and empirically, to consider nonstationary long memory processes. A specific type of nonstationary long memory process is the I(d) process with $1/2< d <1$. An I(d) process with $1/2< d <1$ is nonstationary but still mean-reverting. This contrasts with a stationary long memory process, which is stationary and mean-reverting; also with a unit root process, which is nonstationary and not mean-reverting. The discussion above applies to such series after differencing, since "y_t is I(d) with $1/2< d <1$" is equivalent to "Δy_t is I(d) with $-1/2< d <0$."

In this dissertation we will investigate how we can distinguish among three different kinds of processes, namely short memory, long memory and unit root processes. This is empirically relevant because for some data one can reject both the null hypothesis of a unit root and the null hypothesis of short memory. It is possible that such series may follow a long memory process, and we need to test this possibility. Also we will review estimation of the differencing parameter which determines the main stochastic properties of a long memory process. Lastly the case that the error in a cointegrating relationship of unit root series is I(d) will be considered.

The plan of this dissertation is as follows. In Chapter 2 we will check whether the KPSS test, introduced by Kwiatkowski, Phillips, Schmidt and Shin (1992), is useful in distinguishing short memory, long memory and unit root processes. We will show that the KPSS test can not consistently distinguish a unit root from a nonstationary long memory process, since the order in probability of the KPSS statistic is equivalent under these two processes. However, the KPSS statistic can distinguish consistently between short memory, stationary long memory, and either unit root or nonstationary long memory. We provide simulation results which support our asymptotic results and we compare our results with other results of Diebold and Rudebusch (1991) and Hassler and Wolters (1994) for Dickey-Fuller type tests.

In Chapter 3 we will consider the problem of Minimum Distance Estimation (MDE) of the differencing parameter of a fractionally integrated long memory process. The simple MDE proposed by Tieslau, Schmidt and Baillie (1995), which minimizes the difference between sample and population autocorrelations, is useful because it does not require a distributional assumption and it is easy to compute, which is also true in the Adjusted MDE of Chung and Schmidt (1995). Furthermore in the general ARFIMA model it provides a way to estimate the differencing parameter separately from the ARMA parameters which determine short-run dynamics. However, such a non-parametric treatment of short-run dynamics will cause asymptotic bias, and we investigate ways of decreasing the bias due to ignored short-run dynamics. We investigate how the size of the bias is affected by the value of d and of the ARMA parameters, the number of moment conditions used, and the order of autocorrelations considered. Our computations show that, for certain methods of expressing the moment conditions suggested by Chung and Schmidt

(1995), the asymptotic bias becomes small when only high-order autocorrelations are used or the short-run dynamics are not strong.

In Chapter 4 we consider the estimation of the cointegrating vector in the case that the error in a cointegrating relationship is I(d) with $0< d <1$, rather than in the usual case with I(0) errors. We find that OLS in this case is still consistent and its order in probability depends on the value of d. Specifically, for $0< d <1$ OLS is $O_p(T^{1-d})$. For comparison we also consider OLS in differences. We find that it is not consistent if the errors and regressors are correlated, and it converges at the usual rate $T^{1/2}$ for all values of $d \in [0, 1]$. We provide some simulation results that support these asymptotic results.

Finally in Chapter 5 we summarize our results and make some suggestions for future research.

CHAPTER 2

CONSISTENCY OF THE KPSS UNIT ROOT TEST AGAINST FRACTIONALLY INTEGRATED ALTERNATIVES

1. INTRODUCTION

Since Nelson and Plosser (1982), there has been an enormous body of theoretical and empirical work seeking to distinguish whether economic time series are trend stationary or have a unit root. This distinction is important for both economic and statistical reasons. For a survey, see Diebold and Nerlove (1992).

There are two main approaches to this problem. The most traditional approach is to test the null hypothesis of a unit root against the alternative hypothesis of trend stationarity. For this problem, the Dickey-Fuller tests were introduced by Dickey (1976), Fuller (1976), and Dickey and Fuller (1979). The standard Dickey-Fuller tests are extended to allow general ARMA error processes by Said and Dickey (1984), Phillips (1987) and Phillips and Perron (1988). Dejong, Nankervis, Savin and Whiteman (1992) found that the standard Dickey-Fuller tests and the extensions of Said-Dickey, Phillips-Perron, and Choi-Phillips (1991) have trouble distinguishing unit root processes with substantial short-run dynamics from trend stationary alternatives.

Conversely, a more recent approach is to test the null hypothesis of stationarity against the alternative of a unit root. Tests of the null of stationarity have been suggested by Park and Choi (1988), Kwiatkowski, Phillips, Schmidt, and Shin (1992) (hereafter, KPSS), Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994). In this chapter we will consider the KPSS test, which is a test of the null hypothesis of stationarity around a deterministic trend, and which controls for short-run dynamics using a non-parametric correction similar to those used by Phillips and Perron (1988) or Schmidt and Phillips (1992). Since many simulation results show that the traditional Dickey-Fuller tests are not reliable in the presence of MA errors whose coefficient is not close to zero [for details see Agiakloglou and Newbold (1992), Schwert (1989), Pantula (1991)], Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994) suggest tests of the stationary null hypothesis that are similar to KPSS, but which differ from KPSS in the way they deal with autocorrelation under the null hypothesis.

The asymptotic analysis of the Dickey-Fuller type unit root tests, including those extended versions which allow error autocorrelation, shows that those tests are consistent against stationary alternatives. Also, the KPSS stationarity test is consistent against unit root alternatives. Although the KPSS test was originally intended as a test of the null of stationarity against the unit root alternative, it can also be used as a test of the unit root null against the alternative of stationarity. This has been suggested by Shin and Schmidt (1992) and Stock (1990). Shin and Schmidt (1992) show that the KPSS unit root test is consistent against the alternative hypothesis of stationarity.

A common empirical puzzle is what to conclude when one rejects both the null of a unit root (e.g., using the Dickey-Fuller tests) and the null of stationarity (e.g., using the KPSS test). To understand this outcome, suppose that z_t $(t = 1, 2, \ldots)$ is the series in question and that Z_t is its cumulation (partial sum), i.e.,

$$Z_t = \sum_{j=1}^{t} z_j .$$

Then we follow Lee and Schmidt (1995) in saying that z_t is a short memory process if it satisfies the following two requirements (Assumption B):

(B1) $\sigma^2 = \lim_{T\to\infty} T^{-1}E(Z_T^2)$ exists and is non-zero.

(B2) $\forall r \in [0,1]$, $T^{-1/2}Z_{[rT]} \Rightarrow \sigma W(r)$.

Here [rT] denotes the integer part of rT, $\Rightarrow$ denotes weak convergence, and W(r) is the standard Wiener process (Brownian motion). The concept of short memory is important because the asymptotic analysis of the KPSS test actually assumes that under the null the series is short memory, and the asymptotic analysis of unit root tests actually assumes that under the null the first difference of the series is short memory. Thus we can rationalize rejections of both null hypotheses by postulating series that are not short memory either in levels or in first differences.

These arguments lead to the consideration of long memory processes that are more persistent than a short memory process, but less persistent than a unit root process. Accordingly, they are not short memory either in levels or in first differences. The consideration of such long memory time series has mostly taken place in the physical sciences. They have been applied extensively in hydrology (Hurst, 1951, 1956) and have also been used to model data on temperatures and growth of tree rings (Seater, 1993). The Beveridge wheat price index from 1500 through 1869 (Beveridge, 1921) and U.S. monthly consumer price index inflation rates are examples of economic data that exhibit typical long memory features. There are also studies of long memory in a spatial context; e.g., Whittle (1956) and Beran

(1992). A good survey of long memory from the point of view of economics and econometrics is given by Baillie (1995).

We will use the fractionally integrated process defined by Granger (1980), Granger and Joyeux (1980) and Hosking (1981), and considered by Lee and Schmidt (1995), which is introduced in Chapter 1:

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

where L is the lag operator, d is the differencing parameter and $\{\varepsilon_t\}$ is a short memory process with zero mean and finite variance σ_ε^2.

There has been some recent research on tests related to the fractionally integrated long memory process. Lo (1991) finds that his "rescaled range" test, for which the null hypothesis is short memory, is consistent against I(d) processes with $d \in (-1/2, 1/2)$. Cheung (1993) investigated the finite sample performance of the GPH test, the modified rescaled range test and two LM type tests of the null of short memory against the alternative of fractional integration. Lee and Schmidt (1995) show that the KPSS "stationarity" test is actually a test of the null hypothesis of short memory, and that it is consistent against stationary long memory alternatives (I(d) for $-1/2 < d < 1/2$ and $d \neq 0$). They also provide simulation results on the power of the KPSS test. They found the power of the KPSS short memory test in finite samples to be comparable to that of Lo's rescaled range test. Their results suggest that the KPSS test can be used to distinguish a short memory and stationary long memory processes but a rather large sample size is required to do so reliably.

There are several studies on the power of unit root tests against fractionally integrated alternatives. Diebold and Rudebusch (1991) give Monte Carlo evidence of the low power of the Dickey-Fuller test against fractionally integrated alternatives with $d > 1/2$; that is, nonstationary long memory alternatives. Sowell (1990) derives the asymptotic distribution of the Dickey-Fuller tests under the hypothesis of an I(d) process with $1/2 < d < 3/2$ and shows the consistency of these tests against nonstationary long memory alternatives. Hassler and Wolters (1994) show that the Dickey-Fuller type tests, including the Said-Dickey and Phillips-Perron extensions, have low power in finite samples against I(d) alternatives with $0 < d < 1$, and especially that the augmented Dickey-Fuller test works poorly.

The purpose of this chapter is to investigate whether the KPSS test is useful in distinguishing short memory, long memory and unit root processes. Specifically, we want to ask whether the KPSS test can distinguish the following four types of processes: (i) short memory ($d = 0$); (ii) stationary long memory ($|d| < 1/2$, $d \neq 0$); (iii) nonstationary long memory ($1/2 < d < 1$); and (iv) unit root ($d = 1$). Asymptotics for the KPSS statistic are previously known for cases (i), (ii) and (iv), but not for (iii). Therefore we need to derive the asymptotic distribution of the KPSS statistic when $1/2 < d < 1$.

In the following it will be shown that the asymptotic distribution of the KPSS statistic in the case of a nonstationary long memory process ($1/2 < d < 1$) is different from the other cases, but its order in probability is the same as in the case of a unit root. Therefore, the KPSS unit root test is inconsistent against nonstationary long memory alternatives. More generally, the KPSS test can not consistently distinguish a unit root from a nonstationary long memory process. Using the KPSS statistic we can only distinguish consistently between the following three cases: (i) short memory; (ii) stationary long memory; and (iii) either nonstationary long memory or unit root.

Some Monte Carlo evidence on finite sample power is also provided. It is generally in agreement with the asymptotic results.

2. THEORETICAL RESULTS

A. The KPSS Test Under Short Memory and Unit Root

We consider the data generating process:

$$y_t = \phi + \xi t + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{2}$$

where $\{y_t\}$ is the observed series and $\{\varepsilon_t\}$ is the deviation from deterministic linear trend. Let e_t be the residuals from a regression of y_t on intercept and trend (t), and let S_t be the partial sum of the e_t:

$$S_t = \sum_{j=1}^{t} e_j .$$

Let σ^2 be the long-run variance of the ε_t, as in (B1) above and let $s^2(\ell)$ be the Newey-West estimator of σ^2:

$$s^2(\ell) = \frac{1}{T}\sum_{t=1}^{T} e_t^2 + \frac{2}{T}\sum_{s=1}^{\ell} w(s,\ell) \sum_{t=s+1}^{T} e_t e_{t-s} . \tag{3}$$

Here $w(s, \ell) = 1 - \frac{s}{\ell + 1}$, and ℓ is chosen so that $\ell \to \infty$ but $\ell / T \to 0$ as $T \to \infty$. We will later also consider the case that $\ell = 0$, in which case the second term on the right hand side of (3) is set to zero and $s^2(0) = \frac{1}{T} \sum_{t=1}^{T} e_t^2$.

The KPSS statistic is then defined as:

$$\text{(4)} \qquad \hat{\eta}_\tau(\ell) = \frac{T^{-2} \sum_{t=1}^{T} S_t^2}{s^2(\ell)}.$$

The KPSS statistic $\hat{\eta}_\mu(\ell)$ is defined similarly except that we set $\xi = 0$ in (2), which implies use of the residuals $e_t = y_t - \bar{y}$ in defining S_t and $s^2(\ell)$.

Under the hypothesis that ε_t is a short-memory process, KPSS show that

$$T^{-2} \sum_{t=1}^{T} S_t^2 \Rightarrow \int_0^1 V_2(r)^2 \, dr,$$

where $V_2(r)$ is a second-level Brownian bridge, as defined by KPSS, equation (16). Also $s^2(\ell)$ is a consistent estimator of σ^2. Therefore,

$$\hat{\eta}_\tau(\ell) \Rightarrow \int_0^1 V_2(r)^2 \, dr.$$

Similar statements hold for $\hat{\eta}_\mu(\ell)$, with $V_2(r)$ replaced by the standard Brownian bridge, $V_1(r) = W(r) - rW(1)$. For the purpose of the present chapter, the important result is that $\hat{\eta}_\tau(\ell)$ and $\hat{\eta}_\mu(\ell)$ are $O_p(1)$ when ε_t is short memory.

Next consider the case that ε_t is a unit root process, in the sense that $\Delta\varepsilon_t$ is short-memory. In this case KPSS show that

(5) $$T^{-4}\sum_{t=1}^{T} S_t^2 \Rightarrow \sigma^2 \int_0^1 \left(\int_0^a W^*(s)ds \right)^2 da,$$

where W*(s) is a demeaned and detrended Wiener process, as defined in Park and Phillips (1988, p.474), and σ^2 is the long run variance of $\Delta\varepsilon_t$. Furthermore,

(6) $$\frac{s^2(\ell)}{\ell T} \Rightarrow \sigma^2 \int_0^1 W^*(s)^2 ds.$$

This implies that

(7) $$(\ell/T)\hat{\eta}_\tau(\ell) \Rightarrow \frac{\int_0^1 \left(\int_0^a W^*(s)ds \right)^2 da}{\int_0^1 W^*(s)^2 ds}.$$

Therefore $\hat{\eta}_\tau(\ell)$ is $O_p(T/\ell)$ when ε_t is a unit root process. If we set ξ=0 in (2), then $\hat{\eta}_\mu(\ell)$ is also $O_p(T/\ell)$: in fact, we have the same result as in (7) except that W*(s) is replaced by the demeaned Brownian motion, $\underline{W}(s)$:

$$\underline{W}(s) = W(s) - \int_0^1 W(r)dr.$$

The KPSS unit root test suggested by Shin and Schmidt (1992) sets $\ell = 0$, since the distribution in (7) is independent of the nuisance parameter σ^2 for all values of ℓ, including $\ell = 0$, under the unit root hypothesis. Then $T^{-1}\hat{\eta}_\tau(0)$ has the same distribution as on the right hand side of (7) above, and $\hat{\eta}_\tau(0)$ is $O_p(T)$ under the hypothesis that ε_t is a unit root process.

These results are easy to summarize. (1) When $\ell = 0$, $\hat{\eta}_\tau(0)$ and $\hat{\eta}_\mu(0)$ are $O_p(1)$ if ε_t is short memory and $O_p(T)$ if ε_t has a unit root. (2) If $\ell \to \infty$ but $\ell/T \to 0$ as $T\to\infty$, $\hat{\eta}_\tau(\ell)$ and $\hat{\eta}_\mu(\ell)$ are $O_p(1)$ if ε_t is short memory and $O_p(T/\ell)$ if ε_t has a unit root. Thus, in either case, the KPSS statistic distinguishes consistently (correctly with probability one as $T\to\infty$) between short memory and unit root processes.

B. Asymptotics and Consistency of the KPSS Unit Root Test Under I(d)

First we will show the consistency of the lower tail KPSS unit root test against the stationary long memory alternative hypothesis ($-1/2< d <1/2$). Thus we suppose that $(1-L)^d \varepsilon_t = u_t$, with $-1/2< d <1/2$, and with Assumption B satisfied. Under these assumptions, Lee and Schmidt (1995) derive the asymptotic distribution of the KPSS statistics. In the level stationary case ($e_t = y_t - \bar{y}$), from their Lemma 1, Theorem 1 and Theorem 3:

(8A) $$T^{-(2d+1)}\sum_{t=1}^{T} S_t^2 \Rightarrow \omega_d^2 \int_0^1 B_d(r)^2\,dr\text{, where } B_d(r) = W_d(r) - rW_d(1);$$

(8B) $$s^2(0) \xrightarrow{p} \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_t) = \sigma_u^2 \frac{\Gamma(1-2d)}{\{\Gamma(1-d)\}^2} \quad (\ell = 0);$$

(8C) $$\frac{s^2(\ell)}{\ell^{2d}} \xrightarrow{p} \omega_d^2 \quad (\ell\to\infty \text{ but } \ell/T\to 0 \text{ as } T\to\infty).$$

Therefore, the asymptotic distributions of the KPSS statistics in the level stationary case, when ε_t is I(d) with $-1/2< d <1/2$, are as follows:

(9A) $$\frac{1}{T^{2d}}\hat{\eta}_\mu(0) \Rightarrow \frac{\omega_d^2}{\sigma_\varepsilon^2}\int_0^1 B_d(r)^2 dr \quad (\ell = 0)$$

(9B) $$(\frac{\ell}{T})^{2d}\hat{\eta}_\mu(\ell) \Rightarrow \int_0^1 B_d(r)^2 dr \quad (\ell \to \infty \text{ but } \ell / T \to 0 \text{ as } T \to \infty).$$

Furthermore, $\hat{\eta}_\tau$ has the same orders in probability as $\hat{\eta}_\mu$, and its asymptotic distribution is exactly the same as in (9) except that $B_d(r)$ should be replaced by $V_d(r)$, defined as:

(10) $$V_d(r) = W_d(r) + (2r - 3r^2)W_d(1) + (-6r + 6r^2)\int_0^1 W_d(s)ds.$$

Thus the KPSS statistics $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ are $O_p(T^{2d})$ for $\ell = 0$ and $O_p((T/\ell)^{2d})$ for $\ell \to \infty$ under the stationary long memory alternative, while they are respectively $O_p(T)$ and $O_p(T/\ell)$ under the null hypothesis of a unit root.

This implies the following result.

THEOREM 1:

Suppose that ε_t is I(d) with $d \in (-1/2, 1/2)$ and Assumption B is satisfied. Then

$$\frac{1}{T}\hat{\eta}_\mu(0) \xrightarrow{p} 0, \quad \frac{1}{T}\hat{\eta}_\tau(0) \xrightarrow{p} 0 \quad (\ell = 0)$$

$$\frac{\ell}{T}\hat{\eta}_\mu(\ell) \xrightarrow{p} 0, \quad \frac{\ell}{T}\hat{\eta}_\tau(\ell) \xrightarrow{p} 0 \quad (\ell \to \infty \text{ but } \ell/T \to 0 \text{ as } T \to \infty).$$

Proof: The KPSS test statistics $(1/T)\hat{\eta}_\mu$ and $(\ell/T)\hat{\eta}_\mu$ (also, $(1/T)\hat{\eta}_\tau$ and $(\ell/T)\hat{\eta}_\tau$) are each $O_p(T^{2d-1})$ and $O_p((T/\ell)^{2d-1})$, and 2d is less than 1 because $|d|<1/2$. ■

Theorem 1 implies that the lower tail KPSS unit root test is consistent against the stationary long memory alternative hypothesis. However, as d approaches 1/2, the order in probability of the KPSS statistic under the I(d) alternative approaches the same order in probability as under the unit root null hypothesis. This suggests that when d is close to 1/2 the power of KPSS unit root test would be small. There is also an issue of the continuity of the power of the KPSS test against I(d) alternatives as d $\rightarrow 1/2$.

We now turn to the main theoretical contribution of this chapter, which is the derivation of the asymptotic distribution of the KPSS statistics when ε_t is a nonstationary long memory process. Thus we wish to consider the case that ε_t is I(d) with $1/2 < d < 3/2$.

Define $d^* = d-1$, so that $\Delta\varepsilon_t$ is I(d^*) with $|d^*| < 1/2$; that is, $\Delta\varepsilon_t$ is a stationary long memory process. We assume that Assumption A in Chapter 1 holds with $v_t = \Delta\varepsilon_t$. Then ε_t is the cumulation of the stationary I(d^*) variables $\Delta\varepsilon_t$, and we have the invariance principle:

$$\text{(11)} \qquad \frac{\varepsilon_{[rT]}}{T^{d^*+1/2}} \Rightarrow \omega_{d^*} W_{d^*}(r).$$

Note that this is really the same invariance principle as equation (27) in Chapter 1, with d^* replacing d because $1/2 < d < 3/2$ and $|d^*| < 1/2$.

We will first consider the $\hat{\eta}_\mu$ test. Thus we consider $e_t = y_t - \bar{y} = \varepsilon_t - \bar{\varepsilon}$ and $S_t = \sum_{j=1}^{t} e_j$. Then we can derive the following results.

LEMMA 1:

Suppose $(1-L)^d \varepsilon_t = u_t$ for 1/2< d <3/2, d*=d-1, and $\Delta\varepsilon_t$ satisfies Assumption A in Chapter 1. Then

(i) $$\frac{1}{T^{d^*+3/2}}\sum_{t=1}^{[rT]} \varepsilon_t \Rightarrow \omega_{d^*}\int_0^r W_{d^*}(a)da ;$$

(ii) $$\frac{1}{T^{d^*+3/2}} S_{[rT]} \Rightarrow \omega_{d^*}\int_0^r \underline{W_{d^*}}(a)da, \text{ where } \underline{W_{d^*}}(a) = W_{d^*}(a) - \int_0^1 W_{d^*}(b)db ;$$

(iii) $$\frac{1}{T^{2d^*+4}}\sum_{t=1}^{T} S_t^2 \Rightarrow \omega_{d^*}^2\int_0^1\left[\int_0^r \underline{W_{d^*}}(a)da\right]^2 dr .$$

Proof: See Appendix. ■

THEOREM 2:

Under the same assumptions as in *LEMMA* 1,

(i) When $\ell = 0$, then $$\frac{1}{T^{1+2d^*}} s^2(0) \Rightarrow \omega_{d^*}^2\left\{\int_0^1 \underline{W_{d^*}}(a)^2 da\right\}$$

(ii) When $\ell \to \infty$ and $\ell/T \to 0$ as $T \to \infty$, then

$$\frac{s^2(\ell)}{\ell T^{1+2d^*}} \Rightarrow \omega_{d^*}^2\left\{\int_0^1 \underline{W_{d^*}}(a)^2 da\right\}.$$

Proof: See Appendix. ■

Then we can prove the following theorem.

THEOREM 3:

Under the same assumptions as in *LEMMA* 1,

$$\frac{1}{T}\hat{\eta}_\mu(0) \Rightarrow \frac{\int_0^1 \left\{\int_0^r \underline{W_{d*}}(a)da\right\}^2 dr}{\left\{\int_0^1 \underline{W_{d*}}(a)^2 da\right\}} \quad \text{(for } \ell = 0\text{)}$$

$$\frac{\ell}{T}\hat{\eta}_\mu(\ell) \Rightarrow \frac{\int_0^1 \left\{\int_0^r \underline{W_{d*}}(a)da\right\}^2 dr}{\left\{\int_0^1 \underline{W_{d*}}(a)^2 da\right\}} \quad \text{(when } \ell \to \infty \text{ and } \ell/T \to 0 \text{ as } T \to \infty\text{)}$$

Proof: Since $\frac{1}{T}\hat{\eta}_\mu(0) = \frac{T^{-(2d^*+4)}\sum_{t=1}^{T} S_t^2}{T^{-(2d^*+1)}s^2(0)}$ and $\frac{\ell}{T}\hat{\eta}_\mu(\ell) = \frac{T^{-(2d^*+4)}\sum_{t=1}^{T} S_t^2}{\ell T^{-(2d^*+1)}s^2(\ell)}$, the asymptotic distribution of the numerator is given by part (iii) of *LEMMA* 1 and that of each denominator is given by part (i) and (ii) of *THEOREM* 2. ■

The analysis of $\hat{\eta}_\tau$ is very similar. We just need the generalization of *LEMMA* 1 for the case of the residuals from OLS of y_t on constant and t, t = 1, 2, ..., T.

LEMMA 3: Let e_t be the residuals from an OLS regression of y_t on (1, t), t = 1, 2, ..., T. Then, under the same assumptions as in *LEMMA* 1,

$$T^{-(d^*+3/2)}S_{[rT]} \Rightarrow \omega_{d^*}\int_0^r W_{d^*}^*(a)da,$$

where $W^{*}_{d*}(a) = W_{d*}(a) + (6a-4)\int_0^1 W_{d*}(b)db + (-12a+6)\int_0^1 bW_{d*}(b)db$.

Proof: See Appendix. ■

Given *LEMMA* 3, it is easy to establish the same asymptotic results for the KPSS $\hat{\eta}_\tau$ statistic under the nonstationary long memory process as are given for $\hat{\eta}_\mu$ in *THEOREM* 2 and 3. All that is necessary is to replace the demeaned fractional Brownian motion, $\underline{W_{d*}}(a)$, with the demeaned and detrended fractional Brownian motion, $W^{*}_{d*}(a)$, in *THEOREM* 2 and 3.

Those theorems have several interesting implications. First, even though the KPSS unit root test is consistent against stationary long memory alternatives, I(d) for -1/2< d <1/2, the KPSS unit root test is not consistent against nonstationary long memory alternatives, I(d) for 1/2< d <3/2, because the KPSS statistics have the same orders in probability under both the null and alternative hypothesis. This is the main theoretical result of this chapter. Second, Lee and Schmidt (1995) show that the KPSS short memory test is consistent against a stationary long memory process (-1/2 < d <1/2), and here we can now see that it is also consistent against a nonstationary long memory process (1/2< d <3/2). Below we will show higher power in finite samples against nonstationary long memory alternatives than against stationary long memory alternatives, as we should expect. Third, under the hypothesis of a non-stationary long memory process, the orders in probability of the KPSS statistics are independent of the value of d, even although the form of their asymptotic distributions are affected by the value of d. This is in contrast to the case of a stationary long memory process, where both the order in probability and the form of the asymptotic distribution depends on d. Also by way of contrast, the order in probability of the Dickey-Fuller statistics depends on d for $d < 1$ but not for $1 < d < 3/2$; see Sowell (1990).

3. SIMULATION RESULTS

In this section we provide simulation evidence on the power of the KPSS stationarity (short memory) and unit root tests. The computations are done in FORTRAN using the normal random number generator GASDEV/RAN3 of Press, Flannery, Teukolsky and Vetterling (1989), as in Lee and Schmidt (1995). The data on an I(d) process for $d < 1/2$ are generated using the Levinson algorithm [Levinson (1947), Durbin (1960), Whittle (1963), Brockwell and Davis (1991)]. For a nonstationary long memory process ($1/2 < d < 3/2$) the data are generated by cumulating I(d*) random variates, where $d^* = d-1$. Given the I(d) process ε_t, data on the observable series y_t are generated according to equation (2) with $\phi = \xi = 0$. The value of ϕ and ξ do not matter for the power of any of the tests that we consider, except that the $\hat{\eta}_\mu$ test requires $\xi = 0$.

We have considered only positive values of d because we are primarily interested in testing unit roots against nonstationary long memory processes. The lag truncation parameters are chosen as $\ell 0 = 0$, $\ell 4 =$ integer $[4(T/100)^{1/4}]$, and $\ell 12 =$ integer $[12(T/100)^{1/4}]$ as in Schwert (1989), KPSS (1992) and Lee and Schmidt

(1995). We consider sample sizes 50, 150, 250, 500 and 1000, and the number of iterations is 10000. All of our tests are based on the 5% significance level.

Table 2.1 gives the powers of the 5% upper tail KPSS short memory tests against the alternatives d =0.0, .1, .2, ..., .9, 1.0, and also d =.45 and .499. These are similar to the values considered by Lee and Schmidt (1995), except that we add some cases with d >1. Where they overlap, our results are very similar to those of Lee and Schmidt. There are no surprise in these results, so we will not discuss them in detail. Power increases with T for fixed d or with d for fixed T.

Table 2.2-1 gives the power of the lower-tail KPSS unit root test against I(d) for $0 \leq d < 1/2$, and Table 2.3-1 does the same for the two-tailed KPSS unit root test. Basically these results are as we would expect from our asymptotics and from the previous limited simulations of Shin and Schmidt (1992). (i) The lower tail tests are more powerful than the corresponding two-tailed tests. (ii) For a given d, power increases with T. This reflects the consistency of the tests against stationary long memory alternatives. (iii) Power is largest when $\ell = 0$ and smallest when $\ell = \ell 12$. (There are a few exceptions, for small values of T, due to large size distortions.) Again, this is consistent with the relevant asymptotics, which indicate that power depends on ℓ / T, even asymptotically, for $d < 1/2$. (iv) Power is larger when d is farther from unity. (v) The power of $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ are similar.

Table 2.2-2 gives the power of the lower tail KPSS unit root test against I(d) processes with $1/2 \leq d < 3/2$, while Table 2.3-2 does the same for the two tailed test. The most important result is that, with d fixed, power does not approach one as T increases. This is a reflection of our theoretical result that the KPSS unit root test is not consistent against nonstationary long memory processes. For example, for d =.7 and ℓ = 0, and for the one-tailed test, power grows from .169 with T = 50 to only .256 with T = 1000, and would not be expected to approach one even for arbitrarily large values of T.

Some other results in Tables 2.2-2 and 2.3-2 are as follows. (i) For the lower tail test, power is always lower when ℓ is larger, and is very small for $\ell 12$ for $T \leq 250$. This is due to the large size distortion of the test (too few rejections) for $T \leq 250$. For example, the size of the lower tail test based on $\hat{\eta}_\mu(\ell 12)$ is zero for T = 50, and still only .026 for T = 250. However, for the two tail test, power first decreases as the number of lags grows ($\ell 4$), and then increases with more lags ($\ell 12$). Again, this is due to large size distortions in the two tail test. (ii) The lower tail $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests have similar powers against I(d) for any d in the range (1/2, 3/2). However, the two tail tests have similar powers only against I(d) with d <1. If d is greater than 1 there is a distinct difference in the powers of two statistics. The $\hat{\eta}_\mu$ test is generally more powerful. (iii) For a given sample size and number of lags, power increases monotonically with $|1-d|$ for both the lower and two tail tests for d <1, and the lower tail test has little power against d >1. The power of two tail test is asymmetric around d =1, as in Diebold and Rudebusch (1991) and Lee (1994).

Lastly, our results can be compared with the previous results of other authors for Dickey-Fuller type unit root tests against I(d) alternatives. These comparisons are given in Table 2.4. Diebold and Rudebusch (1991) show that Dickey-Fuller test is not very powerful against I(d) alternatives for d >0.6. This is despite the fact that Sowell has proved that the test is consistent against such alternatives. Lee (1994) gives the power of Dickey-Fuller tests against I(d), and finds it to be small for d >1/2; essentially, his results are the same as those of Diebold and Rudebusch. He also found a discontinuity of the power function of the simple (no constants and no trends) Dickey-Fuller tests at d =1/2. Hassler and Wolters (1994) show that the Phillips-Perron and Dickey-Fuller tests have similar power against I(d) processes but the augmented Dickey-Fuller test is not powerful; they argue that it is inconsistent. Our results are not directly comparable with the others because our data is demeaned while the others' are not. However, two statements seem correct. First, the KPSS unit root test has lower power than Dickey-Fuller tests (except the augmented Dickey-Fuller tests) against nonstationary long memory processes. Second, the unsurprising implication of all of these results is that it is very difficult to distinguish between a unit root and nonstationary long memory. This is unfortunate, because these two types of series differ in fundamental ways, notably their degree of mean reversion.

4. CONCLUSION

In this chapter we have asked whether the KPSS unit root test can be used to distinguish long memory processes from unit root processes. We have shown that the KPSS unit root test is consistent against stationary long memory alternatives, namely I(d) processes for $d \in (-1/2, 1/2)$; but it is not consistent against nonstationary long memory alternatives, I(d) for $d \in (1/2, 3/2)$. This implies that the KPSS statistic can not distinguish nonstationary long memory processes from unit root processes, even though it can consistently distinguish between short memory processes, <u>stationary</u> long memory processes, and nonstationary processes. Also, we have provided the simulation results on power in finite samples. These support the relevance of our asymptotic results.

Dickey-Fuller tests can consistently distinguish a unit root from an I(d) process with $1/2 < d < 1$, but not from an I(d) process with $1 < d < 3/2$; see Sowell (1990). Thus distinguishing a unit root from nonstationary long memory is a difficult and not completely solved problem that is worthy of further attention.

TABLE 2.1

Power of KPSS Short Memory Test against I(d), d∈[0.0, 1.5)

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=0.0												
$\ell0$	.044	.053	.050	.046	.052	.053	.051	.052	.052	.053	.052	.051
$\ell4$	.036	.047	.046	.044	.051	.051	.041	.045	.048	.053	.050	.050
$\ell12$	.012	.033	.039	.041	.049	.048	.044	.034	.040	.044	.048	.050
d=.1												
$\ell0$	.124	.165	.195	.216	.262	.323	.140	.185	.215	.265	.331	.409
$\ell4$	.072	.100	.119	.131	.169	.195	.071	.098	.122	.145	.196	.226
$\ell12$	.020	.055	.075	.085	.117	.141	.049	.055	.071	.092	.121	.152
d=.2												
$\ell0$	.240	.356	.402	.485	.592	.693	.267	.380	.461	.577	.722	.839
$\ell4$	.123	.182	.221	.265	.348	.418	.121	.166	.224	.280	.407	.488
$\ell12$	.032	.093	.121	.158	.211	.282	.060	.086	.113	.149	.218	.295
d=.3												
$\ell0$	.387	.542	.631	.730	.845	.928	.418	.608	.714	.829	.932	.983
$\ell4$	.193	.281	.351	.413	.555	.636	.168	.267	.361	.444	.625	.735
$\ell12$	.047	.139	.193	.251	.343	.425	.064	.115	.159	.226	.345	.473
d=.4												
$\ell0$	.543	.707	.801	.884	.958	.990	.574	.775	.870	.941	.991	.999
$\ell4$	.278	.384	.468	.549	.714	.797	.235	.362	.490	.600	.794	.877
$\ell12$	.072	.198	.255	.329	.455	.576	.072	.154	.219	.316	.476	.626
d=.45												
$\ell0$	.613	.773	.858	.930	.981	.997	.639	.837	.921	.972	.997	1.000
$\ell4$	.320	.431	.531	.621	.775	.850	.259	.417	.559	.661	.848	.919
$\ell12$	.088	.224	.297	.381	.515	.645	.077	.168	.257	.352	.536	.685
d=.49												
$\ell0$	.664	.833	.893	.951	.989	.999	.687	.880	.944	.986	.999	1.000
$\ell4$	.357	.466	.572	.669	.821	.887	.293	.467	.599	.717	.886	.945
$\ell12$	.101	.245	.324	.419	.556	.680	.078	.192	.267	.389	.578	.737
d=.499												
$\ell0$	.659	.826	.902	.959	.990	.999	.700	.885	.949	.988	.999	1.000
$\ell4$	.352	.479	.589	.663	.819	.850	.289	.470	.606	.721	.893	.947
$\ell12$	.095	.250	.336	.416	.559	.697	.083	.192	.275	.388	.597	.742
d=.5												
$\ell0$	.667	.837	.901	.960	.991	.999	.704	.888	.951	.984	.999	1.000
$\ell4$	.356	.476	.588	.680	.826	.898	.306	.465	.616	.721	.891	.951
$\ell12$	.101	.254	.333	.433	.570	.696	.088	.187	.278	.392	.593	.748

TABLE 2.1, CONTINUED

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=.51												
$\ell 0$	.688	.845	.910	.962	.992	.999	.711	.901	.954	.990	.999	1.000
$\ell 4$	.370	.493	.602	.690	.833	.905	.315	.483	.617	.723	.898	.953
$\ell 12$	.103	.268	.343	.439	.579	.713	.086	.196	.282	.410	.604	.752
d=.6												
$\ell 0$	.778	.908	.961	.987	.999	1.000	.807	.951	.983	.998	1.000	1.000
$\ell 4$	.437	.560	.689	.771	.900	.951	.379	.556	.709	.812	.949	.984
$\ell 12$	.134	.315	.414	.509	.661	.791	.791	.101	.234	.340	.484	.681
d=.7												
$\ell 0$	.848	.952	.981	.997	1.000	1.000	.872	.976	.995	1.000	1.000	1.000
$\ell 4$	.507	.643	.766	.835	.946	.975	.450	.645	.792	.882	.978	.994
$\ell 12$	.171	.384	.491	.583	.737	.855	.121	.283	.392	.564	.776	.897
d=.8												
$\ell 0$	.905	.978	.992	.999	1.000	1.000	.924	.992	.999	1.000	1.000	1.000
$\ell 4$	.586	.714	.823	.891	.967	.991	.511	.714	.849	.928	.990	.998
$\ell 12$	.229	.455	.553	.644	.806	.907	.131	.335	.461	.636	.825	.936
d=.9												
$\ell 0$	.935	.990	.996	1.000	1.000	1.000	.953	.997	1.000	1.000	1.000	1.000
$\ell 4$	.652	.776	.865	.928	.984	.995	.574	.771	.892	.950	.995	1.000
$\ell 12$	.286	.523	.609	.713	.861	.938	.150	.365	.514	.692	.876	.963
d=.95												
$\ell 0$	.946	.992	.998	1.000	1.000	1.000	.967	.998	1.000	1.000	1.000	1.000
$\ell 4$	.676	.800	.892	.942	.990	.997	.606	.810	.908	.961	.998	1.000
$\ell 12$	.310	.548	.636	.739	.884	.951	.169	.400	.545	.717	.896	.972
d=.99												
$\ell 0$	.960	.994	.999	1.000	1.000	1.000	.971	.999	1.000	1.000	1.000	1.000
$\ell 4$	.701	.823	.906	.948	.991	.998	.623	.814	.922	.969	.998	1.000
$\ell 12$	.341	.574	.666	.751	.892	.962	.173	.414	.561	.732	.905	.975
d=1.0												
$\ell 0$	.959	.994	.999	1.000	1.000	1.000	.973	.999	1.000	1.000	1.000	1.000
$\ell 4$	.712	.821	.913	.954	.992	.998	.628	.824	.925	.970	.998	1.000
$\ell 12$	.346	.583	.669	.766	.898	.960	.177	.421	.570	.746	.911	.977
d=1.1												
$\ell 0$	.975	.997	1.000	1.000	1.000	1.000	.987	.999	1.000	1.000	1.000	1.000
$\ell 4$	.759	.871	.934	.967	.995	.999	.685	.859	.946	.982	.999	1.000
$\ell 12$	.421	.648	.732	.808	.933	.974	.205	.451	.616	.783	.933	.985

TABLE 2.1, CONTINUED

	$\hat{\eta}_{\mu}$ Test						$\hat{\eta}_{\tau}$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=1.2												
$\ell 0$	.985	.999	1.000	1.000	1.000	1.000	.991	1.000	1.000	1.000	1.000	1.000
$\ell 4$	.810	.903	.960	.982	.999	1.000	.685	.859	.946	.982	.999	1.000
$\ell 12$	.504	.706	.784	.854	.952	.985	.205	.451	.616	.783	.933	.985
d=1.3												
$\ell 0$	.991	1.000	1.000	1.000	1.000	1.000	.996	1.000	1.000	1.000	1.000	1.000
$\ell 4$	.851	.932	.970	.988	.999	1.000	.768	.917	.971	.992	1.000	1.000
$\ell 12$	.598	.771	.828	.891	.967	.991	.265	.561	.707	.856	.968	.995
d=1.4												
$\ell 0$	.995	1.000	1.000	1.000	1.000	1.000	.996	1.000	1.000	1.000	1.000	1.000
$\ell 4$	.902	.958	.983	.994	1.000	1.000	.809	.939	.983	.995	1.000	1.000
$\ell 12$	.719	.846	.887	.931	.982	.995	.304	.598	.744	.881	.979	.997
d=1.45												
$\ell 0$	.997	1.000	1.000	1.000	1.000	1.000	.998	1.000	1.000	1.000	1.000	1.000
$\ell 4$	.933	.972	.991	.995	1.000	1.000	.809	.939	.983	.995	1.000	1.000
$\ell 12$	.805	.887	.925	.954	.989	.997	.304	.598	.744	.881	.979	.997
d=1.499												
$\ell 0$	1.000	1.000	1.000	1.000	1.000	1.000	.999	1.000	1.000	1.000	1.000	1.000
$\ell 4$	.991	.997	.999	1.000	1.000	1.000	.834	.948	.986	.996	1.000	1.000
$\ell 12$	.975	.988	.991	.993	.998	1.000	.353	.629	.774	.903	.984	.999

TABLE 2.2-1

Power of KPSS Lower Tail Unit Root Test against I(d), d∈[0.0, 1/2)

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=0.0												
ℓ0	.964	.996	1.000	1.000	1.000	1.000	.969	.999	1.000	1.000	1.000	1.000
ℓ4	.623	.799	.915	.964	.998	1.000	.286	.708	.910	.969	1.000	1.000
ℓ12	.000	.156	.395	.617	.851	.954	.000	.000	.013	.333	.815	.964
d=.1												
ℓ0	.890	.976	.993	.999	1.000	1.000	.903	.990	.999	1.000	1.000	1.000
ℓ4	.514	.695	.816	.886	.975	.989	.222	.585	.808	.898	.990	.997
ℓ12	.000	.114	.305	.504	.735	.865	.000	.000	.010	.248	.667	.878
d=.2												
ℓ0	.780	.896	.945	.977	.997	1.000	.790	.942	.977	.996	1.000	1.000
ℓ4	.425	.553	.694	.761	.897	.929	.170	.467	.682	.784	.942	.975
ℓ12	.000	.087	.234	.390	.602	.725	.000	.000	.010	.173	.514	.744
d=.3												
ℓ0	.640	.751	.804	.867	.934	.963	.656	.831	.890	.944	.982	.996
ℓ4	.338	.449	.558	.615	.748	.798	.124	.344	.537	.630	.824	.877
ℓ12	.000	.071	.176	.294	.459	.585	.000	.000	.006	.119	.386	.574
d=.4												
ℓ0	.486	.579	.637	.694	.757	.806	.499	.666	.730	.798	.877	.922
ℓ4	.260	.351	.435	.478	.591	.622	.091	.261	.406	.477	.658	.714
ℓ12	.000	.051	.127	.223	.346	.432	.000	.000	.005	.085	.263	.419
d=.45												
ℓ0	.412	.502	.537	.587	.647	.690	.435	.572	.617	.693	.777	.832
ℓ4	.226	.304	.372	.404	.509	.525	.083	.223	.343	.414	.569	.618
ℓ12	.000	.045	.111	.182	.298	.362	.000	.000	.004	.070	.225	.355
d=.49												
ℓ0	.361	.435	.474	.498	.554	.608	.380	.488	.551	.608	.680	.738
ℓ4	.201	.269	.334	.358	.448	.477	.073	.191	.307	.355	.502	.537
ℓ12	.000	.041	.102	.162	.260	.328	.000	.000	.004	.057	.191	.299
d=.499												
ℓ0	.365	.424	.451	.503	.548	.570	.374	.478	.533	.589	.661	.719
ℓ4	.205	.266	.319	.362	.443	.452	.069	.182	.297	.351	.481	.532
ℓ12	.000	.037	.095	.161	.256	.312	.000	.000	.004	.052	.186	.297

TABLE 2.2-2

Power of KPSS Lower Tail Unit Root Test against I(d), d∈[1/2, 3/2)

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=.5												
ℓ0	.358	.426	.450	.485	.535	.576	.360	.499	.528	.590	.647	.714
ℓ4	.197	.259	.327	.346	.430	.454	.073	.186	.291	.347	.481	.523
ℓ12	.000	.037	.097	.155	.252	.310	.000	.000	.004	.057	.182	.290
d=.51												
ℓ0	.338	.397	.433	.465	.515	.549	.350	.456	.510	.566	.634	.687
ℓ4	.189	.248	.310	.336	.420	.434	.068	.173	.286	.339	.469	.509
ℓ12	.000	.035	.090	.155	.243	.293	.000	.000	.004	.054	.180	.283
d=.6												
ℓ0	.243	.295	.298	.322	.347	.364	.250	.328	.351	.379	.430	.456
ℓ4	.147	.195	.230	.252	.309	.315	.046	.130	.209	.242	.344	.359
ℓ12	.000	.027	.063	.113	.180	.214	.000	.000	.003	.038	.125	.195
d=.7												
ℓ0	.169	.190	.193	.207	.213	.223	.169	.209	.212	.226	.240	.256
ℓ4	.112	.139	.166	.183	.212	.213	.035	.094	.138	.160	.220	.228
ℓ12	.000	.021	.048	.084	.121	.148	.000	.000	.002	.021	.076	.125
d=.8												
ℓ0	.108	.119	.126	.125	.129	.127	.107	.122	.130	.130	.144	.150
ℓ4	.080	.099	.124	.125	.145	.134	.026	.062	.100	.104	.151	.151
ℓ12	.000	.012	.037	.054	.082	.096	.000	.000	.001	.015	.052	.081
d=.9												
ℓ0	.072	.074	.082	.075	.078	.078	.068	.073	.076	.075	.078	.077
ℓ4	.059	.070	.088	.083	.093	.091	.018	.041	.066	.072	.095	.092
ℓ12	.000	.009	.025	.037	.055	.064	.000	.000	.001	.010	.035	.047
d=.95												
ℓ0	.063	.059	.061	.059	.058	.059	.049	.056	.058	.060	.055	.057
ℓ4	.053	.059	.069	.068	.072	.071	.013	.034	.055	.060	.073	.072
ℓ12	.000	.008	.021	.030	.042	.051	.000	.000	.000	.007	.025	.037
d=.99												
ℓ0	.048	.050	.049	.050	.049	.045	.043	.044	.044	.046	.048	.049
ℓ4	.043	.051	.059	.060	.066	.055	.013	.028	.045	.049	.067	.062
ℓ12	.000	.006	.017	.028	.038	.040	.000	.000	.001	.006	.025	.033
d=1.0												
ℓ0	.048	.049	.045	.044	.045	.048	.042	.040	.042	.044	.041	.044
ℓ4	.042	.049	.056	.055	.061	.059	.011	.026	.043	.047	.057	.056
ℓ12	.000	.006	.016	.026	.036	.042	.000	.000	.000	.007	.019	.030

TABLE 2.2-2, CONTINUED

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=1.1												
$\ell 0$	.030	.028	.029	.030	.028	.029	.022	.026	.026	.024	.025	.025
$\ell 4$	.029	.032	.038	.037	.039	.040	.006	.020	.030	.029	.040	.036
$\ell 12$	.000	.005	.010	.017	.023	.028	.000	.000	.000	.005	.013	.019
d=1.2												
$\ell 0$	.019	.017	.017	.015	.015	.017	.014	.015	.013	.013	.015	.013
$\ell 4$	.021	.022	.025	.021	.024	.023	.005	.013	.018	.019	.025	.021
$\ell 12$	.000	.003	.006	.010	.014	.017	.000	.000	.000	.003	.009	.011
d=1.3												
$\ell 0$	.011	.008	.012	.010	.010	.010	.008	.009	.010	.009	.008	.008
$\ell 4$	.013	.012	.019	.015	.016	.013	.004	.009	.015	.012	.015	.013
$\ell 12$	.000	.001	.006	.007	.009	.010	.000	.000	.000	.002	.005	.007
d=1.4												
$\ell 0$	.006	.005	.006	.004	.005	.005	.006	.004	.005	.005	.005	.005
$\ell 4$	.007	.007	.009	.008	.009	.008	.003	.005	.008	.009	.009	.008
$\ell 12$	.005	.001	.002	.003	.004	.005	.000	.000	.000	.001	.003	.004
d=1.45												
$\ell 0$	.004	.004	.003	.003	.003	.003	.004	.003	.004	.004	.005	.003
$\ell 4$	.006	.005	.005	.006	.005	.005	.002	.004	.006	.007	.009	.006
$\ell 12$	.000	.001	.001	.002	.003	.003	.000	.000	.000	.001	.003	.003
d=1.499												
$\ell 0$	.000	.000	.001	.000	.000	.000	.003	.003	.004	.004	.003	.003
$\ell 4$	.001	.000	.001	.001	.000	.001	.002	.004	.007	.006	.007	.004
$\ell 12$	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.002

TABLE 2.3-1

Power of KPSS Two Tail Unit Root Test against I(d), d∈[0.0, 1/2)

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=0.0												
$\ell 0$	.919	.989	.999	1.000	1.000	1.000	.932	.998	1.000	1.000	1.000	1.000
$\ell 4$	.456	.679	.843	.916	.991	.998	.119	.556	.833	.933	.998	1.000
$\ell 12$	.011	.034	.228	.469	.753	.903	.472	.006	.000	.169	.702	.922
d=.1												
$\ell 0$	.808	.940	.975	.997	1.000	1.000	.828	.974	.994	1.000	1.000	1.000
$\ell 4$	.358	.551	.709	.800	.938	.971	.090	.423	.700	.825	.972	.991
$\ell 12$	.017	.024	.164	.353	.608	.771	.464	.011	.004	.119	.530	.790
d=.2												
$\ell 0$	.672	.817	.887	.946	.985	.998	.689	.894	.951	.987	.998	1.000
$\ell 4$	.276	.421	.571	.647	.819	.862	.064	.317	.549	.676	.891	.943
$\ell 12$	.027	.020	.116	.256	.467	.610	.484	.019	.001	.082	.379	.632
d=.3												
$\ell 0$	.518	.649	.712	.788	.877	.932	.536	.738	.820	.901	.958	.988
$\ell 4$	.211	.323	.427	.493	.639	.706	.045	.220	.410	.505	.730	.800
$\ell 12$	.042	.017	.082	.176	.332	.452	.481	.028	.002	.052	.260	.442
d=.4												
$\ell 0$	.367	.468	.520	.586	.656	.719	.379	.553	.619	.707	.802	.870
$\ell 4$	.150	.239	.312	.359	.471	.508	.029	.158	.287	.352	.542	.606
$\ell 12$	.063	.013	.059	.125	.231	.321	.501	.045	.007	.034	1.67	.302
d=.45												
$\ell 0$	.301	.391	.426	.463	.537	.582	.320	.455	.507	.590	.680	.752
$\ell 4$	.130	.199	.260	.291	.396	.413	.027	.127	.237	.301	.454	.507
$\ell 12$	.019	.014	.049	.100	.194	.255	.504	.053	.010	.027	.136	.244
d=.49												
$\ell 0$	.254	.326	.360	.390	.446	.497	.271	.381	.436	.497	.579	.639
$\ell 4$	.108	.169	.231	.255	.335	.362	.025	.107	.202	.243	.391	.421
$\ell 12$	.089	.014	.042	.089	.161	.223	.510	.065	.014	.020	.114	.203
d=.499												
$\ell 0$	.257	.318	.335	.391	.434	.465	.262	.367	.423	.480	.547	.619
$\ell 4$	.112	.170	.216	.251	.327	.344	.022	.106	.197	.236	.366	.416
$\ell 12$	.084	.014	.040	.084	.163	.211	.506	.069	.013	.020	.108	.199

TABLE 2.3-2

Power of KPSS Two Tail Unit Root Test against I(d), d∈[1/2, 3/2)

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=.5												
ℓ0	.245	.311	.344	.372	.422	.466	.257	.364	.415	.480	.548	.613
ℓ4	.110	.163	.223	.239	.322	.343	.026	.097	.193	.243	.369	.414
ℓ12	.092	.013	.040	.081	.160	.212	.517	.066	.011	.022	.107	.191
d=.51												
ℓ0	.237	.291	.322	.358	.403	.439	.251	.347	.403	.459	.526	.582
ℓ4	.106	.154	.211	.232	.311	.320	.021	.098	.188	.236	.357	.397
ℓ12	.093	.015	.038	.082	.154	.197	.521	.072	.015	.022	.104	.191
d=.6												
ℓ0	.162	.199	.206	.228	.251	.267	.164	.233	.258	.283	.325	.354
ℓ4	.079	.115	.143	.163	.214	.217	.012	.065	.133	.157	.245	.265
ℓ12	.121	.019	.025	.054	.106	.138	.538	.091	.025	.016	.067	.122
d=.7												
ℓ0	.105	.121	.122	.135	.139	.146	.105	.138	.144	.153	.166	.183
ℓ4	.059	.077	.098	.110	.135	.139	.009	.046	.080	.095	.144	.155
ℓ12	.155	.029	.021	.038	.067	.091	.573	.122	.039	.020	.037	.072
d=.8												
ℓ0	.063	.073	.077	.073	.076	.077	.064	.076	.079	.078	.088	.097
ℓ4	.038	.052	.072	.070	.084	.082	.007	.028	.052	.057	.091	.093
ℓ12	.212	.055	.028	.027	.045	.054	.581	.156	.068	.027	.028	.046
d=.9												
ℓ0	.046	.047	.051	.047	.046	.049	.046	.050	.051	.053	.051	.051
ℓ4	.025	.035	.047	.043	.052	.049	.004	.017	.033	.038	.055	.048
ℓ12	.267	.088	.043	.030	.030	.033	.604	.191	.094	.048	.030	.035
d=.95												
ℓ0	.046	.042	.047	.045	.041	.044	.044	.047	.048	.048	.047	.046
ℓ4	.023	.027	.037	.034	.037	.037	.003	.013	.026	.030	.038	.038
ℓ12	.291	.111	.053	.035	.024	.030	.617	.219	.110	.060	.036	.034
d=.99												
ℓ0	.050	.050	.051	.049	.046	.044	.047	.047	.048	.047	.050	.049
ℓ4	.016	.023	.030	.030	.032	.027	.002	.013	.021	.022	.035	.033
ℓ12	.324	.135	.072	.043	.023	.025	.622	.233	.127	.074	.040	.038
d=1.0												
ℓ0	.049	.053	.051	.049	.049	.047	.047	.047	.045	.051	.047	.053
ℓ4	.017	.024	.026	.027	.032	.029	.002	.009	.018	.022	.029	.030
ℓ12	.325	.144	.076	.042	.025	.027	.625	.239	.133	.076	.043	.043

TABLE 2.3-2, CONTINUED

	$\hat{\eta}_\mu$ Test						$\hat{\eta}_\tau$ Test					
T	50	100	150	250	500	1000	50	100	150	250	500	1000
d=1.1												
$\ell0$	.082	.086	.084	.086	.085	.081	.066	.071	.070	.072	.071	.070
$\ell4$	.012	.014	.017	.017	.019	.018	.001	.008	.013	.014	.018	.020
$\ell12$	.403	.213	.129	.083	.047	.042	.661	.275	.175	.113	.066	.059
d=1.2												
$\ell0$	.146	.154	.150	.150	.146	.147	.100	.104	.105	.098	.103	.102
$\ell4$	.006	.009	.011	.010	.010	.011	.001	.005	.007	.007	.011	.014
$\ell12$	.487	.304	.213	.148	.086	.081	.684	.332	.217	.143	.096	.086
d=1.3												
$\ell0$	.249	.248	.252	.248	.250	.253	.141	.147	.145	.141	.143	.149
$\ell4$	.003	.003	.008	.005	.006	.005	.001	.003	.006	.006	.006	.019
$\ell12$	.582	.415	.321	.251	.175	.166	.693	.381	.261	.185	.132	.125
d=1.4												
$\ell0$	.409	.417	.417	.413	.414	.414	.190	.197	.193	.195	.190	.191
$\ell4$	.003	.002	.003	.002	.003	.003	.001	.002	.003	.006	.004	.030
$\ell12$	.708	.574	.500	.417	.332	.316	.722	.422	.311	.239	.179	.162
d=1.45												
$\ell0$	.556	.551	.554	.560	.553	.561	.213	.212	.215	.218	.218	.219
$\ell4$	.002	.002	.001	.002	.002	.002	.001	.002	.002	.006	.004	.036
$\ell12$	.796	.698	.628	.573	.477	.462	.739	.434	.327	.259	.204	.193
d=1.499												
$\ell0$	.936	.931	.928	.932	.931	.934	.242	.239	.238	.237	.234	.241
$\ell4$	.000	.000	.000	.000	.000	.000	.000	.001	.003	.000	.003	.044
$\ell12$	.973	.958	.942	.934	.917	.913	.747	.465	.348	.283	.222	.210

TABLE 2.4

Power Comparison with Dickey-Fuller Type Tests

		(1)		(2)		(3)			(4)Low $\hat{\eta}_\mu$		(5)Two $\hat{\eta}_\mu$	
d	T	τ	ρ	τ	ρ	τ_0	ADF	PP	$\ell 0$	$\ell 12$	$\ell 0$	$\ell 12$
d=.45	100	.88	.86	.88	.87	.999	.111	.999	.502	.045	.391	.014
	250	.99	.99	.99	.99	1.00	.336	1.00	.587	.182	.463	.100
d=.6	100	.71	.71	.71	.71	.927	.069	.926	.295	.027	.199	.019
	250	.90	.90	.90	.90	.999	.186	.996	.322	.113	.228	.054
d=.9	100	.10	.10	.09	.09	.138	.038	.136	.074	.009	.047	.088
	250	.14	.14	.13	.14	.199	.060	.173	.075	.037	.047	.030
d=1.0	100	.05	.05	.04	.04	.051	.045	.053	.049	.006	.053	.144
	250	.06	.05	.05	.05	.046	.045	.050	.044	.026	.049	.042
d=1.3	100	.54	.22	.54	.21	--	--	--	.008	.010	.248	.415
	250	.62	.25	.62	.25	--	--	--	.001	.007	.248	.251

(1) Lee's dissertation (1994): two tail D-F test (5%)
(2) Diebold and Rudebusch (1991): two tail D-F test (5%)
(3) Hassler and Wolters (1994): one-sided (5%)
τ_0 = t-type simple D-F test
ADF = t-type Augmented D-F test with lags ($\ell 12$)
PP = t-type Phillips-Perron test with lags ($\ell 12$)
(4) KPSS unit root test: lower tail (5%)
(5) KPSS unit root test: two tail (5%)

APPENDIX

Proof of *LEMMA* 1: Using equation (11),

$$\frac{1}{T^{d^*+3/2}}\sum_{t=1}^{[rT]}\varepsilon_t = \frac{1}{T}\sum_{t=1}^{[rT]}\left(\frac{\varepsilon_t}{T^{d^*+1/2}}\right) \Rightarrow \omega_{d^*}\int_0^r W_{d^*}(a)da,$$

$$\frac{1}{T^{d^*+3/2}}S_{[rT]} = \frac{1}{T}\sum_{t=1}^{[rT]}\frac{1}{T^{d^*+1/2}}(\varepsilon_t - \bar{\varepsilon}) = \frac{1}{T}\sum_{t=1}^{[rT]}\left(\frac{\varepsilon_t}{T^{d^*+1/2}}\right) - \frac{[rT]}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t}{T^{d^*+1/2}}\right)$$

$$\Rightarrow \omega_{d^*}\left[\int_0^r W_{d^*}(a)da - r\int_0^1 W_{d^*}(a)da\right] = \omega_{d^*}\int_0^r\left[W_{d^*}(a) - \int_0^1 W_d{}^*(b)db\right]da,$$

which proves parts (i) and (ii). For part (iii),

$$\frac{1}{T^{2d^*+4}}\sum_{t=1}^{T}S_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{S_t}{T^{d^*+3/2}}\right]^2 \Rightarrow \omega_{d^*}^2\int_0^1\left[\int_0^r \underline{W_{d^*}(a)da}\right]^2 dr, \text{ by (ii) and the}$$

continuous mapping theorem. ■

Proof of *THEOREM* 2: For the proof we use the following Lemmas. In each, we make the same assumptions as in *LEMMA* 1 of the main text, and results are as $T\rightarrow\infty$.

LEMMA 2.1:

$$\frac{1}{T^{2+2d^*}}\sum_{t=1}^{T}\varepsilon_{t-1}\Delta\varepsilon_t \xrightarrow{p} 0.$$

Proof: Since $\sum_{t=1}^{T}\varepsilon_{t-1}\Delta\varepsilon_t = \frac{1}{2}(\varepsilon_T^2 - \varepsilon_0^2 - \sum_{t=1}^{T}(\Delta\varepsilon_t)^2)$

$$\frac{1}{T^{2+2d^*}}\sum_{t=1}^{T}\varepsilon_{t-1}\Delta\varepsilon_t = \frac{1}{2}\frac{1}{T}\left[\frac{\varepsilon_T}{T^{d^*+1/2}}\right]^2 - \frac{1}{2}\frac{1}{T}\left[\frac{\varepsilon_0}{T^{d^*+1/2}}\right]^2 - \frac{1}{2}\frac{1}{T^{1+2d^*}}\left[\frac{1}{T}\sum_{t=1}^{T}(\Delta\varepsilon_t)^2\right]$$

$$\xrightarrow{p} 0$$

becau

where

Proc

sinc

bec

usin

(Sov

because $\left[\frac{\varepsilon_T}{T^{d^*+1/2}}\right]^2 \Rightarrow \omega_{d^*}^2 W_{d^*}(1)^2$, $\left[\frac{1}{T}\sum_{t=1}^{T}(\Delta\varepsilon_t)^2\right] \xrightarrow{p} \gamma_0$ and ε_0 is $O_p(1)$,

where γ_j is the j-th autocovariance of $\Delta\varepsilon_t$. ∎

LEMMA 2.2:

$$\frac{1}{T^{2+2d^*}}\sum_{t=1}^{T}\varepsilon_{t-1}\Delta\varepsilon_{t+s} \xrightarrow{p} 0 \text{ for any nonnegative integer s.}$$

Proof: *LEMMA* 2.1 is the special case of s =0. For any given positive integer s,

$$\sum_{t=1}^{T}\varepsilon_{t-1}\Delta\varepsilon_{t+s} = \sum_{t=1}^{T}\varepsilon_{t+s-1}\Delta\varepsilon_{t+s} - \sum_{j=0}^{s-1}\sum_{t=1}^{T}\Delta\varepsilon_{t+j}\Delta\varepsilon_{t+s},$$

since $\varepsilon_{t-1} = \varepsilon_{t+s-1} - (\Delta\varepsilon_{t+s-1} + \Delta\varepsilon_{t+s-2} + \ldots\ldots + \Delta\varepsilon_t)$. Then,

$$\frac{1}{T^{2+2d^*}}\sum_{t=1}^{T}(\varepsilon_{t-1}\Delta\varepsilon_{t+s}) = \frac{1}{T^{2+2d^*}}\sum_{t=1}^{T}(\varepsilon_{t+s-1}\Delta\varepsilon_{t+s}) - \frac{1}{T^{1+2d^*}}\sum_{j=0}^{s-1}(\frac{1}{T}\sum_{t=1}^{T}\Delta\varepsilon_{t+j}\Delta\varepsilon_{t+s})$$

$$\xrightarrow{p} 0,$$

because $\frac{1}{T^{2+2d^*}}\sum_{t=1}^{T}\varepsilon_{t+s-1}\Delta\varepsilon_{t+s} \xrightarrow{p} 0$ by *LEMMA* 2.1 and

$$\frac{1}{T^{1+2d^*}}\sum_{j=0}^{s-1}(\frac{1}{T}\sum_{t=1}^{T}\Delta\varepsilon_{t+j}\Delta\varepsilon_{t+s}) \xrightarrow{p} \frac{1}{T^{1+2d^*}}\left(\sum_{j=0}^{s-1}\gamma_{s-j}\right) \to 0,$$

using $\sum_{j=0}^{s-1}\gamma_{s-j} = \sum_{j=0}^{s}\gamma_j = \left[\frac{\Gamma(1-2d^*)}{\Gamma(d^*)\Gamma(1-d^*)}\sigma_u^2\right]\sum_{j=1}^{s}\left(\frac{\Gamma(d^*+j)}{\Gamma(1-d^*+j)}\right)$

$$= \left[\frac{\Gamma(1-2d^*)}{\Gamma(d^*)\Gamma(1-d^*)}\sigma_u^2\right]\left\{\frac{1}{2d^*}\left[\frac{\Gamma(1+d^*+s)}{\Gamma(1-d^*+s)} - \frac{\Gamma(1+d^*)}{\Gamma(1-d^*)}\right]\right\},$$

(Sowell, 1990). ∎

LEMMA 2.3:

$$\frac{1}{T^{2+2d^*}}\sum_{j=0}^{s-1}\sum_{t=s+1}^{T}\varepsilon_{t-s}\Delta\varepsilon_{t-j}\xrightarrow{p}0 \text{ for any nonnegaitve integer s.}$$

Proof: $\frac{1}{T^{2+2d^*}}\sum_{j=0}^{s-1}\sum_{t=s+1}^{T}\varepsilon_{t-s}\Delta\varepsilon_{t-j}=\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\left[\varepsilon_{t-s}\Delta\varepsilon_{t}+\varepsilon_{t-s}\Delta\varepsilon_{t-1}+\dots\dots+\varepsilon_{t-s}\Delta\varepsilon_{t-s+1}\right].$

Then, $\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\left[\varepsilon_{t-s}\Delta\varepsilon_{t}+\varepsilon_{t-s}\Delta\varepsilon_{t-1}+\dots\dots+\varepsilon_{t-s}\Delta\varepsilon_{t-s+2}\right]\xrightarrow{p}0$ by *LEMMA* 2.2 and

$\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\left[\varepsilon_{t-s}\Delta\varepsilon_{t-s+1}\right]\xrightarrow{p}0$ by *LEMMA* 2.1. ■

LEMMA 2.4:

$$\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\varepsilon_t\varepsilon_{t-s}\Rightarrow\omega_{d^*}^2\int_0^1 W_{d^*}(a)^2\,da.$$

Proof: $\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\varepsilon_t\varepsilon_{t-s}=\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\varepsilon_{t-s}^2+\frac{1}{T^{2+2d^*}}\sum_{j=0}^{s-1}\sum_{t=s+1}^{T}\varepsilon_{t-s}\Delta\varepsilon_{t-j}$,

because $\varepsilon_t=\varepsilon_{t-s}-(\Delta\varepsilon_{t-s}+\Delta\varepsilon_{t-s+1}+\dots\dots+\Delta\varepsilon_t)$. Also,

$\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\varepsilon_{t-s}^2=\frac{1}{T}\sum_{t=s+1}^{T}\left(\frac{\varepsilon_{t-s}}{T^{d^*+1/2}}\right)^2\Rightarrow\omega_{d^*}^2\int_0^1 W_{d^*}(a)^2\,da$ and

$\frac{1}{T^{2+2d^*}}\sum_{j=0}^{s-1}\sum_{t=s+1}^{T}\varepsilon_{t-s}\Delta\varepsilon_{t-j}\xrightarrow{p}0$ by *LEMMA* 2.3. ■

LEMMA 2.5:

For any nonnegative integer ℓ,

$$\frac{1}{\ell}\left[1+2\sum_{s=1}^{\ell}(1-\frac{s}{\ell+1})\right]=\frac{\ell+1}{\ell}.$$

Proof:

$$\frac{1}{\ell(\ell+1)}[(\ell+1)+2(\ell+(\ell-1)+......+2+1)]=\frac{1}{\ell(\ell+1)}\left[(\ell+1)+2\frac{1}{2}\ell(\ell+1)\right]=\frac{\ell+1}{\ell}. \qquad \blacksquare$$

We now will use Lemmas 2.1-2.5 to prove Theorem 2. First consider part (i), with $\ell=0$. Then,

$$s^2(0)=\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^2=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2-\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\right)^2 \text{ and}$$

$$\frac{1}{T^{2d^*+1}}s^2(0)=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t}{T^{d^*+1/2}}\right)^2-\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)^2$$

$$\Rightarrow \omega_{d^*}^2\left\{\int_0^1 W_{d^*}(a)^2da-\left[\int_0^1 W_{d^*}(b)db\right]^2\right\}$$

$$=\omega_{d^*}^2\left\{\int_0^1 \underline{W_{d^*}}(a)^2da\right\}$$

since $\int_0^1 \underline{W_{d^*}}(a)^2da=\int_0^1\left[W_{d^*}(a)-\int_0^1 W_{d^*}(b)db\right]^2da=\int_0^1 W_{d^*}(a)^2da-\left[\int_0^1 W_{d^*}(b)db\right]^2$.

For proving part (ii), from (3) and $e_t=\varepsilon_t-\bar{\varepsilon}$,

$$s^2(\ell)=\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^2+\frac{2}{T}\sum_{s=1}^{\ell}w(s,\ell)\sum_{t=s+1}^{T}(\varepsilon_t-\bar{\varepsilon})(\varepsilon_{t-s}-\bar{\varepsilon})$$

$$= \left\{ \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\right)^2 \right\} + 2\sum_{s=1}^{\ell} w(s,\ell)\left\{ \frac{1}{T}\sum_{t=s+1}^{T}(\varepsilon_t\varepsilon_{t-s} - \varepsilon_t\bar{\varepsilon} - \bar{\varepsilon}\varepsilon_{t-s} - \bar{\varepsilon}^2) \right\}.$$

Thus,

$$\frac{s^2(\ell)}{T^{1+2d^*}} = \left\{ \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t}{T^{d^*+1/2}}\right)^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)^2 \right\} + 2\sum_{s=1}^{\ell} w(s,\ell) \times$$

$$\left\{ \sum_{t=s+1}^{T}\left[\left(\frac{\varepsilon_t\varepsilon_{t-s}}{T^{2+2d^*}}\right) - \left(\frac{\varepsilon_t\bar{\varepsilon}}{T^{2+2d^*}}\right) - \left(\frac{\bar{\varepsilon}\varepsilon_{t-s}}{T^{2+2d^*}}\right) + \left(\frac{\bar{\varepsilon}^2}{T^{2+2d^*}}\right) \right] \right\}$$

$$= (A) + 2\sum_{s=1}^{\ell} w(s,\ell) \times (B).$$

Part (A) is the same as $s^2(0)$ above. For fixed ℓ, (B) equals

$$\left(\sum_{t=s+1}^{T}\frac{\varepsilon_t\varepsilon_{t-s}}{T^{2+2d^*}}\right) - \frac{1}{T}\left(\sum_{t=s+1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)\left(\frac{\bar{\varepsilon}}{T^{d^*+1/2}}\right) - \frac{1}{T}\left(\sum_{t=s+1}^{T}\frac{\varepsilon_{t-s}}{T^{d^*+1/2}}\right)\left(\frac{\bar{\varepsilon}}{T^{d^*+1/2}}\right) + \left(\frac{\bar{\varepsilon}}{T^{d^*+1/2}}\right)^2.$$

From *LEMMA* 2.4, $\frac{1}{T^{2+2d^*}}\sum_{t=s+1}^{T}\varepsilon_t\varepsilon_{t-s} \Rightarrow \omega_{d^*}^2\int_0^1 W_{d^*}(a)^2\,da$,

$$\frac{1}{T}\left(\sum_{t=s+1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)\left(\frac{\bar{\varepsilon}}{T^{d^*+1/2}}\right) = \left(\frac{1}{T}\sum_{t=s+1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)\left(\frac{1}{T}\sum_{t=s+1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)$$

$$\Rightarrow \omega_{d^*}^2\left(\int_0^1 W_{d^*}(b)\,db\right)^2,$$

$$\frac{1}{T}\left(\sum_{t=s+1}^{T}\frac{\varepsilon_{t-s}}{T^{d^*+1/2}}\right)\left(\frac{\bar{\varepsilon}}{T^{d^*+1/2}}\right) = \left(\frac{1}{T}\sum_{t=s+1}^{T}\frac{\varepsilon_{t-s}}{T^{d^*+1/2}}\right)\left(\frac{1}{T}\sum_{t=s+1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right)$$

$$\Rightarrow \omega_{d^*}^2\left(\int_0^1 W_{d^*}(b)\,db\right)^2,$$

and $\left(\frac{\bar{\varepsilon}}{T^{d^*+1/2}}\right)^2 = \left(\frac{1}{T}\sum_{t=s+1}^{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right) \Rightarrow \omega_{d^*}^2\left(\int_0^1 W_{d^*}(b)db\right)^2$.

Therefore, (B) $\Rightarrow \omega_{d^*}^2\int_0^1\left\{W_{d^*}(a)^2 - \left[\int_0^1 W_{d^*}(b)db\right]^2\right\}da$.

Since $\ell \to \infty$ and $(\ell/T) \to 0$, as $T \to \infty$,

$$\frac{s^2(\ell)}{\ell T^{1+2d^*}} \Rightarrow \omega_{d^*}^2\int_0^1\left\{W_{d^*}(a)^2 - \left[\int_0^1 W_{d^*}(b)db\right]^2\right\}da \text{, by the above argument and}$$

LEMMA 2.5. ■

Proof of *LEMMA* 3: Let $\hat{\phi}$ and $\hat{\xi}$ be the coefficients of intercept and trend in the OLS regression of y_t on $(1, t)$. Then

$$\begin{bmatrix}\hat{\phi}-\phi\\ \hat{\xi}-\xi\end{bmatrix} = \begin{bmatrix}T & \Sigma t\\ \Sigma t & \Sigma t^2\end{bmatrix}^{-1}\begin{bmatrix}\Sigma\varepsilon_t\\ \Sigma t\varepsilon_t\end{bmatrix}$$

and by the same algebra as in Lee and Schmidt (1994) we can show the following:

$$(\hat{\phi}-\phi) = \left(\frac{4}{T}+\frac{2}{T^2}\right)\sum_{t=1}^{T}\varepsilon_t - \frac{6}{T^2}\sum_{t=1}^{T}t\varepsilon_t + o_p(1) \text{, then}$$

$$\frac{(\hat{\phi}-\phi)}{T^{d^*+1/2}} = \left(4+\frac{2}{T}\right)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t}{T^{d^*+1/2}}\right) - \frac{6}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right) + \frac{o_p(1)}{T^{d^*+1/2}}$$

$$\Rightarrow \omega_{d^*}\left\{4\int_0^1 W_{d^*}(a)da - 6\int_0^1 aW_{d^*}(a)da\right\}$$

$$= \omega_{d^*}\left\{-2\int_0^1 W_{d^*}(a)da + 6\int_0^1\left(\int_0^a W_{d^*}(b)db\right)da\right\},$$

since $\int_0^1 aW_{d^*}(a)da = \int_0^1 W_{d^*}(a)da - \int_0^1\left(\int_0^a W_{d^*}(b)db\right)da$.

Also, $(\hat{\xi}-\xi) = \frac{-6}{T^2}\sum_{t=1}^{T}\varepsilon_t + \frac{12}{T^3}\sum_{t=1}^{T}t\varepsilon_t + o_p(1)$,

$$\frac{(\hat{\xi}-\xi)}{T^{d^*-1/2}} = -6\frac{1}{T}\left(\frac{\sum_{t=1}^{T}\varepsilon_t}{T^{d^*+1/2}}\right) + \frac{12}{T}\left(\sum_{t=1}^{T}\frac{t}{T}\frac{\varepsilon_t}{T^{d^*+1/2}}\right) + \frac{o_p(1)}{T^{d^*-1/2}}$$

$$\Rightarrow \omega_{d^*}\left\{-6\int_0^1 W_{d^*}(a)da + 12\int_0^1 aW_{d^*}(a)da\right\}$$

$$= \omega_{d^*}\left\{6\int_0^1 W_{d^*}(a)da - 12\int_0^1\left(\int_0^a W_{d^*}(b)db\right)\right\}.$$

Then $S_{[rT]} = \sum_{t=1}^{[rT]}(\varepsilon_t - \hat{\varepsilon}_t) = \sum_{t=1}^{[rT]}\varepsilon_t - [rT](\hat{\phi}-\phi) - \frac{1}{2}[rT]([rT]+1)(\hat{\xi}-\xi)$,

$$\frac{S_{[rT]}}{T^{d^*+3/2}} = \frac{1}{T}\sum_{t=1}^{[rT]}\frac{\varepsilon_t}{T^{d^*+1/2}} - \frac{[rT]}{T}\left(\frac{\hat{\phi}-\phi}{T^{d^*+1/2}}\right) - \frac{1}{2}\frac{[rT]}{T}\frac{([rT]+1)}{T}\left(\frac{\hat{\xi}-\xi}{T^{d^*-1/2}}\right)$$

$$\Rightarrow \omega_{d^*}\left\{\int_0^r W_{d^*}(a)da - r\left(4\int_0^1 W_{d^*}(a)da - 6\int_0^1 aW_{d^*}(a)da\right)\right.$$

$$\left. - \frac{1}{2}r^2\left(-6\int_0^1 W_{d^*}(a)da + 12\int_0^1 aW_{d^*}(a)da\right)\right\}$$

$$= \omega_{d^*} \int_0^r W_{d^*}^*(a)da \,,$$

which is similar to the result of Shin and Schmidt (1992, p.388). ■

CHAPTER 3

ASYMPTOTIC BIAS OF THE MDE WHEN SHORT-RUN DYNAMICS ARE IGNORED

1. INTRODUCTION

Suppose that an observed series $\{y_t\}$ follows an ARFIMA(p,d,q) process:

$$(1) \qquad (1-L)^d y_t = \varepsilon_t, \quad \phi(L)\varepsilon_t = \theta(L)u_t,$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ...... - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$, all of the roots of $\phi(L)$ and $\theta(L)$ lie outside the unit circle, $\phi(L)$ and $\theta(L)$ have no common roots, and $\{u_t\}$ is white noise. This is the same model and the same notation as in Chapter 1. When ε_t itself is white noise, y_t is a fractionally integrated white noise, or ARFIMA(0,d,0), process, also called an I(d) process. In the ARFIMA model, the differencing parameter d determines the long run properties of the series, such as its persistence and the persistence of its autocorrelations, while θ and ϕ influence short-run dynamics. In this chapter we will consider the estimation of d, with particular attention to whether we can estimate d separately from the ARMA parameters that determine short-run dynamics.

The first systematic treatment of estimation of d was by Geweke and Porter-Hudak (1983), hereafter GPH, who suggested a simple semi-parametric two step procedure for estimating d. Their estimator is based on a spectral regression and linear filter theory. If $\{y_t\}$ follows the ARFIMA process (1), its spectral density is:

$$(2) \qquad f_y(\omega) = \frac{\sigma^2}{2\pi}\left|1 - e^{-i\omega}\right|^{-2d} f_\varepsilon(\omega) = \frac{\sigma^2}{2\pi}\left\{4\sin^2\left(\frac{\omega}{2}\right)\right\}^{-d} f_\varepsilon(\omega),$$

where $f_\varepsilon(\omega)$ is the spectral density of ε_t, which is finite, bounded away from zero and continuous on the interval $[-\pi, \pi]$. Taking logarithms,

$$(3) \qquad \log\left\{f_y(\omega)\right\} = \log\left\{\frac{\sigma^2}{2\pi} f_\varepsilon(0)\right\} - d\log\left\{4\sin^2(\frac{\omega}{2})\right\} + \log\left\{\frac{f_\varepsilon(\omega)}{f_\varepsilon(0)}\right\}.$$

Then GPH suggested an OLS regression based on $I(\omega_j)$, which denotes the periodogram at the harmonic ordinate, $\omega_j = 2\pi j/T$ for j = 1, ..., m, where T is the sample size. The number of ordinates used is m = g(T), where $g(T) \to \infty$ but $g(T)/T \to 0$ as $T \to \infty$. The regression model is:

$$(4) \qquad \log\left\{I(\omega_j)\right\} = \alpha - d\log\left\{4\sin^2(\frac{\omega_j}{2})\right\} + v_j,$$

where α = constant and

$$v_j = \log\left\{\frac{I(\omega_j)}{f_y(\omega_j)}\right\}.$$

This implies that $E(v_j) = 0$, $var(v_j) = \pi^2/6$, j = 1, 2, ..., m, and that $cov(v_i, v_j) = 0$, $i \neq j$. The GPH estimator, say $\hat{d}$, is then defined as the OLS estimator of d in (4). GPH show that $\hat{d}$ is consistent, for d <0, and Robinson (1990) shows consistency for $0 \leq d < 1/2$. Under the further condition $\lim_{T\to\infty}\left\{(\log(T)^2)/g(T)\right\} = 0$, $\hat{d}$ is asymptotically normal. However it is not $\sqrt{T}$-consistent; asymptotically its variance is of order m^{-1}, not T^{-1}. The important features of the GPH estimator is that we can disregard the last term in (3), which involves the unknown short-run dynamics parameters (ϕ's and θ's), because it is asymptotically nearly constant for sufficiently low frequencies.

However, the GPH estimator can be badly biased even for moderately large sample sizes, especially when there is substantial autocorrelation in the ε_t process.

Agiakloglou, Newbold and Wohar (1993) show that the GPH estimator of d under AR(1) or MA(1) errors with quite large short-run dynamics, is seriously biased when $m = T^{1/2}$. They conclude that tests based on the GPH estimator are significantly misleading and we may need joint estimation of d and short-run dynamics. A further difficulty is that the sampling distribution of the estimated ARMA parameters after d is replaced with the GPH estimator is currently unknown.

Maximum Likelihood Estimation (MLE) under the assumption of normality has been suggested by a number of authors, apparently starting with Hosking (1984). Sowell (1992) suggested the exact MLE of the general ARFIMA(p,d,q) process with normal disturbances. This estimator maximizes the log likelihood function:

$$(5) \qquad \log L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}Y'\Omega^{-1}Y,$$

where $\Omega_{ij} = \gamma_{|i-j|}$ (with γ_j = j-th order autocovariance of $\{y_t\}$), and Y is the Tx1 vector of observations. In the MLE any ARMA parameters in $\theta(L)$ and $\phi(L)$ must be estimated jointly with d. The exact MLE is computationally difficult, and probably not feasible for sample sizes larger than 1000 or so, because of the need to invert the TxT covariance matrix Ω. Several approximate MLEs which do not require the inversion of the covariance matrix Ω have been suggested. A conditional sum-of-squares estimator (CSS) was proposed by Li and McLeod (1986). It truncates the infinite sum in $(1-L)^d$ to a finite sum for estimation. Chung and Baillie (1993) investigated the small sample performance of the CSS estimator and showed that for the I(d) process the CSS estimator is very close to Sowell's exact MLE, even for

T<100. They show that the estimation of the mean can make a considerable difference to the small sample bias. Fox and Taqqu (1986) and Dalhaus (1989) used an approximation formula for the spectral density given by Whittle (1951, 1953), where the autocovariance matrix is diagonalized by transforming the $\{y_t\}$ process into the frequency domain and the asymptotic properties of Toeplitz matrices and an equicontinuity property of quadratic forms are used to show consistency and asymptotic normality. The approximate log likelihood is as follows:

$$(6) \qquad \log L = \sum_{j=1}^{T-1} \log\left\{2\pi \cdot f(\omega_j)\right\} + \sum_{j=1}^{T-1} \frac{I(\omega_j)}{f(\omega_j)},$$

where $I(\omega_j)$ is the periodogram and $f(\omega_j)$ is a spectral density of y as above.

The main focus of this chapter will be on minimum distance estimation (MDE). Let

$$(7) \qquad \rho = \left[\rho_1, \rho_2, \ldots, \rho_n\right]'$$

be the vector of the first n population autocorrelations, and let $\hat{\rho}$ be the corresponding vector of estimated (sample) autocorrelations. Tieslau, Schmidt, and Baillie (1994), hereafter TSB, suggest minimizing the distance between ρ and $\hat{\rho}$. More precisely, let

$$(8) \qquad \lambda = [d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]'$$

so that ρ depends on λ. TSB suggest minimization of a criterion function $\{\hat{\rho} - \rho(\lambda)\}' W \{\hat{\rho} - \rho(\lambda)\}$, where W is a positive definite matrix. This estimator is consistent, and it is $\sqrt{T}$-consistent for -1/2< d <1/4. A similar GMM estimator is suggested by Dueker and Startz (1992).

Chung and Schmidt (1995) provide a modification of the TSB estimator, "adjusted MDE" (AMDE), to achieve $\sqrt{T}$-consistency for -1/2< d <1/2, not just for -1/2< d <1/4. Define the vector

(9) $\delta = [\delta_1, \delta_2, \ldots, \delta_n]'$

where $\delta_1 = 1 - \rho_1$, $\delta_2 = \rho_1 - \rho_2$, ..., $\delta_n = \rho_{n-1} - \rho_n$. The information in δ is the same as in ρ and the MDE based on δ is asymptotically the same as the MDE based on ρ. However, Chung and Schmidt suggest an MDE based on (n-1) functions of <u>ratios</u> of elements of δ. More specifically, they consider ratios of the form

(10) $r_j = a_j'\delta / b_j'\delta$, j = 1, 2, ..., n-1,

where a_j and b_j are vectors of known constants; they also allow general differentiable functions of these ratios. Using results from Hosking (1995), they show that such an MDE, which they call an AMDE, is $\sqrt{T}$-consistent for -1/2< d <1/2. Its asymptotic variance does not depend on the choice of a_j, b_j, nor on which function of the ratios (10) are taken. For d <1/4, the AMDE is less efficient than TSB's MDE because it does not use information on the levels of the δ's (or ρ's).

Now consider the problem of estimating the differencing parameter d separately from the ARMA parameters that determine short-run dynamics. Every element of δ (or ρ) depends on a non-trivial way on the ARMA parameters. However, ratios like δ_j/δ_{j-1} (or ρ_j/ρ_{j-1}) depend only on d, in the limit as $j \to \infty$. This suggests that we could ignore short-run dynamics by considering ratios of sufficiently high-order autocorrelations, and this is easy to do in the AMDE of Chung and Schmidt. Let n be the number of δ's considered, and let ℓ be the number of elements of δ that are not considered, so that we consider only

$$(11) \qquad \delta_{n,\ell} \equiv \left[\delta_{\ell+1}, \delta_{\ell+2}, \ldots, \delta_{\ell+n}\right]'.$$

We will consider the AMDE based on $\delta_{n,\ell}$, ignoring short-run dynamics; that is, the AMDE assuming the I(d) or ARFIMA(0,d,0) model. When the data are generated by the ARFIMA(p,d,q) model, the AMDE based on the (0,d,0) model will yield asymptotically biased (inconsistent) estimates. However, we expect this asymptotic bias to be small when ℓ is large. In this chapter we will calculate this asymptotic bias, as a function of n, ℓ and the parameters. The idea is to see whether approximately unbiased estimates of d can be obtained, without the need to model short-run dynamics. The idea of doing so using high-order autocorrelations obviously is similar to the GPH idea of using the periodogram at low frequencies. Lobato and Robinson (1993) used the same idea.

2. MOMENT CONDITIONS FOR MDE

In this section, we give the explicit form of the functions of δ that are used in the AMDE estimators that we will consider. To do so, we first need to give a little more detail on the TSB and Chung-Schmidt procedures. Consider the case of an I(d) process; that is, a fractionally integrated white noise. The MDE of d, say $\hat{d}$, is defined as the value which minimizes the following criterion function:

$$(12) \quad S(d)=\{F[\hat{\rho}]-F[\rho(d)]\}' \, W \, \{F[\hat{\rho}]-F[\rho(d)]\},$$

where $\rho(d)=[\rho_1(d), \rho_2(d), \ldots, \rho_n(d)]'$ is a vector of population autocorrelations that depends on the value of d, $\hat{\rho}$ is the corresponding vector of sample autocorrelations, F is a m-dimensional vector of transformation functions of the autocorrelations (m≤n), and W is a symmetric and positive-definite weighting matrix. The asymptotically optimal weighting matrix is as follows:

$$(13) \quad W = AV(F(\hat{\rho}))^{-1} = [P\, C\, P']^{-1}.$$

Here $P = \partial F/\partial \rho'$ (mxn). C is the asymptotic variance-covariance matrix of the sample autocorrelations (nxn). For -1/2< d <1/4 it can be defined by Bartlett's formula,

$$(14) \quad C_{i,j} = \sum_{s=1}^{\infty}(\rho_{s+i}+\rho_{s-i}-2\rho_i\rho_s)(\rho_{s+j}+\rho_{s-j}-2\rho_j\rho_s), \; j = 1, 2, \ldots, n,$$

from Hosking (1995, 1984) and Brockwell and Davis (1991). Define $D = \partial\rho / \partial d$ (nx1). Then the asymptotic variance of the MDE $\hat{d}$ defined in (12) above is:

$$(15) \quad AV(\hat{d}) = \left\{[D'P'WPD]^{-1}D'P'W(PCP')WPD[D'P'WPD]^{-1}\right\}.$$

When $W = [PCP']^{-1}$, this simplifies to

$$(16) \quad AV(\hat{d}) = \left[D'P'(PCP')^{-1}PD\right]^{-1}.$$

We can note that P cancels if it is nonsingular, which will be the case if n = m and the elements of F are functionally independent. In this case the transformation F is asymptotically irrelevant. However, F clearly matters when m < n.

This MDE is similar to the GMM estimator introduced by Hansen (1982). The GMM estimator of a parameter, say β (hx1), is based on conditions of the form:

$$E[g_i(y_t, \beta)] = 0,\ i = 1, 2, \ldots, k\ (h \leq k).$$

The GMM estimator minimizes a criterion function:

$$\min\ S(\beta) = \{\bar{g}(\beta)'W\ \bar{g}(\beta)\}$$

where $\bar{g}_i(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_i(y_t, \beta)$. Therefore, the above MDE of d is not a GMM estimator because $F(\hat{\rho})$ is not an average, but it still has the basic properties of a GMM estimator.

TSB (1995) introduced the simplest case of the MDE using the trivial transformation $F\{\rho(d)\} = \rho(d)$, so that m = n. Then the criterion in equation (12) can be simplified as:

$$(17) \quad S(d)=\{\hat{\rho} - \rho(d)\}'\ W\ \{\hat{\rho} - \rho(d)\}.$$

The optimal weighting matrix is $W = C^{-1}$ since P = I. The asymptotic distribution of $\sqrt{T}(\hat{d} - d)$, for -1/2< d <1/4, is $N(0, [D'CD]^{-1})$. For d =1/4, $\hat{d}$ converges to a normal distribution, but at a rate of $(T/\ln T)^{1/2}$; for 1/4< d <1/2, the MDE converges to a non-normal asymptotic distribution at a rate of $T^{(1-2d)}$. Thus for d $\geq$1/4 convergence is slower than the usual $T^{1/2}$ rate.

These asymptotic results basically depend on the asymptotic distribution of $\hat{\rho}$ (for details, see Hosking (1995)). However, even for $d \in [1/4, 1/2)$, Hosking (1995) has shown that the following normalized differences of the sample autocorrelations are $\sqrt{T}$-consistent and asymptotically normal:

$$(18) \quad \sqrt{T}\left[\frac{\hat{\rho}_k - \rho_k}{1-\rho_k} - \frac{\hat{\rho}_\ell - \rho_\ell}{1-\rho_\ell}\right].$$

Therefore, TSB (1995) suggested an MDE based on such normalized differences of the sample autocorrelations, which is $\sqrt{T}$-consistent for the whole stationary and invertible range ($-1 < d < 1/2$) of the long-memory process.

Also, Chung and Schmidt (1995) introduced an AMDE that is also $\sqrt{T}$-consistent for the entire range $-1 < d < 1/2$. The MDE based on the quantity in (18) is an AMDE. More generally, Chung and Schmidt rely on Hosking's result that differences of sample auto<u>covariances</u> are $\sqrt{T}$-consistent: $\sqrt{T}\left[(\hat{\gamma}_i - \hat{\gamma}_j) - (\gamma_i - \gamma_j)\right]$ is asymptotically normal with variance given by Hosking (1995, equation (16)). Ratios of differences of autocorrelations are functions of differences of autocovariances; for example, $(\rho_3 - \rho_2)/(1-\rho_1) = (\gamma_3 - \gamma_2)/(\gamma_0 - \gamma_1)$. With δ as defined in (9) above, the ratios $r_j = a_j'\delta / b_j'\delta$ given in (10) above are ratios of differences of autocorrelations and provide the basis for a $\sqrt{T}$-consistent AMDE.

Chung and Schmidt (1995) provide several different but asymptotically equivalent forms of the AMDE. They correspond to different choices of the constants a_j, b_j, and different functions of the ratios r_j. Formally, they also correspond to choices of the function F in (12) above. In all cases, if ρ (or δ) has n elements, F is of dimension

(n-1). The formulas given earlier in this section depend on the matrix C, which is defined only for d <1/4, but Hosking (1995) and Chung and Schmidt (1995) give the necessary modifications that are well-defined for d in the entire range -1< d <1/2.

We will discuss three different specializations of F. Because we will be interested in the general ARFIMA(p,d,q) process, we will take the parameter vector as:

(19) $\lambda = [d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]' = [d, \Theta']'$

as in (8) above, with $\Theta = [\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]'$; hence we distinguish the differencing parameter d from the ARMA parameters Θ.

The GLS estimator introduced by Chung and Schmidt (1995) is based on the function F^1 defined by:

(20) $F_j^1[\rho(d, \Theta)] = 1 - j \cdot b_j(d, \Theta)$, $j = (\ell+1), (\ell+2), \ldots, (\ell+n)$.

where

(21) $b_j(d, \Theta) = \dfrac{\delta_j(d, \Theta) - \delta_{j+1}(d, \Theta)}{\delta_j(d, \Theta) + \delta_{j+1}(d, \Theta)}$, with $\delta_j(d, \Theta) = \rho_{j-1}(d, \Theta) - \rho_j(d, \Theta)$.

Thus $b_j(d, \Theta) = [\rho_{j-1}(d, \Theta) - 2\rho_j(d, \Theta) + \rho_{j+1}(d, \Theta)] / [\rho_{j-1}(d, \Theta) - \rho_{j+1}(d, \Theta)]$. As in the previous section, ℓ is the lowest-order autocorrelation used in the calculation; $(n+\ell+1)$ is the highest-order autocorrelation used; and estimation is based on n moment conditions, derived from n ratios of linear combinations of $\delta_{n+1,\ell}$ as defined in (11) above. Chung and Schmidt (1995) consider only the case $\ell = 0$ (which is well defined, with the convention $\rho_0 = 1$). For this pure I(d) process without short-run dynamics (i.e., $\Theta = 0$), $F_j^1[\rho(d, 0)] = 1 - j \cdot b_j(d, 0) = d$ for all j, and the MDE is expressible as a GLS regression of F^1 on a vector of ones. See Chung and Schmidt (1995) for more detail. They refer to this as the GLS version of the AMDE.

A second possibility is to use the following transformation F^2:

$$(22) \quad F_j^2[\rho(d, \Theta)] = \frac{\rho_j(d, \Theta)}{\rho_{j-1}(d, \Theta)}, \; j = (\ell+1), (\ell+2), \ldots, (\ell+n).$$

For the pure I(d) process without short-run dynamics we have:

$$(23) \quad F_j^2[\rho(d)] = \frac{(j-1)+d}{j-d}.$$

A third possibility is similar to the second, but uses the same denominator for each ratio. That is, the function F^3 is defined by:

$$(24) \quad F_j^3[\rho(d, \Theta)] = \frac{\rho_j(d, \Theta)}{\rho_\ell(d, \Theta)}, \; j = (\ell+1), (\ell+2), \ldots, (\ell+n).$$

For the pure I(d) process without short-run dynamics we have:

$$F_j^3[\rho(d)] = \prod_{k=\ell+1}^{j} \frac{(k-1+d)}{(k-d)}.$$

There is presumably an efficiency loss from ignoring low-order autocorrelations, so that the asymptotic variances of the AMDE or the MDE grow with ℓ. However, larger values of ℓ are useful to avoid the asymptotic bias that results from ignoring or misspecifying short-run dynamics. We now turn to the calculation of this asymptotic bias.

3. CALCULATION OF ASYMPTOTIC BIAS

Suppose that $\{y_t\}$ follows an ARFIMA(p,d,q) process as in (1) above. Let $\lambda = [d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]' = [d, \Theta']'$ as above. We can apply the MDE or AMDE estimators described in the previous section to obtain a consistent estimator of λ. However, we now consider the case that we assume (incorrectly) that $\{y_t\}$ follows a pure I(d), or ARFIMA(0,d,0), process and we calculate the MDE or AMDE estimator of d. This estimate will in general be inconsistent. Let $\bar{d}$ represent the probability limit of the estimate $\hat{d}$, so that the asymptotic bias is $(\bar{d} - d_0)$, where d_0 is the true (population) value of d. We wish to evaluate this asymptotic bias.

The MDE based on the assumed (0,d,0) model minimizes the criterion function:

$$\text{(26)} \quad \min S(d) = \{F[\hat{\rho}] - F[\rho(d, 0)]\}' \, W \, \{F[\hat{\rho}] - F[\rho(d, 0)]\}.$$

We can calculate $\bar{d} = \operatorname{plim} \hat{d}$ by invoking the general principle that $\bar{d}$ should minimize in the population the same criterion that $\hat{d}$ minimizes in the sample. Since $\operatorname{plim} \hat{\rho} = \rho(d_0, \Theta)$, $\bar{d}$ minimizes the criterion function:

$$\text{(27)} \quad \min S(\bar{d}) = \{F[\rho(d_0, \Theta_0)] - F[\rho(\bar{d}, 0)]\}' \, W \, \{F[\rho(d_0, \Theta_0)] - F[\rho(\bar{d}, 0)]\}.$$

Thus we can calculate $\bar{d}$ by a numerical minimization of the criterion (27), and from it calculate the asymptotic bias ($\bar{d}$ - d_0). This will depend on the function F, the weighting matrix W, the differencing parameter d_0, and the ARMA parameters Θ.

The transformation $F[\rho] = \rho$ defined the (ordinary) MDE of TSB. Suppose that the MDE is based on $(\rho_\ell, \ldots, \rho_{\ell+n})$, where the standard case of TSB is $\ell = 1$. For this MDE there is no reason to expect the asymptotic bias to decrease when ℓ increases, since every autocorrelation $\rho_j(d, \Theta)$ depends on Θ, no matter how large j is. However, ratios like $\rho_j(d, \Theta)/\rho_{j-1}(d, \Theta)$ depend only on d, not on Θ, for sufficiently large j. Thus, for the AMDE based on the function F^1 or the MDE based on F^2 and F^3 of the previous section, each of which involves ratios, we do expect the asymptotic bias to decrease as ℓ increases.

For the MDE based on F^2 or F^3, defined in equations (22) and (24) above, $\bar{d}$ must be calculated by a numerical minimization. For the GLS version of the AMDE, based on F^1 defined in equation (20), there is a closed form (GLS) solution for $\bar{d}$:

(28) $$\bar{d} = i_n' W\, F^1[\rho(d_0, \Theta)] / [i_n' W\, i_n],$$

where i_n is a vector of ones, of dimension n.

Finally, we need to discuss the relevant form of the weighting matrix, W. The optimal weighting matrix, $(PCP')^{-1}$ above, depends on population quantities and would typically be estimated using the results of some initial estimation. Since the initial estimates will also generally be biased when short-run dynamics are misspecified, there is some ambiguity in how we should handle the weighting matrix

W in our bias calculations. Therefore we will consider three different cases corresponding to different treatment of the weighting matrix. In Case 1, we use the identity matrix as the weighting matrix, so $W = I_n$. In Case 2, we use the weighting matrix in which P and C are evaluated using the true autocorrelations, $\rho(d_0, \Theta_0)$. This is unambiguous, but does not correspond to any feasible method of estimation that misspecifies Θ as zero. In Case 3, we assume that the weighting matrix is evaluated in the sample based on the initial estimate $\tilde{d} = \hat{\rho}_1/(1+\hat{\rho}_1)$, which would be consistent if the (0,d,0) model were true. When the (p,d,q) model is true, $\tilde{d} \rightarrow d^* \equiv \rho_1(d_0, \Theta)/[1+\rho_1(d_0, \Theta)]$ and we evaluate P and C using $d = d^*$.

4. RESULTS

In this section, we provide the results of our calculations of the asymptotic bias of the MDE of d, when we ignore short-run dynamics in the true ARFIMA(p,d_0,q) process. We consider three forms of transformation functions F^1, F^2, and F^3, and three cases that differ in the evaluation of the weighting matrix, as described at the end of last section. The minimization problem (27) that defined $\bar{d}$ was solved (for the MDE based on F^2 and F^3) using the optimization procedure in GAUSS 2.0. For simplicity we consider only ARFIMA(1,d_0,0) and ARFIMA(0,d_0,1) processes, where just one short-run parameter exists (ϕ or θ).

Tables 3.1-1, 3.1-2 and 3.1-3 give the asymptotic bias of the AMDE and the MDE for the case of an ARFIMA(1,d_0,0) process, for three parameter values: (d_0 =

.2, ϕ = .4), (d_0 = .2, ϕ = .8) and (d_0 = .4, ϕ = .4). Tables 3.2-1 and 3.2-2 do the same for the case of an ARFIMA(0,d_0,1) process, for two parameter values (d_0 = .2, θ =-.4) and (d_0 = .2, θ = -.8). We consider numbers of moment conditions (n) equal to 1, 3, 5 and 10. We consider lags (ℓ, equal to the lowest order autocorrelation used) equal to 0, 2, 3, 4, 5, 6, 7, 8, 10, 20 and 30. We consider the AMDE based on F^1 and the MDE based on F^2, and F^3, and three cases corresponding to the treatment of the weighting matrix, as we discussed previously.

Table 3.1 and 3.2 give us some clear and interesting results. (1) For fixed n, the asymptotic bias of the AMDE or the MDE that ignore short-run dynamics decreases and becomes close to zero as we use higher order of autocorrelations (larger ℓ). This is as expected given the characteristics of autocorrelations of our long memory process. This is our main result. It essentially implies that semi-parametric estimation of d through the MDE principle is possible, if we choose an appropriate form of transformation function of the autocorrelations and use high order autocorrelations. (2) The three different transformations (F^1, F^2, F^3) show similar results for larger values of ℓ. This especially true for F^2 and F^3, for which the results are quite similar even when ℓ is not very large. The absolute bias for the estimator based on F^1 is generally larger than for F^2 or F^3. (3) The choice of method of evaluating the weighting matrix (Case 1, 2 or 3) does not usually make much difference. It matters more for F^1 that for F^2 or F^3. (4) The asymptotic bias depends more on the order of autocorrelations used (ℓ) than on the number of moment conditions (n). Thus increasing n with fixed ℓ does not decrease asymptotic bias very much. Especially for large ℓ it has almost a negligible effect.

Tables 3.3-1, 3.3-2 and 3.3-3 give the asymptotic bias for many values of d_0 (= -.49, -.4, -.3, -.2, -.1, .1, .2, .24, .25, .3, .4, .49) and two different values of n (= 1, 10), with ϕ = .4. For n =1 the estimation problem is "exactly identified" in the sense that the number of moment conditions equals the number of parameters estimated. Therefore the choice of weighting matrix does not matter, and the results are the same for Cases 1, 2 and 3. For n =10, however, the choice of weighting matrix matters.

These results show that the asymptotic bias decreases as ℓ increases, as expected. The pattern of absolute bias as a function of d, holding constant n and ℓ, is complicated when ℓ is small. For larger values of ℓ (and both values of n), absolute bias decreases as d increases.

Table 3.4 provides the opposite comparison as in Table 3. It gives the asymptotic bias for many values of ϕ (= -.6, -.4, -.2, .2, .4, .6, .8, .9) for two different values of n (= 5, 10), with d_0 = .2. Results are given only for two relatively large values of ℓ, ℓ=20 and ℓ=30. The asymptotic bias is generally larger in absolute value when ϕ is larger in absolute value, as we would expect. For large $|\phi|$ (i.e., strong short-run dynamics), the asymptotic bias is discouragingly large even for ℓ=30. For example, for ϕ =.9 the asymptotic bias is about -0.1 or -0.2 with ℓ=30 (and d_0 = .2). This reflects the fact that any non-parametric treatment of short-run dynamics will have problems if they are strong enough; it is intrinsically difficult to distinguish long-run properties of the model from very strong short-run dynamics.

5. CONCLUDING REMARKS

In this chapter we have considered the MDE including the adjusted MDE (AMDE) estimator of Chung and Schmidt (1995) for the differencing parameter in the general ARFIMA model. In applying the MDE, one can estimate the ARFIMA model, which amounts to modeling short-run dynamics with an ARMA model; or one can estimate the pure I(d) model, but not using low-order autocorrelations, which is a non-parametric treatment of short-run dynamics. This non-parametric treatment is similar in spirit to the frequency-domain approach of Geweke and Porter-Hudak, based on the periodogram at low frequencies only. We expect a non-parametric treatment of short-run dynamics to have some cost in terms of efficiency, and Chung and Schmidt's results show that this is so. We also expect a non-parametric treatment to lead to finite sample bias, especially when the nuisance parameters (ARMA parameters) take on extreme values; i.e., when short-run dynamics are very strong. In this chapter, we do not evaluate finite samples biases, but we evaluate the asymptotic bias that results from ignoring a fixed number (ℓ) of low-order autocorrelations. The asymptotic bias is larger when short-run dynamics are stronger and when ℓ is smaller. This is as expected. It supports the conjecture of Tieslau, Schmidt and Baillie (1995) and Chung and Schmidt (1995) that a consistent nonparametric estimate of the differencing parameter results from letting ℓ grow with the sample size. A rigorous proof of this conjecture, and a derivation of the asymptotic properties of the estimate when ℓ grow with T, are important topics for future research.

TABLE 3.1-1

Asymptotic Bias of MDE in ARFIMA(1,$d_{\bullet}$,0)
[$d_{\bullet} = 0.2$, $\phi = 0.4$]

GLS (F^1)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.5348	.5348	.5348	.2323	.4798	.4788	.0399	.4573	.4418	-.0748	.4287	.3920
2	-.0466	-.0466	-.0466	-.1813	-.0997	-.1307	-.2137	-.1203	-.1668	-.1467	-.1395	-.1825
3	-.2123	-.2123	-.2123	-.2604	-.2376	-.2491	-.2416	-.2401	-.2509	-.1611	-.2292	-.2229
4	-.2851	-.2851	-.2851	-.2699	-.2790	-.2770	-.2266	-.2638	-.2541	-.1432	-.2295	-.2032
5	-.2838	-.2838	-.2838	-.2370	-.2561	-.2491	-.1894	-.2314	-.2163	-.1173	-.1889	-.1620
6	-.2410	-.2410	-.2410	-.1881	-.2057	-.1990	-.1471	-.1802	-.1672	-.0912	-.1408	-.1205
7	-.1863	-.1863	-.1863	-.1408	-.1536	-.1487	-.1096	-.1321	-.1229	-.0690	-.1006	-.0871
8	-.1371	-.1371	-.1371	-.1027	-.1111	-.1079	-.0806	-.0948	-.0888	-.0520	-.0715	-.0630
10	-.0721	-.0721	-.0721	-.0556	-.0588	-.0576	-.0452	-.0506	-.0483	-.0311	-.0387	-.0354
20	-.0115	-.0115	-.0115	-.0104	-.0106	-.0105	-.0096	-.0098	-.0097	-.0079	-.0084	-.0082
30	-.0050	-.0050	-.0050	-.0047	-.0047	-.0047	-.0044	-.0045	-.0045	-.0039	-.0040	-.0039

Ratios (F^2)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.1861	.1861	.1861	.1372	.2209	.3642	.1200	.2211	.3701	.1101	.2213	.3627
2	-.1126	-.1126	-.1126	-.1204	-.1072	-.1199	-.1113	-.1058	-.1159	-.0954	-.1053	-.1216
3	-.1355	-.1355	-.1355	-.1184	-.1252	-.1207	-.1034	-.1181	-.1101	-.0847	-.1092	-.1072
4	-.1166	-.1166	-.1166	-.0936	-.1005	-.0952	-.0790	-.0904	-.0830	-0628	-.0780	-.0755
5	-.0864	-.0864	-.0864	-.0665	-.0690	-.0674	-.0553	-.0593	-.0572	-.0431	-.0494	-.0495
6	-.0593	-.0593	-.0593	-.0453	-.0434	-.0457	-.0375	-.0378	-.0383	-.0290	-.0312	-.0318
7	-.0403	-.0403	-.0403	-.0308	-.0277	-.0329	-.0254	-.0234	-.0259	-.0198	-.0182	-.0210
8	-.0275	-.0275	-.0275	-.0212	-.0172	-.0214	-.0171	-.0144	-.0180	-.0139	-.0108	-.0144
10	-.0142	-.0142	-.0142	-.0115	-.0068	-.0115	-.0099	-.0057	-.0093	-.0079	-.0040	-.0075
20	-.0013	-.0013	-.0013	-.0015	-.0000	-.0011	-.0016	-.0000	-.0009	-.0018	-.0000	-.0008
30	-.0000	-.0000	-.0000	-.0002	.0000	.0000	-.0000	.0000	-.0001	-.0005	.0000	-.0001

Common Denominator Ratios (F^3)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.1861	.1861	.1861	.1449	.2072	.2378	.1244	.2055	.2394	.1023	.2128	.2397
2	-.1126	-.1126	-.1126	-.1197	-.1128	-.1092	-.1159	-.1106	-.1242	-.1031	-.1081	-.1384
3	-.1355	-.1355	-.1355	-.1236	-.1219	-.1419	-.1118	-.1190	-.1600	-.0922	-.1112	-.1680
4	-.1166	-.1166	-.1166	-.0989	-.1008	-.1115	-.0863	-.0923	-.1243	-.0683	-.0808	-.1371
5	-.0864	-.0864	-.0864	-.0706	-.0694	-.0757	-.0605	-.0638	-.0792	-.0469	-.0520	-.0902
6	-.0593	-.0593	-.0593	-.0481	-.0470	-.0487	-.0410	-.0398	-.0482	-.0316	-.0339	-.0531
7	-.0403	-.0403	-.0403	-.0326	-.0314	-.0319	-.0279	-.0268	-.0194	-.0216	-.0224	-.0314
8	-.0276	-.0276	-.0276	-.0225	-.0168	-.0216	-.0194	-.0185	-.0101	-.0152	-.0154	-.0195
10	-.0142	-.0142	-.0142	-.0120	-.0074	-.0115	-.0106	-.0033	-.0099	-.0086	-.0056	-.0091
20	-.0019	-.0019	-.0019	-.0025	-.0000	-.0023	-.0023	-.0000	-.0019	-.0021	-.0000	-.0010
30	-.0007	-.0007	-.0007	-.0011	.0000	.0000	-.0010	.0000	-.0009	-.0010	.0000	-.0001

TABLE 3.1-2

Asymptotic Bias of MDE in ARFIMA(1,$d_\bullet$,0)
[$d_\bullet = 0.2$, $\phi = 0.8$]

GLS (F^1)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.8472	.8472	.8472	.7435	.8204	.8246	.6468	.8078	.8079	.4300	.7908	.7780
2	.6419	.6419	.6419	.5484	.5986	.5857	.4603	.5749	.5409	.2659	.5428	.4557
3	.5471	.5471	.5471	.4576	.5014	.4863	.3737	.4752	.4363	.1909	.4390	.3413
4	.4563	.4563	.4563	.3709	.4096	.3937	.2962	.3820	.3415	.1209	.3430	.2422
5	.3695	.3695	.3695	.3883	.3229	.3069	.2132	.2946	.2544	.0558	.2540	.1544
6	.2869	.2869	.2869	.2101	.2411	.2254	.1398	.2125	.1739	-.0040	.1713	.0761
7	.2085	.2085	.2085	.1365	.1641	.1491	.0714	.1360	.0992	-.0583	.0952	.0060
8	.1349	.1349	.1349	.0679	.0920	.0785	.0082	.0648	.0313	-.1070	.0253	-.0554
10	.0026	.0026	.0026	-.0534	-.0349	-.0462	-.1016	-.0588	-.0861	-.1872	-.0934	-.1567
20	-.3129	-.3129	-.3129	-.3138	-.3136	-.3139	-.3100	-.3112	-.3113	-.2174	-.2995	.2952
30	-.2228	-.2228	-.2228	-.2085	-.2104	-.2092	-.1947	-.1996	-.1964	-.1641	-.1774	.1687

Ratios (F^2)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.2783	.2783	.2783	.2677	.2955	-.0841	.2616	.2955	-.1472	.2535	.2954	-.2293
2	.1623	.1623	.1623	.1269	.2256	-.0011	.1030	.2243	-.4822	.0697	.2235	-.8270
3	.1062	.1062	.1062	.0702	.1790	.0419	.0455	.1778	-.0807	.0117	.1768	-.5908
4	.0549	.0549	.0549	.0210	.1292	.0095	-.0022	.1284	-.0415	-.0330	.1278	-.1879
5	.0094	.0094	.0094	-.0210	.0783	-.0264	-.0415	.0782	-.0586	-.0673	.0778	-.1184
6	-.0299	-.0299	-.0299	-.0561	.0288	-.0587	-.0731	.0292	-.0813	-.0931	.0291	-.1142
7	-.0628	-.0628	-.0628	-.0843	-.0171	-.0860	-.0979	-.0165	-.1020	-.1117	-.0166	-.1209
8	-.0896	-.0896	-.0896	-.1065	-.0575	-.1074	-.1167	-.0570	-.1185	-.1245	-.0572	-.1283
10	-.1261	-.1261	-.1261	-.1344	-.1121	-.1348	-.1379	-.1149	-.1385	-.1360	-.1143	-.1361
20	-.1040	-.1040	-.1040	-1.003	-.0856	-.1008	-.0937	-.0816	-.0941	-.0799	-.0729	-.0804
30	-.0371	-.0371	-.0371	-.0394	-.0074	-.0404	-.0363	-.0217	-.0373	-.0307	-.0184	-.0316

Common Denominator Ratios (F^3)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.2783	.2783	.2783	.2617	.2968	.2890	.2480	.2968	.2892	.2224	.2987	.2892
2	.1623	.1623	.1623	.1318	.2209	.2513	.1090	.2207	.2672	.0709	2192	.2743
3	.1062	.1062	.1062	.0764	.1122	.1928	.0544	.1617	.2238	.0187	.1598	.2343
4	.0549	.0549	.0549	.0275	.1030	.1235	.0074	.1025	.1586	-.0242	.1006	.1587
5	.0094	.0094	.0094	-.0149	.0476	.0585	-.0324	.0472	.0906	-.0589	.0455	.0842
6	-.0299	-.0299	-.0299	-.0508	-.0016	.0020	-.0654	-.0021	.0293	-.0863	-.0036	.0210
7	-.0628	-.0628	-.0628	-.0801	-.0435	-.0445	-.0917	-.0442	-.0232	-.1071	-.0452	-.0317
8	-.0896	-.0896	-.0896	-.1033	-.0778	-.0804	-.1120	-.0784	-.0653	-.1221	-.0786	-.0744
10	-.1261	-.1261	-.1261	-.1331	-.1227	-.1260	-.1365	-.1221	-.1219	-.1375	-.1202	-.1344
20	-.1040	-.1040	-.1040	-.1025	-.0953	-.1023	-.0972	-.0872	-.1000	-.0860	-.0896	-.1084
30	-.0371	-.0371	-.0371	-.0412	-.0243	-.0407	-.0388	-.0352	-.0382	-.0339	-.0271	-.0356

TABLE 3.1-3

Asymptotic Bias of MDE in ARFIMA(1,d_0,0)
[d_0 = 0.4, ϕ = 0.4]

GLS (F^1)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.4902	.4902	.4902	.2560	.4415	.4378	.1221	4208	.4061	.0228	.3937	.3658
2	.0483	.0483	.0483	-.0365	.0105	-.0062	-.0621	-.0040	-.0297	-.0555	-.0184	-.0458
3	-.0568	-.0568	-.0568	-.0882	-.0741	-.0805	-.0869	.0774	.0854	-.0626	-.0756	-.0796
4	-.1010	-.1010	-.1010	-.1001	-.1014	-.1019	-.0874	-.0964	-.0954	-.0588	-.0848	-.0790
5	-.1067	-.1067	-.1067	-.0923	-.0976	-.0960	-.0764	-.0890	-.0851	-.0502	-.0742	-.0664
6	-.0941	-.0941	-.0941	-.0765	-.0819	-.0800	-.0621	-.0729	-.0690	-.0409	-.0548	-.0522
7	-.0761	-.0761	-.0761	-.0603	-.0646	-.0630	-.0489	-.0568	-.0537	-.0326	-.0451	-.0401
8	-.0593	-.0593	-.0593	-.0468	-.0498	-.0487	-.0382	-.0436	-.0454	-.0260	-.0344	-.0310
10	-.0355	-.0355	-.0355	-.0287	-.0300	-.0295	-.0241	-.0265	-.0255	-.0172	-.0211	-.0195
20	-.0072	-.0072	-.0072	-.0065	-.0066	-.0066	-.0060	-.0061	-.0061	-.0050	-.0053	-.0052
30	-.0032	-.0032	-.0032	-.0030	-.0030	-.0030	-.0028	-.0028	-.0028	-.0024	-.0025	-.0025

Ratios (F^2)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.0671	.0671	.0671	.0574	.0937	-.2947	.0547	.0932	-.2981	.0529	.0926	-.2092
2	-.0143	-.0143	-.0143	-.0164	.0047	-.0145	-.0156	-.0007	-.0180	-.0139	-.0080	-.0320
3	-.0196	-.0196	-.0196	-.0176	-.0324	-.0194	-.0157	-.0388	-.0226	-.0133	-.0489	-.0330
4	-.0173	-.0173	-.0173	-.0143	-.0245	-.0160	-.0124	-.0287	-.0173	-.0102	-.0371	-.0237
5	-.0132	-.0132	-.0132	-.0106	-.0148	-.0114	-.0091	-.0179	-.0118	-.0073	-.0234	-.0151
6	-.0096	-.0096	-.0096	-.0077	-.0099	-.0081	-.0066	-.0108	-.0080	-.0053	-.0149	-.0094
7	-.0070	-.0070	-.0070	-.0055	-.0067	-.0057	-.0048	-.0072	-.0054	-.0038	-.0088	-.0060
8	-.0050	-.0050	-.0050	-.0041	-.0048	-.0041	-.0036	-.0049	-.0039	-.0029	-.0058	-.0041
10	-.0030	-.0030	-.0030	-.0024	-.0026	-.0024	-.0022	-.0025	-.0023	-.0018	-.0028	-.0022
20	-.0005	-.0005	-.0005	-.0006	-.0000	-.0006	-.0004	-.0000	-.0005	-.0004	-.0000	-.0004
30	-.0000	-.0000	-.0000	-.0000	.0000	-.0000	-.0000	.0000	-.0002	-.0000	.0000	-.0000

Common Denominator Ratios (F^3)

	n=1			n=3			n=5			n=10		
ℓ	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3	Case.1	Case.2	Case.3
0	.0671	.0671	.0671	.0526	.0996	.0834	.0451	.0996	.0838	.0366	.0997	.0839
2	-.0143	-.0143	-.0143	-.0164	.0039	-.0023	-.0160	.0010	-.0165	-.0140	-.0060	-.0284
3	-.0196	-.0196	-.0196	-.0180	-.0268	-.0335	-.0162	-.0338	-.0440	-.0132	-.0470	-.0466
4	-.0173	-.0173	-.0173	-.0148	-.0186	-.0296	-.0129	-.0204	-.0408	-.0102	-.0303	-.0430
5	-.0132	-.0132	-.0132	-.0111	-.0124	-.0189	-.0096	-.0126	-.0277	-.0075	-.0158	-.0326
6	-.0096	-.0096	-.0096	-.0080	-.0084	-.0117	-.0069	-.0086	-.0166	-.0054	-.0098	-.0221
7	-.0070	-.0070	-.0070	-.0058	-.0062	-.0075	-.0051	-.0056	-.0099	-.0040	-.0065	-.0142
8	-.0050	-.0050	-.0050	-.0044	-.0046	-.0050	-.0038	-.0042	-.0062	-.0030	-.0043	-.0091
10	-.0030	-.0030	-.0030	-.0026	-.0027	-.0027	-.0023	-.0025	-.0030	-.0019	-.0019	-.0041
20	-.0005	-.0005	-.0005	-.0006	-.0003	-.0005	-.0006	-.0000	-.0006	-.0005	-.0003	-.0006
30	-.0000	-.0000	-.0000	-.0001	-.0000	-.0000	-.0003	.0000	-.0000	-.0002	-.0000	-.0000

TABLE 3.2-1

Asymptotic Bias of MDE in ARFIMA(0,$d_\bullet$,1)
[$d_\bullet$ = 0.2, θ = -0.4]

GLS (F^1)

	n=1			n=3			n=5			n=10		
ℓ	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
0	.6103	.6103	.6103	-.0153	.5947	.4591	-.0291	.5640	.4070	-.0222	.5178	.3548
2	-.1285	-.1285	-.1285	-.0761	-.1061	-.0957	-.0544	-.0929	-.0797	-.0318	-.0753	-.0610
3	-.0624	-.0624	-.0624	-.0471	-.0513	-.0480	-.0314	-.0445	-.0402	-.0195	-.0353	-.0305
4	-.0376	-.0376	-.0376	-.0270	-.0311	-.0297	-.0211	-.0270	-.0251	-.0137	-.0213	-.0191
5	-.0253	-.0253	-.0253	-.0191	-.0212	-.0204	-.0154	-.0185	-.0175	-.0103	-.0146	-.0134
6	-.0182	-.0182	-.0182	-.0143	-.0154	-.0150	-.0117	-.0136	-.0130	-.0081	-.0107	-.0100
7	-.0138	-.0138	-.0138	-.0111	-.0118	-.0116	-.0093	-.0104	-.0101	-.0066	-.0083	-.0078
8	-.0108	-.0108	-.0108	-.0089	-.0093	-.0092	-.0075	-.0083	-.0081	-.0055	-.0067	-.0063
10	-.0072	-.0072	-.0072	-.0061	-.0063	-.0062	-.0053	-.0057	-.0055	-.0040	-.0046	-.0044
20	-.0019	-.0019	-.0019	-.0018	-.0018	-.0018	-.0016	-.0017	-.0017	-.0014	-.0014	-.0014
30	-.0007	-.0007	-.0007	-.0008	-.0008	-.0008	-.0008	-.0008	-.0008	-.0007	-.0007	-.0007

Ratios (F^2)

	n=1			n=3			n=5			n=10		
ℓ	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
0	.1575	.1575	.1575	.0754	.2162	.3050	-.0678	.2192	.2981	.0628	.2193	.2853
2	-.0411	-.0398	-.0409	-.0283	-.0328	-.0298	-.0240	-.0297	-.0265	-.0202	-.0224	-.0250
3	-.0182	-.0176	-.0178	-.0133	-.0145	-.0135	-.0113	-.0129	-.0117	-.0093	-.0099	-.0103
4	-.0104	-.0059	-.0102	-.0079	-.0066	-.0080	-.0068	-.0059	-.0067	-.0056	-.0049	-.0057
5	-.0068	-.0029	-.0059	-.0053	-.0008	-.0051	-.0046	-.0022	-.0045	-.0038	-.0010	-.0038
6	-.0046	-.0020	-.0036	-.0039	-.0004	-.0038	-.0034	-.0004	-.0033	-.0027	-.0004	-.0027
7	-.0035	-.0014	-.0011	-.0029	-.0002	-.0029	-.0026	-.0002	-.0026	-.0021	-.0002	-.0021
8	-.0027	-.0008	-.0009	-.0023	-.0001	-.0015	-.0020	-.0001	-.0020	-.0017	-.0001	-.0017
10	-.0011	-.0008	-.0007	-.0015	-.0009	-.0006	-.0014	-.0000	-.0006	-.0011	.0000	-.0006
20	-.0000	-.0084	-.0001	-.0000	.0000	-.0000	-.0000	-.0000	.0000	-.0000	.0000	-.0000
30	-.0000	-.0200	-.0000	-.0000	.0102	-.0000	-.0002	.0074	.0000	-.0002	.0000	.0000

Common Denominator Ratios (F^3)

	n=1			n=3			n=5			n=10		
ℓ	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
0	.1575	.1575	.1575	.0957	.2015	.2026	.0801	.2021	.2028	.0656	.1985	.2011
2	-.0411	-.0405	-.0409	-.0310	-.0344	-.0376	-.0263	-.0310	-.0399	-.0209	-.0274	-.0422
3	-.0182	-.0176	-.0178	-.0142	-.0144	-.0149	-.0122	-.0123	-.0145	-.0097	-.0107	-.0154
4	-.0104	-.0083	-.0102	-.0084	-.0083	-.0082	-.0073	-.0069	-.0075	-.0059	-.0052	-.0076
5	-.0068	-.0009	-.0059	-.0056	-.0023	-.0054	-.0049	-.0046	-.0048	-.0040	-.0026	-.0044
6	-.0046	-.0005	-.0036	-.0041	-.0012	-.0038	-.0036	-.0033	-.0033	-.0029	-.0011	-.0029
7	-.0035	-.0003	-.0025	-.0031	-.0005	-.0029	-.0027	.0006	-.0025	-.0023	-.0004	-.0019
8	-.0024	-.0001	-.0009	-.0024	-.0002	-.0017	-.0022	.0003	-.0015	-.0018	-.0009	-.0014
10	-.0011	-.0008	-.0000	-.0016	.0003	-.0011	-.0015	.0000	-.0013	-.0012	-.0003	-.0009
20	-.0000	.0114	.0001	-.0004	.0006	-.0002	-.0004	.0012	-.0002	-.0004	.0024	-.0002
30	-.0000	.0200	.0000	-.0002	.0168	.0005	-.0002	.0259	.0003	-.0002	.0018	-.0003

TABLE 3.2-2

Asymptotic Bias of MDE in ARFIMA(0,$d_\bullet$,1)
[$d_\bullet$ = 0.2, θ = -0.8]

GLS (F^1)

	n=1			n=3			n=5			n=10		
ℓ	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
0	.8469	.8469	.8469	.0403	.9504	.6385	.0002	.9867	.5670	-.0091	1.0164	.4956
2	-.1513	-.1513	-.1513	-.0903	-.1318	-.1127	-.0647	-.1151	-.0939	-.0379	-.1004	-.2768
3	-.0745	-.0745	-.0745	-.0500	-.0639	-.0573	-.0377	-.0569	-.0480	-.0234	-.0465	-.0366
4	-.0451	-.0451	-.0451	-.0325	-.0386	-.0356	-.0254	-.0342	-.0301	-.0165	-.0276	-.0230
5	-.0304	-.0304	-.0304	-.0230	-.0261	-.0246	-.0185	-.0232	-.0210	-.0124	-.0186	-.0161
6	-.0220	-.0220	-.0220	-.0172	-.0190	-.0181	-.0142	-.0169	-.0156	-.0098	-.0136	-.0121
7	-.0166	-.0166	-.0166	-.0134	-.0145	-.0139	-.0112	-.0129	-.0121	-.0080	-.0104	-.0095
8	-.0130	-.0130	-.0130	-.0107	-.0114	-.0111	-.0091	-.0102	-.0097	-.0066	-.0083	-.0077
10	-.0086	-.0086	-.0086	-.0073	-.0077	-.0075	-.0064	-.0069	-.0067	-.0048	-.0057	-.0054
20	-.0023	-.0023	-.0023	-.0021	-.0021	-.0022	-.0020	-.0020	-.0020	-.0017	-.0017	-.0017
30	-.0011	-.0011	-.0011	-.0010	-.0010	-.0010	-.0009	-.0010	-.0009	-.0008	-.0008	-.0008

Ratios (F^2)

	n=1			n=3			n=5			n=10		
ℓ	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
0	.1970	.1970	.1970	.1110	.2656	.4270	.1012	.2714	.4125	.0945	.2740	.3973
2	-.0494	-.0479	-.0493	-.0340	-.0428	-.0359	-.0289	-.0410	-.0322	-.0243	-.0398	-.0312
3	-.0219	-.0213	-.0218	-.0161	-.0186	-.0164	-.0137	-.0172	-.0144	-.0113	-.0157	-.0130
4	-.0125	-.0082	-.0125	-.0096	-.0086	-.0097	-.0082	-.0079	-.0083	-.0067	-.0069	-.0072
5	-.0082	-.0023	-.0078	-.0065	-.0037	-.0063	-.0056	-.0034	-.0056	-.0046	-.0030	-.0047
6	-.0056	-.0014	-.0056	-.0047	-.0013	-.0046	-.0041	-.0012	-.0041	-.0033	-.0011	-.0034
7	-.0043	-.0010	-.0042	-.0036	-.0009	-.0035	-.0031	-.0008	-.0031	-.0025	-.0007	-.0026
8	-.0033	-.0007	-.0033	-.0028	-.0006	-.0028	-.0025	-.0006	-.0025	-.0020	-.0005	-.0020
10	-.0015	-.0012	-.0011	-.0019	-.0004	-.0019	-.0017	-.0003	-.0017	-.0014	-.0003	-.0014
20	-.0000	.0083	-.0000	-.0001	.0080	-.0001	-.0001	.0000	-.0001	-.0001	-.0000	-.0001
30	-.0000	.0058	-.0000	-.0000	.0056	-.0000	-.0000	.0000	.0000	-.0000	.0000	-.0000

Common Denominator Ratios (F^3)

	n=1			n=3			n=5			n=10		
ℓ	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
0	.1970	.1970	.1970	.1191	.2712	.2405	.0995	.2894	.2414	.0814	.2963	.2410
2	-.0494	-.0494	-.0493	-.0372	-.0444	-.0557	-.0316	-.0424	-.0602	-.0252	-.0399	-.0585
3	-.0219	-.0213	-.0218	-.0172	-.0186	-.0201	-.0148	-.0173	-.0217	-.0118	-.0148	-.0234
4	-.0125	-.0103	-.0125	-.0102	-.0105	-.0107	-.0088	-.0092	-.0108	-.0071	-.0074	-.0119
5	-.0082	-.0065	-.0081	-.0068	-.0052	-.0068	-.0060	-.0060	-.0065	-.0048	-.0043	-.0070
6	-.0056	-.0015	-.0056	-.0049	-.0022	-.0048	-.0043	-.0042	-.0044	-.0035	-.0022	-.0045
7	-.0043	-.0010	-.0042	-.0037	-.0013	-.0036	-.0033	.0001	-.0031	-.0027	-.0017	-.0033
8	-.0033	-.0007	-.0033	-.0029	-.0007	-.0024	-.0026	.0005	-.0025	-.0022	-.0010	-.0023
10	-.0015	-.0012	-.0015	-.0019	-.0001	-.0016	-.0018	.0008	-.0016	-.0015	-.0007	-.0014
20	-.0000	.0113	-.0000	-.0005	.0005	-.0003	-.0005	.0010	-.0003	-.0004	.0022	-.0003
30	-.0000	.0078	-.0000	-.0002	.0080	-.0004	-.0002	.0009	-.0002	-.0002	.0018	-.0001

TABLE 3.3-1

Asymptotic Bias of MDE in ARFIMA(1,$d_\bullet$,0) in GLS (F^1)
$\phi = 0.4$

[n = 1]

$d_\bullet$	-0.49	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.24	0.25	0.3	0.4	0.49
						$\ell = 0$						
I	.6359	.6262	.6144	.6014	.5872	.5541	.5348	.5266	.5245	.5137	.4902	.4668
II	.6359	.6262	.6144	.6014	.5872	.5541	.5348	.5266	.5245	.5137	.4902	.4668
III	.6359	.6262	.6144	.6014	.5872	.5541	.5348	.5266	.5245	.5137	.4902	.4668
						$\ell = 2$						
I	1.908	4.009	-5.974	-1.282	-.5697	-.1362	-.0466	-.0207	-.0149	.0107	.0483	.0707
II	1.908	4.009	-5.974	-1.282	-.5697	-.1362	-.0466	-.0207	-.0149	.0107	.0483	.0707
III	1.908	4.009	-5.974	-1.282	-.5697	-.1362	-.0466	-.0207	-.0149	.0107	.0483	.0707
						$\ell = 5$						
I	.1688	.2654	.4174	.7338	2.489	-.5301	-.2838	-.2301	-.2188	-.1713	-.1067	-.0687
II	.1688	.2654	.4174	.7338	2.489	-.5301	-.2838	-.2301	-.2188	-.1713	-.1067	-.0687
III	.1688	.2654	.4174	.7338	2.489	-.5301	-.2838	-.2301	-.2188	-.1713	-.1067	-.0687
						$\ell = 10$						
I	-.1633	-.1264	-.0907	-.0542	.0051	-.1210	-.0721	-.0617	-.0595	-.0499	-.0355	-.0259
II	-.1633	-.1264	-.0907	-.0542	.0051	-.1210	-.0721	-.0617	-.0595	-.0499	-.0355	-.0259
III	-.1633	-.1264	-.0907	-.0542	.0051	-.1210	-.0721	-.0617	-.0595	-.0499	-.0355	-.0259
						$\ell = 20$						
I	-.0355	-.0314	-.0272	-.0235	-.0200	-.0141	-.0115	-.0106	-.0103	-.0092	-.0072	-.0056
II	-.0355	-.0314	-.0272	-.0235	-.0200	-.0141	-.0115	-.0106	-.0103	-.0092	-.0072	-.0056
III	-.0355	-.0314	-.0272	-.0235	-.0200	-.0141	-.0115	-.0106	-.0103	-.0092	-.0072	-.0056
						$\ell = 30$						
I	-.0147	-.0131	-.0115	-.0100	-.0086	-.0061	-.0050	-.0046	-.0045	-.0040	-.0032	-.0025
II	-.0147	-.0131	-.0115	-.0100	-.0086	-.0061	-.0050	-.0046	-.0045	-.0040	-.0032	-.0025
III	-.0147	-.0131	-.0115	-.0100	-.0086	-.0061	-.0050	-.0046	-.0045	-.0040	-.0032	-.0025

Note: I (=Case 1), II (=Case 2) and III (=Case 3).

TABLE 3.3-1, CONTINUED

[n = 10]

d_e	-0.49	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.24	0.25	0.3	0.4	0.49
						$\ell = 0$						
I	.1721	.5791	-.2207	2.714	-1.448	.2075	-.0748	-.0453	-.0391	-.0128	.0228	.0430
II	.9486	-1.457	.5816	.5671	.5260	.4506	.4287	.4211	.4192	.3837	.3937	.3793
III	.6779	.6236	.4388	.8594	.2579	.3985	.3920	.3878	.3866	.3657	.3658	.3504
						$\ell = 2$						
I	.2630	.5569	-.2777	2.646	-1.517	-.3017	-.1467	-.1304	-.1237	-.0953	-.0555	-.0314
II	.6601	.6371	1.464	-.7537	-.2933	-.2613	-.1395	-.1058	-.0984	-.0658	-.0184	-.0109
III	.8593	1.562	-.9605	.6082	-3.122	-.3478	-.1825	-.1427	-.1342	-.0975	-.0458	-.0149
						$\ell = 5$						
I	-.0903	-.0441	.0119	.0945	.3993	-.2136	-.1173	-.0973	-.0931	-.0752	-.0502	-.0348
II	-.1095	.0156	.0932	.2102	.4810	-.3674	-.1889	-.1534	-.1461	-.1154	-.0742	-.0497
III	-.1027	-.0142	.0615	.1812	.6322	-.3039	-.1620	-.1330	-.1269	-.1014	-.0664	-.0452
						$\ell = 10$						
I	-.0892	-.0740	-.0594	-.0455	-.0277	-.0452	-.0311	-.0275	-.0267	-.0231	-.0172	-.0130
II	-.1095	-.0895	-.0706	-.0526	-.0289	-.0579	-.0387	-.0341	-.0331	-.0285	-.0211	-.0159
III	-.1027	-.0847	-.0676	-.0514	-.0306	-.0518	-.0354	-.0313	-.0304	-.0262	-.0195	-.0146
						$\ell = 20$						
I	-.0237	-.0211	-.0184	-.0159	-.0136	-.0096	-.0079	-.0073	-.0071	-.0063	-.0050	-.0038
II	-.0253	-.0224	-.0196	-.0169	-.0146	-.0102	-.0084	-.0077	-.0075	-.0067	-.0053	-.0041
III	-.0247	-.0220	-.0192	-.0166	-.0142	-.0100	-.0082	-.0075	-.0074	-.0066	-.0051	-.0040
						$\ell = 30$						
I	-.0112	-.0101	-.0089	-.0077	-.0066	-.0047	-.0039	-.0035	-.0035	-.0031	-.0024	-.0019
II	-.0116	-.0104	-.0091	-.0079	-.0068	-.0048	-.0040	-.0037	-.0036	-.0032	-.0025	-.0020
III	-.0115	-.0103	-.0090	-.0078	-.0067	-.0048	-.0039	-.0036	-.0035	-.0032	-.0025	-.0019

TABLE 3.3-2

Asymptotic Bias of MDE in ARFIMA(1,d_o,0) in Ratios (F^2)
$\phi = 0.4$

[n = 1]

d_o	-0.49	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.24	0.25	0.3	0.4	0.49
						$\ell = 0$						
I	.4613	.4353	.4033	.3671	.3289	.2382	.1861	.1639	.1582	.1291	.0671	.0069
II	.4613	.4353	.4033	.3671	.3289	.2382	.1861	.1639	.1582	.1291	.0671	.0069
III	.4613	.4353	.4033	.3671	.3289	.2382	.1861	.1639	.1582	.1291	.0671	.0069
						$\ell = 2$						
I	.7322	.8796	1.258	-.0700	-2.120	-.2401	-.1126	-.0813	-.0747	-.0474	-.0143	-.0008
II	.7322	.8796	1.258	-.0700	-2.120	-.2401	-.1126	-.0813	-.0747	-.0474	-.0143	-.0008
III	.7322	.8796	1.258	-.0700	-2.120	-.2401	-.1126	-.0813	-.0747	-.0474	-.0143	-.0008
						$\ell = 5$						
I	-.0106	.0401	.0986	.1803	.3999	-.2030	-.0864	-.0623	-.0574	-.0378	-.0132	-.0010
II	-.0106	.0401	.0986	.1803	.3999	-.2030	-.0864	-.0623	-.0574	-.0378	-.0132	-.0010
III	-.0106	.0401	.0986	.1803	.3999	-.2030	-.0864	-.0623	-.0574	-.0378	-.0132	-.0010
						$\ell = 10$						
I	-.0900	-.0694	-.0509	-.0347	-.0171	.0062	-.0142	-.0107	-.0101	-.0074	-.0030	-.0001
II	-.0900	-.0694	-.0509	-.0347	-.0171	.0062	-.0142	-.0107	-.0101	-.0074	-.0030	-.0001
III	-.0900	-.0694	-.0509	-.0347	-.0171	.0062	-.0142	-.0107	-.0101	-.0074	-.0030	-.0001
						$\ell = 20$						
I	-.0168	-.0148	-.0102	-.0077	-.0038	.0001	-.0013	-.0020	.0020	-.0014	-.0005	-.0001
II	-.0168	-.0148	-.0102	-.0077	-.0038	.0001	-.0013	-.0020	.0020	-.0014	-.0005	-.0001
III	-.0168	-.0148	-.0102	-.0077	-.0038	.0001	-.0013	-.0020	.0020	-.0014	-.0005	-.0001
						$\ell = 30$						
I	-.0060	-.0037	-.0018	.0008	-.0006	.0010	-.0000	-.0009	-.0009	-.0006	-.0000	-.0001
II	-.0060	-.0037	-.0018	.0008	-.0006	.0010	-.0000	-.0009	-.0009	-.0006	-.0000	-.0001
III	-.0060	-.0037	-.0018	.0008	-.0006	.0010	-.0000	-.0009	-.0009	-.0006	-.0000	-.0001

TABLE 3.3-2, CONTINUED

[n = 10]

d_e	-0.49	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.24	0.25	0.3	0.4	0.49
						$\ell = 0$						
I	.9485	-.8312	.2217	-153.2	.5731	.0842	.1101	.1063	.1046	.0921	.0529	.0057
II	.4357	.4807	.5512	-.7021	.1923	.2620	.2213	.2021	.1970	.1678	.0926	.0099
III	.3767	.3708	.3692	9999	.2847	.3865	.3627	.3395	.3331	.2992	-.2092	.0094
						$\ell = 2$						
I	.2955	.4097	.6994	-37.40	1.640	-.2135	-.0954	-.0696	-.0642	-.0419	-.0139	-.0009
II	.3775	.4229	.3904	.3860	.4067	-.2247	-.1053	-.0768	-.0724	-.0472	-.0080	.0064
III	.3573	.4301	.6850	9999	1.819	-.2320	-.1216	-.0940	-.0878	-.0664	-.0320	-.0009
						$\ell = 5$						
I	-.0741	-.0399	-.0044	.0380	.1316	-.0986	-.0431	-.0319	-.0296	-.0200	-.0073	-.0006
II	-.0682	-.0257	.0011	.0524	.1460	-.0951	-.0494	-.0394	-.0374	-.0292	-.0234	.0038
III	.0317	.0281	-.0076	.0262	.1033	-.0938	-.0495	-.0384	-.0359	-.0275	-.0151	-.0008
						$\ell = 10$						
I	-.0526	-.0422	-.0326	-.0241	-.0155	-.0133	-.0079	-.0064	-.0059	-.0043	-.0018	-.0001
II	-.0154	.0055	-.0033	-.0090	.0060	.0107	-.0040	-.0066	-.0063	-.0049	-.0028	-.0003
III	.0600	-.0066	-.0024	-.0089	-.0082	-.0116	-.0075	-.0066	-.0060	-.0046	-.0022	-.0002
						$\ell = 20$						
I	-.0125	-.0106	-.0086	-.0068	-.0054	-.0025	-.0018	-.0016	-.0015	-.0010	-.0004	-.0001
II	.0600	-.0300	.0300	.0400	-.0094	.0103	-.0000	-.0038	-.0015	-.0011	-.0000	-.0000
III	.0600	-.0300	-.0068	-.0008	-.0021	-.0024	-.0008	-.0016	-.0015	-.0010	-.0004	-.0000
						$\ell = 30$						
I	-.0058	-.0049	-.0038	-.0026	-.0014	-.0012	-.0005	-.0008	-.0007	-.0005	-.0000	-.0000
II	.0600	-.0300	.0300	-.0700	-.0700	.0079	-.0000	-.0007	-.0007	-.0005	-.0000	-.0000
III	.0600	-.0300	.0300	.0008	.0002	-.0012	-.0001	-.0008	-.0007	-.0005	-.0000	-.0000

TABLE 3.3-3

Asymptotic Bias of MDE in ARFIMA(1,$d_\bullet$,0) in Common Denominator Ratios (F^3)
$\phi = 0.4$
[n= 1]

$d_\bullet$	-0.49	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.24	0.25	0.3	0.4	0.49
						$\ell = 0$						
I	.4613	.4353	.4033	.3671	.3289	.2382	.1861	.1639	.1582	.1291	.0671	.0069
II	.4613	.4353	.4033	.3671	.3289	.2382	.1861	.1639	.1582	.1291	.0671	.0069
III	.4613	.4353	.4033	.3671	.3289	.2382	.1861	.1639	.1582	.1291	.0671	.0069
						$\ell = 2$						
I	.7322	.8796	1.258	-.0700	-2.120	-.2401	-.1126	-.0813	-.0747	-.0474	-.0143	-.0008
II	.7322	.8796	1.258	-.0700	-2.120	-.2401	-.1126	-.0813	-.0747	-.0474	-.0143	-.0008
III	.7322	.8796	1.258	-.0700	-2.120	-.2401	-.1126	-.0813	-.0747	-.0474	-.0143	-.0008
						$\ell = 5$						
I	-.0106	.0401	.0986	.1803	.3999	-.2030	-.0864	-.0623	-.0574	-.0378	-.0132	-.0010
II	-.0106	.0401	.0986	.1803	.3999	-.2030	-.0864	-.0623	-.0574	-.0378	-.0132	-.0010
III	-.0106	.0401	.0986	.1803	.3999	-.2030	-.0864	-.0623	-.0574	-.0378	-.0132	-.0010
						$\ell = 10$						
I	-.0900	-.0694	-.0509	-.0347	-.0171	.0062	-.0142	-.0107	-.0101	-.0074	-.0030	-.0001
II	-.0900	-.0694	-.0509	-.0347	-.0171	.0062	-.0142	-.0107	-.0101	-.0074	-.0030	-.0001
III	-.0900	-.0694	-.0509	-.0347	-.0171	.0062	-.0142	-.0107	-.0101	-.0074	-.0030	-.0001
						$\ell = 20$						
I	-.0168	-.0148	-.0102	-.0077	-.0038	.0001	-.0019	-.0020	.0020	-.0014	-.0005	-.0001
II	-.0168	-.0148	-.0102	-.0077	-.0038	.0001	-.0019	-.0020	.0020	-.0014	-.0005	-.0001
III	-.0168	-.0148	-.0102	-.0077	-.0038	.0001	-.0019	-.0020	.0020	-.0014	-.0005	-.0001
						$\ell = 30$						
I	-.0060	-.0037	-.0018	.0008	-.0006	.0010	-.0007	-.0009	-.0009	-.0006	-.0000	-.0001
II	-.0060	-.0037	-.0018	.0008	-.0006	.0010	-.0007	-.0009	-.0009	-.0006	-.0000	-.0001
III	-.0060	-.0037	-.0018	.0008	-.0006	.0010	-.0007	-.0009	-.0009	-.0006	-.0000	-.0001

TABLE 3.3-3, CONTINUED

[n= 10]

d_θ	-0.49	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.24	0.25	0.3	0.4	0.49
						$\ell = 0$						
I	.4018	.3596	.3109	.2624	.2159	.1360	.1023	.0894	.0862	.0700	.0366	.0038
II	.3124	.2997	.2849	.2692	.2528	.2181	.2128	.1964	.1944	.1795	.0997	.0100
III	.3892	.3756	.3594	.3412	.3200	.2816	.2397	.2110	.2036	.1652	.0839	.0084
						$\ell = 2$						
I	.4114	.4653	.5960	-24.45	-2.189	-.2458	-.1031	-.0736	-.0676	-.0432	-.0140	-.0000
II	.3197	.3594	.4121	.2500	-1.468	-.2426	-.1081	-.0791	-.0731	-.1179	-.0060	.0076
III	.4321	.4180	.5319	-14.88	-1.496	-.2718	-.1384	-.1055	-.0982	-.0686	-.0284	-9999
						$\ell = 5$						
I	-.0673	-.0282	.0109	.0562	.1525	-.1164	-.0469	-.0341	-.0315	-.0209	-.0075	.0006
II	-.0733	-.0334	.0016	.0490	.0859	-.1057	-.0520	-.0395	-.0368	-.0264	-.0158	.0044
III	.0600	.0107	.0018	.0351	.1090	-.1662	-.0902	-.0711	-.0670	-.0541	-.0326	-.0021
						$\ell = 10$						
I	-.0627	-.0492	-.0371	-.0269	-.0170	-.0150	-.0086	-.0069	-.0065	-.0047	-.0019	-.0000
II	-.0347	-.0034	-.0277	.0216	.0219	-.0122	-.0056	-.0068	-.0054	-.0043	-.0019	-.0000
III	.0600	.0700	-.0271	-.0088	-.0024	-.0128	-.0091	-.0077	-.0075	-.0063	-.0041	-.0000
						$\ell = 20$						
I	-.0137	-.0116	-.0094	-.0075	-.0059	-.0031	-.0021	-.0017	-.0016	-.0012	-.0005	-.0000
II	.0600	.0700	-.0300	-.0700	-.0700	.0251	-.0000	-.0016	-.0011	-.0003	-.0003	-.0000
III	.0600	.0700	-.0300	-.0023	.0181	-.0004	-.0000	-.0015	-.0011	-.0012	-.0006	-.0000
						$\ell = 30$						
I	-.0062	-.0052	-.0043	-.0034	-.0027	-.0015	-.0010	-.0008	-.0007	-.0006	-.0002	-.0000
II	.0600	.0700	-.0300	-.0700	-.0700	.0699	.0000	.0000	.0002	.0002	-.0000	-.0000
III	.0600	.0700	-.0300	.0398	.0141	-.0004	.0001	-.0001	-.0004	.0001	-.0000	-.0000

TABLE 3.4

Asymptotic Bias of MDE in ARFIMA(1,d$_\bullet$,0)
d$_\bullet$ = 0.2

[n = 5]

ϕ	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.9
					GLS (F^1)				
					$\ell = 20$				
I	-.2849	.0016	.0011	.0000	-.0025	-.0096	-.0461	-.3100	.0113
II	-.0652	.0016	.0011	.0000	-.0026	-.0098	-.0483	-.3112	.0283
III	-.0813	.0016	.0011	.0000	-.0026	-.0097	-.0471	-.3113	.0163
					$\ell = 30$				
I	-.0042	.0008	.0005	-.0000	-.0012	-.0044	-.0165	-.1947	-.2147
II	.0002	.0008	.0005	-.0000	-.0012	-.0045	-.0168	-.1996	-.2065
III	-.0006	.0008	.0005	-.0000	-.0012	-.0045	-.0166	-.1964	-.2128
					Ratios (F^2)				
					$\ell = 20$				
I	-.0057	.0000	.0000	-.0000	-.0000	-.0016	-.0092	-.0937	-.1261
II	-.0000	.0000	.0000	.0000	.0000	-.0000	.0017	-.0816	-.0876
III	-.0000	.0000	.0000	.0000	.0000	-.0009	-.0092	-.0941	-.1267
					$\ell = 30$				
I	.0000	.0000	.0000	.0000	-.0001	-.0000	-.0028	-.0363	-.1411
II	.0000	.0000	.0000	.0000	-.0000	-.0000	.0000	-.0217	-.1182
III	.0000	.0000	.0000	.0000	-.0000	-.0001	-.0033	-.0373	-.1433
					Common Denominator Ratios (F^3)				
					$\ell = 20$				
I	-.0071	.0004	.0003	.0000	-.0006	-.0023	-.0097	-.0972	-.1242
II	.0000	.0000	.0000	.0000	.0000	-.0000	-.0030	-.0872	-.1062
III	.0000	.0000	.0000	.0000	.0000	-.0019	-.0092	-.1000	-.1199
					$\ell = 30$				
I	.0001	.0001	.0000	.0000	-.0003	-.0011	-.0038	-.0388	-.1436
II	.0000	.0000	.0000	.0000	.0000	-.0000	.0000	-.0352	-.1290
III	.0000	.0000	.0000	.0000	.0000	-.0009	-.0030	-.0382	-.1440

TABLE 3.4, CONTINUED

[n = 10]

ϕ	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.9
					GLS (F^1)				
					$\ell = 20$				
I	-.1208	.0013	.0009	.0000	-.0021	-.0079	-.0357	-.2174	-.0535
II	-.0168	.0014	.0010	.0000	-.0022	-.0084	-.0398	-.2995	-.0035
III	-.0095	.0014	.0010	.0000	-.0022	-.0082	-.0373	-.2952	-.0389
					$\ell = 30$				
I	-.0015	.0007	.0005	-.0000	-.0011	-.0039	-.0142	-.1641	-.2461
II	.0006	.0007	.0005	-.0000	-.0011	-.0040	-.0148	-.1774	-.2247
III	.0006	.0007	.0005	-.0000	-.0011	-.0039	-.0145	-.1687	-.2419
					Ratios (F^2)				
					$\ell = 20$				
I	-.0030	-.0000	.0000	-.0000	-.0004	-.0018	-.0077	-.0799	-.1310
II	-.0000	.0000	.0000	.0000	-.0000	-.0000	-.0017	-.0729	-.0862
III	.0000	.0000	.0000	.0000	-.0000	-.0008	-.0077	-.0804	-.1137
					$\ell = 30$				
I	.0000	.0000	.0000	-.0000	-.0000	-.0005	-.0025	-.0307	-.1356
II	-.0000	.0000	.0000	.0000	.0000	-.0000	.0000	-.0184	-.1164
III	-.0000	.0000	.0000	.0000	.0000	-.0001	-.0029	-.0316	-.1377
					Common Denominator Ratios (F^3)				
					$\ell = 20$				
I	-.0038	.0004	.0002	.0000	-.0006	-.0021	-.0084	-.0860	-.1310
II	-.0000	-.0000	-.0000	.0000	.0000	-.0000	.0007	-.0896	-.1032
III	.0000	.0000	.0000	.0000	-.0000	-.0010	-.0084	-.1084	-.1136
					$\ell = 30$				
I	.0002	.0002	.0001	-.0000	-.0003	-.0010	-.0035	-.0339	-.1401
II	.0000	.0000	-.0000	.0000	-.0000	-.0000	.0000	-.0271	-.1342
III	.0000	.0000	.0000	.0000	-.0000	-.0001	-.0029	-.0356	-.1444

CHAPTER 4

REGRESSION IN FRACTIONAL COINTEGRATION

1. INTRODUCTION

It is now well established that many economic time series contain a unit root. Such series are I(1) in the sense that they are nonstationary but their first difference is stationary. Such series can move in random directions over arbitrarily long time periods. However, in some cases economic theory indicates that certain pairs of series are related and should not diverge too much from each other. This may be reasonable because, when certain economic variables begin to diverge, market forces or government intervention may reestablish their long run relationship.

Suppose that $\{x_t\}$ and $\{y_t\}$ are nonstationary unit root processes where y_t is a scalar but x_t may be a vector. Consider a linear combination of those processes:

$$z_t = y_t - x_t'A,$$

where A is a nonrandom vector. Generally such a linear combination z_t will also be a unit root process. As long as z_t is a unit root process, whether A is zero or not does not make much difference (as we will see later) and we need first-differencing to deal with such cases.

However, when a linear combination of the unit root processes y_t and x_t is an I(d) process with $d < 1$, it is said that $\{x_t\}$ and $\{y_t\}$ are '<u>cointegrated</u>' and A is a <u>cointegrating vector (or coefficient)</u>; see Engle and Granger (1987). An alternative definition of cointegration is that a linear combination of the unit root processes y_t and x_t is stationary. This would rule out the case $1/2 \le d < 1$. <u>Cointegration</u> implies that although there are permanent changes in the individual series x and y over time, there is some long-run equilibrium relationship tying them together, which is represented by the linear combination z_t.

There are several standard examples of cointegration relationships. Davidson, Hendry, Srba and Yeo (1978) show that even though both consumption and income are unit root processes, in the long run the difference between the log of consumption and the log of income appears to be a stationary process. Kremers (1989) proposes that the difference between the log of government debt and the log of GNP is a stationary process even though each is not stationary. Also, although many empirical studies show significant deviations from the Purchasing Power Parity (PPP) hypothesis in the short run, it is argued that the PPP hypothesis works in the long run, in the sense that a cointegration relationship exists among the foreign price index, the domestic price index and the nominal exchange rate; alternatively between a relative price index and the nominal exchange rate. For further details, see Cheung and Lai (1993) and Baillie and Selover (1987).

The standard statistical treatments of cointegration deal with the case of a short memory error; i.e., a linear combination of nonstationary I(1) processes becomes a stationary I(0) process. Under short memory error, the properties of cointegrating coefficient estimates are well known. OLS is consistent and converges in probability at the rate of T rather than the usual rate of $T^{1/2}$. However, in general OLS is asymptotically biased, and it does not lead to asymptotically valid inference. There are many other efficient estimates of cointegrating coefficients. For example, see Johansen (1988, 1991), Stock and Watson (1988, 1993), Phillips and Hansen

(1990), Saikkonen (1991) and Park (1992). These methods also lead to asymptotically valid inference.

However, it is possible that the linear combination of nonstationary unit root processes may be an I(d) process with 0< d <1. This case is referred to as fractional cointegration in Baillie and Bollerslev (1994) and Cheung and Lai (1993). If the error in the cointegrating relationship is I(d) with 0< d <1/2, it is still stationary but it has more persistent autocorrelations than in the usual short memory case. If 1/2< d <1, the error in the cointegrating relationship is not stationary but it is mean-reverting, so that a shock in a given time period will finally disappear in the long run.

Therefore, in this chapter our interest is in the case that the error in a cointegrating relationship is I(d) with 0< d <1, rather than in the usual model of cointegration with errors that are I(0). For this case, we derive the asymptotic distribution of the least squares estimator. Least squares is consistent, and has a rate of convergence to its asymptotic distribution that depends on d. This can be compared to least squares in differences, which is not consistent if the errors and regressors are correlated, and which converges at the usual $T^{1/2}$ rate for all values of d in the range 0< d <1. We also provide some simulations that support the relevance of these asymptotics in samples of moderate size.

2. COINTEGRATION

First we define cointegration in a way similar to that in Engle and Granger (1987).

Definition of Cointegration: The components of an Nx1 vector z_t are said to be cointegrated of order a, b, i.e. $z_t \sim CI(a,b)$, if

(i) z_t is I(a) for a >1/2 and

(ii) there exists a non-zero vector α such that $\alpha' z_t \sim I(a-b)$, $a \geq b > 0$.

If z_t has more than 2 components ($N \geq 2$), then we may have more than one cointegrating vector. When there exist h linearly independent cointegrating vectors with $h \leq N-1$, we can combine them to make a cointegrating matrix C (Nxh). The rank of C is h and is called as the cointegrating rank.

There are at least two reasons why cointegration is important. First, in a regression with nonstationary variables, cointegration is a useful way of distinguishing a meaningful regression from a 'nonsense' (Yule (1926)) or 'spurious' (Granger and Newbold (1978)) regression. In a spurious regression the error is I(1) and least squares does not have useful properties. Second, the Error Correction Representation (ECR) exists only when the nonstationary variables are cointegrated. This is important since the ECR provides a sensible way of combining the information contained both in levels and differences. The ECR models the dynamics of both short-run changes and the long-run adjustment process simultaneously.

In the short memory case (i.e., a = b =1), the problem of estimation of the cointegrating vector has been studied by many economists. OLS is consistent and converges at rate T, which is faster than the usual $T^{1/2}$ rate; see Stock (1987) and Phillips and Durlauf (1986). Phillips and Park (1988) showed that when the error in cointegrating relationship follows a stationary AR process, OLS and Generalized

Least Square (GLS) are asymptotically equivalent. There are many asymptotically efficient estimates, which also lead to asymptotically valid inference. Johansen (1988) derived a maximum likelihood estimator of the dimension of the space of the cointegrating vectors and tests of linear hypotheses on those vectors. Phillips and Hansen (1990) suggested a 'fully modified' least squares estimator. The method of adding leads and lags was suggested by Saikkonen (1991), Phillips and Loretan (1991) and Stock and Watson (1993). Park (1992) proposed an OLS procedure after transforming both regressors and dependent variables.

There are many useful ways of representing cointegrated variables, including the ECR. The Vector Autoregressive Representation (VAR) is a basic tool for analyzing nonstationary variables by making them stationary through first-differencing. For details, see Engle and Yoo (1987) and Ogaki and Park (1992). Johansen (1988) provided the Interim Multiplier Representation (IMR) by modifying the error correction representation. The Triangular Representation (TR) by Phillips (1991) divides the cointegrated system into exactly cointegrated variables and other non-cointegrated variables. The Common Trend Representation (CTR) in Stock and Watson (1988) decomposes the cointegrated nonstationary system into a stationary component plus linear combinations of common deterministic trends and common random walk variables. The Granger Representation Theorem in Engle and Granger (1987) and Johansen (1991) gives several interesting results on the representation of the cointegrated system.

3. ASYMPTOTICS FOR OLS ESTIMATES OF COINTEGRATING COEFFICIENTS

We consider the following data generating process:

$$y_t = x_t\beta + u_t,\quad t = 1, 2, \ldots, T, \tag{1}$$

$$x_t = x_{t-1} + v_t, \tag{2}$$

$$(1-L)^d u_t = \varepsilon_t, \tag{3}$$

where $\{x_t\}$ and $\{y_t\}$ are the observed unit root series and $\{v_t\}$ and $\{\varepsilon_t\}$ are assumed to be short memory processes. For simplicity we consider the case that x_t is a scalar; thus there is at most one cointegrating relationship between y_t and x_t. In general, the error process $\{u_t\}$, which is a linear combination of unit root processes, may be another unit root process. However, when β is not zero and $\{u_t\}$ is an I(d) process with $0 \leq d < 1$, $\{x_t\}$ and $\{y_t\}$ are cointegrated as in Engle and Granger (1987). This model does not include intercept or deterministic time trend. To do so is a feasible but non-trivial extension of this analysis.

A. Short Memory Case (d = 0)

We first consider the case that $d = 0$ in (3) above. Therefore $u_t = \varepsilon_t$ in (3) above, and we have the standard case of cointegration considered in the literature. We will give a brief summary of the results for this case, for purposes of comparison with our results for the case of fractional cointegration.

For simplicity, and to ensure comparability with our treatment of the fractional case, we consider the special case in which the errors $\{v_t\}$ and $\{\varepsilon_t\}$ are only contemporaneously correlated; i.e., $\begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} \sim$ iid $(0, \Sigma)$ with $\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{v\varepsilon} \\ \sigma_{v\varepsilon} & \sigma_v^2 \end{bmatrix}$.

We begin with the following lemma which can be obtained from Phillips (1988) and Phillips and Durlauf (1986):

LEMMA 1:

(i) $\frac{1}{T}\sum_t^T x_{t-1}\varepsilon_t \Rightarrow \int_0^1 B_1(r)dB_2(r)$,

(ii) $\frac{1}{T^2}\sum_t^T x_t^2 \Rightarrow \int_0^1 B_1^2(r)dr$,

(iii) $\frac{1}{T}\sum_t^T v_t\varepsilon_t \xrightarrow{p} \sigma_{v\varepsilon}$,

where $B(r) \equiv \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix}$ = Brownian motion with covariance matrix Σ.

We note that, in a more general setting, the Brownian motion B(r) would have as its covariance matrix the "long-run covariance matrix" of v_t and ε_t, defined as:

$$\Omega = \lim_{T\to\infty} E\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_t^T v_t \\ \frac{1}{\sqrt{T}}\sum_t^T \varepsilon_t \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_t^T v_t \\ \frac{1}{\sqrt{T}}\sum_t^T \varepsilon_t \end{bmatrix}'.$$

However, because we have assumed that v_t and ε_t are iid, the long-run covariance matrix Ω and the contemporaneous covariance matrix Σ are equal.

The OLS estimate of the cointegrating coefficient β is:

$$(4)\qquad \hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t \varepsilon_t}{\sum_{t=1}^{T} x_t^2}.$$

Thus

$$(5)\qquad T(\hat{\beta} - \beta) = \frac{T^{-1}\sum x_{t-1}\varepsilon_t}{T^{-2}\sum x_t^2} + \frac{T^{-1}\sum v_t\varepsilon_t}{T^{-2}\sum x_t^2},$$

since $\sum x_t\varepsilon_t = \sum x_{t-1}\varepsilon_t + \sum v_t\varepsilon_t$. Then, using *LEMMA* 1, one obtains the following asymptotic result:

$$(6)\qquad T(\hat{\beta} - \beta) \Rightarrow \frac{\int_0^1 B_1(r)dB_2(r) + \sigma_{v\varepsilon}}{\int_0^1 B_1^2(r)dr}.$$

Therefor $\hat{\beta} - \beta$ is $O_p(1/T)$ when the error u_t is short memory. OLS is consistent whether or not v_t and ε_t are correlated (i.e., whether or not $\sigma_{v\varepsilon} = 0$), but there is a bias in the asymptotic distribution when $\sigma_{v\varepsilon} \neq 0$. These are well-known, standard results.

B. Spurious Regression Case (d =1)

We next consider the case of a spurious regression, as defined in Granger and Newbold (1978), for which a rigorous asymptotic analysis was given by Phillips (1986). This is the case in which the error process u_t in (1) above is I(d) with d = 1; i.e., a unit root process. Thus we write:

$$u_t = u_{t-1} + \eta_t ,$$

where η_t is I(0). We will consider that $\{v_t\}$ and $\{\eta_t\}$ are only contemporaneously correlated, so that $\begin{bmatrix} \eta_t \\ v_t \end{bmatrix} \sim \text{ iid } (0,\ \Sigma_1),\ \Sigma_1 = \begin{bmatrix} \sigma_\eta^2 & \sigma_{v\eta} \\ \sigma_{\eta v} & \sigma_v^2 \end{bmatrix}$.

Then the following lemma can be obtained from Philips (1986):

LEMMA 2: Define $\begin{bmatrix} B_1(r) \\ B3(r) \end{bmatrix}$ = Brownian motion with covariance matrix Σ_1.

Then

$$\frac{1}{T^2}\sum_t^T x_t u_t \Rightarrow \int_0^1 B_1(r) B_3(r) dr .$$

Again, in a more general setting the Brownian motion in *LEMMA* 2 would have as its covariance matrix of v_t and η_t, say Ω_1 defined as

$$\Omega_1 = \lim_{T\to\infty} E\left[\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_t^T v_t \\ \frac{1}{\sqrt{T}}\sum_t^T \eta_t \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_t^T v_t \\ \frac{1}{\sqrt{T}}\sum_t^T \eta_t \end{bmatrix}' \right] .$$

However, in the iid case Ω_1 and Σ_1 are the same.

Using *LEMMA* 1 and *LEMMA* 2, we obtain Phillips' result:

$$(7) \qquad (\hat{\beta}-\beta)=\frac{\frac{1}{T^2}\sum_t^T x_t u_t}{\frac{1}{T^2}\sum_t^T x_t^2} \Rightarrow \frac{\int_0^1 B_1(r)B_3(r)dr}{\int_0^1 B_1^2(r)dr}.$$

This result implies that $\hat{\beta}$ is not consistent, since $(\hat{\beta}-\beta)$ is $O_p(1)$ and therefore does not diminish as $T\to\infty$.

The case discussed differs slightly from the spurious regression in Granger and Newbold (1978), in that β is not necessarily zero. However, this is not important, since the asymptotic distribution of $(\hat{\beta}-\beta)$ does not depend on β.

C. Stationary Long Memory Error Case (0< d <1/2)

We now turn to the case that the error u_t is I(d) $0< d <1/2$, so that it is a stationary, long memory process. This is the leading case of fractional cointegration. An asymptotic analysis of OLS or other methods of estimation of the cointegrating vector β has not previously been done.

As above, we consider the case that the innovations $(v_t, \varepsilon_t)'$ are iid $(0, \Sigma)$, with Σ as in section 3.A above. We have the following (standard) result for the joint convergence of partial sums of v_t and ε_t:

$$(8) \qquad \frac{1}{\sqrt{T}}\begin{bmatrix}\sum_{t=1}^{[rT]} v_t \\ \sum_{t=1}^{[rT]} \varepsilon_t\end{bmatrix} \Rightarrow B(r)=\begin{bmatrix}B_1(r)\\ B_2(r)\end{bmatrix},$$

where $B(r)$ is a Brownian motion with covariance matrix Σ. The basic result that we need, however, is an expression for the joint limit of the partial sums of v_t and u_t. The problem is the need for a joint limit. The marginal limiting distributions follow from existing results in the literature. For v_t, we have

$$(9) \qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]} v_t \Rightarrow B_1(r),$$

where $B_1(r)$ is a Brownian motion with variance σ_v^2; this is the marginal statement corresponding to (8) above. For u_t, we have convergence to a fractional Brownian motion, as given by equation (27) of Chapter 1:

$$(10) \qquad \frac{1}{T^{d+1/2}}\sum_{t=1}^{[rT]} u_t \Rightarrow \omega_d W_d(r),$$

where $\omega_d^2 = \sigma_\varepsilon^2 \Gamma(1-2d)/[(1+2d)\Gamma(1+d)\Gamma(1-d)]$ and $W_d(r)$ is the solution to

$$(11) \qquad W_d(r) = \frac{1}{\Gamma(d+1)}\int_0^r (r-s)^d \, dW(s),$$

with $W(s)$ a standard Wiener process.

With these marginal results in hand, the only question is how to express the joint result so that it properly reflects the covariance between the two limiting processes. This covariance is also reflected in the covariance between $B_1(r)$ and $B_2(r)$ in (8) above. Thus the standard Wiener process $W(d)$ is in (11) above should in fact be the specific process $\sigma_\varepsilon^{-1}B_2(r)$, to capture this covariance. More specifically, define

$$(12) \quad F_d(r) = \frac{\sqrt{k(d)}}{\Gamma(1+d)} \int_0^r (r-s)^d \, dB_2(s),$$

where $k(d) = \omega_d^2 / \sigma_\varepsilon^2 = \Gamma(1-2d)/[(1+2d)\Gamma(1-d)\Gamma(1+d)]$. Then we have the joint convergence result as given in the following lemma.

LEMMA 3: Suppose that $(v_t, \varepsilon_t)'$ are iid $(0, \Sigma)$, that $[B_1(r), B_2(r)]'$ is a Brownian motion with covariance matrix Σ, that $F_d(r)$ is the fractional Brownian motion defined in (12), and that the model (1)-(3) above holds with $0< d <1/2$. Then

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} v_t \\ \frac{1}{T^{d+1/2}} \sum_{t=1}^{[rT]} u_t \end{bmatrix} \Rightarrow \begin{bmatrix} B_1(r) \\ F_d(r) \end{bmatrix}.$$

Perhaps surprisingly, the joint convergence result of *LEMMA* 3 for a vector of ordinary and fractional Brownian motions does not seem to exist in the statistical literature. Our argument leading to *LEMMA* 3 was somewhat heuristic, but we believe that it captures the essential ideas that would be part of a more rigorous proof. In any case, with *LEMMA* 3 in hand we can proceed to the analysis of least squares for the fractionally integrated model with $0< d <1/2$.

LEMMA 4: Let the same conditions hold as in *LEMMA* 3. Then

$$\frac{1}{T^{1+d}} \sum_{t}^{T} x_t u_t \Rightarrow \int_0^1 B_1(r) dF_d(r).$$

Proof: See Appendix. ■

Therefore, under the conditions of *LEMMA* 3, we have (using *LEMMA* 1, part (ii), and *LEMMA* 4) the following result for the asymptotic distribution of the OLS estimate $\hat{\beta}$ in (1):

$$(14)\quad T^{(1-d)}(\hat{\beta}-\beta)=\frac{\frac{1}{T^{1+d}}\sum_{t}^{T}x_t u_t}{\frac{1}{T^2}\sum_{t}^{T}x_t^2}\Rightarrow\frac{\int_0^1 B_1(r)dF_d(r)}{\int_0^1 B_1^2(r)dr}.$$

Thus, for the case that the errors in the cointegrating relationship are I(d) with $0< d <1/2$, $\hat{\beta}-\beta$ is $O_p(1/T^{1-d}) = O_p(T^{d-1})$. In particular, the value of d affects the order in probability of the OLS estimate.

D. Nonstationary Long Memory Error Case (1/2< d <1)

We now suppose that u_t, the error in the cointegrating relationship (1), is I(d) with $1/2< d <1$. Define $d^* = d-1$, so $-1/2< d^* <0$. Then $\Delta u_t \equiv p_t$ is I(d^*). Let the innovations be represented by ε_t as in (3) above, so that

$$(15)\quad (1-L)^d u_t = (1-L)^{d^*} p_t = \varepsilon_t.$$

As in section C, we assume that the innovations $(v_t, \varepsilon_t)'$ are iid $(0, \Sigma)$, so that the partial sums of v_t and ε_t converge jointly to $[B_1(r), B_2(r)]'$ as in (8) above. We define:

$$(16)\quad F_{d^*}(r)=\frac{\sqrt{k(d^*)}}{\Gamma(1+d^*)}\int_0^r (r-s)^{d^*}dB_2(s),$$

corresponding to (12) above. Then

$$(17) \qquad \begin{bmatrix} \dfrac{x_{[rT]}}{\sqrt{T}} \\ \dfrac{u_{[rT]}}{T^{d^*+1/2}} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} v_t \\ \dfrac{1}{T^{d^*+1/2}} \sum_{t=1}^{[rT]} p_t \end{bmatrix} \Rightarrow \begin{bmatrix} B_1(r) \\ F_{d^*}(r) \end{bmatrix},$$

which is similar to the result of *LEMMA* 4.

LEMMA 5: Under the assumptions listed above in this section,

$$\frac{1}{T^{d^*+2}} \sum_t^T x_t u_t = \frac{1}{T^{d+1}} \sum_t^T x_t u_t \Rightarrow \int_0^1 B_1(r) F_{d^*}(r) dr .$$

Proof: See Appendix. ■

From *LEMMA* 1 and *LEMMA* 5, we therefore have the following result for the asymptotic distribution of the OLS estimate $\hat{\beta}$:

$$(18) \qquad T^{(1-d)}(\hat{\beta} - \beta) = \frac{1}{T^{d^*}}(\hat{\beta} - \beta) = \frac{\dfrac{1}{T^{d^*+2}} \sum_t^T x_t u_t}{\dfrac{1}{T^2} \sum_t^T x_t^2} \Rightarrow \frac{\int_0^1 B_1(r) F_{d^*}(r) dr}{\int_0^1 B_1^2(r) dr} .$$

Thus, for the case that the errors in the cointegrating relationship are I(d) with 1/2< d <1, $\hat{\beta} - \beta$ is $O_p(1/T^{1-d}) = O_p(T^{d-1})$.

E. Remarks

We have considered the regression of y_t on x_t, where y_t and x_t are I(1), and where the error is I(d). We have considered the cases: d = 0, d = 1, 0< d <1/2, and 1/2< d <1. These correspond to all values of d in [0, 1] except d = 1/2, for which the necessary convergence result for partial sums of an I(d) process is apparently not available. While the form of the asymptotic distribution varies across cases, we have the interesting result that $\hat{\beta}-\beta$ is $O_p(1/T^{1-d})$ for all d in [0, 1] (except perhaps d=1/2). Our findings can be compared with the result of Cheung and Lai (1993). They showed that $T^{(1-d-\delta)}(\hat{\beta}-\beta)$ converges in probability to 0 for all $\delta > 0$. Our findings confirm their results and provide the exact asymptotics of the OLS estimates in fractional cointegration relationships.

It is interesting to compare these results to those for another simple estimator; namely, least squares in first differences. Thus suppose that (1) is differenced to yield

$$(19) \quad \Delta y_t = \Delta x_t \beta + \Delta u_t$$

and a least-squares estimator

$$(20) \quad \tilde{\beta} = \frac{\sum_{t=2}^{T} \Delta x_t \Delta y_t}{\sum_{t=2}^{T} \Delta x_t^2} = \beta + \frac{\sum_{t=2}^{T} \Delta x_t \Delta u_t}{\sum_{t=2}^{T} \Delta x_t^2}.$$

If u_t is I(d) with $0< d <1$, then Δu_t is I(d*) with $-1< d^* <0$. From Odaki (1993), this is a stationary and invertible process. Since Δx_t is also a stationary and invertible process, standard results indicate the following. First, $T^{-1}\sum \Delta x_t^2$ converges in probability to $\gamma_{xx} \equiv E(\Delta x_t^2)$. Second, if $\gamma_{xu} \equiv E(\Delta x_t \Delta u_t)$, then $T^{-1}\sum \Delta x_t \Delta u_t$ converges in probability to γ_{xu}, and $T^{-1/2}\sum(\Delta x_t \Delta u_t - \gamma_{xu})$ is asymptotically normal with zero mean. This implies that $\tilde{\beta}$ converges in probability to $\beta_* \equiv \beta + \gamma_{xu}/\gamma_{xx}$, and $\sqrt{T}(\tilde{\beta} - \beta_*)$ is asymptotically normal with zero mean. Thus $\tilde{\beta} - \beta_*$ is $O_p(T^{-1/2})$. Thus, unlike the OLS estimator in levels ($\hat{\beta}$), $\tilde{\beta}$ is inconsistent when x_t and u_t are correlated. Also unlike $\hat{\beta}$, the rate of convergence of $\tilde{\beta}$ does not depend on d. $\tilde{\beta}$ converges faster than $\hat{\beta}$ when d >1/2 (since 1/2> 1-d) but slower than $\hat{\beta}$ when d <1/2. Thus differencing is a poor idea when d <1/2, but it may be a good idea when d >1/2.

4. SIMULATION RESULTS

In this section we provide some simulation results that support the relevance of our asymptotic results of the previous section. The data are generated according the equations (1)-(3) above. We choose $\beta = 1$ but this choice is not substantive. We also choose $\sigma_{v\varepsilon} = 0$ so that the v_t and ε_t processes are not correlated, even contemporaneously. The sample sizes considered are T = 50, 100, 250, 500, 1000 and 1500. The number of iterations in the simulation was 10000. The computations were done in FORTRAN, using the normal random number generator GASDEV/RAN3 as in Chapter 2.

Table 4.1 gives results for the OLS estimator of β. It presents the mean, the standard deviation, and the standard deviation multiplied by T^{1-d}. Since asymptotically the least squares estimator has estimation error that is $O_p(T^{d-1})$, we expect the normalized standard deviation to approach a limit as T increases. This appears to be true in Table 4.1, and in fact the normalized standard deviation does not change much over the range from T = 100 to T = 1500. This supports the relevance of our asymptotic theory, even for only moderate sample sizes. All of the estimates are essentially unbiased, as would be expected given the strict exogeneity of the regressors. As d increases with T fixed, the standard deviation of the estimate increases and the normalized standard deviation of the estimate decreases.

Table 4.2 gives similar results, for a smaller set of values of d, for the estimate of β obtained by least squares in differences. Once again the estimates are essentially unbiased. Now the normalized standard deviation is the standard deviation multiplied by $T^{1/2}$, since asymptotically least squares in differences has an estimation error that is $O_p(T^{-1/2})$. The relevance of the asymptotic theory is supported again, since the normalized standard deviation is more or less constant over different values of T for any given d. For given T, the standard deviation of the estimate does not depend strongly on d.

Comparing results in Tables 4.1 and 4.2, we see that, in terms of the standard deviation of the estimates, least squares in levels dominates least squares in differences for $d < .5$, while the opposite is true for $d > .5$. The estimators have similar variability when d is close to .5, but the difference between them increases as d moves away from .5 in either direction. This result is also as expected from the asymptotics, based on the differing rates of convergence of the two estimators.

5. CONCLUDING REMARKS

This section has considered the case of fractional cointegration, defined as the case in which a set of variables is I(1) but the regression error is I(d), d <1. This case is empirically relevant, and very little is previously known about the properties of estimates of the regression. We have derived the asymptotic distribution of the ordinary least squares estimate, under a fairly strong set of assumptions, and performed simulations that support the relevance of the asymptotic theory.

We assume that similar results would hold under weaker assumptions. In particular, it would be worthwhile to extend these results to a more general model in which there are multiple regressors, possibly including intercept and trend, and in which the innovations are a general short memory process rather than white noise.

The results that we have derived are similar to the results for the usual cointegration model, in that least squares is consistent, but the asymptotic distribution is not necessarily centered at zero, and there is no reason to think that the estimator is efficient or that it leads to asymptotically valid inference. In the cointegration literature, these findings for the least squares estimator were followed by a large volume of research that established asymptotically efficient estimators and asymptotically valid methods of inference. The same considerations should apply to the case of fractional cointegration, and this would appear to be a valuable future line of research.

TABLE 4.1

Mean and Standard Deviation of OLS

Mean

	d=.0	d=.1	d=.3	d=.49	d=.51	d=.7	d=.9	d=1.0
T=50	1.0000	1.0019	.9998	.9992	1.0008	.9934	.9992	.9998
T=100	1.0000	1.0005	.9999	1.0000	.9994	1.0036	.9997	1.0037
T=250	.9999	1.0001	.9997	1.0002	1.0009	1.0022	.9970	1.0010
T=500	1.0000	1.0001	1.0001	.9998	.9993	1.0013	1.0005	.9873
T=1000	1.0001	1.0001	1.0001	.9999	.9994	.9993	.9995	.9942
T=1500	1.0000	1.0000	1.0001	.9996	1.0002	.9994	.9957	.9982

Standard Deviation

	d=.0	d=.1	d=.3	d=.49	d=.51	d=.7	d=.9	d=1.0
T=50	.0668	.0757	.1092	.1612	.1703	.2703	.4694	.6517
T=100	.0326	.0411	.0675	.1125	.1197	.2181	.4339	.6295
T=250	.0130	.0177	.0350	.0705	.0766	.1658	.3989	.6230
T=500	.0066	.0094	.0216	.0494	.0546	.1335	.3689	.6255
T=1000	.0033	.0052	.0133	.0345	.0389	.1081	.3356	.6355
T=1500	.0022	.0036	.0098	.0282	.0319	.0972	.3298	.6341

$T^{1-d}\cdot$ Standard Deviation

	d=.0	d=.1	d=.3	d=.49	d=.51	d=.7	d=.9	d=1.0
T=50	3.34	2.56	1.69	1.19	1.16	.874	.694	.652
T=100	3.26	2.59	1.70	1.18	1.14	.868	.688	.630
T=250	3.25	2.55	1.67	1.18	1.15	.869	.693	.623
T=500	3.30	2.52	1.67	1.18	1.15	.861	.687	.626
T=1000	3.30	2.61	1.67	1.17	1.15	.859	.670	.636
T=1500	3.30	2.60	1.64	1.18	1.15	.872	.685	.634

TABLE 4.2

Mean and Standard Deviation of OLS in Differences

Mean

	d=.1	d=.3	d=.49	d=.51	d=.7	d=.9	d=1.0
T=50	1.0029	1.0072	1.0000	1.0007	1.0028	1.0004	1.0003
T=100	1.0016	1.0012	1.0012	.9987	1.0016	1.0007	1.0008
T=250	.9997	1.0003	1.0005	.9999	.9991	1.0014	.9999
T=500	.9989	.9997	.9998	1.0001	1.0006	1.0002	.9995
T=1000	.9996	.9997	1.0009	1.0002	.9996	1.0002	.9995
T=1500	.9995	.9999	.9997	1.0003	1.0006	1.0001	.9999

Standard Deviation

	d=.1	d=.3	d=.49	d=.51	d=.7	d=.9	d=1.0
T=50	.1992	.1828	.1694	.1690	.1560	.1505	.1470
T=100	.1394	.1260	.1156	.1160	.1091	.1033	.1017
T=250	.0856	.0773	.0717	.0715	.0666	.0644	.0638
T=500	.0614	.0545	.0508	.0503	.0468	.0453	.0447
T=1000	.0432	.0390	.0360	.0356	.0340	.0318	.0316
T=1500	.0350	.0320	.0296	.0292	.0274	.0258	.0259

$T^{1/2}\cdot$ Standard Deviation

	d=.1	d=.3	d=.49	d=.51	d=.7	d=.9	d=1.0
T=50	1.41	1.29	1.20	1.20	1.10	1.06	1.04
T=100	1.39	1.26	1.16	1.16	1.09	1.03	1.02
T=250	1.35	1.22	1.13	1.13	1.05	1.02	1.01
T=500	1.37	1.22	1.14	1.12	1.05	1.01	1.00
T=1000	1.37	1.23	1.14	1.13	1.08	1.01	1.00
T=1500	1.36	1.24	1.15	1.13	1.06	1.00	1.00

APPENDIX

Proof of *LEMMA* 4: Note that x_t is the partial sum of the innovations v_t. Define Z_t to be the partial sum of the u_t: $Z_t = \sum_{j=1}^{t} u_j$. For $r \in [0, 1]$, define the sample version of the processes $B_1(r)$ and $F_d(r)$ as follows:

$$x_T(r) = \frac{1}{\sqrt{T}} x_{[rT]} = \begin{cases} 0, & r < \frac{1}{T} \\ \frac{1}{\sqrt{T}} x_t, & \frac{t}{T} \le r < \frac{t+1}{T} \\ \frac{1}{\sqrt{T}} x_T, & r = 1 \end{cases}$$

$$Z_T(r) = \frac{1}{T^{d+1/2}} Z_{[rT]} = \begin{cases} 0, & r < \frac{1}{T} \\ \frac{1}{\sqrt{T}} Z_t, & \frac{t}{T} \le r < \frac{t+1}{T} \\ \frac{1}{\sqrt{T}} Z_T, & r = 1. \end{cases}$$

Then

$$\begin{bmatrix} x_T(r) \\ Z_T(r) \end{bmatrix} \Rightarrow \begin{bmatrix} B_1(r) \\ F_d(r) \end{bmatrix}$$

$$\int_0^1 x_T(r) dZ_T(r) \Rightarrow \int_0^1 B_1(r) dF_d(r).$$

Thus,

$$\int_0^1 x_T(r) dZ_T(r) = \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} x_{t-1}\right)\left(\frac{1}{T^{d+1/2}} u_t\right) \qquad \text{(e.g., Phillips (1986, p327))}$$

$$= \frac{1}{T^{1+d}} \sum_{t}^{T} x_{t-1} u_t \Rightarrow \int_0^1 B_1(r) dF_d(r).$$

But $\frac{1}{T^{1+d}} \sum_t x_t u_t = \frac{1}{T^{1+d}} \sum_t x_{t-1} u_t \frac{1}{T^{1+d}} \sum_t v_t u_t$. So we need to show that

$\frac{1}{T^{1+d}} \sum_t v_t u_t \to 0$. To do so, we write

$$\left(\frac{1}{T} \sum_t v_t u_t\right)^2 \leq \left(\frac{1}{T} \sum_t v_t^2\right)\left(\frac{1}{T} \sum_t x_t^2\right)$$

as in Cheung and Lai (1993, p106). Then $\frac{1}{T} \sum_t v_t^2 \to \sigma_v^2$, $\frac{1}{T} \sum_t u_t^2 \to \sigma_u^2$, where the

first result is standard and the second result follows from Hosking (1995). So

$\frac{1}{T} \sum_t v_t u_t$ is bounded in probability and $\frac{1}{T^{1+d}} \sum_t v_t u_t \to 0$ for d >0. ■

Proof of *LEMMA* 5: Define $x_T(r)$ as above, and

$$U_T(r) = \frac{1}{T^{d^*+1/2}} u_{[rT]}.$$

Then

$$\int_0^1 x_T(r) U_T(r) dr \Rightarrow \int_0^1 B_1(r) F_{d^*}(r) dr$$

because of the joint convergence result (17). But

$$\int_0^1 x_T(r) U_T(r) dr = \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} x_T(r) U_T(r) dr$$

$$= \sum_{t=1}^{T} \frac{1}{T} \cdot \frac{1}{\sqrt{T}} x_t \cdot \frac{1}{T^{d^*+1/2}} u_t$$

$$= \frac{1}{T^{d^*+2}} \sum_{t=1}^{T} x_t u_t$$

$$\Rightarrow \int_0^1 B_1(r) F_{d^*}(r) dr \qquad \blacksquare$$

CHAPTER 5

CONCLUSIONS

This dissertation considered the ARFIMA(p,d,q) process, which can apply to many economic time series. The long-run characteristics of an ARFIMA(p,d,q) process, such as stationarity, mean-reversion and persistence of autocorrelations, are determined by the differencing parameter value d. We consider the case that the value of d is in the range $-1< d \leq 1$. If $d = 1$ such a series is a unit root process and if $-1< d <1$ and $d \neq 0$ it is a fractionally integrated process or long memory process. When $d=0$, it is a usual stationary short memory process.

In this dissertation we showed how the KPSS unit root test works in distinguishing long memory processes from unit root processes. Our asymptotic findings indicate that the KPSS unit root test is consistent against stationary long memory alternatives with $-1/2< d <1/2$, but it is not consistent against nonstationary long memory alternatives with $1/2< d <3/2$. This implies that the KPSS statistic can consistently distinguish between short memory processes, stationary long memory processes and nonstationary processes. Dickey-Fuller type tests can consistently distinguish a unit root from an I(d) process with $-1/2< d <1$ but not from an I(d) process with $1< d< 3/2$. Further work is needed on ways to distinguish unit root processes from nonstationary (but mean-reverting) long memory processes.

The estimation of the differencing parameter d is an interesting problem, and there are many different estimators, such as the GPH estimator, the maximum likelihood estimator, the CSS estimator and the MDE. The MDE does not require distributional assumptions and is relatively simple in computation. We considered MDEs including the AMDE of Chung and Schmidt (1995) for the general ARFIMA model. In applying the MDEs, we can estimate the model by letting short-run dynamics follow an ARMA model, or we can estimate the pure I(d) model but omit the first ℓ low-order autocorrelations, which is a nonparametric approach. In this nonparametric method we can expect some bias, especially when the ARMA parameters have extreme values, due to misspecifying the short-run dynamics. We compute the asymptotic bias that results from ignoring a fixed number (ℓ) of low-order autocorrelations in the case of simple ARFIMA(1,d,0) and ARFIMA(0,d,1) processes. The asymptotic bias of the AMDE or the MDE is small when ℓ is large. A derivation of the asymptotic properties of the MDE when ℓ grows with T is an important future task.

Fractional cointegration, defined as the case in which a set of variables is I(1) but the regression error is I(d) with d <1, is empirically important but little is known. We found that OLS is consistent and derived its asymptotic distribution. Its order in probability and asymptotic distribution are affected by the value of d. The asymptotic distribution of the OLS estimate in the case of fractional cointegration is not necessarily centered at zero and there is no reason to think that OLS is efficient or that it leads to asymptotically valid inference. Finding an efficient estimate that leads to asymptotically valid inference is another important topic for further research.

LIST OF REFERENCES

REFERENCES

Agiakloglou, C. and P. Newbold (1992), Empirical Evidence on Dickey-Fuller-Type Test, Journal of Time Series Analysis, 13, 471-483.

Agiakloglou, C., P. Newbold and M. Wohar (1993), Bias in an Estimator of the Fractional Difference Parameter, Journal of Time Series Analysis, 14, 235-246.

Andrews, D.W.K. (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, Econometrica, 59, 817-858.

Baillie, R.T. (1995), Long Memory Processes and Fractional Integration in Econometrics, Journal of Econometrics, forthcoming.

Baillie, R.T. and T. Bollerslev (1994), Cointegration, Fractional Cointegation and Exchange Rate Dynamics, Journal of Finance, 49, 737-745.

Baillie, R.T. and D.D. Selover (1987), Cointegration and Models of Exchange Rate Determination, International Journal of Forecasting, 3, 43-51.

Beran, J.A. (1992), Statistical Methods for Data with Long-Range Dependence, Statistical Science, 7, 404-427.

Beveridge, W.H. (1921), Weather and Harvest Cycles, Economic Journal, 31, 429-452.

Boes, D.C., R.A. Davis and S.N. Gupta (1989), Parameter Estimation in Low Order Fractionally Differenced ARMA Models, Stochastic Hydrology and Hydrolics, 3, 97-110.

Brockwell, P.J. and R.A. Davis (1991), Time Series: Theory and Methods, 2nd edition, Springer-Verlag, New York.

Campbell, J.Y. and N.G. Mankiw (1987), Are Output Fluctuations Transitory?, Quarterly Journal of Economics, 102, 857-880.

Cheung, Y.-W. (1993), Tests for Fractional Integration: A Monte Carlo Investigation, Journal of Time Series Analysis, 14, 331-345.

Cheung, Y.-W. and F.X. Diebold (1994), On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally Interated Noise with Unknown Mean, Journal of Econometrics, 62, 301-316.

Cheung, Y.-W. and K.S. Lai (1993), A Fractional Cointegration Analysis of Purchasing Power Parity, Journal of Business and Economic Statistics, 11, 103-112.

Choi, I and P.C.B. Phillips (1993), Testing for a Unit Root by Frequency Domain Regression, Journal of Econometrics, 59, 263-286.

Chung, C.-F. (1994a), A Note on Calculating the Autocovariances of Fractionally Integrated ARMA Models, Economic Letters, 45, 293-297.

Chung, C.-F. (1994b), Calculating and Analyzing Impulse Responses and Their Asymptotic Distributions for the ARFIMA and VARMA Models, unpublished manuscript, Michigan State University.

Chung, C.-F. and R.T. Baillie (1993), Small Sample Bias in Conditional Sum of Squares Estimators of Fractionally Integrated ARMA models, Empirical Economics, 18, 791-806.

Chung, C.-F. and P. Schmidt (1995), The Minimum Distance Estimator for Fractionally Integrated ARMA Models, unpublished manuscript, Michigan State University.

Cox, D.R. (1984), Long-Range Dependence: A Review, Statistics: An Appraisal, Proceedings 50th Anniversary Conference (Iowa State Statistical Laboratory).

Dalhaus, R. (1989), Efficient Parameter Estimation for Self Similar Processes, Annals of Statistics, 17, 1749-1766.

Davidson, J.E.H., D.F. Hendry, F. Srba and S. Yeo (1978), Econometric Modelling of the Aggregate Time-series Relationship between Consumer's Expenditure and Income in the United Kingdom, Economic Journal, 88, 661-692.

Davydov, Y.U. (1970), The Invariance Principle for Stationary Processes, Theory of Probability and Its Applications, 15, 487-498.

Dejong, D.N., J.C. Nankervis, N.E. Savin and C.H. Whiteman (1992), The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors, Journal of Econometrics, 53, 323-343.

Dickey, D.A. (1976), Estimation and Hypothesis Testing in Nonstationary Time Series, Ph.D. dissertation (Iowa State University, Ames, IA).

Dickey, D.A. and W.A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74, 427-431.

Diebold, F.X. and M. Nerlove (1990), Unit Roots in Economic Time Series: A Selective Survey, Advances in Econometrics, 8, 3-69.

Diebold, F.X. and G.D. Rudebusch (1991), On the Power of Dickey-Fuller Type Tests against Fractional Alternatives, Economics Letters, 35, 155-160.

Durbin, J. (1960), The Fitting of Time Series Models, Review of the International Statistical Institute, 28, 233-243.

Engle, R.F. and C.W.J. Granger (1987), Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 251-276.

Engle, R.F. and B.S. Yoo (1987), Forecasting and Testing in Co-integrated Systems, Journal of Econometrics, 35, 143-159.

Fox, R. and M.S. Taqqu (1986), Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series, Annals of Statistics, 14, 517-532.

Fuller W.A. (1976), Introduction to Statistical Time Series, Wiley, New York.

Geweke, J. and S. Porter-Hudak (1983),The Estimation and Application of Long Memory Time Series Models, Journal of Time Series Analysis, 4, 221-238.

Granger, C.W.J. (1980), Long Memory Relationships and the Aggregation of Dynamic Models, Journal of Econometrics, 14, 227-238.

Granger, C.W.J. and R. Joyeux (1980), An Introduction to Long-Memory Time Series Models and Fractional Differencing, Journal of Time Series Analysis, 1, 115-129.

Granger, C.W.J. and P. Newbold (1974), Spurious Regressions in Econometrics, Journal of Econometrics, 2, 111-120.

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

Hassler, U. and J. Wolters (1994), On the Power of Unit Root Tests against Fractional Alternatives, Economics Letters, 45, 1-5.

Hauser, M.A. (1992), Long Range Dependence in International Output Series: A Reexamination, unpublished manuscript, University of Economics and Business Administartion, Vienna, Austria.

Hosking, J.R.M. (1981), Fractional Differencing, Biometrika, 68, 165-176.

Hosking, J.R.M. (1984), Modeling Persistence in Hydrological Time Series Using Fractional Differencing, Water Resources Research, 20, 1898-1908.

Hosking, J.R.M. (1995), Asymptotic Distributions of the Sample Mean, Autocovariance and Autocorrelations of Long-memory Time Series, Journal of Econometrics, forthcoming.

Hurst, H.E. (1951), Long-term Storage Capacity of Reservoirs, Transactions of the American Society of Civil Engineers, 116, 770-779.

Hurst, H.E. (1956), Methods of Using Long Term Storage in Reservoirs, Proceedings of the Institute of Civil Engineers, 1, 519-543.

Johansen, S. (1988), Statistical Analysis of Cointegrating Vectors, Journal of Economic Dynamics and Control, 12, 231-254.

Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica, 59, 1551-1580.

Kremers, J.J.M. (1989), U.S. Federal Indebtness and the Conduct of Fiscal Policy, Journal of Monetary Economics, 23, 219-238.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root ?, Journal of Econometrics, 54, 159-178.

Lee, D. (1994), Asymptotic Theory for Long-Memory Time Series, Ph.D. dissertation (Michigan State University, East Lansing, MI).

Lee, D. and P. Schmidt (1995), On the Power of the KPSS Test of Stationarity against Fractionally-Integrated Alternatives, Journal of Econometrics, forthcoming.

Levinson, N. (1947), The Wiener RMS Criterion in Filter Design and Prediction, Journal of Mathematics and Physics, 25, 261-278.

Leybourne, B.J. and B.P.M. McCabe (1994), A Consistent Test for a Unit Root, Journal of Business and Economic Statistics, 12, 157-166.

Li, W.K. and A.I. McLeod (1986), Fractional Time Series Modeling, Biometrika, 73, 217-221.

Lo, A.W. (1991), Long-Term Memory in Stock Market Prices, Econometrica, 59, 1279-1313.

Lobato, I. and P.M. Robinson (1993), Averaged Periodogram Estimation of Long Memory, unpublished manuscript, London School of Economics.

Mandelbrot, B.B. (1971), A Fast Fractional Gaussian Noise Generator, Water Resources Research, 7, 543-553.

Mandelbrot, B.B. and J.W. Van Ness (1968), Fractional Brownian Motions, Fractional noises and Applications, SIAM Review, 10, 422-437.

Mandelbrot, B.B. and J.Wallis (1968), Noah, Joseph and Operational Hydrology, Water Resources Research, 4, 909-918.

McLeod, A.I. and K.W. Hipel (1978), Preservation of the Rescaled Adjusted Range 1: A Reassesment of the Hurst Phenomenon, Water Resources Research, 14, 491-508.

Nelson, C.R. and C.I. Plosser (1982), Trends Versus Random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics, 10, 139-162.

Newey, W.K. and K.D. West (1987), A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703-708.

Odaki, M. (1993), On the Invertibility of Fractionally Differenced ARIMA Processes, Biometrika, 80, 703-709.

Ogaki, M. and J.Y. Park (1992), A Cointegration Approach to Estimating Preference Parameters, unpublished manuscipt, University of Rochester.

Pantula, S.G. (1991), Asymptotic Distributions of Unit-Root Tests When the Process is Nearly Stationary, Journal of Business and Economic Statistics, 9, 63-71.

Park, J.Y. (1992), Canonical Cointegrating Regressions, Econometrica, 60, 119-143.

Park, J.Y. and B. Choi (1988), A New Approach to Testing for a Unit Root, Working paper No. 88-23 (Center for Analytical Economics, Cornell University, Ithaca, NY).

Park, J.Y. and P.C.B. Phillips (1989), Statistical Inference in Regressions with Integrated Processes:Part 2, Econometric Theory, 5, 95-131.

Phillips, P.C.B. (1986), Understanding Spurious Regressions in Econometrics, Journal of Econometrics, 33, 311-340.

Phillips, P.C.B. (1987), Time Series Regression with a Unit Root, Econometrica, 55, 277-301.

Phillips, P.C.B. (1988a), Weak Convergence of Sample Covariance Matrix to Stochastic Integrals via Martingale Approximations, Econometric Theory, 4, 528-533.

Phillips, P.C.B. (1988b), Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$, Journal of Multivariate Analysis, 24, 252-264.

Phillips, P.C.B. (1991), Optimal Inference in Cointegrated Systems, Econometrica 59, 283-306

Phillips, P.C.B. and S.N. Durlauf (1986), Multiple Time Series Regression with Integrated Porcesses, Review of Economic Studies, 53, 473-495.

Phillips, P.C.B. and B.E. Hansen (1990), Statistical Inference in Instrumental Variables Regression with I(1) Processes, Review of Economic Studies, 57, 99-125.

Phillips, P.C.B. and J.Y. Park (1988), Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors, Journal of the American Statistical Association, 83, 111-115.

Phillips, P.C.B. and M. Loretan (1991), Estimating Long-Run Economic Equilibria, Review of Economic Studies, 58, 407-436.

Phillips, P.C.B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, Biometrika, 75, 335-346.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1989), Numerical Recipes: the Art of Scientific Computing, Cambridge University Press.

Robinson, P.M. (1994), Time Series with Strong Dependence, Advances in Econometrics, 6th World Congress, vol. I, Cambridge University Press.

Rosenblatt, M (1956), A Central Limit Theorem and a Strong Mixing Condtion, Proceedings of the National Academy of Sciences, 42, 43-47.

Said, S.E. and D.A. Dickey (1984), Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order, Biometrika, 71, 599-608.

Saikkonen, P (1991), Asymptotically Efficient Estimation of Cointegrating Regression, Econometric Theory, 7, 1-21.

Saikkonen, P. and R. Luukkonen (1993), Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models, Journal of the American Statistical Association, 88, 596-601.

Schmidt, P. and P.C.B. Phillips (1992), LM Tests for a Unit Root in the Presence of Deterministic Trend, Oxford Bulletin of Economics and Statistics, 54, 257-287.

Schwert, G.W. (1989), Tests for Unit Roots: A Monte Carlo Investigation, Journal of Business and Economic Studies, 7, 147-159.

Seater, J.J. (1993), World Temperature-Trend Uncertainties and Their Implications for Economic Policy, Journal of Business and Economic Studies, 11, 265-277.

Shin, Y. and P. Schmidt (1992), The KPSS Stationary Test as a Unit Root Test, Economics Letters, 38, 387-392.

Sowell, F. (1990), The Fractional Unit Root Distribution, Econometrica, 58, 495-505.

Sowell, F. (1992a), Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models, Journal of Econometrics, 53, 165-188.

Sowell, F. (1992b), Modelling Long-run Behavior with the Fractional ARIMA model, Journal of Monetary Economics, 29, 2770302.

Stock, J.H. (1987), Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors, Econometrica, 55, 1035-1056.

Stock, J.H. (1990), A Class of Tests for Integration and Cointegration, unpublished manuscript, Harvard University.

Stock, J.H. and M.W. Watson (1988), Testing for Common Trends, Journal of the American Statistical Association, 83, 1097-1107.

Stock, J.H. and M.W. Watson (1993), A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, Econometrica, 61, 783-820.

Tieslau, M.A., P. Schmidt and R.T. Baillie (1995), A Minimum-Distance Estimator for Long Memory Processes, Journal of Econometrics, forthcoming.

Whittle, P. (1951), Hypothesis Testing in Time Series Analysis, Almquist and Wiksells, Uppsala, Sweden.

Whittle, P. (1956), Variation of Yield Variance with Plot Size, Biometrika, 43, 337-343.

Whittle, P. (1963), On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix, Biometrika, 50, 129-134.

Yule, G.U. (1926), Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series, Journal of the Royal Statistical Society, 89, 1-64.