This is to certify that the

dissertation entitled

ASYMPTOTIC THEORY FOR LONG-MEMORY TIME SERIES

presented by

DONGIN LEE

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Peter Schmidt
Major professor

Date June 27, 1994

ASYMPTOTIC THEORY FOR LONG-MEMORY TIME SERIES

By

Dongin Lee

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Economics

1994

ABSTRACT

ASYMPTOTIC THEORY FOR LONG-MEMORY TIME SERIES

By

Dongin Lee

Economic time series are nonstationary rather than stationary around deterministic trends in most cases. Usually nonstationary time series are analyzed by integrated process models.

This dissertation considers a generalized integrated process in the sense that the differencing parameter is allowed to be a fractional value. For $d \in (-1/2, 1/2)$, the I(d) process is stationary and invertible. For $0 < d < 1/2$, the autocorrelations of the I(d) process are positive and decline so slowly that the sum of autocorrelations is infinite in the limit, while for $-1/2 < d < 0$ the autocorrelations of the I(d) process are negative for all lags and the sum of autocorrelations goes to zero. Therefore as long as $d \in (-1/2, 1/2)$ and $d \neq 0$, the standard ARIMA model cannot be applied to the I(d) process.

Chapter 2 considers a stationarity test against I(d) alternatives. Kwiatkowski, Phillips, Schmidt and Shin (KPSS) proposed a test of the null hypothesis of stationarity. It is shown in Chapter 2 that the KPSS test is consistent against an I(d) processes for $d \in (-1/2, 1/2)$. It can therefore be used to distinguish short memory and long memory stationary processes. The simulation results show that a rather large sample size, such as

T = 1000, will be necessary to distinguish reliably between a long memory process and a short memory process with comparable short-term autocorrelation.

Chapter 3 considers the power of Dickey-Fuller unit root tests against I(d) alternatives with $d \in (-0.5, 0.5)$. The Dickey-Fuller tests are shown to be consistent against these alternatives. Simulations show high power of the tests against stationary fractionally integrated alternatives, and reveal some interesting features of the power function at and around the boundary ($d = 0.5$) of the stationary region.

Chapter 4 considers several estimators for the differencing parameter in the I(d) model. Specifically the minimum distance estimator (MDE) suggested by Tieslau, Schmidt and Baillie (1994) is compared to the exact MLE and the approximate MLE of various forms. Both the exact MLE and approximate MLE of d are $\sqrt{T}$-consistent and asymptotically normal for $d \in (-1/2, 1/2)$, while this is true for the MDE only for $d \in (-1/2, 1/4)$. Simulations show that if the mean of the process is unknown, the MDE is comparable to the MLE in a reasonable sized sample when the number of autocorrelations is more than two or three.

ACKNOWLEDGMENTS

First of all, I would like to thank the members of my dissertation committee for their guidance and advice. Without their assistance the completion of this dissertation would not be possible. Especially I would like to express my thanks to Professor Peter Schmidt, the committee chair. Throughout the entire process of writing dissertation he provided helpful guidance and careful comments which were essential to completing this dissertation. Also I would like to thank Professor Richard Baillie who introduced me to the main subjects of this dissertation as well as taught me time series econometrics.

There are so many other people, both professors and graduate students, who have been important for me during the whole program. Some of them are so special to me that I would like to name them. I gratefully acknowledge that I learned statistics from Professor James Stapleton and Professor James Hannan and econometrics from Professor Peter Schmidt, Professor Richard Baillie and Professor Ching-Fan Chung. I am indebted to those professors forever. Also I am grateful that I have met Junsoo Lee, Yongcheol Shin and Kyungso Im, the graduate students who studied econometrics together . Especially, I am grateful to meet Kyungso who has been my classmate and officemate for five years. The tie between us is more than that of officemate or classmate. Without him, studying econometrics would not have been fun and profitable.

Special thanks also goes to the administrative staff in the department. I especially acknowledge Mrs. Ann Feldman who gave me practical advice from time to time which made the whole process of my study smooth and enjoyable.

Finally, I would like to thank my wife, Jaekyung, and my two children, Minyoung and Changwoo. Words would not be enough for their love, understanding and encourangement. Also I wish to express my thanks to my parents for their prayer and moral support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Many macroeconomic times series are nonstationary processes rather than stationary processes around deterministic trends, as first found in Nelson and Plosser (1982). In the recent literature these nonstationary macro series are modeled by integrated processes. Economic theory such as the real business cycle theory, the permanent income-rational expectation theory of consumption or the efficient market hypothesis in financial economics provided theoretical grounds for integrated time series processes.

However, if the first order autocorrelation of the series is too small for an integrated process and the autocorrelations for large lags are too persistent for a stationary ARMA process, it will be hard to decide whether the series is stationary or not. Or, if a series looks like a unit root process (integrated process of order one), but the first differenced series has small negative autocorrelations so the differenced series looks overdifferenced, what is a natural guess for the series? It might be neither a usual short-memory stationary process nor a unit root process.

This dissertation considers an alternative type of series, called long-memory processes. In a typical long-memory process the autocorrelations of the process are persistent, but it is neither a stationary ARMA process, nor is it a nonstationary integrated process. Long-memory persistence in a time series was observed in hydrology and referred to as a "Hurst effect" quite a long time ago. In the mid 1960s it was modeled as a "fractional noise process" proposed by Mandelbrot and Van Ness (1968). In economics, the possibility of long-memory processes was implied in some early literature, for example Granger (1966), and this kind of process was investigated in a formal way in the early 1980s after Granger and Joyeux (1980) and Hosking (1981) provided an alternative definition of the long-memory process. Their process is called a "fractionally integrated process". Granger (1980) provided an argument for the theoretical possibility of a fractionally integrated process in the economic time series. He showed that the aggregated series of heterogeneous but persistent AR(1) processes follows a fractionally integrated process. Geweke and Porter-Hudak (1983) proved that the two classes of long-memory processes (Mandelbrot and Van Ness versus Granger-Joyeux-Hosking) are equivalent. In the dissertation we will follow the definition of the fractionally integrated process of Granger (1980), Granger and Joyeux (1980) and Hosking (1981).

A time series $\{y_t\}$ said to be a fractionally integrated process of order d, or I(d) with zero mean, if it has the following from;

(1) $$(1 - L)^d \, y_t = \varepsilon_t \quad \text{with } d \in (-1/2, 1/2),$$

where L is the lag operator, d is the differencing parameter and ε_t is a white noise process with zero mean and finite variance σ^2. The expression $(1 - L)^d$ is defined by means of the binomial expansion:

(2) $$(1 - L)^d = \sum_{i=0}^{\infty} \pi_i L^i, \quad \pi_i = \Gamma(i-d)/[\Gamma(i+1)/\Gamma(-d)],\ i = 0, 1, 2, \ldots,$$

where $\Gamma(\bullet)$ is the gamma function.

Note that if d is a positive integer, $\{y_t\}$ is a nonstationary integrated process. However, if $d \in (-1/2, 1/2)$, $\{y_t\}$ is stationary and invertible. The AR(∞) and MA(∞) representations of I(d) are as follows:

(3) $$y_t = \sum_{i=1}^{\infty} \phi_i \, y_{t-i} + \varepsilon_t, \quad \phi_i = -\Gamma(i-d)/[\Gamma(i+1)\Gamma(-d)],\ i = 1, 2, 3, \ldots.$$

(4) $$y_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}, \quad \theta_i = \Gamma(i+d)/[\Gamma(i+1)\Gamma(d)],\ i = 0, 1, 2, \ldots.$$

The variance σ_y^2 and autocorrelations ρ_i of the I(d) process are also expressed in terms of gamma functions as follows:

(5) $$\sigma_y^2 = \sigma^2\, \Gamma(1-2d)/\Gamma^2(1-d)$$

(6) $$\rho_i = \{\Gamma(i+d)\Gamma(1-d)\}/\{\Gamma(i-d+1)\Gamma(d)\}$$

$$= \prod_{k=1}^{i} (k-1+d)/(k-d),\ \ i = 1, 2, 3, \ldots.$$

To determine the partial autocorrelations, we write the best linear predictor $\hat{y}_{t+1}$ of y_{t+1} given $y_1, y_2, y_3, \ldots, y_t$ as

$$\hat{y}_{t+1} = \phi_{t1}\, y_1 + \phi_{t2}\, y_2 + \cdots + \phi_{tt}\, y_t,$$

where the coefficient ϕ_{ti} is computed by the Durbin-Levinson algorithm of Levinson (1947), Durbin (1960) and Whittle (1963) as

(7) $$\phi_{ti} = -\binom{t}{i} \{\Gamma(i-d)\, \Gamma(t-d-i+1)\}/\{\Gamma(-d)\Gamma(t-d+1)\},\quad i = 1, 2, \ldots, t.$$

So the partial autocorrelations α_i are as follows:

(8) $$\alpha_i = \phi_{ii} = -\{\Gamma(i-d)\, \Gamma(1-d)\}/\{\Gamma(-d)\Gamma(i-d+1)\} = d/(i-d),\ \ i = 1,2,3,\ldots.$$

Since $\Gamma(x) \sim \sqrt{2\pi}\, e^{-x+1} (x-1)^{x-1/2}$ as $x \to \infty$, we can find the asymptotic behavior of the coefficients in the AR(∞) and MA(∞) representations, and also of the autocorrelations for large lags. Thus

(9) $\phi_i \sim -i^{-d-1}/\Gamma(-d)$ as $i \rightarrow \infty$,

(10) $\theta_i \sim i^{d-1}/\Gamma(d)$ as $i \rightarrow \infty$,

(11) $\rho_i \sim \{i^{2d-1}\,\Gamma(1-d)\}/\Gamma(d)$ as $i \rightarrow \infty$.

Comparing the asymptotic behavior of autocorrelations between an I(d) process and a stationary ARMA process, the autocorrelations of an I(d) process satisfy $\rho_i \sim C_1\, i^{2d-1}$, while the autocorrelations of an ARMA process satisfy $\rho_i \sim C_2\, r^{-i}$, where C_1, C_2, and r are some constants. In other words, the autocorrelations in an ARMA process decrease rapidly (exponentially), while the autocorrelations in an I(d) process decrease very slowly (hyperbolically).

Sometimes the spectral density at zero frequency is used as a measure of persistence in a time series. The spectral density of the I(d) process is,

(12) $f(\lambda) = |1 - z|^{-2d}\,\sigma^2/(2\pi) = |2\sin(\lambda/2)|^{-2d}\,\sigma^2/(2\pi)$, for $-\pi < \lambda < \pi$,

where $z = e^{-i\lambda}$. From Equation (12), f(0) is zero for $d < 0$, and is infinite for $d > 0$. For the case of $d > 0$, since $\sin(\lambda) \sim \lambda$ as $\lambda \rightarrow 0$, asymptotically the behavior of f(0) is as follows:

(13) $f(0) \sim \lambda^{-2d}\,\sigma^2/(2\pi)$ as $\lambda \rightarrow 0$.

We can generalize the I(d) process in such a way that we can apply it to more general times series models for economics data.

A time series $\{y_t\}$ is said to be an autoregressive fractionally integrated moving average process of order p, d, q, or ARFIMA(p,d,q), with zero mean, if it has the following form:

(14) $$\Phi(L)(1 - L)^d y_t = \Theta(L)\varepsilon_t \quad \text{with } d \in (-1/2, 1/2),$$

where $\Phi(L)$ is a p^{th} order lag polynomial of autoregressive parameters, $\Theta(L)$ is a q^{th} order lag polynomial of moving average parameters, and ε_t is a white noise process as before. Furthermore we assume that all the roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit cycle, for stationarity and invertiblity respectively, and also we assume that no roots are common in $\Phi(L)$ and $\Theta(L)$, for identification of the parameters.

This is a generalization of the ARIMA process in the sense that the order of integration is allowed to be a fractional value. Comparing the I(d) process in Equation (1) with the ARFIMA(p,d,q) process in Equation (14), since $(1 - L)^d y_t = [\Theta(L)/\Phi(L)]\, \varepsilon_t \equiv u_t$, where u_t is ARMA(p,q), and since $\Phi(L)y_t = \Theta(L)(1 - L)^{-d} \varepsilon_t \equiv \Theta(L)z_t$, where z_t is I(d), an ARFIMA process y_t is an I(d) process with ARMA(p,q) error and it is also an ARMA process with I(d) error. Therefore the characteristics of an ARFIMA process are similar to those of an I(d) process.

The ARFIMA process is stationary and invertible for $d \in (-1/2, 1/2)$. The autocovariances γ_i of the ARFIMA process are expressed in terms of the autocorrelations of the ARMA process u_t and the autocovariances of the I(d) process z_t as followings:

(15) $$\gamma_i = \sum_k \rho'_k\, \gamma'_{i-k}, \quad i = 0, 1, 2, \cdots,$$

where ρ'_i are the autocorrelations of the ARMA process u_t, and γ'_i are the autocovariances of the I(d) process z_t. The autocovariances given in Equation (15) involve an infinite sum; however, if all the roots in $\Phi(L)$ are distinct, Sowell (1992a) provided a simpler form.

Similarly to the autocorrelations of the I(d) process, the autocorrelations of the ARFIMA process decrease very slowly. In fact $\rho_i \sim C_3\, i^{2d-1}$ as $i \to \infty$, just as for the I(d) process. This occurs because in Equation (15) the autocorrelations of the ARMA process, ρ_i' decrease quickly, while the autocovariances of the I(d), γ_i' decrease slowly as i increases. Thus the asymptotic behavior of the autocorrelations of the ARFIMA process is dominated by the γ_i'. For a formal proof, see Brockwell and Davis (1991), for example.

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

(16) $$f(\lambda) = |\Theta(z)|^2\, |\Phi(z)|^{-2}\, |1 - z|^{-2d}\, \sigma^2/(2\pi), \quad z = e^{-i\lambda} \quad \text{for } -\pi < \lambda < \pi.$$

Similarly to the case for the I(d) process, in Equation (16), $f(0) \sim [\Theta(1)/\Phi(1)]^2\, \sigma^2/(2\pi)\, \lambda^{-2d}$ as $\lambda \to 0$ for $d > 0$, and $f(0) = 0$ for $d < 0$.

This dissertation investigates two basic concerns about the stationary I(d) process. First, if we apply a unit root test or a stationarity test, as is common practice in time series applications, to a stationary I(d) process, what will be the results? This is not a trivial question because in both tests the usual alternative hypothesis is not a long memory process; the alternative is an I(0) process for the unit root test and an I(1) process for the stationarity test. Second, how can we measure the long-memory characteristics of a given data sets? Because any statistic based on I(d) data depends on the value of d, the differencing parameter, the second question is directly related to the estimation of the differencing parameter d.

The plan of this dissertation is as follows. In Chapter 2 we will prove the consistency of the KPSS test against a stationary I(d) alternative, where the KPSS test, suggested by Kwiatkowski, Phillips, Schmidt and Shin (1992), is a test of stationarity against an I(1) alternative. Simulations are performed to provide evidence on the power of the test in finite samples. Also we will compare the power of the KPSS test against I(d) alternatives to the power of the modified rescaled range test suggested by Lo (1991), which is another type of stationarity test that is designed to have power against stationary long-memory alternatives. Furthermore in Chapter 2 we will compare the power of the KPSS test against a stationary I(d) process to the size of the KPSS test in the presence of stationary AR(1) errors. From these results we can have some idea about the ability of the KPSS test to distinguish a long-memory process, such as I(d), from an autocorrelated but short-memory process, such as AR(1).

In Chapter 3 we will prove the consistency of the Dickey-Fuller test against a stationary I(d) alternative. In a previous article, Sowell (1990) provided the asymptotic distribution of the Dickey-Fuller statistics when the true process is I(d) with $d \in (1/2, 3/2)$. So our asymptotic theory is a natural extension of Sowell's results. The finite sample performance of the Dickey-Fuller tests against an I(d) alternative with some values of $d \in (0, 3/2)$ will be investigated, similarly to Diebold and Rudebusch (1991a), but more extensively. Also in Chapter 3 we will compare the power of the Dickey-Fuller tests against stationary I(d) alternatives to the power of the tests against stationary AR(1) alternatives.

Chapter 4 will consider the estimation of the differencing parameter in the stationary long-memory model. In the recent literature several methods of estimation for the stationary long memory model have been proposed. These include regression based estimation procedures, a conditional sum of squares estimator, exact MLE, several types of approximate MLE, and a minimum distance estimate (MDE). We discuss the asymptotic properties of the MDE and MLE, and also we compare the finite sample performances of the estimates using simulations. In addition we will consider the estimates of the mean, autocorrelations and autocovariances of the I(d) process, because they are the basis for the minimum distance estimates, and the estimates of these parameters are not $\sqrt{T}$-consistent for values of d in some range.

Finally in Chapter 5 we summarize our findings and make some suggestions for further research.

CHAPTER 2

POWER OF THE KPSS TEST OF STATIONARITY AGAINST FRACTIONALLY-INTEGRATED ALTERNATIVES

1. Introduction

Let $\{z_t\}_1^\infty$ be a time series with zero mean, and let $Z_t = \sum_{j=1}^{t} z_j$ be its cumulation (partial sum), for t = 1,2,.... Then we will say that z_t is a short memory process if it satisfies the following two requirements.

(A1) $\sigma^2 = \lim_{T\to\infty} T^{-1}E(Z_T^2)$ exists and is non-zero.

(A2) $\forall\, r \in [0,1]$, $T^{-1/2} Z_{[rT]} \Rightarrow \sigma W(r)$.

In assumption (A2) and throughout this chapter, [rT] denotes the integer part of rT, $\Rightarrow$ denotes weak convergence, and W(r) is the standard Wiener process (Brownian motion).

According to this definition, a short memory process need not be covariance stationary; some heterogeneity in the z_t process is allowed. If z_t is stationary, the "long run variance" σ^2 is proportional to the spectral density at zero frequency, which is required to be neither zero nor infinite. Assumption (A2) is just the usual "invariance principle" for convergence of partial sums to a Wiener process. Several sets of sufficient conditions for such an invariance principle to hold can be found in the literature. Many authors have used Assumption 2.1 of Phillips (1987, p. 280), which requires the existence of absolute moments of order β, for some $\beta > 2$, and strong mixing with mixing coefficients α_m such that $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$. For example, Lo (1991) defines a short memory process as one that satisfies these assumptions. Our definition above is slightly more general.

At a semantic level, one might object to our definition of short memory, because it implicitly involves conditions on existence of moments as well as restrictions on the persistence of dependence. (For example, an iid Cauchy series is not short memory by our definition.) However, no matter what name they are given, conditions (A1) and (A2) are important, because the enormous recent literature on the problem of distinguishing integrated and stationary series has relied heavily on asymptotics involving Wiener processes, established using the invariance principle (A2). For example, the asymptotic properties of the usual Dickey-Fuller tests and of their various autocorrelation-corrected versions are routinely established in terms of Wiener processes. This asymptotic analysis establishes that the common unit root tests are consistent against short-memory alternatives. Conversely, Kwiatkowski, Phillips, Schmidt and Shin (1992) -- hereafter KPSS -- consider a test of the null hypothesis of stationarity, and show its consistency against unit root alternatives. They also assume the conditions of Phillips (1987) to establish asymptotics in terms of Wiener processes, so their null hypothesis is implicitly that the series is short memory, and they prove consistency against alternatives that are integrated in the sense of being short-memory in first differences.

Some recent papers have considered the properties of tests when neither the data nor its first difference are short memory. These papers have typically assumed that the data are fractionally integrated, or I(d), in the sense of Granger (1980), Granger and Joyeux (1980) and Hosking (1981), and have involved asymptotics in terms of fractional Brownian motion. For example, Sowell (1990) derived the asymptotic distribution of the Dickey-Fuller unit root tests when the first difference of the variable is I(d), and Diebold and Rudebusch (1991a) demonstrated by simulations the low power of the Dickey-Fuller tests against I(d) alternatives. Lo (1991) showed that a modified version of the rescaled range test of the null hypothesis of short memory is consistent against I(d) alternatives, and provided simulation evidence of its power in finite samples.

Our objective in this chapter is similar to that of Lo. We consider the KPSS test as a test of the null hypothesis of short memory, and we prove that it is consistent against I(d) alternatives. We provide simulation evidence of its power in finite samples, and show that its power compares favorably to the power of Lo's test. We also compare its power against I(d) alternatives to its size distortion in the presence of short memory autocorrelation. Unsurprisingly, a rather large sample size is required to distinguish reliably between a long memory process and a highly autocorrelated short memory process.

2. Preliminaries

KPSS describe their test as a test of the trend stationarity hypothesis. More precisely, we wish to test the hypothesis that deviations of a series from deterministic trend are short memory. We therefore consider the data generating process (DGP):

(1) $$y_t = \psi + \xi t + z_t\ , \ t = 1, 2, \ldots, T\ ,$$

where $\{y_t\}$ is the observed series and $\{z_t\}$ represents its deviations from deterministic (linear) trend. KPSS assume the components representation $z_t = r_t + \varepsilon_t$, where r_t is a random walk ($r_t = r_{t-1} + v_t$, with $r_0 = 0$, and where the v_t are iid with zero mean and finite variance), and ε_t is a short memory process that satisfies Assumption 2.1 of Phillips (1987, p. 280), and therefore satisfies assumptions (A1) and (A2) above. They test the "stationarity" hypothesis H_0: $\sigma_v^2 = 0$, which implies that $z_t = \varepsilon_t$ is short memory.

Let e_t be the residuals from a regression of y_t on intercept and time (t), and let S_t be the partial sum process of the e_t: $S_t = \sum_{j=1}^{t} e_j$, t= 1,...,T. Let σ^2 be the long run variance of the errors ε_t, and consider the Newey-West (1987) estimator of σ^2:

(2) $$s^2(\ell) = T^{-1}\sum_{t=1}^{t} e_t^2 + 2T^{-1}\sum_{s=1}^{\ell} W(s,\ell)\sum_{t=s+1}^{T} e_t e_{t-s}$$

Here $w(s,\ell) = 1 - s/(\ell+1)$, which guarantees the non-negativity of $s^2(\ell)$. For consistency of $s^2(\ell)$ under the null hypothesis it is necessary that the lag truncation parameter $\ell \rightarrow \infty$ as $T \rightarrow \infty$. The rate $\ell = o(T^{1/2})$ will usually be satisfactory [see, e.g., Andrews (1991)]. The KPSS statistic for testing the null of stationarity can then be expressed as follows:

(3) $$\hat{\eta}_\tau = T^{-2}\sum_{t=1}^{T} S_t^2 / s^2(\ell).$$

The KPSS statistic $\hat{\eta}_\mu$ is defined in exactly the same way, except that it is based on the residuals $e_t = y_t - \bar{y}$. This corresponds to a regression of y_t on intercept only, and is appropriate if we set $\xi = 0$ in (1), so that deterministic trend is assumed to be absent. That is, the $\hat{\eta}_\mu$ test allows for non-zero level of y_t but not for trend. In that respect it is similar to Lo's modified rescaled range statistic Q_n, which also allows for level but not trend. (Of course, Lo's statistic could easily be modified to allow for linear trend.)

Under the null hypothesis that $z_t = \varepsilon_t$ is a short memory process, $T^{-2}\sum_{t=1}^{T} S_t^2 \Rightarrow \sigma^2 \int_0^1 V_2(r)^2 dr$, where $V_2(r)$ is a so-called second level Brownian bridge, as defined by KPSS, equation (16). Also $s^2(\ell)$ is a consistent estimator of σ^2. Therefore $\hat{\eta}_\tau \Rightarrow$

$\int_0^1 V_2(r)^2 dr$, which KPSS tabulate. Similar statements hold for the $\hat{\eta}_\mu$ test, with $V_2(r)$ replaced by the standard Brownian bridge $V_1(r) = W(r) - rW(1)$.

Under the alternative that Δz_t is a short memory process, KPSS show that $(\ell/T)\ \hat{\eta}_\tau \Rightarrow \int_0^1 [\int_0^a W^*(s)ds]^2 da / \int_0^1 W^*(s)^2 ds$, where $W^*(s)$ is a demeaned and detrended Wiener process, as defined in Park and Phillips (1988, p. 474). Thus the statistic $\hat{\eta}_\tau$ is $O_p(1)$ under the null hypothesis and is $O_p(T/\ell)$ under the unit root alternative. Since $T/\ell \rightarrow \infty$ as $T \rightarrow \infty$, the test is consistent. A very similar asymptotic distribution result and the same conclusion hold for the $\hat{\eta}_\mu$ test.

In this chapter we are concerned not with unit root alternatives, but rather with the alternative that the z_t are fractionally integrated, or I(d), in the sense of Granger (1980), Granger and Joyeux (1980) and Hosking (1981). As a matter of definition, z_t is I(d) if it has the representation

(4) $$(1 - L)^d z_t = u_t$$

where the series $\{u_t\}$ is short memory. Equivalently, $z_t = (1 - L)^{-d} u_t$. The usual binomial expansion of $(1 - L)^{-d}$ yields the infinite moving average expression $z_t = \sum_{j=1}^{\infty} b_j u_{t-j}$ where $b_j = \Gamma(j+d)/[\Gamma(d)\Gamma(j+1)]$. Some well known properties of I(d) processes include the following. An I(d) process is stationary and invertible for d in the range (-1/2, 1/2). Its autocorrelations decline slowly, at a hyperbolic rate rather than the usual exponential rate, and so an I(d) process is natural to consider when a series appears to exhibit persistent autocorrelation ("long memory"). For $d > 0$ the series is so strongly positively autocorrelated that the sum of the autocorrelations diverges and the spectral density of the series at frequency zero is infinite. However, the spectral density at zero of the first differenced series equals zero, so that the first differenced series will appear to be overdifferenced. Thus an analysis of either z_t or Δz_t using standard ARMA models is unlikely to be satisfactory. For $d < 0$ the converse statements are true: the spectral density at zero of the series equals zero, and yet the spectral density at zero of its partial sum is infinite.

We will proceed under the following Assumption.

ASSUMPTION 1: (i) z_t is I(d) with $d \in (-1/2, 1/2)$. (ii) The u_t are iid $N(0, \sigma_u^2)$.

This assumption is slightly stronger than is needed, and slightly stronger than others have made. For example, Sowell (1990, p. 498) does not assume normality, but does assume that the u_t are iid with zero mean and a finite r^{th} absolute moment for some $r \geq \max [4, -8d/(1+2d)]$. Lo (1991, p. 1294) follows Taqqu (1975) in assuming normality and stationarity of u_t, but he does not assume that the u_t are iid. Hosking (1984) assumes that the u_t are iid, and he makes a variety of other assumptions ranging from finite second moment to normality; a consistency result that we will quote below relies on u_t having a finite fourth moment. We have deliberately made Assumption 1 strong enough that we can take useful intermediate results from a variety of sources.

The basic tools that we need follow directly from Sowell. Define the partial sum process corresponding to z_t as $Z_t = \sum_{j=1}^{t} z_j$. Define $\sigma_T^2 = \text{var}(Z_T)$. Then Sowell shows that

$$(5) \qquad \sigma_T^2 = \sigma_u^2 \{\Gamma(1-2d)/[(1+2d)\Gamma(1+d)\Gamma(1-d)]\} \bullet [\Gamma(1+d+T)/\Gamma(T-d) - \Gamma(1+d)/\Gamma(-d)]$$

and that, as $T \to \infty$,

(6) $\sigma_T^2 / T^{1+2d} \to \sigma_u^2\, \Gamma(1\text{-}2d)/[(1+2d)\Gamma(1+d)\Gamma(1\text{-}d)] \equiv \omega_d^2$.

(Thus, for $d \neq 0$, requirement (A1) above fails, and the series is not short memory.) Furthermore, given Assumption 1, Sowell (Theorem 2) shows that, for $r \in [0, 1]$,

(7) $\sigma_T^{-1} Z_{[rT]} \Rightarrow W_d(r)$

where the "fractional Brownian motion" $W_d(r)$ of Mandelbrot and Van Ness (1968) is defined by the stochastic integral

(8) $$W_d(r) = \int_0^1 (r - s)^d\, dW(s) / \Gamma(d+1).$$

(Thus, for $d \neq 0$, requirement (A2) above for the series to be short memory also fails.) Using equation (6), we will rewrite the weak convergence result (7) in a slightly more convenient form:

(9) $$T^{-(d+1/2)} Z_{[rT]} \Rightarrow \omega_d\, W_d(r)\ .$$

Note that if z_t is I(d), its partial sum Z_t is $O_p(T^{d+1/2})$; in contrast, if z_t is short memory, its partial sum is $O_p(T^{1/2})$. This difference in orders in probability drives the consistency of tests based on partial sums against I(d) alternatives.

3. Consistency Against I(d) Alternatives

In this section we prove that the KPSS $\hat{\eta}_\tau$ and $\hat{\eta}_\mu$ tests are consistent against I(d) alternatives with $d \in (-1/2, 1/2)$ and $d \neq 0$. To do so, we show that the statistics are $O_p(T/\ell)^{2d}$, and so as $T \to \infty$ they $\xrightarrow{p} \infty$ for $d > 0$ and they $\xrightarrow{p} 0$ for $d < 0$. Thus an upper tail test (which is standard when unit root alternatives are considered) is consistent against $d \in (0, 1/2)$, while a two-tailed test is consistent against $d \in (-1/2, 0)$ and against $d \in (0, 1/2)$.

For simplicity, we will first consider the the $\hat{\eta}_\mu$ test, based on the residuals $e_t = y_t - \bar{y}$. Thus we assume that the DGP is of the form of equation (1) with $\xi = 0$, so that $e_t = z_t - \bar{z}$. Assumption 1 is maintained throughout this section, so that the invariance principle (9) is assumed to hold.

LEMMA 1: (i) $T^{-(d+1/2)} S_{[rT]} \Rightarrow \omega_d B_d(r)$, where $B_d(r) = W_d(r) - rW_d(1)$.

(ii) $T^{-2(d+1)} \sum_{t=1}^{T} S_t^2 \Rightarrow \omega_d^2 \int_0^1 B_d(r)^2 dr$.

Proof: $T^{-(d+1/2)} S_{[rT]} = T^{-(d+1/2)} \sum_{j=1}^{[rT]} (z_j - \bar{z})$

$$= T^{-(d+1/2)} \sum_{j=1}^{[rT]} z_j - \{[rT]/T\}\, T^{-(d+1/2)} \sum_{j=1}^{T} z_j$$

$$\Rightarrow \omega_d W_d(r) - \omega_d r W_d(1) = \omega_d B_d(r),$$

which proves part (i). For part (ii),

$$T^{-2(d+1)} \sum_{t=1}^{T} S_t^2 = T^{-1} \sum_t \{T^{-(d+1/2)} S_t\}^2 \Rightarrow \omega_d^2 \int_0^1 B_d(r)^2 dr$$

by the continuous mapping theorem. ■

THEOREM 1: Suppose that $\ell = 0$. Then $T^{-2d}\, \hat{\eta}_\mu \Rightarrow (\omega_d^2/\sigma_z^2) \int_0^1 B_d(r)^2 dr$, where $\sigma_z^2 \equiv \mathrm{var}(z_t) = \sigma_u^2 \Gamma(1-2d)/\Gamma^2(1-d)$.

Proof: $T^{-2d}\,\hat{\eta}_\mu = T^{-2(d+1)} \sum_{t=1}^{T} S_t^2 \,/\, s^2(0)$. The asymptotic distribution of the numerator is given by part (ii) of Lemma 1. The denominator, $s^2(0) =, T^{-1} \sum_{t=1}^{T} e_t^2$ converges in probability to σ_z^2; for example, see Hosking (1984, Theorem 2, p. 6). The result then follows by the joint convergence of the numerator and denominator. ■

The case just treated, $\ell = 0$, is appropriate if one is interested in testing the null of white noise against I(d) alternatives, but not if one is interested in testing the null of short memory against I(d) alternatives. For the asymptotic distribution of the statistic under the null of short memory to be free of nuisance parameters, we must pick ℓ such that $\ell \to \infty$ but $\ell / T \to 0$ as $T \to \infty$. We now proceed to consider this case.

THEOREM 2: Suppose that, as $T \to \infty$, $\ell \to \infty$ but $\ell / T \to 0$. Then, for $d \in (0, 1/2)$, $\hat{\eta}_\mu \xrightarrow{p} \infty$; for $d \in (-1/2, 0)$, $\hat{\eta}_\mu \xrightarrow{p} 0$.

Proof: $\hat{\eta}_\mu = T^{-2(d+1)} \sum_{t=1}^{T} S_t^2 \,/\, T^{-2d} s^2(\ell)$. The asymptotic distribution of the numerator is given by part (ii) of Lemma 1. For $d \in (0, 1/2)$, $T^{-2d} s^2(\ell) \xrightarrow{p} 0$ according to Lo, p. 1309, equation (A.5). Similarly, for $d \in (-1/2, 0)$, $T^{-2d} s^2(\ell) \xrightarrow{p} \infty$ according to Lo, p. 1310. The result follows immediately. ■

Theorem 2 implies that the two-tailed $\hat{\eta}_\mu$ test is consistent against I(d) alternatives for $d \in (-1/2, 1/2)$, $d \neq 0$. Obviously the upper tail test is consistent against $d \in (0, 1/2)$, while the lower tail test is consistent against $d \in (-1/2, 0)$.

In fact, we can say more about $s^2(\ell)$ than the limiting results used to prove Theorem 2. By doing so, we can establish the following theorem giving the asymptotic distribution of the $\hat{\eta}_\mu$ statistic under the alternative, from which Theorem 2 would follow immediately as a corollary.

THEOREM 3: Suppose that, as $T \to \infty$, $\ell \to \infty$ but $\ell/T \to 0$. Then, for $d \in (-1/2, 1/2)$, $(\ell/T)^{2d}\ \hat{\eta}_\mu \Rightarrow \int_0^1 B_d(r)^2\,dr$.

Proof: $(\ell/T)^{2d}\ \hat{\eta}_\mu = T^{-2(d+1)} \sum_t S_t^2 \ /\ \ell^{-2d}\ s^2(\ell)$.

The numerator converges to $\omega_d{}^2 \int_0^1 B_d(r)^2\,dr$ according to Lemma 1. To prove the theorem, we therefore show that the denominator converges in probability to $\omega_d{}^2$.

To do so, we first note that, as $T \to \infty$ with ℓ fixed, $\ell^{-2d}\ s^2(\ell) \xrightarrow{p} \ell^{-2d}\ \sigma^2(\ell)$, where as a matter of definition

$$\sigma^2(\ell) = \gamma_0 + 2 \sum_{s=1}^{\ell} w_{s,\ell}\gamma_s \text{ with } \gamma_j = j^{\text{th}} \text{ autocovariance of } z_t \text{ and}$$

$w_{s,\ell} = 1 - s/(\ell+1)$. This is an implication of the fact that the sample autocovariances are consistent estimates of the population autocovariances [see, for example, Hosking (1984)]. We next note that

$$(\ell+1)\,\sigma^2(\ell) = (\ell+1)\,\gamma_0 + 2 \sum_{s=1}^{\ell} (\ell+1-s)\,\gamma_s = \text{var}(Z_{\ell+1}).$$

Taking the limit as $\ell \to \infty$, $(\ell+1)^{-2d}\,\sigma^2(\ell) = (\ell+1)^{-(2d+1)}\,\text{var}(Z_{\ell+1}) \to \omega_d{}^2$, using equation (6) above. Since $(\ell+1)^{-2d}\,\sigma^2(\ell)$ and $\ell^{-2d}\,\sigma^2(\ell)$ have the same limit, this proves the result. ∎

The analysis for the $\hat{\eta}_\tau$ test is very similar. It rests on the following generalization of Lemma 1.

LEMMA 2: Let e_t be the residuals from a regression of y_t on (1,t), t = 1, 2, ..., T, and let $S_t = \sum_{j=1}^{T} e_j$. Then $T^{-(d+1/2)} S_{[rT]} \Rightarrow \omega_d V_d(r)$, where

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

Proof: Let $\hat{\psi}$ and $\hat{\xi}$ be the coefficents of intercept and trend in the regression of y_t on (1,t). Then

$$(10) \qquad T^{-(d+1/2)} S_{[rT]} = T^{-(d+1/2)} \sum_{j=1}^{[rT]} z_j - \{[rT]/T\} T^{1/2-d}(\hat{\psi} - \psi)$$

$$- 1/2 \{[rT]/T\}\{([rT]+1)/T\} T^{1.5-d}(\hat{\xi} - \xi).$$

Furthermore, by the same algebra as in Schmidt and Phillips (1992, pp. 285-286), specialized to their case p = 2, we have

$$(11) \qquad T^{1/2-d}(\hat{\psi} - \psi) = 4 T^{-(d+1/2)} - 6 T^{-(1.5+d)} \sum_t t z_y + o_p(1)$$

$$\Rightarrow 4\omega_d W_d(1) - 6\omega_d \int_0^1 r dW_d(r)$$

$$= \omega_d\{-2W_d(1) + 6 \int_0^1 W_d(r)dr\}.$$

Here we have made use of $\int_0^1 r dW_d(r) = W_d(1) - \int_0^1 W_d(r)dr$, which follows from Jonas (1983, p. 29). Also

$$(12) \qquad T^{1.5-d}(\hat{\xi} - \xi) = -6 T^{-(d+1/2)} \sum_t z_t + 12 T^{-(1.5+d)} \sum_t t z_t + o_p(1)$$

$$\Rightarrow -6\omega_d W_d(1) + 12\omega_d \int_0^1 r dW_d(r)$$

$$= \omega_d \{6 W_d(1) - 12 \int_0^1 W_d(r)dr\}.$$

Combining (9), (10), (11) and (12),

$$T^{-(d+1/2)} S_{[rT]} \Rightarrow \omega_d W_d(r) - \omega_d r \{-2 W_d(1) + 6 \int_0^1 W_d(r)dr\}$$

$$- 1/2\, \omega_d r^2 \{6 W_d(1) - 12 \int_0^1 W_d(r)dr\}$$

$$= \omega_d V_d(r).$$ ■

We may note that, for d = 0, $V_d(r)$ is the "second-level Brownian bridge" defined by MacNeill (1978) and Schmidt and Phillips (1992).

Given Lemma 2, it is easy to establish the same conclusions for the $\hat{\eta}_\tau$ test as were given for the $\hat{\eta}_\mu$ test in Theorems 1, 2 and 3. All that is necessary is to replace $B_d(r)$ in Theorems 1 and 3 with $V_d(r)$.

4. Power in Finite Samples

In this section we provide some evidence on the power of the $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests against I(d) alternatives. This evidence is based on simulations. The calculations were done in FORTRAN using the normal random number generator GASDEV/RAN3 of Press, Flannery, Teukolsky and Vetterling (1986). Observations on an I(d) process for d $\in$ [-1/2, 1/2) were generated using the Levinson algorithm [Levinson (1947), Durbin (1960), Whittle (1963)]. We also performed some simulations using I(d) observations generated using the Cholesky decomposition of the error covariance matrix, and got essentially the same results as using the Levinson algorithm. For d $\in$ [1/2, 1], observations were generated by cumulating I(d-1) random variates. (Thus, as a matter of definition, z_t is I(.8) if Δz_t is I(-.2), for example.) Given the I(d) series z_t, t = 1, 2, ..., T, data on the observable series y_t were generated according to equation (1), with $\psi = \xi = 0$. The values of ψ and ξ do not matter for any of the tests that we consider, except that the $\hat{\eta}_\mu$ test and Lo's modified R/S test assume $\xi = 0$.

Tables 2-1 and 2-2 give the powers of the 5% upper tail $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests, respectively, against the alternatives d = 0.1, 0.2, ..., 0.9, 1.0, and also d = 0.45 and 0.499. The results are based on 5,000 replications, except that 10,000 replications were used for d = 0.4, 0.45 and 0.499. We have considered only positive values of d because we are primarily interested in testing short memory against long memory, and thus we consider only upper tail tests. Following KPSS, the number of lags used in the denominator of the statistic (ℓ) was chosen as $\ell 0 = 0$, $\ell 4 = \text{integer}[4(T/100)^{1/4}]$, and $\ell 12 = \text{integer}[12(T/100)^{1/4}]$. We consider samples sizes T = 50, 100, 250 and 500.

Some patterns in Tables 2-1 and 2-2 are clear, and in accord with our expectations. With other things held constant: (i) Power increases as T increases. This is a reflection of the consistency of the test. The rate of growth of power as T increases depends strongly on the choice of ℓ; it is higher when ℓ is lower. (ii) Power is lower when ℓ is higher. Note that this is true even for large sample sizes, in accord with the asymptotics of the previous section, which indicate that power depends on (ℓ/T) even asymptotically. (iii) Power is not very different for $\hat{\eta}_\mu$ than for $\hat{\eta}_\tau$. Allowing for deterministic trend does not cost power. (iv) Power is higher when d is larger; that is, as the alternative hypothesis becomes further from the null.

With respect to point (iv), it is interesting that there is no apparent discontinuity in the power function at or near d = 1/2. As d $\uparrow$ 1/2, the series z_t approaches nonstationarity, the one-period autocorrelation approaches unity, and the covariance matrix of $(z_1,\ldots,z_T)$ approaches singularity. The asymptotic results in the previous section do not hold for $d \geq 1/2$, and, if we were to derive the appropriate asymptotic distribution results, they would look rather different for $d \geq 1/2$ than for $d \in (-1/2, 1/2)$. For $d > 1/2$, it would not be difficult to derive the relevant asymptotic results, using our asymptotic results and the fact that an I(d) series is the cumulation of an I(d-1) series. However, we established the asymptotic distribution of the KPSS test statistics only for $d \in (-1/2, 1/2)$, and in particular not for d = -1/2, so our results cannot be extended in any straightforward way to the case of d = 1/2, and it is not clear that some sort of discontinuity at d = 1/2 can be ruled out. Nevertheless, the power function is smooth in d over the whole range that we consider (from zero to one).

This is not a trivial result. For example, in Chapter 3 we finds that the powers of the Dickey-Fuller $\hat{\rho}_\mu$, $\hat{\rho}_\tau$, $\hat{\tau}_\mu$, and $\hat{\tau}_\tau$ tests are continuous at d = 1/2, while the powers of the Dickey-Fuller $\hat{\rho}$ and $\hat{\tau}$ tests have a discontinuity at d = 1/2. Thus a discontinuity arises only when the series has zero mean and correspondingly level and trend are not extracted. The same appears to be true for the KPSS tests. The KPSS $\hat{\eta}_\mu$ test involves extraction of a mean, and the $\hat{\eta}_\tau$ test involves extraction of level and trend, and in both cases the power function is continuous at d = 1/2. However, suppose we define a KPSS-type test $\hat{\eta}$ in the same way as the $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests, except that level and trend are not extracted; that is, the statistic is based on the raw series rather than the demeaned or detrended series. Interestingly, this test's power function is discontinuous at d = 1/2. For example, for T = 50 and ℓ = 0, power is .753 for d = .4; .837 for d = .45; .977 for d = .499; .800 for d = .5; and .824 for d = .6. Similar results occur for other values of T and ℓ; at d = 1/2, the power function is continuous from the right but not from the left. The reason why this discontinuity should occur, for both Dickey-Fuller and KPSS tests but not when the data are demeaned or detrended, is an interesting puzzle that remains to be solved.

How optimistic the results in Tables 2-1 and 2-2 are depends largely on the choice of ℓ. With ℓ = 0, both tests show reasonable power against d ≥ 0.3 for T ≥ 100; for example, the power of $\hat{\eta}_\mu$ against d = 0.3 is 0.54 for T = 100 and 0.73 for T = 250. However, with ℓ = 0 the tests are susceptible to considerable size distortions in the presence of short-memory autocorrelation. Choosing ℓ large enough to more or less remove these possible size distortions will reduce power very substantially. KPSS provide some evidence on size distortions in the presence of short-memory errors. Specifically, they consider the size of the $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests in the presence of AR(1) errors, with autoregressive parameter ρ = 0, ±0.5 and ±0.8. For T ≥ 100 and ρ = 0.5, the choice ℓ = $\ell 4$ is sufficient to keep the size of the 5% test below 0.10, but ℓ = $\ell 12$ is required if ρ = 0.8. In Tables 2-1 and 2-2, we see that, with ℓ = $\ell 4$, a fairly large sample size is necessary to attain reasonable power against d ≥ 0.3. For example, the power against d=0.3 of the $\hat{\eta}_\mu(\ell 4)$ test is 0.41 for T = 250 and 0.55 for T = 500; these are about the same as the power of the $\hat{\eta}_\mu(\ell 0)$ test for T = 50 and T = 100, respectively. With ℓ =

$\ell 12$, even larger sample sizes are necessary for reasonable power. For example, for T = 500 the power of the $\hat{\eta}_\mu(\ell 12)$ test is only 0.35 for d = 0.3 and 0.46 for d = 0.4.

The good power properties of the tests with $\ell = 0$ basically reflect the fact that it is not hard to distinguish an I(d) series with d > 0 from white noise, while the poorer power properties with larger values of ℓ reflect the fact that it is harder to distinguish an I(d) series from a substantially autocorrelated short memory series. To elaborate on this last point, Table 2-3 compares the power of the $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests against I(d) alternatives to their size in the presence of AR(1) errors. Specifically, we compare power against an I(d) alternative with d = 1/3 to size in the presence of AR(1) errors with $\rho = 0.5$. Both series have a one-period autocorrelation equal to 0.5, but the autocorrelations of the I(1/3) series are much more persistent than those of the AR(1) series. Power against the I(d) series is calculated by simulation as above, using 20,000 replications, while size in the presence of AR(1) errors is taken from KPSS, Table 3.

In Table 2-3, it is clear that the powers of the $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests against the I(1/3) alternative are larger than the corresponding sizes in the presence of AR(1) errors with $\rho = 0.5$, with a few exceptions for the $\hat{\eta}_\tau$ test when $\ell = 0$ and T is small. The difference is most substantial when T is moderately large. For example, for the $\hat{\eta}_\mu(\ell 4)$ test with T = 500, compare power of 0.612 to size of 0.090; for the $\hat{\eta}_\mu(\ell 12)$ test with T = 500, compare power of 0.383 to size of 0.058. It appears that we can hope to distinguish a long memory process from a short memory process with approximately equivalent short-run autocorrelation, but it will require a rather large sample size to do so reliably.

Finally, we compares the power of the $\hat{\eta}_\mu$ test to the power of Lo's modified rescaled range test. Table 2-4 gives the power of the 5% upper tail test using the Lo's rescaled range statistic, with the critical value given by Table II (p.1288) of Lo (1991). The format of Table 2-4 is the same as those of Table 2-1 and 2-2. We use the same specifications for simulations in terms of d, T and ℓ, and we use the same generated data series for the simulation results as in Table 2-1, 2-2 and 2-4.

As a general statement, the powers of the $\hat{\eta}_\mu$ test and Lo's modified R/S test are fairly similar. However, the $\hat{\eta}_\mu$ test is clearly less powerful than the R/S test when power is high, and more powerful when power is low. Thus the $\hat{\eta}_\mu$ test is more powerful when T is small and d is close to zero, or ℓ is $\ell 4$ or $\ell 12$; and Lo's modified R/S test is more powerful when T is large, d is close to one and ℓ is $\ell 0$. Especially, when we choose ℓ = $\ell 4$ or $\ell 12$, Lo's modified R/S test has little power in small samples. Thus the $\hat{\eta}_\mu$ test seems to enjoy an advantage in power over the R/S test in cases in which ℓ is picked large enough to protect against severe size distortions from short-memory autocorrelation, but this is not necessarily an optimistic conclusion, since these are cases in which neither test has high power.

5. Concluding Remarks

In this chapter we have shown that the KPSS $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ statistics can be used to distinguish short memory processes from long memory processes. Specifically, we showed that tests of the null hypothesis of short memory based on these statistics are consistent against long memory alternatives, and we have provided Monte Carlo evidence on their power in finite samples. Their power compares favorably to the power of Lo's modified rescaled range test, which is also consistent against long memory alternatives.

An important practical conclusion that can be drawn from our simulations is that a rather large sample size, such as T = 500 or 1000, will be required to distinguish a long memory process from a short memory process with any reasonable degree of reliability. It is interesting and important to note that this conclusion does not depend much on the strength of the autocorrelation of the series, since what is important is not the size of the autocorrelations, but their persistence. For example, we noted above that an AR(1) process with $\rho = 0.5$ and an I(d) process with d = 1/3 each imply a one-period autocorrelation of 0.5. With T = 500, choosing $\ell = \ell 4$ for the $\hat{\eta}_\mu$ test yields size of .09 with the AR errors and power of .61 with the I(d) errors. Now consider a more strongly autocorrelated series, with one-period autocorrelation equal to 0.8, which could be generated by an AR(1) process with $\rho = 0.8$ or an I(d) process with d = .444. Again with T = 500, results from KPSS and our Table 2-1 indicate that choosing $\ell = \ell 12$ yields size of .09 with the AR errors and power of approximately .51 with the I(d) errors. Finally, consider a less strongly autocorrelated series, with one-period autocorrelation of 0.2, which could be generated by an AR(1) process with $\rho = 0.2$ or an I(d) process with d = .167. With T = 500, picking $\ell = \ell 0$ implies size of .13 with the AR errors and power of approximately .47 (found by interpolating in Table 2-1) with the I(d) errors. Size and power are approximately the same (perhaps to a surprising degree, in fact) in all three cases. The reason is straightforward: with a less strongly autocorrelated series, a smaller value of ℓ is required to keep the size under the null close to its nominal value, but the relevant value of d under the alterantive is also smaller. Conversely, with a more strongly autocorrelated series, the relevant value of d under the alternative is larger, but a larger value of ℓ is required to control size distortions under the null.

The KPSS tests and Lo's test do not have any known optimality properties in the present context. An important avenue of future research will be to try to find more powerful tests, perhaps through a systematic application of standard principles of testing to the I(d) model. For example, Robinson (1993) considers the LM test of the hypothesis d = 0 in the I(d) model, and his statistic can apparently be made (asymptotically) robust to short memory autocorrelation using parametric or nonparametric corrections. We might anticipate a gain in finite sample power from this or similar tests, but that remains to be seen.

TABLE 2-1

POWER OF THE $\hat{\eta}_\mu$ TEST AGAINST I(d) ALTERNATIVES

	VALUE OF d												
ℓ	0.0	0.1	0.2	0.3	0.4	0.45	0.499	0.5	0.6	0.7	0.8	0.9	1.0
						T = 50							
ℓ 0	.042	.129	.251	.392	.544	.610	.675	.672	.771	.849	.897	.938	.960
ℓ 4	.034	.075	.129	.197	.277	.314	.360	.372	.444	.522	.583	.645	.715
ℓ 12	.012	.024	.034	.054	.070	.087	.099	.105	.131	.180	.229	.275	.343
						T = 100							
ℓ 0	.054	.168	.347	.535	.723	.777	.830	.832	.910	.955	.976	.988	.993
ℓ 4	.048	.099	.185	.272	.386	.429	.481	.474	.566	.645	.708	.767	.826
ℓ 12	.037	.053	.090	.132	.196	.219	.250	.244	.316	.380	.449	.509	.595
						T = 250							
ℓ 0	.048	.212	.472	.728	.882	.934	.958	.959	.987	.995	.999	1.000	1.000
ℓ 4	.045	.129	.258	.408	.555	.621	.676	.677	.772	.833	.892	.930	.948
ℓ 12	.040	.084	.161	.244	.335	.384	.428	.427	.511	.581	.644	.715	.760
						T = 500							
ℓ 0	.049	.267	.598	.836	.960	.982	.990	.991	1.000	1.000	1.000	1.000	1.000
ℓ 4	.051	.174	.357	.552	.724	.773	.833	.833	.903	.946	.969	.986	.994
ℓ 12	.049	.122	.219	.352	.462	.511	.573	.578	.662	.747	.910	.864	.898

TABLE 2-2

POWER OF THE $\hat{\eta}_\tau$ TEST AGAINST I(d) ALTERNATIVES

VALUE OF d

ℓ	0.0	0.1	0.2	0.3	0.4	0.45	0.499	0.5	0.6	0.7	0.8	0.9	1.0
							T = 50						
ℓ0	.053	.138	.262	.417	.581	.640	.705	.700	.801	.880	.923	.952	.976
ℓ4	.039	.076	.116	.179	.242	.268	.306	.306	.374	.441	.510	.577	.621
ℓ12	.040	.051	.050	.065	.078	.076	.085	.092	.098	.109	.128	.157	.178
							T = 100						
ℓ0	.051	.180	.377	.609	.780	.842	.888	.889	.952	.979	.990	.996	.997
ℓ4	.043	.090	.175	.272	.364	.413	.461	.461	.565	.653	.714	.771	.824
ℓ12	.032	.057	.079	.112	.146	.165	.190	.185	.247	.288	.320	.362	.415
							T = 250						
ℓ0	.053	.269	.584	.832	.948	.973	.987	.989	.997	1.000	1.000	1.000	1.000
ℓ4	.050	.149	.286	.448	.598	.665	.722	.733	.810	.878	.921	.955	.969
ℓ12	.044	.094	.160	.230	.304	.353	.390	.383	.471	.552	.624	.710	.742
							T = 500						
ℓ0	.049	.323	.724	.930	.991	.997	.999	.998	1.000	1.000	1.000	1.000	1.000
ℓ4	.049	.189	.411	.623	.798	.846	.892	.885	.948	.975	.990	.994	.999
ℓ12	.044	.115	.219	.339	.476	.531	.595	.590	.681	.782	.835	.879	.914

TABLE 2-3

POWER OF THE $\hat{\eta}_\mu$ AND $\hat{\eta}_\tau$ TESTS AGAINST I(d) ALTERNATIVES VERSUS SIZE IN THE PRESENCE OF AR(1) ERRORS

$\hat{\eta}_\mu$ TEST

	Size with AR(1) Errors, $\rho = 0.5$			Power against I(d), $d = 1/3$		
T	$\ell 0$	$\ell 4$	$\ell 12$	$\ell 0$	$\ell 4$	$\ell 12$
30	.321	.114	.005	.344	.184	.009
50	.331	.098	.021	.451	.227	.058
80	.350	.108	.042	.555	.312	.128
100	.352	.090	.043	.604	.310	.154
120	.359	.092	.047	.645	.344	.189
200	.367	.099	.053	.752	.452	.247
500	.370	.090	.058	.895	.612	.383

$\hat{\eta}_\tau$ TEST

	Size with AR(1) Errors, $\rho = 0.5$			Power against I(d), $d = 1/3$		
T	$\ell 0$	$\ell 4$	$\ell 12$	$\ell 0$	$\ell 4$	$\ell 12$
30	.425	.129	.178	.335	.149	.189
50	.486	.113	.047	.473	.194	.069
80	.521	.124	.046	.606	.290	.101
100	.538	.107	.047	.673	.297	.123
120	.542	.114	.054	.717	.340	.155
200	.559	.121	.054	.838	.485	.223
500	.586	.110	.062	.964	.681	.384

TABLE 2-4

POWER OF THE LO'S MODIFIED R/S TEST AGAINST I(d) ALTERNATIVES

	VALUE OF d												
ℓ	0.0	0.1	0.2	0.3	0.4	0.45	0.499	0.5	0.6	0.7	0.8	0.9	1.0
							T = 50						
ℓ 0	.012	.064	.170	.341	.519	.604	.675	.681	.794	.874	.919	.950	.967
ℓ 4	.000	.001	.002	.001	.002	.002	.002	.002	.003	.001	.003	.002	.002
ℓ 12	.006	.002	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
							T = 100						
ℓ 0	.021	.141	.359	.611	.803	.860	.900	.904	.964	.984	.991	.997	.999
ℓ 4	.007	.017	.043	.090	.155	.199	.244	.231	.341	.435	.516	.600	.676
ℓ 12	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
							T = 250						
ℓ 0	.032	.255	.625	.880	.966	.984	.994	.996	.999	.999	1.000	1.000	1.000
ℓ 4	.019	.090	.220	.408	.558	.641	.701	.701	.803	.864	.910	.950	.963
ℓ 12	.005	.013	.020	.036	.055	.067	.087	.082	.119	.150	.213	.274	.319
							T = 500						
ℓ 0	.034	.367	.796	.960	.997	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ℓ 4	.028	.175	.417	.652	.823	.870	.911	.910	.955	.979	.989	.995	.998
ℓ 12	.016	.065	.131	.250	.378	.424	.496	.502	.598	.705	.772	.837	.876

CHAPTER 3

POWER OF DICKEY-FULLER UNIT ROOT TESTS AGAINST STATIONARY FRACTIONALLY-INTEGRATED ALTERNATIVES

1. Introduction

In recent years the econometric literature has shown a growing concern for the long run properties of time series data. For example, there has been an enormous amount of work on testing for unit roots and on cointegration. Virtually all of this work has assumed that the data series are either I(0) or I(1) processes. However, this framework is too restrictive for some applications. Following Granger (1980), Granger and Joyeux (1980) and Hosking (1981), we can generate a fractionally integrated, or I(d), model by allowing for a fractional value of the differencing parameter. The I(d) model has been successfully applied to a number of "long memory" series that are stationary and yet display very considerable dependence over long time horizons.

One of the standard topics in the unit root literature is the problem of distinguishing I(1) and I(0) processes. The unit root testing literature typically considers tests of the null hypothesis that the series in question is I(1) against the alternative that it is I(0). The most common tests have been the Dickey-Fuller (hereafter DF) tests of Dickey (1976) and Dickey and Fuller (1979), and various elaborations including the augmented DF test of Said and Dickey (1984) and the DF tests with Phillips-Perron corrections proposed by Phillips (1987) and Phillips and Perron (1988).

In this chapter, we consider the power of the DF unit root tests against I(d) alternatives. We will be mostly concerned with the empirically relevant case that the data are stationary but long memory, so we will consider data generated by the I(d) model with $d \in (-0.5, 0.5)$. We will derive the asymptotic distribution of the DF statistics when the data generating process is I(d) with $d \in (-0.5, 0.5)$, and show that the tests are consistent against these I(d) alternatives. Our results involve fractional Brownian motion, and are somewhat similar to results of Sowell (1990). However, Sowell considered I(d) alternatives with d in the range (0.5,1.5); our results are a useful addition to his.

We also provide simulation evidence on the finite sample power of the DF tests against I(d) alternatives. Similar results have been presented by Diebold and Rudebusch (1991a) and Hassler and Wolters (1993). However, our results cover more values of d than theirs, and in doing so we uncover some interesting results that had previously been missed. In particular, we discover a discontinuity in the power functions of the DF $\hat{\rho}$ and $\hat{\tau}$ tests at d = 0.5.

2. Preliminaries

Let $\{z_t\}$ be a time series with zero mean, and let $Z_t=\sum_{j=1}^{t} z_j$ be its cumulation (partial sum), for t=1,2, ... Assume that Z_t satisfies the following two conditions for some $d \in (-0.5,0.5)$:

(A1) $\sigma^2 = \lim_{t\to\infty} T^{-(1+2d)} E(Z_T^2)$ exists and is non-zero,

(A2) $\forall\, r \in [0,1],\ T^{-(1/2+d)} Z_{[rT]} \Rightarrow \sigma W_d(r)$

In assumption (A2) and throughout this chapter, [rT] denotes the integer part of rT, $\Rightarrow$ denotes weak convergence, and $W_d(r)$ is the fractional Brownian motion on [0,1] of Mandelbrot and Van Ness (1968), which is defined by the stochastic integral

(1) $W_d(r) = \int_0^r (r-s)^d \, dW(s) \,/\, \Gamma(d+1),$

where W(s) is the standard Brownian motion. Note that $W_d(r) = W(r)$ for d=0.

Note that for d = 0 assumption (A1) is the definition of "the long run variance". So, if d = 0, the long run variance σ^2 is finite and non-zero, and z_t can be called a "short memory" process. See Chapter 2 for a more detailed discussion. However, a short memory process need not be covariance stationary to satisfy assumption (A2), which is (for d=0) an "invariance principle" for convergence of the partial sum to a standard Brownian motion. Some heterogeneity in the z_t process is allowed. A sufficient set of conditions commonly assumed in the time series literature for such an invariance principle is assumption 2.1 of Phillips (1987, p. 280), which requires the existence of absolute moments of order β, for β > 2, and strong mixing with mixing coefficients α_m such that

$$\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty.$$

When d ≠ 0, a wide range of series that satisfy assumptions (A1) and (A2) may be found. Many recent papers focus on the I(d), or fractionally integrated of order d process. As a matter of definition, z_t is I(d) if it has the representation

(2) $$(1 - L)^d z_t = u_t,$$

where L represents the lag operator, and u_t is iid with zero mean and finite variance. A generalization of the I(d) process is the ARFIMA(p,d,q) model, which is also of the form given in equation (2) but where u_t follows a stationary ARMA(p,q) process.

Several sets of sufficient conditions for the series to satisfy the assumptions (A1) and (A2) for d ≠ 0 can be found in the literature. For example, Sowell (1990, p. 498) assumes that the u_t are iid with zero mean, and a finite r^{th} absolute moment for some $r \geq \max[4,-8d/(1+2d)]$. Following Taqqu (1975), Lo (1991, p.1294) assumes normality and stationarity of u_t, but does not assume that u_t are iid. This is actually slightly stronger than Taqqu (1975), who assumes that z_t is strictly stationary and that the absolute $2a^{th}$ moment of the partial sum Z_t is $O_p[a(1+2d)]$ for some $a > 1/(1+2d)$ for $d \leq 0$, and with $a = 1$ for $d > 0$. We will follow the way in Chapter 2 by assuming that the u_t are iid $N(0,\sigma_u^2)$, which is somewhat stronger than the other sets of assumptions, and sufficient for (A1) and (A2).

For $d \in (-0.5,0.5)$, the I(d) process is stationary and invertible, but if $d \neq 0$ it differs from the usual short memory stationary process. The autocorrelations of an I(0) stationary process decrease exponentially after some lags, so that the sum of the autocovariances is finite, and is proportional to the spectral density at zero frequency. The stationary I(d) process with $d > 0$, however, is so strongly positively autocorrelated that the sum of the autocovariances diverges, and the spectral density at zero frequency is infinite, which explains why it is often called a "long memory" process in the literature. The I(d) process with $d < 0$ is negatively autocorrelated and the spectral density at zero frequency is zero.

Furthermore, for $d > 0$, the spectral density at zero frequency of the differenced series is zero; and for $d < 0$, the spectral density at zero frequency of the partial sum process is infinite. Therefore if d is in the range of (-0.5,0.5), neither first differencing nor cumulation is a relevant transformation, since the central limit theorem does not hold for the transformed observations or for the original data.

3. Consistency of DF Tests against I(d) Alternatives

The DF unit root tests are based on the following regression equation:

(3) $$y_t = \mu + \beta[t-(T+1)/2] + \rho y_{t-1} + e_t,\ t = 1,2,\ldots,T$$

In equation (3), y_0 can be any random variable with an arbitrary distribution including fixed constant, but must be independent of the sample size T. The error process $\{e_t\}$ can be iid, or stationary ARMA, or any short memory process which satisfies the conditions (A1) and (A2) with d = 0.

The DF test statistics are formulated using the OLS estimate of ρ (coefficient-type test) and its usual t-statistic (t-statistic-type test). The null hypothesis of a unit root is $\rho = 1$. There are three kinds of tests based on different assumptions about level and trend in the stationary alternative. If we restrict $\mu = 0$ and $\beta = 0$ in equation (3), which presumes that the alternative hypothesis is that y_t is a zero mean short memory process, $T(\hat{\rho}-1)$ and $\hat{\tau}$ are the statistics for the test. If we restrict $\beta = 0$ only, so the alternative is that y_t is a short memory process with constant but possibly non-zero mean, $T(\hat{\rho}_\mu-1)$ and $\hat{\tau}_\mu$ are used. When we do not restrict the parameter values for μ and β, so that in the alternative we allow a non-zero level and a deterministic linear trend, the test statistics are $T(\hat{\rho}_\tau-1)$ and $\hat{\tau}_\tau$. Here $\hat{\rho}$, $\hat{\rho}_\mu$, $\hat{\rho}_\tau$ are the OLS estimates of ρ, and $\hat{\tau}$, $\hat{\tau}_\mu$, $\hat{\tau}_\tau$ are the usual t-statistics for the hypothesis $\rho = 1$, in the respective regression equations.

Under the null hypothesis that $\rho = 1$, the OLS estimates of ρ are consistent and of order $O_p(T^{-1})$. Thus to obtain an asymptotic distribution we normalize them by T, and consider $T(\hat{\rho}-1)$. The limiting distributions are not normal but rather are functions of

Brownian motion. The t-statistics are $O_p(1)$ but do not follow the t-distribution; again the limiting distributions are functions of Brownian motion.

Sowell (1990) considered the asymptotic distribution of the DF statistics under the assumption that the data are generated by equation (3) with $\rho = 1$ and the errors e_t follow an $I(d^*)$ process with $d^* \in (-0.5,0.5)$. (For $\hat{\rho}$ and $\hat{\tau}$, it is also assumed that $\mu = \beta = 0$, while for $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ it is assumed that $\beta = 0$.) Thus the observed data y_t are $I(d)$ with $d \in (0.5,1.5)$. Note that, to avoid confusion, we let d^* represent the value of the fractional differencing parameter of the e_t process, and $d = 1+d^*$ represent the value of the fractional differencing parameter of the y_t process. Sowell's proofs only apply to the $\hat{\rho}$ and $\hat{\tau}$ tests, but the same results should hold for the tests based on $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ or $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$.

Consider first the case that e_t is $I(d^*)$ with $d^* \in (-0.5,0)$, so that y_t is $I(d)$ with $d \in (0.5,1)$. Then $\hat{\rho}$ is a consistent estimate of $\rho = 1$, but $\hat{\rho}-1$ is $O_p[T^{-(1+2d^*)}]$, so that $T(\hat{\rho}-1)$ diverges. The asymptotic distribution of $T^{(1+2d^*)}(\hat{\rho}-1)$ has non-positive support, so $T(\hat{\rho}-1)$ diverges to $-\infty$. Furthermore $\hat{\tau} \to -\infty$. Thus the DF tests are consistent against $d \in (0.5,1)$. Next consider the case that e_t is $I(d^*)$ with $d^* \in (0,0.5)$, so that y_t is $I(d)$ with $d \in (1,1.5)$. Then $\hat{\tau} \to \infty$, so that the DF t-statistic based tests are consistent against d in this range. However, $\hat{\rho}-1$ is $O_p(T^{-1})$ and $T(\hat{\rho}-1)$ converges to a limiting distribution that is a function of fractional Brownian motion. Thus the limiting distribution but not the normalization differs from the case that $d^* = 0$, and the DF coefficient based tests are not consistent against $d \in (1,1.5)$.

In this chapter we consider the Dickey-Fuller tests for the case that $d \in (-0.5,0.5)$. This corresponds to the case that we are testing the null hypothesis of a unit root against the alternative of a stationary long-memory process, and this is an empirically relevant case. We will formally state our assumptions, as follows.

ASSUMPTION 1.

1. The data generating process is of the form:

(4) $$y_t = \mu + \beta[t-(T+1)/2] + e_t, \quad e_t = (1 - L)^d u_t$$

for $d \in (-0.5,0.5)$.

2. The u_t are iid $N(0,\sigma_u^2)$.

3. $\mu = \beta = 0$.

We note the following features of these assumptions. First, in this representation y_t and e_t are fractionally integrated of the same order. Second, the assumption of normality in 2. is stronger than necessary. Third, for tests based on $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ we can allow $\mu \neq 0$, while for tests based on $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ we can allow both μ and $\beta \neq 0$.

Notice that according to Theorem 1 of Sowell (1990), under these assumptions, $\{e_t\}$ satisfies conditions (A1) and (A2). In (A1), $\sigma^2 = \sigma_u^2\,\Gamma(1-2d)/[(1+2d)\Gamma(1+d)\Gamma(1-d)]$, where σ_u^2 is the variance of u_t, and $\Gamma(\cdot)$ is the gamma function.

LEMMA 1: Let $\tilde{t} = t-(T+1)/2$ and $\tilde{y}_t = y_t - \bar{y}$, where $\bar{y} = \sum_{t=1}^{T} y_t/T$. Then under Assumption 1, $\sum_{t=1}^{T} \tilde{t}\,\tilde{y}_t = O_p(T^{d+3/2})$.

Proof: $\sum_{t=1}^{T} \tilde{t}\,\tilde{y}_t = \sum_{t=1}^{T} t y_t - [(T+1)/2]\sum_{t=1}^{T} y_t$. Then

$$\sum_{t=1}^{T} t y_t/T^{d+3/2} = \sum_{t=1}^{T} \frac{[rT]}{T} y_t/T^{d+1/2} \Rightarrow \int_0^1 r\,dW_d(r) = W_d(1) - \int_0^1 W_d(r)\,dr, \text{ where}$$

the last equality follows from Jonas (1983, p. 29). Also $\sum_{t=1}^{T} y_t / T^{d+1/2} \Rightarrow W_d(1)$. Then the result follows immediately. ■

THEOREM 1: Under Assumption 1 $\hat{\rho}$, $\hat{\rho}_\mu$, $\hat{\rho}_\tau \xrightarrow{p} \rho_1$, the first order autocorrelation of $\{y_t\}$, and $\hat{\beta}_\tau \xrightarrow{p} \beta = 0$.

Proof: $\hat{\rho}$ and $\hat{\rho}_\mu$ are the first order sample autocorrelations using the known mean of zero and the sample mean, respectively, and are known to be consistent estimates of the population first order autocorrelation [see, for example, Hosking (1984), and Brockwell and Davis (1991)]. So we just need to prove the consistency of $\hat{\beta}_\tau$ and $\hat{\rho}_\tau$. First consider $\hat{\beta}_\tau$. After some algebra,

$$\hat{\beta}_\tau = \frac{\sum_t \tilde{y}_{t-1}^2 \sum_t \tilde{t}\tilde{y}_t - \sum_t \tilde{t}\tilde{y}_{t-1} \sum_t \tilde{y}_t \tilde{y}_{t-1}}{\sum_t \tilde{t}^2 \sum_t \tilde{y}_{t-1}^2 - (\sum_t \tilde{t}\tilde{y}_{t-1})^2}$$

$$= \frac{O_p(T)\, O_p(T^{d+2/3}) - O_p(T^{d+2/3})\, O_p(T)}{O(T^3)\, O_p(T) - O_p(T^{2d+3})},$$ since the consistency of the sample autocovariances provided by Hosking (1984) imply $\sum_t \tilde{y}_{t-1}^2$ and $\sum_t \tilde{y}_t \tilde{y}_{t-1}$ are $O_p(T)$; by Lemma 1 $\sum_t \tilde{t} y_t$ and $\sum_t \tilde{t}\tilde{y}_{t-1}$ are $O_p(T^{d+3/2})$; and $\sum_t \tilde{t}^2$ is $O(T^3)$. Finally from the facts that $O_p(T^\alpha)O_p(T^\beta) = O_p(T^{\alpha+\beta})$ and $O_p(T^\gamma) + O_p(T^\gamma) = O_p(T^\gamma)$ for any real numbers α, β and γ:

$$\hat{\beta}_\tau = \frac{O_p(T^{d+5/2})}{O_p(T^4)} = O_p(T^{d-3/2}) \xrightarrow{p} 0.$$

Similarly,

$$\hat{\rho}_\tau = \frac{\sum_t \tilde{t}^2 \sum_t \tilde{y}_t \tilde{y}_{t-1} - \sum_t \tilde{t}\tilde{y}_t \sum_t \tilde{t}\tilde{y}_{t-1}}{\sum_t \tilde{t}^2 \sum_t \tilde{y}_{t-1}^2 - (\sum_t \tilde{t}\tilde{y}_{t-1})^2}$$

$$= \frac{(T^{-3}\sum_t \tilde{t}^2)(T^{-1}\sum_t \tilde{y}_t \tilde{y}_{t-1}) - (T^{-2}\sum_t \tilde{t}\tilde{y}_t)(T^{-2}\sum_t \tilde{t}\tilde{y}_{t-1})}{(T^{-3}\sum_t \tilde{t}^2)(T^{-1}\sum_t \tilde{y}_{t-1}^2) - (T^{-2}\sum_t \tilde{t}\tilde{y}_{t-1})^2}.$$

Then $T^{-3}\sum_t \tilde{t}^2 \longrightarrow 1/12$, $T^{-1}\sum_t \tilde{y}_t \tilde{y}_{t-1}$, $T^{-1}\sum_t \tilde{y}_{t-1}^2 \xrightarrow{p} \gamma_1, \gamma_0$ respectively, and $T^{-2}\sum_t \tilde{t}\tilde{y}_t$, $T^{-2}\sum_t \tilde{t}\tilde{y}_{t-1} \xrightarrow{p} 0$ since $\sum_t \tilde{t}y_t$ and $\sum_t \tilde{t}\tilde{y}_{t-1}$ are $O_p(T^{d+3/2})$ by Lemma 1.

Therefore

$$\hat{\rho}_\tau \xrightarrow{p} \frac{(1/12)(\gamma_1)}{(1/12)(\gamma_0)} = \rho_1 \qquad \blacksquare$$

Note that even though Theorem 1 tells us that the OLS estimates $\hat{\rho}$, $\hat{\rho}_\mu$, and $\hat{\rho}_\tau$ are consistent for the one-period population autocorrelation, they are not guaranteed to have asymptotic normal distributions. From Hosking (1984) it is known that $\hat{\rho}$ and $\hat{\rho}_\mu$ are $\sqrt{T}$-consistent and asymptotically normal for $d < 0.25$ but not for $d \geq 0.25$. For $d = 0.25$, the asymptotic distribution is normal, but the asymptotic variance is of order $(\ln T)/T$ instead of $1/T$. For $d > 0.25$, the asymptotic variance is of order $T^{-(1-2d)}$.

LEMMA 2: Let denote $\hat{s}^2$, $\hat{s}_\mu^{\,2}$ and $\hat{s}_\tau^{\,2}$ be the usual error variance estimates from the regressions that yield $\hat{\rho}$, $\hat{\rho}_\mu$ and $\hat{\rho}_\tau$, respectively. Then $\hat{s}^2$, $\hat{s}_\mu^{\,2}$, $\hat{s}_\tau^{\,2} \xrightarrow{p} \gamma_0(1-\rho_1^2)$

Proof: The proof for $\hat{s}^2$ is straightforward, as follows.

$$\hat{s}^2 = \frac{1}{T-1}\sum_t (y_t - \hat{\rho} y_{t-1})^2$$

$$= \frac{1}{T-1}(\sum_t y_t^2 - 2\hat{\rho}\sum_t y_t y_{t-1} + \hat{\rho}^2 \sum_t y_{t-1}^2)$$

$\xrightarrow{p} \gamma_0 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0 = \gamma_0(1 - \rho_1^2)$, by Theorem 1 and the consistency of the sample autocovariances. The proofs for $\hat{s}_\mu^2$ and $\hat{s}_\tau^2$ are essentially the same. ■

THEOREM 2: Under Assumption 1 all of the DF test statistics [$T(\hat{\rho}-1)$, $T(\hat{\rho}_\mu-1)$, $T(\hat{\rho}_\tau-1)$, $\hat{\tau}$, $\hat{\tau}_\mu$, and $\hat{\tau}_\tau$] $\to -\infty$ as $T \to \infty$.

Proof: Consider $T(\hat{\rho} - 1) = T(\hat{\rho} - \rho_1) + T(\rho_1 - 1)$. Clearly $T(\rho_1 - 1)$ is $O(T)$ and $\to -\infty$ as $T \to \infty$. Now we want to claim the first term [$T(\hat{\rho} - \rho_1)$] in the expression is dominated by the second term [$T(\rho_1 - 1)$] as $T \to \infty$ so that the whole expression $T(\hat{\rho} - 1) \to -\infty$ as $T \to \infty$. Consider $T(\hat{\rho} - \rho_1)$. If $-0.5 < d < 0.25$, $(\hat{\rho} - \rho_1)$ is $O_p(T^{-1/2})$ and $T(\hat{\rho} - \rho_1)$ is $O_p(T^{1/2})$. If $d = 0.25$, $\sqrt{T/\ln T}(\hat{\rho} - \rho_1) \xrightarrow{p}$ a normal distribution, thus $(\hat{\rho} - \rho_1)$ is $O_p(\sqrt{(\ln T)/T})$, and $T(\hat{\rho} - \rho_1)$ is $O_p(\sqrt{T\ln T})$. Finally if $0.25 < d < 0.5$, $T^{1-2d}(\hat{\rho} - \rho_1) \xrightarrow{p}$ a non-normal limiting distribution, so $T(\hat{\rho} - \rho_1)$ is $O_p(T^{2d})$. Therefore for $d \in (-0.5,0.5)$, in the expression of $T(\hat{\rho} - 1)$ the second term [$T(\rho_1 - 1)$] always dominates the first term [$T(\hat{\rho} - \rho_1)$] as $T \to \infty$. For the other cases of the coefficient tests, $T(\hat{\rho}_\mu-1)$ and $T(\hat{\rho}_\tau-1)$, the proofs are basically the same.

For the t-statistic-type tests $\hat{\tau}$, $\hat{\tau}_\mu$, and $\hat{\tau}_\tau$, first consider $\frac{1}{\sqrt{T}}\hat{\tau}$.

$$\frac{1}{\sqrt{T}}\hat{\tau} = \frac{\hat{\rho}-1}{\sqrt{[\hat{s}^2/(\frac{1}{T}\sum_t y_{t-1}^2)]}} \xrightarrow{P} \frac{\rho_1-1}{\sqrt{\frac{\gamma_0(1-\rho_1^2)}{\gamma_0}}} = \frac{\rho_1-1}{\sqrt{(1-\rho_1^2)}},$$

since $\hat{s}^2 \xrightarrow{P} \gamma_0(1-\rho_1^2)$ by Lemma 1, $(\hat{\rho} - 1) \xrightarrow{P} (\rho_1 - 1)$ by Theorem 1, and $\frac{1}{T}\sum_t y_{t-1}^2$ $\xrightarrow{P} \gamma_0$ by the consistency of the sample autocovariances given in Hosking (1984). So $\frac{1}{\sqrt{T}}\hat{\tau} \xrightarrow{P} \frac{\rho_1-1}{\sqrt{(1-\rho_1^2)}} < 0$. Thus $\hat{\tau} \to -\infty$, as $T \to \infty$. The proof for $\hat{\tau}_\mu$ is the same as the proof for $\hat{\tau}$, after replacing $\hat{\rho}$, $\hat{s}^2$ and y_{t-1} with $\hat{\rho}_\mu$, $\hat{s}_\mu^{\,2}$ and $\tilde{y}_{t-1}$ respectively.

Considering $\hat{\tau}_\tau$, after some algebra,

$$\frac{1}{\sqrt{T}}\hat{\tau}_\tau = \frac{\hat{\rho}_\tau - 1}{\sqrt{\hat{s}_\tau^{\,2}\frac{\sum_t \tilde{t}^2}{\frac{1}{T}\sum_t \tilde{t}^2\sum_t \tilde{y}_{t-1}^{\,2} - \frac{1}{T}(\sum_t \tilde{t}\tilde{y}_{t-1})^2}}}$$

$$= \frac{\hat{\rho}_\tau - 1}{\sqrt{\hat{s}_\tau^2/[(\frac{1}{T}\sum_t \tilde{y}_{t-1}^{\,2}) - (\sum_t \tilde{t}\tilde{y}_{t-1})^2/(T\sum_t \tilde{t}^2)]}}$$

$$= \frac{\hat{\rho}_\tau - 1}{\sqrt{\hat{s}_\tau^2/[(\frac{1}{T}\sum_t \tilde{y}_{t-1}^{\,2}) - o_p(1)]}},$$ since $(\sum_t \tilde{t}\tilde{y}_{t-1})^2$ is $O_p(T^{2d+3})$ and $\sum_t \tilde{t}^2$ is $O(T^3)$, thus $(\sum_t \tilde{t}\tilde{y}_{t-1})^2/(T\sum_t \tilde{t}^2) = O_p(T^{2d+3})/[O(T)\,O_p(T^3)] =$ $O_p(T^{2d+3})/O_p(T^4) = O_p(T^{2d-1}) = o_p(1)$. Therefore similarly to the proof for the $\hat{\tau}$,

ar

rc

T

4.

D

t,

us

T

є

D

cu

ac

de

sir

co

tha

$$\frac{1}{\sqrt{T}}\hat{\tau}_\tau \xrightarrow{p} \frac{\rho_1 - 1}{\sqrt{\frac{\gamma_0(1-\rho_1^2)}{\gamma_0}}} = \frac{\rho_1 - 1}{\sqrt{(1-\rho_1^2)}} < 0, \text{ again by Theorem 1, Lemma 1}$$

and consistency of the autocovarince. So as $T \rightarrow \infty$, $\hat{\tau}_\tau \rightarrow -\infty$. ■

The Theorem 2 is intuitively natural, because the value of d is one under the unit root hypothesis, and it is less than one under the I(d) alternatives of this chapter. From Theorem 2, both lower tail tests and two tail tests are consistent.

4. Power in Finite Samples

In this section we provide some evidence on the power in finite samples of the the DF coefficient type tests [$T(\hat{\rho}-1)$, $T(\hat{\rho}_\mu-1)$, $T(\hat{\rho}_\tau-1)$] and the t-statistic-type tests ($\hat{\tau}$, $\hat{\tau}_\mu$, $\hat{\tau}_\tau$) against I(d) alternatives with $d \in (-0.5, 1.5)$. This evidence is based on simulations, using the normal random number generator GASDEV/RAN3 of Press, Flannery, Teukolsky and Vetterling (1989). Observations on the I(d) process $\{e_t\}$, t=1,2,..., for $d \in [-0.5, 0.5)$ were generated using the recursion algorithm given by Levinson (1947), Durbin (1960), and Whittle (1963). For $d \in [0.5, 1.5)$, the observations were generated by cumulating observations from an I(d-1) process. The observed series $\{y_t\}$ was generated according to the DGP (4) with $\mu = 0$ and $\beta = 0$, so that $y_t \equiv e_t$ and the parameter "d" is the degree of fractional integration of the observed series y_t.

Diebold and Rudebusch (1991a) performed a similar though less extensive set of simulations. They generated I(d) series using the Cholesky decomposition of the error covariance matrix. Our results agree quite closely with their results (Table1, p. 158) for those parameter values that are common to both experiments.

d

10

1.

si

th

of

au

th

kn

on

$\hat{\tau}_{\tau}$

so

for

po

tha

see

doe

2 fo

mea

Tables 3-1 and 3-2 give the powers of 5% two tailed tests against alternatives with d = 0.4, ..., 1.499. The critical value of the tests were taken from Fuller (1976) for T=50, 100, 250, 500. The results are based on 5,000 replications except for d = 0.4, 0.45, 0.499, 1.4, 1,45 and 1.499 where the results were based on 10,000 replications. We did simulations for d = 0.0, 0.1, 0.2 and 0.3, in which the power of the tests is so close to one that we did not report these cases in the tables. Note that we consider only positive values of d, including the case where $d \geq 0.5$, since we are primarily interested in positively autocorrelated series.

There are several important results in Tables 3-1 and 3-2. First, with d constant, the power of the tests increases. This is certainly not surprising for those tests that are known to be consistent. (Recall that all of the tests are consistent against d < 1, while only the t-statistic based tests are consistent against $d \in (1.0,1.5)$; furthermore, for $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$ consistency against $d \in (1.0,1.5)$ has been conjectured but not formally proved.) In some cases power grows rather slowly as T increases. See in particular the $\hat{\rho}$ and $\hat{\rho}_\mu$ tests for d > 1.

Second, the power functions of all of the tests are generally monotonic, so that power grows as d diverges from unity. An interesting and possibly important exception is that the power functions of the $\hat{\rho}$ and $\hat{\tau}$ tests are discontinuous from the left at d = 0.5; see the low powers of these tests against d = 0.499 for all sample sizes. This discontinuity does not occur for the $\hat{\rho}_\mu$, $\hat{\tau}_\mu$, $\hat{\rho}_\tau$ or $\hat{\tau}_\tau$ tests. A similar discontinuity was found in Chapter 2 for the tests of the stationarity hypothesis, for statistics not involving correction for mean or trend (and in the absence of mean or trend). The power of the $\hat{\rho}$ and $\hat{\rho}_\mu$ tests also falls as d increases to 1.499, so that it is natural to suspect a discontinuity of the power function from the left at $d = 1.5$. However, we did not consider values of $d \geq 1.5$ so we cannot confirm such a discontinuity. In the case of the discontinuity at $d = 0.5$, we should note that the asymptotic distributions of the statistics for $d < 0.5$, derived in this chapter, are naturally different from the asymptotic distributions for $d > 0.5$, derived by Sowell. Also the asymptotic distributions for $d = 0.5$ are unknown. From this perspective a discontinuity of the power function at $d = 0.5$ is not surprising. What is surprising is that it occurs for some but not all of the tests.

Third, it is worth stressing that the power of unit root tests against stationary long memory processes [$d \in (-0.5,0.5)$] is quite high, except for the $\hat{\rho}$ test with d very close to 0.5. Previous papers, such as Diebold and Rudebusch (1991a), have tended to stress the low power of unit root tests against fractionally integrated alternatives, but this is because they have not focused on d in the stationary range. It is true that power is not high against d in the range (0.5,1.0), especially for d close to unity, and it is even lower against d in the range (1.0,1.5). However, power against stationary long-memory processes is quite good.

Fourth, the power of coefficient-based tests is quite similar to the power of the corresponding t-statistic based tests for $d < 1.0$. However, the t-statistic based tests are generally more powerful for $1.0 < d < 1.5$.

Finally, we can compare the power of the tests that do not make mean or trend corrections ($\hat{\rho}$ and $\hat{\tau}$) to those that make a mean correction ($\hat{\rho}_\mu$ and $\hat{\tau}_\mu$) or to those that make both mean and trend corrections ($\hat{\rho}_\tau$ and $\hat{\tau}_\tau$). The tests that do not make mean or trend corrections are generally less powerful than those that do, for $d < 1.0$, and this is perhaps surprising given that no mean or trend is present. We might suppose that the flexibility to allow for mean or trend would cost power, but it does not. The same pattern is true for the coefficient-based tests for d > 1. However, for the t-statistic based tests for d > 1, the pattern is reversed, and the τ test is more powerful than the $\hat{\tau}_\mu$ or $\hat{\tau}_\tau$ tests. There is no apparent explanation for these interesting results.

We also did some experiments to compare the power of Dickey-Fuller tests against I(d) alternatives to their power against stationary short-memory alternatives. Specifically, we consider power against AR(1) alternatives. In each case, we considered 5% two tailed tests. We consider AR(1) coefficients ρ = 0.8, 0.9, 0.95 and 0.98. We consider I(d) processes with values of d that imply the same one-period correlation as these values of ρ; that is, we choose d = 0.8/1.8 (= 0.444), 0.9/1.9 (= 0.474), 0.95/1.95 (= 0.487), and 0.98/1.98 (= 0.495).

The results are given in Tables 3-3 and 3-4, based on simulations with 10,000 replications. Comparing parameter values that imply equal one-period autocorrelations (e.g., ρ = 0.8 versus d = 0.8/1.8), the power of all tests against the I(d) process is almost always higher than the power of the same test against the corresponding AR(1) process. These differences in power are often substantial. The few exceptions that we find to this general rule are not substantial, and occur when power is high.

The higher power of the tests against long-memory alternatives than against short-memory alternatives is perhaps surprising. Although we have picked values of ρ and d that equate the one-period autocorrelation, the I(d) processes are much more persistent, and their high-order autocorrelations are much larger than the corresponding high-order autocorrelations for the AR(1) processes. In terms of the pattern of autocorrelations exhibited over moderate to long periods, an I(d) process with d = 0.444 is much more similar to a unit root process than is an AR(1) process with $\rho = 0.8$, for example. Why unit root tests should be more powerful against the I(d) process with $d = 0.444$ than against the AR(1) process with $\rho = 0.8$ is certainly not clear. It may simply reflect the fact that unit root tests, at least in the forms we consider them (with no corrections for autocorrelation), basically rely on the one-period autocorrelation. With corrections for autocorrelation, especially with data-driven choices of lag lengths, these results might well reverse. For example, if we considered the augmented Dickey-Fuller test with a data-driven rule for choosing the number of augmentations, the higher persistence of the I(d) process would likely lead to a larger number of augmentations than would occur for the corresponding AR(1) process. Since more augmentations lead to lower power, the power of the augmented test against the I(d) process would quite possibly be lower than against the AR(1) process with equal one-period autocorrelation. This is an interesting topic for further research.

5. Conclusion

In this chapter we show that the DF unit root tests can be used to distinguish an I(1) process from a stationary I(d) process. We prove the consistency of the tests against I(d) alternatives for $d \in (-0.5, 0.5)$, and the finite sample performance of the tests is investigated in a Monte Carlo simulation.

The DF tests are quite powerful against stationary I(d) alternatives, even in moderate sized samples. They are less powerful against I(d) alternatives with $d > 0.5$, as has also been shown by Diebold and Rudebusch (1991a).

We usually found apparent continuity in the power function between the tests against stationary I(d) alternatives and the tests against nonstationary I(d) alternatives. However we also found somewhat strange discountinuities in the power function for some tests when the value of d approaches 0.5 from the left or 1.5 from the left. These discontinuities were related to the treatment of unknown mean and deterministic trend, in ways that are not at present understandable.

We also compared the power of the DF tests against stationary I(d) alternatives to the power against stationary AR(1) processes, picking the values of d and of the autoregressive parameter ρ so as to imply the same one-period autocorrelation. A surprising result is that the DF tests usually had higher power against the I(d) process than against the corresponding AR(1) process. We conjecture that this result might be reversed by considering DF tests with autocorrelation corrections, such as the augmented DF tests or the Phillips-Perron corrected versions of the tests.

In fact, the asymptotic and finite sample properties of augmented and Phillips-Perron corrected DF tests in the presence of I(d) data are a very important and natural topic for further research. In a recent unpublished paper, Hassler and Wolters (1993) argue that the augmented DF test is inconsistent against I(d) alternatives if the number of augmentations grows with sample size, and they support their argument with some limited simulations. These simulations show much higher powers for the Phillips-Perron corrected tests than for the augmented DF tests, and this is true for both stationary and nonstationary I(d) alternatives. However, they do not present rigorous asymptotics for either type of test, nor do they consider data-driven choices of the lag length in either type of correction.

TABLE 3-1

POWER OF COEFFICIENT TYPE DF UNIT ROOT TESTS AGAINST I(d) ALTERNATIVES

VALUE OF d

TESTS	0.4	0.45	0.499	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.45	1.499
								T = 50							
$T(\hat{\rho}-1)$	.87	.63	.09	.65	.49	.31	.16	.08	.05	.07	.13	.18	.24	.25	.08
$T(\hat{\rho}_\mu-1)$	.98	.95	.89	.88	.67	.41	.19	.08	.05	.08	.14	.20	.25	.26	.07
$T(\hat{\rho}_\tau-1)$	.94	.88	.81	.80	.57	.34	.16	.08	.05	.07	.15	.23	.33	.37	.42
								T = 100							
$T(\hat{\rho}-1)$	.99	.86	.13	.85	.71	.47	.25	.10	.05	.08	.15	.22	.28	.28	.10
$T(\hat{\rho}_\mu-1)$	1.0	1.0	1.0	1.0	.92	.66	.32	.11	.05	.09	.17	.25	.30	.29	.08
$T(\hat{\rho}_\tau-1)$	1.0	1.0	.99	1.0	.91	.64	.33	.11	.05	.10	.20	.33	.47	.51	.56
								T = 250							
$T(\hat{\rho}-1)$	1.0	.99	.23	.98	.90	.69	.40	.14	.05	.10	.17	.25	.31	.31	.13
$T(\hat{\rho}_\mu-1)$	1.0	1.0	1.0	1.0	1.0	.92	.55	.17	.05	.11	.21	.30	.33	.32	.09
$T(\hat{\rho}_\tau-1)$	1.0	1.0	1.0	1.0	1.0	.94	.60	.19	.05	.13	.32	.50	.62	.66	.69
								T = 500							
$T(\hat{\rho}-1)$	1.0	1.0	.32	1.0	.98	.84	.51	.18	.05	.10	.21	.27	.31	.32	.13
$T(\hat{\rho}_\mu-1)$	1.0	1.0	1.0	1.0	1.0	.98	.72	.23	.06	.12	.24	.32	.34	.32	.10
$T(\hat{\rho}_\tau-1)$	1.0	1.0	1.0	1.0	1.0	1.0	.79	.26	.05	.17	.39	.60	.71	.74	.75

TABLE 3-2

POWER OF T-STATISTIC TYPE DF UNIT ROOT TESTS AGAINST I(d) ALTERNATIVES

VALUE OF d

TESTS	0.4	0.45	0.499	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.45	1.499
T = 50															
$\hat{\tau}$	.88	.64	.09	.65	.49	.31	.16	.08	.05	.11	.28	.47	.68	.78	.97
$\hat{\tau}_\mu$	.95	.89	.80	.80	.55	.30	.13	.06	.05	.09	.18	.29	.42	.49	.61
$\hat{\tau}_\tau$	.90	.83	.75	.73	.50	.29	.13	.06	.05	.07	.13	.20	.28	.32	.36
T = 100															
$\hat{\tau}$	.99	.88	.14	.85	.71	.47	.25	.10	.05	.14	.33	.54	.74	.83	.98
$\hat{\tau}_\mu$	1.0	1.0	.99	.99	.87	.55	.25	.08	.05	.11	.22	.37	.51	.58	.69
$\hat{\tau}_\tau$	1.0	1.0	.99	.99	.88	.58	.27	.09	.05	.09	.18	.28	.38	.42	.46
T = 250															
$\hat{\tau}$	1.0	.99	.24	.98	.90	.69	.40	.14	.06	.17	.39	.62	.80	.88	.98
$\hat{\tau}_\mu$	1.0	1.0	1.0	1.0	1.0	.86	.46	.13	.05	.13	.29	.47	.62	.68	.78
$\hat{\tau}_\tau$	1.0	1.0	1.0	1.0	1.0	.91	.54	.16	.05	.11	.25	.40	.49	.54	.59
T = 500															
$\hat{\tau}$	1.0	1.0	.33	1.0	.98	.83	.51	.17	.05	.20	.47	.68	.83	.89	.99
$\hat{\tau}_\mu$	1.0	1.0	1.0	1.0	1.0	.97	.64	.18	.05	.15	.34	.53	.68	.75	.83
$\hat{\tau}_\tau$	1.0	1.0	1.0	1.0	1.0	.99	.74	.22	.05	.14	.30	.46	.58	.63	.67

TABLE 3-3

POWER OF COEFFICIENT TYPE DF TESTS AGAINST STATIONARY AR(1) ALTERNATIVES AND AGAINST STATIONARY I(d) ALTERNATIVES

	AR(1)			I(d)		
			TESTS			
T	$T(\hat{\rho}-1)$	$T(\hat{\rho}_\mu-1)$	$T(\hat{\rho}_\tau-1)$	$T(\hat{\rho}-1)$	$T(\hat{\rho}_\mu-1)$	$T(\hat{\rho}_\tau-1)$
		$\rho = 0.8$			$d = 0.8/1.8$	
50	.58	.30	.13	.67	.95	.89
100	.99	.87	.57	.89	1.0	1.0
250	1.0	1.0	1.0	1.0	1.0	1.0
		$\rho = 0.9$			$d = 0.9/1.9$	
50	.17	.10	.05	.45	.92	.84
100	.57	.29	.13	.68	1.0	1.0
250	1.0	.97	.78	.91	1.0	1.0
		$\rho = 0.95$			$d = 0.95/1.95$	
50	.06	.05	.04	.30	.90	.82
100	.16	.09	.05	.49	1.0	1.0
250	.75	.44	.21	.73	1.0	1.0
		$\rho = 0.98$			$d = 0.98/1.98$	
50	.02	.04	.05	.19	.89	.81
100	.04	.05	.04	.30	1.0	.99
250	.16	.09	.06	.50	1.0	1.0

TABLE 3-4

POWER OF T-STATISTIC TYPE DF TESTS AGAINST STATIONARY AR(1) ALTERNATIVES AND AGAINST STATIONARY I(d) ALTERNATIVES

	AR(1)			I(d)		
	TESTS					
T	$\hat{\tau}$	$\hat{\tau}_\mu$	$\hat{\tau}_\tau$	$\hat{\tau}$	$\hat{\tau}_\mu$	$\hat{\tau}_\tau$
		ρ = 0.8			d = 0.8/1.8	
50	.61	.20	.10	.68	.90	.84
100	.99	.74	.48	.90	1.0	1.0
250	1.0	1.0	1.0	1.0	1.0	1.0
		ρ = 0.9			d = 0.9/1.9	
50	.22	.07	.05	.46	.85	.78
100	.61	.20	.10	.69	1.0	.99
250	1.0	.91	.69	.92	1.0	1.0
		ρ = 0.95			d = 0.95/1.95	
50	.09	.04	.04	.31	.83	.76
100	.21	.06	.05	.50	.99	.99
250	.78	.31	.16	.74	1.0	1.0
		ρ = 0.98			d = 0.98/1.98	
50	.05	.04	.05	.20	.81	.75
100	.08	.04	.04	.31	.99	.99
250	.21	.07	.05	.51	1.0	1.0

CHAPTER 4

FINITE SAMPLE PERFORMANCE OF THE MINIMUM DISTANCE ESTIMATOR IN THE FRACTIONALLY-INTEGRATED MODEL

1. Introduction

In this chapter we will consider the finite sample properties of several estimators of the differencing parameter in the autoregressive fractional integrated moving average (ARFIMA) process of Granger (1980), Granger and Joyeux (1980) and Hosking (1981).

A time series $\{y_t\}$ is said to be an autoregressive fractionally integrated moving average process of order p,d,q or ARFIMA(p,d,q) if

(1) $$\Phi(L)\,(1 - L)^d\,(y_t - \mu) = \Theta(L)\,\varepsilon_t,$$

where L is the lag operator, $(1 - L)^d$ is defined by the binomial series;

(2) $$(1 - L)^d = \sum_{k=0}^{\infty}\binom{d}{k}(-L)^k,$$

$\Phi(L)$ is a polynomial in L of order p containing the autoregressive parameters, $\Theta(L)$ is a polynomial in L of order q containing the moving average parameters, d is the differencing parameter, μ is the mean of the process, and ε_t is a white noise process. Furthermore all the roots of $\Phi(L)$ and $\Theta(L)$ lie outside of the unit circle, and $\Phi(L)$ and $\Theta(L)$ contain no common roots. When p=q=0, the ARFIMA(p,d,q) process becomes a fractionally integrated process of order d, or I(d) process.

In this model, the differencing parameter, d, is of special interest, because the long run properties of the process only depend only on the value of d, while the AR and MA parameters capture the short run dynamics. The value of d can be any real number, but most of literature focuses on d in the range between -1/2 and 1/2. The series is stationary for $d < 1/2$ and is invertible for $d > -1/2$, and it is common to assume that the series has been differenced or cumulated sufficiently that d is in this range. For $0 < d < 1/2$, the series is so strongly positively autocorrelated that the sum of the autocorrelations diverges, which is why this kind of process is called a "long-memory process" in the literature. If $-1/2 < d < 0$ the series is so strongly negatively autocorrelated that the sum of autocorrelations goes to zero in the limit. So as long as d is not an integer value, the usual ARIMA models are not suitable for these kinds of series.

In the recent literature, basically two types of estimates are proposed for the fractionally integrated model. The first type is a two step procedure in which the differencing parameter d is estimated consistently in the first step, and the other parameters of the model are estimated in the second step using the consistent estimate of the differencing parameter. The best known example of this kind of estimator is Geweke and Porter-Hudak (1983). They proposed a least squares estimation method for the differencing parameter in the first stage, followed by usual methods for ARIMA models applied to the series filtered by $(1 - L)^d$. These procedures are computationally simple, but they are not efficient asymptotically, and their finite-sample properties are poor in the presence of significant short-run dynamics. In the simulation study of Agiakloglou, Newbold, and Wohar (1992) it is shown that the Geweke and Porter-Hudak estimate of the differencing parameter has a severe bias in finite samples.

The second type of procedures for estimating the parameters in the long memory model are the methods in which all the parameters are estimated simultaneously, except sometimes the mean μ which can be estimated by the sample mean. The typical example for this case is the maximum likelihood estimator (MLE), called the exact MLE. Assuming normality, the log likelihood function is the following:

(3) $\ln L = -T/2 \ln(2\pi) - 1/2 \ln|\Sigma| - 1/2 (Y_T'\Sigma^{-1}Y_T)$,

where Y_T is the $T \times 1$ vector of demeaned data series, so $Y_T = [(y_1-\mu)\ (y_2-\mu)\ \cdots\ (y_T-\mu)]'$, Σ is the covariance matrix of Y_T, and T is the sample size. The covariance matrix Σ depends on d and on the ARMA parameters. Often μ is replaced by the sample mean $\bar{y}$.

The exact MLE is intuitively appealing, but it has some shortcomings. Specifically, the calculation of the MLE is time-consuming and demanding because of the need to calculate and invert the $T \times T$ covariance matrix Σ. So several approximate MLEs have been proposed, which do not require the inversion of Σ.

The first approximate MLE is the conditional sum of squares estimator (CSS) which was proposed by Li and McLeod (1986). It truncates the infinite sum in the definition of $(1 - L)^d$ to a finite sum, and estimate the parameters ignoring the truncated parts which are negligible when the sample sizes is large enough.

The second method avoiding the inversion of the covariance matrix is to use an approximation to the sum of squares in the likelihood function using the formula suggested by Whittle (1951), based on the spectral density. These kinds of MLEs are called approximate MLEs in the literature. Fox and Taqqu (1986) used the Whittle approximation on only the sum of squares $Y_T'\ \Sigma^{-1}Y_T$ in Equation (3). Dahlhaus (1989) and Hauser (1992) used the Whittle approximation for $|\Sigma|$ as well as the sum of squares in Equation (3).

Several other methods have been proposed in the literature, based on different principles than MLE. They estimate all parameters at once except possibly μ. Dueker and Startz (1992) utilized the GMM principle to estimate the parameters in the long memory model using the orthogonality conditions of $E(y_{t-i}\,\varepsilon_t)$ for i = 1, 2, 3,.... Tieslau, Schmidt and Baillie (1994) proposed a minimum distance estimator (MDE) for the I(d) process, minimizing the difference between population and sample autocorrelations. These estimators require relatively weak assumptions compared to MLE, and under some specific conditions they may be asymptotically equivalent to MLE.

In this chapter we will compare the finite sample performance of the MDE and various types of MLE. In our study we focus on the differencing parameter in the I(d) model, because it is natural starting point for comparison and presumably, it gives some general idea about the performance of the different estimators in more general cases.

Also in this chapter we will provide a detailed comparison of several version of MLE. Several authors investigated the finite sample performance of some types of MLE. Chung and Baillie (1994) studied the finite sample properties of the CSS estimator. Sowell (1992a) compared the exact MLE with known mean to the Fox and Taqqu approximate MLE and the Geweke and Porter-Hudak estimate. Cheung and Diebold (1994) showed that the finite sample performance of the Fox and Taqqu approximate MLE compares favorably to that of the exact MLE when the mean of the process is unknown. Hauser (1992) constructed an approximate likelihood which is similar to the Fox and Taqqu approximation, but more accurate in finite samples, and showed that MLE based on his likelihood has smaller bias and similar variance, compared to the other MLE, when the mean is unknown.

The scheme of this Chapter is as follows. In the next section we discuss the MDE, while the following section discusses various MLEs. Then we report the finite sample properties of these estimates of d and finally we add concluding remarks.

2. The MDE and the Asymptotic Properties of the Estimate

The MDE in the ARFIMA model is based on the consistency of the sample autocorrelations under relatively weak assumptions on the process. The idea of the MDE in the ARFIMA(p,d,q) model is to find the value of the parameter which minimizes the distance between the true autocorrelation function and its sample counterpart. The following discussion is a brief summary of Tieslau, Schmidt and Baillie (1994) for the MDE, but in the general ARFIMA model.

Let ρ_i be the i^{th} order autocorrelation of the y_t process of Equation (1), and $\hat{\rho}_i$ be the i^{th} order sample autocorrelation function in the usual way as

(4) $$\hat{\rho}_i = \sum_{t=1}^{T-i}(y_t - \bar{y})(y_{t+i} - \bar{y}) / \sum_{t=1}^{T}(y_t - \bar{y})^2,$$

where $\bar{y}$ is the sample mean.

Obviously ρ_i is a function of the differencing parameter d and the p+q AR and MA parameters. Thus we write ρ_i as $\rho_i(\theta)$, where θ is the p+q+1 vector of parameters to be estimated. Sowell (1992a) derived the closed form of the autocovariance function for the ARFIMA process in terms of the hypergeometric function, so it is not too difficult to calculate $\rho_i(\theta)$. Now define vectors of the first n population and the sample autocorrelations as

(5) $$\rho(\theta) = [\rho_1(\theta) \ \ \rho_2(\theta) \cdots \rho_n(\theta)]', \ \ \hat{\rho} = [\hat{\rho}_1 \ \ \hat{\rho}_2 \cdots \hat{\rho}_n]',$$

where $n \geq p+q+1$. Then the MDE estimator, $\hat{\theta}$, of θ is the value of θ which minimizes the criterion function,

(6) $$S(\theta) = [\hat{\rho} - \rho(\theta)]' W [\hat{\rho} - \rho(\theta)],$$

where W is an $n \times n$ symmetric, positive-definite weighting matrix. The asymptotically optimal choice for W is the inverse of the covariance matrix of $\hat{\rho}$.

The asymptotic properties of MDE depend on the asymptotic properties of the sample autocorrelation function of the ARFIMA process. These were derived by Hosking (1984) and Brockwell and Davis (1991), and we summarize them in the following lemma.

LEMMA 1: Let y_t follow the ARFIMA(p,d,q) process of Equation (1), where the white noise process satisfies either (a) iid(o,σ^2) with finite 4^{th} moment; or (b) iid $N(0,\sigma^2)$. Then

(i) for $d \in (-1/2, 1/4)$, $\sqrt{T}[\hat{\rho} - \rho(\theta)] \xrightarrow{d} N(0, V_1)$ under condition (a);

(ii) for $d = 1/4$, $\sqrt{T/\ln(T)}\,[\hat{\rho} - \rho(\theta)] \xrightarrow{d} N(0, V_2)$ under condition (b);

(iii) for $d \in (1/4,1/2)$, $T^{(1-2d)}[\hat{\rho} - \rho(\theta)] \xrightarrow{d}$ non-normal distribution with zero mean and covariance matrix V_3, under condition (b); where T is the sample size, V_1, V_2, V_3, are the $n \times n$ covariance matrices of the limiting distributions, and $\xrightarrow{d}$ means convergence in distribution. Specifically ij^{th} element of V_1 defined as

(7) $$V_{1,ij} = \sum_{k=1}^{\infty} \{\rho_{k+i}(\theta) + \rho_{k-i}(\theta) - 2\rho_i(\theta)\rho_k(\theta)\}\{\rho_{k+j}(\theta) + \rho_{k-j}(\theta) - 2\rho_j(\theta)\rho_k(\theta)\} \blacksquare$$

Now consider the asymptotic properties of the MDE $\hat{\theta}$. The following theorem summarizes the consistency and asymptotic normality of the estimate.

THEOREM 1: Let y_t follow the ARFIMA(p,d,q) process of Equation (1), and satisfy the conditions in Lemma 1. Let $\hat{\theta}$ be MDE of θ. Then

(i) for $d \in (-1/2, 1/4)$, $\sqrt{T}[\hat{\theta} - \theta] \xrightarrow{d} N(0, C_1)$ under condition (a);

(ii) for $d = 1/4$, $\sqrt{T/\ln(T)}[\hat{\theta} - \theta] \xrightarrow{d} N(0, C_2)$ under condition (b);

(iii) for $d \in (1/4,1/2)$, $T^{(1-2d)}[\hat{\theta} - \theta] \xrightarrow{d}$ non-normal distribution with zero mean and covariance matrix C_3, under condition (b). Here T is the sample size and C_1, C_2, C_3, are the $(p+q+1) \times (p+q+1)$ covariance matrices of the limiting distributions. Specifically they are of following form:

(8) $$C_i = (D' W D)^{-1} D' W V_i W D (D' W D)^{-1}, \quad i = 1,2,3,$$

where W is the weighting matrix in Equation (6), V_i is the covariance matrix of the limiting distribution of $\hat{\rho}$ defined in Lemma 1, and D is $n \times (p+q+1)$ derivative matrix of $\rho(\theta)$ with respect to θ, so $D(\theta) = \partial\rho(\theta)/\partial\theta'$.

Proof: Since $\hat{\theta}$ is the value at which the criterion function $S(\theta)$ is minimized, at $\theta = \hat{\theta}$

(9) $$\partial S(\theta)/\partial\theta = -2D' W [\hat{\rho} - \rho(\theta)] = 0$$

(10) $$\partial^2 S(\theta)/\partial\theta\partial\theta' = 2D' W D - 2(\partial D/\partial\theta)' W [\hat{\rho} - \rho(\theta)] = 2D' W D + o_p(1),$$

so the second derivative matrix is asymptotically positive definite. Taking the Taylor expansion of the first derivative of $S(\theta)$ around the true value of θ, say θ_0,

(11) $$\partial S(\theta)/\partial\theta|_{\hat{\theta}} = \partial S(\theta)/\partial\theta|_{\theta_0} + (\partial^2 S(\theta)/\partial\theta\partial\theta'|_{\theta^*})(\hat{\theta} - \theta_0),$$

where θ^* is between θ_0 and $\hat{\theta}$. So provided $\partial^2 S(\theta)/\partial\theta\partial\theta'|_{\theta^*}$ is nonsingular, after substituting the second derivative in Equation (11) into Equation (10), we get

$$\hat{\theta} - \theta_0 = [D(\theta^*)' W D(\theta^*) + o_p(1)]^{-1} \times [D(\theta_0)' W (\hat{\rho} - \rho(\theta_0))],$$

$$\sqrt{T}(\hat{\theta} - \theta_0) = [D(\theta^*)' W D(\theta^*) + o_p(1)]^{-1} \times [D(\theta_0)' W \sqrt{T}(\hat{\rho} - \rho(\theta_0))]$$

$$= [D(\theta^*)' W D(\theta^*)]^{-1}[D(\theta_0)' W \sqrt{T}(\hat{\rho} - \rho(\theta_0))] + o_p(1).$$

Finally, since $\hat{\rho}$ converges in probability to $\rho(\theta_0)$, $\hat{\theta}$ and θ^* converge in probability to θ_0, and we get the following equation:

(12) $$\sqrt{T}(\hat{\theta} - \theta) = [D(\theta)' W D(\theta)]^{-1}[D(\theta)' W \sqrt{T}(\hat{\rho} - \rho(\theta))] + o_p(1),$$

where we drop the subscript from θ_0 for simplicity. From Equation (12) it is clear that as long as $\hat{\rho}$ is consistent, and has the asymptotic normal distribution for $d \in (-1/2, 1/4)$ given by Lemma 1, $\hat{\theta}$ does also, and the covariance of the limiting distribution for $\hat{\theta}$ is as given in Equation (8). For other ranges for d, after replace the normalizing factor $\sqrt{T}$ with the proper one, we have the same type of result. ■

From Theorem 1, it is clear that the optimal weighting matrix is the inverse of the covariance matrix of the limiting distribution of $\hat{\rho}$, which is either V_1, V_2 or V_3, according to the range of d. If we choose the optimal weighting matrix, the covariance of the limiting distribution for $\hat{\theta}$ is

(13) $$C_i = [D(\theta)' V_i^{-1} D(\theta)]^{-1}, \quad i=1, 2, 3.$$

A few points should be made about the implementation of MDE. First, because the criterion function $S(\theta)$ is nonlinear in θ, it is generally not possible to have a closed form solution for the estimator, and we have to use numerical optimization to get the estimate. For the initial value for θ in the numerical optimization, one possible suggestion is the Geweke and Porter-Hudak (1983) estimate. However, if there are no AR and MA terms in the process, so that p=q=0 and d is the only parameter to be estimated, we can get a simple consistent estimate for d from the one-period autocorrelation. If y_t follows the I(d) process and if we use only the one-period autocorrelation, the MDE of d is given by

$$(14) \qquad \hat{d} = \hat{\rho}_1 / (1 + \hat{\rho}_1).$$

Therefore we can use this estimate as an initial value of d for more general cases which use more than one autocorrelation.

Second, to calculate $S(\theta)$, we need to construct the weighting matrix first. In general it involves an infinite sum or integral. If we consider only the case for $d \in (-1/2, 1/4)$, where the MDE is $\sqrt{T}$-consistent, the optimal weighting matrix is the inverse of the covariance matrix of the sample autocorrelations as given by Equation (7). The expression involves an infinite sum. When d is less than zero, there is little persistence in the autocorrelations and the infinite sum can be approximated with less than 100 terms. But if $d > 0$, especially $d > 0.1$, the infinite sum cannot be approximated very well even though we allow more than 1000 terms in Equation (7).

Third, for the estimate of the asymptotic covariance of the MDE $\hat{\theta}$, if we know the closed form of the covariance matrix, we can evaluate it at $\hat{\theta}$. However even if we do not know the closed form of the covariance matrix, we can estimate it consistently through the numerical second derivatives of the criterion function. Since $[\partial^2 S(\theta)/\partial\theta\partial\theta']/2$ converges in probability to $D(\theta)' V_1^{-1} D(\theta)$ in Equation (10) when we use the optimal weighting matrix, a consistent estimate of the covariance matrix of the limiting distribution of the $\hat{\theta}$ is given by

(15) $$\hat{C}_i = 2[\partial^2 S(\theta)/\partial\theta\partial\theta']^{-1} \text{ evaluated at } \hat{\theta},$$

which can be provided by the numerical optimization procedure in most computer software.

3. The Exact MLE, the Approximate MLE and Their Asymptotic Properties

The exact MLE for the model given by Equation (1) is the value of the parameters at which the likelihood function of Equation (3) is maximized. In calculating the MLE we can substitute the sample mean for the population mean μ if we are only interested in the differencing parameter and p+q ARMA parameters, or we can estimate the mean μ together with the other parameters, in which case the MLE of μ is the GLS estimate with the covariance matrix Σ. Because the MLE of μ is asymptotically independent of the other estimates, the choice of the estimate of μ does not affect the asymptotic properties of the estimates of the other parameters.

An alternative form of the likelihood function for the exact MLE, which is mentioned in Yajima (1985) and formally suggested in Brockwell and Davis (1991), is numerically more convenient, since it reduce the number of calculations. It is numerically equivalent to the exact MLE likelihood function of Equation (3). It is of the form:

(16) $$\ln L = -T/2 \ln(2\pi) - 1/2 \sum_{t=1}^{T} \ln(v_t^2) - 1/2 [\sum_{t=1}^{T} (x_t - \hat{x}_t)^2 / v_t^2],$$

where x_t is the demeaned data series (so if we know the mean it is $y_t - \mu$, and if we do not know the mean it is $y_t - \bar{y}$), $\hat{x}_t$ is the one step predictor $\hat{x}_t = E[x_t \mid x_{t-1}, x_{t-2}, \cdots x_1]$, $t = 1, 2, 3, \ldots T$, and v_t^2 is the variance of the $\hat{x}_t$. The formula for $\hat{x}_t$ and v_t^2 are provided in Brockwell and Davis (1991, Proposition 5.2.2 in p.172).

An approximate MLE based on the frequency domain can be defined as the value of the parameters at which the following function is minimized.

(17) $$L_1 = \sum_{j=1}^{m} \ln[f(\lambda_j)] + \sum_{j=1}^{m} I(\lambda_j)/f(\lambda_j),$$

where $\lambda_j = 2\pi j/m$, is the j^{th} Fourier frequency for $j = 1, 2, \ldots, m$; m is the largest integer in $(T - 1)/2$; $f(\lambda_j)$ is the spectral density at λ_j; and $I(\lambda_j)$ is the periodogram at λ_j,. An asymptotically equivalent form of the Fox and Taqqu approximate MLE is the value of the parameters which minimizes following:

(18) $$L_2 = \sum_{j=1}^{m} I(\lambda_j)/f(\lambda_j)$$

Several authors have provided the asymptotic theory for the MLE in the long memory model, using the likelihood functions or objective functions for minimization based on (3), (16), (17) and (18) or equivalent ones. Yajima (1985) considered the exact MLE and the approximate MLE of Fox and Taqqu form based on the I(d) model. He called the second estimator a "least squares estimator" but the objective function of the minimization is the same as the likelihood function used by Fox and Taqqu (1986). For the exact MLE, he showed the $\sqrt{T}$-consistency and asymptotic normality of the MLE $\hat{d}$ for $d \in (0, 1/2)$. For the Fox and Taqqu approximate MLE, he proved $\sqrt{T}$-consistency holds only for $d \in (0, 1/4)$; for $d = 1/4$, $(\hat{d} - d_0) \sim O_p[(1/T \ln T)^{1/2}]$, and for $d \in (1/4, 1/2)$, $(\hat{d} - d_0) \sim O_p(T^{2d-1})$, where d_0 is the true value of d. These results were extended by Yajima (1988) to a regression setup in which μ is replaced by a regression function $z_t'\beta$, where z_t are non stochastic regressors and β is the vector of coefficients. Fox and Taqqu (1986) studied the approximate MLE based on two type of long memory processes, one of which is the ARFIMA process. They proved $\sqrt{T}$-consistency and asymptotic normality of the approximate MLE for $d \in (0, 1/2)$, which appears to contradict Yajima's result. Dahlhaus (1989) improved the Fox and Taqqu (1986) results for the exact MLE and the approximate MLE based on the self-similar process, which is a generalization of the long memory process. He confirmed the $\sqrt{T}$-consistency and asymptotic normality of the two estimates for $d \in (0, 1/2)$ and proved the efficiency of the MLE. Möhring (1990) extended these results to the case that $d < 0$. He proved the $\sqrt{T}$-consistency and asymptotic normality of the exact MLE and the approximate MLE for $d \in (-1/2,0)$.

4. The Sample Mean, Sample Autocovariances and Sample Autocorrelations

In this section, we consider the finite sample properties of the sample mean, sample autocovariances and sample autocorrelations for the I(d) process. For the autocovariances and autocorrelations, we consider both the case in which the mean is known and the case in which it is unknown. This is of interest because the sampling properties of the MDE largely depend on those of the sample autocorrelations, and in general if the unknown mean is replaced with the sample mean the properties of the sample autocorrelations are quite different.

All the results in this section are based on Monte Carlo simulations with 10,000 replications. The I(d) data series are generated by the Durbin-Levinson algorithm with p=q=0, μ=0 in Equation (1), using the normal random number generator GASDEV/RAN3 of Press, Flannery, Teukolsky and Vetterlimg (1986) in FORTRAN. See Chapter 2 for details. We considered d = -.49, -.4, -.3, -.2, -.1, 0, .1, .2, .24, .25, .3, .4, .45, .49, and sample size T = 50, 100, 250. We considered the sample autocovariances and sample autocorrelations up to 5th order. Tables 4-1, 4-2 and 4-3 show the simulation results for the sample mean, sample autocovariances and sample autocorrelations.

First consider the results for the sample mean $\bar{y}$, as given in Table 4-1. The Table gives the mean of the sample mean, and its variance multiplied by $T^{(1-2d)}$. From Hosking (1984) it is known that $T^{(1/2-d)}(\bar{y} - \mu)$ has an asymptotic distribution, so that var($\bar{y}$) is asymptotically of order $T^{(2d-1)}$. Therefore $T^{(1-2d)}$ times the variance of $\bar{y}$ should approach a limiting value as $T \to \infty$. This limiting value, given by Hosking (1984), is presented in Table 4-1 where it is called the "asymptotic variance".

The asymptotic theory for the sample mean seems to be a good approximation for sample sizes 50, 100 and 250, except for the cases where d is close to -.5 or .5. Except the case for d = .49, the mean (of the sample mean) is close to the true value of zero. The variances (of the sample mean), normalized by $T^{(1-2d)}$, are also close to the asymptotic variances except the case for d = -.4 and d = -.49 where the normalized variances of the sample mean are smaller than the asymptotic variances. For example, for d = -.49, the theoretical variance in the limiting distribution is 32.195 but when T = 50, 100 and 250, the normalized finite sample variances are 3.600, 3.942 and 4.483 respectively, as given in Table 4-1.

We next consider the results for the sample autocovariances, which are given in Table 4-2. We begin with the zero-period autocovariance γ_0, the population variance, for which the estimate $\hat{\gamma}_0$ is just the sample variance. These results are given in Table 4-2-0, made up of two pages. Table 4-2-0(a) gives the mean, the normalized variance, the finite sample and the theoretical asymptotic bias, and the mean squared error (MSE) of $\hat{\gamma}_0$. The normalized variance is the finite sample variance, multiplied by T. For $d < .25$, the variance of $\hat{\gamma}_0$ is asymptotically of order T^{-1}, so T times the variance of $\hat{\gamma}_0$ should approach a limiting value. For $d \geq .25$ this is not the appropriate normalization to approach an asymptotic limit, but it is used to avoid the confusion that could result from two different normalizations in the same table. In Table 4-2-0(a) it is assumed that $\hat{\gamma}_0$ is calculated using the sample mean, as it would be when the mean is unknown. Table 4-2-0(b) gives the same information as Table 4-2-0(a) (mean, normalized variances, bias and MSE) for $\hat{\gamma}_0$ calculated using the true mean of the series. Finally, Table 4-2-1 through 4-2-5 give the same information as Table 4-2-0, but for the autocovariances of order one through five.

The results for the autocovariances at lags one through five are quite similar to those for the variance, so we will discuss only the results for $\hat{\gamma}_0$, as given in Table 4-2-0. The sample autocovariances using the sample mean are downward biased in general, except the zero-period autocovariance which is upward biased for $d < 0$, and downward biased for $d > 0$, while the sample autocovariances using the true mean are not biased systematically. When $d < 0$, the variances of the sample autocovariances are also quite similar whether the sample mean or the true mean is used, and the normalized variances do not change much with T. Thus it appears that, for $d < 0$, the finite sample behavior of the sample autocovariances is similar to what would be expected from the asymptotics. However, for d > 0, and especially for d close to .5, things are rather different. When the mean is unknown and the sample mean is used, the sample autocovariances have a severe bias. This bias grows quickly as d approaches .5. For example, for d = .45, $\gamma_0 = 3.642$ but the mean of $\hat{\gamma}_0$ is only 1.308 for T = 50, 1.456 for T = 100 and 1.635 for T = 250. The bias disappears very slowly as T grows; however, this is as predicted by the asymptotic theory in Hosking (1984) as reported in Table 4-2-0(a). When the mean of the process is known and the sample autocovariances are calculated using the true mean, however, this situation exactly reverses: for d close to .5, the bias is much smaller than when the mean is unknown, but the variance is much larger. For example, with d = .49 and T = 50 (and $\gamma_0 = 16.36$), $\hat{\gamma}_0$ has a mean of 1.401 and variance of 11.533/50 = .231 when the mean of the process is unknown, and a mean of 16.415 and variance of 23,154.981/50 = 463.1 when the true mean is known. Mean square error is of comparable magnitude in the two cases (224 versus 463) but the division of mean square error into squared bias and variance is strikingly different.

The results for sample autocorrelations are given Table 4-3, which is similar in format to Table 4-2. We will discuss the results for the one-period autocorrelation ρ_1, as given in Table 4-3-1, but the results for higher-order autocorrelations, given in Table 4-3-2 through 4-3-5, are very similar.

For d < 0 the properties of the sample autocorrelations are more or less the same whether the mean is known or unknown. The sample autocorrelations are only very slightly biased, and their variance is essentially the same whether the true mean or the sample mean is used. When d > 0, the bias is larger, especially when the sample mean is used. When the mean is known, the bias of the sample autocorrelations is not very large, and it tends to disappear fairly quickly as T grows. Furthermore, the variances of the sample autocorrelations with known mean are of reasonable magnitude. This is strikingly different than the situation for the sample autocovariances with known mean, whose variances became very large as d approached .5. The sample autocorrelations based on the sample mean are downward biased. Especially when d is close to .5, the bias is large, and this bias largely persists as T increases from 50 to 250. For $d > 0$, the sample autocorrelations based on the sample mean usually have smaller variances than the sample autocorrelations based on the true mean.

5. The Finite Sample Properties of the MDE and MLE in the I(d) Model

The main purpose of this section is to investigate the adequacy of the asymptotic theory provided in the previous sections for the I(d) process. First, we want to know how reliable the asymptotic theory for the MDE is in finite samples. Second, we want to compare the MDE to the MLE, which is asymptotically efficient. Note that a comparison of the asymptotic efficiency of the MDE and the MLE is presented in Tieslau, Schmidt and Baillie (1994). They showed numerically that for values of $d \in (-1/2, 1/4)$, the variance of the estimate approaches the variance of the MLE, $6/\pi^2$, as the number of autocorrelations in the criterion function increases. So we believe that for $d \in (-1/2, 1/4)$, so that the MDE and MLE are $\sqrt{T}$-consistent and have asymptotic normal distributions, those two estimates are asymptotically equivalent when we use enough autocorrelations in computing the MDE. However, we now ask whether this is approximately true in finite samples.

We begin our simulations of the MDE in the I(d) model with the simplest case in which we use only the first-order autocorrelation. Then $\hat{d}_1 = \hat{\rho}_1/(1+\hat{\rho}_1)$ is the MDE, as given above in Equation (14). Tables 4-4-1 and Table 4-4-2 show the results for these simulations based on 10,000 replications using the same data series as for Table 4-1, 4-2 and 4-3, for the cases that $\hat{\rho}_1$ is based on the sample mean and on the true mean respectively.

For $d \leq 0$, the bias of $\hat{d}_1$ is small, and it goes to zero quickly as T increases. The asymptotic theory is reliable in the sense that the finite sample variance of $\hat{d}_1$ is close to the asymptotic variance, especially for $T \geq 100$. For positive values of d, the bias of $\hat{d}_1$ is small when the mean is known, but when the mean is unknown, the bias of $\hat{d}_1$ is larger and goes to zero more slowly than it did when $d < 0$. Unsurprisingly, the bias becomes worse as d approaches .5. The asymptotic variance becomes a less and less accurate guide to the finite sample variances of $\hat{d}_1$, especially for $d \geq .2$, whether the mean is known or not. For $d \geq .25$, $\hat{d}$ is not $\sqrt{T}$-consistent and the asymptotic distribution is of a different form, and Tieslau, Schmidt and Baillie do not provide asymptotic variances. However, because $\hat{d}_1$ converges to d more slowly for $d \geq .25$, normalized variance defined as T times finite sample variance should increase with T. This does not appear to happen in Table 4-4-1 (case of mean unknown), though it does in Table 4-4-2 (case of mean known). Thus an overall summary of these results is that the asymptotic theory seems quite reliable in moderate sized samples for $d < .2$ but not for larger values of d.

We next consider the MDE based on a larger number of autocorrelations, and compare the results to those for various forms of the MLE. The results for these simulations are given in Table 4-5 for T = 50, 100 and 250, and for values of d between -.4 and .4. More specifically, in Table 4-5, MLE_μ denotes the time-domain exact MLE when the population mean is known; $MLE\bar{y}$ represents the exact MLE when the data are demeaned using the sample mean; F&T represents the Fox-Taqqu approximate MLE; and WL represents the approximate MLE based on the Whittle likelihood. For i = 1, 2, ..., 5, both MDE_i and MDE_i^* represent the MDE based on $(\rho_1, \cdots, \rho_i)$; that is, on the first i autocorrelations. They differ in how the weighting matrix is evaluated. Both MDE_i and MDE_i^* use the optimal weighting matrix as given in Equation (7), but MDE_i evaluates the weighting matrix at $d = \hat{d}_1$ (the consistent estimate base on $\hat{\rho}_1$) whereas MDE_i^* evaluates the weighting matrix at the true value of d. This does not matter for i = 1, since $\hat{d}_1$ is the MDE without specification of the weighting matrix, but it matters for $i \geq 2$. Clearly MDE_i^* is not a feasible estimator in practice, but we include it to understand the extent to which any poor performance of the MDE might be due only to the use of $\hat{d}_1$ in evaluating the weighting matrix. It should be noted that the form of the weighting matrix given in Equation (7) is optimal for d < .25. For d ≥ .25, and in particular for d= .3 and .4 in our simulations, this is not necessarily the optimal weighting matrix, and some other version of the MDE might be better. Also, Equation (7) contains an infinite sum and this sum converges very slowly for d > 0. For d = .2, .3 and .4, our evaluation of Equation (7) was accurate only to about 10^{-2}; however, this did not seem to matter much in the simulations.

The results in Table 4-5 are based on 1,000 replications, using the GAUSS random number generator. For the numerical optimizations, we used the GAUSS maximization procedure with a convergence tolerance of 10^{-5} for the gradient. In most cases we used the Davidon, Fletcher and Powell (DFP) algorithm. In a few cases in which DFP could not find the optimum, we used the Broyden, Fletcher, Goldfard and Shanno (BFGS) algorithm and/or the Newton-Raphson algorithm provided by GAUSS. We used $\hat{d}_1$ as the starting value for all optimizations, except for a few replications of the exact MLE in which we could not find the maximum starting from $\hat{d}_1$, and so we used the true value of d or the true value of d ± .05 as a starting value.

Except for the exact MLE we did not have any particular problems in the numerical optimizations. However, for the exact MLE we faced a problem which is worth noting. As long as d is less than .5, we can evaluate the likelihood function - either the original one given by Equation (3) or the alternative form given by Equation (16). However, if the value of d equals or exceeds .5 during the search for the optimum, we cannot evaluate the likelihood function directly, because the covariance matrix Σ in Equation (3) fails to be positive definite. If the estimate tries to go above the value of .5, we can still evaluate the likelihood by differencing the data and then evaluating the likelihood based on the differenced data and the value (d-1) for the differencing parameter. This is legitimate if we assume the unobservable past observations are fixed at the mean of the series. When we did this, we found a small jump in the likelihood function around d = .5.

Table 4-6 shows the number of irregular replications in the exact MLE - both MLE_μ and $MLE\bar{y}$. In a quite large number of replications, the final estimates are not in the range of (-1/2, 1/2). This happens particularly for d = -.4 or .4 and T = 50. However, comparing our results for the exact MLE with similar simulations given by Sowell (1992a), Cheung and Diebold (1994), Hauser (1992) and Smith, Sowell and Zin (1993), we cannot find any significant differences for the cases where they use the same value of d and sample size.

The MDE using the weighting matrix evaluated at $\hat{d}_1$ (MDE_i, i = 1, 2, ..., 5 in Tables 4-5) is generally biased downward. For each sample size we considered, for d < 0 the bias of the estimate is small and decreases quickly as the number of autocorrelations increases from one to five, while for d > 0 the bias of the estimate is large in general, increases as d approaches to .5, and does not decrease quickly as the number of autocorrelations increases. For a given value of d, as the sample size increases from 50 to 250, the bias of the estimate is reduced substantially. Therefore when T = 250 and the number of autocorrelations is four or five, the bias of the MDE is of reasonable size. For example in Table 4-5-3(a), with T = 250 and five autocorrelations used (MDE_5), the absolute bias of the MDE using the weighting matrix evaluated at $\hat{d}_1$ is less than .01 for d < 0 and less than .025 for d > 0.

The variance and the mean squared error (MSE) of the MDE using the weighting matrix evaluated at $\hat{d}_1$ are given in Tables 4-5-1(b), 4-5-2(b) and 4-5-3(b) for T = 50, 100, 250, respectively. To see how well asymptotic theory works in finite samples we reported T times finite sample variance and T times finite sample MSE as "Normalized Variance" and "Normalized MSE" respectively in these tables. As we mentioned in the previous section, for $d \in (-1/2, 1/4)$ the MDE is $\sqrt{T}$-consistent and has an asymptotic normal distribution, while for $d \in [1/4, 1/2)$ the MDE is consistent but the convergence rate is slower than for $d \in (-1/2, 1/4)$. Thus if the asymptotic distribution theory is applicable to the finite samples sizes we considered, the normalized variance or normalized

MSE should be stable for $d \in (-1/2, 1/4)$ and they should be increasing with T for $d \in [1/4, 1/2)$. Also we note that for $d \in (-1/2, 1/4)$ we use the optimal weighting matrix for the criterion function, but for $d \in [1/4, 1/2)$ the weighting matrix is not optimal. Therefore the variance of the MDE for $d \in (-1/2, 1/4)$ is not compatible to that of the MDE for $d \in [1/4, 1/2)$, not only because of the difference between the two limiting distributions but also because of the difference between the two weighting matrices in the criterion functions.

In Tables 4-5-1(b), 4-5-2(b) and 4-5-3(b), for the variance and MSE of the MDE using the weighting matrix evaluated at $\hat{d}_1$, there seem to be two patterns according to the values of d: one is for $d \in (-1/2, 1/4)$, the other is for $d \in [1/4, 1/2)$, reflecting the different asymptotic distributions, for these two ranges of d.

First, for a given number of autocorrelations and sample size, the variance of the MDE using the weighting matrix evaluated at $\hat{d}_1$ is decreasing as d increases from -.4 to .2. For a given sample size and value of d, as the number of autocorrelations increases, the variance of the estimate decreases rapidly for $d < 0$, and tends to be stabilized or increases a little for $d > 0$. The MSE has a similar pattern to the variance. In the theoretical variance of the limiting distribution for the MDE as given by Tieslau, Schmidt and Baillie (1994) for $d \in (-1/2, 1/4)$ a similar pattern was found. As the value of d increases from -.5 to .25 the asymptotic variance decreases and as the number of autocorrelations increases the asymptotic variance decreases quickly for $d < 0$, but decreases slowly for $d > 0$.

Comparing the magnitude of the normalized finite sample variance or MSE of the estimate to the asymptotic variance, in all the cases the normalized finite sample variance or MSE is slightly bigger than the asymptotic variance. However as the sample size increases the difference between the normalized variance or MSE and the asymptotic variance decreases, as expected. For example, when the number of autocorrelations is five and d = -.4, -.2, 0, and 2, the theoretical variance in the limiting distribution is 1.137, .892, .683, and .676, respectively; the normalized variance of MDE_5 for T = 250 is 1.192, .953, .771, and .733, respectively; and the normalized MSE of MDE_5 for T = 250 is 1.191, .953, .784, and .833, respectively.

Second, for d = .3 or .4, the normalized variance or MSE of the MDE using the weighting matrix evaluated at $\hat{d}_1$ is generally smaller than that of the same estimator of d for $d \leq .2$. For a given sample size, as the number of autocorrelations increases, the variance of the estimate increases for d = .3, and generally decreases for d = .4; the MSE of the estimate generally decreases for both d = .3 and .4 as the number of autocorrelations increases.

For a given number of autocorrelations, as the sample size increases from 50 to 250, the normalized variance (finite sample variance × T) of the estimate increases for both d = .3 and .4, except for MDE_1. But the normalized MSE (finite sample MSE × T) of the estimate does not change much as the sample size increases. If the asymptotics are relevant for these sample sizes, T times the finite sample variance should increase as T increases.

The MDE using the weighting matrix evaluated at the true value of d (MDE_i^*, i = 1, 2, ...,5 in Table 4-5) is also biased downward and the bias is increasing as d increases.

Contrary to the MDE using the weighting matrix evaluated at $\hat{d}_1$, the bias, variance and MSE of the MDE using the weighting matrix evaluated at the true value of d do not generally decrease as the number of autocorrelations increases. Comparing the MDE using the weighting matrix evaluated at $\hat{d}_1$ (MDE_i) to the MDE using the weighting matrix evaluated at the true value of d (MDE_i^*), it is surprising the MDE using the weighting matrix evaluated at $\hat{d}_1$ is better in terms of bias and variance in most the cases we considered. However as the sample size increases form 50 to 250, the differences between the two estimates decreases, as we would expected, and for T = 250 it makes little difference whether the weighting matrix is evaluated at $\hat{d}_1$ or at the true value of d.

We next consider the properties of the various exact and approximate MLEs in our experiments. With the exception of WL (the MLE based on the Whittle likelihood), the MLEs are all biased downward. The absolute bias decreases as T increases, as would be expected. MLE_μ (exact MLE with μ known) has smaller absolute bias than $MLE\bar{y}$ (exact MLE using $\bar{y}$), F&T (fox and Taqqu approximate MLE) or WL (the MLE based on the Whittle likelihood). MLE_μ clearly has the smallest variance, though its variance is not much smaller than that of $MLE\bar{y}$. In terms of MSE, MLE_μ is clearly best, and WL is generally best among the estimators that do not assume knowledge of μ.

It is worth noting that the exact MLEs (MLE_μ and $MLE\bar{y}$) are biased downward as d approaches .5, even though the range of d is not restricted in our numerical maximization. Thus the argument of Smith, Sowell and Zin (1993) for the source of this bias (slow convergence of the sample mean, and restriction of d to the range d < .5 in maximization) are not supported by our results.

Even for T = 250, the normalized variances of the MLEs are not very close to the asymptotic variance of $6/\pi^2$, except the case for d = .4. The convergence to the asymptotic distribution is obviously fairly slow.

We next compare the properties of the MDE to the various MLEs. We will consider only MDE_5, the MDE using five moments and the weighting matrix evaluated at $\hat{d}_1$, which is generally the best of the MDEs in our experiments. In terms of absolute bias, MDE_5 is generally worse than WL, and sometimes better and sometimes worse than MLE_μ, but better than any of the other MLEs. The variance of MDE_5 is larger than the variance of the MLEs for d < 0, but it is generally smaller than the variance of any of the MLEs except MLE_μ for d > 0 and worse than WL for d < 0.

As a general statement, the MDE is dominated by the exact MLE based on the true value of μ (MLE_μ). However, it compares favorably with the exact and approximate MLEs based on $\bar{y}$. Since convergence to the asymptotic distribution is slow for all of these estimators, further simulations with T > 250 are really needed to say more about the comparisons of these estimators.

6. Concluding remarks

In this chapter we discussed the asymptotic theory for the MDE in a general ARFIMA setup, and also we surveyed the asymptotic theory for the exact MLE and approximate MLEs. To see the finite sample behavior of the MDE and MLE, we performed the simulations for the MDE and MLE, as well as for the sample mean, sample autocovariances and sample autocorrelations. All of these simulations were in the context of the simple I(d) model.

In our simulations for the autocovariances and the autocorrelations, we found strange behavior of the sample autocovariances and of the sample autocorrelations using the sample mean when the value of d is close to .5. In our simulations for the MLEs and MDE, we found that the MDE is comparable to the MLEs if we use more than two or three autocorrelations in the criterion function in the MDE.

Our results could profitably be extended in at least two ways. First, the largest sample size that we considered is only T = 250. For series with a strong degree of persistence, convergence to asymptotic results is slow, and larger sample sizes may be relevant. Second, many of the estimates have a substantial and persistent finite sample bias, which makes it hard to compare the finite sample and asymptotic results. It would be desirable to develop higher-order asymptotic approximations which would provide asymptotic expressions for the bias as well as for the variance of the estimates. This is an important topic for further research.

TABLE 4-1

THE SAMPLE MEAN OF THE I(d) PROCESS AND ITS NORMALIZED VARIANCE

	T = 50		T = 100		T = 250		
d	Mean	Normalized Variance	Mean	Normalized Variance	Mean	Normalized Variance	Asymptotic Variance
-.49	-.001	3.600	.000	3.942	.000	4.483	32.195
-.40	.000	2.459	.000	2.568	.000	2.719	3.525
-.30	.000	1.766	.000	1.790	.000	1.806	1.918
-.20	.000	1.357	.001	1.363	.000	1.366	1.383
-.10	-.001	1.135	-.001	1.137	.000	1.140	1.129
.00	-.001	.999	.000	1.012	.002	1.010	1.000
.10	.005	.946	-.002	.961	.000	.938	.954
.20	.004	1.002	.002	1.000	.000	.991	.995
.24	-.006	1.065	-.005	1.024	-.002	1.048	1.047
.25	-.004	1.088	-.001	1.069	.003	1.078	1.064
.30	-.007	1.185	.000	1.212	.002	1.209	1.190
.40	.000	1.966	-.008	1.939	-.015	1.919	1.930
.45	.003	3.509	-.004	3.492	-.014	3.419	3.498
.49	.027	16.267	-.032	15.874	-.042	16.275	16.213

Note : "Normalized variance" denotes (finite sample variance of $\bar{y}$) $\times$ $T^{(1-2d)}$. "Asymptotic variance" denotes the theoretical variance in the limiting distribution of the sample mean, based on Hosking (1984).

TABLE 4-2

THE SAMPLE AUTOCOVARIANCES OF THE I(d) PROCESS

0(a). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_0$ using Sample Mean

		T = 50					T = 100					T = 250				
d	γ_0	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	1.263	1.290	4.212	.026	-.014	.085	1.276	4.087	.012	-.004	.041	1.269	3.914	.005	-.001	.016
-.40	1.183	1.204	3.377	.021	-.003	.068	1.196	3.335	.013	-.001	.034	1.190	3.347	.007	.000	.013
-.30	1.109	1.128	2.858	.019	-.004	.057	1.123	2.827	.013	-.001	.028	1.114	2.776	.005	.000	.011
-.20	1.052	1.066	2.465	.014	-.006	.049	1.060	2.362	.007	-.002	.024	1.056	2.340	.003	-.001	.009
-.10	1.014	1.026	2.232	.012	-.010	.045	1.020	2.105	.006	-.004	.021	1.017	2.132	.003	-.002	.009
.00	1.000	.998	2.066	-.002	-.020	.041	1.002	2.040	.002	-.010	.020	1.001	1.977	.001	-.004	.008
.10	1.019	.993	2.075	-.027	-.042	.042	1.005	2.146	-.015	-.024	.022	1.011	2.129	-.009	-.011	.009
.20	1.099	1.023	2.502	-.076	-.095	.056	1.043	2.645	-.055	-.063	.030	1.067	2.851	-.032	-.036	.012
.24	1.161	1.048	2.798	-.112	-.137	.069	1.078	3.169	-.083	-.095	.039	1.107	3.578	-.053	-.059	.017
.25	1.180	1.052	2.821	-.128	-.150	.073	1.086	3.278	-.094	-.106	.042	1.117	3.744	-.063	-.067	.019
.30	1.316	1.087	3.413	-.229	-.249	.121	1.138	4.262	-.178	-.189	.074	1.191	5.452	-.126	-.131	.038
.40	2.070	1.214	6.072	-.856	-.883	.854	1.310	9.201	-.760	-.768	.670	1.437	16.092	-.633	-.640	.465
.45	3.642	1.308	8.763	-2.334	-2.366	5.623	1.456	14.859	-2.186	-2.207	4.929	1.635	29.430	-2.008	-2.014	4.149
.49	16.360	1.401	11.533	-14.959	-14.993	224.017	1.595	22.336	-14.765	-14.787	218.229	1.853	51.018	-14.507	-14.518	210.665

Note : "γ_0" denotes the true value of the (population) variance. "Mean" denotes the mean of $\hat{\gamma}_0$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_0$) × T. "Sample Bias" denotes the mean of ($\hat{\gamma}_0 - \gamma_0$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4.3 (CONTINU

TABLE 4-2 (CONTINUED)

0(b). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_0$ using True Mean

		T = 50				T = 100				T = 250			
d	γ_0	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	1.263	1.266	4.047	.002	.081	1.263	4.006	.000	.040	1.264	3.883	.000	.016
-.40	1.183	1.182	3.244	-.001	.065	1.184	3.269	.001	.033	1.185	3.320	.002	.013
-.30	1.109	1.109	2.746	-.001	.055	1.113	2.771	.003	.028	1.110	2.753	.001	.011
-.20	1.052	1.051	2.372	-.002	.047	1.051	2.316	-.001	.023	1.052	2.322	.000	.009
-.10	1.014	1.016	2.150	.002	.043	1.015	2.065	.000	.021	1.015	2.116	.000	.008
.00	1.000	.998	2.018	-.002	.040	1.002	2.009	.002	.020	1.001	1.962	.001	.008
.10	1.019	1.015	2.141	-.005	.043	1.019	2.200	-.001	.022	1.018	2.170	-.002	.009
.20	1.099	1.098	3.297	.000	.066	1.096	3.304	-.003	.033	1.099	3.442	.000	.014
.24	1.161	1.167	4.761	.006	.095	1.160	4.779	.000	.048	1.162	5.326	.002	.021
.25	1.180	1.185	5.074	.005	.102	1.182	5.511	.002	.055	1.181	6.087	.000	.024
.30	1.316	1.313	9.385	-.003	.188	1.319	11.096	.002	.111	1.319	14.531	.002	.058
.40	2.070	2.089	85.054	.019	1.701	2.069	123.218	-.001	1.232	2.067	221.133	-.003	.885
.45	3.642	3.655	538.914	.013	10.778	3.645	997.704	.002	9.977	3.597	1967.092	-.046	7.870
.49	16.360	16.415	23154.981	.054	463.103	16.056	42832.779	-.304	428.421	16.420	106275.802	.059	425.107

Note : "γ_0" denotes the true value of the (population) variance. "Mean" denotes the mean of $\hat{\gamma}_0$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_0$) × T. "Sample Bias" denotes (Mean of $\hat{\gamma}_0 - \gamma_0$). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

1(a). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_1$ using Sample Mean

d	γ_1	T = 50 Mean	T = 50 Normalized Variance	T = 50 Sample Bias	T = 50 Asymp. Bias	T = 50 MSE	T = 100 Mean	T = 100 Normalized Variance	T = 100 Sample Bias	T = 100 Asymp. Bias	T = 100 MSE	T = 250 Mean	T = 250 Normalized Variance	T = 250 Sample Bias	T = 250 Asymp. Bias	T = 250 MSE
-.49	-.416	-.426	2.115	-.010	-.014	.042	-.420	2.017	-.005	-.004	.020	-.417	1.972	-.002	-.001	.008
-.40	-.338	-.347	1.679	-.009	-.003	.034	-.342	1.636	-.004	-.001	.016	-.340	1.659	-.002	.000	.007
-.30	-.256	-.264	1.334	-.008	-.004	.027	-.263	1.358	-.007	-.001	.014	-.258	1.289	-.002	.000	.005
-.20	-.175	-.183	1.157	-.008	-.006	.023	-.179	1.148	-.004	-.002	.011	-.177	1.101	-.002	-.001	.004
-.10	-.092	-.108	1.051	-.016	-.010	.021	-.099	1.025	-.006	-.004	.010	-.094	.986	-.001	-.002	.004
.00	.000	-.020	1.040	-.020	-.020	.021	-.011	.997	-.011	-.010	.010	-.004	1.030	-.004	-.004	.004
.10	.113	.072	1.109	-.041	-.042	.024	.090	1.183	-.024	-.024	.012	.102	1.221	-.011	-.011	.005
.20	.275	.182	1.564	-.092	-.095	.040	.212	1.738	-.062	-.063	.021	.239	1.980	-.035	-.036	.009
.24	.367	.233	1.911	-.133	-.137	.056	.272	2.276	-.094	-.095	.032	.308	2.681	-.058	-.059	.014
.25	.393	.249	1.957	-.144	-.150	.060	.290	2.371	-.104	-.106	.035	.329	2.910	-.065	-.067	.016
.30	.564	.319	2.578	-.245	-.249	.112	.377	3.365	-.187	-.189	.069	.435	4.678	-.129	-.131	.035
.40	1.380	.509	5.213	-.871	-.883	.862	.612	8.158	-.768	-.768	.671	.745	15.072	-.635	-.640	.464
.45	2.980	.629	7.811	-2.351	-2.366	5.685	.783	13.947	-2.197	-2.207	4.967	.968	28.681	-2.012	-2.014	4.163
.49	15.719	.739	10.470	-14.979	-14.993	224.594	.944	21.302	-14.775	-14.787	218.506	1.208	50.025	-14.511	-14.518	210.773

Note : "γ_1" denotes the true value of the (population) one-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_1$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_1$) × T. "Sample Bias" denotes the mean of ($\hat{\gamma}_1 - \gamma_1$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

1(b). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_1$ using True Mean

		T = 50				T = 100				T = 250			
d	γ_1	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.416	-.416	2.031	.000	.041	-.416	1.977	.000	.020	-.415	1.957	.000	.008
-.40	-.338	-.338	1.613	.000	.032	-.338	1.604	.000	.016	-.339	1.646	-.001	.007
-.30	-.256	-.256	1.282	.000	.026	-.259	1.330	-.003	.013	-.257	1.279	-.001	.005
-.20	-.175	-.174	1.114	.002	.022	-.175	1.125	.000	.011	-.176	1.092	.000	.004
-.10	-.092	-.096	1.021	-.003	.020	-.093	1.009	-.001	.010	-.092	.979	.000	.004
.00	.000	.001	1.035	.001	.021	-.001	.999	-.001	.010	.000	1.026	.000	.004
.10	.113	.112	1.229	-.001	.025	.113	1.283	.000	.013	.113	1.274	.000	.005
.20	.275	.275	2.392	.001	.048	.274	2.472	-.001	.025	.275	2.620	.000	.010
.24	.367	.369	3.925	.002	.079	.363	3.953	-.003	.040	.366	4.477	.000	.018
.25	.393	.399	4.258	.005	.085	.394	4.715	.000	.047	.396	5.334	.002	.021
.30	.564	.562	8.647	-.002	.173	.566	10.450	.002	.105	.566	13.876	.002	.056
.40	1.380	1.401	85.092	.021	1.702	1.380	122.929	.000	1.229	1.378	220.810	-.002	.883
.45	2.980	2.993	540.053	.013	10.801	2.981	999.176	.001	9.992	2.934	1968.875	-.047	7.878
.49	15.719	15.772	23172.872	.053	463.460	15.416	42869.201	-.303	428.784	15.778	106293.800	.059	425.179

Note : "γ_1" denotes the true value of the (population) one-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_1$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_1$) × T. "Sample Bias" denotes (Mean of $\hat{\gamma}_1$ - γ_1). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

2(a). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_2$ using Sample Mean

		T = 50					T = 100					T = 250				
d	γ_2	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.085	-.088	2.162	-.003	-.014	.043	-.085	2.021	.000	-.004	.020	-.087	2.031	-.002	-.001	.008
-.40	-.085	-.089	1.794	-.004	-.003	.036	-.084	1.713	.001	-.001	.017	-.085	1.677	-.001	.000	.007
-.30	-.078	-.084	1.423	-.006	-.004	.028	-.077	1.435	.001	-.001	.014	-.077	1.364	.001	.000	.005
-.20	-.064	-.069	1.216	-.005	-.006	.024	-.066	1.195	-.002	-.002	.012	-.065	1.157	-.001	-.001	.005
-.10	-.040	-.048	1.105	-.009	-.010	.022	-.044	1.071	-.005	-.004	.011	-.041	1.077	-.001	-.002	.004
.00	.000	-.020	1.042	-.020	-.020	.021	-.011	1.040	-.011	-.010	.011	-.003	1.011	-.003	-.004	.004
.10	.066	.024	1.103	-.041	-.042	.024	.044	1.166	-.022	-.024	.012	.055	1.121	-.011	-.011	.005
.20	.183	.089	1.392	-.094	-.095	.037	.121	1.617	-.063	-.063	.020	.147	1.843	-.036	-.036	.009
.24	.258	.122	1.720	-.136	-.137	.053	.164	2.086	-.094	-.095	.030	.199	2.527	-.059	-.059	.014
.25	.281	.132	1.746	-.149	-.150	.057	.176	2.157	-.105	-.106	.033	.215	2.690	-.066	-.067	.015
.30	.431	.182	2.257	-.250	-.249	.108	.243	3.092	-.188	-.189	.066	.302	4.418	-.130	-.131	.035
.40	1.208	.330	4.692	-.877	-.883	.864	.438	7.555	-.770	-.768	.669	.570	14.578	-.638	-.640	.465
.45	2.788	.427	7.014	-2.361	-2.366	5.713	.588	13.150	-2.200	-2.207	4.972	.774	28.243	-2.013	-2.014	4.167
.49	15.511	.520	9.566	-14.991	-14.993	224.916	.732	20.397	-14.779	-14.787	218.608	.997	49.308	-14.514	-14.518	210.842

Note : "γ_2" denotes the true value of the (population) two-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_2$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_2$) × T. "Sample Bias" denotes the mean of ($\hat{\gamma}_2 - \gamma_2$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

TABLE 4-2 (CONTINUED)

2(b). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_2$ using True Mean

		T = 50				T = 100				T = 250			
d	γ_2	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.085	-.085	2.074	.000	.041	-.084	1.980	.001	.020	-.086	2.014	-.001	.008
-.40	-.085	-.085	1.721	-.001	.034	-.082	1.678	.002	.017	-.085	1.663	.000	.007
-.30	-.078	-.080	1.366	-.002	.027	-.075	1.406	.003	.014	-.077	1.353	.001	.005
-.20	-.064	-.062	1.171	.002	.023	-.063	1.173	.001	.012	-.064	1.148	.000	.005
-.10	-.040	-.037	1.073	.003	.021	-.039	1.051	.000	.011	-.039	1.069	.000	.004
.00	.000	.001	1.033	.001	.021	.000	1.035	.000	.010	.001	1.011	.001	.004
.10	.066	.066	1.219	.000	.024	.068	1.277	.002	.013	.066	1.183	.000	.005
.20	.183	.184	2.261	.001	.045	.183	2.346	.000	.023	.183	2.476	.000	.010
.24	.258	.260	3.746	.002	.075	.256	3.802	-.002	.038	.258	4.336	.000	.017
.25	.281	.285	4.063	.004	.081	.282	4.523	.001	.045	.282	5.126	.001	.021
.30	.431	.428	8.424	-.003	.168	.434	10.275	.002	.103	.434	13.646	.002	.055
.40	1.208	1.228	85.452	.021	1.709	1.208	122.838	.000	1.228	1.205	221.563	-.003	.886
.45	2.788	2.799	540.853	.011	10.817	2.790	1000.523	.002	10.005	2.741	1969.776	-.047	7.881
.49	15.511	15.561	23185.422	.050	463.711	15.209	42897.040	-.301	429.061	15.569	106308.888	.058	425.239

Note : "γ_2" denotes the true value of the (population) two-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_2$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_2$) × T. "Sample Bias" denotes (Mean of $\hat{\gamma}_2 - \gamma_2$). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

3(a). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_3$ using Sample Mean

d	γ_3	T = 50 Mean	T = 50 Normalized Variance	T = 50 Sample Bias	T = 50 Asymp. Bias	T = 50 MSE	T = 100 Mean	T = 100 Normalized Variance	T = 100 Sample Bias	T = 100 Asymp. Bias	T = 100 MSE	T = 250 Mean	T = 250 Normalized Variance	T = 250 Sample Bias	T = 250 Asymp. Bias	T = 250 MSE
-.49	-.037	-.039	2.209	-.002	-.014	.044	-.037	2.030	.000	-.004	.020	-.035	2.071	.001	-.001	.008
-.40	-.040	-.042	1.820	-.002	-.003	.036	-.044	1.757	-.004	-.001	.018	-.040	1.661	.000	.000	.007
-.30	-.040	-.041	1.553	-.001	-.004	.031	-.042	1.460	-.001	-.001	.015	-.042	1.392	-.002	.000	.006
-.20	-.036	-.044	1.272	-.009	-.006	.026	-.039	1.211	-.004	-.002	.012	-.036	1.206	.000	-.001	.005
-.10	-.024	-.036	1.155	-.012	-.010	.023	-.029	1.093	-.005	-.004	.011	-.026	1.068	-.002	-.002	.004
.00	.000	-.021	1.069	-.021	-.020	.022	-.009	1.024	-.009	-.010	.010	-.003	1.006	-.003	-.004	.004
.10	.047	.006	1.107	-.042	-.042	.024	.024	1.115	-.023	-.024	.012	.036	1.127	-.011	-.011	.005
.20	.144	.045	1.372	-.099	-.095	.037	.081	1.511	-.062	-.063	.019	.106	1.768	-.038	-.036	.008
.24	.210	.072	1.568	-.138	-.137	.050	.115	1.936	-.095	-.095	.028	.151	2.420	-.058	-.059	.013
.25	.230	.079	1.673	-.151	-.150	.056	.124	2.055	-.106	-.106	.032	.163	2.622	-.067	-.067	.015
.30	.368	.114	2.070	-.254	-.249	.106	.178	2.958	-.190	-.189	.066	.237	4.209	-.130	-.131	.034
.40	1.115	.231	4.221	-.883	-.883	.865	.343	7.130	-.771	-.768	.666	.476	14.246	-.639	-.640	.465
.45	2.679	.312	6.393	-2.366	-2.366	5.727	.475	12.479	-2.203	-2.207	4.978	.665	27.702	-2.014	-2.014	4.165
.49	15.387	.390	8.652	-14.997	-14.993	225.087	.604	19.541	-14.783	-14.787	218.723	.872	48.366	-14.515	-14.518	210.877

Note : "γ_3" denotes the true value of the (population) three-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_3$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_3$) × T. "Sample Bias" denotes the mean of ($\hat{\gamma}_3 - \gamma_3$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

3(b). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_3$ using True Mean

		T = 50				T = 100				T = 250			
d	γ_3	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.037	-.037	2.116	.000	.042	-.037	1.989	.000	.020	-.035	2.054	.002	.008
-.40	-.040	-.039	1.744	.001	.035	-.043	1.721	-.003	.017	-.040	1.647	.000	.007
-.30	-.040	-.037	1.490	.003	.030	-.040	1.431	.000	.014	-.041	1.380	-.001	.006
-.20	-.036	-.038	1.220	-.002	.024	-.037	1.187	-.001	.012	-.035	1.197	.001	.005
-.10	-.024	-.025	1.119	-.001	.022	-.024	1.076	.000	.011	-.024	1.061	.000	.004
.00	.000	-.001	1.068	-.001	.021	.002	1.028	.002	.010	.001	1.010	.001	.004
.10	.047	.048	1.235	.000	.025	.048	1.221	.001	.012	.048	1.184	.000	.005
.20	.144	.142	2.269	-.002	.045	.144	2.232	.001	.022	.142	2.426	-.002	.010
.24	.210	.212	3.681	.002	.074	.208	3.676	-.002	.037	.210	4.253	.001	.017
.25	.230	.233	3.968	.003	.079	.231	4.468	.001	.045	.231	5.092	.001	.020
.30	.368	.363	8.433	-.005	.169	.369	10.288	.002	.103	.370	13.427	.002	.054
.40	1.115	1.133	85.647	.019	1.713	1.116	122.772	.001	1.228	1.111	221.426	-.003	.886
.45	2.679	2.689	541.641	.010	10.833	2.680	1001.726	.002	10.017	2.633	1972.249	-.046	7.891
.49	15.387	15.436	23183.144	.049	463.665	15.085	42913.354	-.302	429.225	15.445	106335.497	.058	425.345

Note : "γ_3" denotes the true value of the (population) three-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_3$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_3$) × T. "Sample Bias" denotes (Mean of $\hat{\gamma}_3$ - γ_3). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

4(a). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_4$ using Sample Mean

d	γ_4	T = 50 Mean	T = 50 Normalized Variance	T = 50 Sample Bias	T = 50 Asymp. Bias	T = 50 MSE	T = 100 Mean	T = 100 Normalized Variance	T = 100 Sample Bias	T = 100 Asymp. Bias	T = 100 MSE	T = 250 Mean	T = 250 Normalized Variance	T = 250 Sample Bias	T = 250 Asymp. Bias	T = 250 MSE
-.49	-.021	-.023	2.211	-.002	-.014	.044	-.025	2.035	-.004	-.004	.020	-.021	1.996	.000	-.001	.008
-.40	-.023	-.025	1.890	-.001	-.003	.038	-.022	1.758	.001	-.001	.018	-.025	1.705	-.001	.000	.007
-.30	-.025	-.031	1.601	-.005	-.004	.032	-.028	1.517	-.003	-.001	.015	-.026	1.412	-.001	.000	.006
-.20	-.024	-.030	1.316	-.006	-.006	.026	-.027	1.227	-.003	-.002	.012	-.024	1.207	.000	-.001	.005
-.10	-.017	-.030	1.167	-.013	-.010	.023	-.022	1.099	-.005	-.004	.011	-.019	1.068	-.002	-.002	.004
.00	.000	-.020	1.069	-.020	-.020	.022	-.010	1.045	-.010	-.010	.011	-.004	1.023	-.004	-.004	.004
.10	.038	-.003	1.132	-.041	-.042	.024	.014	1.140	-.024	-.024	.012	.026	1.112	-.012	-.011	.005
.20	.121	.023	1.312	-.099	-.095	.036	.060	1.493	-.061	-.063	.019	.084	1.694	-.037	-.036	.008
.24	.181	.041	1.545	-.140	-.137	.050	.086	1.917	-.094	-.095	.028	.122	2.357	-.058	-.059	.013
.25	.199	.047	1.609	-.153	-.150	.055	.093	1.937	-.106	-.106	.031	.131	2.539	-.068	-.067	.015
.30	.328	.071	1.954	-.256	-.249	.105	.137	2.804	-.190	-.189	.064	.197	4.052	-.131	-.131	.033
.40	1.053	.164	3.860	-.888	-.883	.866	.279	6.822	-.773	-.768	.666	.414	13.897	-.639	-.640	.464
.45	2.603	.233	5.856	-2.370	-2.366	5.736	.397	11.831	-2.206	-2.207	4.983	.590	27.085	-2.014	-2.014	4.163
.49	15.299	.296	7.746	-15.003	-14.993	225.250	.514	18.648	-14.785	-14.787	218.787	.783	47.275	-14.517	-14.518	210.921

Note : "γ_4" denotes the true value of the (population) four-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_4$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_4$) × T. "Sample Bias" denotes the mean of ($\hat{\gamma}_4 - \gamma_4$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

4(b). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_4$ using True Mean

		T = 50				T = 100				T = 250			
d	γ_4	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.021	-.021	2.117	-.001	.042	-.024	1.993	-.004	.020	-.021	1.979	.000	.008
-.40	-.023	-.023	1.811	.001	.036	-.021	1.722	.002	.017	-.024	1.691	-.001	.007
-.30	-.025	-.027	1.533	-.002	.031	-.027	1.486	-.002	.015	-.026	1.401	-.001	.006
-.20	-.024	-.024	1.262	.000	.025	-.024	1.203	.000	.012	-.024	1.197	.000	.005
-.10	-.017	-.019	1.127	-.002	.023	-.017	1.079	.000	.011	-.018	1.060	.000	.004
.00	.000	.001	1.061	.001	.021	.000	1.049	.000	.010	.000	1.023	.000	.004
.10	.038	.039	1.253	.001	.025	.038	1.245	.000	.012	.037	1.169	.000	.005
.20	.121	.120	2.278	-.001	.046	.123	2.244	.002	.022	.120	2.348	-.001	.009
.24	.181	.182	3.715	.001	.074	.180	3.681	-.001	.037	.181	4.165	.001	.017
.25	.199	.203	3.942	.003	.079	.201	4.417	.001	.044	.199	5.017	.000	.020
.30	.328	.321	8.273	-.006	.165	.330	10.125	.002	.101	.329	13.339	.002	.053
.40	1.053	1.069	85.593	.016	1.712	1.053	122.838	.001	1.228	1.050	221.536	-.003	.886
.45	2.603	2.613	543.514	.010	10.870	2.604	1002.759	.001	10.028	2.559	1972.631	-.045	7.893
.49	15.299	15.349	23186.893	.050	463.740	14.999	42931.116	-.301	429.401	15.357	106334.001	.058	425.339

Note : "γ_4" denotes the true value of the (population) four-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_4$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_4$) × T. "Sample Bias" denotes (Mean of $\hat{\gamma}_4$ - γ_4). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

5(a). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_5$ using Sample Mean

d	γ_5	T = 50					T = 100					T = 250				
		Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.013	-.014	2.198	.000	-.014	.044	-.012	2.111	.001	-.004	.021	-.013	1.973	.000	-.001	.008
-.40	-.016	-.017	1.851	-.002	-.003	.037	-.017	1.741	-.001	-.001	.017	-.015	1.717	.000	.000	.007
-.30	-.018	-.023	1.602	-.006	-.004	.032	-.019	1.489	-.002	-.001	.015	-.017	1.395	.001	.000	.006
-.20	-.017	-.025	1.338	-.007	-.006	.027	-.019	1.286	-.001	-.002	.013	-.018	1.218	.000	-.001	.005
-.10	-.013	-.025	1.189	-.012	-.010	.024	-.017	1.110	-.004	-.004	.011	-.014	1.057	-.001	-.002	.004
.00	.000	-.018	1.125	-.018	-.020	.023	-.012	1.046	-.012	-.010	.011	-.003	1.023	-.003	-.004	.004
.10	.032	-.012	1.133	-.044	-.042	.025	.008	1.095	-.024	-.024	.012	.019	1.110	-.012	-.011	.005
.20	.106	.011	1.301	-.095	-.095	.035	.044	1.485	-.062	-.063	.019	.069	1.693	-.037	-.036	.008
.24	.161	.024	1.472	-.137	-.137	.048	.065	1.858	-.096	-.095	.028	.101	2.230	-.060	-.059	.013
.25	.178	.025	1.556	-.153	-.150	.055	.072	1.899	-.107	-.106	.030	.111	2.451	-.067	-.067	.014
.30	.300	.044	1.853	-.256	-.249	.102	.111	2.701	-.189	-.189	.063	.168	3.945	-.131	-.131	.033
.40	1.007	.118	3.570	-.889	-.883	.862	.232	6.427	-.775	-.768	.665	.368	13.534	-.639	-.640	.463
.45	2.546	.174	5.273	-2.372	-2.366	5.733	.340	11.253	-2.206	-2.207	4.980	.530	26.554	-2.016	-2.014	4.170
.49	15.231	.223	7.028	-15.008	-14.993	225.382	.444	17.746	-14.787	-14.787	218.830	.713	46.145	-14.519	-14.518	210.980

Note : "γ_5" denotes the true value of the (population) five-period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_5$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_5$) × T. "Sample Bias" denotes the mean of ($\hat{\gamma}_5 - \gamma_5$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-2 (CONTINUED)

5(b). Mean, Normalized Variance, Bias and MSE of $\hat{\gamma}_5$ using True Mean

		T = 50				T = 100				T = 250			
d	γ_5	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.013	-.013	2.102	.001	.042	-.011	2.067	.002	.021	-.013	1.957	.000	.008
-.40	-.016	-.016	1.770	.000	.035	-.016	1.705	-.001	.017	-.015	1.703	.000	.007
-.30	-.018	-.020	1.533	-.002	.031	-.018	1.458	-.001	.015	-.017	1.384	.001	.006
-.20	-.017	-.019	1.283	-.002	.026	-.017	1.259	.001	.013	-.017	1.208	.000	.005
-.10	-.013	-.014	1.149	-.001	.023	-.013	1.092	.000	.011	-.012	1.049	.001	.004
.00	.000	.002	1.110	.002	.022	-.002	1.046	-.002	.010	.001	1.024	.001	.004
.10	.032	.030	1.265	-.001	.025	.032	1.194	.001	.012	.030	1.176	-.001	.005
.20	.106	.109	2.279	.003	.046	.108	2.208	.002	.022	.105	2.336	-.001	.009
.24	.161	.165	3.686	.004	.074	.159	3.646	-.001	.036	.160	4.079	-.001	.016
.25	.178	.182	3.992	.004	.080	.179	4.367	.001	.044	.179	4.949	.001	.020
.30	.300	.296	8.214	-.004	.164	.304	10.019	.004	.100	.302	13.212	.002	.053
.40	1.007	1.024	85.713	.017	1.715	1.007	122.755	.000	1.228	1.004	221.141	-.003	.885
.45	2.546	2.557	545.241	.011	10.905	2.549	1004.582	.003	10.046	2.500	1974.605	-.046	7.901
.49	15.231	15.278	23183.092	.047	463.664	14.932	42953.214	-.299	429.622	15.288	106334.339	.057	425.341

Note : "γ_5" denotes the true value of the (population) five period autocovariance. "Mean" denotes the mean of $\hat{\gamma}_5$. "Normalized Variance" denotes (finite sample variance of $\hat{\gamma}_5$) × T. "Sample Bias" denotes (Mean of $\hat{\gamma}_5 - \gamma_5$). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3

THE SAMPLE AUTOCORRELATIONS OF THE I(d) PROCESS

1(a). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_1$ using Sample Mean

		T = 50					T = 100					T = 250				
d	ρ_1	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.329	-.322	.689	.007	-.015	.014	-.325	.675	.004	-.004	.007	-.327	.697	.002	-.001	.003
-.40	-.286	-.281	.729	.005	-.003	.015	-.283	.734	.003	-.001	.007	-.284	.747	.001	.000	.003
-.30	-.231	-.228	.766	.002	-.004	.015	-.231	.789	-.001	-.001	.008	-.231	.763	.000	.000	.003
-.20	-.167	-.167	.838	.000	-.006	.017	-.167	.861	.000	-.002	.009	-.167	.834	.000	-.001	.003
-.10	-.091	-.102	.912	-.011	-.011	.018	-.095	.923	-.004	-.005	.009	-.091	.900	-.001	-.002	.004
.00	.000	-.021	.998	-.021	-.020	.020	-.011	.978	-.011	-.010	.010	-.004	1.018	-.004	-.004	.004
.10	.111	.069	1.035	-.043	-.036	.023	.087	1.091	-.024	-.021	.012	.100	1.112	-.011	-.010	.005
.20	.250	.168	1.131	-.082	-.065	.029	.198	1.190	-.052	-.043	.015	.222	1.267	-.028	-.025	.006
.24	.316	.209	1.194	-.106	-.081	.035	.244	1.271	-.071	-.056	.018	.274	1.353	-.042	-.035	.007
.25	.333	.223	1.164	-.110	-.085	.035	.258	1.265	-.075	-.060	.018	.290	1.375	-.044	-.038	.007
.30	.429	.275	1.235	-.153	-.108	.048	.320	1.314	-.109	-.082	.025	.359	1.505	-.069	-.057	.011
.40	.667	.391	1.245	-.276	-.142	.101	.448	1.396	-.218	-.124	.062	.507	1.706	-.160	-.103	.032
.45	.818	.446	1.245	-.372	-.118	.163	.514	1.422	-.304	-.110	.107	.578	1.742	-.241	-.101	.065
.49	.961	.489	1.226	-.471	-.036	.247	.564	1.355	-.396	-.035	.171	.634	1.697	-.326	-.035	.113

Note : "ρ_1" denotes the true value of the (population) one-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_1$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_1$) × T. "Sample Bias" denotes the mean of ($\hat{\rho}_1 - \rho_1$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

1(b). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_1$ using True Mean

		T = 50				T = 100				T = 250			
d	ρ_1	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.329	-.320	.688	.009	.014	-.325	.675	.004	.007	-.327	.697	.002	.003
-.40	-.286	-.278	.729	.007	.015	-.282	.734	.004	.007	-.284	.747	.002	.003
-.30	-.231	-.225	.764	.006	.015	-.230	.788	.001	.008	-.230	.763	.000	.003
-.20	-.167	-.161	.836	.006	.017	-.164	.859	.002	.009	-.166	.834	.001	.003
-.10	-.091	-.091	.909	.000	.018	-.090	.921	.001	.009	-.090	.899	.001	.004
.00	.000	.000	.997	.000	.020	-.001	.979	-.001	.010	.000	1.015	.000	.004
.10	.111	.106	1.061	-.005	.021	.108	1.122	-.003	.011	.110	1.127	-.001	.005
.20	.250	.235	1.229	-.015	.025	.241	1.331	-.009	.013	.246	1.430	-.004	.006
.24	.316	.293	1.374	-.023	.028	.300	1.477	-.016	.015	.309	1.677	-.007	.007
.25	.333	.311	1.359	-.022	.028	.318	1.558	-.015	.016	.328	1.752	-.006	.007
.30	.429	.390	1.526	-.039	.032	.404	1.784	-.024	.018	.416	2.301	-.013	.009
.40	.667	.584	1.755	-.082	.042	.601	2.357	-.065	.028	.621	3.598	-.046	.016
.45	.818	.703	1.733	-.115	.048	.720	2.473	-.098	.034	.739	4.045	-.079	.022
.49	.961	.852	1.342	-.109	.039	.863	1.936	-.098	.029	.878	3.427	-.083	.021

Note : "ρ_1" denotes the true value of the (population) one-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_1$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_1$) × T. "Sample Bias" denotes (Mean of $\hat{\rho}_1 - \rho_1$). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

2(a). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_2$ using Sample Mean

		T = 50					T = 100					T = 250				
d	ρ_2	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.067	-.071	1.218	-.004	-.012	.024	-.069	1.218	-.001	-.003	.012	-.069	1.257	-.002	-.001	.005
-.40	-.071	-.076	1.166	-.004	-.003	.023	-.071	1.165	.001	-.001	.012	-.072	1.177	-.001	.000	.005
-.30	-.070	-.075	1.054	-.005	-.004	.021	-.068	1.097	.002	-.001	.011	-.069	1.081	.001	.000	.004
-.20	-.061	-.064	1.022	-.004	-.006	.020	-.062	1.022	-.001	-.002	.010	-.062	1.021	-.001	-.001	.004
-.10	-.039	-.047	1.010	-.008	-.011	.020	-.043	.997	-.004	-.005	.010	-.040	1.022	-.001	-.002	.004
.00	.000	-.020	1.010	-.020	-.020	.021	-.010	1.012	-.010	-.010	.010	-.003	.997	-.003	-.004	.004
.10	.064	.021	1.067	-.043	-.038	.023	.042	1.101	-.022	-.022	.012	.054	1.056	-.011	-.011	.004
.20	.167	.079	1.126	-.088	-.072	.030	.111	1.263	-.056	-.048	.016	.136	1.351	-.031	-.028	.006
.24	.222	.106	1.263	-.117	-.092	.039	.145	1.395	-.078	-.064	.020	.176	1.543	-.047	-.040	.008
.25	.238	.114	1.250	-.124	-.097	.040	.154	1.395	-.084	-.069	.021	.188	1.567	-.050	-.043	.009
.30	.328	.151	1.356	-.177	-.127	.058	.202	1.552	-.126	-.096	.031	.247	1.861	-.081	-.067	.014
.40	.583	.242	1.590	-.341	-.178	.148	.313	1.878	-.271	-.155	.092	.383	2.499	-.200	-.129	.050
.45	.765	.289	1.689	-.476	-.152	.260	.376	2.074	-.390	-.142	.173	.455	2.761	-.310	-.130	.107
.49	.948	.327	1.782	-.621	-.048	.421	.425	2.134	-.523	-.047	.295	.516	2.836	-.432	-.046	.198

Note : "ρ_2" denotes the true value of the (population) two-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_2$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_2$) × T. "Sample Bias" denotes the mean of ($\hat{\rho}_2$ - ρ_2). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

2(b). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_2$ using True Mean

		T = 50				T = 100				T = 250			
d	ρ_2	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.067	-.070	1.213	-.003	.024	-.068	1.216	-.001	.012	-.069	1.257	-.002	.005
-.40	-.071	-.074	1.160	-.003	.023	-.070	1.163	.001	.012	-.072	1.177	-.001	.005
-.30	-.070	-.072	1.047	-.002	.021	-.067	1.094	.003	.011	-.069	1.080	.001	.004
-.20	-.061	-.059	1.015	.002	.020	-.060	1.020	.001	.010	-.061	1.020	.000	.004
-.10	-.039	-.036	1.000	.003	.020	-.038	.990	.001	.010	-.038	1.020	.001	.004
.00	.000	.001	1.002	.001	.020	.000	1.008	.000	.010	.001	.996	.001	.004
.10	.064	.061	1.100	-.004	.022	.064	1.148	.000	.011	.064	1.088	.000	.004
.20	.167	.153	1.318	-.013	.027	.159	1.462	-.008	.015	.163	1.567	-.004	.006
.24	.222	.200	1.575	-.023	.032	.208	1.717	-.015	.017	.215	1.984	-.007	.008
.25	.238	.216	1.587	-.023	.032	.223	1.822	-.015	.018	.231	2.094	-.007	.008
.30	.328	.285	1.861	-.043	.039	.302	2.310	-.026	.024	.313	3.001	-.015	.012
.40	.583	.482	2.606	-.101	.062	.503	3.531	-.080	.042	.526	5.590	-.058	.026
.45	.765	.618	2.802	-.147	.078	.640	4.004	-.125	.056	.664	6.713	-.102	.037
.49	.948	.804	2.313	-.144	.067	.819	3.367	-.129	.050	.839	5.950	-.109	.036

Note : "ρ_2" denotes the true value of the (population) two-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_2$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_2$) $\times$ T. "Sample Bias" denotes (Mean of $\hat{\rho}_2 - \rho_2$). "MSE" denotes (Sample Bias $\times$ Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

3(a). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_3$ using Sample Mean

		T = 50					T = 100					T = 250				
d	ρ_3	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.029	-.032	1.230	-.003	-.011	.025	-.030	1.213	-.001	-.003	.012	-.028	1.273	.001	-.001	.005
-.40	-.034	-.035	1.177	-.002	-.003	.024	-.037	1.190	-.004	-.001	.012	-.034	1.156	.000	.000	.005
-.30	-.036	-.037	1.159	-.001	-.003	.023	-.037	1.132	-.001	-.001	.011	-.038	1.104	-.001	.000	.004
-.20	-.034	-.041	1.054	-.007	-.006	.021	-.037	1.049	-.003	-.002	.010	-.034	1.068	.000	-.001	.004
-.10	-.024	-.035	1.046	-.011	-.010	.021	-.028	1.024	-.005	-.004	.010	-.026	1.023	-.002	-.002	.004
.00	.000	-.021	1.036	-.021	-.020	.021	-.009	.997	-.009	-.010	.010	-.003	.994	-.003	-.004	.004
.10	.047	.005	1.069	-.042	-.039	.023	.023	1.067	-.024	-.022	.011	.035	1.075	-.011	-.011	.004
.20	.131	.039	1.180	-.092	-.075	.032	.074	1.232	-.057	-.050	.016	.097	1.376	-.034	-.029	.007
.24	.181	.060	1.231	-.120	-.097	.039	.100	1.378	-.080	-.067	.020	.133	1.582	-.048	-.042	.009
.25	.195	.065	1.274	-.129	-.103	.042	.107	1.418	-.088	-.073	.022	.142	1.652	-.053	-.046	.009
.30	.279	.092	1.367	-.188	-.136	.062	.146	1.644	-.134	-.103	.034	.193	1.985	-.086	-.072	.015
.40	.538	.163	1.655	-.375	-.197	.174	.241	2.079	-.298	-.171	.110	.316	2.934	-.222	-.143	.061
.45	.735	.203	1.858	-.532	-.172	.321	.298	2.401	-.438	-.160	.216	.387	3.382	-.349	-.146	.135
.49	.941	.235	1.982	-.705	-.054	.537	.343	2.581	-.597	-.054	.383	.446	3.579	-.495	-.053	.259

Note : "ρ_3" denotes the true value of the (population) three-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_3$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_3$) × T. "Sample Bias" denotes the mean of ($\hat{\rho}_3$ - ρ_3). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

3(b). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_3$ using True Mean

d	ρ_3	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.029	-.031	1.224	-.002	.024	-.030	1.212	.000	.012	-.028	1.272	.001	.005
-.40	-.034	-.034	1.169	.000	.023	-.037	1.189	-.003	.012	-.034	1.155	.000	.005
-.30	-.036	-.035	1.151	.002	.023	-.036	1.130	.000	.011	-.037	1.103	-.001	.004
-.20	-.034	-.036	1.041	-.002	.021	-.035	1.044	-.001	.010	-.034	1.068	.001	.004
-.10	-.024	-.024	1.037	-.001	.021	-.024	1.020	.000	.010	-.024	1.021	.000	.004
.00	.000	-.001	1.035	-.001	.021	.002	1.000	.002	.010	.001	.998	.001	.004
.10	.047	.045	1.111	-.002	.022	.045	1.117	-.001	.011	.046	1.103	.000	.004
.20	.131	.117	1.414	-.014	.028	.124	1.455	-.007	.015	.126	1.634	-.005	.007
.24	.181	.159	1.639	-.022	.033	.167	1.771	-.014	.018	.174	2.102	-.006	.008
.25	.195	.173	1.671	-.022	.034	.180	1.946	-.015	.020	.188	2.271	-.007	.009
.30	.279	.234	2.066	-.045	.043	.252	2.577	-.027	.027	.264	3.312	-.015	.013
.40	.538	.428	3.063	-.111	.074	.451	4.225	-.087	.050	.475	6.774	-.064	.031
.45	.735	.571	3.464	-.164	.096	.595	4.965	-.140	.069	.621	8.446	-.114	.047
.49	.941	.776	2.982	-.164	.087	.793	4.366	-.148	.065	.815	7.783	-.125	.047

Note : "ρ_3" denotes the true value of the (population) three-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_3$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_3$) $\times$ T. "Sample Bias" denotes (Mean of $\hat{\rho}_3$ - ρ_3). "MSE" denotes (Sample Bias $\times$ Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

4(a). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_4$ using Sample Mean

d	ρ_4	T = 50					T = 100					T = 250				
		Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.016	-.018	1.234	-.002	-.011	.025	-.020	1.209	-.003	-.003	.012	-.016	1.223	.000	-.001	.005
-.40	-.020	-.021	1.221	-.001	-.003	.024	-.019	1.195	.001	-.001	.012	-.021	1.189	-.001	.000	.005
-.30	-.023	-.028	1.189	-.005	-.003	.024	-.026	1.166	-.003	-.001	.012	-.024	1.121	-.001	.000	.004
-.20	-.023	-.028	1.108	-.005	-.006	.022	-.025	1.065	-.002	-.002	.011	-.023	1.075	.000	-.001	.004
-.10	-.017	-.028	1.063	-.011	-.010	.021	-.021	1.028	-.005	-.004	.010	-.019	1.025	-.002	-.002	.004
.00	.000	-.020	1.019	-.020	-.020	.021	-.010	1.022	-.010	-.010	.010	-.004	1.013	-.004	-.004	.004
.10	.037	-.004	1.096	-.041	-.039	.024	.013	1.094	-.025	-.023	.012	.025	1.070	-.012	-.011	.004
.20	.110	.018	1.158	-.092	-.077	.032	.054	1.255	-.057	-.051	.016	.077	1.351	-.033	-.029	.007
.24	.156	.032	1.226	-.123	-.100	.040	.074	1.410	-.081	-.069	.021	.107	1.624	-.049	-.043	.009
.25	.169	.037	1.268	-.132	-.106	.043	.080	1.392	-.089	-.075	.022	.114	1.689	-.055	-.047	.010
.30	.249	.055	1.363	-.194	-.142	.065	.111	1.646	-.138	-.108	.035	.159	2.048	-.090	-.075	.016
.40	.509	.111	1.675	-.397	-.210	.191	.192	2.218	-.316	-.182	.122	.273	3.182	-.235	-.152	.068
.45	.715	.145	1.882	-.570	-.185	.363	.244	2.558	-.470	-.173	.247	.340	3.759	-.375	-.158	.155
.49	.935	.171	2.031	-.764	-.059	.624	.286	2.832	-.649	-.059	.450	.396	4.076	-.539	-.058	.307

Note : "ρ_4" denotes the true value of the (population) four-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_4$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_4$) × T. "Sample Bias" denotes the mean of ($\hat{\rho}_4$ - ρ_4). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

4(b). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_4$ using True Mean

d	ρ_4	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE
-.49	-.016	-.018	1.227	-.001	.025	-.019	1.207	-.003	.012	-.016	1.223	.000	.005
-.40	-.020	-.020	1.213	.000	.024	-.019	1.193	.001	.012	-.021	1.189	-.001	.005
-.30	-.023	-.025	1.179	-.002	.024	-.025	1.163	-.002	.012	-.023	1.120	-.001	.004
-.20	-.023	-.023	1.095	.000	.022	-.023	1.062	.000	.011	-.023	1.073	.000	.004
-.10	-.017	-.018	1.050	-.001	.021	-.017	1.021	.000	.010	-.017	1.023	.000	.004
.00	.000	.001	1.013	.001	.020	.000	1.026	.000	.010	.000	1.014	.000	.004
.10	.037	.036	1.147	-.001	.023	.036	1.145	-.001	.011	.036	1.099	-.001	.004
.20	.110	.097	1.455	-.013	.029	.105	1.524	-.005	.015	.106	1.625	-.004	.007
.24	.156	.134	1.703	-.022	.035	.143	1.847	-.013	.019	.150	2.165	-.006	.009
.25	.169	.147	1.739	-.022	.035	.155	2.003	-.014	.020	.161	2.357	-.008	.009
.30	.249	.204	2.146	-.045	.045	.222	2.689	-.027	.028	.233	3.533	-.016	.014
.40	.509	.392	3.403	-.117	.082	.416	4.767	-.092	.056	.442	7.594	-.067	.035
.45	.715	.539	3.922	-.175	.109	.564	5.697	-.150	.080	.593	9.736	-.122	.054
.49	.935	.757	3.485	-.178	.101	.775	5.153	-.161	.077	.799	9.203	-.136	.055

Note : "ρ_4" denotes the true value of the (population) four-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_4$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_4$) $\times$ T. "Sample Bias" denotes (Mean of $\hat{\rho}_4 - \rho_4$). "MSE" denotes (Sample Bias $\times$ Sample Bias + Normalized Variance / T)

TABLE 4-3 (CONTINUED)

5(a). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_5$ using Sample Mean

		T = 50					T = 100					T = 250				
d	ρ_5	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE	Mean	Normalized Variance	Sample Bias	Asymp. Bias	MSE
-.49	-.010	-.011	1.247	-.001	-.011	.025	-.010	1.258	.001	-.003	.013	-.011	1.210	.000	-.001	.005
-.40	-.013	-.015	1.209	-.002	-.003	.024	-.014	1.183	-.001	-.001	.012	-.013	1.202	.000	.000	.005
-.30	-.016	-.021	1.192	-.005	-.003	.024	-.017	1.150	-.001	-.001	.012	-.015	1.112	.001	.000	.004
-.20	-.017	-.024	1.127	-.007	-.006	.023	-.018	1.108	-.001	-.002	.011	-.017	1.084	.000	-.001	.004
-.10	-.013	-.024	1.085	-.011	-.010	.022	-.017	1.045	-.004	-.004	.010	-.013	1.009	.000	-.002	.004
.00	.000	-.018	1.083	-.018	-.020	.022	-.012	1.027	-.012	-.010	.010	-.003	1.009	-.003	-.004	.004
.10	.031	-.013	1.093	-.044	-.040	.024	.007	1.060	-.024	-.023	.011	.018	1.065	-.013	-.011	.004
.20	.096	.007	1.151	-.089	-.078	.031	.039	1.268	-.057	-.052	.016	.063	1.373	-.034	-.030	.007
.24	.139	.018	1.197	-.121	-.102	.039	.056	1.405	-.083	-.071	.021	.088	1.569	-.051	-.044	.009
.25	.151	.018	1.261	-.133	-.108	.043	.060	1.399	-.091	-.076	.022	.096	1.671	-.055	-.048	.010
.30	.228	.033	1.335	-.195	-.146	.065	.088	1.656	-.139	-.111	.036	.136	2.079	-.092	-.077	.017
.40	.486	.076	1.650	-.410	-.219	.201	.157	2.255	-.329	-.191	.131	.241	3.349	-.245	-.159	.074
.45	.699	.103	1.828	-.596	-.196	.392	.205	2.645	-.494	-.182	.270	.303	4.033	-.396	-.167	.173
.49	.931	.123	2.020	-.808	-.063	.693	.243	2.977	-.688	-.062	.503	.358	4.447	-.573	-.061	.346

Note : "ρ_5" denotes the true value of the (population) four-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_5$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_5$) × T. "Sample Bias" denotes the mean of ($\hat{\rho}_5 - \rho_5$). "Asymptotic Bias" denotes the theoretical finite sample bias for d > 0, based on Hosking (1984). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T).

TABLE 4-3 (CONTINUED)

5(b). Mean, Normalized Variance, Bias and MSE of $\hat{\rho}_5$ using True Mean

d	ρ_5	T = 50 Mean	T = 50 Normalized Variance	T = 50 Sample Bias	T = 50 MSE	T = 100 Mean	T = 100 Normalized Variance	T = 100 Sample Bias	T = 100 MSE	T = 250 Mean	T = 250 Normalized Variance	T = 250 Sample Bias	T = 250 MSE
-.49	-.010	-.011	1.239	.000	.025	-.010	1.256	.001	.013	-.011	1.209	.000	.005
-.40	-.013	-.014	1.199	-.001	.024	-.014	1.181	-.001	.012	-.013	1.201	.000	.005
-.30	-.016	-.018	1.181	-.002	.024	-.016	1.147	-.001	.011	-.015	1.111	.001	.004
-.20	-.017	-.018	1.112	-.002	.022	-.016	1.102	.001	.011	-.016	1.082	.000	.004
-.10	-.013	-.014	1.071	-.001	.021	-.012	1.041	.000	.010	-.012	1.007	.001	.004
.00	.000	.002	1.074	.002	.021	-.002	1.028	-.002	.010	.001	1.011	.001	.004
.10	.031	.028	1.161	-.003	.023	.030	1.113	-.001	.011	.029	1.104	-.002	.004
.20	.096	.088	1.467	-.009	.029	.092	1.536	-.005	.015	.093	1.649	-.004	.007
.24	.139	.121	1.726	-.018	.035	.126	1.896	-.013	.019	.131	2.175	-.007	.009
.25	.151	.130	1.798	-.021	.036	.137	2.044	-.014	.021	.144	2.394	-.007	.010
.30	.228	.185	2.221	-.043	.046	.202	2.791	-.026	.029	.212	3.653	-.016	.015
.40	.486	.367	3.645	-.119	.087	.391	5.147	-.096	.061	.417	8.214	-.069	.038
.45	.699	.516	4.260	-.183	.119	.541	6.294	-.158	.088	.570	10.876	-.129	.060
.49	.931	.742	3.898	-.189	.114	.761	5.774	-.170	.087	.786	10.394	-.145	.063

Note : "ρ_5" denotes the true value of the (population) five-period autocorrelation. "Mean" denotes the mean of $\hat{\rho}_5$. "Normalized Variance" denotes (finite sample variance of $\hat{\rho}_5$) $\times$ T. "Sample Bias" denotes (Mean of $\hat{\rho}_5 - \rho_5$). "MSE" denotes (Sample Bias $\times$ Sample Bias + Normalized Variance / T)

TABLE 4-4

THE MDE $\hat{d}_1$ IN THE I(d) MODEL

1. $\hat{d}_1$ using Sample Mean

	T = 50				T = 100				T = 250				
d	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Asymp. Variance
-.49	-.521	3.843	-.031	.078	-.504	3.541	-.014	.036	-.495	3.506	-.005	.014	3.401
-.40	-.432	3.240	-.032	.066	-.415	3.028	-.015	.030	-.406	2.930	-.006	.012	2.821
-.30	-.331	2.552	-.031	.052	-.319	2.469	-.019	.025	-.307	2.255	-.007	.009	2.251
-.20	-.231	2.056	-.031	.042	-.215	1.945	-.015	.020	-.206	1.780	-.006	.007	1.758
-.10	-.141	1.653	-.041	.035	-.118	1.503	-.018	.015	-.105	1.355	-.005	.005	1.341
.00	-.044	1.290	-.044	.028	-.022	1.110	-.022	.012	-.008	1.072	-.008	.004	1.000
.10	.046	.944	-.054	.022	.071	.853	-.029	.009	.088	.785	-.012	.003	.749
.20	.129	.714	-.071	.019	.158	.627	-.042	.008	.179	.588	-.021	.003	.688
.24	.159	.658	-.081	.020	.190	.574	-.050	.008	.213	.531	-.027	.003	.863
.25	.169	.618	-.081	.019	.199	.547	-.051	.008	.222	.512	-.028	.003	
.30	.203	.554	-.097	.020	.236	.472	-.064	.009	.262	.452	-.038	.003	
.40	.271	.388	-.129	.024	.305	.345	-.095	.013	.334	.343	-.066	.006	
.45	.300	.341	-.150	.029	.335	.298	-.115	.016	.364	.292	-.086	.009	
.49	.321	.297	-.169	.035	.357	.246	-.133	.020	.387	.247	-.103	.012	

Note : "Mean" denotes the mean of $\hat{d}_1$. "Normalized Variance" denotes (finite sample variance of $\hat{d}_1$) × T. "Sample Bias" denotes (Mean of $\hat{d}_1$ - d). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T). "Asymptotic Variance" is the theoretical variance in the limiting distribution, as given by Tieslau, Schmidt and Baillie (1994).

TABLE 4-4 (CONTINUED)

2. $\hat{d}_1$ using True Mean

	T = 50				T = 100				T = 250				
d	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Mean	Normalized Variance	Sample Bias	MSE	Asymp. Variance
-.49	-.517	3.800	-.027	.077	-.503	3.532	-.013	.036	-.495	3.504	-.005	.014	3.401
-.40	-.427	3.196	-.027	.065	-.413	3.017	-.013	.030	-.405	2.928	-.005	.012	2.821
-.30	-.324	2.496	-.024	.051	-.317	2.449	-.017	.025	-.306	2.251	-.006	.009	2.251
-.20	-.221	1.979	-.021	.040	-.212	1.916	-.012	.019	-.205	1.773	-.005	.007	1.758
-.10	-.126	1.560	-.026	.032	-.112	1.467	-.012	.015	-.103	1.343	-.003	.005	1.341
.00	-.021	1.184	-.021	.024	-.011	1.066	-.011	.011	-.004	1.052	-.004	.004	1.000
.10	.079	.839	-.021	.017	.089	.808	-.011	.008	.096	.766	-.004	.003	.749
.20	.176	.618	-.024	.013	.187	.605	-.013	.006	.195	.609	-.005	.002	.688
.24	.213	.574	-.027	.012	.224	.553	-.016	.006	.233	.585	-.007	.002	.863
.25	.225	.542	-.025	.011	.234	.553	-.016	.006	.244	.572	-.006	.002	
.30	.268	.488	-.032	.011	.281	.491	-.019	.005	.290	.576	-.010	.002	
.40	.359	.338	-.041	.008	.370	.395	-.030	.005	.380	.533	-.020	.003	
.45	.405	.264	-.045	.007	.413	.329	-.037	.005	.422	.474	-.028	.003	
.49	.455	.165	-.035	.005	.460	.206	-.030	.003	.465	.330	-.025	.002	

Note : "Mean" denotes the mean of $\hat{d}_1$. "Normalized Variance" denotes (finite sample variance of $\hat{d}_1$) × T. "Sample Bias" denotes (Mean of $\hat{d}_1$ - d). "MSE" denotes (Sample Bias × Sample Bias + Normalized Variance / T). "Asymptotic Variance" is the theoretical variance in the limiting distribution, as given by Tieslau, Schmidt and Baillie (1994).

TABLE 4-5

MLE AND MDE IN THE I(d) MODEL

1(a). Mean and Bias for T = 50

d	MLEμ	MLE$\bar{y}$	F&T	WL	MDE$_1$	MDE$_2$	MDE$_3$	MDE$_4$	MDE$_5$	MDE$_2^*$	MDE$_3^*$	MDE$_4^*$	MDE$_5^*$
						Mean							
-.4	-.422	-.452	-.499	-.404	-.453	-.426	-.415	-.411	-.411	-.461	-.472	-.472	-.475
-.3	-.322	-.364	-.406	-.310	-.350	-.327	-.318	-.315	-.315	-.359	-.364	-.366	-.371
-.2	-.223	-.274	-.311	-.214	-.249	-.230	-.223	-.221	-.221	-.257	-.263	-.266	-.269
-.1	-.124	-.182	-.214	-.117	-.150	-.134	-.130	-.129	-.129	-.159	-.165	.168	-.171
.0	-.025	-.088	-.115	-.018	-.054	-.042	-.039	-.039	-.040	-.065	-.071	-.073	-.075
.1	.073	.006	-.014	.082	.037	.047	.049	.048	.048	.023	.020	.019	.018
.2	.171	.099	.088	.184	.123	.131	.133	.133	.134	.101	.108	.112	.114
.3	.269	.192	.191	.286	.200	.209	.216	.217	.221	.200	.228	.235	.240
.4	.372	.292	.297	.390	.268	.278	.291	.297	.302	.311	.336	.341	.343
						Bias × 100							
-.4	-2.173	-5.211	-9.875	-.376	-5.327	-2.633	-1.527	-1.065	-1.058	-6.114	-7.179	-7.169	-7.528
-.3	-2.242	-6.421	-10.600	-1.005	-5.044	-2.731	-1.845	-1.495	-1.513	-5.867	-6.445	-6.641	-7.117
-.2	-2.320	-7.389	-11.093	-1.438	-4.909	-2.978	-2.316	-2.082	-2.130	-5.735	-6.344	-6.645	-6.944
-.1	-2.410	-8.171	-11.385	-1.701	-4.993	-3.437	-2.992	-2.873	-2.946	-5.948	-6.545	-6.800	-7.093
.0	-2.527	-8.832	-11.491	-1.821	-5.398	-4.186	-3.914	-3.894	-3.978	-6.549	-7.065	-7.279	-7.539
.1	-2.690	-9.449	-11.437	-1.796	-6.260	-5.272	-5.141	-5.184	-5.234	-7.714	-8.018	-8.073	-8.217
.2	-2.910	-10.088	-11.238	-1.644	-7.741	-6.883	-6.670	-6.683	-6.586	-9.887	-9.162	-8.780	-8.612
.3	-3.134	-10.797	-10.858	-1.363	-10.008	-9.065	-8.443	-8.259	-7.904	-10.011	-7.186	-6.459	-5.987
.4	-2.768	-10.780	-10.326	-.950	-13.208	-12.247	-10.921	-10.343	-9.806	-8.863	-6.392	-5.926	-5.711

Note : "Mean" denotes the mean of the estimate. "Bias" denotes the mean of (estimate - d).

TABLE 4-5 (CONTINUED)

1(b). Normalized MSE and Normalized Variance for T = 50

d	MLEμ	MLE$\bar{y}$	F&T	WL	MDE_1	MDE_2	MDE_3	MDE_4	MDE_5	MDE_2^*	MDE_3^*	MDE_4^*	MDE_5^*
						Normalized MSE							
-.4	.901	1.101	1.747	1.183	3.703	2.243	1.691	1.546	1.517	4.863	5.624	4.942	4.947
-.3	.876	1.183	1.795	1.169	2.997	1.898	1.473	1.379	1.361	3.864	3.751	3.652	3.747
-.2	.850	1.262	1.830	1.162	2.401	1.599	1.294	1.241	1.230	2.995	2.922	2.909	2.885
-.1	.820	1.331	1.850	1.159	1.910	1.353	1.154	1.131	1.126	2.379	2.292	2.265	2.253
.0	.784	1.388	1.857	1.157	1.525	1.168	1.061	1.059	1.058	1.736	1.826	1.808	1.803
.1	.741	1.427	1.849	1.156	1.248	1.061	1.018	1.029	1.035	1.725	1.591	1.543	1.528
.2	.699	1.458	1.827	1.154	1.091	1.046	1.043	1.063	1.073	1.897	1.571	1.438	1.392
.3	.642	1.493	1.781	1.147	1.086	1.169	1.141	1.140	1.143	2.436	1.668	1.436	1.360
.4	.583	1.639	1.718	1.131	1.295	1.513	1.303	1.240	1.190	1.955	1.379	1.214	1.151
						Normalized Variance							
-.4	.878	.967	1.260	1.184	3.565	2.211	1.681	1.542	1.513	4.680	5.372	4.689	4.668
-.3	.852	.978	1.235	1.166	2.873	1.863	1.458	1.369	1.351	3.695	3.547	3.435	3.497
-.2	.824	.990	1.216	1.153	2.282	1.557	1.268	1.220	1.209	2.833	2.723	2.691	2.646
-.1	.792	.999	1.204	1.146	1.787	1.295	1.110	1.091	1.084	2.204	2.080	2.036	2.003
.0	.752	.998	1.197	1.141	1.379	1.082	.984	.983	.979	1.521	1.577	1.543	1.519
.1	.706	.982	1.196	1.141	1.053	.923	.887	.896	.899	1.429	1.271	1.218	1.192
.2	.657	.950	1.197	1.142	.793	.810	.821	.841	.857	1.409	1.153	1.053	1.022
.3	.593	.911	1.193	1.139	.586	.759	.785	.800	.831	1.937	1.411	1.229	1.181
.4	.545	1.059	1.186	1.127	.424	.764	.708	.706	.710	1.564	1.176	1.039	.989

Note : "Normalized MSE" denotes T × the mean of (estimate - d)^2. "Normalized Variance" denotes T × (MSE - Bias × Bias).

TABLE 4-5 (CONTINUED)

2(a). Mean and Bias for T = 100

d	MLEμ	MLE$\bar{y}$	F&T	WL	MDE_1	MDE_2	MDE_3	MDE_4	MDE_5	MDE_2^*	MDE_3^*	MDE_4^*	MDE_5^*
						Mean							
-.4	-.416	-.431	-.447	-.402	-.430	-.415	-.409	-.405	-.403	-.425	-.427	-.427	-.427
-.3	-.316	-.338	-.352	-.308	-.328	-.315	-.310	-.307	-.306	-.325	-.327	-.327	-.327
-.2	-.216	-.243	-.256	-.211	-.227	-.216	-.213	-.210	-.209	-.226	-.228	-.228	-.229
-.1	-.116	-.147	-.158	-.113	-.128	-.119	-.116	-.115	-.114	-.128	-.130	-.131	-.132
.0	-.017	-.050	-.059	-.014	-.030	-.023	-.022	-.021	-.021	-.032	-.035	-.036	-.037
.1	.083	.048	.041	.086	.063	.069	.070	.070	.070	.060	.057	.056	.055
.2	.182	.145	.142	.187	.152	.157	.158	.159	.160	.144	.146	.146	.147
.3	.281	.240	.245	.290	.232	.239	.245	.248	.250	.235	.255	.259	.262
.4	.383	.336	.349	.393	.302	.313	.327	.332	.337	.345	.363	.365	.366
						Bias × 100							
-.4	-1.581	-3.062	-4.670	-.245	-3.013	-1.452	-.907	-.533	-.337	-2.496	-2.709	-2.688	-2.667
-.3	-1.598	-3.799	-5.249	-.776	-2.838	-1.487	-1.049	-.731	-.573	-2.482	-2.715	-2.715	-2.721
-.2	-1.617	-4.326	-5.634	-1.131	-2.742	-1.609	-1.280	-1.027	-.913	-2.554	-2.804	-2.834	-2.877
-.1	-1.636	-4.702	-5.849	-1.340	-2.778	-1.870	-1.650	-1.472	-1.399	-2.762	-3.037	-3.099	-3.183
.0	-1.661	-4.986	-5.948	-1.425	-3.035	-2.344	-2.220	-2.110	-2.079	-3.202	-3.481	-3.582	-3.699
.1	-1.701	-5.236	-5.913	-1.401	-3.651	-3.104	-3.033	-2.969	-2.951	-4.048	-4.261	-4.363	-4.469
.2	-1.764	-5.530	-5.756	-1.265	-4.819	-4.335	-4.168	-4.057	-3.960	-5.647	-5.449	-5.368	-5.296
.3	-1.861	-5.961	-5.476	-1.017	-6.786	-6.130	-5.500	-5.237	-4.959	-6.519	-4.541	-4.112	-3.779
.4	-1.734	-6.417	-5.077	-.666	-9.806	-8.675	-7.319	-6.823	-6.324	-5.461	-3.656	-3.452	-3.435

Note : "Mean" denotes the mean of the estimate. "Bias" denotes the mean of (estimate - d).

TABLE 4-5 (CONTINUED)

2(b). Normalized MSE and Normalized Variance for T = 100

d	MLEμ	MLE$\bar{y}$	F&T	WL	MDE_1	MDE_2	MDE_3	MDE_4	MDE_5	MDE_2^*	MDE_3^*	MDE_4^*	MDE_5^*
						Normalized MSE							
-.4	.905	.987	1.252	1.021	3.459	2.009	1.618	1.470	1.403	2.589	2.259	2.105	2.001
-.3	.887	1.047	1.283	1.005	2.784	1.700	1.410	1.305	1.250	2.192	1.957	1.845	1.758
-.2	.866	1.101	1.311	1.000	2.208	1.428	1.226	1.157	1.113	1.822	1.667	1.591	1.523
-.1	.842	1.142	1.330	.998	1.729	1.198	1.070	1.028	.998	1.495	1.395	1.351	1.308
.0	.814	1.165	1.334	.996	1.351	1.019	.953	.935	.919	1.239	1.173	1.153	1.133
.1	.779	1.173	1.326	.992	1.080	.914	.900	.904	.904	1.109	1.058	1.045	1.039
.2	.734	1.171	1.308	.989	.940	.919	.939	.962	.980	1.296	1.144	1.096	1.072
.3	.672	1.186	1.285	.989	.985	1.164	1.137	1.150	1.155	2.285	1.565	1.376	1.279
.4	.613	1.311	1.266	1.003	1.345	1.700	1.396	1.296	1.222	1.947	1.396	1.253	1.158
						Normalized Variance							
-.4	.881	.894	1.035	1.021	3.372	1.990	1.612	1.469	1.403	2.529	2.188	2.035	1.932
-.3	.862	.904	1.009	1.000	2.707	1.679	1.400	1.301	1.248	2.133	1.885	1.773	1.685
-.2	.840	.915	.994	.988	2.135	1.404	1.211	1.147	1.105	1.759	1.590	1.512	1.442
-.1	.816	.922	.989	.981	1.654	1.164	1.043	1.007	.979	1.421	1.304	1.257	1.207
.0	.786	.916	.981	.976	1.258	.964	.903	.891	.875	1.136	1.052	1.025	.996
.1	.751	.900	.978	.973	.948	.818	.809	.817	.818	.946	.878	.856	.840
.2	.704	.867	.978	.973	.708	.732	.766	.799	.824	.978	.848	.809	.793
.3	.638	.831	.986	.980	.525	.789	.835	.877	.910	1.862	1.361	1.209	1.137
.4	.584	.900	1.009	.999	.384	.948	.861	.832	.823	1.650	1.264	1.135	1.041

Note : "Normalized MSE" denotes T × the mean of (estimate - d)^2. "Normalized Variance" denotes T × (MSE - Bias × Bias).

TABLE 4-5 (CONTINUED)

3(a). Mean and Bias for T = 250

d	MLEμ	MLE$\bar{y}$	F&T	WL	MDE_1	MDE_2	MDE_3	MDE_4	MDE_5	MDE_2^*	MDE_3^*	MDE_4^*	MDE_5^*
							Mean						
-.4	-.406	-.411	-.415	-.397	-.413	-.404	-.401	-.400	-.399	-.408	-.407	-.407	-.408
-.3	-.306	-.315	-.319	-.301	-.312	-.304	-.302	-.301	-.300	-.307	-.307	-.308	-.308
-.2	-.206	-.217	-.221	-.203	-.211	-.204	-.203	-.202	-.202	-.207	-.207	-.208	-.208
-.1	-.106	-.119	-.123	-.105	-.110	-.105	-.104	-.104	-.104	-.108	-.109	-.109	-.110
.0	-.006	-.020	-.023	-.005	-.011	-.008	-.007	-.007	-.007	-.010	-.011	-.012	-.012
.1	.094	.079	.077	.095	.085	.087	.087	.087	.087	.085	.084	.083	.083
.2	.193	.178	.178	.196	.177	.179	.180	.180	.180	.175	.175	.176	.176
.3	.293	.275	.279	.297	.261	.265	.271	.274	.275	.266	.278	.283	.285
.4	.392	.370	.382	.399	.333	.351	.364	.371	.374	.375	.387	.396	.396
							Bias × 100						
-.4	-.612	-1.118	-1.455	.306	-1.285	-.439	-.092	-.004	.051	-.770	-.699	-.741	-.765
-.3	-.617	-1.504	-1.864	-.085	-1.155	-.422	-.154	-.089	-.041	.730	-.701	-.755	-.784
-.2	-.622	-1.746	-2.120	-.330	-1.059	-.445	-.253	-.211	-.172	-.731	-.744	-.807	-.841
-.1	-.626	-1.904	-2.267	-.470	-1.029	-.544	-.425	-.405	-.379	-.808	-.860	-.932	-.972
.0	-.628	-2.015	-2.326	-.528	-1.132	-.785	-.733	-.737	-.727	-1.035	-1.122	-1.196	-1.246
.1	-.639	-2.121	-2.319	-.521	-1.491	-1.267	-1.216	-1.277	-1.277	-1.518	-1.619	-1.689	-1.727
.2	-.651	-2.244	-2.235	-.443	-2.322	-2.116	-2.049	-2.007	-2.007	-2.497	-2.477	-2.433	-2.392
.3	-.690	-2.457	-2.073	-.289	-3.939	-3.503	-2.943	-2.611	-2.455	-3.443	-2.163	-1.656	-1.454
.4	-.841	-2.973	-1.827	-.055	-6.711	-4.926	-3.585	-2.913	-2.629	-2.526	-1.327	-.365	-.407

Note : "Mean" denotes the mean of the estimate. "Bias" denotes the mean of (estimate - d).

TABLE 4-5 (CONTINUED)

3(b). Normalized MSE and Normalized Variance for T = 250

d	MLEμ	MLE$\bar{y}$	F&T	WL	MDE_1	MDE_2	MDE_3	MDE_4	MDE_5	MDE_2^*	MDE_3^*	MDE_4^*	MDE_5^*
						Normalized MSE							
-.4	.770	.815	.921	.859	2.997	1.752	1.406	1.270	1.191	1.797	1.475	1.327	1.228
-.3	.759	.851	.933	.836	2.386	1.475	1.222	1.122	1.066	1.515	1.282	1.173	1.100
-.2	.748	.879	.949	.831	1.868	1.231	1.057	.990	.953	1.268	1.109	1.033	.983
-.1	.735	.895	.961	.830	1.442	1.023	.914	.876	.857	1.057	.959	.911	.882
.0	.718	.903	.965	.829	1.108	.857	.802	.788	.784	.889	.839	.812	.800
.1	.697	.900	.962	.827	.874	.750	.738	.747	.753	.781	.762	.751	.748
.2	.668	.894	.949	.823	.766	.750	.791	.824	.833	.816	.792	.774	.766
.3	.622	.884	.930	.819	.878	1.279	1.235	1.209	1.190	2.092	1.429	1.225	1.152
.4	.547	.895	.914	.824	1.495	1.966	1.370	1.161	1.059	2.025	1.436	1.360	1.271
						Normalized Variance							
-.4	.761	.785	.869	.857	2.959	1.749	1.407	1.271	1.192	1.784	1.464	1.315	1.214
-.3	.751	.796	.847	.837	2.355	1.472	1.222	1.123	1.067	1.504	1.271	1.160	1.085
-.2	.739	.803	.837	.829	1.842	1.228	1.056	.989	.953	1.256	1.097	1.018	.967
-.1	.726	.805	.833	.826	1.417	1.017	.911	.872	.854	1.041	.942	.890	.859
.0	.708	.801	.830	.822	1.076	.842	.789	.775	.771	.862	.807	.777	.761
.1	.688	.788	.828	.821	.819	.710	.699	.707	.713	.724	.697	.680	.675
.2	.658	.769	.825	.818	.631	.639	.687	.724	.733	.661	.639	.627	.624
.3	.611	.734	.823	.817	.491	.973	1.019	1.039	1.040	1.798	1.314	1.158	1.100
.4	.530	.674	.832	.825	.369	1.361	1.049	.950	.888	1.867	1.393	1.358	1.268

Note : "Normalized MSE" denotes T × the mean of (estimate - d)^2. "Normalized Variance" denotes T × (MSE - Bias × Bias).

TABLE 4-6

IRREGULAR REPLICATIONS IN THE EXACT MLE FOR THE I(d) MODEL

	MLEμ						MLE$\bar{y}$					
	T = 50		T = 100		T = 250		T = 50		T = 100		T = 250	
d	**lt**	**ge**	**lt**	**ge**	**lt**	**ge**	**lt**	**ge**	**lt**	**ge**	**lt**	**ge**
-.4	249	0	183	0	51	0	341	0	233	0	62	0
-.3	81	0	26	0	2	0	154	0	44	0	3	0
-.2	30	0	7	0	0	0	63	0	10	0	0	0
-.1	8	0	0	0	0	0	27	0	1	0	0	0
.0	1	0	0	0	0	0	6	0	0	0	0	0
.1	0	0	0	0	0	0	2	0	0	0	0	0
.2	0	3	0	1	0	0	0	3	0	0	0	0
.3	0	14	0	7	0	0	0	16	0	6	0	0
.4	0	81	0	66	0	14	0	117	0	70	0	13

Note : The numbers in the "lt" columns show the number of replications where the estimates are less than -.5; the numbers in the "ge" columns show the number of replications where the estimates are greater than or equal to .5. Note that the total number of replications is 1,000 for each parameter value (d, T pair).

CHAPTER 5

CONCLUSION

This dissertation considered a stationary long-memory process for economic time series. In a long-memory process, the autocorrelations of the process are so persistent that the sum of the autocorrelations does not converge to a finite non-zero constant. In the literature it is shown that a fractional value for the differencing parameter in a ARIMA(p,d,q) process implies a long memory process for some range of the differencing parameter. To distinguish this process from the standard ARIMA(p,d,q) series, the long-memory ARIMA process is called the autoregressive fractionally integrated moving average process of order p, d, q, or ARFIMA(p,d,q).

The simplest form of the ARFIMA(p, d, q) process is the ARFIMA(0,d,0) or I(d) process. If $d \in (-1/2, 1/2)$ it is stationary and invertible. For $0 < d < 1/2$ the autocorrelations of the I(d) process are positive for all lags, and they decrease so slowly that the sum of the autocorrelations is infinity in the limit, while for $-1/2 < d < 0$ all autocorrelations are negative, and the sum of autocorrelations goes to zero in the limit. Therefore the spectral density at zero frequency is infinity for $d > 0$ and is zero for $d < 0$.

In the dissertation we considered the behavior of a stationarity test and a unit root test when the series is a stationary I(d) process We found that the KPSS stationary test is consistent against stationary I(d) alternatives with $d \in (-1/2, 1/2)$. However, we found using simulations that to distinguish a stationary autocorrelated I(0) process, such as an AR(1) process with coefficient close to unity, from a stationary I(d) process with $d \in (-1/2, 1/2)$, the sample size must be very large. We also found that the power of the KPSS test against a stationary I(d) process is comparable to power of the modified rescaled range test, which is also a test of stationarity against an I(d) alternative.

We considered the Dickey-Fuller unit root tests, and showed that they are also consistent against a stationary I(d) alternative with $d \in (-1/2, 1/2)$. We can use either the coefficient-type test or the t-statistic-type test to distinguish an I(1) process from a stationary I(d) process. Our simulation study showed that the powers of the tests are high even in relatively small samples. However, this might not be true if we used the test statistics which allow for autocorrelation in the error process, such as augmented Dickey-Fuller tests or the Phillips-Perron versions of the Dickey-Fuller tests.

In the simulations for the KPSS test and Dickey-Fuller tests against I(d) alternatives, we chose values of d that allowed for nonstationary cases as well as stationary cases. In the KPSS case we chose values of d from the range [0, 1], and in the Dickey-Fuller case we chose values of d from the range (0, 1.5). Note that if $0.5 \leq d \leq 1.5$ the I(d) process is nonstationary and the consistency of the KPSS test against an I(d) process with d in this range is not proved, while the consistency of the Dickey-Fuller tests is not guaranteed for all tests against an I(d) process with the value of d in this range. The power function of some tests is seen to be continuous over the whole range of d we considered, while for other tests it is discontinuous from the left at $d = 1/2$. We guess that these strange phenomena are caused by the strange behavior in the autocorrelation function of the I(d) process when d is close to 1/2.

Our consistency results for stationarity and unit root tests were shown for I(d) alternatives. We believe that consistency of the stationarity and unit root tests would hold against general stationary ARFIMA alternatives, because the fractional functional limit theorem holds for the general ARFIMA process and this theorem is a main building block for the asymptotic distribution theory for the test statistics. But the finite sample behaviors of the tests against ARFIMA alternatives might be substantially different than against I(d) alternatives. The usefulness of stationarity tests and unit root tests to identify a general long-memory process (including a possibly nonstationary fractionally integrated process) is a quite interesting and challenging topic for further research.

We also compared the finite sample properties of different estimates of the differencing parameter in the I(d) model. Especially we compared the minimum distance estimate of d to various forms of the MLE of d. We found that the minimum distance estimate of d is favorably comparable to the MLE when the mean of the process is not known, even though for $d \geq 1/4$ the minimum distance estimate is slow to converge compared to the MLE. In addition, we confirmed previous findings that the approximate MLE based on the Whittle likelihood function is better than the exact MLE in terms of MSE and bias when the mean is unknown.

Finally, we note that even though the estimates of d we considered are consistent, they are biased in finite samples. The bias is usually negative. It is quite persistent as sample size increases, and is a serious practical problem even for fairly large sample sizes such as $T = 500$. A distribution theory that would explain the size of the bias in the I(d) model or the more general ARFIMA model is another important area for further research.

LIST OF REFERENCES

List of references

Agiakloglou, C., P. Newbold and M. Wohar (1992), "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., C.F. Chung and M. Tieslau (1992), "The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis", unpublished manuscript.

Baillie, R.T. and T. Bollerslev (1993), "The Long Memory of the Forward Premium", unpublished manuscript.

Baillie, R.T., T. Bollerslev and Hans-Ole Mikkelsen (1993), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity", unpublished manuscript.

Baillie, R.T., C.F. Chung and M. Tieslau (1992), "Analyzing Industrialized Countries Inflation by the Fractionally Integrated ARFIMA-GRACH Model", unpublished manuscript.

Beverridge, S. and C. Nelson (1981), "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle", *Journal of Monetary Economics*, 7, 151-174.

Box, G.E.P and G.M. Jenkins (1976), *Forecasting and Control*, Revised Edition, Holden-Day, San Francisco

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*, Second Edition, Springer-Verlag, New York.

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

Campbell, J.Y. and P. Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Root", unpublished manuscript.

Cheung, Y. and F.X. Diebold (1994), "On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally-Integrated Noise With Unknown Mean", *Journal of Econometrics*, forthcoming.

Choi, S and M. Wohar (1991), "The Performance of the GPH Estimator of the Fractional Difference Parameter: Simulation Results", unpublished manuscript.

Chung, C.F. (1993), "Estimating a Generalized Long Memory Process", unpublished manuscript.

Chung, C.F. and R.T. Baillie (1994), "Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models", *Empirical Economics*, forthcoming.

Dahlhaus, R. (1989), "Efficient Parameter Estimation for Self-Similar Processes", *The Annals of Statistics*, 17, 4, 1749-1766.

Davies, R.B. and D.S. Harte (1987), "Test for Hurst Effect", *Biometrika*, 74, 1, 95-101.

Davydov, YU.A. (1970), "The Invariance Principle for Stationary Process", *Theory of Probability and Its Applications*, 15, 3, 487-489.

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, 366, 427-431.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, 4, 1057-1072.

Diebold F.X. and G.D. Rudebusch (1989), "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics*, 24, 189-209.

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 (1991a), "On the Power of Dickey-Fuller Tests against Fractional Alternatives", *Economics Letters*, 35, 155-160.

Diebold, F.X. and G.D. Rudebusch (1991b), "Is Consumption too Smooth? Long Memory and the Deaton Paradox", *The Review of Economics and Statistics*, LXXIII, 1, 1-9.

Dueker, M. and R. Startz (1992), "On Fractional Integration and Cointegration", unpublished manuscript.

Durbin, J. (1960), "The Fitting of Time Series Models", *Review of the International Statistical Institute*, 28, 3, 233-243.

Evans, G.B.A and N.E. Savin (1981), "Testing for Unit Root: 1", *Econometrica*, 49, 3, 753-779.

Evans, G.B.A. and N.E. Savin (1984), "Testing for Unit Root: 2", *Econometrica*, 52, 5, 1241-1269.

Fox, R. and M.S. Taqqu (1986), "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series", *The 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. (1966), "The Typical Spectral Shape of an Economic Variable", *Econometrica*, 34, 1, 150-161.

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

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, 1, 15-29.

Hassler, U. and J. Wolters (1993), "On the Power of Unit Root Tests Against Fractional Alternatives", Discussion Paper No. 2, Fachbereich Wirtschaftswissenschaft, Free University of Berlin.

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

Hosking, J.R.M. (1981), "Fractional Differencing", *Biometrika*, 68, 1, 165-176.

Hosking, J.R.M. (1984), "Asymptotic Distributions of the Sample Mean, Autocovariances and Autocorrelations of Long-Memory Time Series", Mathematics Research Center Technical Summary Report #2752, University of Wisconsin-Madison.

Jonas, A.B. (1983), "Persistent Memory Random Processes", unpublished Ph.D. dissertation, Department of Statistics, Harvard University.

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.

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

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

Lo, A.W. (1991), "Long-Term Memory in Stock Market Prices", *Econometrica*, 59, 5, 1279-1313.

MacNeill, I. (1978), "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times", *Annals of Statistics*, 6, 422-433.

Mandelbrot, B.B. and J.W. Van Ness (1968), "Fractional Brownian Motion, Fractional Noises and Applications", *S.I.A.M. Review*, 10, 4, 422-437.

Möhring, R. (1990), "Parameter Estimation in Gaussian Intermediate-Memory Time Series", unpublished manuscript, Universität Hamburg, Germany.

Nelson, C.R. and C.I. Plosser (1982), "Trends and 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.

Park, J.Y. and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part I", *Econometric Theory*, 4, 468-498.

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, 2, 272-301.

Phillips, P.C.B. and P. Perron (1988) "Testing for a Unit Root in Time Series Regression", *Biometrika*, 75, 2, 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, Cambridge.

Robinson, P.M. (1993), "Time Series with Strong Dependence", invited paper for the 1990 World Congress of the Econometric Society, forthcoming in *Advances in Econometrics: Sixth World Congress*, Cambridge University Press.

Said, S.E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", *Biometrika*, 71, 3, 599-607.

Said, S.E. and D.A. Dickey (1985), "Hypothesis Testing in ARIMA(p,1,q) Models", *Journal of the American Statistical Association*, 80, 390, 369-374.

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

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 & Economic Statistics*, 7, 2, 147-159.

Smith, A.A., F. Sowell and S.E. Zin (1993), "Fractional Integration with Drift: Estimation in Small Samples", unpublished manuscript.

Sowell, F. (1990), "The Fractional Unit Root Distribution", *Econometrica*, 58, 2, 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), "Modeling Long-Run Behavior with the Fractional ARIMA Model", *Journal of Monetary Economics*, 29, 277-302.

Taqqu, M. (1975), "Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process", *Z. Wahrscheinlichkeits-theorie verw. Gebiete*, 31, 287-302.

Tieslau, M.A., P. Schmidt and R.T. Baillie (1994), "A Minimum Distance Estimator for Long-Memory Processes", unpublished manuscript, Michigan State University.

White H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, Inc., New York.

Whittle, P. (1951), *Hypothesis Testing in Time Series Analysis*, Hafner, New York.

Whittle, P. (1963), "On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix", *Biometrika*, 50, 1 and 2, 129-134.

Wooldridge, J.M. (1993), "Estimation and Inference for Dependent Processes", Econometrics and Economic Theory Paper No. 9201, Department of Economics, Michigan State University.

Yajima, Y. (1985), "On Estimation of Long-Memory Time Series Models", *Australian Journal of Statistics*, 27, 3, 303-320.

Yajima, Y. (1988), "On Estimation of a Regression Model with Long-Memory Stationary Errors", *The Annals of Statistics*, 16, 2, 791-807.