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

$$\phi_0(B)(1-B)^{d_0} y_t = \psi_0(B)\varepsilon_{0t}, \tag{1.1}$$

where $\{\varepsilon_{0t}\}$ is a sequence of independently and identically distributed (i.i.d.) random variables with mean 0, variance 1, and $E|\varepsilon_{0t}|^{2+\alpha} < \infty$ 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

$$(1 - B)^{d_0} = \sum_{k=0}^{\infty} a_{0k} B^k, \qquad a_{0k} = \frac{(k - d_0 - 1)!}{k!(-d_0 - 1)!}, \tag{1.2}$$

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 $\rho(k) \sim ck^{2d_0 - 1}$ 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 $u_t = (1 - B)^{-d_0}\varepsilon_{0t}$ 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 $\phi_0(B)u_t = \psi_0(B)\varepsilon_{0t}$ 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., $(1 - B)(1 - B)^{d_0}y_t = \varepsilon_{0t}$ with initial values $y_t = 0$ as $t \leq 0$. From DGM1, it follows that

$$y_t = y_{t-1} + u_t, \qquad u_t = \sum_{k=0}^{\infty} \tilde{a}_{0k} \varepsilon_{0t-k}, \tag{1.4}$$

where $\tilde{a}_{0k} = (k + d_0 - 1)! / [k!(d_0 - 1)!]$, i.e., the coefficient of the kth 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 $\sum_{k=0}^{t-1} a_{0k} y_{t-k} = u_t$, where $(1 - B)u_t = \varepsilon_t$, $t = 1, 2, \ldots$, and $u_0 = 0$. Thus, we have

$$y_t = y_{t-1} + u_t^*, \qquad u_t^* = \sum_{k=0}^{t-1} \tilde{a}_{0k} \varepsilon_{0t-k} \tag{1.5}$$

(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 $t, k \geq 1$, $E(u_t^* u_{t+k}^*) = E(u_t u_{t+k}) + O(t^{-\alpha}) = O(k^{2d_0-1} + t^{-\alpha})$, 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 ε_{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 $x_t = \sum_{k=0}^{t-1} a_{0k} y_{t-k}$. Under DGM1, $(1 - B)x_t \neq \varepsilon_{0t}$, Jeganathan (1999) shows that $n^{-\max\{1/2, 1/2+d_0\}} x_{[n\tau]}$ converges to $F(\tau)$, a functional of the fractional Brownian motion. However, under DGM2, we have $(1 - B)x_t = \varepsilon_{0t}$, and hence $n^{-1/2} x_{[n\tau]}$ 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); $\rightarrow_{\mathcal{L}}$ 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 λ_0 is the true parameter that together with $\{\varepsilon_{0t}\}$ generates y_t through (1.1) with DGM2. To estimate λ_0, we need the unknown parameter model

$$\phi(B)(1 - B)^d y_t = \psi(B)\varepsilon_t, \tag{2.1}$$

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 λ_0 and λ are the interior points of the compact set Θ, $|d| < \frac{1}{2}$ and all the roots of $\psi(z)$ are outside the unit circle. The CSS estimator is the parameter λ in Θ that minimizes the objective function,

$$S(\lambda) = \sum_{t=1}^{n} \varepsilon_t^2, \tag{2.2}$$

where $\varepsilon_t = \psi^{-1}(B)\phi(B)(1 - B)^d y_t$, a function of λ and the observed data $\{y_t\}$. Note that $\varepsilon_{0t} = \varepsilon_t|_{\lambda=\lambda_0}$. When $\{\varepsilon_{0t}\}$ 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

moving-average (ARMA) models. Hosking (1984) considers the CSS method for stationary FARIMA models. A similar method is also used by Li and McLeod (1986) for stationary FARIMA models. More recently Chung and Baillie (1993) and Chung (1996) use the CSS estimation for some general long memory fractional ARIMA models.

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

$$\frac{\partial S(\lambda)}{\partial \phi} = 2 \sum_{t=1}^{n} \frac{\partial \varepsilon_t}{\partial \phi} \varepsilon_t, \qquad \frac{\partial S(\lambda)}{\partial d} = 2 \sum_{t=1}^{n} \varepsilon_t \log(1-B) \varepsilon_t, \qquad (2.3)$$

$$\frac{\partial S(\lambda)}{\partial \psi} = 2 \sum_{t=1}^{n} \frac{\partial \varepsilon_t}{\partial \psi} \varepsilon_t, \qquad \frac{\partial^2 S(\lambda)}{\partial \phi \partial \phi'} = 2 \sum_{t=1}^{n} \frac{\partial \varepsilon_t}{\partial \phi} \frac{\partial \varepsilon_t}{\partial \phi'}, \qquad (2.4)$$

$$\frac{\partial^2 S(\lambda)}{\partial \psi \partial \psi'} = 2 \sum_{t=1}^{n} \frac{\partial \varepsilon_t}{\partial \psi} \frac{\partial \varepsilon_t}{\partial \psi'} + 2 \sum_{t=1}^{n} \frac{\partial^2 \varepsilon_t}{\partial \psi \partial \psi'} \varepsilon_t, \qquad (2.5)$$

$$\frac{\partial^2 S(\lambda)}{\partial d^2} = 2 \sum_{t=1}^{n} \{[\log(1-B)\varepsilon_t]^2 + \varepsilon_t \log^2(1-B)\varepsilon_t\}, \qquad (2.6)$$

where $\partial \varepsilon_t / \partial \phi = -\psi^{-1}(B)(1-B)^d \mathbf{y}_{t-1}$, $\mathbf{y}_t = (y_t, \ldots, y_{t-p+1})'$, and $\partial \varepsilon_t / \partial \psi_i = -\psi^{-1}(B)\varepsilon_{t-i}$, $i = 1, \ldots, q$. 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 $J_n = \mathrm{diag}(N_n, \tilde{N}_n, L_{1,n}, \ldots, L_{l,n}, n^{1/2} I_{p^* \times p^*})$,

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

where $N_n = \mathrm{diag}(n, \ldots, n^a)$, $\tilde{N}_n = \mathrm{diag}(n, \ldots, n^b)$, $L_{k,n} = \mathrm{diag}(nI_{2 \times 2}, \cdots, n^{d_k} I_{2 \times 2})$, $k = 1, \ldots, l$, 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 $Z_t^0 = (Z_{\phi^* t}^{0'}, Z_{\psi t}^{0'}, Z_{dt}^0)'$,

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

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 $\partial S(\lambda)/\partial \lambda = 0$ such that*

$$n^{\alpha-1/2} G_n'^{-1}(\hat{\lambda}_n - \lambda_0) = o_p(1);$$

(b) *for such a sequence,*

$$G_n'^{-1}(\hat{\lambda}_n - \lambda_0) \to_{\mathcal{L}} ((F^{-1}\xi)', (\tilde{F}^{-1}\tilde{\xi})', (H_1^{-1}\zeta_1)', \ldots, (H_l^{-1}\zeta_l)', (\Sigma^{-1}N)')',$$

where $0 < \alpha < \frac{1}{2}$, the limiting distributions F, ξ, $\tilde{F}$, $\tilde{\xi}$, H_k, and ζ_k are defined in Appendix A, $\Sigma = E(Z_t^0 Z_t^{0'})$, N is a $(p^* + q + 1)$-dimensional standard normal vector, and (F, ξ), $(\tilde{F}, \tilde{\xi})$, (H_1, ζ_1), ..., (H_l, ζ_l), 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

$$\Sigma = \begin{pmatrix} E(Z_{1t}^0 Z_{1t}^{0'}) & E(Z_{1t}^0 Z_{dt}^0) \\ E(Z_{1t}^{0'} Z_{dt}^0) & \pi^2/6 \end{pmatrix},$$

where $Z_{1t}^0 = (Z_{\phi^* t}^{0'}, Z_{\psi t}^{0'})'$. 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,

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

where $\phi_0 = 1$. Suppose that $(\hat{\phi}, \hat{d})$ is the CSS estimator of (ϕ_0, d_0). Then, by Theorem 2.1,

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

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

and

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

where $u_{t-1} = (1 - B)^{\hat{d}} y_{t-1}$, $W(\tau)$ is a standard Brownian motion, and ρ and $\sqrt{n}(\hat{d} - d_0)$ are asymptotically independent.

Remark. The limiting distributions of ρ and τ 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 $\hat{\lambda}_n^*$. In a similar manner to that given by Ling and Li (1998), we can show that

$$[n(\hat{\gamma}_1 - 1),\sqrt{n}(\hat{\lambda}_n^* - \lambda_0^*)'] \xrightarrow{\mathcal{L}} \mathrm{diag}(c\xi,N'), \tag{2.12}$$

where $c = 1/(1 - \sum_{i=2}^{p}\gamma_{0i})$, ξ is defined by the right-hand side of (2.8), N is a normal random vector with mean zero and variance Σ, which is independent of ξ. 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$(1, d_0, 0)$ 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, ± 0.1, and ± 0.3, and $\phi_0 = 1.0$. For various parameter pairs of (d_0,ϕ_0) and each of the 1,000 replications, samples of series lengths $n = 200$ and 300 are used, respectively. In the estimation, the LS estimator of ϕ_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 $\hat{u}_t = (1 - \hat{\phi}B)y_t$ 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$(1,d_0,0)$ model

| | | $n = 200$ | | | | $n = 300$ | | | |
| | | Mean | | $\sqrt{MSE}$ | | Mean | | $\sqrt{MSE}$ | |
d_0	ϕ_0	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$
0.3	1.0	0.303	0.991	0.071	0.023	0.301	0.995	0.053	0.012
0.1	1.0	0.105	0.989	0.072	0.029	0.101	0.994	0.052	0.013
0.0	1.0	−0.002	0.989	0.073	0.024	0.000	0.994	0.053	0.013
−0.1	1.0	−0.099	0.990	0.071	0.020	−0.100	0.994	0.053	0.012
−0.3	1.0	−0.296	0.991	0.072	0.026	−0.299	0.995	0.053	0.014

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 ϕ_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 ρ and τ 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, ϕ_0) and each of the 1,000 replications, samples of series lengths $n = 200$ and 300 are considered. The critical values of ρ and τ 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$(1, d_0, 1)$ 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 2. Power and size of lower tail Dickey–Fuller test based on the 5% significance level for the nonstationary FARIMA$(1, d_0, 0)$ model

d_0	ϕ_0		$n = 200$					$n = 300$				
			0.8	0.9	0.95	0.99	1.0	0.8	0.9	0.95	0.99	1.0
0.3	1.0	ρ	0.992	0.898	0.624	0.151	0.066	1.000	0.985	0.840	0.164	0.049
		τ	0.992	0.898	0.627	0.156	0.067	0.999	0.986	0.846	0.160	0.049
0.1	1.0	ρ	1.000	0.964	0.688	0.176	0.067	1.000	1.000	0.888	0.221	0.068
		τ	1.000	0.959	0.688	0.182	0.065	1.000	0.999	0.890	0.216	0.067
0.0	1.0	ρ	0.994	0.926	0.682	0.171	0.073	1.000	0.986	0.879	0.222	0.066
		τ	0.993	0.927	0.682	0.171	0.073	1.000	0.988	0.874	0.223	0.062
−0.1	1.0	ρ	0.993	0.930	0.689	0.162	0.070	1.000	0.986	0.873	0.220	0.062
		τ	0.992	0.929	0.688	0.164	0.069	0.999	0.986	0.870	0.215	0.060
−0.3	1.0	ρ	0.999	0.917	0.639	0.142	0.063	1.000	0.981	0.842	0.180	0.050
		τ	0.998	0.919	0.634	0.141	0.063	1.000	0.982	0.843	0.175	0.054

Note: Replications = 1,000.

TABLE 3. The empirical mean and standard deviation of CSS estimators for the nonstationary FARIMA$(1, d_0, 1)$ Model

			Mean			$\sqrt{MSE}$		
d_0	ϕ_0	ψ_0	$\hat{d}$	$\hat{\phi}$	$\hat{\psi}$	$\hat{d}$	$\hat{\phi}$	$\hat{\psi}$
0.3	1.0	0.5	0.290	0.985	0.484	0.064	0.013	0.045
−0.3	1.0	0.5	−0.311	0.995	0.482	0.060	0.010	0.049
0.3	1.0	0.3	0.291	0.985	0.286	0.060	0.013	0.042
−0.3	1.0	0.3	−0.308	0.995	0.283	0.053	0.009	0.042
0.3	1.0	−0.3	0.297	0.984	−0.310	0.050	0.013	0.027
−0.3	1.0	−0.3	−0.300	0.995	−0.309	0.043	0.009	0.026
0.3	1.0	−0.5	0.299	0.984	−0.507	0.048	0.013	0.021
−0.3	1.0	−0.5	−0.299	0.994	−0.507	0.044	0.009	0.021

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

± 0.3. Various combinations of values d_0 and ψ_0 are used. The means and the standard errors of the estimators of d_0, ϕ_0, and ψ_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

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

where $u_{t-1} = (\partial \varepsilon_t / \partial \phi)_{(\phi, \psi, d)=(\hat{\phi}, \hat{\psi}, \hat{d})}$. Here, ρ and τ have the same asymptotic distributions as those in (2.8) and (2.9). To examine the power and size of ρ and τ, we take $\phi_0 = 0.8, 0.9, 0.95, 1.0$ with (ψ_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 ρ and τ are 0.053 and 0.054, respectively. This means that we can obtain almost the exact sizes of ρ and τ 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

ψ_0		$d_0 = 0.3$					$d_0 = -0.3$				
		0.8	0.9	0.95	0.99	1.0	0.8	0.9	0.95	0.99	1.0
0.5	ρ	1.000	0.995	0.955	0.251	0.057	1.000	1.000	0.960	0.255	0.060
	τ	1.000	0.994	0.958	0.253	0.053	1.000	0.999	0.962	0.260	0.058
0.3	ρ	1.000	0.994	0.937	0.254	0.053	1.000	1.000	0.943	0.253	0.059
	τ	1.000	0.993	0.937	0.260	0.051	1.000	0.999	0.944	0.247	0.061
-0.3	ρ	1.000	1.000	0.968	0.253	0.052	1.000	1.000	0.968	0.251	0.073
	τ	1.000	0.999	0.966	0.247	0.054	1.000	0.999	0.966	0.253	0.070
-0.5	ρ	1.000	1.000	0.970	0.255	0.051	1.000	1.000	0.970	0.255	0.072
	τ	1.000	1.000	0.969	0.255	0.052	1.000	1.000	0.969	0.255	0.071

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*

$$\phi_0^*(B)z_t = \psi_0(B)\varepsilon_{0t} \tag{4.1}$$

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 ε_{0t} are defined as in model (1.1). Denote $Z_{t-1} = -(\partial\varepsilon_{0t}/\partial\phi^{*'}, \partial\varepsilon_{0t}/\partial\psi', \partial\varepsilon_{0t}/\partial d)'$, where $\phi_0^* = (\phi_{01}^*,\ldots,\phi_{0p^*}^*)'$, $\psi_0 = (\psi_{01},\ldots,\psi_{0q})'$, $\partial\varepsilon_{0t}/\partial\phi^{*'} = \partial\varepsilon_t/\partial\phi^{*'}|_{(\phi_0^*,\psi_0,d_0)}$, and similarly define $\partial\varepsilon_{0t}/\partial\psi$ and $\partial\varepsilon_{0t}/\partial d$. Let

$$S_t = (\varepsilon_{0t}, (-1)^t\varepsilon_{0t}, \sqrt{2}\sin t\theta_1\varepsilon_{0t}, \sqrt{2}\cos t\theta_1\varepsilon_{0t},\ldots,\sqrt{2}\cos t\theta_l, Z_{t-1}'\varepsilon_{0t})',$$

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]} S_t \xrightarrow{\mathcal{L}} (W'(\tau), N'(\tau))' \quad \text{in } D^{2(l+1)+p^*+q+1}, \tag{4.2}$$

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 $Z_{t-1}'\varepsilon_t$. 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. *$n^{-1}\sum_{k=2}^{[n\tau]} e^{ik\theta} E(\varepsilon_{0k}^2 Z_{k-1}|\mathcal{F}_{k-1}) = o_p(1)$, where $i = \sqrt{-1}$, $\tau \in [0,1]$, and $\mathcal{F}_t = \sigma\{\varepsilon_{01},\ldots,\varepsilon_{0t}\}$, the σ-field generated by $\varepsilon_{01},\ldots,\varepsilon_{0t}$.*

LEMMA 4.2. *$n^{-1}\sum_{k=2}^{[n\tau]} E[\|Z_k\varepsilon_{0k+1}\|^2 I(\|Z_k\varepsilon_{0k+1}\| > \sqrt{n}\delta|\mathcal{F}_k)] = o_p(1)$.*

Proof of Theorem 4.1. First,

$$\frac{1}{n}\sum_{t=1}^{n} E\left(\frac{\partial\varepsilon_{0t}}{\partial d} - Z_{dt}^0\right)^2 = \frac{1}{n}\sum_{t=1}^{n} E\left(\sum_{i=t}^{\infty} \frac{1}{i}\varepsilon_{0t-i}\right)^2$$

$$= \frac{1}{n}\sum_{t=1}^{n}\sum_{i=t}^{\infty}\frac{1}{i^2} = o(1).$$

Similarly, we can show that $n^{-1} \sum_{t=1}^{n} E \| \partial \varepsilon_{0t} / \partial \phi^* - Z_{\phi^* t}^0 \|^2 = o(1)$ and $n^{-1} \sum_{t=1}^{n} E \| \partial \varepsilon_{0t} / \partial \psi - Z_{\psi t}^0 \|^2 = o(1)$. Thus,

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

Because Z_t^0 is strictly stationary and ergodic with $E(Z_t^0 Z_t^{0'}) < \infty$, by the ergodic theorem, $n^{-1} \sum_{t=1}^{n} Z_t^0 Z_t^{0'} = \Sigma + o_p(1)$. Furthermore, by (4.3), we can show that $n^{-1} \sum_{t=1}^{n} (Z_t Z_t' - Z_t^0 Z_t^{0'}) = o_p(1)$ and hence

$$\frac{1}{n} \sum_{t=2}^{[n\tau]} Z_{t-1} Z_{t-1}' = \tau \Sigma + o_p(1). \tag{4.4}$$

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

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

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

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

$$\phi_0(B) x_t = \varepsilon_{0t}, \tag{4.5}$$

with $x_t = 0$ as $t \leq 0$.[1] 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

$$(1 - B)^a u_t = \varepsilon_{0t}, \qquad (1 + B)^b v_t = \varepsilon_{0t},$$

$$\phi_0^*(B) z_t = \varepsilon_{0t},$$

$$(1 - 2\cos\theta_k B + B^2)^{d_k} x_{t,k} = \varepsilon_{0t}, \qquad k = 1, \ldots, l,$$

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

$$\mathbf{u}_t = (u_t, \ldots, u_{t-a+1})', \qquad \mathbf{v}_t = (v_t, \ldots, v_{t-b+1})',$$

$$\mathbf{z}_t = (z_t, \ldots, z_{t-p^*+1})',$$

$$\mathbf{x}_{t,k} = (x_{t,k}, \ldots, x_{t-d_k+1,k})', \qquad k = 1, \ldots, l.$$

As shown in Chan and Wei (1988), there exists a nonsingular matrix Q^* such that

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

where $\mathbf{x}_t = (x_t, \ldots, x_{t-p+1})'$.

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 $j = 0, 1, \ldots, b-1$, $V_t = (v_t(1), \ldots, v_t(b))'$, $x_{t,k}(j+1) = (1 - 2B \cos \theta_k + B^2)^{d_k-j} x_{t,k}$ 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, $\widetilde{M}$, C_k, $k = 1, \ldots, l$, which can be found in Chan and Wei (1988), such that

$$M\mathbf{u}_t = U_t, \qquad \widetilde{M}\mathbf{v}_t = V_t, \qquad C_k \mathbf{x}_{t,k} = X_{t,k}, \qquad k = 1, \ldots, l. \tag{4.7}$$

Denote $G^* = \mathrm{diag}(M, \widetilde{M}, C_1, \ldots, C_l, I_{p^*}) Q^*$. Then

$$G^* \mathbf{x}_t = (U_t', V_t', X_{t,1}', \ldots, X_{t,l}', \mathbf{z}_t')'. \tag{4.8}$$

Thus, $\mathbf{x}_t$ 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 ϕ_0^* in the information matrix converge to zero in probability.

LEMMA 4.3.

(a) $N_n^{-1} \sum\limits_{t=2}^{n} U_{t-1} \varepsilon_{0t} \to_{\mathcal{L}} \xi,$

(b) $\widetilde{N}_n^{-1} \sum\limits_{t=2}^{n} V_{t-1} \varepsilon_{0t} \to_{\mathcal{L}} \widetilde{\xi},$

(c) $L_{k,n}^{-1} \sum\limits_{t=2}^{n} X_{t-1,k} \varepsilon_{0t} \to_{\mathcal{L}} \zeta_k,$

(d) $N_n^{-1} \sum\limits_{t=2}^{n} U_{t-1} U_{t-1}' N_n^{-1} \to_{\mathcal{L}} F,$

(e) $\widetilde{N}_n^{-1} \sum\limits_{t=2}^{n} V_{t-1} V_{t-1}' \widetilde{N}_n^{-1} \to_{\mathcal{L}} \widetilde{F},$

(f) $L_{k,n}^{-1} \sum\limits_{t=2}^{n} X_{t-1,k} X_{t-1,k}' L_{k,n}^{-1} \to_{\mathcal{L}} H_k,$

where $k = 1,\ldots,l$, ξ, F, $\tilde{\xi}$, $\tilde{F}$, ζ_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:*

$$N_n^{-1} \sum_{t=2}^{n} U_{t-1} V'_{t-1} \tilde{N}_n^{-1}, \qquad N_n^{-1} \sum_{t=2}^{n} U_{t-1} X'_{t-1,k} L_{k,n}^{-1},$$

$$\tilde{N}_n^{-1} \sum_{t=2}^{n} V_{t-1} X'_{t-1,k} L_{k,n}^{-1}, \qquad k = 1,\ldots,l,$$

$$L_{k,n}^{-1} \sum_{t=2}^{n} X_{t-1,k} X'_{t-1,k_1} L_{k_1,n}^{-1}, \qquad k \neq k_1, \qquad k, k_1 = 1,\ldots,l,$$

$$N_n^{-1} \sum_{t=2}^{n} U_{t-1} \mathbf{z}'_{t-1}/\sqrt{n}, \; \tilde{N}_n^{-1} \sum_{t=2}^{n} V_{t-1} \mathbf{z}'_{t-1}\sqrt{n},$$

$$L_{k,n}^{-1} \sum_{t=2}^{n} X_{t-1,k} \mathbf{z}'_{t-1}/\sqrt{n}, \qquad k = 1,\ldots,l.$$

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 ψ_0. Lemma 4.6 shows the asymptotic distributions of the score function and the information matrix.

LEMMA 4.5. *Suppose that ς_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) $n^{-1/2} \sum_{t=2}^{n} \varsigma_{t-1} \left(\sum_{i=1}^{t-1} \dfrac{1}{i} \varepsilon_{0t-i} \right) + n^{-1/2} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} \dfrac{1}{i} \varsigma_{t-i} \right) \varepsilon_{0t} = o_p(1),$

(b) $n^{-1/2} \sum_{t=2}^{n} \varsigma_{t-1} \left(\sum_{i=1}^{t-1} v_{\psi_0}(i) \varepsilon_{0t-i} \right) + n^{-1/2} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} v_{\psi_0}(i) \varsigma_{t-i} \right) \varepsilon_{0t} = o_p(1),$

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

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

$$E\left(\sum_{t=1}^{i-1} \eta_t \right)^2 = E\left(\sum_{t=1}^{i-1} \mu_t - \sum_{t=1}^{i-1} \sum_{k=t}^{\infty} \frac{1}{k} \varepsilon_{0t-k} \right)^2$$

$$\leq 2E\left(\sum_{t=1}^{i-1} \mu_t \right)^2 + 2E\left(\sum_{t=1}^{i-1} \sum_{k=t}^{\infty} \frac{1}{k} \varepsilon_{0t-k} \right)^2. \qquad (4.9)$$

Furthermore, we have

$$E\left(\sum_{t=1}^{i}\mu_t\right)^2 = iE\mu_t^2 + 2\sum_{r=1}^{i-1}(i-r)E\mu_t\mu_{t-r}$$

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

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

$$= iE\mu_t^2 + O\left(\sum_{r=1}^{i-1}(i-r)r^{-1/2}\right) = o(i^{3/2+1/8}). \qquad (4.10)$$

By Minkowski's inequality,

$$E\left(\sum_{t=1}^{i-1}\sum_{k=t}^{\infty}\frac{1}{k}\varepsilon_{0t-k}\right)^2 \leq \left[\sum_{t=1}^{i-1}\sqrt{E\left(\sum_{k=t}^{\infty}\frac{1}{k}\varepsilon_{0t-k}\right)^2}\right]^2$$

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

$$= o(i^{3/2}), \qquad (4.11)$$

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

$$E\left(\sum_{t=1}^{i-1}\eta_t\right)^2 = o(i^{3/2+1/8}). \qquad (4.12)$$

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

$$u_t(j) = \sum_{i=1}^{t}u_i(j-1), \qquad (4.13)$$

and hence

$$\sum_{t=2}^{n}u_{t-1}(j)\eta_t = \left[\sum_{t=2}^{n}u_{t-1}(j-1)\right]\left(\sum_{i=1}^{n}\eta_i\right) - \sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\eta_i\right)u_{t-1}(j-1)$$

$$= I_{1n} - I_{2n}, \quad \text{say}. \qquad (4.14)$$

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

$$n^{-j-1/2}I_{1n} = [n^{-(j-1)-1/2}u_n(j)]\left(n^{-1}\sum_{i=1}^{n}\eta_t\right) = o_p(1). \qquad (4.15)$$

When $j = 1$, by (4.12),

$$EI_{2n}^2 = \sum_{t=2}^{n} E\left(\sum_{i=1}^{t-1} \eta_i\right)^2 = \sum_{t=2}^{n} o(t^{3/2+1/8}) = o(n^{5/2+1/8}). \tag{4.16}$$

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 $j > 1$,

$$E|I_{2n}| \leq \sum_{t=2}^{n} E\left|\left(\sum_{i=1}^{t-1} \eta_i\right)u_{t-1}(j-1)\right|$$

$$\leq n^{j-3/2} \sum_{t=2}^{n}\left\{E\left(\sum_{i=1}^{t-1} \eta_i\right)^2 E[n^{-(j-2)-1/2}u_t(j-1)]^2\right\}^{1/2}$$

$$= o\left(n^{j-3/2}\sum_{t=2}^{n} t^{3/4+1/16}\right) = o(n^{j+1/4+1/16}), \tag{4.17}$$

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

$$n^{-1/2}N_n^{-1}\sum_{t=2}^{n} U_{t-1}\eta_t = o_p(1). \tag{4.18}$$

Next, we consider the jth element of the second term in (a), that is, $n^{-j-1/2}\sum_{t=2}^{n}(\sum_{i=1}^{t-1} i^{-1}u_{t-i}(j))\varepsilon_{0t}$. 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

$$n^{-1/2}N_n^{-1}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\frac{1}{i}U_{t-i}\right)\varepsilon_{0t} = o_p(1). \tag{4.19}$$

By (4.18) and (4.19), (a) holds. For (b), because $v_{\psi_0}(i) = O(\rho^i)$ 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) $G_n D_n \to_{\mathcal{L}} -2(\xi', \tilde{\xi}', \zeta'_1 \dots, \zeta'_l, N')'$,

(b) $G_n I_n G'_n \to_{\mathcal{L}} 2 \operatorname{diag}(F, \tilde{F}, H_1, \dots, H_l, \Sigma)$,

where ξ, F, $\tilde{\xi}$, $\tilde{F}$, ζ_i, and H_i, $i = 1, \dots, l$ are defined in Appendix A, N is a normal random vector with mean 0 and covariance Σ defined as in Theorem 2.1, and (F, ξ), $(\tilde{F}, \tilde{\xi})$, (H_1, ζ_1), $\dots$, (H_l, ζ_l), and N are independent.

Proof. For (a), note that

$$G_n D_n = -2 \sum_{t=2}^n \left((N_n^{-1} U_t)', (\tilde{N}_n^{-1} V_t)', (L_{1,n}^{-1} X_{1,n})', \dots, (L_{l,n}^{-1} X_{l,n})', \frac{1}{\sqrt{n}} Z'_{t-1} \varepsilon_{0t} \right)',$$

(4.20)

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 $t \le 0$, i.e., model (4.1). By (4.4),

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

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

$$= 2E \begin{pmatrix} Z_{\psi t}^0 Z_{\psi t}^{0'} & Z_{\psi t}^0 Z_{dt}^0 \\ Z_{dt}^0 Z_{\psi t}^{0'} & \dfrac{\pi^2}{6} \end{pmatrix} + o_p(1), \tag{4.22}$$

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

$$J_n^{-1} G_n^* \left[\frac{\partial^2 S(\lambda_0)}{\partial \phi \partial \phi'} \right] G_n^{*'} J_n^{'-1}$$

$$= 2 \sum_{t=2}^n \operatorname{diag} \left\{ N_n^{-1} U_t U_t' N_n^{-1}, \tilde{N}_n^{-1} V_t V_t' \tilde{N}_n^{-1}, L_{1,n}^{-1} X_{1,n} X_{1,n}' L_{1,n}^{-1}, \right.$$

$$\left. \dots, L_{l,n}^{-1} X_{l,n} X_{l,n}' L_{l,n}^{-1}, \frac{1}{n} \frac{\partial \varepsilon_{0t}}{\partial \phi^*} \frac{\partial \varepsilon_{0t}}{\partial \phi^{*'}} \right\}$$

$$+ o_p(1). \tag{4.23}$$

By Lemma 4.5 and (4.4), we can show that

$$J_n^{-1} G_n^* \frac{\partial^2 S(\lambda_0)}{\partial \phi \partial d} = 2[0, E(Z_{\phi^* t}^{0'} Z_{dt}^0)]' + o_p(1), \tag{4.24}$$

$$J_n^{-1} G_n^* \frac{\partial^2 S(\lambda_0)}{\partial \phi \partial \psi'} = 2[0, E(Z_{\phi^* t}^0 Z_{\psi t}^{0'})']' + o_p(1), \tag{4.25}$$

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, ξ), $(\tilde{F}, \tilde{\xi})$, (H_1, ζ_1), …, (H_l, ζ_l), 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* $\|n^{\alpha-1/2} G_n'^{-1}(\lambda - \lambda_0)\| < 1$,

$$G_n[I_n(\lambda) - I_n] G_n' = O_p\left(\left\| \frac{1}{\sqrt{n}} G_n'^{-1}(\lambda - \lambda_0) \right\|\right),$$

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 Π.

Let v and ε 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 $(F, \tilde{F}, H_1, \ldots, H_l, \Sigma)$ is positive definite in probability. By Lemma 4.6(b), there are a constant c_1 and an integer N_1 such that, as $n > N_1$,

$$P\{G_n I_n G_n' < -c_1 I_{m \times m}\} > 1 - v,$$

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

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

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

$$P\{\Pi < -c\varepsilon^2\} > 1 - \nu. \tag{4.28}$$

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 $n > N_2$,

$$P\left\{n^{\alpha - 1/2}\|G_n D_n\| < \frac{c}{2}\varepsilon\right\} > 1 - \nu. \tag{4.29}$$

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

$$P\left\{[n^{\alpha - 1/2}G_n'^{-1}(\lambda - \lambda_0)]'[n^{\alpha - 1/2}G_n D_n] < \frac{c}{2}\varepsilon^2\right\} > 1 - \nu. \tag{4.30}$$

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 $1 - \nu$,

$$n^{2\alpha - 1}(\lambda - \lambda_0)'\frac{\partial S(\lambda)}{\partial \lambda} < -c\varepsilon^2 + \frac{c}{2}\varepsilon^2 < 0. \tag{4.31}$$

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

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

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

$$G_n'^{-1}(\hat{\lambda}_n - \lambda_0) = -\left[G_n I_n G_n' + O_p\left(\left\|\frac{1}{\sqrt{n}}G_n'^{-1}(\hat{\lambda}_n - \lambda_0)\right\|\right)\right]^{-1}(G_n D_n). \tag{4.32}$$

By part (a) of this theorem, $n^{\alpha - 1/2}G_n'^{-1}(\hat{\lambda}_n - \lambda_0)$ 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). ∎

NOTE

1. Denote $x_t^* = (1 - B)^{d_0}y_t$. Then $\phi_0(B)x_t^* = \psi_0(B)\varepsilon_{0t}$. The term x_t^* also has two different DGMs. The first one, namely, DGM1*, is to generate $u_t = \psi_0(B)\varepsilon_{0t}$ and then generate x_t^* through $\phi_0(B)x_t^* = u_t$ with $x_t^* = 0$ as $t \le 0$. The second one, namely DGM2*, is to generate $\phi_0(B)u_t = \varepsilon_{0t}$ with $u_t = 0$ as $t \le 0$ and then generate x_t^* through $x_t^* = \psi_0(B)u_t$. 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_t 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

Aitchison, J. & S.D. Silvey (1958) Maximum-likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics* 29, 813–828.

Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society, Series B* 57, 659–672.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Box, G.E.P. & G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control,* 2nd ed. San Francisco: Holden-Day.

Chan, N.H. & N. Terrin (1996) Inference for unstable long memory processes with applications to fractional unit root autoregressive. *Annals of Statistics* 23, 1662–1683.

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

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

Chung, C.F. & R.T. Baillie (1993) Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. In Jean-Marie Dufour (ed.), *New Developments in Time Series Econometrics*. New York: Springer-Verlag.

Chung, K.L. (1968) *A Course in Probability Theory*. New York: Academic Press.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

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

Granger, C.W.J. & R. Joyeux (1980) An introduction to long-memory time series models and fractional difference. *Journal of Time Series Analysis* 1, 15–39.

Helland, I.S. (1982) Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics* 9, 79–94.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–76.

Hosking, J.R.M. (1984) Modelling persistence in hydrological time series using fractional differencing. *Water Resources Research* 20, 1898–1908.

Jeganathan, P. (1999) On the asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* 15, 583–621.

Li, W.K. & A.I. McLeod (1986) Fractional time series modelling. *Biometrika* 73, 217–221.

Ling, S. (1997). *Stationary and Non-Stationary Time Series Models with Conditional Heteroscedasticity*. Ph.D. Dissertation, University of Hong Kong.

Ling, S. (1998) Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Annals of Statistics* 26, 741–754.

Ling, S. & W.K. Li (1997a) Fractional ARIMA-GARCH time series models. *Journal of the American Statistical Association* 92, 1184–1194.

Ling, S. & W.K. Li (1997b) Asymptotic Properties of the Conditional Sum of Squares Estimator for Nonstationary Fractional ARIMA Models. Technical report, Department of Statistics, the University of Hong Kong.

Ling, S. & W.K. Li (1998) Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Annals of Statistics* 26, 84–125.

McLeod, A.I. & K.W. Hipel (1978) Preservation of the rescaled adjusted range, part I: A reassessment of the Hurst phenomenon. *Water Resources Research* 14, 491–508.

Robinson, P. (1994a) Time series with strong dependence. In C. Sims (ed.), *Advances in Conometrics*, Sixth World Congress. Cambridge: Cambridge University Press.

Robinson, P. (1994b) Efficient tests of nonstationary hypothesis. *Journal of the American Statistical Association* 89, 1420–1437.

Schmidt, P. & P.C.B. Phillips (1992) LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.

Sowell, F. (1990) Fractional unit root distribution. *Econometrica* 50, 495–505.

Xiao, Z. & P.C.B. Phillips (1998) A primer on unit root testing. *Journal of Economic Surveys* 12, 423–469.

Yap, S.F. & G.C. Reinsel (1995) Results on estimation and testing for unit roots in the nonstationary autoregressive moving-average model. *Journal of Time Series Analysis* 16, 339–353.

APPENDIX A: LIMITING DISTRIBUTIONS

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

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

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

$$\tilde{F}_0(\tau) = W_2(\tau), \qquad \tilde{F}_j(\tau) = \int_0^\tau \tilde{F}_{j-1}(s)\,ds, \qquad j = 1,\ldots,b,$$

$$\tilde{F} = (\tilde{\sigma}_{ij}), \qquad \tilde{\sigma}_{ij} = \int_0^\tau \tilde{F}_{i-1}(s)\tilde{F}_{j-1}(s)\,ds, \qquad i,j = 1,\ldots,b,$$

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

$$\zeta = (\xi_1,\ldots,\xi_{2d})', \qquad H = (\sigma_{ij})_{2d\times 2d},$$

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

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

$$\left. - \sin\theta\left[\int_0^1 f_{j-1}(s)\,dW_1(s) + \int_0^1 g_{j-1}(s)\,dW_2(s) \right] \right\},$$

$$\sigma_{2k-1,2j-1} = \sigma_{2k,2j}$$

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

$$\sigma_{2k-1,2j} = \sigma_{2j,2k-1}$$

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

$$\left. - \sin\theta \left[\int_0^1 f_{j-1}(s) g_{k-1}(s) ds - \int_0^1 g_{j-1}(s) f_{k-1}(s) ds \right] \right\},$$

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

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

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

where $W_i(\tau)$ is the ith element of $W(\tau)$ defined in Theorem 4.1. Finally, define H_k and ζ_k as earlier with θ replaced by θ_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

$$1 \equiv (1-z)^{d_0}(1-z)^{-d_0}$$

$$\equiv \left(\sum_{i=0}^{\infty} a_{0i} z^i \right) \left(\sum_{i=0}^{\infty} \tilde{a}_{0i} z^i \right)$$

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

Thus, we have

$$\sum_{i=0}^{k} a_{0i} \tilde{a}_{0,k-i} = 0, \qquad k = 1,2,\dots. \tag{B.1}$$

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

$$\sum_{k=0}^{t-1} a_{0k}(y_{t-k} - y_{t-k-1}) = u_t - u_{t-1} = \varepsilon_{0t}.$$

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

$$\sum_{k=0}^{t-1} a_{0k} u_{t-k}^* = \varepsilon_{0t}. \tag{B.2}$$

From (B.2), it follows that

$$\sum_{k=0}^{t-1} \tilde{a}_{0k}\varepsilon_{0t-k} = \sum_{k=0}^{t-1} \tilde{a}_{0k}\left(\sum_{i=0}^{t-k-1} a_{0i}u^*_{t-k-i}\right)$$

$$= u^*_t + \sum_{k=1}^{t-1}\left(\sum_{i=0}^{k} a_{0i}\tilde{a}_{0,k-i}\right)u^*_{t-k}. \tag{B.3}$$

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

APPENDIX C: THE REMAINING PROOFS

Proof of Lemma 4.1.

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

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

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

Note that

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

By the triangle inequality, $|\sum_{k=1}^{[n\tau]-j} e^{ik\theta}\varepsilon_{0k}| \leq |\sum_{k=1}^{[n\tau]-j}\sin k\theta\varepsilon_{0k}| + |\sum_{k=1}^{[n\tau]-j}\cos k\theta\varepsilon_{0k}|$. It is easy to show that $\max_{1\leq j\leq n} n^{-1/2}|\sum_{k=1}^{j}\sin k\theta\varepsilon_{0k}|$ and $\max_{1\leq j\leq n} n^{-1/2}|\sum_{k=1}^{j}\cos k\theta\varepsilon_{0k}|$ converge to $\max_{0\leq\tau\leq 1}|B_{1k}(\tau)|$ and $\max_{0\leq\tau\leq 1}|B_{2k}(\tau)|$ in distribution, where $B_{1k}(\tau)$ and $B_{2k}(\tau)$ are Brownian motions. Thus

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

where $0 < \alpha < \frac{1}{2}$, and the last equation holds because $\sum_{j=1}^{\infty} j^{-1-\alpha} < \infty$.

Note that $\partial\varepsilon_{0t}/\partial\phi^*_i = -\sum_{k=1}^{t-1} c_\psi(k)\varepsilon_{0t-i-k}$ and $\partial\varepsilon_{0t}/\partial\psi_i = -\sum_{k=1}^{t-1} c_{\phi^*}(k)\varepsilon_{0t-i-k}$, with $c_\psi(k) = O(\rho^k)$ and $c_{\phi^*}(k) = O(\rho^k)$, where $0 \leq \rho < 1$. Similar to the proof of (C.2), we have $-(1/n)\sum_{k=2}^{j} e^{ik\theta}(\partial\varepsilon_{0t}/\partial\phi^*) = o_p(1)$ and $-(1/n)\sum_{k=2}^{j} e^{ik\theta}(\partial\varepsilon_{0t}/\partial\psi) = o_p(1)$. This completes the proof.   ∎

Proof of Lemma 4.2. We first show that

$$\frac{1}{n} \sum_{k=2}^{[n\tau]} E\left[\varepsilon_{0k+1}^2 \left(\frac{\partial \varepsilon_{0k}}{\partial d} \right)^2 I\left(\left| \varepsilon_{0k+1}\left(\frac{\partial \varepsilon_{0k}}{\partial d} \right) \right| > \sqrt{n}\delta \bigg| \mathcal{F}_k \right) \right] = o_p(1). \tag{C.3}$$

Note that $\partial \varepsilon_{0k}/\partial d = -\sum_{i=1}^{k-1} \varepsilon_{0k-i}/i$ and

$$\frac{1}{n} \sum_{k=2}^{n} \left(\sum_{i=1}^{k-1} \frac{1}{i} \varepsilon_{0k-i} \right)^2 \leq \frac{2}{n} \sum_{k=2}^{n} Z_{dk}^{02} + \frac{2}{n} \sum_{k=2}^{n} \left(\sum_{i=k}^{\infty} \frac{1}{i} \varepsilon_{0k-i} \right)^2$$

$$= 2A_1 + 2A_2, \text{ say}, \tag{C.4}$$

where Z_{dt}^0 is defined before Theorem 2.1.

$$E|A_2| = \frac{1}{n} \sum_{k=2}^{n} E\left(\sum_{i=k}^{\infty} \frac{1}{i} \varepsilon_{0k-i} \right)^2 = \frac{1}{n} \sum_{k=2}^{n} \sum_{j=k}^{\infty} j^{-2}$$

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

where $0 < \alpha < 1$. Because Z_{dk}^0 is strictly stationary and ergodic with $EZ_{dk}^{02} < \infty$, by the ergodic theorem, $A_1 = n^{-1} \sum_{k=2}^{n} Z_{dk}^{02}$ converges to $\pi^2/6$ almost surely (a.s.). Thus, by (C.4) and (C.5),

$$\frac{1}{n} \sum_{k=2}^{n} \left(\sum_{i=1}^{k-1} \frac{1}{i} \varepsilon_{0k-i} \right)^2 = O_p(1). \tag{C.6}$$

Note that

$$\left(\sum_{i=1}^{k-1} \frac{1}{i} \varepsilon_{0k-i} \right)^2 \leq 2Z_{dt}^{02} + 2\left(\sum_{i=k}^{\infty} \frac{1}{i} \varepsilon_{0k-i} \right)^2. \tag{C.7}$$

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

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

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

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

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

$$\frac{1}{n} \sum_{k=2}^{[n\tau]} E\left[\left(\varepsilon_{0k} \sum_{i=1}^{k-1} \frac{1}{i} \varepsilon_{0k-i} \right)^2 I\left(\left| \varepsilon_{0k} \sum_{i=1}^{k-1} \frac{1}{i} \varepsilon_{0k-i} \right| > \sqrt{n}\delta \right) \bigg| \mathcal{F}_{k-1} \right]$$

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

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

$$= o_p(1), \tag{C.9}$$

where K is some constant. That is, (C.3) holds. Similarly we can show that

$$\frac{1}{n}\sum_{k=2}^{[n\tau]} E\left[\varepsilon_{0k+1}^2 \left\|\frac{\partial\varepsilon_{0k}}{\partial\phi^*}\right\|^2 I\left(|\varepsilon_{0k+1}|\left\|\frac{\partial\varepsilon_{0k}}{\partial\phi^*}\right\| > \sqrt{n}\delta|\mathcal{F}_k\right)\right] = o_p(1), \tag{C.10}$$

$$\frac{1}{n}\sum_{k=2}^{[n\tau]} E\left[\varepsilon_{0k+1}^2 \left\|\frac{\partial\varepsilon_{0k}}{\partial\psi}\right\|^2 I\left(|\varepsilon_{0k+1}|\left\|\frac{\partial\varepsilon_{0k}}{\partial\psi}\right\| > \sqrt{n}\delta|\mathcal{F}_k\right)\right] = o_p(1). \tag{C.11}$$

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

Proof of Lemma 4.7. By a direct differentiation,

$$\frac{\partial^3 S(\lambda)}{\partial d^3} = 2\sum_{t=1}^{n}\left(3\frac{\partial^2\varepsilon_t}{\partial d^2}\frac{\partial\varepsilon_t}{\partial d} + \varepsilon_t\frac{\partial^3\varepsilon_t}{\partial d^3}\right), \tag{C.12}$$

$$\frac{\partial\varepsilon_t}{\partial d} = \phi(B)\psi^{-1}(B)\log(1-B)(1-B)^d y_t, \tag{C.13}$$

$$\frac{\partial^2\varepsilon_t}{\partial d^2} = \phi(B)\psi^{-1}(B)\log^2(1-B)(1-B)^d y_t, \tag{C.14}$$

$$\frac{\partial^3\varepsilon_t}{\partial d^3} = \phi(B)\psi^{-1}(B)\log^3(1-B)(1-B)^d y_t. \tag{C.15}$$

Note that

$$\begin{aligned}
\frac{\partial\varepsilon_t}{\partial d} &= (\phi - \phi_0)'\psi^{-1}(B)\psi_0(B)\log(1-B)(1-B)^{d-d_0}\mathbf{x}_t \\
&\quad + \log(1-B)(1-B)^{d-d_0}\varepsilon_{0t} \\
&\quad + [\psi^{-1}(B) - \psi_0^{-1}(B)]\log(1-B)(1-B)^{d-d_0}\psi_0(B)\varepsilon_{0t} \\
&= I_{1t} + I_{2t} + I_{3t}, \text{ say}, \tag{C.16}
\end{aligned}$$

where $\mathbf{x}_t$ is defined as in (4.6) with $x_t = \psi_0^{-1}(B)(1-B)^{d_0}y_t$. By the given condition, $|d - d_0| < n^{-\alpha}$ and $\|n^{-1/2}\tilde{G}_n'(\phi - \phi_0)\| < n^{-\alpha}$, where $\tilde{G}_n$ is defined after (2.6). Thus

$$\begin{aligned}
|I_{1t}| &= |[n^{-1/2}\tilde{G}_n'^{-1}(\phi - \phi_0)]'\log(1-B)(1-B)^{d-d_0}\psi(B)^{-1}\psi_0(B)[n^{1/2}\tilde{G}_n\mathbf{x}_t]| \\
&\leq n^{-\alpha}\|\log(1-B)(1-B)^{d-d_0}[\psi^{-1}(B)\psi_0(B)[n^{1/2}\tilde{G}_n\mathbf{x}_t]]\|.
\end{aligned}$$

By Lemma 2.1(d) in Ling (1998), $\max_{1\leq t\leq n} E\|\tilde{G}_n\mathbf{x}_t\|^2 = O(n^{-1})$. Furthermore, by Minkowski's inequality,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} EI_{1t}^2 &\leq \frac{1}{n}\sum_{t=1}^{n} n^{-\alpha} O\left(\sum_{i=1}^{t}\frac{1}{i}\sum_{k=1}^{i}k^{d-d_0-1}\right)\left(\max_{1\leq t\leq n} E\|n^{1/2}\tilde{G}_n\mathbf{x}_t\|^2\right)^{1/2} \\
&= \frac{1}{n}\sum_{t=1}^{n} O\left(n^{-\alpha}\sum_{i=1}^{t}\frac{1}{i}\sum_{k=1}^{i}k^{d-d_0-1}\right) \\
&\leq O\left(\sum_{i=1}^{n}\frac{1}{i^{1+\alpha/2}}\sum_{k=1}^{i}k^{|d-d_0|-1-\alpha/2}\right) = O(1). \tag{C.17}
\end{aligned}$$

Now, we consider the second term I_{2t} in (C.16). Let $\eta_t = (1 - B)^{\tilde{d}}\varepsilon_{0t}$, where $\tilde{d} = d - d_0$. By the given condition, $|\tilde{d}| < \frac{1}{2}$ for large enough n and hence by Theorem 1.1(d) in Hosking (1981), $E\eta_t^2 = $ a constant and for large enough k, $E(\eta_t\eta_{t-k}) = O(k^{2\tilde{d}-1})$. Thus

$$\frac{1}{n}\sum_{t=1}^{n} EI_{2t}^2 = E\left(\sum_{i=1}^{\infty}\frac{1}{i}\eta_{t-i}\right)^2$$

$$= \sum_{i=1}^{\infty}\frac{1}{i^2}E\eta_{t-i}^2 + 2\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\frac{1}{i(i+k)}E(\eta_{t-i}\eta_{t-i-k})$$

$$= O(1) + 2\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}O\left(\frac{1}{i(i+k)k^{1-2\tilde{d}}}\right)$$

$$\leq O(1) + 2\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}O\left(\frac{1}{i^{1+1/2}k^{1+1/2-2\tilde{d}}}\right) = O(1). \tag{C.18}$$

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

$$\frac{1}{n}\frac{\partial^3 S(\lambda)}{\partial d^3} = O_p(1). \tag{C.19}$$

Similarly we can show

$$\frac{1}{n}\frac{\partial^3 S(\lambda)}{\partial d^2\partial\psi} = O_p(1) \quad \text{and} \quad \frac{1}{n}\left[\sqrt{n}\tilde{G}_n\frac{\partial^3 S(\lambda)}{\partial d^2\partial\phi}\right] = O_p(1). \tag{C.20}$$

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

$$\frac{1}{n}\left[\frac{\partial^2 S(\lambda)}{\partial d^2} - \frac{\partial^2 S(\lambda_0)}{\partial d^2}\right] = \frac{1}{n}\left[\sqrt{n}\tilde{G}_n\frac{\partial^3 S(\lambda^*)}{\partial d^2\partial\phi}\right]'[n^{-1/2}\tilde{G}_n'^{-1}(\phi - \phi_0)]$$

$$+ \frac{1}{n}\frac{\partial^3 S(\lambda^*)}{\partial d^2\partial\psi'}(\psi - \psi_0) + \frac{1}{n}\frac{\partial^3 S(\lambda^*)}{\partial d^3}(d - d_0)$$

$$= O_p(\|n^{-1/2}J_n'^{-1}(\phi - \phi_0)\|) + O_p(\|\psi - \psi_0\|) + O_p(|d - d_0|)$$

$$= O_p(\|n^{-1/2}G_n'^{-1}(\lambda - \lambda_0)\|), \tag{C.21}$$

where λ^* is an intermediate point between λ and λ_0. Similar to the proof of (C.21), we can show that the following quantities are equal to $O_p(\|n^{-1/2}G_n'^{-1}(\lambda - \lambda_0)\|)$:

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

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

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

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