The Nonstationary Fractional Unit Root

Katsuto Tanaka *

Hitotsubashi University, Tokyo

Abstract

This paper deals with a scalar I(d) process $\{y_j\}$, where the integration order d is any real number. Under this setting, we first explore asymptotic properties of various statistics associated with $\{y_j\}$, assuming that d is known and is greater than or equal to 1/2. Note that $\{y_j\}$ becomes stationary when $d < 1/2$, whose case is not our concern here. It turns out that the case of $d = 1/2$ needs a separate treatment from $d > 1/2$. We then consider, under the normality assumption, testing and estimation for d, allowing for any value of d. The tests suggested here are asymptotically uniformly most powerful invariant, while the maximum likelihood estimator is asymptotically efficient. The asymptotic theory for these results will not assume normality. Unlike in the usual unit root problem based on autoregressive models, standard asymptotic results hold for test statistics and estimators, where d need not be restricted to $d \geq 1/2$. Simulation experiments are conducted to examine the finite sample performance of both the tests and estimators.

---

* An earlier version of this paper was presented at the first meeting of the New Zealand Econometrics Study Group held at University of Auckland in February, 1997. I am very grateful to Professor P.C.B. Phillips and three anonymous referees for their useful comments and suggestions.

1. Introduction

The unit root problem is usually discussed in connection with autoregressive (AR) models. In formulating a nonstationary AR model, it is implicitly assumed that there exists a positive integer d such that differencing the series d times produces a stationary AR process.

In this paper we consider a scalar I(d) process $\{y_j\}$, where the integration order d is any real number. More specifically we deal with the process $\{y_j\}$ defined by

$$(1-L)^d y_j = \varepsilon_j\,, \qquad (j=1,\cdots,T)\,, \tag{1}$$

where L is the lag-operator, while $\{\varepsilon_j\}$ is an i.i.d.$(0,\sigma^2)$ sequence which we shall extend to the stationary case later. It is known (e.g., Hosking (1981)) that $\{y_j\}$ becomes stationary if $d < 1/2$, whose case with the additional restriction $d > -1/2$ has been studied to a large extent in the literature. Beran (1994) gives a good survey for this case.

The model (1) may be related to the usual integrated process of integer order in the following way. Let $d = \bar{d} + d^*$, where $\bar{d}$ is an integer closest to d. By taking $\bar{d} = d + 1/2$ when the decimal part of d is 1/2, this decomposition is unique and it holds that $-1/2 \le d^* < 1/2$. Thus the above model may be rewritten as

$$(1-L)^{\bar{d}} y_j = v_j\,, \qquad (1-L)^{d^*} v_j = \varepsilon_j\,. \tag{2}$$

It is seen that the fractional I(d) process $\{y_j\}$ with an independent error term is equivalent to the usual I($\bar{d}$) process with a long-range dependent error term $\{v_j\}$. The model (2) with $\bar{d} = 1$ is discussed by Sowell (1990), while the general case is considered by Chan and Terrin (1995), where the AR unit root asymptotics discussed in Chan and Wei (1988) are extended to the fractional case. Jeganathan (1996) developes inference on fractionally cointegrated systems. It is noticed that the integration order d^* of the error process is restricted to $|d^*| < 1/2$ in these studies.

We return to the model (1) which allows for any value of d. The present model and its extended version with stationary error terms are analyzed in Robinson (1994), who considers testing hypotheses for d in the frequency domain, claiming that the analysis is more amenable to the frequency domain approach than to the usual time domain approach. We, however, find the latter still effective, and consider the estimation as well as testing problems for d in the time domain.

In Section 2 we first explore asymptotic properties of various statistics associated with $\{y_j\}$, assuming that d is known and is greater than or equal to 1/2. Note that $\{y_j\}$ becomes stationary when $d < 1/2$, whose case is not our concern here. It turns out that the case of $d = 1/2$ need be treated separately from the case of $d > 1/2$. Section 3 considers testing for d without restricting the parameter space of d. We suggest a locally best invariant (LBI) test equivalent to the Lagrange multiplier (LM) or Rao's score test. It turns out that the resulting statistic tends to normality, unlike in testing for an AR unit root. The above result is parallel to the one obtained in Robinson (1994) by the frequency domain approach. We shall show that the LBI test is asymptotically uniformly most powerful invariant (UMPI) in the sense that the test achieves asymptotically the highest power or power envelope of all the invariant tests under a sequence of local alternatives.

Section 4 discusses estimation of d, where any value of d is allowed for. It is shown that the maximum likelihood estimator (MLE) of d tends to normality and is asymptotically efficient. We then suggest the Wald test based on the MLE, which is also asymptotically UMPI. Section 5 examines, via simulations, the finite sample behavior of the LBI or LM and Wald tests as well as the MLE. Section 6 concludes this paper. Proofs of theorems and corollaries are given in the Appendix.

2. Asymptotic Properties of Various Statistics

In this section we explore asymptotic properties of various statistics associated with $\{y_j\}$ in (1) when $d \geq 1/2$. Since $\{y_j\}$ becomes nonstationary when $d \geq 1/2$, we need to impose restrictions on y_k for $k \leq 0$ to initiate the process $\{y_j\}$. Here we assume, for simplicity, that $y_k = 0$ for $k \leq 0$, although the other initializations are possible (see Diebold and Rudebusch (1991)) and the behavior of $\{y_j\}$ does depend on the initial condition. Then, $\{y_j\}$ is assumed to be generated by

$$(1-L)^d y_j = \sum_{k=0}^{j-1} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}\, y_{j-k} = \varepsilon_j \,. \tag{3}$$

It can be checked that (3) is equivalent to

$$(1-L)^d y_j = \sum_{k=0}^{j-1} A_k y_{j-k}, \qquad A_0 = 1, \qquad A_k = \frac{k-d-1}{k} A_{k-1} \quad (k \geq 1)\,. \tag{4}$$

Alternatively, we can rewrite (3) and (4) as

$$y_j = (1-L)^{-d}\varepsilon_j = \sum_{k=0}^{j-1} \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}\, \varepsilon_{j-k} = \sum_{k=0}^{j-1} B_k \varepsilon_{j-k}\,, \tag{5}$$

where $B_0 = 1$ and $B_k = (k+d-1)B_{k-1}/k$ for $k \geq 1$. This last expression is useful for simulating $\{y_j\}$. Since, by Stirling's formula, $\Gamma(k+d)/\Gamma(k+1) = O(k^{d-1})$ as $k \to \infty$, we have

$$y_T = \begin{cases} O_p\left(\sqrt{\log T}\right) & (d = 1/2)\,, \qquad (6) \\ O_p\left(T^{d-1/2}\right) & (d > 1/2)\,. \qquad (7) \end{cases}$$

It is seen that $\{y_j\}$ becomes nonstationary if $d \geq 1/2$, and that the behavior of $\{y_j\}$ is different between $d = 1/2$ and $d > 1/2$. Because of this, the case of $d = 1/2$ is discussed in Section 2.1, while the case of $d > 1/2$ is treated in Section 2.2. Some of the limiting distributions obtained in these two subsections are graphically presented in Section 2.3.

2.1. Case of $d = 1/2$

We first construct a partial sum process $X_T = \{X_T(t)\}$ defined on [0, 1], which is given by

$$X_T(t) = \frac{1}{s_T} y_j + \frac{t s_T^2 - s_j^2}{s_j^2 - s_{j-1}^2} \frac{1}{s_T} (y_j - y_{j-1})\,, \qquad \left(\frac{s_{j-1}^2}{s_T^2} \leq t \leq \frac{s_j^2}{s_T^2} \right)\,, \tag{8}$$

where $X_T(0) = 0$, $X_T(1) = y_T/s_T$ and

$$s_j^2 = V(y_j) = V\left((1-L)^{-1/2}\varepsilon_j\right) = \frac{\sigma^2}{\pi} \sum_{k=0}^{j-1} \frac{\Gamma^2\left(k + \frac{1}{2}\right)}{\Gamma^2(k+1)}\,. \tag{9}$$

The process $\{X_T(t)\}$ belongs to the function space C that is the space of all real-valued continuous functions defined on [0, 1]. Note that $s_j^2 = O(\log j)$, as is described in (6).

The following weak convergence for $\{X_T(t)\}$ holds because of the functional central limit theorem (FCLT) due to Brown (1971).

Theorem 2.1. *Let $\{y_j\}$ be given by (1) whose generating process is defined in (5), where $d = 1/2$ and $\{\varepsilon_j\}$ is an i.i.d.$(0, \sigma^2)$ sequence. Then, for the partial sum process $X_T = \{X_T(t)\}$ defined in (8), it holds that, as $T \to \infty$,*

$$\mathcal{L}(X_T) \longrightarrow \mathcal{L}(w)\,,$$

where $\mathcal{L}(X)$ denotes the probability law of X, while $w = \{w(t)\}$ is the standard Brownian motion defined on [0, 1].

The following results are a consequence of the above theorem.

Corollary 2.1. *Under the same conditions as in Theorem 2.1, it holds that, as* $T \to \infty$,

$$\frac{1}{\sqrt{\log T}}\, y_T \longrightarrow N\left(0, \frac{\sigma^2}{\pi}\right), \tag{10}$$

$$\mathcal{L}\left(\frac{1}{(\log T)^2}\sum_{j=1}^{T}\frac{1}{j}y_j^2\right) \longrightarrow \mathcal{L}\left(\frac{\sigma^2}{\pi}\int_0^1 w^2(t)dt\right), \tag{11}$$

$$\operatorname{plim}\frac{1}{T}\sum_{j=2}^{T}(y_j - y_{j-1})^2 = \frac{4\sigma^2}{\pi}\,. \tag{12}$$

The result in (10) is an immediate consequence of Theorem 2.1, while (11) follows from

$$\mathcal{L}\left(\sum_{j=1}^{T} X_T^2\left(\frac{s_j^2}{s_T^2}\right)\frac{s_j^2 - s_{j-1}^2}{s_T^2}\right) \longrightarrow \mathcal{L}\left(\int_0^1 w^2(t)dt\right). \tag{13}$$

The result in (12) comes from the variance of the stationary process $\{y_j - y_{j-1}\} = \{(1-L)^{1/2}\varepsilon_j\}$.

Extensions to the case where the error term is dependent are straightforward. Suppose that (1) is replaced by

$$(1-L)^d y_j = u_j\,, \qquad (j = 1, \cdots, T)\,, \tag{14}$$

where $d = 1/2$ and

$$u_j = \sum_{\ell=0}^{\infty}\phi_\ell\,\varepsilon_{j-\ell}\,, \quad \sum_{\ell=0}^{\infty}\ell\,|\phi_\ell| < \infty\,, \quad \phi \equiv \sum_{j=0}^{\infty}\phi_j \neq 0\,, \tag{15}$$

with $\{\varepsilon_j\} \sim \text{i.i.d.}(0, \sigma^2)$. Then we apply the Beveridge-Nelson (B-N) decomposition (Phillips and Solo (1992)) to get

$$u_j = \phi\varepsilon_j - (1-L)\tilde{\varepsilon}_j\,, \tag{16}$$

where $\{\tilde{\varepsilon}_j\}$ is a stationary process defined by

$$\tilde{\varepsilon}_j = \sum_{\ell=0}^{\infty}\tilde{\phi}_\ell\,\varepsilon_{j-\ell}\,, \qquad \tilde{\phi}_\ell = \sum_{k=\ell+1}^{\infty}\phi_k\,. \tag{17}$$

We now have

$$y_j = (1-L)^{-1/2}u_j = \phi(1-L)^{-1/2}\varepsilon_j + O_p(1)\,,$$

$$(1-L)y_j = (1-L)^{1/2}u_j = O_p(1)\,.$$

Then, for the partial sum process $\{X_T(t)\}$ in (8) with $s_j^2 = V\left((1-L)^{-1/2}\varepsilon_j\right)$ defined in (9), we have

$$\mathcal{L}(X_T) \longrightarrow \mathcal{L}(\phi w)\,.$$

Asymptotic results described in Corollary 2.1 are modified accordingly in the following Corollary.

Corollary 2.2. *Let $\{y_j\}$ be given by (14) and (15) with $d = 1/2$. Then it holds that, as $T \to \infty$,*

$$\frac{1}{\sqrt{\log T}} y_T \longrightarrow N\left(0, \frac{\phi^2\sigma^2}{\pi}\right)\,, \tag{18}$$

$$\mathcal{L}\left(\frac{1}{(\log T)^2}\sum_{j=1}^{T}\frac{1}{j}y_j^2\right) \longrightarrow \mathcal{L}\left(\frac{\phi^2\sigma^2}{\pi}\int_0^1 w^2(t)dt\right)\,, \tag{19}$$

$$\operatorname{plim}\frac{1}{T}\sum_{j=2}^{T}(y_j - y_{j-1})^2 = 4\int_0^\pi f_u(\lambda)\sin\frac{\lambda}{2}d\lambda\,, \tag{20}$$

where $f_u(\lambda)$ is the spectrum of $\{u_j\}$.

2.2. Case of $d > 1/2$

Weak convergence results for $d > 1/2$ can be studied in two ways. One is based on $\{X_T(t)\}$ defined in (8), while the other is based on another partial sum process in function space C constructed later.

Let us deal with $\{y_j\}$ defined in (14) and (15), with $d > 1/2$. Again using the B-N decomposition, we obtain $\mathcal{L}(X_T) \to \mathcal{L}(\phi w)$, where we now have

$$s_j^2 = V\left((1-L)^{-d}\varepsilon_j\right) = \frac{\sigma^2}{\Gamma^2(d)}\sum_{k=0}^{j-1}\frac{\Gamma^2(k+d)}{\Gamma^2(k+1)}\,. \tag{21}$$

Note that $s_j^2 = O(j^{2d-1})$, as is seen from (7).

Proceeding in the same way as in the case of $d = 1/2$, we obtain the following results for $d > 1/2$ which specialize to well known results for $d = 1$.

Corollary 2.3. *Let $\{y_j\}$ be given by (14) and (15) with $d > 1/2$. Then it holds that, as $T \to \infty$,*

$$\frac{1}{T^{d-1/2}}\, y_T \longrightarrow N\left(0,\ \frac{\phi^2\sigma^2}{(2d-1)\Gamma^2(d)}\right), \tag{22}$$

$$\mathcal{L}\left(\frac{1}{T^{4d-2}} \sum_{j=1}^{T} j^{2d-2} y_j^2\right) \longrightarrow \mathcal{L}\left(\frac{\phi^2\sigma^2}{(2d-1)^2\,\Gamma^2(d)} \int_0^1 w^2(t)dt\right). \tag{23}$$

The other approach is based on the partial sum process

$$Y_T(t) = \frac{1}{\sigma T^{d-1/2}} y_j + T\left(t - \frac{j}{T}\right)\frac{y_j - y_{j-1}}{\sigma T^{d-1/2}}, \qquad \left(\frac{j-1}{T} \le t \le \frac{j}{T}\right), \tag{24}$$

where y_j is defined in (14) and (15) with $d > 1/2$. When d is a positive integer, it can be shown (see Chan and Wei, 1988 and Tanaka 1996) that

$$\mathcal{L}(Y_T) \longrightarrow \mathcal{L}(\phi F_{d-1}), \tag{25}$$

where $F_g = \{F_g(t)\}$ is the g-fold integrated Brownian motion defined by

$$F_g(t) = \int_0^t F_{g-1}(s)ds, \qquad F_0(t) = w(t), \qquad (g = 1, 2, \cdots). \tag{26}$$

When d is any real number greater than $1/2$, (25) still holds with

$$F_g(t) = \frac{1}{\Gamma(g+1)} \int_0^t (t-s)^g dw(s), \qquad g > -\frac{1}{2}. \tag{27}$$

We have $E(F_g(t)) = 0$ and $V(F_g(t)) = t^{2g+1}/((2g+1)\Gamma^2(g+1))$ for each fixed t. Note that (26) and (27) are equivalent when g is a positive integer.

The following properties can be derived on the basis of the FCLT described in (25).

Corollary 2.4. *Under the same conditions as in Corollary 2.3, it holds that, as $T \to \infty$,*

$$\mathcal{L}\left(\frac{1}{T^{d-1/2}}\, y_T\right) \longrightarrow \mathcal{L}(\phi\sigma F_{d-1}(1)) \sim N\left(0,\ \frac{\phi^2\sigma^2}{(2d-1)\Gamma^2(d)}\right), \tag{28}$$

$$\mathcal{L}\left(\frac{1}{T^{2d}} \sum_{j=1}^{T} y_j^2\right) \longrightarrow \mathcal{L}\left(\phi^2\sigma^2 \int_0^1 F_{d-1}^2(t)dt\right), \tag{29}$$

$$\mathcal{L}\left(T^{2d-1}(\hat{\rho}-1)\right) \longrightarrow \mathcal{L}\left(\frac{-M(d)/2}{\phi^2\sigma^2\int_0^1 F_{d-1}^2(t)dt}\right) \qquad (1/2 < d < 1)\,, \tag{30}$$

$$\mathcal{L}\left(T(\hat{\rho}-1)\right) \longrightarrow \begin{cases} \mathcal{L}\left(\dfrac{\left(w^2(1)-r\right)/2}{\displaystyle\int_0^1 w^2(t)dt}\right) & (d=1)\,, \quad (31) \\[2ex] \mathcal{L}\left(\dfrac{F_{d-1}^2(1)/2}{\displaystyle\int_0^1 F_{d-1}^2(t)dt}\right) & (d>1)\,, \quad (32) \end{cases}$$

where $r = \sum_{j=0}^{\infty}\phi_j^2/\phi^2$, $\hat{\rho} = \sum_{j=2}^{T} y_{j-1}y_j / \sum_{j=2}^{T} y_{j-1}^2$ *and*

$$M(d) = \int_{-\pi}^{\pi}\left|1-e^{i\lambda}\right|^{2-2d} f_u(\lambda)d\lambda\,, \tag{33}$$

with $f_u(\lambda)$ *being the spectrum of* $\{u_j\}$.

The statistic $\hat{\rho}$ in (30), (31) and (32) can be interpreted as the least squares estimator (LSE) of ρ applied to the following model:

$$y_j = \rho y_{j-1} + v_j\,, \qquad (1-L)^{d-1}v_j = u_j\,, \qquad (j=1,\cdots,T)\,, \tag{34}$$

where the true value of ρ is 1 and d is greater than 1/2, while $\{u_j\}$ is a stationary process defined in (15).

Noting that

$$\hat{\rho}-1 = (A_T - B_T)\Big/\sum_{j=2}^{T} y_{j-1}^2\,,$$

where

$$A_T = y_T^2/2 = O_p(T^{2d-1}) \qquad (d>1/2)\,,$$

$$B_T = \sum_{j=2}^{T}(y_j - y_{j-1})^2/2 = \begin{cases} O(T) & (1/2 < d < 3/2)\,, \\ O(T\log T) & (d = 3/2)\,, \\ O_p(T^{2d-2}) & (d > 3/2)\,, \end{cases}$$

it is seen that A_T dominates B_T when $d > 1$, while B_T dominates A_T when $1/2 < d < 1$. When $d = 1$, A_T and B_T have the same stochastic order T. This is the reason why

$\hat{\rho}$ behaves differently depending on the value of d. This fact is also discussed in Sowell (1990) when d takes values between 1/2 and 3/2, where $\{u_j\}$ in (34) is assumed to be i.i.d.(0,σ^2). In that case, $M(d)$ in (33) becomes the variance of the stationary process $\{(1-L)^{1-d}\varepsilon_j\}$ and we have

$$\begin{aligned} M(d) &= V\left((1-L)^{1-d}\varepsilon_j\right) = \sigma^2 \sum_{k=0}^{\infty} \frac{\Gamma^2(k+d-1)}{\Gamma^2(d-1)\Gamma^2(k+1)} \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \left|1-e^{i\lambda}\right|^{2-2d} d\lambda = \sigma^2 \frac{\Gamma(3-2d)}{\Gamma^2(2-d)}\,, \qquad (1/2 < d < 3/2)\,. \end{aligned}$$

2.3. Graphs of Some Limiting Distributions

In this subsection we present graphically some of the limiting distributions obtained in the previous subsections. For this purpose we derive the characteristic function (c.f.) of the limiting distribution, and then invert it numerically.

Let us first deal with the quantity appearing in (29):

$$G_d = \int_0^1 F_{d-1}^2(t)dt = \int_0^1\int_0^1 K_{d-1}(s,t)dw(s)dw(t)\,, \tag{35}$$

where $d > 1/2$ and

$$K_{d-1}(s,t) = \frac{1}{\Gamma^2(d)} \int_{\max(s,t)}^1 \left((u-s)(u-t)\right)^{d-1} du\,.$$

Note that G_d is a limit in distribution of the sample second moment arising from the I(d) process. It is clear that

$$E(G_d) = \int_0^1 K_{d-1}(t,t)dt = \frac{1}{2d(2d-1)\Gamma^2(d)}\,,$$

which tends to 0 as $d \to \infty$. We also have

$$V(G_d) = 2\int_0^1\int_0^1 K_{d-1}^2(s,t)dsdt\,,$$

which is 1/3 for $d = 1$, 1.31×10^{-2} for $d = 2$, 1.36×10^{-4} for $d = 3$, 4.87×10^{-7} for $d = 4$, and so on.

Figure 1 draws the probability density of $G_1 = \int_0^1 w^2(t)dt$, while Figures 2 and 3 those of G_2 and G_3, respectively. These were computed by numerical integration via inversion of the c.f.'s of G_d (see, for details, Tanaka 1996, Chap.6). In Figures 2 and 3, we also present approximate distributions of G_d, where the approximations are based on the distributional relation:

$$\mathcal{L}(G_d) = \mathcal{L}\left(\sum_{n=1}^{\infty} \frac{1}{\lambda_n(d)}\, Z_n^2\right)\,,$$

where $\{Z_n\}$ is i.i.d. normal with common mean 0 and variance 1, which is abbreviated as $\{Z_n\} \sim$ NID(0,1), while $0 < \lambda_1(d) < \lambda_2(d) < \cdots$ are the eigenvalues of the integral equation:

$$f(t) = \lambda \int_0^1 K_{d-1}(s,t) f(s) ds \,.$$

Then G_d is approximated as $Z_1^2/\lambda_1(d)$, where $\lambda_1(2) = 12.36236$ and $\lambda_1(3) = 121.259$ (Tanaka (1996, Chap.6)). It is seen that G_2 and G_3 are well approximated by a constant multiple of $\chi^2(1)$. Note that G_1 cannot be well approximated in this way because of its nonmonotonic distributuional nature. The reason may be partly explained by considering the ratio of $E(Z_1^2/\lambda_1(d))$ to $E(G_d)$:

$$R(d) = 1/\lambda_1(d) \Bigg/ \sum_{n=1}^{\infty} 1/\lambda_n(d) \ ,$$

which is 0.8106 for $d = 1$, 0.9707 for $d = 2$, and 0.9896 for $d = 3$.

Figure 1 Figure 2 Figure 3

We also consider the quantity appearing in (32):

$$H_d = \frac{F_{d-1}^2(1)/2}{\displaystyle\int_0^1 F_{d-1}^2(t) dt} \ , \tag{36}$$

which is well defined for $d > 1/2$ and is a limit in distribution of

$$H_{Td} = \frac{T}{2} y_T^2 \Bigg/ \sum_{j=2}^{T} y_{j-1}^2 \ ,$$

where $(1-L)^d y_j = u_j$ with the stationary process $\{u_j\}$ defined in (15).

Figure 4 draws probability densities of H_d for $d = 1$, 2, and 3. Note that $H_1 = \frac{1}{2} w^2(1) \Big/ \int_0^1 w^2(t)\, dt$. These densities were also computed by numerical integration. We can show by numerical integration that $E(H_d) = d$ for $d = 1$, 2, and 3. It is conjectured that this holds for any d $(> 1/2)$. We also obtain $V(H_d) = 0.891$ for $d = 1$, 1.313 for $d = 2$, and 1.716 for $d = 3$.

Figure 4

3. Testing for d

This section discusses testing for the integration order d without restricting the parameter space of d. For this purpose we consider the model:

$$z_j = x_j'\beta + y_j\,, \quad (1-L)^{d+\theta} y_j = u_j\,, \qquad (j = 1, \cdots, T)\,, \tag{37}$$

where $\{x_j\}$ is a sequence of $K \times 1$ fixed, nonstochastic variables and β is a $K \times 1$ unknown vector, while d is any preassigned value and $\{u_j\}$ is a stationary process defined in (17). Note that $\{y_j\}$ becomes stationary when $d + \theta < 1/2$, whose case is also treated here.

Then the testing problem considered here is, as in Robinson (1994) (see also Diebold and Rudebusch (1991)),

$$H_0 : \theta = 0 \qquad \text{against} \qquad H_1 : \theta > 0 \ \text{ or } \ H_2 : \theta < 0\,. \tag{38}$$

Note that $\{z_j\}$ becomes more nonstationary under $H_1 : \theta > 0$, while the reverse is true under $H_2 : \theta < 0$. This is a very general testing procedure, allowing for a test of the stationary hypothesis ($d < 1/2$), the nonstationary unit root hypothesis ($d = 1$) and the $I(0)$ hypothesis (d=0).

In Section 3.1 we deal with the simplest case where $\{u_j\}$ in (37) is an i.i.d.$(0, \sigma^2)$ sequence, while the stationary case is discussed in Section 3.2. Our test can be easily adapted to deal with the model:

$$z_j = x_j'\beta + y_j\,, \quad y_j = (1-L)^{d+\theta} u_j\,, \qquad (j = 1, \cdots, T)\,. \tag{39}$$

This case will also be discussed as a by-product of the original test.

3.1. i.i.d. Case

Let us put $u_j = \varepsilon_j$ in (37) and assume that $\{\varepsilon_j\} \sim \text{NID}(0, \sigma^2)$, though the asymptotic theory developed later will not assume normality. Since $\{y_j\}$ is generated by (5) with d replaced by $d + \theta$, it follows that

$$z = X\beta + y \ \sim \ N\left(X\beta,\, \sigma^2\Omega(\theta)\right)\,, \tag{40}$$

where $z = (z_1, \cdots, z_T)'$, $X = (x_1, \cdots, x_T)'$, $y = (y_1, \cdots, y_T)'$ and $\Omega(\theta) = V(y)/\sigma^2$ with $rank(X) = K(< T)$. Note that d is a given constant and the testing problem (38) is invariant under the group of transformations:

$$z \longrightarrow az + Xb \quad and \quad (\theta, \beta, \sigma^2) \longrightarrow (\theta, a\beta + b, a^2\sigma^2)\,,$$

where $0 < a < \infty$ and b is a $K \times 1$ vector.

For the model (40), the log-likelihood $L(\theta, \beta, \sigma^2)$ is given by

$$\begin{aligned} L(\theta, \beta, \sigma^2) &= -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(z - X\beta)'\Omega^{-1}(\theta)(z - X\beta) \\ &= -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{T}\left\{(1-L)^{d+\theta}(z_j - x_j'\beta)\right\}^2 . \end{aligned} \tag{41}$$

Note that the second line in (41) is an approximated version of the exact log-likelihood given in the first line if $d + \theta < 1/2$ because of introducing an explicit conditioning on initial values. We, however, use the approximated version as the true log-likelihood, which does not affect the asymptotic theory developed subsequently. Then the LBI test is shown to be equivalent to the LM test (Tanaka (1996, Chap.9)). Thus the LBI test for $H_0 : \theta = 0$ against $H_1 : \theta > 0$ rejects H_0 when

$$\begin{aligned} S_{T1} &= \left.\frac{\partial L(\theta, \beta, \sigma^2)}{\partial \theta}\right|_{H_0:\theta=0,\beta=\hat{\beta},\sigma^2=\hat{\sigma}^2} \\ &= -\frac{1}{\hat{\sigma}^2}\sum_{j=1}^{T}\left\{\log(1-L) \times (1-L)^d(z_j - x_j'\hat{\beta})\right\}(1-L)^d(z_j - x_j'\hat{\beta}) \end{aligned} \tag{42}$$

becomes large, where $\hat{\beta} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{z}$ with $\tilde{X} = (\tilde{x}_1, \cdots, \tilde{x}_T)'$, $\tilde{x}_j = (1-L)^d x_j$, $\tilde{z} = (\tilde{z}_1, \cdots, \tilde{z}_T)'$, and $\tilde{z}_j = (1-L)^d z_j$, while $\hat{\sigma}^2 = (\tilde{z} - \tilde{X}\hat{\beta})'(\tilde{z} - \tilde{X}\hat{\beta})/T$. The LBI test for $H_0 : \theta = 0$ against $H_2 : \theta < 0$ rejects H_0 when S_{T1} becomes small.

If we define $\hat{\varepsilon}_j = (1-L)^d(z_j - x_j'\hat{\beta}) = \tilde{z}_j - \tilde{x}_j'\hat{\beta}$ and consider the expansion:

$$-\log(1-L) = L + \frac{1}{2}L^2 + \frac{1}{3}L^3 + \cdots ,$$

S_{T1} can be rewritten more compactly as

$$S_{T1} = \frac{1}{\hat{\sigma}^2}\sum_{j=2}^{T}\left(\sum_{k=1}^{j-1}\frac{1}{k}\hat{\varepsilon}_{j-k}\right)\hat{\varepsilon}_j = T\sum_{k=1}^{T-1}\frac{1}{k}\hat{\rho}_k , \tag{43}$$

where $\hat{\rho}_k = \sum_{j=k+1}^{T}\hat{\varepsilon}_{j-k}\,\hat{\varepsilon}_j / \sum_{j=1}^{T}\hat{\varepsilon}_j^2$ is the k-th order autocorrelation of residuals $\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_T$. It is noticed that the statistic S_{T1} has some similarity to the portmanteau Q statistic (Box and Pierce (1970)) for diagnostic checking of time series models, although Q takes the form of an unweighted sum of squares of $\hat{\rho}_k$'s.

The finite sample distribution of S_{T1} is intractable even under H_0; so we consider the limiting distribution of S_{T1} under a sequence of local alternatives.

Theorem 3.1. *Under $\theta = \delta/\sqrt{T}$ with δ fixed, it holds that, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} S_{T1} = \sqrt{T} \sum_{k=1}^{T-1} \frac{1}{k} \hat{\rho}_k \longrightarrow N\left(\frac{\pi^2}{6} \delta\, ,\ \frac{\pi^2}{6} \right) .$$

It can be checked that

$$\lim_{T\to\infty} E\left(-\frac{1}{T} \frac{\partial^2 L(\theta, \beta, \sigma^2)}{\partial \theta^2} \bigg|_{\theta=\delta/\sqrt{T}, \beta=\hat{\beta}, \sigma^2=\hat{\sigma}^2} \right) = \frac{\pi^2}{6} ,$$

which is the limiting value of the normalized Fisher information. Thus it is seen that standard results apply to the present problem, unlike in the AR unit root test case. In practice we compute

$$S'_{T1} = \frac{1}{\sqrt{T}} S_{T1} \Big/ \sqrt{\frac{\pi^2}{6}} = \sqrt{T} \sum_{k=1}^{T-1} \frac{1}{k} \hat{\rho}_k \Big/ \sqrt{\frac{\pi^2}{6}} , \tag{44}$$

and compare this with the upper or lower $100\alpha\%$ point of N(0,1), which gives the right-sided ($H_1 : \theta > 0$) or left-sided ($H_2 : \theta < 0$) LBI test of approximate size α.

The limiting power of the LBI test under $\theta = \delta/\sqrt{T}$ can be easily computed.

Corollary 3.1. *Let z_α be the upper $100\alpha\%$ point of N(0,1). Then, it holds that, as $T \to \infty$ under $\theta = \delta/\sqrt{T}$,*

$$P\left(S'_{T1} > z_\alpha\right) \longrightarrow \Phi\left(-z_\alpha + \delta \sqrt{\frac{\pi^2}{6}} \right) \qquad \text{for} \quad \delta > 0 ,$$

$$P\left(S'_{T1} < -z_\alpha\right) \longrightarrow \Phi\left(-z_\alpha - \delta \sqrt{\frac{\pi^2}{6}} \right) \qquad \text{for} \quad \delta < 0 ,$$

where $\Phi(z)$ is the distribution function of N(0,1).

We next show, following Elliott, Rothenberg and Stock (1996) and Tanaka (1996), that the LBI test obtained above is asymptotically UMPI in the sense that its power attains the highest power of all the invariant tests as $T \to \infty$ under $\theta = \delta/\sqrt{T}$. For this purpose we assume that the data generating process (DGP) is

$$z_j = x'_j \beta + y_j\, , \quad (1-L)^{d+\theta_0} y_j = \varepsilon_j\, , \qquad (j = 1, \cdots, T)\, , \tag{45}$$

where $\theta_0 = c/\sqrt{T}$ with $c\,(> 0)$ fixed.

We now consider testing for

$$H_0 :\ \theta = 0 \qquad \text{against} \qquad H_1 :\ \theta = \delta/\sqrt{T}\, , \tag{46}$$

where $\delta\,(>0)$ is a known constant. This is a test of a simple null against a simple alternative with nuisance parameters β and σ^2 which can be eliminated by invariance arguments. Then we have that the test which rejects H_0 when

$$M_T = T \times \frac{\displaystyle\sum_{j=1}^{T} \hat{\varepsilon}_j^2 - \sum_{j=1}^{T} \tilde{\varepsilon}_j^2}{\displaystyle\sum_{j=1}^{T} \hat{\varepsilon}_j^2} \tag{47}$$

becomes large is MPI, where $\hat{\varepsilon}_j$ and $\tilde{\varepsilon}_j$ are residuals under H_0 and H_1, respectively.

The limiting distribution of M_T in (47) as $T \to \infty$ under $\theta_0 = c/\sqrt{T}$ and $\theta = \delta/\sqrt{T}$ is given by the following theorem.

Theorem 3.2. *The MPI statistic M_T in (47) has the following limiting distribution as $T \to \infty$ under $\theta_0 = c/\sqrt{T}$ and $\theta = \delta/\sqrt{T}$:*

$$\mathcal{L}(M_T) \longrightarrow \mathcal{L}(M(c,\delta)) = \mathcal{L}\left(\delta\left(2\sqrt{\frac{\pi^2}{6}}Z + (2c-\delta)\frac{\pi^2}{6}\right)\right), \tag{48}$$

where $Z \sim$ N(0,1).

Let $x_\alpha(\delta)$ be the upper $100\alpha\%$ point of $M(0,\delta)$. Then, the highest power of all the invariant tests of size α is given asymptotically by $P(M(\delta,\delta) > x_\alpha(\delta))$. Then it turns out that the power envelope coincides with the power of the LBI test.

Corollary 3.2. *The power envelope of all the invariant tests of size α for $\theta = 0$ against $\theta = \delta/\sqrt{T}$ is given asymptotically by $\Phi(-z_\alpha + \delta\sqrt{\pi^2/6}\,)$ for $\delta > 0$, and $\Phi(-z_\alpha - \delta\sqrt{\pi^2/6}\,)$ for $\delta < 0$. Thus each LBI test is asymptotically UMPI.*

The fact that the LBI test for d is asymptotically UMPI contrasts with the LBI tests for an AR unit root (Elliott, Rothenberg and Stock (1996)) and an MA unit root (Tanaka (1996, Chap.10)). It holds that

$$y_T^2 \Bigg/ \sum_{j=1}^{T} (y_j - y_{j-1})^2$$

is the LBI statistic for testing $H_0 : \rho = 1$ against $H_1 : \rho < 1$ in the AR(1) model: $y_j = \rho y_{j-1} + \varepsilon_j$, where $y_0 = 0$ and $\{\varepsilon_j\} \sim \mathrm{NID}(0, \sigma^2)$. It also holds that

$$\frac{\sum_{j=1}^{T}\left(\sum_{k=1}^{j-1}(j-k)\,y_k \;-\; \frac{j}{T+1}\sum_{k=1}^{T}(T-k+1)\,y_k\right)^2}{\sum_{j=1}^{T}\frac{1}{j(j+1)}\left(\sum_{k=1}^{j} k\,y_k\right)^2}$$

is the LBI statistic for testing $H_0 : \alpha = 1$ against $\alpha < 1$ in the MA(1) model: $y_j = \varepsilon_j - \alpha\varepsilon_{j-1}$, where $\varepsilon_0, \varepsilon_1, \cdots \sim$ NID(0, σ^2). The fact that the LBI test for d attains the power envelope, while the LBI test for the unit root in the AR(1) or MA(1) model does not, reflects the standard nature of the present problem.

The present test can be easily adapted to test for the integration order in the MA part. For this purpose we deal with the model:

$$z_j = x_j'\beta + y_j\,, \quad y_j = (1-L)^{d+\theta}\varepsilon_j\,, \qquad (j = 1, \cdots, T)\,, \tag{49}$$

where d is any preassigned value and $\{\varepsilon_j\} \sim \text{NID}(0, \sigma^2)$, for which the same testing problem as in (38) is considered.

The log-likelihood is now given by

$$L(\theta, \beta, \sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{T}\left\{(1-L)^{-d-\theta}(z_j - x_j'\beta)\right\}^2\,,$$

so that

$$\left.\frac{\partial L(\theta, \beta, \sigma^2)}{\partial\theta}\right|_{H_0} = -T\sum_{k=1}^{T-1}\frac{1}{k}\hat{\rho}_k\,, \tag{50}$$

where $\hat{\rho}_k$ is the k-th order autocorrelation of residuals $\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_T$ with $\hat{\varepsilon}_j = (1 - L)^{-d}(z_j - x_j'\hat{\beta})$. Note the similar form in (50) to (43) except for the negative sign, which leads us to the following result.

Corollary 3.3. *For the model (49), the test which rejects $H_0 : \theta = 0$ against $H_1 : \theta > 0$ ($H_2 : \theta < 0$), when $\tilde{S}_{T1}' < -z_\alpha$ ($\tilde{S}_{T1}' > z_\alpha$), is an LBI test of approximate size α, where $\tilde{S}_{T1}'$ is the same as S_{T1}' in (44) with $\hat{\rho}_k$ defined in (50). The limiting power of the $\tilde{S}_{T1}'$-test, as $T \to \infty$ under $\theta = \delta/\sqrt{T}$, is given by $\Phi(-z_\alpha + \delta\sqrt{\pi^2/6})$ for $\delta > 0$, and $\Phi(-z_\alpha - \delta\sqrt{\pi^2/6})$ for $\delta < 0$. Moreover, each $\tilde{S}_{T1}'$-test is asymptotically UMPI.*

3.2. Stationary Case

Here we consider the model (37), where $\{u_j\}$ is assumed to follow an autoregressive moving average (ARMA(p, q)) process, namely, $a(L)u_j = b(L)\varepsilon_j$, where $\{\varepsilon_j\} \sim \text{NID}(0, \sigma^2)$ and

$$a(L) = 1 - a_1L - \cdots - a_pL^p\,, \quad b(L) = 1 - b_1L - \cdots - b_qL^q\,,$$

with $a(x) \neq 0$ and $b(x) \neq 0$ for $|x| \leq 1$.

The log-likelihood is now given by

$$L(\theta,\beta,\psi,\sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{T}\left\{a(L)b^{-1}(L)(1-L)^{d+\theta}(z_j - x_j'\beta)\right\}^2 ,$$

where $\psi = (a_1, \cdots, a_p, b_1, \cdots, b_q)'$, and the LM principle yields the same statistic as S_{T1} in (43), but with $\hat{\varepsilon}_j$ defined in the present case by

$$\hat{\varepsilon}_j = \hat{a}(L)\hat{b}^{-1}(L)(1-L)^d(z_j - x_j'\hat{\beta}) , \tag{51}$$

where $\hat{\beta}$ is the MLE of β under H_0, while $\hat{a}(L)$ and $\hat{b}(L)$ are estimated from $a(L)u_j = b(L)\varepsilon_j$ with $u_j = (1-L)^d(z_j - x_j'\beta)$ replaced by $\hat{u}_j = (1-L)^d(z_j - x_j'\hat{\beta})$.

Then we obtain the following result.

Theorem 3.3. *Consider the statistic*

$$S_{T2} = \left.\frac{\partial L(\theta,\beta,\psi,\sigma^2)}{\partial\theta}\right|_{H_0} = T\sum_{k=1}^{T-1}\frac{1}{k}\hat{\rho}_k , \tag{52}$$

where $\hat{\rho}_k$ is the k-th order autocorrelation of residuals $\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_T$ defined in (51). Then it holds that, as $T \to \infty$ under $\theta = \delta/\sqrt{T}$ with δ fixed,

$$\frac{1}{\sqrt{T}}S_{T2} = \sqrt{T}\sum_{k=1}^{T-1}\frac{1}{k}\hat{\rho}_k \longrightarrow N\left(\delta\omega^2, \omega^2\right) ,$$

where

$$\omega^2 = \frac{\pi^2}{6} - (\kappa_1, \cdots, \kappa_p, \lambda_1, \cdots, \lambda_q)\Phi^{-1}(\kappa_1, \cdots, \kappa_p, \lambda_1, \cdots, \lambda_q)' , \tag{53}$$

$$\kappa_i = \sum_{j=i}^{\infty}\frac{1}{j}c_{j-i} , \qquad \lambda_i = -\sum_{j=i}^{\infty}\frac{1}{j}d_{j-i} , \tag{54}$$

with c_j and d_j the coefficients of L^j in the expansion of $1/a(L)$ and $1/b(L)$, respectively, and Φ the Fisher information matrix for a and b.

Note that this theorem reduces to Theorem 3.1 when $u_j = \varepsilon_j$ and no estimation of ψ is attempted. Since S_{T2} depends on ω^2 that is a function of ψ, we suggest as a test statistic

$$S'_{T2} = \frac{1}{\sqrt{T}}S_{T2}/\hat{\omega} ,$$

where $\hat{\omega}$ is the MLE of ω under H_0, which can be computed from (53) by inserting $\hat{\psi}$ into the κ_i's and Γ_{p+q}. Then it holds that $S'_{T2} \to N(\delta\omega, 1)$ as $T \to \infty$ under $\theta = \delta/\sqrt{T}$.

The computation of $\hat{\omega}$ in the above way will become much involved as $p+q$ gets large. A computationally simpler method will be suggested in Section 5. For the simplest case where $\{u_j\}$ follows an AR(1) process: $u_j = au_{j-1} + \varepsilon_j$ or an MA(1) process: $u_j = \varepsilon_j - a\varepsilon_{j-1}$, however, $\hat{\omega}$ is easily computed. We have $v_j = av_{j-1} + \varepsilon_j$ in both cases so that $c_j = a^j$ and

$$\kappa_1 = \sum_{j=1}^{\infty} \frac{1}{j} a^{j-1} = -\frac{1}{a} \log(1-a)\,, \quad \sigma^2 \Gamma_1^{-1} = 1 - a^2\,.$$

Thus we have

$$\omega^2 = \frac{\pi^2}{6} - \frac{1-a^2}{a^2} (\log(1-a))^2\,, \tag{55}$$

and $\hat{\omega}$ can be computed using $\hat{a}$.

Because of estimating ψ, the limiting power of the S'_{T2}-test is lower than that of S'_{T1}-test. In particular, if we suppose that $\{u_j\}$ follows AR(1) or MA(1) while the DGP of $\{u_j\}$ is $\{\varepsilon_j\}$, the asymptotic efficiency of the S'_{T2}-test is $\omega^2/(\pi^2/6) = 0.392$.

The S'_{T2}-test possesses asymptotic properties similar to the S'_{T1}-test. Namely, the power function of the former is asymptotically the one given in Corollary 3.1 with $\sqrt{\pi^2/6}$ replaced by ω. Moreover, the power function coincides asymptotically with the power envelope of all the invariant tests for the model (37).

The present test can also be implemented to test for the integration order in the MA part, as in the i.i.d. case. The model now takes the form in (39), and we can easily conduct a test similar to the S'_{T2}-test.

The finite sample performance of the S'_{T1}-test discussed in the previous subsection and the S'_{T2}-test discussed here will be examined in Section 5.

4. Estimation of *d*

Here we discuss the asymptotic theory for the ML estimation of d when d takes any value. The discussion is divided into two subsections by considering first the i.i.d. case followed by the stationary case, as in the testing problem. For simplicity of presentation, however, we concentrate on models without any regressor, and discussions are given on an intuitive basis. Formal proofs for the asymptotic properties of the MLE are much more involved (see Hosoya(1997)).

4.1. i.i.d. Case

Let us consider the model

$$(1-L)^d y_j = \varepsilon_j\,, \qquad (j=1,\cdots,T)\,, \tag{56}$$

where d is any value and $\{\varepsilon_j\} \sim \mathrm{NID}(0,\sigma^2)$. The parameters estimated here are d and σ^2, and the concentrated log-likelihood for d is given, except for constants, by

$$\ell(d) = -\frac{T}{2}\log\left[\sum_{j=1}^{T}\left\{(1-L)^d y_j\right\}^2\right].$$

Let d_0 be the true value of d. Then the maximization of $\ell(d)$ is equivalent to that of

$$\begin{aligned} g(d) &= \ell(d)-\ell(d_0) = -\frac{T}{2}\log\left[\sum_{j=1}^{T}\left\{(1-L)^d y_j\right\}^2 \Big/ \sum_{j=1}^{T}\left\{(1-L)^{d_0} y_j\right\}^2\right] \\ &= -\frac{T}{2}\log\left[1-\frac{1}{T}\frac{\sum_{j=1}^{T}\left\{(1-L)^{d_0} y_j\right\}^2-\sum_{j=1}^{T}\left\{(1-L)^d y_j\right\}^2}{\frac{1}{T}\sum_{j=1}^{T}\{(1-L)^{d_0} y_j\}^2}\right]. \end{aligned} \tag{57}$$

Here it holds that

$$\operatorname{plim}\frac{1}{T}\sum_{j=1}^{T}\left\{(1-L)^{d_0} y_j\right\}^2 = \operatorname{plim}\frac{1}{T}\sum_{j=1}^{T}\varepsilon_j^2 = \sigma^2\,.$$

Moreover, if there exists a constant δ such that $d = d_0+\delta/\sqrt{T}$, then we have the following theorem. The existence of such a δ as $T\to\infty$ will be proved shortly.

Theorem 4.1. *For the model (56), it holds that, as $T\to\infty$ under $d=d_0+\delta/\sqrt{T}$,*

$$\mathcal{L}(g(d)) \longrightarrow \mathcal{L}(W(\delta)) = \mathcal{L}\left(\frac{\delta}{2}\left(2\sqrt{\frac{\pi^2}{6}}\,Z-\frac{\pi^2}{6}\delta\right)\right),$$

$$\mathcal{L}\left(\frac{\partial g(d)}{\partial\delta}\right) \longrightarrow \mathcal{L}(W'(\delta)) = \mathcal{L}\left(\sqrt{\frac{\pi^2}{6}}\,Z-\frac{\pi^2}{6}\delta\right),$$

$$\operatorname{plim}\left(\frac{\partial^2 g(d)}{\partial\delta^2}\right) = -\frac{\pi^2}{6}\,,$$

where $Z\sim N(0,1)$.

Note that $\mathcal{L}(W(\delta)) = \mathcal{L}(M(\delta,\delta)/2)$, where $M(\delta,\delta)$ appeared in (48) when we derived the MPI test.

We now prove that there exists an MLE $\hat{d}$ of d_0 such that $\sqrt{T}(\hat{d}-d_0) = \hat{\delta} = O_p(1)$. We first show the existence of a local MLE $\tilde{d}$ such that $\sqrt{T}(\tilde{d}-d_0) = O_p(1)$. For this purpose it is sufficient to show that, for any $\varepsilon > 0$ and for all $T \geq T_0$ with T_0 fixed, there exists a positive constant δ_0 such that

$$P\left(|\tilde{d} - d_0| \leq \frac{\delta_0}{\sqrt{T}}\right) \geq 1 - \varepsilon\,. \tag{58}$$

This can be proved in the same lines as in Sargan and Bhargava (1983), and it is sufficient to show that

$$P\left(g'\left(d_0 + \delta_1/\sqrt{T}\right) \geq 0\right) = P\left(\frac{\partial g(d_0 + \delta_1/\sqrt{T})}{\partial \delta_1} \geq 0\right) \leq \varepsilon\,,$$

$$P\left(g'\left(d_0 + \delta_2/\sqrt{T}\right) \leq 0\right) = P\left(\frac{\partial g(d_0 + \delta_2/\sqrt{T})}{\partial \delta_2} \leq 0\right) \leq \varepsilon\,,$$

for all $T \geq T_0$ by taking suitably δ_1 (> 0), δ_2 (< 0) and a corresponding T_0.

In fact, for $d = d_0 + \delta_1/\sqrt{T} > d_0$ with $\delta = \delta_1 > 0$, it follows from Theorem 4.1 that:

$$\begin{aligned} P\left(\frac{\partial g(d)}{\partial \delta} \geq 0\right) \quad &\longrightarrow \quad P\Big(W'(\delta) - E(W'(\delta)) \geq -E(W'(\delta))\Big) \\ &\leq \quad \frac{V\left(W'(\delta)\right)}{\left(E\left(W'(\delta)\right)\right)^2} = \frac{6}{\pi^2\delta^2}\,, \end{aligned} \tag{59}$$

by Chebyshev's inequality. Similarly, for $d = d_0 + \delta_2/\sqrt{T} < d_0$ with $\delta = \delta_2 < 0$,

$$\begin{aligned} P\left(\frac{\partial g(d)}{\partial \delta} \leq 0\right) \quad &\longrightarrow \quad P\Big(-W'(\delta) + E(W'(\delta)) \geq E(W'(\delta))\Big) \\ &\leq \quad \frac{6}{\pi^2\delta^2}\,. \end{aligned} \tag{60}$$

Thus it is ensured that the probabilities in (59) and (60) can be made smaller than ε by taking $\delta_0 = \max(\delta_1, |\delta_2|, \sqrt{6/(\pi^2\varepsilon)})$. Then T_0 can be chosen so that (58) holds for all $T \geq T_0$.

Returning to Theorem 4.1, we can assert that a local maximizer $\tilde{d}$ of $g(d)$ is asymptotically the MLE of d_0 since $W(\delta)$ has a unique maximum which occurs at $\hat{\delta} = Z/\sqrt{\pi^2/6}$. Thus we have the following result.

Theorem 4.2. *Let $\hat{d}$ be the MLE of d_0 for the model (56). Then it holds that, as $T \to \infty$,*

$$\sqrt{T}\left(\hat{d} - d_0\right) \longrightarrow N\left(0, 6/\pi^2\right)\,.$$

Note that $\pi^2/6$ is the limiting value of the normalized Fisher information for d. Hence the above result turns out to be quite standard, on the basis of which we can also suggest a test for d. Namely, for the right-sided alternative $H_1 : d > d_0$, we can reject $H_0 : d = d_0$ when $\sqrt{T}(\hat{d} - d_0)/\sqrt{6/\pi^2}$ exceeds the upper $100\alpha\%$ point of N(0,1). The test may be called the Wald test, and is a test of asymptotic size α. It holds that the limiting local power of the Wald test is the same as that of the LBI test. The left-sided Wald test can also be conducted similarly.

4.2. Stationary Case

We next consider the model

$$(1-L)^d a(L) y_j = b(L)\varepsilon_j\,, \qquad (j = 1, \cdots, T)\,, \tag{61}$$

where $\{\varepsilon_j\} \sim \mathrm{NID}(0, \sigma^2)$, $a(L) = 1 - a_1 L - \cdots - a_p L^p$ and $b(L) = 1 - b_1 L - \cdots - b_q L^q$ with $a(x) \neq 0$ and $b(x) \neq 0$ for $|x| \leq 1$. The parameters to be estimated are d, $\psi = (a_1, \cdots, a_p, b_1, \cdots, b_q)'$ and σ^2. The concentrated log-likelihood for d and ψ is now given, except for constants, by

$$\ell(d, \psi) = -\frac{T}{2} \log \left[\sum_{j=1}^{T} \left\{ a(L) b^{-1}(L)(1-L)^d y_j \right\}^2 \right]. \tag{62}$$

If we consider $g(d, \psi) = \ell(d, \psi) - \ell(d_0, \psi_0)$ as in the i.i.d. case, where d_0 and ψ_0 are the true values of d and ψ, respectively, it can be checked that $g(d, \psi)$ is asymptotically a concave function of $\delta = \sqrt{T}(d - d_0)$ and $\gamma = \sqrt{T}(\psi - \psi_0)$. Hence the MLE's of d_0 and ψ_0 are asymptotically unique, with an asymptotic distribution given by the following theorem.

Theorem 4.3. *Let $\hat{\tau} = (\hat{d}, \hat{\psi}')'$ be the MLE of $\tau_0 = (d_0, \psi_0')'$ for the model (61). Then it holds that, as $T \to \infty$,*

$$\sqrt{T}\,(\hat{\tau} - \tau_0) \longrightarrow N\left(0,\, \Xi^{-1}\right),$$

where

$$\Xi = \begin{pmatrix} \pi^2/6 & \kappa' \\ \kappa & \Gamma_{p+q}/\sigma^2 \end{pmatrix} \tag{63}$$

with $\kappa = (\kappa_1, \cdots, \kappa_{p+q})'$ and Γ_{p+q} defined in (54).

The above result is also standard since Ξ turns out to be the limiting matrix of the normalized Fisher information for d and ψ. Thus $\hat{d}$ and $\hat{\psi}$ are asymptotically efficient. Since we have

$$\sqrt{T}(\hat{d}-d_0) \longrightarrow N\left(0,\, \omega^{-2}\right) = N\left(0,\, \left(\pi^2/6 - \sigma^2\kappa'\Gamma_{p+q}^{-1}\kappa\right)^{-1}\right),$$

it is recognized that the asymptotic efficiency of $\hat{d}$ is decreased in the present case where ψ is estimated. The same is true for $\hat{\psi}$. This is because $\hat{d}$ and $\hat{\psi}$ are asymptotically correlated.

It can also be checked that

$$\Xi = \frac{1}{4\pi}\int_0^{2\pi} \frac{\partial \log h(\lambda,\tau)}{\partial \tau}\, \frac{\partial \log h(\lambda,\tau)}{\partial \tau'}\, d\lambda\,, \tag{64}$$

where $h(\lambda,\tau)$ is the spectrum-like quantity for $y_j = b(L)\varepsilon_j/\left(a(L)(1-L)^d\right)$ defined by

$$h(\lambda,\tau) = \frac{|b(e^{i\lambda})|^2}{|1-e^{i\lambda}|^{2d}\,|a(e^{i\lambda})|^2}\,.$$

The expression in (64) is well known (Walker (1964)), except for the first column and row.

To justify (64), consider

$$\begin{aligned} \frac{\partial \log h(\lambda,\tau)}{\partial d} &= -\log|1-e^{i\lambda}|^2 = -2\log\left(2\sin\frac{\lambda}{2}\right) \\ &= \sum_{n=1}^{\infty}\frac{2}{n}\cos n\lambda\,, \qquad (0<\lambda<2\pi)\,. \end{aligned} \tag{65}$$

It is known (see, for example, Zygmund, 1968, p.180) that, for any square integrable function $f(\lambda)$,

$$\begin{aligned} \frac{1}{4\pi}\int_0^{2\pi}\frac{\partial \log h(\lambda,\tau)}{\partial d} f(\lambda)d\lambda &= -\frac{1}{2\pi}\int_0^{2\pi}\log\left(2\sin\frac{\lambda}{2}\right) f(\lambda)d\lambda \\ &= \sum_{n=1}^{\infty}\frac{1}{2n}c_n\,, \end{aligned} \tag{66}$$

where c_n's are coefficients in the Fourier expansion of $f(\lambda)$:

$$f(\lambda) \sim \frac{1}{2}c_0 + \sum_{n=1}^{\infty}(c_n\cos n\lambda + d_n \sin n\lambda)\,.$$

In particular, if $f(\lambda) = \partial \log h(\lambda,\tau)/\partial d$, then (65) yields $c_n = 2/n$ so that (66) gives us $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$, which is the (1,1)-element of Ξ. If $f(\lambda) = \partial \log h(\lambda,\tau)/\partial \psi_j$

$(j = 1, \cdots, p+q)$, the computation in (66) yields κ_j in (63), although it is much involved in general.

We can also devise a test based on the MLE of d, as in the i.i.d. case. The right-sided (left-sided) Wald test rejects $H_0 : d = d_0$ when $\sqrt{T}\,\hat{\omega}(\hat{d} - d_0)$ exceeds the upper (lower) $100\alpha\%$ point of N(0,1). It holds that the Wald test has the same limiting local power as the LM test, as in the i.i.d. case. The finite sample performance of these tests as well as the MLE of d will also be examined in the next section.

5. Simulation Experiments

In this section we examine, by simulations, the finite sample performance of the LBI or LM and Wald tests suggested in previous sections. In the course of the simulations the behavior of the MLE can also be examined.

The models employed here are

$$\text{Model A:} \qquad (1-L)^d y_j = \varepsilon_j\,, \qquad (j = 1, \cdots, T)$$

$$\text{Model B:} \qquad (1-L)^d (1-aL) y_j = \varepsilon_j\,, \qquad (j = 1, \cdots, T)\,,$$

where $\{\varepsilon_j\} \sim \text{NID}(0, \sigma^2)$. For both models we consider testing for $H_0 : d = d_0$ against $H_1 : d > d_0$ or $H_2 : d < d_0$. Throughout simulations, we fix the sample size at $T = 100$, the number of replications at 1,000, and the significance level at the nominal 5% based on N(0,1).

Let us first deal with Model A. Table 1 reports percentage powers of various tests for $H_0 : d = 1/2$ against $H_1 : d > 1/2$, where

$$S'_{T1} = \sqrt{T} \sum_{j=1}^{T-1} \frac{1}{j} \hat{\rho}_j \Big/ \sqrt{\frac{\pi^2}{6}}\,,$$

$$S''_{T1} = -\sqrt{T} \sum_{j=1}^{T} \hat{\varepsilon}_j \frac{\partial \hat{\varepsilon}_j}{\partial d} \Big/ \sqrt{\sum_{j=1}^{T} \hat{\varepsilon}_j^2 \sum_{j=1}^{T} \left(\frac{\partial \hat{\varepsilon}_j}{\partial d} \right)^2} \;\Bigg|_{H_0}\,,$$

$$W'_{T1} = \sqrt{T} \left(\hat{d} - d_0 \right) \Big/ \sqrt{6/\pi^2}\,,$$

$$W''_{T1} = \sqrt{T} \left(\hat{d} - d_0 \right) \times \sqrt{\sum_{j=1}^{T} \left(\frac{\partial \hat{\varepsilon}_j}{\partial d} \right)^2 \Big/ \sum_{j=1}^{T} \hat{\varepsilon}_j^2} \;\Bigg|_{H_1}\,.$$

Here the statistics S'_{T1} and W'_{T1} are respectively the original LBI and Wald statistics suggested before, while S''_{T1} and W''_{T1} are their modified versions, which can be justified by observing that

$$\begin{aligned}\sum_j \hat{\varepsilon}_j \frac{\partial \hat{\varepsilon}_j}{\partial d} &= \sum_j \hat{\varepsilon}_j \{\log(1-L) \times \hat{\varepsilon}_j\} = -\sum_j \hat{\varepsilon}_j \sum_k \frac{1}{k} \hat{\varepsilon}_{j-k} \\ &= -\sum_k \frac{1}{k} \sum_j \hat{\varepsilon}_{j-k} \hat{\varepsilon}_j = -\sum_k \frac{1}{k} \hat{\rho}_k \sum_j \hat{\varepsilon}_j^2 ,\end{aligned}$$

$$\operatorname{plim} \sum_j \left(\frac{\partial \hat{\varepsilon}_j}{\partial d} \right)^2 \Big/ \sum_j \hat{\varepsilon}_j^2 = \frac{\pi^2}{6} \qquad \text{under} \quad H_0 .$$

To compute S''_{T1} and W''_{T1} we need the partial derivative $\partial \hat{\varepsilon}_j / \partial d$, which can be computed numerically from

$$\frac{\partial \hat{\varepsilon}_j}{\partial d} = \frac{\hat{\varepsilon}_j(d + \triangle d) - \hat{\varepsilon}_j(d)}{\triangle d},$$

where $\hat{\varepsilon}_j(d)$ is computed under the assumption that the integration order is d. The value of $\triangle d = 0.001$ has been used in our simulation studies. Note that the partial derivative should be computed under H_0 for S''_{T1}, and under H_1 for W''_{T1}. The MLE of d has been obtained from the Gauss-Newton procedure:

$$\hat{d}_{(i)} = \hat{d}_{(i-1)} - \sum_j \hat{\varepsilon}_j \frac{\partial \hat{\varepsilon}_j}{\partial d} \Big/ \sum_j \left(\frac{\partial \hat{\varepsilon}_j}{\partial d} \right)^2 \qquad (i = 1, 2, \cdots) .$$

To initiate the iterative scheme we need an initial value $\hat{d}_{(0)}$, for which we used the true value of d.

In Table 1 we also report under the heading 'Limit' the theoretical limiting powers derived from Corollary 3.1. It is seen that the power of each test is reasonably well approximated by the limiting power. In particular, the S'_{T1}-test seems to behave best in the present case. Note that the column entries corresponding to $d = 0.50$ are type I errors. Table 2 reports percentage powers of the left-sided tests for the same model as in Table 1. The S'_{T1}-test does not behave well in the present case because of size distortion towards 0. Robinson (1994) also deals with Model A and reports powers of the tests derived from the frequency domain approach. Overall the present tests behave better, in terms of size and power, than the frequency domain-based tests.

Table 1 Table 2

Tables 3 to 6 are concerned with Model B, with an AR coefficient of $a = 0.6$. Tables 3 and 4 present results for the test of $H_0 : d = 1$, while Tables 5 and 6 present results for $H_0 : d = 1.5$. Tables 7 and 8 deal with Model B with $a = -0.8$ and present results for $H_0 : d = 1$. The statistics S'_{T2} and W'_{T2} are respectively the LM and Wald statistics defined by

$$S'_{T2} = \sqrt{T} \sum_{k=1}^{T-1} \frac{1}{k} \hat{\rho}_k \Big/ \hat{\omega}_0 , \qquad W'_{T2} = \sqrt{T}\, \hat{\omega}_1 \left(\hat{d} - d_0 \right) ,$$

where $\hat{\omega}_0$ and $\hat{\omega}_1$ are the estimators of ω under H_0 and H_1, respectively, with ω^2 defined in (55). The statistics S''_{T2} and W''_{T2} are defined in the same way as S''_{T1} and W''_{T1}, respectively. Theoretical limiting powers derived from Theorem 3.3, namely, $\Phi\left(-z_\alpha + \sqrt{T}\omega(d - d_0)\right)$ are also reported under the heading 'Limit', where $T = 100$, $z_\alpha = 1.645$ and ω^2 is given in (55).

Table 3 Table 4 Table 5 Table 6 Table 7 Table 8

The general feature of Tables 3 to 6 is that size distortion is serious. Hence there is much discrepancy between finite sample and limiting powers. On the other hand, in Tables 7 and 8, the finite sample powers are well approximated by the limiting powers. This is closely related to the fact that the estimators of a and d are negatively correlated and the correlation is much higher for $a = 0.6$, as Table 9 below shows. Thus the inference on d is affected by the confusing impact of a, though the degree of influence depends on the value of a. It is also noticed that the limiting powers are quite low in Tables 3 to 6, in comparison with those in Tables 7 and 8. For example, the theoretical power of the test for $H_0 : d = 1$ against $d = 0.8$ is 19.4% in the former, while it is 77.8% in the latter. Given the sample size T, the significance level α and the distance from H_0, $|d - d_0|$, the theoretical power depends only on ω and becomes higher as ω gets large or ω^{-1} gets small. Note that ω^{-1} is the standard error of the limiting distribution of $\sqrt{T}(\hat{d} - d)$, where $\hat{d}$ is the MLE of d. The value of ω^{-1} will also be reported in Table 9.

Figures 5 to 8 draw histograms of the normalized MLE $\sqrt{T}(\hat{d}-d)$ obtained from the same simulations as above, and the correponding limiting densities of $\mathrm{N}(0,\, 6/\pi^2)$ or $\mathrm{N}(0,\, \omega^{-2})$. Figure 5 is for the i.i.d. case, while the others are for stationary cases. The behavior of the MLE for the i.i.d. case agrees quite well with the asymptotic theory, while that for stationary cases varies depending on the value of a. In particular, the performance of the MLE's in Figures 6 and 7 is very poor, where $a = 0.6$ is assumed, while the MLE in Figure 8 performs well, where $a = -0.8$ is assumed.

Figure 5 Figure 6 Figure 7 Figure 8

To see why, Table 9 reports ω^{-1}, the standard error of the limiting distribution of $\sqrt{T}(\hat{d}-d)$, for various values of a. The correlation coefficient of the limiting distributions of $\sqrt{T}(\hat{d}-d)$ and $\sqrt{T}(\hat{a}-a)$ is also reported under the heading 'Corr'. It is seen that ω^{-1} increases with a up to $a = 0.7$, and then decreases. Thus d can be estimated much better for $a = -0.8$ than for $a = 0.6$. It is also noticed that the correlation coefficient is negative for all values of a. It increases in absolute value with a up to $a = 0.7$, and then decreases. This is another source of the poor performance of the MLE in Figures 7 and 8. The source of the size distortion in the Wald test is also evident. The nonmonotonic behavior of ω^{-1} and 'Corr' beyond $a = 0.7$ seems to reflect the AR unit root situation as a approaches 1.

Table 9

6. Concluding Remarks

In this paper we dealt with a fractional $\mathrm{I}(d)$ process $\{y_j\}$ with any real number d. After investigating asymptotic properties of various statistics associated with $\{y_j\}$ when d is known and is greater than or equal to $1/2$, we discussed, under the normality assumption, testing and estimation for d without restricting the parameter space of d. It was shown that standard asymptotic results hold for tests and estimators. Namely, the LBI and Wald tests are asymptotically UMPI, and the MLE of d is asymptotically efficient. These asymptotic results hold without the normality assumption. The finite sample behavior of the tests and estimators were examined by simulations, and the source of different behavior was made clear in terms of the asymptotic theory.

This paper discussed nonseasonal time series only, but the discussion may be extended, as in Chung (1996), to deal with seasonality by considering

$$\left(1 - 2\cos\lambda\, L + L^2\right)^d y_j = u_j\,,$$

where $\lambda\,(0 < \lambda < \pi)$ is the seasonal frequency.

The analysis can also be extended to deal with fractional cointegration, as in Jeganathan (1996), where variables follow I(d) processes with d greater than or equal to 1/2. As an example, we can consider the following fractionally cointegrated system:

$$y_{j2} = \beta y_{j1} + u_{j2}\,, \qquad (1-L)^d y_{j1} = u_{j1}\,, \qquad (j = 1, \cdots, T)\,,$$

where $\{u_{j1}\}$ and $\{u_{j2}\}$ are stationary processes. As was observed in Section 1, inference on β requires a separate treatment of the two cases $d = 1/2$ and $d > 1/2$. If $\hat{\beta}$ is the LSE of β, it holds that $\sqrt{T \log T}(\hat{\beta} - \beta)$ converges to a nondegenerate distribtuion when $d = 1/2$, while $T^d(\hat{\beta} - \beta)$ has a nondegenerate limiting distribution when $d > 1/2$. The same is true for the MLE of d. A detailed analysis is currently being undertaken.

References

Anderson, T.W. (1971) *The Statistical Analysis of Time Series.* New York: Wiley.

Beran, J. (1994) *Statistics for Long-Memory Processes.* New York: Chapman & Hall.

Box, G.E.P. & D.A. Pierce (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509-1526.

Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* **42**, 59-66.

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

Chan, N.H. & N. Terrin (1995) Inference for unstable long-memory processes with applications to fractional unit root autoregressions.

Chung, C.-F. (1996) Estimating a generalized long memory process. *Journal of Econometrics* **73**, 237-259.

Diebold, F.X. & G.D. Rudebusch (1991) On the power of Dickey-Fuller tests against fractional alternatives. *Economics Letters* **35**, 155-160.

Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* **64**, 813-836.

Hannan, E.J. & C.C. Heyde (1972) On limit theorems for quadratic functions of discrete time series. *Annals of Mathematical Statistics* **43**, 2058-2066.

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

Hosoya, Y. (1997) A limit theory for long-range dependence and statistical inference on related models. *Annals of Statistics* **25**, 105-137.

Jeganathan, P. (1996) On asymptotic inference in cointegrated time series with fractionally integrated errors. Preprint, Department of Statistics, University of Michigan, Ann Arbor.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* **20**, 971-1001.

Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* **89**, 1420-1437.

Sargan, J.D. & A. Bhargava (1983) Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle. *Econometrica* **51**, 799-820.

Sowell, F. (1990) The fractional unit root distribution. *Econometrica* **58**, 495-505.

Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory.* New York: Wiley.

Walker, A.M. (1964) Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society* **4**, 363-384.

Zygmund, A. (1968) *Trigonometric Series, Vol. I.* Cambridge: Cambridge University Press.

Appendix

Proof of Theorem 2.1.

We first rewrite (5) as $y_j = \sum_{k=1}^{j} a_{kj}\varepsilon_k$, where $a_{kj} = \Gamma(j-k+d)/(\Gamma(d)\Gamma(j-k+1))$ with $d = 1/2$. Since $\{a_{kj}\varepsilon_k\}$ is an independent sequence for each j, Theorem 2.1 is established by Brown's (1971) FCLT if we show that the Lindeberg condition holds, namely,

$$\frac{1}{s_T^2}\sum_{k=1}^{T} E\left[a_{kT}^2\varepsilon_k^2\, I\left(\,|a_{kT}\varepsilon_k|\, > \delta s_T\right)\right] \longrightarrow 0 \qquad \text{for every} \quad \delta > 0\,,$$

where $I(A)$ is the indicator function of the set A. Since $s_T^2 = \sigma^2\sum_{k=1}^{T} a_{kT}^2$, $|a_{kT}| \leq 1$ and

$$E\left[a_{kT}^2\varepsilon_k^2\, I\left(\,|a_{kT}\varepsilon_k|\, > \delta s_T\right)\right] \leq a_{kT}^2 E\left[\varepsilon_1^2\, I\left(\,|\varepsilon_1|\, > \delta s_T\right)\right]\,,$$

the Lindeberg condition is seen to hold, which establishes Theorem 2.1.

Proof of Corollary 2.1.

The result in (10) is an immediate consequence of Theorem 2.1, while (11) comes from the continuous mapping theorem (CMT). To see this, consider

$$Q_T = \int_0^1 X_T^2(t)dt + R_T\,,$$

where

$$\begin{aligned} R_T &= \sum_{j=1}^{T} X_T^2\left(\frac{s_j^2}{s_T^2}\right)\frac{s_j^2 - s_{j-1}^2}{s_T^2} - \int_0^1 X_T^2(t)dt \\ &= \sum_{j=1}^{T}\int_{s_{j-1}^2/s_T^2}^{s_j^2/s_T^2}\left[X_T^2\left(\frac{s_j^2}{s_T^2}\right) - X_T^2(t)\right]dt\,. \end{aligned}$$

We have

$$|R_T| \,\leq 2\sup_t\, |X_T(t)|\,\frac{1}{s_T}\max_j\, |y_j - y_{j-1}|\,,$$

where $\sup_t\, |X_T(t)|\, = O_p(1)$. Since

$$P\left(\frac{1}{s_T}\max_j\, |y_j - y_{j-1}|\, > \delta\right) = P\left(\frac{1}{s_T^2}\sum_{j=2}^{T}(y_j - y_{j-1})^2\, I(\,|y_j - y_{j-1}|\, > \delta s_T\,) > \delta^2\right)$$

for any $\delta > 0$, and $\{y_j - y_{j-1}\} = \{(1-L)^{1/2}\varepsilon_j\}$ is a second-order stationary process, it can be checked that $\max_j |y_j - y_{j-1}|\,/s_T \to 0$ in probability. Thus $R_T \to 0$ in probability, which yields (11) by the CMT and the fact that $(s_j^2 - s_{j-1}^2)/s_T^2 \cong 1/(j\log T)$.

To prove (12), we note that $y_j - y_{j-1} = (1-L)^{1/2}\varepsilon_j$ may be expressed as $\sum_{\ell=0}^{\infty} \alpha_\ell \varepsilon_{j-\ell}$ with $\sum_{\ell=0}^{\infty} |\alpha_\ell| < \infty$. Then the sample variance of $\{y_j - y_{j-1}\}$ converges in probability to $V(y_j - y_{j-1})$ by the result of Hannan and Heyde (1972). Thus we have

$$\begin{aligned} V(y_j - y_{j-1}) &= V\left((1-L)^{1/2}\varepsilon_j\right) \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \left|1 - e^{i\lambda}\right| d\lambda = \frac{4\sigma^2}{\pi}\,. \end{aligned}$$

Proof of Corollary 2.2.

The result in (18) is a consequence of $\mathcal{L}(X_T) \to \mathcal{L}(\phi w)$, which is also used to establish (19). To prove (20), we note that $\{y_j - y_{j-1}\} = \{(1-L)^{1/2}u_j\}$ is a second-order stationary process that may be expressed as $\sum_{\ell=0}^{\infty} \beta_\ell \varepsilon_{j-\ell}$ with $\sum_{\ell=0}^{\infty} |\beta_\ell| < \infty$. Thus the sample variance of $\{y_j - y_{j-1}\}$ converges in probability to

$$\int_{-\pi}^{\pi} \left|1 - e^{i\lambda}\right| f_u(\lambda) d\lambda = 4\int_0^{\pi} f_u(\lambda) \sin\frac{\lambda}{2} d\lambda\,.$$

Proof of Corollary 2.3.

It follows from $\mathcal{L}(X_T) \to \mathcal{L}(\phi w)$ that $X_T(1) = y_T/s_T \to N(0, \phi^2)$, where

$$s_T^2 = \frac{\sigma^2}{\Gamma^2(d)}\sum_{k=0}^{T-1}\frac{\Gamma^2(k+d)}{\Gamma^2(k+1)} \cong \frac{\sigma^2}{\Gamma^2(d)}\frac{T^{2d-1}}{2d-1}\,,$$

which proves (22). Since it can be checked that $(s_j^2 - s_{j-1}^2)/s_T^2 \cong (2d-1)j^{2d-2}/T^{2d-1}$, (23) follows from the same lines as in the proof of Corollary 2.1.

Proof of Corollary 2.4.

The results in (28) and (29) are an immediate consequence of $\mathcal{L}(Y_T) \to \mathcal{L}(\phi F_{d-1})$. To prove (30), we first have

$$\begin{aligned} T^{2d-1}(\hat\rho - 1) &= \frac{1}{2T}\left\{y_T^2 - \sum_{j=2}^{T}(y_j - y_{j-1})^2\right\} \Big/ \frac{1}{T^{2d}}\sum_{j=2}^{T} y_{j-1}^2 \\ &= \left\{-\frac{1}{2}V\left((1-L)^{1-d}u_j\right) + o_p(1)\right\} \Big/ \frac{1}{T^{2d}}\sum_{j=2}^{T} y_{j-1}^2\,, \end{aligned}$$

where $y_T^2 = O_p(T^{2d-1})$ and $y_j - y_{j-1} = (1-L)^{1-d}u_j$. Then (30) follows from (29) and the CMT. The result in (31) is well known in the unit root problem. To prove (32), we have

$$T(\hat\rho - 1) = \frac{1}{2T^{2d-1}}y_T^2 \Big/ \left(\frac{1}{T^{2d}}\sum_{j=2}^{T} y_{j-1}^2\right) + o_p(1)\,,$$

which leads us to (32) by the CMT.

Proof of Theorem 3.1.

We first prove the theorem when $\theta = 0$ and there is no regressor so that $\hat{\varepsilon}_j = \varepsilon_j$. Then it is known (Anderson (1971, p.489)) that the joint distribution of $\sqrt{T}\hat{\rho}_1, \cdots, \sqrt{T}\hat{\rho}_m$ with m fixed tends to $N(0, I_m)$, where I_m is the $m \times m$ identity matrix. Thus $\sqrt{T}\sum_{k=1}^m \hat{\rho}_k/k$ tends to $N(0, \sum_{k=1}^m 1/k^2)$, from which it follows that $S_{T1}/\sqrt{T} \to N(0, \pi^2/6)$ since $\sum_{k=1}^\infty 1/k^2 = \pi^2/6$. Consider next the case for $\theta = \delta/\sqrt{T}$ and with no regressor. We have

$$\begin{aligned}
\hat{\varepsilon}_j &= (1-L)^d y_j = (1-L)^{-\theta}\varepsilon_j \\
&= \varepsilon_j + \frac{\delta}{\sqrt{T}}\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k} + O_p\left(\frac{1}{T}\right),
\end{aligned}$$

$$\text{plim}\,\frac{1}{T}\sum_{j=1}^T \hat{\varepsilon}_j^2 = \sigma^2, \qquad E\left(\frac{1}{\sqrt{T}}\sum_{k=1}^{T-1}\frac{1}{k}\sum_{j=k+1}^T \hat{\varepsilon}_{j-k}\hat{\varepsilon}_j\right) \longrightarrow \frac{\pi^2}{6}\delta\sigma^2,$$

from which we establish that $S_{T1}/\sqrt{T} \to N(\pi^2\delta/6, \pi^2/6)$. When $\theta = \delta/\sqrt{T}$ and there is a regressor, we have $\hat{\varepsilon}_j = \tilde{z}_j - \tilde{x}_j'\hat{\beta} = (1-L)^{-\theta}\varepsilon_j - \tilde{x}_j'(\hat{\beta}-\beta)$. Following the arguments of Robinson (1994) it can be shown that the existence of the regressor does not affect the limiting distribution of $S_{T1}/\sqrt{T}$, which establishes the theorem.

Proof of Corollary 3.1.

It follows from Theorem 3.1 that $S'_{T1} \to N(\delta\sqrt{\pi^2/6}, 1)$ as $T \to \infty$ under $\theta = \delta/\sqrt{T}$, which immediately yields the corollary.

Proof of Theorem 3.2.

Let us first consider the case where there is no regressor. Then it holds that

$$\begin{aligned}
\hat{\varepsilon}_j &= (1-L)^d y_j = (1-L)^{-\theta_0}\varepsilon_j = \varepsilon_j + \frac{c}{\sqrt{T}}\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k} + O_p\left(\frac{1}{T}\right), \\
\tilde{\varepsilon}_j &= (1-L)^{d+\theta} y_j = (1-L)^{\theta-\theta_0}\varepsilon_j = \varepsilon_j + \frac{c-\delta}{\sqrt{T}}\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k} + O_p\left(\frac{1}{T}\right).
\end{aligned}$$

Since it can be checked that

$$\text{plim}\,\frac{1}{T}\sum_{j=1}^T \hat{\varepsilon}_j^2 = \sigma^2,$$

$$\sum_{j=1}^T \hat{\varepsilon}_j^2 - \sum_{j=1}^T \tilde{\varepsilon}_j^2 = \frac{2\delta}{\sqrt{T}}\sum_{k=1}^{T-1}\frac{1}{k}\sum_{j=k+1}^T \varepsilon_{j-k}\varepsilon_j + \frac{\delta(2c-\delta)}{T}\sum_{j=2}^T\left(\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k}\right)^2 + o_p(1),$$

which converges in distribution to $\sigma^2 M(c,\delta)$. Thus we establish the theorem for the case with no regressor. When there is a regressor, the limiting distribution of M_T is unaffected, as in the proof of Theorem 3.1, which proves the theorem.

Proof of Corollary 3.2.

It follows from Theorem 3.2 that, for $\delta > 0$ and $Z \sim \mathrm{N}(0,1)$,

$$\begin{aligned} P(M(\delta,\delta) > x_\alpha(\delta)) &= P\left(\delta\left(2\sqrt{\frac{\pi^2}{6}}\, Z + \frac{\pi^2}{6}\,\delta\right) > x_\alpha(\delta)\right) \\ &= P\left(Z > \left(\frac{1}{\delta}\, x_\alpha(\delta) - \frac{\pi^2}{6}\,\delta\right) \Big/ 2\sqrt{\frac{\pi^2}{6}}\right). \end{aligned}$$

Since $x_\alpha(\delta)$ satisfies

$$\begin{aligned} \alpha &= P(\, M(0,\delta) > x_\alpha(\delta)\,) \\ &= P\left(Z > \left(\frac{\delta\pi^2}{6} + \frac{x_\alpha(\delta)}{\delta}\right) \Big/ 2\sqrt{\frac{\pi^2}{6}}\right), \end{aligned}$$

it must hold that

$$x_\alpha(\delta) = \delta\left(2\sqrt{\frac{\pi^2}{6}}\, z_\alpha - \frac{\delta\pi^2}{6}\right).$$

Thus $P(\, M(\delta,\delta) > x_\alpha(\delta)\,) = P\left(Z > z_\alpha - \delta\sqrt{\pi^2/6}\right)$. The case of $\delta < 0$ can be proved similarly.

Proof of Corollary 3.3.

When $\theta = \delta/\sqrt{T}$ and there is no regressor, we have

$$\hat{\varepsilon}_j = (1-L)^{-d} y_j = (1-L)^{\theta}\varepsilon_j = \varepsilon_j - \frac{\delta}{\sqrt{T}} \sum_{k=1}^{j-1} \frac{1}{k}\varepsilon_{j-k} + O_p\left(\frac{1}{T}\right),$$

$$\operatorname{plim} \frac{1}{T}\sum_{j=1}^{T} \hat{\varepsilon}_j^2 = \sigma^2, \qquad E\left(\frac{1}{\sqrt{T}} \sum_{k=1}^{T-1} \frac{1}{k} \sum_{j=k+1}^{T} \hat{\varepsilon}_{j-k}\hat{\varepsilon}_j\right) \longrightarrow -\frac{\pi^2}{6}\delta\sigma^2.$$

Then it holds that $\tilde{S}'_{T1} \to N\left(-\delta\sqrt{\pi^2/6},\, 1\right)$. The rest of the proof proceeds in much the same way as before.

Proof of Theorem 3.3.

We first consider the case where $\theta = 0$ and there is no regressor. Then $\hat{\rho}_k$ is the k-th order autocorrelation of residuals $\hat{\varepsilon}_j = \hat{a}(L)u_j/\hat{b}(L)$, $(j = 1, \cdots, T)$. Then it follows from Box and Pierce (1970) that the joint distribution of $\sqrt{T}\hat{\rho}_1, \cdots, \sqrt{T}\hat{\rho}_m$ with $m(> p+q)$ fixed tends to $N\left(0,\, I_m - \sigma^2 K_m \Gamma_{p+q}^{-1} K_m'\right)$, where

$$K_m = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ c_1 & 1 & & & & \cdot & -d_1 & -1 & & & & \cdot \\ c_2 & c_1 & & \ddots & & 0 & -d_2 & -d_1 & & \ddots & & 0 \\ \cdot & \cdot & & & & 1 & \cdot & \cdot & & & & -1 \\ \cdot & \cdot & & & & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot & \cdot & & & & \cdot \\ c_{m-1} & c_{m-2} & \cdot & \cdot & \cdot & c_{m-p} & -d_{m-1} & -d_{m-2} & \cdot & \cdot & \cdot & -d_{m-q} \end{pmatrix}$$

Thus $\sqrt{T}\sum_{k=1}^m \hat{\rho}_k/k$ tends to $N(0,\, \omega_m^2)$, where

$$\begin{aligned} \omega_m^2 &= \sum_{k=1}^m \frac{1}{k^2} - \sigma^2(\kappa_{m1}, \cdots, \kappa_{m,p+q})\,\Gamma_{p+q}^{-1}(\kappa_{m1}, \cdots, \kappa_{m,p+q})'\,, \\ \kappa_{m\ell} &= \sum_{k=1}^m \frac{1}{k} K_m(k,\ell) = \sum_{k=\ell}^m \frac{1}{k} c_{k-\ell}\,, \qquad (\ell = 1, \cdots, p+q)\,, \end{aligned}$$

from which it follows that $S_{T2}/\sqrt{T} \to N(0, \omega^2)$. The same asymptotic result holds for the case where $\theta = 0$ and there is a regressor (see Robinson (1994) for the proof by the frequency domain approach). When $\theta = \delta/\sqrt{T}$ and there is no regressor, we have, following the idea of Box and Pierce (1970),

$$\hat{\rho}_k \cong \tilde{\rho}_k - \theta\,\frac{\partial \tilde{\rho}_k}{\partial d} + \frac{\partial \tilde{\rho}_k}{\partial \psi'}\left\{(\tilde{\psi} - \psi) + (\hat{\psi} - \tilde{\psi})\right\},$$

where $\tilde{\rho}_k$ is the kth order autocorrelation for $\{\varepsilon_j\}$, while $\hat{\psi}$ and $\tilde{\psi}$ are the estimators of ψ under H_0 and H_1, respectively. Then it can be checked after some manipulations that $S_{T2}/\sqrt{T} \to N(\delta\omega^2, \omega^2)$. This result is unaffected if there is a regressor.

Proof of Theorem 4.1.

It follows from the proof of Theorem 3.2 that, under $d = d_0 + \delta/\sqrt{T}$,

$$\begin{aligned} &\sum_{j=1}^T \left\{(1-L)^{d_0} y_j\right\}^2 - \sum_{j=1}^T \left\{(1-L)^{d} y_j\right\}^2 \\ &\quad = \sum_{j=1}^T \varepsilon_j^2 - \sum_{j=1}^T \left\{(1-L)^{d-d_0}\varepsilon_j\right\}^2 \\ &\quad = \frac{2\delta}{\sqrt{T}} \sum_{k=1}^{T-1} \frac{1}{k} \sum_{j=k+1}^T \varepsilon_{j-k}\varepsilon_j - \frac{\delta^2}{T}\sum_{j=2}^T \left(\sum_{k=1}^{j-1} \frac{1}{k}\varepsilon_{j-k}\right)^2 + o_p(1)\,, \end{aligned}$$

which converges in distribution to $2\sigma^2 W(\delta)$. Since

$$g(d) = -\frac{T}{2}\log\left[1 - \frac{1}{T}(2W(\delta) + o_p(1)\,)\right] = W(\delta) + o_p(1)\,,$$

the first statement in the theorem is established. Since it can be checked that

$$\begin{aligned}\frac{\partial g(d)}{\partial \delta} &= -\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\left\{\log(1-L)\times(1-L)^d y_j\right\}(1-L)^d y_j \Bigg/ \left[\frac{1}{T}\sum_{j=1}^{T}\left\{(1-L)^d y_j\right\}^2\right] \\ &= \frac{1}{\sigma^2}\left[\frac{1}{\sqrt{T}}\sum_{k=1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T}\varepsilon_{j-k}\varepsilon_j - \frac{\delta}{T}\sum_{j=2}^{T}\left(\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k}\right)^2\right] + o_p(1)\,,\end{aligned}$$

the second statement is established. The last statement can also be proved similarly.

Proof of Theorem 4.2.

It follows from Theorem 4.1 and the subsequent arguments that $\hat{\delta} = \sqrt{T}(\hat{d} - d_0)$ is asymptotically the unique maximizer of $W(\delta)$, which is given by $\hat{\delta} = Z/\sqrt{\pi^2/6}$. This leads us to the conclusion.

Proof of Theorem 4.3.

For simplicity of presentation, we consider the case where $a(L) = 1 - aL$ and $b(L) = 1$. Let us put $d = d_0 + \delta/\sqrt{T}$ and $a = a_0 + \gamma/\sqrt{T}$, and consider $g(d, a) = \ell(d, a) - \ell(d_0, a_0)$, where d_0 and a_0 are the true parameter values of d and a, respectively, while $\ell(d, a)$ is the concentrated log-likelihood given in (62) with ψ replaced by a. Proceeding in the same way as in the i.i.d. case, it is not hard to deduce that

$$\begin{aligned}g(d,a) = 2\,(\delta,\,\gamma)\,U_T - \frac{\gamma^2}{T\sigma^2}\sum_{j=1}^{T}u_{j-1}^2 - \frac{2\delta\gamma}{T\sigma^2}\sum_{j=2}^{T}\left(\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k}\right)u_{j-1} \\ - \frac{\delta^2}{T\sigma^2}\sum_{j=2}^{T}\left(\sum_{k=1}^{j-1}\frac{1}{k}\varepsilon_{j-k}\right)^2 + o_p(1)\,,\end{aligned}$$

where $u_j = \varepsilon_j/(1 - a_0 L)$ and

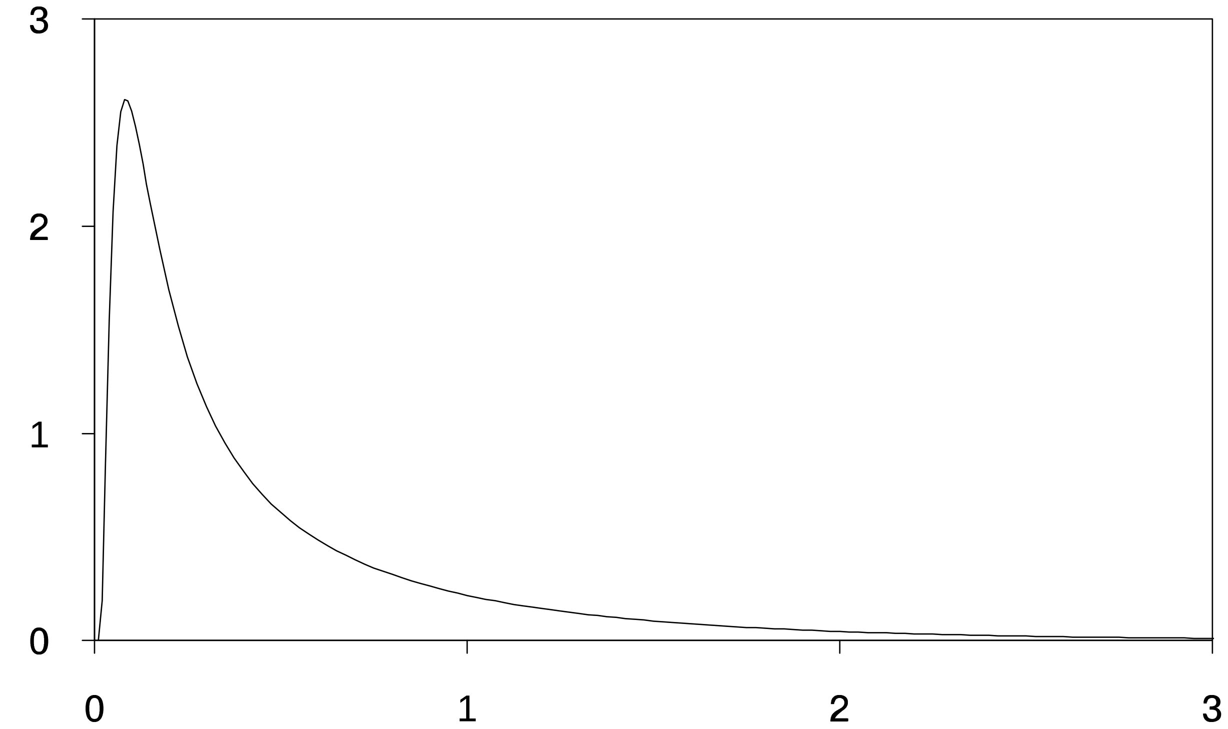
$$U_T = \frac{1}{\sqrt{T}\sigma^2}\left(\sum_{k=1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T}\varepsilon_{j-k}\varepsilon_j\,,\ \sum_{j=1}^{T}u_{j-1}\varepsilon_j\right)'.$$

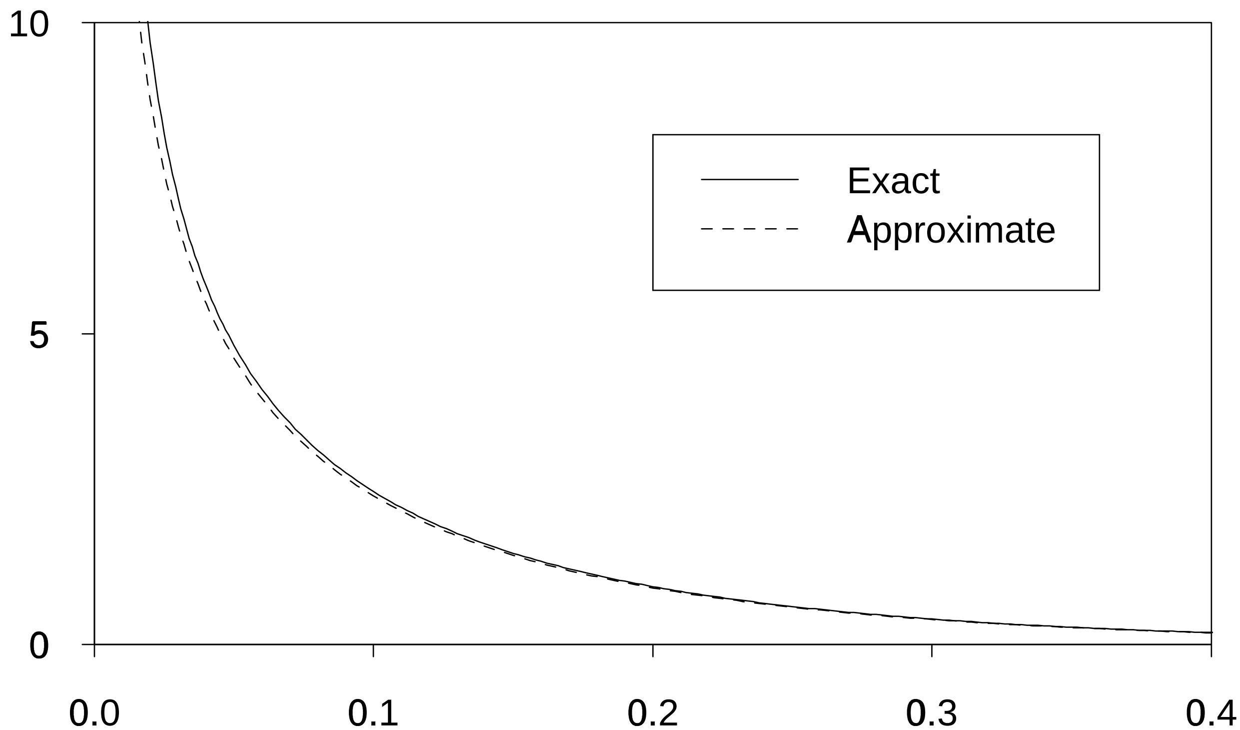
Then it holds that $\mathcal{L}(g(d,a)) \longrightarrow \mathcal{L}(W(\delta,\gamma))$ as $T \longrightarrow \infty$, where

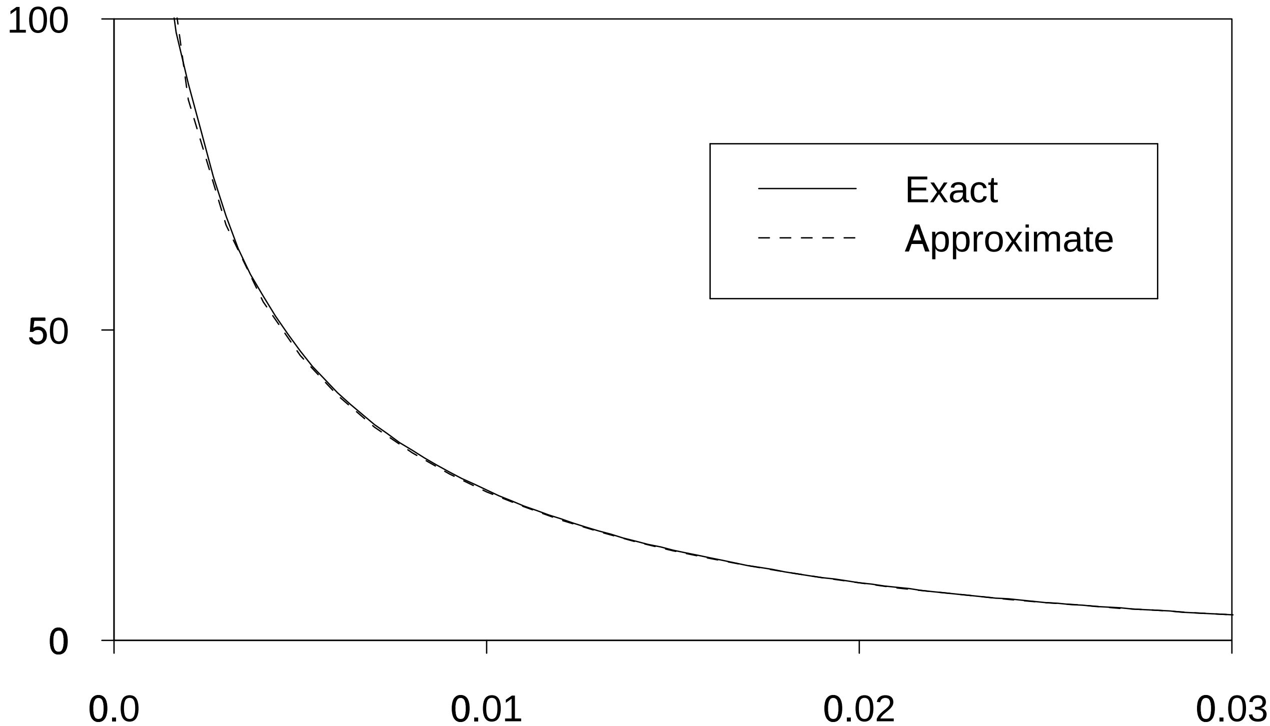
$$W(\delta,\,\gamma) = 2\,(\delta,\,\gamma)\,U - (\delta,\,\gamma)\,\Xi\begin{pmatrix}\delta\\ \gamma\end{pmatrix}, \qquad U \sim N(0,\,\Xi)\,,$$

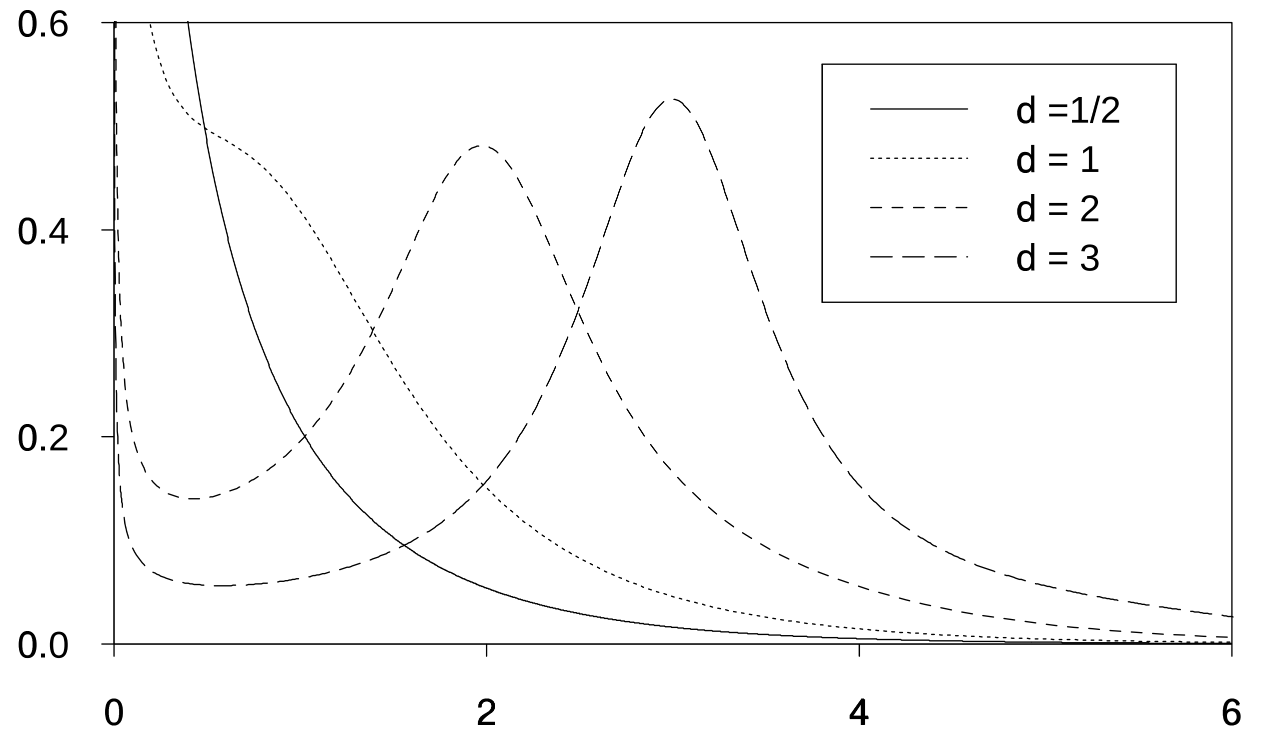
$$\Xi = \begin{pmatrix}\frac{\pi^2}{6} & -\frac{1}{a_0}\log(1-a_0) \\ -\frac{1}{a_0}\log(1-a_0) & \frac{1}{1-a_0^2}\end{pmatrix}.$$

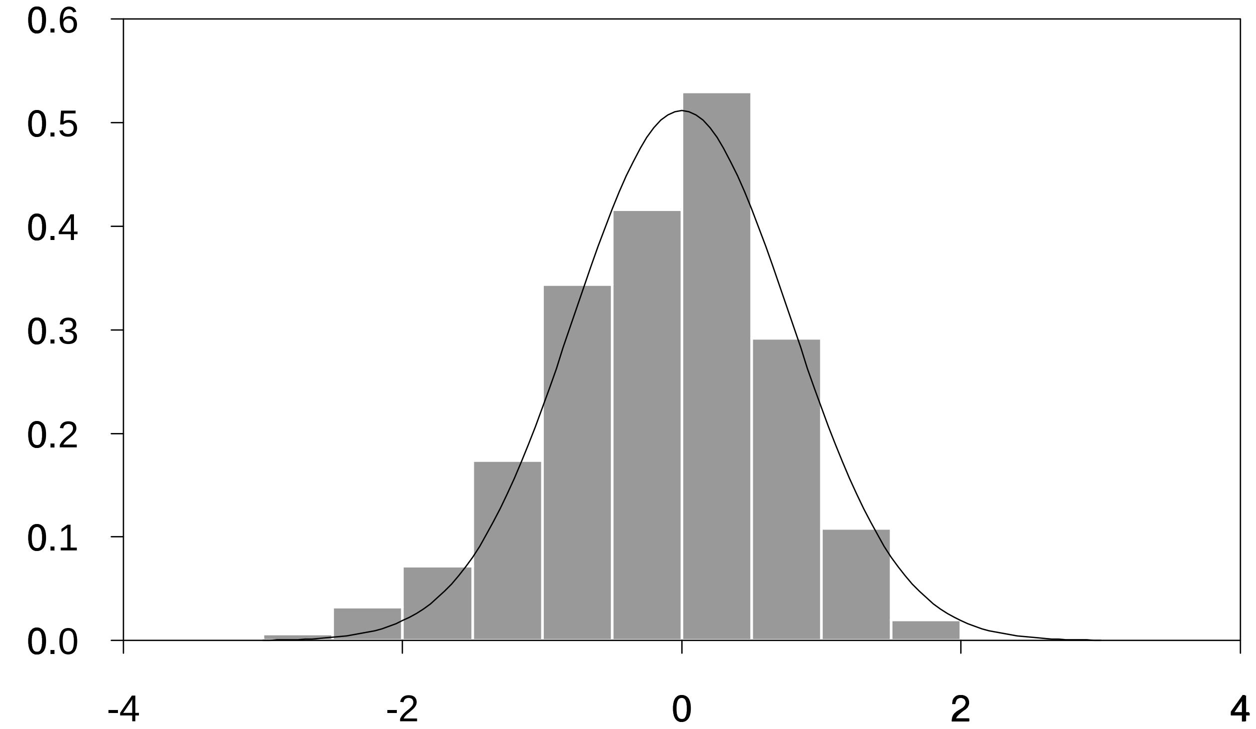
Thus $g(d, a)$ is asymptotically a concave function of $\delta = \sqrt{T}(d - d_0)$ and $\gamma = \sqrt{T}(a - a_0)$, and the MLE's of δ and γ are asymptotically the unique solution to $\Xi(\delta,\,\gamma)' = U$, which establishes the theorem when $a(L) = 1 - aL$ and $b(L) = 1$. The general case can be proved similarly.


3
2
1
0
0
1
2
3

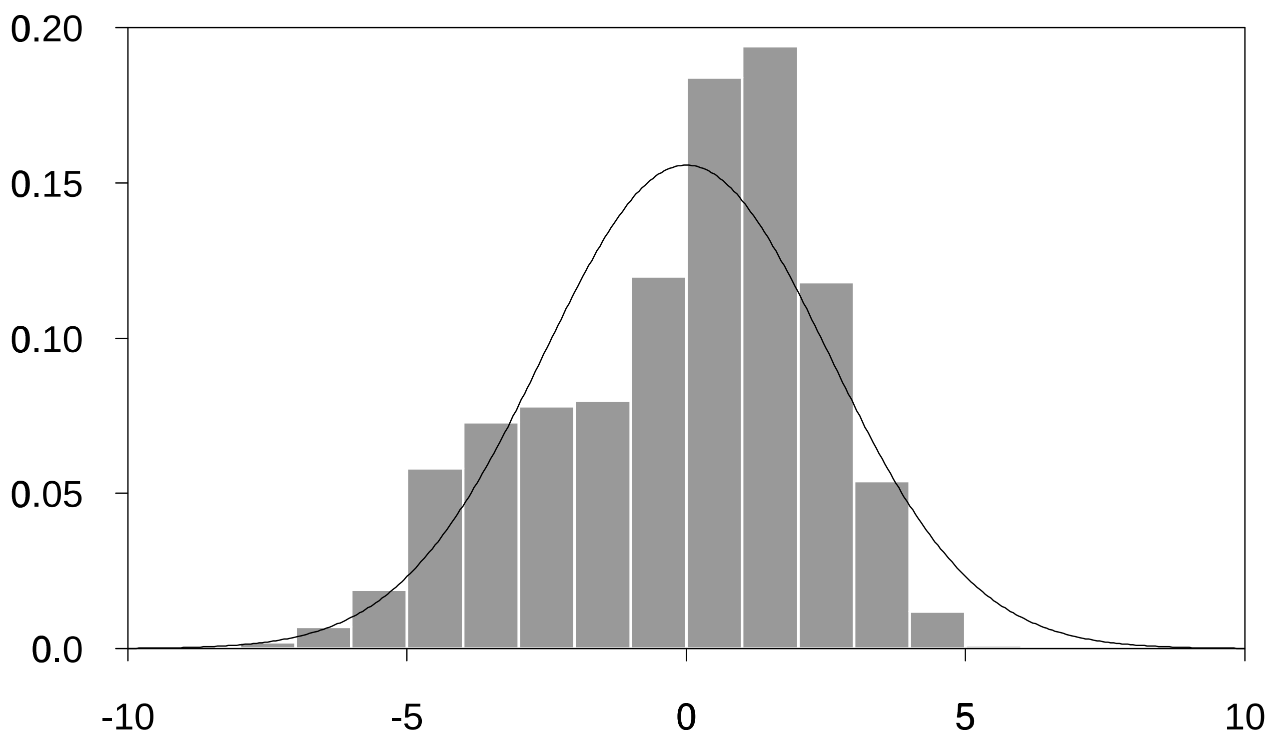

10
5
0
0.0
0.1
0.2
0.3
0.4
Exact
Approximate


100
50
0
0.0
0.01
0.02
0.03
Exact
Approximate

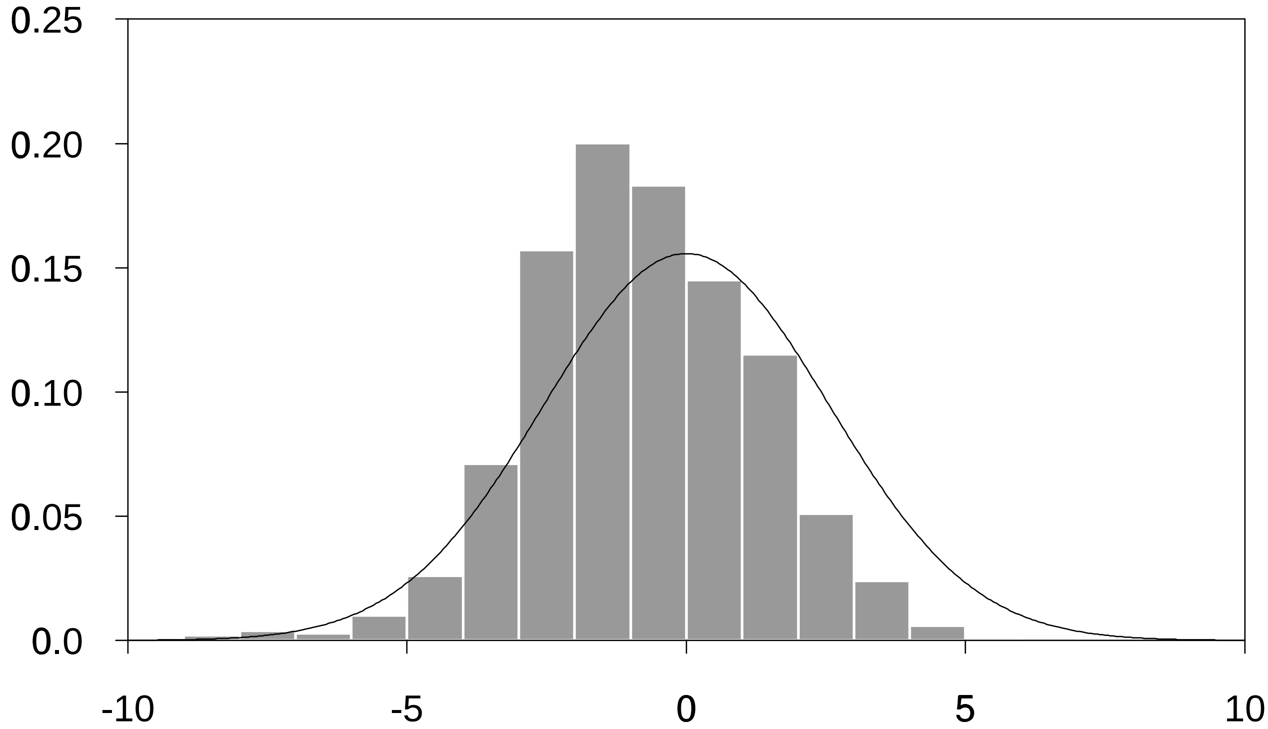

0.6
0.4
0.2
0.0
0
2
4
6
d =1/2
d = 1
d = 2
d = 3

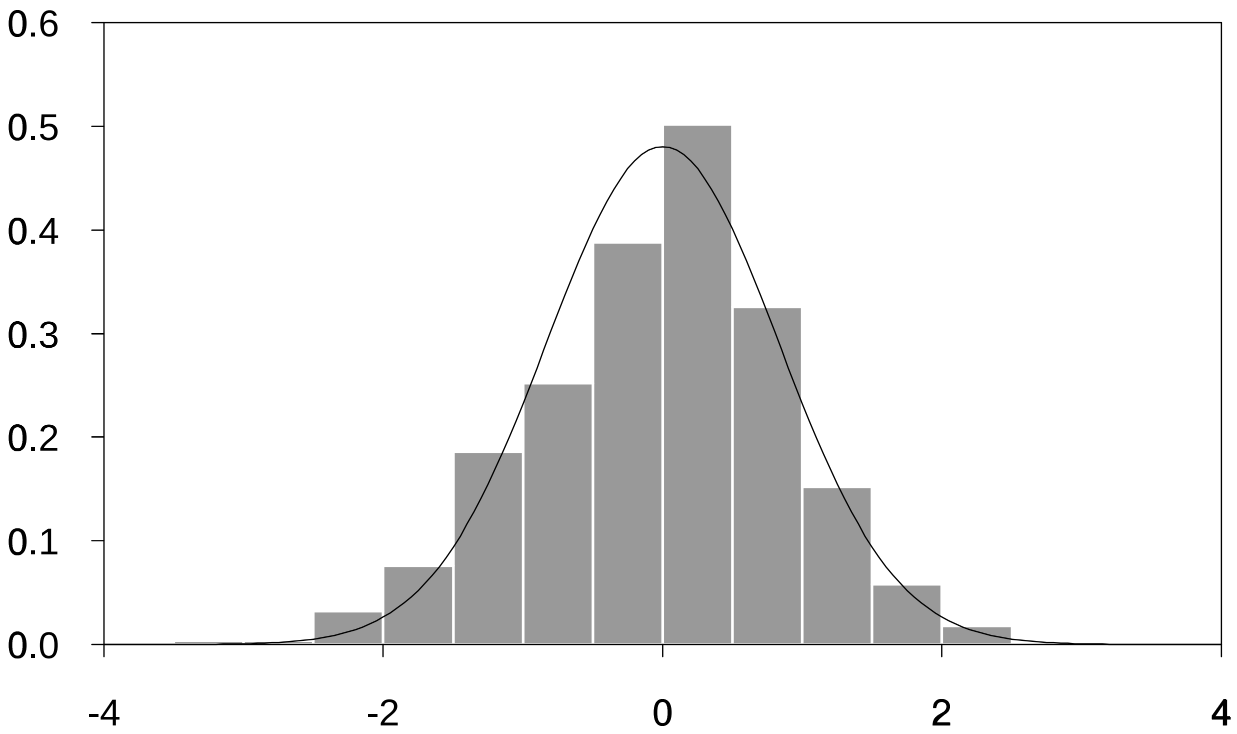


0.6
0.5
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4



0.20
0.15
0.10
0.05
0.0
-10
-5
0
5
10


0.25
0.20
0.15
0.10
0.05
0.0
-10
-5
0
5
10


0.6
0.5
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4