ASYMPTOTIC INFERENCE FOR NONSTATIONARY FRACTIONALLY INTEGRATED AUTOREGRESSIVE MOVING-AVERAGE MODELS

SHIQING LING
Hong Kong University of Science and Technology

W.K. Li
University of Hong Kong

This paper considers nonstationary fractional autoregressive integrated moving-average \( (p,d,q) \) models with the fractionally differencing parameter \( d \in (-\frac{1}{2}, \frac{1}{2}) \) and the autoregression function with roots on or outside the unit circle. Asymptotic inference is based on the conditional sum of squares (CSS) estimation. Under some suitable conditions, it is shown that CSS estimators exist and are consistent. The asymptotic distributions of CSS estimators are expressed as functions of stochastic integrals of usual Brownian motions. Unlike results available in the literature, the limiting distributions of various unit roots are independent of the parameter \( d \) over the entire range \( d \in (-\frac{1}{2}, \frac{1}{2}) \). This allows the unit roots and \( d \) to be estimated and tested separately without loss of efficiency. Our results are quite different from the current asymptotic theories on nonstationary long memory time series. The finite sample properties are examined for two special cases through simulations.

1. INTRODUCTION

Consider the nonstationary fractional autoregressive integrated moving-average (FARIMA) model

\[
\phi_0(B)(1 - B)^{d_0}y_t = \psi_0(B)e_{0t},
\]

where \( \{e_{0t}\} \) is a sequence of independently and identically distributed (i.i.d.) random variables with mean 0, variance 1, and \( E|e_{0t}|^{2+\alpha} < \infty \) for some \( \alpha > 0 \); \( \phi_0(B) = 1 - \phi_{01}B - \cdots - \phi_{0p}B^p \) with \( \phi_{0p} \neq 0 \), \( \psi_0(B) = 1 + \psi_{01}B + \cdots + \psi_{0q}B^q \) with \( \psi_{0q} \neq 0 \), \( |d_0| < \frac{1}{2} \), \( (1 - B)^{d_0} \) is defined by the binomial series.

The authors thank a referee and the co-editor for helpful comments, which led to improvements of this paper. The research of S. Ling was supported in part by a Direct Allocation Grant from the Hong Kong University of Science and Technology. The research of W.K. Li was supported by the Hong Kong Research Grants Council. Address correspondence to: W.K. Li, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong; e-mail: hrntlwk@hku.hk.
\[(1 - B)^{d_0} = \sum_{k=0}^{\infty} a_{0k} B^k, \quad a_{0k} = \frac{(k - d_0 - 1)!}{k!(-d_0 - 1)!}, \tag{1.2}\]

\(B\) is the backward-shift operator; \(\phi_0(z)\) includes unit roots and has the decomposition

\[\phi_0(z) = (1 - z)^a (1 + z)^b \prod_{k=1}^{l} (1 - 2 \cos \theta_k z + z^2)^{d_k}\phi^*_0(z), \tag{1.3}\]

where \(a, b, l, \) and \(d_k\) are nonnegative integers, \(\theta_k \in (0, \pi),\) and \(\phi^*_0(z)\) is a \(p^*\) order polynomial with \(p^* = p - (a + b + 2d_1 + \cdots + 2d_l);\) and \(\phi_0(B)\) and \(\psi_0(B)\) have no common root with all roots outside the unit circle.

When \(\phi_0(B)\) has no unit root, \(\{y_t\}\) defined by (1.1) is stationary and invertible with the autocorrelation function \(\rho(k) \sim c k^{2d_0 - 1}\) as \(k \to \infty,\) where \(c\) is some constant depending on \(d_0.\) When \(d_0 \in (0, 0.5)\) \(\{y_t\}\) possesses the so-called long-memory characteristic because \(\rho(k)\) decreases at a hyperbolic rate that is slower than the exponential rate that characterizes ARMA processes. The stationary FARIMA model was proposed by McLeod and Hipel (1978), Granger and Joyeux (1980), and Hosking (1981). It has been widely applied in modeling long memory behavior in hydrology and economics. Some recent reviews on this topic can be found in Robinson (1994a), Baillie (1996), and Phillips and Xiao (1998, Sect. 6), and the references therein.

When \(\phi_0(B)\) includes unit roots, model (1.1) is nonstationary. Unlike the stationary case, there are two different data generating mechanisms (DGM) for model (1.1) because the AR(\(p\)) part includes a nonstationary component. The first one, namely, DGM1, is to generate \(u_t\) through \(u_t = (1 - B)^{-d_0} \epsilon_{0t}\) and then generate \(y_t\) through \(\phi_0(B) y_t = \psi_0(B) u_t.\) The second, namely, DGM2, is to generate \(u_t\) through \(\phi_0(B) u_t = \psi_0(B) \epsilon_{0t}\) and then generate \(y_t\) through \((1 - B)^{d_0} y_t = u_t.\) It is the preceding different DGMs that result in some different research and conclusions.

To understand the difference between the two DGMs, we consider the simple FARIMA(1, \(d_0\), 0) model, i.e., \((1 - B)(1 - B)^{d_0} y_t = \epsilon_{0t}\) with initial values \(y_t = 0\) as \(t \leq 0.\) From DGM1, it follows that

\[y_t = y_{t-1} + u_t, \quad u_t = \sum_{k=0}^{\infty} \bar{a}_{0k} \epsilon_{0t-k}, \tag{1.4}\]

where \(\bar{a}_{0k} = (k + d_0 - 1)!/[k!(d_0 - 1)!],\) i.e., the coefficient of the \(k\)th term in the expansion of \((1 - B)^{-d_0}.\) From DGM2, because \(y_t = 0\) as \(t \leq 0,\) \(y_t\) is actually generated through the recursion formula \(\sum_{k=0}^{t-1} \bar{a}_{0k} y_{t-k} = u_t,\) where \((1 - B) u_t = \epsilon_{t}, t = 1, 2, \ldots,\) and \(u_0 = 0.\) Thus, we have

\[y_t = y_{t-1} + u^*_t, \quad u^*_t = \sum_{k=0}^{t-1} \bar{a}_{0k} \epsilon_{0t-k}, \tag{1.5}\]
(see Appendix B for the proof of (1.5)). Now, it is clear that observations $y_t$ from model (1.4) are generated by shocks starting at time $-\infty$, whereas $y_t$ from model (1.5) are generated by shocks starting at time 0. Hence $u_t$ is a long memory time series defined as in Hosking (1981), whereas $u_t^*$ is not. From Theorem 1(d) in Hosking (1981), it is not difficult to show that, for any $t, k \geq 1$, $E(u_t^* u_{t+k}^*) = E(u_t u_{t+k}) + O(t^{-\alpha}) = O(k^2 d_0^{-1} + k^{-\alpha})$, where $\alpha > 0$. Thus, $u_t^*$ still have the long memory feature. We call $u_t^*$ a semi–long memory time series.

Model (1.1) from DGM1, with $\phi_0(B) = (1 - B)^m \phi_0^*(B)$, is investigated by Beran (1995). His method is to merge the unit root into the fractional differencing parameter $d_0$, i.e., writing $\phi_0(B)(1 - B)^{d_0}$ as $\phi_0^*(B)(1 - B)^d$, where $d = m + d_0$. He develops an approximate estimation procedure that can estimate $d$ directly. Ling and Li (1997a) extend Beran’s method to permit the noise $e_{0t}$ to follow a GARCH process. It is shown that the estimator of $d$ is asymptotically normal. Other approaches to model (1.4) and its general case are considered by Sowell (1990) and Chan and Terrin (1996), respectively. They treat $u_t$ nonparametrically and investigate the least squares estimator (LSE) of unit roots. Their results indicate that the convergent rates and asymptotic distributions of the LSE of the unit roots on the left hand range $d_0 \in (-\frac{1}{2}, 0)$ and on the right hand range $d_0 \in (0, \frac{1}{2})$ are different. In particular, their asymptotic distributions depend on $d_0$ and involve a sequence of fractional Brownian motions.

Model (1.5), i.e., a special case of model (1.1) under DGM2, is investigated by Robinson (1994b). Because the autocorrelation structure of the series in model (1.5) has the semi–long memory property, Dickey–Fuller tests (see Fuller, 1976; Dickey and Fuller, 1979) and Schmidt–Phillips tests (see Schmidt and Phillips, 1992) are invalid for testing the unit root in model (1.5) unless $d_0 = 0$. Simulation results in Robinson (1994b) show that, even when $d_0 = 0.05$, Dickey–Fuller tests and Schmidt–Phillips tests may reject the unit root null hypothesis in model (1.4) with quite a high probability. As in Beran (1995), Robinson (1994b) also merges the unit root into $d_0$, and he proposes Lagrange multiplier (LM) tests for the unit root against fractional alternatives.

In this paper, we investigate model (1.1) under DGM2. The conditional sum of squares (CSS) estimation procedure is used for $d_0$ and other parameters. Under the general nonstationary framework given by Chan and Wei (1988), we derive the asymptotic distribution of the CSS estimators. Because model (1.1) from DGM2 has only the semi–long memory feature, the asymptotic results are quite different from those in the literature (see the Remarks in Section 2). Our results may be seen as an extension of Chan and Wei (1988), and their results and techniques will be used in our paper. However, our approach needs to estimate $d_0$ and to solve a nonlinear equation. A technical difficulty is to prove that the estimators of unit roots are independent of the estimator of $d_0$. Our asymptotic theory is examined through simulation. The results illustrate that our method is feasible and applicable for the nonstationary FARIMA model.

Jeganathan (1999) considers cointegrating time series with fractional integrated errors. (Our main results were obtained in 1996 and appeared in Ling
and Li [1997b]. We are grateful to N.H. Chan for providing us with a copy of Jeganathan’s [1999] technical report after this paper was completed.) The DGM of his model is principally the same as that of model (1.4), and hence his model includes model (1.4) as a special case. As a result of the different formulations, his results and ours are different. For $y_t$ from the FARIMA$(1, d_0, 0)$ model, both Jeganathan (1999) and the current paper use the transformation $x_t = \sum_{k=1}^{n-1} a_{0k} y_{t-k}$. Under DGM1, $(1 - B)x_t \neq \epsilon_0$, Jeganathan (1999) shows that $n^{-\max\left\{1/2, 1/2 + d_0\right\}} x_{[nt]}$ converges to $F(\tau)$, a functional of the fractional Brownian motion. However, under DGM2, we have $(1 - B)x_t = \epsilon_0$, and hence $n^{-1/2} x_{[nt]}$ converges to the usual Brownian motion. This is the basic reason for the difference between the results in this paper and those of Jeganathan (1999).

This paper proceeds as follows. Section 2 presents the main results. Section 3 reports some simulation results. Section 4 gives the proof of the main results. Throughout the paper, we use the following notations: $U'$ denotes the transpose of the vector $U$; $o(1)(o_p(1))$ denotes a series of numbers (random numbers) converging to zero (in probability); $O(1)(O_p(1))$ denotes a series of numbers (random numbers) that are bounded (in probability); $\rightarrow_{\mathcal{D}}$ denotes convergence in distribution; $D = D[0, 1]$ denotes the space of function $f(s)$ on $[0, 1]$, which is defined and equipped with the Skorokhod topology (Billingsley, 1968); $D^k = D \times D \times \cdots \times D$ ($k$ factors); and $\| \cdot \|$ denotes the Euclidean norm.

2. CSS ESTIMATION AND MAIN RESULTS

We are given observations $y_t$, $t = 1, \ldots, n$, with initial values $y_0 = y_{-1} = \cdots = 0$, which are generated through (1.1) with DGM2. Denote $\lambda_0 = (\phi_0', \psi_0', d_0)'$ with $\phi_0 = (\phi_{10}, \ldots, \phi_{p0})'$ and $\psi_0 = (\psi_{10}, \ldots, \psi_{q0})'$. Note that $\lambda_0$ is the true parameter that together with $\{\epsilon_0\}$ generates $y_t$ through (1.1) with DGM2. To estimate $\lambda_0$, we need the unknown parameter model

$$\phi(B)(1 - B)^d y_t = \psi(B)e_t,$$

(2.1)

where $\phi(B)$, $\psi(B)$, and $(1 - B)^d$ are defined similarly as $\phi_0(B)$, $\psi_0(B)$, and $(1 - B)^{d_0}$, respectively, and the unknown parameter is $\lambda = (\phi', \psi', d)'$ with $\phi = (\phi_1, \ldots, \phi_p)'$ and $\psi = (\psi_1, \ldots, \psi_q)'$.

We assume that $\lambda_0$ and $\lambda$ are the interior points of the compact set $\Theta$, $|d| < \frac{1}{2}$ and all the roots of $\psi(z)$ are outside the unit circle. The CSS estimator is the parameter $\lambda$ in $\Theta$ that minimizes the objective function,

$$S(\lambda) = \sum_{t=1}^{n} e_t^2,$$

(2.2)

where $e_t = \psi^{-1}(B) \phi(B)(1 - B)^d y_t$, a function of $\lambda$ and the observed data $\{y_t\}$. Note that $e_0 = e_t|_{\lambda = \lambda_0}$. When $\{\epsilon_0\}$ is normally distributed, minimizing the CSS function is asymptotically equivalent to maximizing the likelihood function. Box and Jenkins (1976) adopt this method for stationary autoregressive

In the following expressions we show the first-order and second-order partial derivatives of $S(\lambda)$.

\[
\frac{\partial S(\lambda)}{\partial \phi} = 2 \sum_{i=1}^{n} \frac{\partial e_i}{\partial \phi} e_i, \quad \frac{\partial S(\lambda)}{\partial d} = 2 \sum_{i=1}^{n} e_i \log (1 - B) e_i, \quad (2.3)
\]

\[
\frac{\partial^2 S(\lambda)}{\partial \psi \partial \phi'} = 2 \sum_{i=1}^{n} \frac{\partial e_i}{\partial \psi} \frac{\partial e_i}{\partial \phi'}, \quad (2.4)
\]

\[
\frac{\partial^2 S(\lambda)}{\partial \psi \partial \psi'} = 2 \sum_{i=1}^{n} \frac{\partial e_i}{\partial \psi'} e_i + 2 \sum_{i=1}^{n} \frac{\partial^2 e_i}{\partial \psi \partial \phi'} e_i, \quad (2.5)
\]

\[
\frac{\partial^2 S(\lambda)}{\partial d^2} = 2 \sum_{i=1}^{n} \left\{ \left[ \log (1 - B) e_i \right]^2 + e_i \log^2 (1 - B) e_i \right\}, \quad (2.6)
\]

where $\frac{\partial e_i}{\partial \phi} = -\psi^{-1}(B)(1 - B)^{d} y_{i-1}$, $y_i = (y_{1}, \ldots, y_{t-p+1})'$, and $\frac{\partial e_i}{\partial \psi} = -\psi^{-1}(B) e_{i-1}$, $i = 1, \ldots, q$. Let $D_n(\lambda) = \partial S(\lambda)/\partial \lambda$ and $I_n(\lambda) = \partial^2 S(\lambda)/\partial \lambda \partial \lambda'$. Denote $D_n = D_n(\lambda_0)$ and $I_n = I_n(\lambda_0)$. Let $J_n = \text{diag} \left( N_n, \tilde{N}_n, L_{1,n}, \ldots, L_{k,n}, n^{1/2} I_{p' \times p} \right)$.

\[
\tilde{G}_n = J_n^{-1} G^* \quad \text{and} \quad G_n = \text{diag} \left( \tilde{G}_n, n^{-1/2} I_{(q+1) \times (q+1)} \right),
\]

where $N_n = \text{diag} \left( n, \ldots, n^a \right)$, $\tilde{N}_n = \text{diag} \left( n, \ldots, n^b \right)$, $L_{k,n} = \text{diag} \left( n I_{2 \times 2}, \ldots, n^{k-1} I_{2 \times 2} \right)$, $k = 1, \ldots, l$, and $G^*$ is defined as in (4.8) in Section 4.

Now, it is convenient to introduce the unobserved process \{\(Z_t^0: t = 0, \pm 1, \ldots\)\} with \(Z_t^0 = (Z_{\phi t}^0, Z_{\psi t}^0, Z_{dt}^0)\),

\[
Z_{\phi t}^0 = -\phi_0^{-1}(B) \tilde{\varepsilon}_{\phi t}, \quad Z_{\psi t}^0 = -\psi_0^{-1}(B) \tilde{\varepsilon}_{\psi t}, \quad Z_{dt}^0 = -\sum_{k=1}^{\infty} \frac{1}{k} \varepsilon_{0t-k},
\]

where \(\tilde{\varepsilon}_{\phi t} = (\varepsilon_{0t}, \ldots, \varepsilon_{0t-p' + 1})'\) and \(\tilde{\varepsilon}_{\psi t} = (\varepsilon_{0t}, \ldots, \varepsilon_{0t-q+1})'\). Our main results can be stated by the following theorem.

**THEOREM 2.1.** Under the assumptions of model (1.1), it follows that

(a) there exists a sequence of solutions satisfying the equation $\frac{\partial S(\lambda)}{\partial \lambda} = 0$ such that

\[
n^{\alpha - 1/2} G_n^{-1}(\lambda_n - \lambda_0) = o_p(1);
\]

(b) for such a sequence,

\[
G_n^{-1}(\lambda_n - \lambda_0) \rightarrow_{L} ((F^{-1}\xi)'(F^{-1}\xi)',(H_{1}^{-1}\xi_1)'$, $\ldots, (H_{p}^{-1}\xi_p)'$, $(\Sigma^{-1}N)''),
where $0 < \alpha < \frac{1}{4}$, the limiting distributions $F, \xi, \tilde{F}, \tilde{\xi}, H_k,$ and $\xi_k$ are defined in Appendix A, $\Sigma = E(Z_0^T Z_0^t)$, $N$ is a $(p^* + q + 1)$-dimensional standard normal vector, and $(F, \xi), (\tilde{F}, \tilde{\xi}), (H_1, \xi_1), \ldots, (H_1, \xi_1)$, and $N$ are independent.

Remark. The preceding results have three different features from those in Sowell (1990) and Chan and Terrin (1996). First, the estimators of unit roots and $d_0$ are asymptotically independent. This makes it possible to estimate and specify separately the unit roots and the parameter $d_0$ without loss of asymptotic efficiency. Second, the asymptotic distributions of unit roots do not depend on $d_0$, and hence unit root tests based on these distributions can be conducted as those for the nonstationary ARMA model. Third, the convergent rates and asymptotic distributions of the estimators of unit roots are invariant over the entire range $d_0 \in (-\frac{1}{2}, \frac{1}{2})$. This ensures that the unit root tests based on our asymptotic results are robust for different $d_0$.

For various unit roots, the limiting distributions given in Theorem 2.1 have the same form as those given by Chan and Wei (1988) for an unstable autoregressive model. Theorem 2.1 also implies that the limiting distribution of the maximum likelihood estimation for unstable ARMA models is the same as that given by Ling and Li (1998). It is easy to show that

$$
\Sigma = \left( \begin{array}{cc}
E(Z_1^0 Z_1^0) & E(Z_1^0 Z_2^0) \\
E(Z_2^0 Z_2^0) & \pi^2/6 \end{array} \right),
$$

where $Z_{1t} = (Z_{\phi^T}, Z_{\phi A})'$. Thus, the limiting distribution for the stationary component is the same as that in Li and McLeod (1986) for a stationary FARIMA $(p^*, d, q)$ model. The preceding result can be extended to the near unit root case, which can be found in the University of Hong Kong Ph.D thesis by Ling (1997).

**Example 2.1**

Consider the model,

$$(1 - \phi_0 B)(1 - B)^{d_0} y_t = \epsilon_{0t}, \tag{2.7}$$

where $\phi_0 = 1$. Suppose that $(\hat{\phi}, \hat{d})$ is the CSS estimator of $(\phi_0, d_0)$. Then, by Theorem 2.1,

$$
\rho = n(\hat{\phi} - 1) \rightarrow L \int_0^1 W(\tau) dW(\tau) \Big/ \left[ \int_0^1 W^2(\tau) d\tau \right], \quad \equiv \xi \tag{2.8}
$$

$$
\tau = \left( \sum_{i=2}^n u_{i-1}^2 \right)^{1/2} (\hat{\phi} - 1) \rightarrow L \int_0^1 W(\tau) dW(\tau) \Big/ \left[ \int_0^1 W^2(\tau) d\tau \right]^{1/2}, \tag{2.9}
$$

and

$$
\sqrt{n}(\hat{d} - d_0) \rightarrow L N(0, \text{6}/\pi^2), \tag{2.10}
$$

where $u_{i-1} = (1 - B)^{d} y_{i-1}, W(\tau)$ is a standard Brownian motion, and $\rho$ and $\sqrt{n}(\hat{d} - d_0)$ are asymptotically independent.
Remark. The limiting distributions of \( \rho \) and \( \tau \) are the same as those of Dickey–Fuller tests, and hence the critical values given in Tables 8.5.2 and 8.5.3 of Fuller (1976) can be used. The distribution of \( \hat{d} \) is the same as that in Li and McLeod (1986).

Example 2.2

Consider model (1.1) with only a unit root +1, i.e.,

\[
\phi_0(B)(1 - B)^{d_0}y_t = \psi_0(B)\varepsilon_{0t}, \tag{2.11}
\]

and \( \phi_0(B) = (1 - B)\phi_0^*(B) \) with all roots of \( \phi_0^*(B) \) outside the unit circle. Reparameterizing (2.11) as

\[
(1 - B)^{d_0}\left[y_t - \gamma_{01}y_{t-1} - \sum_{i=2}^{p}\gamma_{0i}(y_{t-i+1} - y_{t-i})\right] = \psi_0(B)\varepsilon_{0t},
\]

where \( \gamma_{01} = \sum_{i=1}^{p}\phi_{0i} \) and \( \gamma_{0j} = -\sum_{i=j}^{p}\phi_{0i}, j = 2, \ldots, p. \) Suppose that \( \hat{\phi} \) is the CSS estimator of the parameter \( \phi_0 = (\phi_{01}, \ldots, \phi_{0p})' \). Define \( \hat{\gamma}_1 = \sum_{i=1}^{p}\hat{\phi}_i, \)

\( \hat{\gamma}_j = -\sum_{i=j}^{p}\hat{\phi}_i, j = 2, \ldots, p, \) and \( \lambda_0^* = (\gamma_{02}, \ldots, \gamma_{0p}, \psi_{01}, \ldots, \psi_{0q}, d_0)' \). Similarly define \( \lambda_n^* \). In a similar manner to that given by Ling and Li (1998), we can show that

\[
[n(\hat{\gamma}_1 - 1), \sqrt{n}(\lambda_n^* - \lambda_0^*)'] \overset{D}{\to} \text{diag}(c\xi, \Sigma'), \tag{2.12}
\]

where \( c = 1/(1 - \sum_{i=2}^{p}\gamma_{0i}) \), \( \xi \) is defined by the right-hand side of (2.8), \( \Sigma \) is a normal random vector with mean zero and variance \( \Sigma \), which is independent of \( \xi \). The asymptotic distribution of \( n(\hat{\gamma}_1 - 1) \) has the same form as those given by Dickey and Fuller (1979) and Yap and Reinsel (1995), and hence similar test procedures can be established for testing \( \gamma_{01} = 1 \).

3. SOME SIMULATION RESULTS

In this section, we report some simulation results for two special nonstationary FARIMA models. The first one is the following FARIMA(1, 0, 0) model:

\[
(1 - \phi_0 B)(1 - B)^{d_0}y_t = \varepsilon_{0t},
\]

where \( \phi_0 = 1, |d_0| < \frac{1}{2}, \) and \( \varepsilon_{0t} \sim iid \ N(0,1). \)

We first investigate the finite sample properties of the CSS estimation. In the simulation, the true parameter \( d_0 \) is taken as 0.0, \( \pm 0.1, \) and \( \pm 0.3, \) and \( \phi_0 = 1.0. \) For various parameter pairs of \( (d_0, \phi_0) \) and each of the 1,000 replicates, samples of series lengths \( n = 200 \) and 300 are used, respectively. In the estimation, the LS estimator of \( \phi_0 \) is taken as the initial value of \( \hat{\phi} \). To obtain an initial value of \( \hat{d} \) such that it satisfies the condition in Lemma 4.7 in Section 4, we may take \( \hat{u}_t = (1 - \phi B)y_t \) as artificial observations of the model \( (1 - B)^{d_0}u_t = \varepsilon_{0t} \) and then use the estimation procedure of Li and McLeod.
Table 1. The empirical mean and standard deviation of CSS estimators for the nonstationary FARIMA(1, d0, 0) model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>(\sqrt{\text{MSE}})</th>
<th>Mean</th>
<th>(\sqrt{\text{MSE}})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(d)</td>
<td>(\hat{\phi})</td>
<td>(d)</td>
<td>(\hat{\phi})</td>
</tr>
<tr>
<td>0.3</td>
<td>1.0</td>
<td>0.303 0.991 0.071 0.023</td>
<td>0.301 0.995 0.053 0.012</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>0.105 0.989 0.072 0.029</td>
<td>0.101 0.994 0.052 0.013</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>-0.002 0.989 0.073 0.024</td>
<td>0.000 0.994 0.053 0.013</td>
<td></td>
</tr>
<tr>
<td>-0.1</td>
<td>1.0</td>
<td>-0.099 0.990 0.071 0.020</td>
<td>-0.100 0.994 0.053 0.012</td>
<td></td>
</tr>
<tr>
<td>-0.3</td>
<td>1.0</td>
<td>-0.296 0.991 0.072 0.026</td>
<td>-0.299 0.995 0.053 0.014</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Replications = 1,000.

(1986) to do this. However to save CPU time, here we take \(d_0 + n^{-1/3}\) as the initial value of \(\hat{d}\). The optimization algorithm from Fortran subroutine DBCOAH in the IMSL library is used by constraining \(|d| < \frac{1}{2}\). The results are summarized in Table 1. We can see that the biases of the estimator of \(d_0\) are very small and the empirical standard errors are very close to the asymptotic standard errors \([6/(\pi^2 \times 200)]^{0.5} = 0.0551\) for \(n = 200\) and \([6/(\pi^2 \times 300)]^{0.5} = 0.0450\) for \(n = 300\). For the estimator of \(\phi_0\), all biases and empirical standard errors are small, and they are almost constant for different values of \(d_0\). This is especially the case with \(n = 300\). These findings are consistent with our asymptotic theory. The simulation results indicate that the CSS estimation is feasible and applicable for the nonstationary FARIMA model.

To examine the empirical size and power of the Dickey–Fuller tests \(\rho\) and \(\tau\) in Example 2.1 and the parameters \(d_0 = 0.0, \pm 0.1, \pm 0.3\) and \(\phi_0 = 0.8, 0.90, 0.95, 0.99, 1.0\) are used. For different pairs of the parameter \((d_0, \phi_0)\) and each of the 1,000 replications, samples of series lengths \(n = 200\) and 300 are considered. The critical values of \(\rho\) and \(\tau\) given in Tables 8.5.2 and 8.5.3 of Fuller (1976) are used. The results for the lower 5% Dickey–Fuller tests are reported in Table 2. It is seen that for each case, both size and power are quite satisfactory. In particular, when the sample size \(n = 300\), the empirical sizes are very close to the nominal value of 0.05 and powers are reasonable. The results in Table 2 indicate that Dickey–Fuller tests are robust for different values of \(d_0\).

Now, we simulate the FARIMA(1, d0, 1) model

\[(1 - \phi_0 B)(1 - B)^{d_0} y_t = \varepsilon_{0t} - \psi_0 \varepsilon_{t-1},\]

where \(\phi_0 = 1, |d_0| < \frac{1}{2}, |\psi_0| < 1,\) and \(\varepsilon_{0t} \sim iid N(0,1)\). We choose the sample size \(n = 400\), with 1,000 replications, and take \(d_0 = \pm 0.3\) and \(\psi_0 = \pm 0.5\),
<table>
<thead>
<tr>
<th>$d_0$</th>
<th>$\phi_0$</th>
<th>$\rho$</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>1.0</th>
<th>$\tau$</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>1.0</td>
<td>0.992</td>
<td>0.898</td>
<td>0.624</td>
<td>0.151</td>
<td>0.066</td>
<td></td>
<td>1.000</td>
<td>0.985</td>
<td>0.840</td>
<td>0.164</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.992</td>
<td>0.898</td>
<td>0.627</td>
<td>0.156</td>
<td>0.067</td>
<td></td>
<td>0.999</td>
<td>0.986</td>
<td>0.846</td>
<td>0.160</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>1.000</td>
<td>0.964</td>
<td>0.688</td>
<td>0.176</td>
<td>0.067</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.888</td>
<td>0.221</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.959</td>
<td>0.688</td>
<td>0.182</td>
<td>0.065</td>
<td></td>
<td>1.000</td>
<td>0.999</td>
<td>0.890</td>
<td>0.216</td>
<td>0.067</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.994</td>
<td>0.926</td>
<td>0.682</td>
<td>0.171</td>
<td>0.073</td>
<td></td>
<td>1.000</td>
<td>0.986</td>
<td>0.879</td>
<td>0.222</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.993</td>
<td>0.927</td>
<td>0.682</td>
<td>0.171</td>
<td>0.073</td>
<td></td>
<td>1.000</td>
<td>0.988</td>
<td>0.874</td>
<td>0.223</td>
<td>0.062</td>
<td></td>
</tr>
<tr>
<td>−0.1</td>
<td>1.0</td>
<td>0.993</td>
<td>0.930</td>
<td>0.689</td>
<td>0.162</td>
<td>0.070</td>
<td></td>
<td>1.000</td>
<td>0.986</td>
<td>0.873</td>
<td>0.220</td>
<td>0.062</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.992</td>
<td>0.929</td>
<td>0.688</td>
<td>0.164</td>
<td>0.069</td>
<td></td>
<td>0.999</td>
<td>0.986</td>
<td>0.870</td>
<td>0.215</td>
<td>0.060</td>
<td></td>
</tr>
<tr>
<td>−0.3</td>
<td>1.0</td>
<td>0.999</td>
<td>0.917</td>
<td>0.639</td>
<td>0.142</td>
<td>0.063</td>
<td></td>
<td>1.000</td>
<td>0.981</td>
<td>0.842</td>
<td>0.180</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.998</td>
<td>0.919</td>
<td>0.634</td>
<td>0.141</td>
<td>0.063</td>
<td></td>
<td>1.000</td>
<td>0.982</td>
<td>0.843</td>
<td>0.175</td>
<td>0.054</td>
<td></td>
</tr>
</tbody>
</table>

Note: Replications = 1,000.
Table 3. The empirical mean and standard deviation of CSS estimators for the nonstationary FARIMA$(1, d_0, 1)$ Model

<table>
<thead>
<tr>
<th>$d_0$</th>
<th>$\phi_0$</th>
<th>$\psi_0$</th>
<th>$\hat{d}$</th>
<th>$\hat{\phi}$</th>
<th>$\hat{\psi}$</th>
<th>$\sqrt{\text{MSE}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>1.0</td>
<td>0.5</td>
<td>0.290</td>
<td>0.985</td>
<td>0.484</td>
<td>0.064</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.0</td>
<td>0.5</td>
<td>-0.311</td>
<td>0.995</td>
<td>0.482</td>
<td>0.060</td>
</tr>
<tr>
<td>0.3</td>
<td>1.0</td>
<td>0.3</td>
<td>0.291</td>
<td>0.985</td>
<td>0.286</td>
<td>0.060</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.0</td>
<td>0.3</td>
<td>-0.308</td>
<td>0.995</td>
<td>0.283</td>
<td>0.053</td>
</tr>
<tr>
<td>0.3</td>
<td>1.0</td>
<td>-0.3</td>
<td>0.297</td>
<td>0.984</td>
<td>-0.310</td>
<td>0.050</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.0</td>
<td>-0.3</td>
<td>-0.300</td>
<td>0.995</td>
<td>-0.309</td>
<td>0.043</td>
</tr>
<tr>
<td>0.3</td>
<td>1.0</td>
<td>-0.5</td>
<td>0.299</td>
<td>0.984</td>
<td>-0.507</td>
<td>0.048</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.0</td>
<td>-0.5</td>
<td>-0.299</td>
<td>0.994</td>
<td>-0.507</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Note: Sample size $n = 400$ and replications $= 1,000$.

$\pm 0.3$. Various combinations of values $d_0$ and $\psi_0$ are used. The means and the standard errors of the estimators of $d_0$, $\phi_0$, and $\psi_0$ are reported in Table 3. Again, all biases and empirical standard errors of the estimators are almost constant for different $d_0$.

From Example 2.2, the statistics for testing the unit root in the FARIMA$(1, d_0, 1)$ model are

$$
\rho = n(\hat{\phi} - 1) \quad \text{and} \quad \tau = \left(\frac{\sum_{t=2}^{n} u_{t-1}^2}{\sum_{t=2}^{n} u_{t-1}^2} \right)^{1/2} (\hat{\phi} - 1),
$$

where $u_{t-1} = (\partial e_t / \partial \phi)_{(\phi, \psi, d)} = (\hat{\phi}, \hat{\psi}, \hat{d})$. Here, $\rho$ and $\tau$ have the same asymptotic distributions as those in (2.8) and (2.9). To examine the power and size of $\rho$ and $\tau$, we take $\phi_0 = 0.8$, 0.9, 0.95, 1.0 with $(\psi_0, d_0)$ having the same values as in Table 3. Again, we choose the sample size $n = 400$ with 1,000 replications. The critical values given in Tables 8.5.2 and 8.5.3 of Fuller (1976) are used. The powers and sizes based on the 5% significance level are reported in Table 4. All tests have quite satisfactory powers, and the sizes are also close to the nominal 0.05, except the case $d_0 = -0.3$ with $\psi_0 = -0.3$ and $-0.5$, which is a bit sensitive. To assess the large sample behavior, we perform a simulation for the case $d_0 = -0.3$ and $\psi_0 = -0.5$ with sample size $n = 1,000$ and 1,000 replications, in which case the sizes of $\rho$ and $\tau$ are 0.053 and 0.054, respectively. This means that we can obtain almost the exact sizes of $\rho$ and $\tau$ when the sample size becomes large. These results further indicate that the theory and method developed in this paper are potentially useful for the nonstationary FARIMA model.
Table 4. Power and size of lower tail Dickey–Fuller test based on the 5% significance level for the nonstationary FARIMA(1, \(d_0, 1\)) model

<table>
<thead>
<tr>
<th>(\psi_0)</th>
<th>(d_0 = 0.3)</th>
<th>(d_0 = -0.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>0.5</td>
<td>(\rho)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(\tau)</td>
<td>1.000</td>
</tr>
<tr>
<td>0.3</td>
<td>(\rho)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(\tau)</td>
<td>1.000</td>
</tr>
<tr>
<td>-0.3</td>
<td>(\rho)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(\tau)</td>
<td>1.000</td>
</tr>
<tr>
<td>-0.5</td>
<td>(\rho)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(\tau)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Sample size \(n = 400\) and replications = 1,000.
4. THE PROOF OF MAIN RESULTS

We begin by introducing an invariance principle, i.e., Theorem 4.1. This theorem serves a purpose similar to that of Theorem 2.2 in Chan and Wei (1988) and will be used to establish Lemma 4.6. In Theorem 4.1, the elements of $S_t$ will be basic processes corresponding to the componentwise argument in Theorem 2.1.

**THEOREM 4.1.** Suppose that $\{z_t\}$ is generated by

$$\phi_0^*(B)z_t = \psi_0(B)\epsilon_{0t}$$  \hspace{1cm} (4.1)

with $z_t = 0$ as $t \leq 0$, where $\phi_0^*(B) = 1 - \sum_{i=1}^{p^*} \phi_{0i}^* B^i$ and $\psi_0(B)$ and $\epsilon_{0t}$ are defined as in model (1.1). Denote $Z_{t-1} = -(\partial \epsilon_{0t}/\partial \phi^*, \partial \epsilon_{0t}/\partial \psi, \partial \epsilon_{0t}/\partial d)'$, where $\phi_0^* = (\phi_{01}, \ldots, \phi_{0p^*})'$, $\psi_0 = (\psi_01, \ldots, \psi_0q)'$, $\partial \epsilon_{0t}/\partial \phi^* = \partial \epsilon_{0t}/\partial \phi^*(|_{\phi_0^*, \psi_0, d_0})$, and similarly define $\partial \epsilon_{0t}/\partial \psi$ and $\partial \epsilon_{0t}/\partial d$. Let

$$S_t = (\epsilon_{0t}, (-1)^i \epsilon_{0t}, \sqrt{2} \sin t \theta_i \epsilon_{0t}, \sqrt{2} \cos t \theta_i \epsilon_{0t}, \ldots, \sqrt{2} \cos t \theta_i Z_{t-1} \epsilon_{0t})',$$

where $\theta_i \neq \theta_j$ if $i \neq j$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[nr]} S_t \xrightarrow{\mathcal{L}} (W'(\tau), N'(\tau))' \hspace{1cm} \text{in} \hspace{0.2cm} D^{2(l+1)+p^*+q+1},$$  \hspace{1cm} (4.2)

where $W(\tau)$ is a $(2l+2)$-dimensional standard Brownian motion and $N(\tau)$ is a $(p^*+q+1)$-dimensional Brownian motion, which is independent of $W(\tau)$ and has mean zero and covariance $\tau \Sigma$ defined as in Theorem 2.1.

The random elements in $S_t$ are the same as those in Theorem 2.2 of Chan and Wei (1988), except for $Z_{t-1} \epsilon_t$. We need some special arguments because of this different random element. The following two lemmas are essential for the proof of Theorem 4.1. The first ensures that $W(\tau)$ and $N(\tau)$ are asymptotically independent, and the second implies that the Lindeberg condition given in Helland (1982) is satisfied. The proofs of the two lemmas can be found in Appendix C, which includes some arguments different from those for Theorem 2.2 in Chan and Wei (1988).

**LEMMA 4.1.** $n^{-1} \sum_{k=2}^{[nr]} e^{ik\theta} E(\epsilon_{0k}^2 Z_{k-1} F_k) = o_p(1)$, where $i = \sqrt{-1}$, $\tau \in [0,1]$, and $F_t = \sigma\{\epsilon_{01}, \ldots, \epsilon_{0t}\}$, the $\sigma$-field generated by $\epsilon_{01}, \ldots, \epsilon_{0t}$.

**LEMMA 4.2.** $n^{-1} \sum_{k=2}^{[nr]} E[\|Z_k \epsilon_{0k+1}\|^2 I(\|Z_k \epsilon_{0k+1}\| > \sqrt{n} \delta |F_k|)] = o_p(1)$.

Proof of Theorem 4.1. First,

$$\frac{1}{n} \sum_{i=1}^{n} E \left( \frac{\partial \epsilon_{0t}}{\partial d} - Z_0^0 \right)^2 = \frac{1}{n} \sum_{i=1}^{n} E \left( \sum_{i=1}^{\infty} \frac{1}{i} \epsilon_{0t-i} \right)^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{i=1}^{\infty} \frac{1}{i^2} = o(1).$$
Similarly, we can show that \( n^{-1} \sum_{t=1}^{n} E[\|\partial e_{0t}/\partial \phi^* - Z_0^0\|^2] = o(1) \) and \( n^{-1} \sum_{t=1}^{n} E[\|\partial e_{0t}/\partial \psi - Z_0^0\|^2] = o(1) \). Thus,

\[
\frac{1}{n} \sum_{t=1}^{n} \|Z_t - Z_0^0\|^2 = o_p(1). \tag{4.3}
\]

Because \( Z_0^0 \) is strictly stationary and ergodic with \( E(Z_0^0 Z_0^0^*) < \infty \), by the ergodic theorem, \( n^{-1} \sum_{t=1}^{n} Z_t Z_t^* = \Sigma + \tau \Sigma + o_p(1) \). Furthermore, by (4.3), we can show that \( n^{-1} \sum_{t=1}^{n} (Z_t, Z_t^* - Z_0^0 Z_0^0^*) = o_p(1) \) and hence

\[
\frac{1}{n} \sum_{t=2}^{n} Z_{t-1} Z_{t-1}^* = \tau \Sigma + o_p(1). \tag{4.4}
\]

For \( \theta, \delta \in [0, 2\pi] \), observe that

\[
\frac{1}{n} \sum_{k=2}^{n} \cos k\theta \sin k\delta \to 0, \quad \text{and if} \quad \theta \neq \delta,
\]

\[
\frac{1}{n} \sum_{k=2}^{n} \cos k\theta \cos k\delta = \frac{1}{n} \sum_{k=2}^{n} \sin k\theta \sin k\delta \to 0.
\]

Now, by Lemmas 4.1 and 4.2, (4.4), and applying the standard functional central limit theorem (Helland, 1982, Theorem 3.3), we can complete the proof.

Let \( x_t = \psi_0^{-1}(B)(1 - B)^{d_0} y_t \). Because \( y_t \) are generated by model (1.1) through DGM2, \( \{x_t\} \) satisfies the unstable autoregressive model

\[
\phi_0(B)x_t = e_{0t}, \tag{4.5}
\]

with \( x_t = 0 \) as \( t \leq 0 \). Following Chan and Wei (1988), we may transform \( \{x_t\} \) into various componentwise arguments corresponding to the locations of their roots. Let \( u_t = (1 - B)^{-a} \phi_0(B)x_t \), \( v_t = (1 + B)^{-b} \phi_0(B)x_t \), \( z_t = \phi_0^{-1}(B) \phi_0(B)x_t \), and \( x_{t,k} = (1 - 2 \cos \theta_k B + B^2)^{-d_k} \phi_0(B)x_t \), \( k = 1, \ldots, l \). Then

\[
(1 - B)^a u_t = e_{0t}, \quad (1 + B)^b v_t = e_{0t}, \quad \phi_0^* (B) z_t = e_{0t},
\]

\[
(1 - 2 \cos \theta_k B + B^2)^{d_k} x_{t,k} = e_{0t}, \quad k = 1, \ldots, l,
\]

where \( a, b, d_k \), and \( \phi_0^* (B) \) are defined as in (1.3). Define

\[
u_t = (u_t, \ldots, u_{t-a+1})', \quad v_t = (v_t, \ldots, v_{t-b+1})', \quad z_t = (z_t, \ldots, z_{t-p^*+1})', \quad x_{t,k} = (x_{t,k}, \ldots, x_{t-d_k+1,k})', \quad k = 1, \ldots, l.
\]
As shown in Chan and Wei (1988), there exists a nonsingular matrix $Q^*$ such that

$$Q^* x_t = (u'_t, v'_t, x'_{t,1}, \ldots, x'_{t,l}, z'_t)' \quad (4.6)$$

where $x_t = (x_t, \ldots, x_{t-p+1})'$. Further define $u_t(j + 1) = (1 - B)^{a-j} u_t$ for $j = 0, 1, \ldots, a - 1$, $U_t = (u_t(1), \ldots, u_t(a))'$, $v_t(j + 1) = (1 - B)^{b-j} v_t$ for $j = 0, 1, \ldots, b - 1$, $V_t = (v_t(1), \ldots, v_t(b))'$, $x_{t,k}(j + 1) = (1 - 2B \cos \theta_k + B^2)_{d-k} x_{t,k}$ for $k = 1, \ldots, l$, $j = 0, 1, \ldots, d_k - 1$, $X_{t,k} = (x_{t,k}(1), x_{t-1,k}(1), \ldots, x_{t,k}(d_k), x_{t-1,k}(d_k))'$. There exist nonsingular matrices $M, \tilde{M}, C_k, k = 1, \ldots, l$, which can be found in Chan and Wei (1988), such that

$$M u_t = U_t, \quad \tilde{M} v_t = V_t, \quad C_k x_{t,k} = X_{t,k}, \quad k = 1, \ldots, l. \quad (4.7)$$

Denote $G^* = \text{diag}(M, \tilde{M}, C_1, \ldots, C_l, I_p^*) Q^*$. Then

$$G^* x_t = (U'_t, V'_t, X'_{t,1}, \ldots, X'_{t,l}, z'_t)' \quad (4.8)$$

Thus, $x_t$ has been decomposed into some subvectors corresponding to various unit roots and stationary components. Lemmas 4.3 and 4.4, which follow, come directly from Theorems 3.1.2, 3.2.1, and 3.3.4 of Chan and Wei (1988). In Lemma 4.3, (a)–(c) show the limiting distributions corresponding to various unit roots in the score function, whereas (d)–(f) show the limiting distributions corresponding to various unit roots in the information matrix. Lemma 4.4 shows that the cross product terms between various nonstationary and stationary components involving $\phi_0^*$ in the information matrix converge to zero in probability.

**LEMMA 4.3.**

(a) $N_n^{-1} \sum_{i=2}^{n} U_{i-1} e_{0i} \rightarrow_{\mathcal{L}} \xi$,

(b) $\tilde{N}_n^{-1} \sum_{i=2}^{n} V_{i-1} e_{0i} \rightarrow_{\mathcal{L}} \tilde{\xi}$,

(c) $L^{-1}_{k,n} \sum_{i=2}^{n} X_{i-1,k} e_{0i} \rightarrow_{\mathcal{L}} \xi_k$,

(d) $N_n^{-1} \sum_{i=2}^{n} U_{i-1} U_{i-1}^* N_n^{-1} \rightarrow_{\mathcal{L}} F$,

(e) $\tilde{N}_n^{-1} \sum_{i=2}^{n} V_{i-1} V_{i-1}^* \tilde{N}_n^{-1} \rightarrow_{\mathcal{L}} \tilde{F}$,

(f) $L^{-1}_{k,n} \sum_{i=2}^{n} X_{i-1,k} X_{i-1,k}^* L^{-1}_{k,n} \rightarrow_{\mathcal{L}} H_k$,
where \( k = 1, \ldots, l \), \( \xi, F, \tilde{F}, \tilde{\xi}_k \), and \( H_k \) are defined in Appendix A, and \( F, \tilde{F}, \) and \( H_k \) are positive definite matrices in probability.

**LEMMA 4.4.** The following random variables converge to zero in probability:

\[
N_n^{-1} \sum_{i=2}^{n} U_{t-1} V_{t-1}^{-1} \tilde{N}_n^{-1}, \quad N_n^{-1} \sum_{i=2}^{n} U_{t-1} X_{t-1, k}^{-1} L_{k,n}^{-1},
\]

\[
\tilde{N}_n^{-1} \sum_{i=2}^{n} V_{t-1} X_{t-1, k}^{-1} L_{k,n}^{-1}, \quad k = 1, \ldots, l,
\]

\[
L_{k,n}^{-1} \sum_{i=2}^{n} X_{t-1, k} X_{t-1, k_1}^{-1} L_{k_1,n}^{-1}, \quad k \neq k_1, \quad k, k_1 = 1, \ldots, l,
\]

\[
N_n^{-1} \sum_{i=2}^{n} U_{t-1} z_{t-1}' / \sqrt{n}, \quad \tilde{N}_n^{-1} \sum_{i=2}^{n} V_{t-1} z_{t-1}' / \sqrt{n},
\]

\[
L_{k,n}^{-1} \sum_{i=2}^{n} X_{t-1, k} z_{t-1}' / \sqrt{n}, \quad k = 1, \ldots, l.
\]

Now, we introduce two additional lemmas. Lemma 4.5 is for the proof of Lemma 4.6(b), which together with Theorem 4.1 ensures that the estimators of various unit roots are asymptotically independent of the estimators of \( d_0 \) and \( \psi_0 \). Lemma 4.6 shows the asymptotic distributions of the score function and the information matrix.

**LEMMA 4.5.** Suppose that \( s_t \) is one of the random vectors \( N_n^{-1} U_t \), \( \tilde{N}_n^{-1} V_t \), and \( L_{k,n}^{-1} X_{t, k} \), \( k = 1, \ldots, l \), then

\[
(a) \quad n^{-1/2} \sum_{i=2}^{n} s_{t-1} \left( \sum_{i=1}^{t-1} \frac{1}{i} e_{0t-i} \right) + n^{-1/2} \sum_{i=2}^{n} \left( \sum_{i=1}^{t-1} s_{t-1} \right) e_{0t} = o_p(1),
\]

\[
(b) \quad n^{-1/2} \sum_{i=2}^{n} s_{t-1} \left( \sum_{i=1}^{t-1} v_{\psi_0}(i) e_{0t-i} \right) + n^{-1/2} \sum_{i=2}^{n} \left( \sum_{i=1}^{t-1} v_{\psi_0}(i) s_{t-1} \right) e_{0t} = o_p(1),
\]

where \( \psi_0^{-1}(B) = \sum_{i=1}^{\infty} v_{\psi_0}(i) B^i \).

Proof. We present the proof only for \( s_t = N_n^{-1} U_t \). Other cases are similar. For (a), denote \( \eta_t = \sum_{i=1}^{t-1} (1/i) e_{0t-i} \) and \( \mu_t = \sum_{i=1}^{\infty} (1/i) e_{0t-i} \). We first note that

\[
E \left( \sum_{i=1}^{t-1} \eta_t \right)^2 = E \left( \sum_{i=1}^{t-1} \mu_t - \sum_{i=1}^{t-1} \sum_{k=1}^{\infty} \frac{1}{k} e_{0t-k} \right)^2 \\
\leq 2E \left( \sum_{i=1}^{t-1} \mu_t \right)^2 + 2E \left( \sum_{i=1}^{t-1} \sum_{k=1}^{\infty} \frac{1}{k} e_{0t-k} \right)^2.
\]

(4.9)
Furthermore, we have

\[ E\left( \sum_{i=1}^{i} \mu_{i} \right)^{2} = iE\mu_{i}^{2} + 2 \sum_{r=1}^{i-1} (i - r)E\mu_{i} \mu_{i-r} \]

\[ = iE\mu_{i}^{2} + 2 \sum_{r=1}^{i-1} (i - r) \left( \sum_{k=1}^{\infty} \frac{1}{k(k+r)} \right) \]

\[ \leq iE\mu_{i}^{2} + \sum_{r=1}^{i-1} (i - r) r^{-1/2} \left( \sum_{k=1}^{\infty} \frac{1}{k^{3/2}} \right) \]

\[ = iE\mu_{i}^{2} + O \left( \sum_{r=1}^{i-1} (i - r) r^{-1/2} \right) = o(i^{3/2+1/8}). \tag{4.10} \]

By Minkowski’s inequality,

\[ E\left( \sum_{i=1}^{i} \sum_{k=r}^{\infty} \frac{1}{k} \varepsilon_{i+k} \right)^{2} \leq \left[ \sum_{i=1}^{i} E\left( \sum_{k=r}^{\infty} \frac{1}{k} \varepsilon_{i+k} \right)^{2} \right]^{2} \]

\[ = \left( \sum_{i=1}^{i} \sum_{k=r}^{\infty} \frac{1}{k} \right)^{2} = O \left( \left[ \sum_{i=1}^{i} t^{-(1-\alpha)/2} \right]^{2} \right) \]

\[ = o(i^{3/2}), \tag{4.11} \]

where \( 0 < \alpha < \frac{1}{2} \). By (4.9)–(4.11), we know that

\[ E\left( \sum_{i=1}^{i} \eta_{i} \right)^{2} = o(i^{3/2+1/8}). \tag{4.12} \]

Now we consider the \( j \)th element of the first term in (a), i.e., \( n^{-1/2-j} \sum_{i=2}^{n} u_{i-1}(j) \eta_{i} \). By the definition of \( u_{i}(j) \), we have

\[ u_{i}(j) = \sum_{i=1}^{i} u_{i}(j - 1), \tag{4.13} \]

and hence

\[ \sum_{i=2}^{n} u_{i-1}(j) \eta_{i} = \left[ \sum_{i=2}^{n} u_{i-1}(j - 1) \right] \left( \sum_{i=1}^{n} \eta_{i} \right) - \sum_{i=2}^{n} \left( \sum_{i=1}^{i-1} \eta_{i} \right) u_{i-1}(j - 1) \]

\[ = I_{1n} - I_{2n}, \quad \text{say.} \tag{4.14} \]

By (4.12), \( n^{-1} \sum_{i=1}^{n} \eta_{i} = o_{p}(n^{-1+3/4+1/16}) = o_{p}(1) \), and by Lemma 4.3(d), \( |n^{-1+1} \sum_{i=2}^{n} u_{i-1}(j - 1)| \leq [n^{-2(j-1)+1} \sum_{i=2}^{n} u_{i-1}(j - 1)]^{1/2} = O_{p}(n^{1/2}) \). Thus, we have

\[ n^{-j-1/2} I_{1n} = [n^{-j-1/2} u_{n}(j)] \left( n^{-1} \sum_{i=1}^{n} \eta_{i} \right) = o_{p}(1). \tag{4.15} \]
When $j = 1$, by (4.12),
\[
EI_{2n}^2 = \sum_{t=2}^{n} E \left( \sum_{i=1}^{t-1} \eta_i \right)^2 = \sum_{t=2}^{n} o(t^{3/2+1/8}) = o(n^{5/2+1/8}).
\] (4.16)

Thus, $E(n^{-3/2}I_{2n})^2 = o(n^{-1/2+1/8}) = o(1)$, and hence $n^{-3/2}I_{2n} = o_p(1)$. When $j > 1$,
\[
E|I_{2n}| \leq \sum_{t=2}^{n} E \left( \sum_{i=1}^{t-1} \eta_i \right) u_{t-1}(j - 1) \\
\leq n^{j-3/2} \sum_{t=2}^{n} \left\{ E \left( \sum_{i=1}^{t-1} \eta_i \right)^2 E \left[ n^-(j-2)-1/2 u_t(j - 1) \right]^2 \right\}^{1/2} \\
= o \left( n^{j-3/2} \sum_{t=2}^{n} t^{3/4+1/16} \right) = o(n^{j+1/4+1/16}),
\] (4.17)

where the first equation holds by (2.17) in Ling (1998) (i.e., $Eu_t^2(j) = O(t^{2(j-1)+1})$) and (4.12). Hence $n^{-j-1/2}I_{2n} = o_p(1)$. Further by (4.14) and (4.15), we can claim
\[
n^{-1/2}N^{-1} \sum_{t=2}^{n} U_{t-1} \eta_t = o_p(1).
\] (4.18)

Next, we consider the $j$th element of the second term in (a), that is, $n^{-j-1/2} \sum_{t=2}^{n} (\sum_{i=1}^{t-1} \frac{1}{i} u_{t-i}(j)) \varepsilon_{0t}$. By Minkowski’s inequality and (2.17) in Ling (1998), we have
\[
E \left[ n^{-j-1/2} \sum_{t=2}^{n} \left( \sum_{i=1}^{t-1} \frac{1}{i} u_{t-i}(j) \right) \varepsilon_{0t} \right]^2 \\
= n^{-2j-1} \sum_{t=2}^{n} E \left( \sum_{i=1}^{t-1} \frac{1}{i} u_{t-i}(j) \right)^2 \\
\leq n^{-2j-1} \sum_{t=2}^{n} \left( \sum_{i=1}^{t-1} \frac{1}{i} \sqrt{Eu_{t-i}^2(j)} \right)^2 = O \left( n^{-2j-1} \sum_{t=2}^{n} \left( \sum_{i=1}^{t-1} \frac{n^{j-1/2}}{i} \right)^2 \right) \\
= O \left( n^{-2} \sum_{t=2}^{n} \left( \sum_{i=1}^{t-1} \frac{1}{i} \right)^2 \right) = o(1).
\]

Thus
\[
n^{-1/2}N^{-1} \sum_{t=2}^{n} \left( \sum_{i=1}^{t-1} \frac{1}{i} U_{t-i} \right) \varepsilon_{0t} = o_p(1).
\] (4.19)

By (4.18) and (4.19), (a) holds. For (b), because $v_{\varphi_2}(i) = O(\rho^i)$ with $0 \leq \rho < 1$, it can be more easily proved, and hence the details are omitted. This completes the proof.
LEMMA 4.6.

(a) \( G_n D_n \to_L -2(\xi', \tilde{\xi}', \zeta_1', \ldots, \zeta_l', N')', \)

(b) \( G_n I_n G_n' \to_L 2 \text{ diag}(F, \bar{F}, H_1, \ldots, H_l, \Sigma), \)

where \( \xi, F, \tilde{\xi}, \xi', \) and \( H_i, i = 1, \ldots, l \) are defined in Appendix A, \( N \) is a normal random vector with mean 0 and covariance \( \Sigma \) defined as in Theorem 2.1, and \( (F, \xi), (\tilde{F}, \xi'), (H_1, \xi_1), \ldots, (H_l, \xi_l), \) and \( N \) are independent.

Proof. For (a), note that

\[
G_n D_n = -2 \sum_{t=2}^{n} \left( (N_n^{-1} U_t)', (\tilde{N}_n^{-1} V_t)', (L_1^{-1} X_{1,n})', \ldots, (L_l^{-1} X_{l,n})', \frac{1}{\sqrt{n}} Z_{t-1} Z_{t-1}' \right),
\]

where \( Z_{t-1} \) is defined as in Theorem 4.1. By Lemma 4.3(a)–(c) and Theorem 4.1, (a) holds.

For (b), by the definition of \( \mathbf{z}_t \), its element \( z_t \) satisfies

\[
\phi_0^*(B)z_t = \psi_0(B)\varepsilon_{0t},
\]

with \( z_t = 0 \) as \( t \leq 0 \), i.e., model (4.1). By (4.4),

\[
\frac{1}{n} \sum_{t=2}^{n} \frac{\partial \varepsilon_{0t}}{\partial \phi^*} \frac{\partial \varepsilon_{0t}}{\partial \phi^{*'}} = E\left(Z_{\phi^*t}^0 Z_{\phi^{*'}}^0\right) + o_p(1) \tag{4.21}
\]

\[
\frac{1}{n} \frac{\partial^2 S(\lambda_0)}{\partial \tilde{\psi} \partial \tilde{\psi}'} = 2 \sum_{t=2}^{n} \frac{\partial \varepsilon_{0t}}{\partial \tilde{\psi}} \frac{\partial \varepsilon_{0t}}{\partial \tilde{\psi}'}, + o_p(1)
\]

\[
= 2E\left(\begin{array}{cc}
Z_{\psi^*t}^0 Z_{\psi^{*'}}^0 & Z_{\psi^*t}^0 Z_{\psi^{*'}}^0 \\
Z_{\psi^*t}^0 Z_{\psi^{*'}}^0 & \frac{\pi^2}{6}
\end{array}\right) + o_p(1), \tag{4.22}
\]

where \( \tilde{\psi} = (\psi_1, \ldots, \psi_q, d)' \). By Lemma 4.4, we have

\[
J_n^{-1} G_n^* \left[ \frac{\partial^2 S(\lambda_0)}{\partial \phi \partial \phi'} \right] G_n^* J_n^{-1} = 2 \sum_{t=2}^{n} \text{ diag} \left\{ N_n^{-1} U_t U_t' N_n^{-1}, \tilde{N}_n^{-1} V_t V_t' \tilde{N}_n^{-1}, L_1^{-1} X_{1,n} X_{1,n}' L_1^{-1}, \ldots, L_l^{-1} X_{l,n} X_{l,n}' L_l^{-1}, \frac{1}{n} \frac{\partial \varepsilon_{0t}}{\partial \phi^*} \frac{\partial \varepsilon_{0t}}{\partial \phi^{*'}} \right\}
\]

\[+ o_p(1). \tag{4.23} \]
By Lemma 4.5 and (4.4), we can show that
\[
J_n^{-1} G_n^* \frac{\partial^2 S(\lambda_0)}{\partial \phi \partial \phi'} = 2[0, E(Z_{\phi_1}^0, Z_{\phi_2}^0)]' + o_p(1), \tag{4.24}
\]
\[
J_n^{-1} G_n^* \frac{\partial^2 S(\lambda_0)}{\partial \phi \partial \psi'} = 2[0, E(Z_{\phi_1}^0, Z_{\psi_1}^0)]' + o_p(1), \tag{4.25}
\]
where \( i = 1, \ldots, q \). By (4.21)–(4.25) and Lemma 4.3(d)–(f), we know that (b) holds.

Note that the random vectors and matrices involved in (4.20) and (4.23) are functionals of the basic processes in Theorem 4.1. The independence of \((F, \xi), (\bar{F}, \bar{\xi}), (H_1, \xi_1), \ldots, (H_q, \xi_q)\), and \(N\) follows (4.23) and Theorem 4.1. This completes the proof.

The following is the final lemma. Its proof can be found in Appendix C. This lemma ensures that the remainder term of the Taylor expansion of \(S(\lambda)\) is small enough such that there is a local solution for \(\partial S(\lambda)/\partial \lambda\).

**Lemma 4.7.** When \(\|n^{\alpha - 1/2}G_n' - 1(\lambda - \lambda_0)\| < 1\),
\[
G_n[I_n(\lambda) - I_n]G_n' = O_p\left(\frac{1}{\sqrt{n}} \|G_n' - 1(\lambda - \lambda_0)\|\right),
\]
where \(0 < \alpha < \frac{1}{2}\).

Proof of Theorem 2.1. By Taylor’s expansion,
\[
\frac{\partial S(\lambda)}{\partial \lambda} = D_n + I_n \times (\lambda - \lambda_0) + [I_n(\lambda^*) - I_n](\lambda - \lambda_0), \tag{4.26}
\]
where \(\lambda^* = \lambda_0 + v(\lambda - \lambda_0)\) with \(v = v(n, \lambda)\) satisfying \(|v| \leq 1\). Multiplying \(n^{2\alpha - 1}(\lambda - \lambda_0)'\) to (4.26), we have
\[
n^{2\alpha - 1}(\lambda - \lambda_0)' \frac{\partial S(\lambda)}{\partial \lambda}
= [n^{\alpha - 1/2}G_n' - 1(\lambda - \lambda_0)]' [n^{\alpha - 1/2}G_n D_n]
+ [n^{\alpha - 1/2}G_n' - 1(\lambda - \lambda_0)]' [G_n I_n G_n' + R_n(\lambda)] [n^{\alpha - 1/2}G_n' - 1(\lambda - \lambda_0)], \tag{4.27}
\]
where \(R_n(\lambda) = G_n[I_n(\lambda^*) - I_n]G_n'\). Denote the last term by \(\Pi\).

Let \(\nu\) and \(\varepsilon\) be two given and sufficiently small positive numbers and \(V_n(\varepsilon) = \{\lambda : \|n^{\alpha - 1/2}G_n' - 1(\lambda - \lambda_0)\| = \varepsilon\}\).

Note that \(\Sigma > 0\) (see Li and McLeod, 1986). By Lemma 4.3, the matrix diag \((F, \bar{F}, H_1, \ldots, H_q, \Sigma)\) is positive definite in probability. By Lemma 4.6(b), there are a constant \(c_1\) and an integer \(N_1\) such that, as \(n > N_1\),
\[
P\{G_n I_n G_n' < -c_1 I_{m \times m}\} > 1 - \nu,
\]
where \( m = p + q + 1 \) and \( I_{m \times m} \) is the \( m \times m \) identity matrix. By Lemma 4.7, there exists a constant \( c \) such that, for small enough \( \varepsilon \), as \( n > N_1 \) and \( \lambda \in V_n(\varepsilon) \),

\[
P\{ |G_n I_n G_n' + R_n(\lambda)| < -c I_{m \times m} \} \geq 1 - \nu.
\]

Hence as \( n > N_1 \) and \( \lambda \in V_n(\varepsilon) \),

\[
P\{ \Pi < -c \varepsilon^2 \} > 1 - \nu.
\]

By Lemma 4.6(a), we know that \( G_n D_n = O_p(1) \). Hence there exists an integer \( N_2 \) such that as \( n > N_2 \),

\[
P\left\{ n^{a-1/2} \| G_n D_n \| < \frac{c}{2} \varepsilon \right\} > 1 - \nu.
\]

By (4.29), as \( n > N_2 \) and \( \lambda \in V_n(\varepsilon) \),

\[
P\left\{ \left[ n^{a-1/2} G_n^{-1}(\lambda - \lambda_0) \right]' \left[ n^{a-1/2} G_n D_n \right] < \frac{c}{2} \varepsilon^2 \right\} > 1 - \nu.
\]

Thus by (4.27), (4.28), and (4.30), as \( n > \max\{N_1, N_2\} \) and \( \lambda \in V_n(\varepsilon) \), with at least probability \( 1 - \nu \),

\[
n^{a-1}(\lambda - \lambda_0)' \frac{\partial S(\lambda)}{\partial \lambda} < -c \varepsilon^2 + \frac{c}{2} \varepsilon^2 < 0.
\]

Let \( Y = n^{a-1/2} G_n^{-1}(\lambda - \lambda_0)/\varepsilon \) and \( g(\lambda) = G_n \partial S(\lambda)/\partial \lambda \). Then by (4.31),

\[
\| Y \| = 1 \quad \text{and} \quad Y' g(n^{1/2-\alpha} \varepsilon G_n' Y + \lambda_0) < 0.
\]

Because \( \partial S(\lambda)/\partial \lambda \) is continuous and hence \( g \) is also continuous on \( Y \), by the fixed point theorem (Aitchison and Silvey, 1958) there is a solution \( \hat{\lambda}_n \) satisfying \( g(n^{1/2-\alpha} \varepsilon G_n' Y + \lambda_0) = 0 \), i.e., \( \partial S(\hat{\lambda}_n)/\partial \lambda = 0 \), and \( n^{a-1/2} G_n^{-1}(\hat{\lambda}_n - \lambda_0) < \varepsilon \). Consequently, the proof of part (a) is completed.

For such a sequence of \( \hat{\lambda}_n \), by (4.26) and Lemma 4.7 we have

\[
G_n^{-1}(\hat{\lambda}_n - \lambda_0) = - \left[ G_n I_n G_n' + O_p \left( \left\| \frac{1}{\sqrt{n}} G_n^{-1}(\hat{\lambda}_n - \lambda_0) \right\| \right) \right]^{-1} (G_n D_n).
\]

By part (a) of this theorem, \( n^{a-1/2} G_n^{-1}(\hat{\lambda}_n - \lambda_0) \) converges to zero in probability. By Theorem 2.3 of Chan and Wei (1988) and Lemma 4.1, all random variables in \( G_n D_n \) and \( G_n I_n G_n' \) converge jointly. Again by Lemma 4.6 and (4.32), we complete the proof of part (b).

\[\text{\textit{NOTE}}\]

1. Denote \( x_t^* = (1 - B)^d y_t \). Then \( \phi_0(B)x_t^* = \phi_0(B)e_t \). The term \( x_t^* \) also has two different DGMs. The first one, namely, DGM1\(^*\), is to generate \( u_t = \phi_0(B)e_t \) and then generate \( x_t^* \) through \( \phi_0(B)x_t^* = u_t \) with \( x_t^* = 0 \) as \( t \leq 0 \). The second one, namely DGM2\(^*\), is to generate \( \phi_0(B)u_t = e_t \) with \( u_t = 0 \) as \( t \leq 0 \) and then generate \( x_t^* \) through \( x_t^* = \phi_0(B)u_t \). Under DGM2\(^*\), (4.5) holds for \( x_t \), and under DGM1\(^*\), (4.5) does not hold. However, in the latter case, using Theorems 4.1–4.3 in
Ling and Li (1998) with \( h \) being a constant, we can show that the CSS estimators have the same asymptotic distributions as those in this paper. This means that the effect of the DGM1* and DGM2* in the nonstationary ARMA models is asymptotically ignorable for the CSS approach. For simplicity, we use DGM2* for \( x_t^* \) in this paper.

REFERENCES


APPENDIX A: LIMITING DISTRIBUTIONS

\[ F_0(\tau) = W_1(\tau), \quad F_j(\tau) = \int_0^\tau F_{j-1}(s) \, ds, \quad j = 1, \ldots, a, \]

\[ F = (\sigma_{ij})_{a \times a}, \quad \sigma_{ij} = \int_0^\tau F_{i-1}(s) F_{j-1}(s) \, ds, \quad i, j = 1, \ldots, a, \]

\[ \xi = \left( \int_0^1 F_0(\tau) \, dW_1(\tau), \ldots, \int_0^1 F_{a-1}(\tau) \, dW_1(\tau) \right)'. \]

\[ \bar{F}_0(\tau) = W_2(\tau), \quad \bar{F}_j(\tau) = \int_0^\tau \bar{F}_{j-1}(s) \, ds, \quad j = 1, \ldots, b, \]

\[ \bar{F} = (\bar{\sigma}_{ij}), \quad \bar{\sigma}_{ij} = \int_0^\tau \bar{F}_{i-1}(s) \bar{F}_{j-1}(s) \, ds, \quad i, j = 1, \ldots, b, \]

\[ \bar{\xi} = -\left( \int_0^1 \bar{F}_0(\tau) \, dW_2(\tau), \ldots, \int_0^1 \bar{F}_{b-1}(\tau) \, dW_2(\tau) \right). \]

\[ \xi = (\xi_1, \ldots, \xi_{2d}), \quad H = (\sigma_{ij})_{2d \times 2d}, \]

\[ \xi_{2j-1}(\tau) = \frac{1}{2 \sin \theta} \left( \int_0^1 f_{j-1}(s) \, dW_2(s) - \int_0^1 g_{j-1}(s) \, dW_1(s) \right), \]

\[ \xi_{2j}(\tau) = \frac{1}{2 \sin \theta} \left\{ \cos \theta \left[ \int_0^1 f_{j-1}(s) \, dW_1(s) - \int_0^1 g_{j-1}(s) \, dW_1(s) \right] \right. \\
\left. - \sin \theta \left[ \int_0^1 f_{j-1}(s) \, dW_1(s) + \int_0^1 g_{j-1}(s) \, dW_2(s) \right] \right\}, \]

\[ \sigma_{2k-1,2j-1} = \sigma_{2k,2j} = \frac{1}{4 \sin^2 \theta} \left( \int_0^1 f_{k-1}(s) f_{j-1}(s) \, ds + \int_0^1 g_{k-1}(s) g_{j-1}(s) \, ds \right), \]
\[ \sigma_{2k-1,2j} = \sigma_{2j,2k-1} = \frac{1}{4 \sin^2 \theta} \left\{ \cos \theta \left[ \int_0^1 f_{j-1}(s)f_{j-1}(s)ds + \int_0^1 g_{j-1}(s)g_{j-1}(s)ds \right] \\ - \sin \theta \left[ \int_0^1 f_{j-1}(s)g_{j-1}(s)ds - \int_0^1 g_{j-1}(s)f_{j-1}(s)ds \right] \right\}, \]

\[ f_j(\tau) = \frac{1}{2 \sin \theta} \left( \sin \theta \int_0^\tau f_{j-1}(s)ds - \cos \theta \int_0^\tau g_{j-1}(s)ds \right), \]

\[ g_j(\tau) = \frac{1}{2 \sin \theta} \left( \cos \theta \int_0^\tau f_{j-1}(s)ds + \sin \theta \int_0^\tau g_{j-1}(s)ds \right), \]

\[ f_0(\tau) = W_1(\tau) \quad \text{and} \quad g_0 = W_2(\tau), \]

where \( W_i(\tau) \) is the \( i \)th element of \( W(\tau) \) defined in Theorem 4.1. Finally, define \( H_k \) and \( \xi_k \) as earlier with \( \theta \) replaced by \( \theta_k \) and \((W_1(\tau), W_2(\tau))\) replaced by \((W_{2k+2}(\tau), W_{2k+3}(\tau))\).

**APPENDIX B: THE PROOF OF (1.5)**

First, we note that

\[ 1 = (1 - z)^{d_0}(1 - z)^{-d_0} = \left( \sum_{i=0}^\infty a_{0,i} z^i \right) \left( \sum_{i=0}^\infty \tilde{a}_{0,i} z^i \right) = 1 + \sum_{k=1}^\infty \left( \sum_{i=0}^k a_{0,i} \tilde{a}_{0,k-i} \right) z^k. \]

Thus, we have

\[ \sum_{i=0}^k a_{0,i} \tilde{a}_{0,k-i} = 0, \quad k = 1, 2, \ldots. \quad (B.1) \]

Because \( \sum_{k=0}^{t-1} a_{0,k} y_{t-k} = u_t \) with \((1 - B)u_t = \epsilon_{0t}, t = 1, 2, \ldots, \) and \( u_0 = 0 \), it follows that

\[ \sum_{k=0}^{t-1} a_{0,k} (y_{t-k} - y_{t-k-1}) = u_t - u_{t-1} = \epsilon_{0t}. \]

Let \( u_t^* = y_t - y_{t-1} \). Then \( y_t = y_{t-1} + u_t^* \), and

\[ \sum_{k=0}^{t-1} a_{0,k} u_{t-k}^* = \epsilon_{0t}. \quad (B.2) \]
From (B.2), it follows that

\[
\sum_{k=0}^{t-1} \tilde{a}_{0k} e_{0t-k} = \sum_{k=0}^{t-1} \tilde{a}_{0k} \left( \sum_{i=0}^{t-k-1} a_{0i} u_{t-k-i}^* \right) = u_*^t + \sum_{k=1}^{t-1} \left( \sum_{i=0}^{k} a_{0i} \tilde{a}_{0,k-i} \right) u_{t-k}^*. \tag{B.3}
\]

By (B.1), the second term in (B.3) is equal to zero, and hence (1.5) holds. This completes the proof.

\section*{APPENDIX C: THE REMAINING PROOFS}

\textbf{Proof of Lemma 4.1.}

\[
\frac{1}{n} \sum_{k=2}^{[n\tau]} e^{ik\theta} E(e^2_{0k} Z_{k-1} | \mathcal{F}_{k-1}) = -\frac{1}{n} \sum_{k=2}^{[n\tau]} e^{ik\theta} \left( \frac{\partial e_{0k}}{\partial \theta^* \theta}, \frac{\partial e_{0k}}{\partial \psi}, \frac{\partial e_{0k}}{\partial \psi^*} \right), \tag{C.1}
\]

where \(\theta \in [0,1]\). First, we show that

\[
-\frac{1}{n} \sum_{k=2}^{[n\tau]} e^{ik\theta} \frac{\partial e_{0k}}{\partial d} = -\frac{1}{n} \sum_{k=2}^{[n\tau]} e^{ik\theta} \left( \sum_{j=1}^{k-1} \frac{1}{j} e_{0k-j} \right) = o_p(1). \tag{C.2}
\]

Note that

\[
\frac{1}{n} \sum_{k=2}^{[n\tau]} e^{ik\theta} \left( \sum_{j=1}^{k-1} \frac{1}{j} e_{0k-j} \right) = \frac{1}{n} \sum_{j=1}^{[n\tau]-1} e^{i j \theta} \left( \sum_{k=1}^{[n\tau]-1} \frac{1}{j} e_{0k-j} \right). \]

By the triangle inequality, \(|\sum_{j=1}^{[n\tau]-1} e^{i j \theta} e_{0k-j}| \leq |\sum_{k=1}^{[n\tau]-1} \sin k\theta e_{0k}| + |\sum_{k=1}^{[n\tau]-1} \cos k\theta e_{0k}| + \max_{1 \leq j \leq n^{-1/2}} |\sum_{k=1}^{[n\tau]-1} \cos k\theta e_{0k}| + \max_{1 \leq j \leq n^{-1/2}} |\sum_{k=1}^{[n\tau]-1} \cos k\theta e_{0k}| \text{ converge to max}_{0 \leq \tau \leq 1} |B_{1k}(\tau)| + \max_{0 \leq \tau \leq 1} |B_{2k}(\tau)| \text{ in distribution, where} \ B_{1k}(\tau) \text{ and } B_{2k}(\tau) \text{ are Brownian motions. Thus}

\[
\frac{1}{n} \sum_{k=2}^{[n\tau]} e^{ik\theta} \left( \sum_{j=1}^{[n\tau]-1} \frac{1}{j} e_{0k-j} \right) = o_p \left( \frac{1}{n^{1/2-\alpha}} \sum_{j=1}^{[n\tau]-1} \frac{1}{j^{1+\alpha}} \right) = o_p(1),
\]

where \(0 < \alpha < \frac{1}{4}\), and the last equation holds because \(\sum_{j=1}^{\infty} j^{-1-\alpha} < \infty\).

Note that \(\partial e_{0k}/\partial \theta^* = - \sum_{i=1}^{k-1} c_{\theta}(k) e_{0t-i-k} \) and \(\partial e_{0k}/\partial \psi = - \sum_{i=1}^{k-1} c_{\psi}(k) e_{0t-i-k}\), with \(c_{\theta}(k) = \mathcal{O}(\rho^k)\) and \(c_{\psi}(k) = \mathcal{O}(\rho^k)\), where \(0 \leq \rho < 1\). Similar to the proof of (C.2), we have \(- (1/n) \sum_{k=2}^{[n\tau]} e^{ik\theta} (\partial e_{0k}/\partial \theta^*) = o_p(1)\) and \(- (1/n) \sum_{k=2}^{[n\tau]} e^{ik\theta} (\partial e_{0k}/\partial \psi) = o_p(1)\). This completes the proof.
Proof of Lemma 4.2. We first show that

\[
\frac{1}{n} \sum_{k=2}^{n} \mathbb{E} \left[ e_{0k+1} \left( \frac{\partial e_{0k}}{\partial d} \right)^2 \left( e_{0k+1} \left( \frac{\partial e_{0k}}{\partial d} \right) \right) > \sqrt{n} \delta \right] = o_p(1).
\]  
(C.3)

Note that \( \frac{\partial e_{0k}}{\partial d} \) and

\[
\frac{1}{n} \sum_{k=2}^{n} \left( \sum_{i=1}^{k-1} e_{0k-i} \right)^2 \leq \frac{2}{n} \sum_{k=2}^{n} Z_{0k}^2 + \frac{2}{n} \sum_{k=2}^{n} \left( \sum_{i=k}^{\infty} e_{0k-i} \right)^2
\]

= \( 2A_1 + 2A_2 \), say,  
(C.4)

where \( Z_{0k}^2 \) is defined before Theorem 2.1.

\[
E[|A_2|] = \frac{1}{n} \sum_{k=2}^{n} \mathbb{E} \left[ \frac{1}{i} \sum_{i=1}^{k} e_{0k-i} \right] \leq n \sum_{k=2}^{n} j^{-2}
\]

\[
\leq O \left( \frac{1}{n} \sum_{k=2}^{n} k^{-1+\alpha} \sum_{j=k}^{\infty} j^{-1-\alpha} \right) \leq O \left( \frac{1}{n} \sum_{k=2}^{n} k^{-1+\alpha} \right) = o(1),
\]  
(C.5)

where \( 0 < \alpha < 1 \). Because \( Z_{0k}^2 \) is strictly stationary and ergodic with \( EZ_{0k}^2 < \infty \), by the ergodic theorem, \( A_1 = n^{-1} \sum_{k=2}^{n} Z_{0k}^2 \) converges to \( \pi^2/6 \) almost surely (a.s.). Thus, by (C.4) and (C.5),

\[
\frac{1}{n} \sum_{k=2}^{n} \left( \sum_{i=1}^{k-1} e_{0k-i} \right)^2 = o_p(1).
\]  
(C.6)

Note that

\[
\left( \sum_{i=1}^{k-1} e_{0k-i} \right)^2 \leq 2Z_{0k}^2 + 2 \left( \sum_{i=k}^{\infty} e_{0k-i} \right)^2.
\]  
(C.7)

By (C.5), for any small \( \epsilon > 0 \),

\[
P \left( \frac{1}{n} \max_{2 \leq k \leq n} \left( \sum_{i=1}^{k} e_{0k-i} \right)^2 > \epsilon \right) \leq \frac{1}{n} \mathbb{E} \sum_{k=2}^{n} \left( \sum_{i=k}^{\infty} e_{0k-i} \right)^2 = o(1),
\]  

i.e., \( n^{-1} \max_{2 \leq k \leq n} (\sum_{i=k}^{\infty} e_{0k-i})^2 = o_p(1) \). Note that \( n^{-1/2} \max_{2 \leq k \leq n} |Z_{0k}| = o_p(1) \) (see Chung, 1968, p. 95). From (C.7), we have

\[
\frac{1}{n} \max_{2 \leq k \leq n} \left( \sum_{i=1}^{k-1} e_{0k-i} \right)^2 = o_p(1),
\]  
(C.8)

Now by the conditional Markov inequality, (C.6), and (C.8),

\[
\frac{1}{n} \sum_{k=2}^{n} \mathbb{E} \left[ \left( e_{0k} \sum_{i=1}^{k-1} e_{0k-i} \right)^2 \mathbb{E} \left( e_{0k} \sum_{i=1}^{k-1} e_{0k-i} \right) > \sqrt{n} \delta \right] = o_p(1),
\]

\[
\leq \frac{1}{n} \sum_{k=2}^{n} \left( \sum_{i=1}^{k-1} e_{0k-i} \right)^{2+\alpha} n^{-\alpha/2} K \delta^{-\alpha}
\]

\[
\leq \frac{1}{n} \sum_{k=2}^{n} \left( \sum_{i=1}^{k-1} e_{0k-i} \right)^{2} \left( \frac{1}{n} \max_{2 \leq k \leq n} \left| \sum_{i=1}^{k-1} e_{0k-i} \right|^2 \right)^{\alpha/2} K \delta^{-\alpha}
\]

\[
= o_p(1),
\]  
(C.9)
where $K$ is some constant. That is, (C.3) holds. Similarly we can show that
\[
\frac{1}{n} \sum_{k=2}^{[n/2]} E \left[ \epsilon_{0k+1}^2 \left( \frac{\partial \epsilon_{0k}}{\partial \phi^*} \right)^2 I \left( |\epsilon_{0k+1}| \left| \frac{\partial \epsilon_{0k}}{\partial \phi^*} \right| > \sqrt{n \delta} |F_k| \right) \right] = o_p(1),
\]
(C.10)

\[
\frac{1}{n} \sum_{k=2}^{[n/2]} E \left[ \epsilon_{0k+1}^2 \left( \frac{\partial \epsilon_{0k}}{\partial \psi} \right)^2 I \left( |\epsilon_{0k+1}| \left| \frac{\partial \epsilon_{0k}}{\partial \psi} \right| > \sqrt{n \delta} |F_k| \right) \right] = o_p(1).
\]
(C.11)

By (C.3), (C.10), and (C.11), we complete the proof.

**Proof of Lemma 4.7.** By a direct differentiation,
\[
\frac{\partial^3 S(\lambda)}{\partial d^3} = 2 \sum_{t=1}^n \left( 3 \frac{\partial^2 \epsilon_t}{\partial d^2} \frac{\partial \epsilon_t}{\partial d} + \epsilon_t \frac{\partial^3 \epsilon_t}{\partial d^3} \right),
\]
(C.12)

\[
\frac{\partial \epsilon_t}{\partial d} = \phi(B) \psi^{-1}(B) \log(1-B)(1-B)^d y_t,
\]
(C.13)

\[
\frac{\partial^2 \epsilon_t}{\partial d^2} = \phi(B) \psi^{-1}(B) \log^2(1-B)(1-B)^d y_t,
\]
(C.14)

\[
\frac{\partial^3 \epsilon_t}{\partial d^3} = \phi(B) \psi^{-1}(B) \log^3(1-B)(1-B)^d y_t.
\]
(C.15)

Note that
\[
\frac{\partial \epsilon_t}{\partial d} = (\phi - \phi_0') \psi^{-1}(B) \psi_0(B) \log(1-B)(1-B)^d x_t
\]
\[+ \log(1-B)(1-B)^d \epsilon_{0t} + [\psi^{-1}(B) - \psi_0^{-1}(B)] \log(1-B)(1-B)^d \epsilon_0(B) \epsilon_{0t},
\]
\[= I_{1t} + I_{2t} + I_{3t}, \text{ say},
\]
(C.16)

where $x_t$ is defined as in (4.6) with $x_t = \psi_0^{-1}(B)(1-B)^d y_t$. By the given condition, $|d - d_0| < n^{-\alpha}$ and $n^{-1/2} \tilde{G}_{n}(\phi - \phi_0) \leq n^{-\alpha}$, where $\tilde{G}_n$ is defined after (2.6). Thus
\[
|I_{1t}| = \left| [n^{-1/2} \tilde{G}_{n}(\phi - \phi_0)] \log(1-B)(1-B)^d \psi_0(B) \psi^{-1}(B) \psi_0(B) [n^{-1/2} \tilde{G}_{n} x_t] \right|
\]
\[\leq n^{-\alpha} \log(1-B)(1-B)^d \epsilon_0(B) \epsilon_{0t} \epsilon_0(B) [n^{-1/2} \tilde{G}_{n} x_t] \right|
\]
\[\leq n^{-\alpha} \log(1-B)(1-B)^d \epsilon_0(B) \epsilon_{0t} \epsilon_0(B) [n^{-1/2} \tilde{G}_{n} x_t] \right|
\]
\[\leq n^{-\alpha} \log(1-B)(1-B)^d \epsilon_0(B) \epsilon_{0t} \epsilon_0(B) [n^{-1/2} \tilde{G}_{n} x_t] \right|
\]
\[\leq O(1).
\]
(C.17)
Now, we consider the second term $I_2$ in (C.16). Let $\eta_i = (1 - B)^{2\varepsilon_0}$, where $\bar{a} = d - d_0$. By the given condition, $|\bar{a}| < \frac{1}{2}$ for large enough $n$ and hence by Theorem 1.1(d) in Hosking (1981), $E\eta_i^2 = \text{a constant}$ and for large enough $k$, $E(\eta_i, \eta_{i-k}) = O(k^{2\bar{a}-1})$. Thus

\[
\frac{1}{n} \sum_{i=1}^{n} E^2 = E\left( \sum_{i=1}^{\infty} \frac{1}{i} \eta_{i-i} \right)^2
\]

\[
= \sum_{i=1}^{\infty} \frac{1}{i^2} E\eta_i^2 + 2 \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{i(i+k)} E(\eta_i, \eta_{i-k})
\]

\[
= O(1) + 2 \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} O\left( \frac{1}{i(i+k)k^{1-2\bar{a}}} \right)
\]

\[
\leq O(1) + 2 \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} O\left( \frac{1}{i(i+1/2)k^{1+1/2-2\bar{a}}} \right) = O(1).
\]

(C.18)

Similarly we can show that $n^{-1} \sum_{i=1}^{n} E_3^2 = O(1)$. Further by (C.16)–(C.18), we know that $n^{-1} \sum_{i=1}^{n} E(\partial^2 \varepsilon_i / \partial d^2)^2 = O(1)$. Similarly $n^{-1} \sum_{i=1}^{n} E(\partial^3 \varepsilon_i / \partial d^3)^2 = O(1)$ and $n^{-1} \sum_{i=1}^{n} E(\partial^3 \varepsilon_i / \partial d^3)^2 = O(1)$. By the preceding discussion, we can show that

\[
\frac{1}{n} \partial^3 S(\lambda) \partial d^3 = O_p(1),
\]

(C.19)

Similarly we can show

\[
\frac{1}{n} \partial^3 S(\lambda) \partial d^3 = O_p(1)
\]

and

\[
\frac{1}{n} \sqrt{n} G_n \partial^3 S(\lambda) \partial d^3 \partial \varphi = O_p(1).
\]

(C.20)

By (C.19) and (C.20),

\[
\frac{1}{n} \left[ \frac{\partial^2 S(\lambda)}{\partial d^2} - \frac{\partial^2 S(\lambda_0)}{\partial d^2} \right] = \frac{1}{n} \sqrt{n} G_n \left[ \frac{\partial^3 S(\lambda)}{\partial d^3 \partial \varphi} \right]' [n^{-1/2} G_n^{-1}(\phi - \varphi)]
\]

\[
+ \frac{1}{n} \frac{\partial^3 S(\lambda_0)}{\partial d^3 \partial \varphi} (\psi - \varphi_0) + \frac{1}{n} \frac{\partial^3 S(\lambda_0)}{\partial d^3} (d - d_0)
\]

\[
= O_p(\|(n^{-1/2} J_n^{-1}(\phi - \varphi_0))\|) + O_p(\|\psi - \psi_0\|) + O_p(\|d - d_0\|)
\]

\[
= O_p(\|(n^{-1/2} G_n^{-1}(\lambda - \lambda_0))\|),
\]

(C.21)

where $\lambda_0$ is an intermediate point between $\lambda$ and $\lambda_0$. Similar to the proof of (C.21), we can show that the following quantities are equal to $O_p(\|(1/\sqrt{n}) G_n^{-1}(\lambda - \lambda_0))$: 

\[
\tilde{G}_n \left[ \frac{\partial^2 S(\lambda)}{\partial d \partial \varphi} - \frac{\partial^2 S(\lambda_0)}{\partial d \partial \varphi} \right] \tilde{G}_n',
\]

\[
\frac{1}{\sqrt{n}} \left[ \frac{\partial^2 S(\lambda)}{\partial \psi \partial \varphi} - \frac{\partial^2 S(\lambda_0)}{\partial \psi \partial \varphi} \right] \tilde{G}_n',
\]

\[
\frac{1}{n} \left[ \frac{\partial^2 S(\lambda)}{\partial \psi \partial \varphi} - \frac{\partial^2 S(\lambda_0)}{\partial \psi \partial \varphi} \right] \tilde{G}_n'.
\]

Thus, $G_n[I_n(\lambda) - I_n] = O_p(\|(1/\sqrt{n}) G_n^{-1}(\lambda - \lambda_0))$. This completes the proof. ■