Hitotsubashi Journal of Economics 48 (2007), pp.25-55. © Hitotsubashi University

SEASONALLY AND FRACTIONALLY DIFFERENCED TIME SERIES*

Naoya Katayama

Faculty of Economics, Kyusyu University
Fukuoka, Fukuoka 812-8581, Japan
katayama@en.kyushu-u.ac.jp

Received December 2005; Accepted September 2006

Abstract

This paper presents a generalized seasonally integrated autoregressive moving average (SARIMA) model that allows the two differencing parameters to take on fractional values. We examine the asymptotic properties of the estimators and test statistics when the mean of the model is unknown. The findings show that standard asymptotic results hold for the tests and that the conditional sum of squares estimators are consistent and tends towards normality. The paper provides a modelling application using data on total power consumption in Japan.

Keywords: fractional differencing, Lagrange multiplier test, long memory, seasonal differencing, seasonal persistence.
JEL classification: C12, C13, C22, C50.

I. *Introduction*

In the past decade, there has been burgeoning interest in time series with strong dependence properties, especially hydrological and financial time series. These series generally have the property of slowly declining serial correlations, such that the sum of the absolute values of these correlations may diverge. In response, new classes of time series that have the property of strong dependence have been presented by Granger and Joyeux (1980), Hosking (1981), and Gray et al. (1989), which allow the differencing parameters to take on fractional values. Giraitis and Leipus (1995), Robinson (1994), and Woodward et al. (1998) generalized Gegenbauer autoregressive moving average (GARMA) models, known as k-factor GARMA (p, q) models, which allow the spectral density to be unbounded and peak at an arbitrary k with different frequencies of $\nu \in [0, \pi]$:

$$\phi(L)(1-L)^{d_1}(1+L)^{d_k}\prod_{i=2}^{k-1}(1-2\eta_i L+L^2)^{d_i}(x_t-\mu)=\theta(L)\varepsilon_t \tag{1}$$

where $\{\varepsilon_t\}$ is $iid(0, \sigma^2)$ and $\mathrm{E}[\varepsilon_t^4]<\infty$. The polynomials $\phi(z)=1-\Sigma_{i=1}^{p}\phi_i z^i$ and $\theta(z)=1+$

* The author thanks editors, an anonymous referee, and Professor Katsuto Tanaka for their valuable comments and suggestions. This work was supported by the Japan Society for the Promotion of Science.

$\sum_{i=1}^{q}\theta_i z^i$ have roots outside the unit circle. $\eta_i \equiv \cos(\nu_i)$ and $0=\nu_1<\nu_2<\cdots<\nu_{k-1}<\nu_k=\pi$. When $k=1$, it is known as the fractionally integrated autoregressive moving average model, or ARFIMA(p, d_1, q) for short, by Granger and Joyeux (1980) and Hosking (1981). Giraitis and Leipus (1995) and Woodward et al. (1998) analyzed the k-factor GARMA(p, q) model and showed that $\{x_t\}$ is stationary and invertible if $|d_i|<1/2$ for $i=1, ..., k$.

This paper investigates a special case of the k-factor GARMA model, which is considered by Porter-Hudak (1990) and naturally extends the seasonally integrated autoregressive moving average (SARIMA) model of Box and Jenkins (1976):

$$\phi(L)\Phi(L^s)(1-L)^{d_0}(1-L^s)^{d_s}(x_t-\mu)=\theta(L)\Theta(L^s)\varepsilon_t \tag{2}$$

where s is even, $\Phi(z^s)=1-\sum_{i=1}^{p_s}\Phi_i z^{is}$, $\Theta(z^s)=1+\sum_{i=1}^{q_s}\Theta_i z^{is}$ and $\phi(z)\Phi(z^s)=0$, $\theta(z)\Theta(z^s)=0$ have no roots in common and all roots are outside the unit circle. Since $(1-z)^a(1-z^s)^b=(1-z)^{a+b}(1+z)^b\prod_{j=1}^{s/2-1}(1-2\cos(2\pi j/s)z+z^2)^b$, the model (2) is a $(1+s/2)$-factor GARMA model, which allows the integration order to be a real number, and throughout this paper we refer to the fractional SARIMA$(p, d_0, q)(p_s, d_s, q_s)_s$ model as the SARFIMA or SARFIMA $(p, d_0, q)(p_s, d_s, q_s)_s$ for short.

In Section II, we explain the parameter estimation of the SARFIMA model, using the conditional sum of squares (CSS) method. It is shown that the CSS estimator is consistent and tends to normality. In Section III, testing procedures using residual autocorrelations such as the Lagrange multiplier (LM) test are shown. We also explore the asymptotic properties of the Wald test statistics. We note that Sections II and III impose the condition $\{x_t=\mu, t\leq 0\}$ to simplify the proof of asymptotic normality, but do not impose the conditions of normality of the model. The finite sample performance of these tests and the CSS estimators is examined in Section IV. Section V illustrates the use of the SARFIMA model. Section VI concludes.

Throughout this paper, let L be the lag operator, $\partial f(x)/\partial x|_{x=y}=\partial f(y)/\partial x$. In addition, 'RHS' abbreviates 'right-hand side', 'LHS' abbreviates 'left-hand side', and C_i, $i=1, 2, ...$, is used to denote universal appropriate positive constants to economize on notation. All proofs are given in the Appendix.

II. *Asymptotic Results for CSS Estimation*

In this section, we examine the asymptotic properties of the estimators of the nonstationary SARFIMA model, which is defined by

$$(1-L)^{d_0}(1-L^s)^{d_s}(x_t-\mu)=\vartheta(L)\varepsilon_t, \quad t\geq 1; \quad x_t=\mu, \quad t\leq 0, \tag{3}$$

where $\vartheta(L)=\theta(L)\Theta(L^s)/[\phi(L)\Phi(L^s)]$. We make the assumption that $\{x_t=\mu, t\leq 0\}$ in order to simplify the proof of asymptotic normality. Following Chung (1996) and Beran (1995), we use the sample mean as an estimator of μ, and the CSS method to estimate d_0, d_s, SARMA parameters, and σ^2. For the process $\{x_t\}$ in (3), we assume:

Assumption 1. (a) $\{\varepsilon_t\}_{t=1}^{\infty}$ is $iid(0, \sigma^2)$ and $\mathrm{E}[\varepsilon_t^4]<\infty$. (b) s is known and an even integer. (c) $(d_0, d_s)\in \boldsymbol{D}_{i,j}^s$ for some $i, j=1, 2, 3$ where $\boldsymbol{D}_{i,j}^s=\{(\ddot{d}_0, \ddot{d}_s)\,|\,(\ddot{d}_0+\ddot{d}_s, \ddot{d}_s)\in \boldsymbol{D}_i^s\times \boldsymbol{D}_j^s\}$, $\boldsymbol{D}_1^s=[\tau, 1/2-\tau]$, $\boldsymbol{D}_2^s=[\tau-1/4, 1/4-\tau]$, $\boldsymbol{D}_3^s=[\tau-1/2, -\tau]$, and $\tau\in(0, 1/4)$. (d) Let ϑ be $(\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \Phi_1, ..., \Phi_{p_s}, \Theta_1, ..., \Theta_{q_s})'$ and $\boldsymbol{D}_\vartheta$ be a compact space such that, for any $\vartheta\in \boldsymbol{D}_\vartheta$, $\phi(z)$, $\Phi(z^s)$,

$\theta(z)$, and $\Theta(z^s)$ satisfy conditions given in Section I. In addition, σ^2 is in the interior of the compact space contained in $\mathbb{R}^+$.

Since the model (3) assumes $x_t=\mu$ for $t\leq 0$, the SARFIMA model (3) is nonstationary. However, the model (3), which satisfies Assumption 1, is an approximate version of the stationary and noninvertible SARFIMA model as $t\to\infty$. This is because, when $|d_0+d_s|$, $|d_s|$ $<1/2$, the model (2) for $t=\ldots,\ -1,\ 0,\ 1,\ \ldots$, is stationary and noninvertible as shown by Woodward et al. (1999).

Given a process $\{x_t\}_{t=1}^T$ defined in (3), which satisfies Assumption 1, let δ be a true parameter vector $(d_0, d_s, \vartheta')'$, and let $\ddot{\delta}=(\ddot{d}_0, \ddot{d}_s, \ddot{\vartheta}')'$ be any parameter vector in the parameter space $D_{i,j}^s\times D_\vartheta$, where $(\ddot{d}_0, \ddot{d}_s)'$ is any vector in $D_{i,j}^s$, $\ddot{\vartheta}$ is any vector in D_ϑ, and assume that $\ddot{\delta}$ and δ are in the same compact parameter space defined by Assumption 1. Let $\bar{x}=\sum_{t=1}^T x_t/T$, and let $\pi_k(\ddot{\delta})$ be defined by $\sum_{k=0}^\infty \pi_k(\ddot{\delta})z^k=(1-z)^{\ddot{d}_0}(1-z^s)^{\ddot{d}_s}\ddot{\vartheta}(z)^{-1}$, where $\ddot{\vartheta}(z)$ be given by replacing ϑ in $\vartheta(z)$ by $\ddot{\vartheta}\in D_\vartheta$ in Assumption 1. Then the CSS estimator $(\hat{\delta}', \hat{\sigma}^2)'$ of $(\delta', \sigma^2)'$ is obtained by maximizing the CSS function:

$$S(\ddot{\delta}, \ddot{\sigma}^2)=-\frac{T}{2}\log 2\pi-\frac{T}{2}\log\ddot{\sigma}^2-\frac{1}{2\ddot{\sigma}^2}\sum_{t=1}^T\varepsilon_t^2(\ddot{\delta}), \qquad (4)$$

where $\varepsilon_t(\ddot{\delta})$ is defined by $\varepsilon_t(\ddot{\delta})=\varepsilon_t(\ddot{\delta}, \bar{x})=\sum_{k=0}^{t-1}\pi_k(\ddot{\delta})(x_{t-\mathrm{k}}-\bar{x})$.

Assumption 1 (c) is from Yajima (1985) where he proves strong consistency and asymptotic normality of maximum likelihood estimators (MLE) of the ARFIMA(0, d, 0) model with $d\in(0, 1/2)$. Using the techniques of Yajima's proof, we can prove the consistency of the CSS estimators when $(d_0, d_s)\in D_{1,1}^s$ (see Lemmas B 4 to B 8 in Appendix B) and extend this result to the case of any $D_{i,j}^s$ (see Lemma B 9 in Appendix B). Note that the deviation of the asymptotic distributions of δ's CSS estimator $\hat{\delta}$ is independent on that of σ^2, $\hat{\sigma}^2=\sum_{t=1}^T\varepsilon_t^2(\hat{\delta})/T$, which is obtained in the same way as the MLE for the ARMA model of Box and Jenkins (1976).

Then we have the following result.

Theorem 1. *Let $\hat{\delta}$ and $\hat{\sigma}^2$ be the CSS estimator of the parameter vector $(\delta', \sigma^2)'$ based on a sample $\{x_t\}_{t=1}^T$ given by (3) and Assumption 1. Then it follows that, as $T\to\infty$,*

$$\hat{\delta}\xrightarrow{p}\delta, \quad \hat{\sigma}^2\xrightarrow{p}\sigma^2, \qquad (5)$$

$$\text{and } \sqrt{T}(\hat{\delta}-\delta)\xrightarrow{d}N(0, I_\delta^{-1}), \quad \sqrt{T}(\hat{\sigma}^2-\sigma^2)\xrightarrow{d}N(0, 2\sigma^4+\kappa_4), \qquad (6)$$

where $\kappa_4=\mathrm{E}[\varepsilon_t]^4-3\sigma^4$,

$$I_\delta=\sum_{k=1}^\infty\delta_k\delta_k', \quad \text{and} \quad \frac{\partial\varepsilon_t(\delta, \mu)}{\partial\delta}=\sum_{k=1}^{t-1}\delta_k L^k\varepsilon_t, \qquad (7)$$

and each element of $\{\delta_k\}$ is given by (38) in the proof of Theorem 1.

The proof of Theorem 1 is given in Appendix B. Note that $\bar{x}\xrightarrow{a.c.}\mu$, $\mathrm{E}[\bar{x}-\mu]^2=O(T^{2(d_0+d_s)-1})$ by Lemma B 10, and if μ is known and $\bar{x}$ of $\varepsilon_t(\ddot{\delta})$ is replaced by μ, then $\hat{\delta}$ and $\hat{\sigma}^2$ are strongly consistent and asymptotic normality of (6) holds (see Remark 1).

For the simple case of the process in (3) with $p=p_s=1$, $q=q_s=0$, $\phi(L)=1-\phi L$, and $\Phi(L^s)=1-\Phi L^s$, I_δ can be written as

$$I_\delta = \begin{pmatrix} \pi^2/6 & \pi^2/(6s) & -\log(1-\phi)/\phi & -\log(1-\Phi)/(\Phi s) \\ \cdot & \pi^2/6 & -\log(1-\phi^s)/\phi & -\log(1-\Phi)/\Phi \\ \cdot & \cdot & 1/(1-\phi^2) & \phi^{s-1}/(1-\phi^s\Phi) \\ \cdot & \cdot & \cdot & 1/(1-\Phi^2) \end{pmatrix}. \tag{8}$$

III. *Tests Based on Residual Autocorrelation*

This section discusses testing for the integration order, namely, the LM test, which draws on LM tests for the integration order of the ARFIMA model by Robinson (1991), Robinson (1994), Agiakloglou and Newbold (1994), and Tanaka (1999). For the purposes of practical implementation, Godfrey's (1979) LM approach is also used. Finally, this section shows that the Wald test statistic has the same limiting local power as the LM test. Let $\{\hat{\varepsilon}_t\}$ be the residual sequence of the CSS estimator and let $\hat{r}(j)=\sum_{t=1}^{T-j}\hat{\varepsilon}_t\hat{\varepsilon}_{t+j}/\sum_{t=1}^{T}\hat{\varepsilon}_t^2$, $j=0, 1, ..., T-1$.

For the SARFIMA model, $\{x_t\}_{t=1}^{T}$, given by (3), we consider the testing problem of the null hypothesis H_0: SARFIMA$(p, d_0, q)(p_s, d_s, q_s)_s$ against the alternative

$$H_{A,1}: \text{SARFIMA}(p, d_0+\alpha_0, q)(p_s, d_s, q_s)_s \tag{9}$$

$$\text{or } H_{A,2}: \text{SARFIMA}(p, d_0, q)(p_s, d_s+\alpha_s, q_s)_s, \tag{10}$$

where the sets of the integration orders (d_0, d_s), $(d_0+\alpha_0, d_s)$, and $(d_0, d_s+\alpha_s)$ satisfy Assumption 1. The assumed null model is obtained by imposing the restrictions α_0 $(\alpha_s)=0$ and the alternatives are α_0 $(\alpha_s)>0$ and/or α_0 $(\alpha_s)<0$.

Under the testing problem H_0 against $H_{A,1}$, as in Tanaka (1999), let the CSS function be $S(\alpha_0, \xi, \sigma^2)$, where $\xi=(d_s, \vartheta')'$ is unknown vector, whereas d_0 is any preassigned value. Then the score-like test statistic is given by

$$S_T(\alpha_0|H_{A,1}) = \left.\frac{\partial S(\alpha_0, \xi, \sigma^2)}{\partial \alpha_0}\right|_{H_0:\alpha_0=0,\ \xi=\hat{\xi},\ \sigma^2=\hat{\sigma}^2}$$

$$= \frac{1}{\hat{\sigma}^2}\sum_{i=2}^{T}\left(\sum_{j=1}^{i-1}\frac{\varepsilon_{i-j}((d_0, \hat{\xi}')', \bar{x})}{j}\right)\varepsilon_i((d_0, \hat{\xi}')', \bar{x}) = T\sum_{i=1}^{T-1}\frac{\hat{r}(i)}{i} \tag{11}$$

where carets denote CSS estimators with the null hypothesis imposed.

Similarly, under the testing problem H_0 against $H_{A,2}$, we have the test statistic

$$S_T(\alpha_s|H_{A,2}) = \left.\frac{\partial S(\alpha_s, \xi, \sigma^2)}{\partial \alpha_s}\right|_{H_0:\alpha_s=0,\ \xi=\hat{\xi},\ \sigma^2=\hat{\sigma}^2} = T\sum_{i=1}^{[(T-1)/s]}\frac{\hat{r}(is)}{i}, \tag{12}$$

where $\xi=(d_0, \vartheta')'$ is unknown vector, whereas d_s is any preassigned value. This implies that the residuals $\{\hat{\varepsilon}_t\}$ are defined differently from (11).

To obtain potentially useful measures of power with a fixed significance level, we consider a sequence of local alternatives. Then we obtain the following results, which generalize Tanaka (1999, Theorem 3.3).

Theorem 2. *Under the testing problem H_0 against $H_{A,1}$ defined in (9) and $\alpha_0=c/\sqrt{T}$ with c fixed, it follows that, as $T\to\infty$,*

$$\frac{1}{\sqrt{T}}\frac{S_T(\alpha_0|H_{A,1})}{\sigma_{d_0}} \xrightarrow{d} N(c\sigma_{d_0}, 1) \tag{13}$$

where $S_T(\alpha_0|H_{A,1})$ *is defined in (11),* $\sigma_{d_0}=\sqrt{\sigma^2_{d_0}}$, *and* $1/\sigma^2_{d_0}$*is the* $(1, 1)$ *element of* I_δ^{-1} *defined in Theorem 1.*

Theorem 3. *Under the testing problem* H_0 *against* $H_{A,2}$ *defined in (10) and* $\alpha_s=c/\sqrt{T}$ *with c fixed, it follows that,* as $T\to\infty$,

$$\frac{1}{\sqrt{T}}\frac{S_T(\alpha_s|H_{A,2})}{\sigma_{d_s}} \xrightarrow{d} N(c\sigma_{d_s}, 1) \tag{14}$$

where $S_T(\alpha_s|H_{A,2})$ *is defined in (12),* $\sigma_{d_s}=\sqrt{\sigma^2_{d_s}}$, *and* $1/\sigma^2_{d_s}$ *is the (2, 2) element of* I_δ^{-1} *defined in Theorem 1.*

The proof of Theorem 3 is omitted since it can be obtained similarly to the proof of Theorem 2 in Appendix C. Note that a consistent estimator of σ_{d_0} or σ_{d_s} ($\hat{\sigma}_{d_0}$ or $\hat{\sigma}_{d_s}$) can be obtained by inserting the CSS estimator $\hat{\delta}$ into δ in I_δ. In addition, using $\varepsilon=(\varepsilon_1(\delta), \varepsilon_2(\delta), ..., \varepsilon_T(\delta))'$ and a $T\times(2+p+q+p_s+q_s)$ matrix $X=(\partial\varepsilon/\partial\delta')|_{H_0}$ with each (i, j) element of the partitioned matrix:

$$-\left(\sum_{k=1}^{i-1}\frac{\hat{\varepsilon}_{i-k}}{k} \;\Big|\; \sum_{k=1}^{\left[\frac{i-1}{s}\right]}\frac{\hat{\varepsilon}_{i-ks}}{k} \;\Big|\; \frac{\hat{\varepsilon}_{i-j}}{\hat{\phi}(L)} \;\Big|\; \frac{\hat{\varepsilon}_{i-j}}{\hat{\theta}(L)} \;\Big|\; \frac{\hat{\varepsilon}_{i-js}}{\hat{\Phi}(L^s)} \;\Big|\; \frac{\hat{\varepsilon}_{i-js}}{\hat{\Theta}(L^s)}\right)T, \quad \begin{matrix}1 & 1 & p & q & p_s & q_s\end{matrix} \tag{15}$$

where $(1,j)$ element is zero and $\hat{\varepsilon}_t=0$ for $t\le 0$, we can also obtain a consistent estimator of I_δ, $X'X/(T\hat{\sigma}^2)$ where $\hat{\sigma}^2=\Sigma_{t=1}^T\hat{\varepsilon}_t^2/T$.

Hence, we suggest the following test statistics:

$$S'_T(\alpha_0|H_{A,1})=\frac{S_T(\alpha_0|H_{A,1})}{\sqrt{T}\hat{\sigma}_{d_0}}, \quad \text{and} \quad S'_T(\alpha_s|H_{A,2})=\frac{S_T(\alpha_s|H_{A,2})}{\sqrt{T}\hat{\sigma}_{d_s}} \tag{16}$$

for the testing problems (9) and (10), respectively, which have a standard normal distribution under the null hypothesis. Hence, for example, for the testing problem of (9) with a right-sided alternative ($\alpha_0>0$), we can reject the null hypothesis when $S'_T(\alpha_0|H_{A,1})$ exceeds the upper 100 $a\%$ of $N(0, 1)$ for a test of asymptotic size a.

In many situations, researchers may wish to contemplate the following model:

$$y_t=\varphi_{t\cdot}\beta+x_t, \quad (1-L)^{d_0+\alpha_0}(1-L^s)^{d_s+\alpha_s}x_t=\vartheta(L)\varepsilon_t, \quad t\ge 1, \tag{17}$$

where $\{\varphi_{t\cdot}\}$ is a $1\times r$ sequences of fixed, nonstochastic variables, β is a $r\times 1$ unknown vector, (d_0, d_s) is any preassigned vector $(d_0, d_s>-1/2)$, and $\{x_t\}$ is a mean zero SARFIMA model. We assume that we observe $\{(y_t, \varphi_{t\cdot})\}_{t=1}^T$. The assumed null model H_0 is obtained by imposing the restrictions $\alpha\equiv(\alpha_0, \alpha_s)'=0$ and the alternative, $H_{A,3}$, is $\alpha\ne 0$.

To deduce the LM statistic, let the "differenced" model of (17) be $\tilde{y}_t=\tilde{\varphi}_{t\cdot}\beta+\tilde{x}_t(\alpha)$ and $y=\Phi\beta+x(\alpha)$, where $\tilde{y}_t=(1-L)^{d_0}(1-L^s)^{d_s}y_t$, $\tilde{\varphi}_{t\cdot}=(1-L)^{d_0}(1-L^s)^{d_s}\varphi_{t\cdot}$, $\tilde{x}_t(\alpha)=(1-L)^{d_0}(1-L^s)^{d_s}x_t$, $y=(\tilde{y}_1, ..., \tilde{y}_T)'$, $\Phi=(\tilde{\varphi}'_{1\cdot}, ..., \tilde{\varphi}'_{T\cdot})'$, and $x(\alpha)=(\tilde{x}_1(\alpha), ..., \tilde{x}_T(\alpha))'$. Then the least-squares estimator (LSE) of β is $\hat{\beta}=(\Phi'\Phi)^{-1}\Phi' y=\beta+(\Phi'\Phi)^{-1}\Phi' x(\alpha)$ and CSS estimates of $\hat{\vartheta}$ and $\hat{\sigma}^2$ are obtained by maximizing the CSS function $S((d_0, d_s, \ddot{\vartheta}')', \ddot{\sigma}^2)$ with the residual $\varepsilon_t(\ddot{\vartheta})=(1-L)^{d_0}(1-L^s)^{d_s}\ddot{\vartheta}(L)^{-1}\{y_t-\varphi_{t\cdot}\hat{\beta}\}=\ddot{\vartheta}(L)^{-1}\{\tilde{y}_t-\tilde{\varphi}_{t\cdot}\hat{\beta}\}$ under the null model. To investigate the large sample behaviour of LSE, let the (i, j) element of Φ be $\tilde{\varphi}_{i,j}$ and $D_T=\text{diag}\{(\Sigma_{i=1}^{T}\tilde{\varphi}_{i,1}^2)^{1/2}, ..., (\Sigma_{i=1}^{T}\tilde{\varphi}_{i,r}^2)^{1/2}\}=\text{diag}\{d_{T11}, ..., d_{Trr}\}$. Let

$$S_T=S_T(\alpha\,|\,H_{A,3})=T\left(\sum_{i=1}^{T-1}\frac{\hat{r}(i)}{i}\quad\sum_{i=1}^{[(T-1)/s]}\frac{\hat{r}(is)}{i}\right)' \tag{18}$$

where the $\{\hat{r}(i)\}$ are obtained by imposing the null hypothesis (i.e., $\{\hat{r}(i)\}$ are given by the residuals $\{\varepsilon_t(\hat{\vartheta})\}$. We assume:

Assumption 2. For the model in (17), (a) $\{x_t=y_t=0, \varphi_{t\cdot}=0, t\leq 0\}$. (b) Conditions (a), (b) and (d) in Assumption 1 hold, (d_0, d_s) is known, and $d_0, d_s>-1/2$ for the process $\{x_t\}$ in (17). (c) $\lim_{T\to\infty} d_{Tii}=\infty$, $i=1, 2, ..., r$. (d) $\lim_{T\to\infty} D_T^{-1}\Phi'\Phi D_T^{-1}=A$, where A is nonsingular.

We make Assumption 2 (a) in order to simplify the proof of asymptotic normality. Assumption 2 (b) ensures the assignment of any positive integration orders to be tested, e.g., a testing problem of a SARIMA model with $(d_0, d_s)=(1, 1)$. The role of Assumption 1 (c) corresponds to α, namely, $\alpha\in D_{2,2}^{s}$ under $H_{A,3}$ for appropriately small $\|\alpha\|$. In particular, $\{\tilde{x}_t(\alpha)\}$ is a SARMA model under H_0. Assumption 2 (c) and (d) are well-known conditions to investigate the large sample behaviour of $\hat{\beta}$ [see, e.g., Section 9.1 in Fuller (1996)].

Then we obtain the following theorem.

Theorem 4. *Under the testing problem H_0 against $H_{A,3}$ defined in (17) and Assumption 2, for an LM statistic S_T defined in (18) with $\alpha=c/\sqrt{T}$ where c is a 2×1 constant vector, it follows that, as $T\to\infty$,*

$$S_T'\Sigma^{-1}S_T/T \xrightarrow{d} \chi^2(2, c'\Sigma c), \tag{19}$$

where Σ^{-1} is a 2×2 partitioned matrix in the north-west corner of I_δ^{-1} defined in Theorem 1, and $\chi^2(m, \tau^2)$ denotes a noncentral chi-squared variable with m degrees of freedom and noncentrality parameter τ^2. This variable is given by the relation $\chi^2(m, \tau^2)=(Z_1+\tau)^2+\Sigma_{i=2}^{m}Z_i^2$, where $\{Z_i\}_{i=1}^{m}$ is iid $N(0, 1)$.

The detailed proof of this theorem is given in Appendix C. Results in Theorem 4 not only generalize Tanaka (Theorem 3.3, 1999) to the seasonal long memory case, but also coincide with Robinson (Theorem 4, 1994), which considers frequency-domain LM test statistics.

As discussed above, because the consistent estimator of Σ^{-1}, $\hat{\Sigma}^{-1}$ can be obtained, the test statistic,

$$\lambda_T(\alpha\,|\,H_{A,3})=S_T'\hat{\Sigma}^{-1}S_T/T \tag{20}$$

is asymptotically distributed as $\chi^2(2)$ when the null model H_0 is correct. Hence for the testing problem H_0 against $H_{A,3}$, we can reject the null hypothesis when $\lambda_T(\alpha\,|\,H_{A,3})$ exceeds the upper $100a$ % of $\chi^2(2)$ for a test of asymptotic size a.

Furthermore, for practical implementation, we can calculate $\lambda_T(\alpha\,|\,H_{A,3})$ by using Godfrey's auxiliary regression method. First, imposing the integration order of the null hypothesis, estimate SARMA parameters by the CSS method and calculate the residual vector $\hat{\varepsilon}=(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)'$ as the dependent variable. Next, substitute $\hat{\varepsilon}$ and the CSS estimates for the regressor X as in (15). Then conduct OLS regression and calculate the corresponding T times R^2 statistic, $T\hat{\varepsilon}'X(X'X)^{-1}X'\hat{\varepsilon}/\hat{\varepsilon}'\hat{\varepsilon}$, as $\lambda_T(\alpha|H_{A,3})$.

For an intuitive comparison with the limiting power envelope, we have the following result for the simplest model.

Corollary 1. *For the model,* $(1-L)^{d_0}x_t=\varepsilon_t$, *let* $\hat{\varepsilon}_t=x_t$ *and* $\{x_t=0, t\leq 0\}$. *Then it follows that, as* $T\to\infty$ *under* $d_0=c/\sqrt{T}$, $c>0$, *for an even integer s, and fixed but appropriately large m* such that $\sum_{i=1}^{m} i^{-2}\sim\pi^2/6$,

$$(A):\ \Pr\left(\sqrt{T}\sum_{i=1}^{T-1}\frac{\hat{r}(i)}{i}\Big/\sqrt{\frac{\pi^2}{6}}>z_a\right)\to\Pr\left(Z_1<-z_a+c\sqrt{\frac{\pi^2}{6}}\right),$$

$$(B):\ \Pr\left(\sqrt{T}\sum_{i=1}^{[(T-1)/s]}\frac{\hat{r}(is)}{i}\Big/\sqrt{\frac{\pi^2}{6}}>z_a\right)\to\Pr\left(Z_1<-z_a+\frac{c}{s}\sqrt{\frac{\pi^2}{6}}\right),$$

$$(C):\ \Pr\left(\frac{1}{T}S_T'\Sigma_2^{-1}S_T>\chi^2_{2,a}\right)\to\Pr\left(\chi^2\left(2,\frac{c^2\pi^2}{6}\right)>\chi^2_{2,a}\right),$$

$$(D):\ \Pr\left(T\sum_{i=1}^{m}\hat{r}^2(i)>\chi^2_{m,a}\right)\to\Pr\left(\chi^2\left(m,\frac{c^2\pi^2}{6}\right)>\chi^2_{m,a}\right),$$

where $Z_1\sim N(0,1)$, S_T *is defined by (18),* z_a *is the upper* $100a$ *percent point of* $N(0,1)$, $\chi^2_{m,a}$ *is the upper* $100a$ *percent point of a chi-squared variable with m degrees of freedom, and* Σ_2 *is a* 2×2 *partitioned matrix in the north-west corner of* I_δ.

Result (A) is due to Tanaka (1999, Corollary 3.1), who also shows that it is the locally best invariant test under the local alternative $d_0=c/\sqrt{T}$, $c>0$. We omit the proof since it follows from a slight modification to the proof of Theorem 2.

Corollary 2. *For the model,* $(1-L^s)^{d_s}x_t=\varepsilon_t$, *under the same conditions as in Corollary 1, it follows that, as* $T\to\infty$ *under* $d_s=c/\sqrt{T}$, $c>0$, *for an even integer s, and fixed but appropriately large m* such that $\sum_{i=1}^{[m/s]} i^{-2}\sim\pi^2/6$,

$$(A'):\ \Pr\left(\sqrt{T}\sum_{i=1}^{T-1}\frac{\hat{r}(i)}{i}\Big/\sqrt{\frac{\pi^2}{6}}>z_a\right)\to\Pr\left(Z_1<-z_a+\frac{c}{s}\sqrt{\frac{\pi^2}{6}}\right),$$

$$(B'):\ \Pr\left(\sqrt{T}\sum_{i=1}^{[(T-1)/s]}\frac{\hat{r}(is)}{i}\Big/\sqrt{\frac{\pi^2}{6}}>z_a\right)\to\Pr\left(Z_1<-z_a+c\sqrt{\frac{\pi^2}{6}}\right),$$

$$(C'):\ \Pr\left(\frac{1}{T}S_T'\Sigma_2^{-1}S_T>\chi^2_{2,a}\right)\to\Pr\left(\chi^2\left(2,\frac{c^2\pi^2}{6}\right)>\chi^2_{2,a}\right),$$

$$(D'):\ \Pr\left(T\sum_{i=1}^{m}\hat{r}^2(i)>\chi^2_{m,a}\right)\to\Pr\left(\chi^2\left(m,\frac{c^2\pi^2}{6}\right)>\chi^2_{m,a}\right).$$

The corollaries above relate to the situation in which a researcher doubts that the process is *iid* but cannot clearly determine what kind of long memory process applies. We note that the LHS of (A) through (D) (and (A') through (D')) corresponds to $S_T'(\alpha|H_{A,1})$, $S_T'(\alpha|H_{A,2})$, $\lambda_T(\alpha|H_{A,3})$, and the (modified) Portmanteau test statistic respectively. It seems that not only both (A) and (B') but also (B) and (A'), (C) and (C'), and (D) and (D') have the same

FIG. 1. RHS OF (A) THROUGH (C) IN COROLLARY 1 CHANGING s AND c WITH $a=0.95$

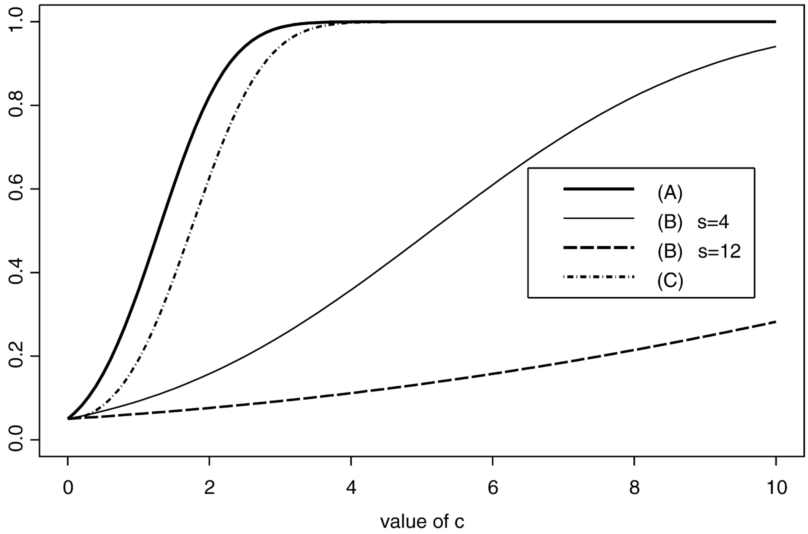
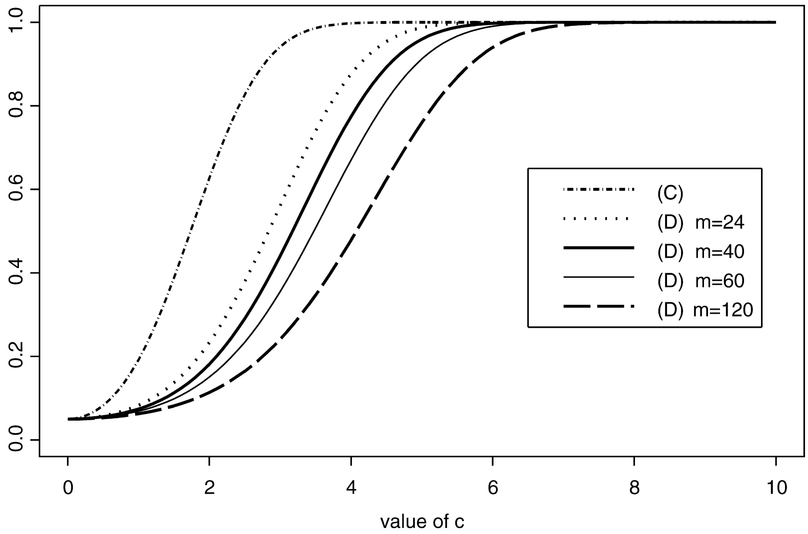


FIG. 2. RHS OF (C) AND (D) IN COROLLARY 1 CHANGING m AND c WITH $a=0.95$



limiting distribution.

Figures 1 and 2 illustrate the RHS of (A) through (D) changing s, m and c with $a=0.95$ by using S-PLUS. For a calculation of (C) and (D), we used Imhof's (1961) formula. It is apparent that (A) is uniformly most powerful in c, (C) is higher than various (D)s, (B) depends on the value of s and tends to (A) as s becomes small. It also indicates, for appropriately large s, that score-like test statistics from incorrect alternatives cannot detect the true long memory model, while correct ones can detect it with high power. Furthermore, (D) decreases as m increases. It also illustrates the difficulty of carrying out the (modified) Portmanteau test since the approximation of a chi-squared variable needs large m while power becomes low as m becomes large. On the whole, (C) has stable power compared to (A), (B), (A'), and (B') under the condition of Corollaries 1 and 2. Therefore, it seems reasonable to use LM test statistics to test for the integration order.

We can also derive the Wald test statistics, which have the same limiting local power as the LM test using the arguments of Remark 3. Let $(\tilde{d}_0, \tilde{d}_s)'$ be the unrestricted CSS estimators of $(d_0, d_s)'$ in (3) by maximizing the CSS function (4). Then it follows that, as $T \to \infty$,

$$\begin{aligned} W_{T,0} &= \sqrt{T}\,\sigma_{d_0}(\tilde{d}_0 - d_0) \xrightarrow{d} N(c\sigma_{d_0}, 1), \text{ under } H_{A,1} \text{ with } \alpha_0 = c/\sqrt{T}, \\ W_{T,S} &= \sqrt{T}\,\sigma_{d_s}(\tilde{d}_s - d_s) \xrightarrow{d} N(c\sigma_{d_s}, 1), \text{ under } H_{A,2} \text{ with } \alpha_s = c/\sqrt{T}, \\ W_{T,0S} &= T\begin{pmatrix} \tilde{d}_0 - d_0 \\ \tilde{d}_s - d_s \end{pmatrix}' \Sigma \begin{pmatrix} \tilde{d}_0 - d_0 \\ \tilde{d}_s - d_s \end{pmatrix} \xrightarrow{d} \chi^2(2, c'\Sigma c), \text{ under } H_{A,3} \text{ with } \alpha = c/\sqrt{T}, \end{aligned} \tag{21}$$

where σ_{d_0}, σ_{d_s}, and Σ are defined by Theorem 2, Theorem 3, and Theorem 4, respectively. The finite sample performance of these tests and the CSS estimates will be also be examined in the next section.

IV. *Some Simulations*

This section provides some evidence on the simulation results of the CSS estimation of the SARFIMA processes and the power of modified Portmanteau tests, LM tests, and Wald tests. All experiments are based on 1000 replications and in each replication, data series of size $T=100$ are generated. The calculations were conducted using S-PLUS. Here observations of both models were generated by Cholesky decomposition of the covariance matrix of the process [see Sections 11.3.1 and 11.3.5 of Beran (1994)]. We also performed some simulations using the Levinson-Durbin algorithm and obtained essentially the same results as those using the Cholesky decomposition. In addition, the Gauss-Newton procedure was used for the maximization of the CSS functions, the procedures of which are provided in Tanaka (1999, Section 5).

1. Results on CSS Estimates

The models employed here are

$$\text{DGP 1: } (1-\phi L)(1-L)^{d_0}(1-L^{12})^{d_s}(x_t - 1) = \varepsilon_t,$$

$$\text{and DGP 2: } (1-\Phi L^{12})(1-L)^{d_0}(1-L^{12})^{d_s}(x_t - 1) = \varepsilon_t.$$

TABLE 1. SIMULATION ON THE ESTIMATION OF SARFIMA(1, d_0, 0)(0, d_s, 0)$_s$ PROCESSES

True value			Simulation results				
d_0	d_s	ϕ	μ	d_0	d_s	ϕ	σ^2
0.35	0.10	0.80	−0.2011	−0.0910	−0.0332	0.0177	0.2181
			(5.8311)	(0.2376)	(0.1018)	(0.1648)	(1.5346)
				[0.2195]	[0.0962]	[0.1638]	[1.5190]
				{0.2327}	{0.0787}	{0.1786}	{0.1414}
0.35	−0.10	0.80	−0.1471	−0.0923	−0.0075	0.0202	0.4389
			(2.1311)	(0.2622)	(0.0919)	(0.1833)	(5.3401)
				[0.2454]	[0.0916]	[0.1822]	[5.3220]
−0.35	0.30	0.80	−0.0012	−0.0864	0.0352	0.0208	0.8343
			(0.1926)	(0.2464)	(0.0998)	(0.1725)	(14.7717)
				[0.2308]	[0.0934]	[0.1712]	[14.7481]
0.35	0.10	−0.80	0.0094	−0.0604	−0.0276	0.0259	−0.0234
			(0.6519)	(0.1218)	(0.1011)	(0.0808)	(0.1417)
				[0.1058]	[0.0973]	[0.0765]	[0.1398]
				{0.0835}	{0.0785}	{0.0641}	{0.1414}
0.35	−0.10	−0.80	−0.0161	−0.0595	−0.0279	0.0320	−0.0290
			(0.2259)	(0.1192)	(0.1031)	(0.0827)	(0.1418)
				[0.1033]	[0.0993]	[0.0763]	[0.1388]
−0.35	0.30	−0.80	−0.0007	−0.0103	0.0282	0.0211	0.0803
			(0.0292)	(0.1031)	(0.0983)	(0.0738)	(0.1988)
				[0.1026]	[0.0942]	[0.0707]	[0.1819]

DGP 1: $(\mu, \sigma^2, s) = (1.00, 1.00, 12)$

Tables 1 and 2 examine the finite sample performance of the estimates discussed in Section II. For each simulated data series, the sample mean, $\bar{x}$, is calculated and subtracted from the data points before the CSS method is applied to obtain the other parameter estimates. For each cell of five columns denoted "Simulation results" in the Tables, the first number is the estimation bias, the number in parentheses is the square root of the mean squared error (SRMSE), the number in brackets is the mean of the asymptotic standard squared errors (MASE),[1] and the number in braces is the true asymptotic standard error (TASE). For the CSS estimates, TASE is computed from Theorem 1. We omitted TASE for some cells since it does not depend on the integration order. The results are quite similar to those obtained by Chung and Baillie (1993) for the ARFIMA case. Since $\bar{x} - \mu = O_p(T^{d_0+d_s-1/2})$ by Lemma B 10 and Leipus and Viano (2000, Lemma 9), the rate of convergence of $\bar{x}$ for true μ depends on the value of $d_0 + d_s$, and the columns of μ reflect this. Estimation bias and SRMSE of $\bar{x}$ gets smaller as $d_0 + d_s$ gets smaller. For the CSS estimates, in this case, if $\phi = \Phi$, we find that both the Fisher information matrix of $(\hat{d}_0 - d_0, \hat{\phi} - \phi)'$ and $(\hat{d}_s - d_s, \hat{\Phi} - \Phi)'$ have the same elements by (8). It follows that the value of TASE in Table 1 is comparable to the corresponding TASE in Table 2. It is also apparent that the MASE and SRMSE in Table 1 and those in Table 2 are similarly symmetrical. Roughly speaking, if we ignore the elements $-\log(1-\phi^s)/\phi$ and $-\log(1-\Phi)/(\Phi s)$ in (8), the TASE of d_0 in Table 1 and d_s in Table 2 correspond to the results

[1] Given the estimate a_j for the true parameter a from the jth simulation trial and the average $\bar{a}$ of a_j, $j = 1, \ldots, 1000$, bias is defined as $\bar{a} - a$, while MASE is the square root of $\Sigma_{j=1}^{1000}(a_j - \bar{a})^2/1000$. The SRMSE is the square root of $\Sigma_{j=1}^{1000}(a_j - a)^2/1000$, which is equal to the square root of $(\text{bias})^2 + (\text{MASE})^2$.

TABLE 2. SIMULATION ON THE ESTIMATION OF SARFIMA$(0, d_0, 0)(1, d_s, 0)_s$ PROCESSES

True value			Simulation results				
d_0	d_s	Φ	μ	d_0	d_s	Φ	σ^2
0.35	0.10	0.80	0.0778	−0.0343	−0.0321	−0.0169	0.4140
			(5.5191)	(0.1201)	(0.2923)	(0.2453)	(1.1311)
				[0.1151]	[0.2905]	[0.2447]	[1.0526]
				{0.0782}	{0.2308}	{0.1775}	{0.1414}
0.35	−0.10	0.80	0.0194	−0.0264	0.0204	−0.0508	0.2242
			(1.7547)	(0.1088)	(0.3307)	(0.2926)	(1.2578)
				[0.1056]	[0.3301]	[0.2882]	[1.2377]
−0.35	0.30	0.80	−0.0045	−0.0676	−0.1502	0.0427	1.1938
			(0.1443)	(0.1616)	(0.2563)	(0.1758)	(1.3322)
				[0.1468]	[0.2077]	[0.1705]	[0.5912]
0.35	0.10	−0.80	0.0074	−0.0509	−0.0455	0.0163	0.1405
			(0.6598)	(0.1083)	(0.1127)	(0.0825)	(0.2307)
				[0.0956]	[0.1031]	[0.0809]	[0.1830]
				{0.0782}	{0.0833}	{0.0639}	{0.1414}
0.35	−0.10	−0.80	−0.0046	−0.0513	0.0142	0.0131	0.2162
			(0.2507)	(0.1112)	(0.0968)	(0.0773)	(0.2984)
				[0.0987]	[0.0958]	[0.0762]	[0.2056]
−0.35	0.30	−0.80	0.0000	−0.0070	0.0036	0.0088	0.1617
			(0.0330)	(0.0979)	(0.1017)	(0.0783)	(0.2372)
				[0.0976]	[0.1016]	[0.0778]	[0.1735]

DGP 2: $(\mu, \sigma^2, s)=(1.00, 1.00, 12)$

of Tanaka (1999, Table 9). It reveals not only a poor performance of the CSS estimates depending on some of the SARMA parameters but also reveals an unstable limiting power of LM tests for the integration order, which is considered in the next subsection.

2. Testing for the Integration Order

Next we examine testing the AR(1) or SAR(1) model against the following DGP 3-6:

$$\text{DGP 3:}\quad (1-\vartheta L)(1-L)^{\alpha}x_t=\varepsilon_t,\qquad \text{DGP 4:}\quad (1-\vartheta L)(1-L^{12})^{\alpha}x_t=\varepsilon_t,$$

$$\text{DGP 5:}\quad (1-\vartheta L^{12})(1-L)^{\alpha}x_t=\varepsilon_t,\qquad \text{DGP 6:}\quad (1-\vartheta L^{12})(1-L^{12})^{\alpha}x_t=\varepsilon_t,$$

where we fixed $\vartheta=0.8$ or -0.8 and assumed $E[x_t]=0$ is known. Tables 3 and 4 are concerned with the rate of rejection of the null hypothesis $\alpha=0$ of no long memory.

In Table 3, the statistics $S_{T,0}$ and $S_{T,S}$ are, respectively, LM statistics defined from (16):

$$S_{T,0}=\sum_{i=1}^{T-1}\sqrt{\frac{T(T+2)}{T-i}}\,\frac{\hat{r}(i)}{i\hat{\sigma}_{d_0}},\quad S_{T,S}=\sum_{i=1}^{[(T-1)/s]}\sqrt{\frac{T(T+2)}{T-is}}\,\frac{\hat{r}(is)}{i\hat{\sigma}_{d_s}},$$

where $\hat{\sigma}_{d_0}$ and $\hat{\sigma}_{d_s}$ are computed from (15). These have the same asymptotic results in Theorems 2 and 3 by (40). The statistics $\lambda_{T,0S}$ are also LM test statistics, obtained using Godfrey's TR^2 statistics, which are asymptotically distributed as (20). The statistics Q_{24}^* and Q_{40}^* denote modified Portmanteau test statistics, which are assumed to be asymptotically chi-squared with 24 and 40 degrees of freedom, respectively, under the null hypothesis. The number in parentheses denotes the theoretical limiting power derived from Theorems 2-4. The general

TABLE 3. THE RATE OF REJECTION OF THE NULL HYPOTHESIS $\alpha=0$ FOR DGP 3-6 AT THE 5% LEVEL

$\vartheta=$			0.8					−0.8		
$\alpha=$	0	0.05	0.10	0.15	0.20	0	0.05	0.10	0.15	0.20
DGP 3										
Q^*_{24}	6.5	6.5	6.3	7.8	7.6	6.7	7.5	12.6	24.3	40.5
Q^*_{40}	6.8	6.1	7.7	8.9	9.9	7.7	9.7	14.0	24.5	39.1
$S_{T,0}$	5.3	7.7	14.1	18.4	25.4	3.1	10.7	30.3	51.4	70.9
	(5.0)	(7.7)	(11.3)	(16.0)	(21.8)	(5.0)	(14.9)	(33.0)	(56.4)	(77.8)
$S_{T,S}$	6.5	5.3	7.1	6.0	5.6	6.9	7.5	9.5	15.4	22.1
$\lambda_{T,0S}$	4.1	4.4	6.7	8.1	11.2	4.7	5.9	13.8	32.9	51.6
	(5.0)	(5.4)	(6.4)	(8.3)	(11.1)	(5.0)	(7.8)	(17.4)	(34.7)	(56.9)
DGP 4										
Q^*_{24}	6.5	8.5	10.5	19.2	35.0	6.7	8.0	10.9	18.8	33.8
Q^*_{40}	6.8	9.7	11.9	20.9	35.2	7.7	10.0	11.5	20.8	34.7
$S_{T,0}$	4.7	7.1	8.6	9.1	10.7	2.9	4.1	5.9	6.0	7.9
$S_{T,S}$	3.8	16.2	34.5	59.0	80.6	3.6	15.3	34.8	58.2	79.9
	(5.0)	(15.8)	(35.8)	(60.9)	(82.1)	(5.0)	(15.8)	(35.8)	(60.9)	(82.1)
$\lambda_{T,0S}$	4.1	6.2	15.7	31.2	56.6	4.7	5.9	14.1	30.9	57.7
	(5.0)	(8.2)	(19.2)	(38.7)	(62.6)	(5.0)	(8.2)	(19.2)	(38.7)	(62.6)
DGP 5										
Q^*_{24}	6.1	6.8	15.5	26.6	48.8	5.2	5.8	12.1	25.3	40.2
Q^*_{40}	6.2	4.8	11.4	22.4	44.5	5.6	5.7	12.9	22.4	35.8
$S_{T,0}$	4.7	18.0	37.7	62.3	82.5	5.0	13.3	30.4	53.3	72.3
	(5.0)	(15.7)	(35.7)	(60.8)	(81.9)	(5.0)	(15.8)	(35.8)	(61.0)	(82.1)
$S_{T,S}$	5.5	5.0	6.0	5.4	8.2	5.3	6.8	7.5	11.6	17.4
$\lambda_{T,0S}$	5.4	6.4	22.1	40.5	65.8	4.1	6.7	17.0	32.3	53.2
	(5.0)	(8.2)	(19.1)	(38.6)	(62.4)	(5.0)	(8.2)	(19.2)	(38.7)	(62.6)
DGP 6										
Q^*_{24}	6.1	3.8	5.5	6.4	6.9	5.2	5.0	7.5	9.8	20.4
Q^*_{40}	6.2	3.5	8.0	4.9	4.7	5.6	4.5	6.1	9.7	19.1
$S_{T,0}$	4.7	5.5	6.4	6.6	8.4	5.0	4.1	6.3	7.5	7.4
$S_{T,S}$	5.5	4.2	4.0	6.0	4.0	5.3	12.1	25.4	42.3	63.6
	(5.0)	(7.7)	(11.3)	(16.0)	(21.8)	(5.0)	(14.9)	(33.0)	(56.4)	(77.8)
$\lambda_{T,0S}$	5.4	3.3	8.9	7.0	8.5	4.1	4.8	8.9	18.6	36.5
	(5.0)	(5.4)	(6.4)	(8.3)	(11.1)	(5.0)	(7.8)	(17.4)	(34.7)	(56.9)

feature of Table 3 is that the modified Portmanteau test statistics perform poorly. $S_{T,0}$ or $S_{T,S}$ is the most powerful if an alternative model is correctly specified, while the other is the least powerful. The powers of $\lambda_{T,0S}$ are monotonically increasing in each case, though it is not the most powerful. It is similar to the corollaries in Section III. It is also worth noting that, and as in Tanaka (1999), the discrepancy between the finite sample and limiting powers is related to the fact that, by (8), the estimators of α and ϑ are negatively correlated, and the correlation is much higher for the case of $(\alpha, \vartheta)=(d_0, \phi)$ (and$=(d_s, \Phi)$) with $\vartheta=0.8$ than for the other cases. In these cases, LM statistics have not only quite low limiting powers but also a large discrepancy between a finite sample and these limiting powers.

Finally, in Table 4, we conducted LM test statistics $\lambda_{T,k}$ assuming alternatives, 7-factor GARMA models with $\nu_j=(j-1)\pi/6$, $j=1$, ..., 7, which is considered by Silvapulle (2001). We also conducted the Wald test statistics $W_{T,0}$, $W_{T,S}$ and $W_{T,0S}$ defined from (21). To compute consistent estimators of σ_{d_0}, σ_{d_s}, and Σ, we used a Hessian (the second-order

TABLE 4. THE RATE OF REJECTION OF THE NULL HYPOTHESIS $\alpha=0$ FOR DGP 3-6 AT THE 5% LEVEL

$\vartheta=$			0.8					0.8		
$\alpha=$	0	0.05	0.10	0.15	0.20	0	0.05	0.10	0.15	0.2
			DGP3					DGP5		
$W_{T,0}$	7.0	10.4	11.2	16.9	22.7	4.8	15.4	37.7	58.8	80.3
	(5.0)	(7.7)	(11.3)	(16.0)	(21.8)	(5.0)	(15.7)	(35.7)	(60.8)	(81.9)
$W_{T,0S}$	11.1	11.8	11.4	13.7	14.3	10.6	12.2	26.5	44.5	67.8
	(5.0)	(5.4)	(6.4)	(8.3)	(11.1)	(5.0)	(8.2)	(19.1)	(38.6)	(62.4)
$\lambda_{T,k}$	2.7	3.1	4.8	4.5	5.5	4.3	5.4	13.5	30.4	50.7
	(5.0)	(5.2)	(5.7)	(6.6)	(7.9)	(5.0)	(6.5)	(11.9)	(23.2)	(41.2)
			DGP4					DGP6		
$W_{T,S}$	3.4	12.0	28.7	47.0	74.5	4.0	4.5	6.5	7.1	9.6
	(5.0)	(15.8)	(35.8)	(60.9)	(82.1)	(5.0)	(7.7)	(11.3)	(16.0)	(21.8)
$W_{T,0S}$	11.1	14.1	22.1	37.2	59.4	10.6	10.5	11.7	12.1	12.8
	(5.0)	(8.2)	(19.2)	(38.7)	(62.6)	(5.0)	(5.4)	(6.4)	(8.3)	(11.1)
$\lambda_{T,k}$	2.7	4.8	7.4	15.5	32.6	4.3	5.1	5.8	9.3	12.1
	(5.0)	(6.5)	(12.0)	(23.3)	(41.4)	(5.0)	(5.2)	(5.7)	(6.6)	(7.9)

derivative) matrix from the Gauss-Newton procedure (see Tanaka, 1999, Section 5). The statistics $W_{T,0}$ and $W_{T,S}$ perform similarly to $S_{T,0}$ and $S_{T,S}$, respectively. The statistics $W_{T,0S}$ and $\lambda_{T,k}$ also perform similarly to $S_{T,0S}$.

It implies that the impact of SARMA parameters on integration orders is quite complicated so that the LM test and the Wald test may perform poorly for testing for the integration order of the SARFIMA model without strong evidence of SARMA parameters when the sample size is 100.

V. *An Example Using Japanese Total Power Consumption*

As an illustration of the use of the SARFIMA model, we consider monthly total power consumption data in Japan from the Federation of Electric Power Companies (FEPC) between January 1995 and December 2004 (sum of the 10 electric power companies, unit: MWh, sample size: 120).[2] Since the storage of a large amount of electricity is impossible, we can regard total power consumption as electric energy demand. A large number of statistical and numerical methods have been applied to modelling Japanese electric energy demand and total power consumption data, including, amongst others, (non)linear regression, Box-Jenkins SARIMA models and neural networks [see Yamamoto (1988) and Honda (2000) and references therein]. One efficient method is SARIMA modelling, however residual analysis by Yamamoto (1988, Section 7.6) and Honda (2000, Section 11.2) provides evidence of cyclical behaviour around the peak and bottom, and the modelling results are generally unsatisfactory.

Figure 3 displays the total power consumption data, $\{x_t\}$. Figure 4 displays the autocorrelation function (ACF) of the transformed data $\{x_t\}$. Note that the ACF decays very slowly and exhibits cyclical behaviour.

To search for the best representation of this data, we first fitted differenced data $y_t=$

[2] These data are available from the website of the FEPC: http://www.fepc.or.jp/.

FIG. 3. JAPANESE TOTAL POWER CONSUMPTION DATA $\{x_t\}$, JANUARY 1995 TO DECEMBER 2004 (sum of the ten electric power companies, unit: MWh, sample size: 120)

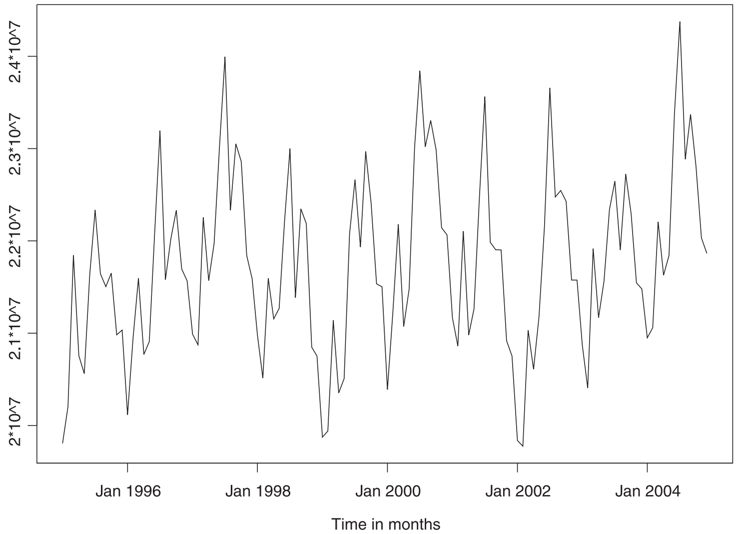


FIG. 4. THE SAMPLE AUTOCORRELATION FUNCTION (ACF) OF THE TRANSFORMED SERIES, WHERE A IS $\{x_t\}$, B IS $\{(1-L)x_t\}$, C IS $\{(1-L^{12})x_t\}$, AND D IS $\{(1-L)(1-L^{12})x_t\}$. DOTTED LINES ARE APPROXIMATE 95% CONFIDENCE LIMITS OF THE ACF OF THE WHITE NOISE RANDOM VARIABLE.

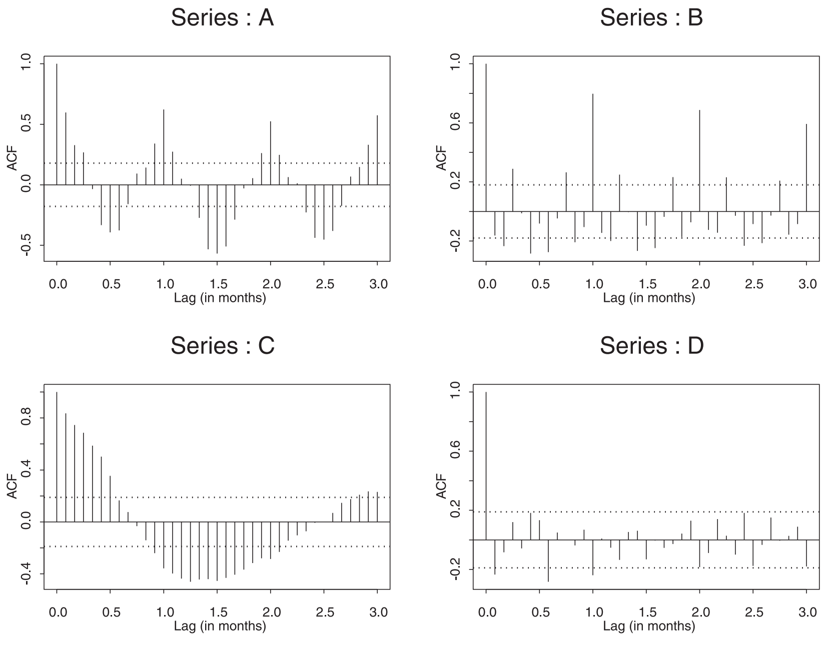


TABLE 5. SUMMARY OF AIC AND BIC MODEL SELECTION AND ESTIMATES

ID	AIC	BIC	d_0	d_s	ϕ_1	θ_1	Φ_1	Θ_1	Θ_2	$\sigma^2(\times 10^{11})$
50	(1) 2933.9	(2) 2944.6	−0.259	NE	NE	NE	NE	−0.510	−0.192	1.195
33	(2) 2934.2	(3) 2944.8	NE	NE	NE	−0.298	NE	−0.498	−0.196	1.198
49	(3) 2934.8	(4) 2945.5	−0.263	NE	NE	NE	0.222	−0.769	NE	1.205
29	(4) 2935.1	(5) 2945.8	NE	NE	NE	−0.306	0.230	−0.765	NE	1.209
32	(5) 2935.9	(7) 2946.6	NE	NE	−0.245	NE	NE	−0.481	−0.217	1.218
54	(6) 2936.0	(1) 2944.0	NE	−0.487	NE	−0.305	NE	NE	NE	1.242

$\bar{y}=697.729$

$(1-L)(1-L^{12})\,x_t$ by the CSS method, where we used a sample mean of $\{y_t\}$, $\bar{y}$ as an estimator of $E[y_t]=\mu$, and set $s=12$. AIC and BIC criteria are also used under the assumption of normality [see, e.g., Brockwell and Davis (1991, Section 9.3)]. Calculations of AIC and BIC are given by $-2S(\hat{\delta}, \hat{\sigma}^2)+2$ (number of estimated parameters) and $-2S(\hat{\delta}, \hat{\sigma}^2)+\log$(sample size used for CSS estimation)$\times$(number of estimated parameters), respectively. Fitting SARFIMA models or SARIMA models is limited to having SARMA parameters with $0\leq p$, $q, p_s, q_s \leq 3$, and where the total number of estimated SARFIMA parameters (d_0, d_s, SARMA parameters, and σ^2) is less than 4. The total number of models is 70. From among these estimation results, we selected models in terms of AIC and BIC that satisfy the following conditions: (i) Modified portmanteau tests are not rejected with the significance level 5% and 10 to 30 degrees of freedom. (ii) The estimated SARFIMA parameters all converged and satisfy Assumption 1 (c) and (d). All calculations were made using S-PLUS.[3]

Table 5 shows the best six models in terms of AIC model selection with estimators. ID denotes the model identification within 70 models. NE indicates the corresponding parameter is not estimated and is set to be 0. The numbers in parentheses in the columns of AIC (BIC) denote the ranking of models in terms of AIC (BIC). These six models show that similar models are selected. Our main concern is whether the $\{x_t\}$ is overdifferenced (ID 50 and ID 49) or not overdifferenced (ID 33, ID 29, and ID 32) because the estimator of d_0 in ID 50 (ID 49) appears to relate to the estimator of θ_1 in ID 33 and ϕ_1 in ID 32 (θ_1 in ID 29). Also, we check whether the $\{x_t\}$ is seasonally overdifferenced (ID 54) or not seasonally overdifferenced (ID 33 and ID 29) because the estimator of d_s in ID 52 appears to relate to the estimator of Θ_1 and Θ_2 in ID 33 (Φ_1 and Θ_1 in ID 29).

Table 6 shows the p-values for testing the integration order corresponding to these six models using the LM test statistics in Section III. In each cell of three columns denoted "Alternative hypotheses", the first number denotes the p-value of the LM test statistics when $E[y_t]=\mu$ is estimated by the sample mean and the number in parentheses denotes the p-value of the LM test statistics when y_t is a linear regression model including deterministic seasonality [see, e.g., Ghysels and Osborn (2001, Section 2.2)]:

$$y_t=\mu+\sum_{k=1}^{6}\left[\alpha_k\cos\left(\frac{2\pi kt}{12}\right)+\beta_k\sin\left(\frac{2\pi kt}{12}\right)\right]+z_t, \tag{22}$$

where $\beta_6=0$ and $\{z_t\}$ is the SARFIMA$(p, \alpha_0, q)(p_s, \alpha_s, q_s)_s$ model and μ, $\{\alpha_k\}$ and $\{\beta_k\}$ are estimated by LSEs. NA denotes a p-value that is not calculated because the estimated SARMA

[3] These programs are available on request.

TABLE 6. P-VALUES OF TESTING FOR $\alpha_0=\alpha_s=0$ OF THE SARFIMA MODELS

Model	Alternative hypotheses		
	$\alpha_0<0, \alpha_s=0$	$\alpha_0=0, \alpha_s<0$	$\alpha_0\neq 0, \alpha_s\neq 0$
SARFIMA$(0, \alpha_0, 0)(0, \alpha_s, 2)_s$	0.0022	0.8133	0.0162(0.0148)
SARFIMA$(0, \alpha_0, 1)(0, \alpha_s, 2)_s$	0.3787	0.5743	0.7579(0.8703)
SARFIMA$(0, \alpha_0, 0)(1, \alpha_s, 1)_s$	0.0000	0.7751	0.0042(NA)
SARFIMA$(0, \alpha_0, 1)(1, \alpha_s, 1)_s$	0.3814	0.6046	0.2844(0.9282)
SARFIMA$(1, \alpha_0, 0)(0, \alpha_s, 2)_s$	0.1327	0.5775	0.3383(0.3923)
SARFIMA$(0, \alpha_0, 1)(0, \alpha_s, 0)_s$	0.4508	0.0001	0.0001(0.0000)

parameters do not satisfy Assumption 1 (d). In this table, models ID 50, ID 49 and ID 54 correspond to some models in alternative hypotheses of the first, third and sixth rows of SARFIMA models, and models ID 33, ID 29 and ID 32 correspond to null hypotheses of the second, fourth and fifth rows of SARFIMA models. Our findings are as follows: (i) Results for SARFIMA$(0, \alpha_0, 0)(0, \alpha_s, 2)_s$, SARFIMA$(0, \alpha_0, 0)(1, \alpha_s, 1)_s$, and SARFIMA$(0, \alpha_0, 1)(0, \alpha_s, 0)_s$ support the estimation of d_0 or d_s for models ID 50, ID 49 and ID 54. (ii) Except for SARFIMA$(0, \alpha_0, 1)(0, \alpha_s, 0)_s$, results for SARFIMA models show large p-values for the alternative $\alpha_0=0$, $\alpha_s<0$. (iii) Results for some SARFIMA models show relatively small p-values for the alternative $\alpha_0<0$, $\alpha_s=0$ and $\alpha_0\neq 0$, $\alpha_s\neq 0$. Therefore, we cannot conclude that $\{x_t\}$ is not overdifferenced and d_0 should be set to zero.

Model ID 50 is the best model in terms of AIC among the 70 model candidates. The estimated model of ID 50 is

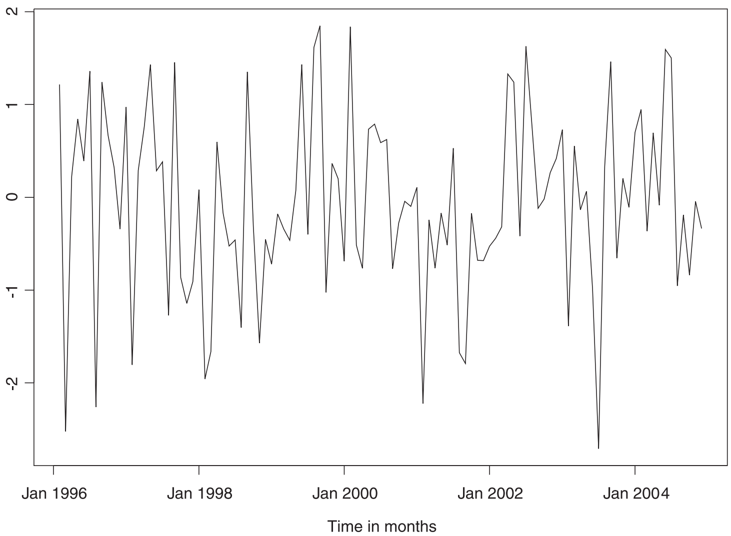
$$(1-L)^{-0.259}(y_t-697.729)=(1-0.510L^{12}-0.192L^{24})\varepsilon_t,$$
$$y_t=(1-L)(1-L^{12})\,x_t, \quad \text{and} \quad \hat{\sigma}^2=1.195\times 10^{11}.$$

Figure 5 shows the standardized residuals of Japanese total power consumption data using this model. The behaviour of this residual sequence resembles a white noise sequence and presents no cyclical pattern.

In place of the sample mean, we specified the sample median because electric energy demand can be affected by excessive changes in air temperature and the sample median is robust to additive outliers. In this case, model ID 33 (a SARIMA model) is selected as the best model in terms of AIC and model ID 54 (a SARFIMA model) is selected as the best model in terms of BIC among the 70 candidates; here the rankings and estimates are similar to those in Table 5. We also considered the time series regression model (22), however, most models are rejected because the estimators do not satisfy Assumption 1 (c) and (d). Nonetheless, models ID 29 and ID 33 (SARIMA models) and model ID 50 (a SARFIMA model) in Table 5 are selected as the best three models in terms of AIC and BIC among the 70 candidates. Note that we also conducted other transformed series $\{(1-L)x_t\}$ and $\{(1-L^{12})x_t\}$. However, the best of these were inferior to models by the series $\{(1-L)(1-L^{12})x_t\}$ in terms of AIC and BIC.

On this basis, we conclude that the SARFIMA model is effective and can be usefully employed as a substitute for the SARIMA model when fitting Japanese total power consumption data.

FIG. 5. STANDARDIZED RESIDUALS FROM THE SARFIMA$(0, 1+d_0, 0)(0, 1, 2)_s$ MODEL (model ID: 50) BASED ON JAPANESE TOTAL POWER CONSUMPTION DATA



VI. *Concluding Remarks*

This paper has examined a seasonal long memory process, denoted as the SARFIMA model. The paper provides evidence of the consistency and asymptotic normality of CSS estimates and the testing procedures of two differencing parameters.

This paper is based on parts of Chapters 1 and 2 in the author's Ph.D. thesis [Katayama (2004a)]. Sections II and III in this paper are an extension of the results of the author's Ph. D. thesis to the case of unknown mean, and can be applied to the k-factor model, though we must assume that Gegenbauer frequencies, $\nu_1, \nu_2, ..., \nu_k$ in (1), are known.

Section II discussed the estimation problem by using the CSS method. We obtain a unified approach to fitting traditional SARIMA processes as well as non-stationary (seasonal) ARFIMA processes [see Box and Jenkins (1976) and Beran (1995)]. However, we cannot extend the model (2) in Section I to the following linear regression model:

$$y_t=\tilde{\varphi}_t.\beta+x_t, \quad (1-L)^{d_0}(1-L^s)^{d_s}x_t=\vartheta(L)\varepsilon_t, \quad (t=1, 2, ..., T).$$

In this case, consistency of the least-squares estimator of β, $\hat{\beta}$, depends on differencing parameters, i.e., $\mathrm{Var}[D_T(\hat{\beta}-\beta)]$ is $O(T^{2d})$ if $d\in(0, 1/2)$; and $O(1)$ if $d\in(-1/2, 0)$, as $T\to\infty$, where $d=\max\{d_0+d_s, d_s\}$ because autocovariances, $\gamma(j)$, is $O(j^{2d-1})$ as $j\to\infty$ and $\mathrm{Var}[D_T(\hat{\beta}-\beta)]=O(\Sigma_{j=0}^{T}|\gamma(j)|)$, as $T\to\infty$ [see, e.g., Section 9.1 in Fuller (1996) and Section 2 in Yajima (1988)]. But we cannot prove consistency and asymptotic normality of CSS estimates $(\hat{\delta}', \hat{\sigma}^2)$. The main difficulty is the case of $\max\{d_0+d_s, d_s\}>0$ and $\max\{-d_0-d_s, -d_s\}>0$, typically, $(d_0, d_s)\in D^s_{1,3}$, which is different from that of the ARFIMA model. In Section III, we cannot formulate a linear regression model as in (17) under the testing problem H_0

against $H_{A,1}$ (or $H_{A,2}$) because the LM test statistics have a differencing parameter d_s (or d_0) in nuisance parameters.

Appendix

A. Results on a Fractional Filter

A recursion formula and asymptotic results for a fractional filter are given by following results. The proof of Lemma A 1 is obtained by Katayama (2004b, Lemma A 1).

Lemma A 1. *Let $F(z)$ be a fractional filter defined in (1) such that*

$$F(z)=(1-z)^{-d_1}(1+z)^{-d_k}\prod_{i=2}^{k-1}(1-2\eta_i z+z^2)^{-d_i}=\prod_{i=1}^{k}(1-2\eta_i z+z^2)^{-D_i}=\sum_{j=0}^{\infty}\phi_j z^j, \quad (23)$$

$|z|<1$, *where* $\eta_i\equiv\cos(\nu_i)$ *and* $0=\nu_1<\nu_2<\cdots<\nu_{k-1}<\nu_k=\pi$, *and* $D_1=d_1/2$, $D_k=d_k/2$, $D_i=d_i$ *for* $i=2, ..., k-1$. *Then*

1. $\phi_0=1$, *and*

$$\phi_j=\frac{2}{j}\sum_{i=0}^{j-1}\sum_{m=1}^{k}D_m\cos[(j-i)\nu_m]\phi_i, \quad \text{for } j\geq 1, \quad (24)$$

2. [Asymptotic results by Giraitis and Leipus (1995, Theorem 1), Leipus and Viano (2000, Lemma 1), and Viano et al. (1995, Proposition 7)].

$$\phi_j\sim\sum_{i=1}^{k}\frac{\kappa_i(j)}{\Gamma(d_i)}j^{d_i-1}, \quad \text{as } j\to\infty, \quad (25)$$

where $\kappa_1=\kappa_1(j)=2^{-d_k}\prod_{i=2}^{k-1}(2-\cos(\nu_i))^{-d_i}$, $\kappa_k=\kappa_k(j)=2^{-d_1}\prod_{i=2}^{k-1}(2+2\cos(\nu_i))^{-d_i}$, *and*

$$\kappa_i(j)=2\left\{2\sin\left(\frac{\nu_i}{2}\right)\right\}^{-d_1}\left\{2\cos\left(\frac{\nu_i}{2}\right)\right\}^{-d_k}\left\{2\sin\left(\nu_i\right)\right\}^{-d_i}$$

$$\times\prod_{\substack{l\neq i\\ l=2,\ldots,k-1}}[|2(\cos(\nu_i)-\cos(\nu_l))|]^{-d_l}\cos\left[\nu_i\left(\frac{d_1+d_k}{2}+\sum_{m=2}^{k-1}d_m+j\right)-\frac{(d_1+d_i)\pi}{2}\right]$$

for $i=2, 3, ..., k-1$.

3. Let $\phi_{1,j}(d)$ *be defined by* $(1-z)^{-d}=\sum_{j=0}^{\infty}\phi_{1,j}(d)z^j$, $\phi_{1,j}(d)=\Gamma(j+d)/\{\Gamma(d)\Gamma(j+1)\}$, *and* $d\in(-1, 0)\cup(0, 1)$. *Then*

$$\sum_{j=0}^{n}\phi_{1,j}(d)=\frac{n+1}{d}\phi_{1,n+1}(d)=\phi_{1,n}(d+1)\sim\frac{n^d}{\Gamma(d+1)}, \quad (26)$$

$$\sum_{j=0}^{n}|\phi_{1,j}(d)|=\begin{cases}\sum_{j=0}^{n}\phi_{1,j}(d)\sim\dfrac{n^d}{\Gamma(d+1)}, & \text{if } d\in(0, 1),\\ 2-\phi_{1,n}(d+1)\sim 2-\dfrac{n^d}{\Gamma(d+1)}, & \text{if } d\in(-1, 0),\end{cases} \quad (27)$$

where $f(n) \sim g(n)$ *means* $f(n)/g(n) \to 1$, *as* $n \to \infty$.

4. [The summability of Gegenbauer polynomials by Theorem (2.1) in Zayed (1980)].[4] *Let Gegenbauer polynomials be* $C_j^d(\eta), j=0, 1, 2, \ldots$, *which are defined by the generating relation* $(1-2\eta z+z^2)^{-d}=\Sigma_{j=0}^{\infty} C_j^d(\eta) z^j$, $|\eta|<1$, $|z|<1$. *If* $d \in (-1, 0) \cup (0, 1)$, $A=\Sigma_{j=0}^{\infty} a_j$ *is convergent,* $B=\Sigma_{j=0}^{\infty} b_j$, and $b_j=\Sigma_{k=0}^{j} a_{j-k} C_k^d(\eta)$, *then* $C=\Sigma_{j=0}^{\infty} C_j^d(\eta)$ *is convergent, and* $B=AC$.
5. *Let* $\phi_{k,j}(d)$ *be defined by* $(1+z)^{-d}=\Sigma_{j=0}^{\infty}\phi_{k,j}(d) z^j$, $\phi_{k,j}(d)=(-1)^j\phi_{1,j}(d)$, *and* $d \in (-1, 0) \cup (0, 1)$, *where* $\phi_{1,j}(d)$ *is given by 3. Then* $\Sigma_{j=0}^{\infty}\phi_{k,j}(d)$ *is convergent.*
6. *Let, in (23),* $d_i \in (-1, 0) \cup (0, 1)$ *for* $i=1, 2, \ldots, k$. *If* $d_1 \in (-1, 0)$, *then* $\Sigma_{j=0}^{\infty}\phi_j$ *is convergent, and*

$$\sum_{j=0}^{n} \phi_j \sim \kappa_1 \phi_{1,n}(d+1) \sim \frac{\kappa_1}{\Gamma(d_1+1)} n^{d_1}, \quad \text{as } n \to \infty, \tag{28}$$

where κ_1 *and* $\phi_{1,n}(d+1)$ *are given by 2 and 3, respectively. If* $d_1 \in (0, 1)$, *then* $\Sigma_{j=0}^{n}\phi_j = O(n^{d_1})$, *as* $n \to \infty$.

B. Asymptotic Results Relating to CSS Estimates

In this appendix we present some details of the proof of Theorem 1 and some remarks. For simplicity we mainly focus on the proof of Theorem 1 with $\vartheta(z)=1$.

From the definitions in Section III, we first introduce some notations. Let $\delta=(d_0, d_s)'$ be the true parameter vector and let $\ddot{\delta}=(\ddot{d}_0, \ddot{d}_s)'$, $\ddot{\delta}$, $\delta \in D_{i,j}^s$ for some $i, j=1, 2, 3$,

$$\varepsilon_t(\ddot{\delta})=\varepsilon_t(\ddot{\delta}, \bar{x})=\sum_{k=0}^{t-1}\pi_k(\ddot{d}_0, \ddot{d}_s)(x_{t-k}-\bar{x}), \quad x_t=\mu+\sum_{k=0}^{t-1}\phi_k(d_0, d_s)\,\varepsilon_{t-k}, \quad \text{for } t=1, 2, \ldots,$$

be the residual process for evaluating the CSS function,

$$\varepsilon_t(\ddot{\delta}, \mu)=\sum_{k=0}^{t-1}\pi_k(\ddot{d}_0, \ddot{d}_s)(x_{t-k}-\mu), \quad u_t(\ddot{\delta})=\sum_{k=0}^{\infty}\pi_k(\ddot{d}_0, \ddot{d}_s)\, v_{t-k}(\delta), \quad v_t(\delta)=\sum_{k=0}^{\infty}\phi_k(d_0, d_s)\,\varepsilon_{t-k},$$

for $t=1, 2, \ldots$, be the counterparts of the residual process,

$$S(\ddot{\delta})=\frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2(\ddot{\delta}), \quad Q(\ddot{\delta})=\frac{1}{2\sigma^2}\sum_{t=1}^{T}u_t^2(\ddot{\delta}), \quad S^{(2)}(\ddot{\delta})=\frac{\partial^2 S(\ddot{\delta})}{\partial\ddot{\delta}\partial\ddot{\delta}'}, \quad Q^{(2)}(\ddot{\delta})=\frac{\partial^2 Q(\ddot{\delta})}{\partial\ddot{\delta}\partial\ddot{\delta}'},$$

$$(1-z)^a(1-z^s)^b=\sum_{j=0}^{\infty}\pi_j(a, b)z^j, \quad \text{and} \quad (1-z)^{-a}(1-z^s)^{-b}=\sum_{j=0}^{\infty}\phi_j(a, b)z^j.$$

We show that $\hat{\delta}$ is a consistent estimator of δ by showing that

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\ddot{\delta})^2 \xrightarrow{p} \mathrm{E}[u_t(\ddot{\delta})]^2, \quad \text{as } T \to \infty \text{ uniformly in } \ddot{\delta} \in D_{i,j}^s \tag{29}$$

because $\hat{\delta}$ is the estimator of δ that minimizes the objective function $\Sigma_{t=1}^{T}\varepsilon_t^2(\ddot{\delta})/T$. This is sufficient condition for weak consistency by Fuller (1996, Lemma 5.5.1 and Lemma 5.5.2) because $\mathrm{E}[u_t^2(\ddot{\delta})]$ reaches its minimum at δ by the fact that $-\Sigma_{k=1}^{\infty}\pi_k(d_0, d_s)v_{t-k}(\delta)$ uniquely determines the best linear predictor of $v_t(\delta)$ on the basis of the mean squared error based on

[4] Theorem (2.1) in Zayed (1980) shows that $\Sigma_{j=0}^{\infty} a_j C_j^d(\eta)$ converges for any $d>0$, where $a_j \le M(j+1)^P$, $j=0, 1, 2, \ldots$, for some integers M and P. However, we assume $d \in (0, 1)$ for simplicity and modify Zayed's results multiplication of summable series.

the infinite past $v_{t-1}(\delta)$, $v_{t-2}(\delta)$, ... (i.e., $\varepsilon_t = u_t(\delta) = v_t(\delta) + \Sigma_{k=1}^{\infty}\pi_k(d_0,\ d_s)v_{t-k}(\delta)$), which establishes the condition (5.5.7) of Lemma 5.5.2 in Fuller (1996).

We prove the following lemmas that are needed subsequently.

Lemma B 1. *Let the* $\{a_j\}$ *and* $\{b_j\}$ *satisfy* $|a_j|,\ |b_j| \leq C_1(j+1)^{-(\tau+1)}$ *for some* $C_1,\ \tau > 0$, *and any* $j \geq 0$ *and let* $\{c_j\}$ *be defined by* $c_j = \Sigma_{k=0}^{j} a_k b_{j-k},\ j \geq 0$. *Then* $|c_j| \leq Cj^{-(\tau+1)}$, *for some* $C > 0$ *and any* $j \geq 2$.

Proof. By the definition of c_j, dividing the inner summation into two: $1 \leq k \leq [j/2]$ and $[j/2]+1 \leq k \leq j$, we have

$$|c_j| \leq \sum_{k=0}^{[j/2]} \frac{C_1|a_k|}{(j-k+1)^{\tau+1}} + \sum_{k=[j/2]+1}^{j} \frac{C_1|b_{j-k}|}{(k+1)^{\tau+1}}$$

$$\leq \frac{C_1}{(j-[j/2]+1)^{\tau+1}} \sum_{k=0}^{[j/2]} |a_k| + \frac{C_1}{([j/2]+2)^{\tau+1}} \sum_{k=[j/2]+1}^{j} |b_{j-k}|$$

$$\leq C_2(j/2+1)^{-\tau-1} \leq Cj^{-\tau-1}$$

for $j \geq 2$ because $j/2-1 \leq [j/2] \leq j/2$ and $\{a_j\}$ and $\{b_j\}$ are absolutely summable. □

Lemma B 2. *Let* $\ddot{\delta} \in D_{1,1}^{s}$. *Then (i) there exist absolutely summable sequences* $\{\pi_{j,0}(\tau)\}$, *which do not depend on* $\ddot{\delta}$, *and which satisfy* $|\pi_j(\ddot{d}_0,\ \ddot{d}_s)| \leq \pi_{j,0}(\tau)$ *for all* $j \geq 0$ *and* $\pi_{j,0}(\tau) = O(j^{-1-\tau})$ *as* $j \to \infty$. *And (ii) there exist absolutely summable sequences* $\{\pi_{j,i+k}(\tau)\}$, *which do not depend on* $\ddot{\delta}$, *and which satisfy* $|\partial^{i+k}\pi_j(\ddot{d}_0,\ \ddot{d}_s)/(\partial d_0^i \partial d_s^k)| \leq \pi_{j,i+k}(\tau)$ *for all* $j \geq 1$, *and* $\pi_{j,i+k}(\tau) = O((\log j)^{i+k}/j^{1+\tau})$ *as* $j \to \infty$ *for* $i+k = 1,\ 2,\ 3$.

Proof. By $(1-z)^a(1-z^s)^b = (1-z)^{a+b}(1+z)^b \Pi_{j=1}^{s/2-1}(1-2\cos(2\pi j/s)\ z+z^2)^b$ and Lemma B 1, it is sufficient to show that absolute value of coefficients of the expanded series of each factor can be dominated by some absolutely summable sequences. Let a_j be defined by $(1-z)^d = \Sigma_{j=0}^{\infty} a_j(d)z^j$. Then, equation (7) of Yajima (1985):

$$(n+1)^{t-1} \leq \frac{\Gamma(n+t)}{\Gamma(n+1)} \leq n^{t-1}, \quad \text{for } 0 \leq t \leq 1, \quad \text{and} \quad n = 1,\ 2,\ ..., \tag{30}$$

implies $|a_j(\ddot{d}_0 + \ddot{d}_s)| \leq C_1(j-1)^{-\tau-1}$ for $j \geq 2$. The coefficients of the expanded series of $(1+z)^{\ddot{d}_s}$ can be treated similarly. By $(1-2\cos(\theta)\ z+z^2)^{-\nu} = \Sigma_{j=0}^{\infty} C_j^{\nu}(\cos\theta)\ z^j$,

$$(2\nu+\mathrm{j})\ C_j^{\nu}(t) = 2\nu[C_j^{\nu+1}(t) - tC_{j-1}^{\nu+1}(t)], \quad \text{and} \quad |C_j^{\nu}(\cos\theta)| \leq 2^{1-\nu}\frac{j^{\nu-1}}{(\sin\theta)^{\nu}\Gamma(\nu)} \tag{31}$$

for $\nu \in (0,\ 1)$, $\theta \in (0,\ \pi)$ from 8.933.3 of Gradshteyn and Ryzhik (2000) and 22.14.3 of Abramowitz and Stegan (1974), it immediately follows that $|C_j^{\nu}(t)| \leq C_2(j-1)^{\nu-1}$ for $j \geq 2$ and $\nu \in (-1/2,\ 0)$. Hence $|C_j^{-d_s}(t)| \leq C_2(j-1)^{-\tau-1}$ for each $t \in (0,\ \pi)$ and thereby demonstrates (i).

We omit the proof of (ii) since these results are obtained in the same way as those in, e.g., Section 2.11 and (8.8.6) of Fuller (1996). □

Next, consider the lemma for the strong law of large numbers (SLLN) by Yajima (1985, Lemma 3.3) and Doob (1953, Theorem X 6.2).

Lemma B 3. [SLLN by Yajima (1985) and Doob (1953)]. *If random variables* $\{x_j\}$ *satisfy*

$E|x_i x_j| < \infty$ *for all* $i, j > 0$ *and* $E(\Sigma_{i=1}^T x_i/T)^2 \le C/T^a$ *for some* $a, C > 0$, *then, as* $T \to \infty$, $\Sigma_{i=1}^T x_i/T$ *almost certainly converges to zero.*

Lemma B 4. $\Sigma_{t=1}^T \varepsilon_t^2(\ddot{\delta}, \mu)/T - \Sigma_{t=1}^T u_t^2(\ddot{\delta})/T \to 0$ *a.c. as* $T \to \infty$ *uniformly in* $\ddot{\delta} \in D_{1,1}^s$.
Proof. Rewriting $\varepsilon_t(\ddot{\delta}, \mu)$ and $u_t(\ddot{\delta})$ as

$$\varepsilon_t(\ddot{\delta}, \mu) = \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) \sum_{k=0}^{t-j-1} \phi_k(d_0, d_s)\varepsilon_{t-j-k} = \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) e_{t-j}(\delta),$$

$$\begin{aligned} u_t(\ddot{\delta}) &= \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\delta) + \sum_{j=t}^{\infty} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\delta) \\ &= \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) e_{t-j}(\delta) + \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) v'_{t-j}(\delta) + \sum_{j=t}^{\infty} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\delta) \\ &= \varepsilon_t(\ddot{\delta}, \mu) + \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) v'_{t-j}(\delta) + \sum_{j=t}^{\infty} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\delta) \\ &= \varepsilon_t(\ddot{\delta}, \mu) + w_{t,1} + w_{t,2}, \quad \text{(say)}, \end{aligned}$$

where $v_t(\delta) = \Sigma_{k=0}^{\infty} \phi_k(d_0, d_s)\varepsilon_{t-k} = \Sigma_{k=0}^{t-1} \phi_k(d_0, d_s)\varepsilon_{t-k} + \Sigma_{k=t}^{\infty} \phi_k(d_0, d_s)\varepsilon_{t-k} = e_t(\delta) + v'_t(\delta)$, (say), by Lemma B 2, we have $|\varepsilon_t(\ddot{\delta}, \mu)| \le \Sigma_{j=0}^{t-1} \pi_{j,0}(\tau)|e_{t-j}(\delta)| = z_{t,0}$, (say), $|w_{t,1}| \le \Sigma_{j=0}^{t-1} \pi_{j,0}(\tau) |v'_{t-j}(\delta)| = z_{t,1}$, (say), $|w_{t,2}| \le \Sigma_{j=t}^{\infty} \pi_{j,0}(\tau)|v_{t-j}(\delta)| = z_{t,2}$, (say), and

$$\left| \frac{1}{T}\sum_{t=1}^T \varepsilon_t^2(\ddot{\delta}, \mu) - \frac{1}{T}\sum_{t=1}^T u_t^2(\ddot{\delta}) \right| \le \frac{2}{T}\sum_{t=1}^T z_{t,0} z_{t,1} + \frac{2}{T}\sum_{t=1}^T z_{t,0} z_{t,2} + \frac{2}{T}\sum_{t=1}^T z_{t,1} z_{t,2} + \frac{1}{T}\sum_{t=1}^T z_{t,1}^2 + \frac{1}{T}\sum_{t=1}^T z_{t,2}^2. \tag{32}$$

Using Lemma B 2 and the Cauchy-Schwarz inequality, we have, as $t \to \infty$,

$$\begin{aligned} E[z_{t,1}]^4 &\le \left[\sum_{j=0}^{t-1} \pi_{j,0}(\tau) \{E v'_{t-j}(\delta)^4\}^{1/4} \right]^4 = O\left(\left\{ \sum_{j=0}^{[t/2]} \frac{\pi_{j,0}(\tau)}{(t-j)^\tau} \right\}^4 + \left\{ \sum_{j=[t/2]+1}^{t-1} \pi_{j,0}(\tau) \right\}^4 \right) \\ &= O\left(t^{-4\tau} \left\{ \sum_{j=0}^{[t/2]} \pi_{j,0}(\tau) \right\}^4 + \left\{ \sum_{j=[t/2]+1}^{\infty} \pi_{j,0}(\tau) \right\}^4 \right) = O(t^{-4\tau}), \\ E[z_{t,2}]^4 &\le \left[\sum_{j=t}^{\infty} \pi_{j,0}(\tau) \left\{ E v_{t-j}(\delta)^4 \right\}^{1/4} \right]^4 = O\left(\left\{ \sum_{j=t}^{\infty} \pi_{j,0}(\tau) \right\}^4 \right) = O(t^{-4\tau}), \end{aligned}$$

and $E[z_{t,0}]^4 = O(1)$. Therefore, using the Cauchy-Schwarz inequality and Lemma B 3, the RHS of (32) almost certainly converges to zero, which proves the lemma. □

Lemma B 5. $\Sigma_{t=1}^T u_t^2(\ddot{\delta})/T \to E[u_t^2(\ddot{\delta})]$ *a.c. as* $T \to \infty$ *uniformly in* $\ddot{\delta} \in D_{1,1}^s$.

Proof. For fixed $\ddot{\delta}$, we have $\Sigma_{t=1}^T u_t^2(\ddot{\delta})/T \to E[u_t^2(\ddot{\delta})]$. Therefore, the rest of the proof is devoted to showing uniformity in $\ddot{\delta} \in D_{1,1}^s$. By Lemma B 2, we have $|u_t(\ddot{\delta})| \le w_{t,1}$ and $|\partial u_t(\ddot{\delta})/\partial\ddot{\delta}| \le w_{t,2}$ where $w_{t,1} = \Sigma_{j=0}^{\infty} \pi_{j,0}(\tau)|v_{t-j}(\delta)|$ and $w_{t,2} = \Sigma_{j=1}^{\infty} \pi_{j,1}(\tau)|v_{t-j}(\delta)|$. Let $\Omega_1 = \{\omega | \lim_{T\to\infty} \Sigma_{t=1}^T w_{t,i}^2/T = E w_{t,i}^2, i = 1, 2\}$ and $D_0 = \{\delta_i | i = 1, 2, ...\}$ be a countable dense subset of $D_{1,1}^s$. Put $\Omega_{\delta(i)} = \{\omega | \lim_{T\to\infty} \Sigma_{t=1}^T u_t^2(\delta_i)/T = E[u_t(\delta_i)]^2\}$ and $\Omega = \Omega_1 \cap_i \Omega_{\delta(i)}$, we have Pr $(\Omega) = 1$ since $u_t(\delta_i)$ and $w_{t,i}$'s are ergodic processes. The rest of the proof is obvious from the proof of

Theorem 1 by Yajima (1985). Hence, the proof is omitted. □

Lemma B 6. $\sum_{t=1}^{T}\varepsilon_t^2(\ddot{\delta}, \mu)/T \rightarrow \mathrm{E}[u_t^2(\ddot{\delta})]$ *a.c. as* $T \rightarrow \infty$ *uniformly in* $\ddot{\delta} \in \boldsymbol{D}_{1,1}^s$.
Proof. Using the triangle inequality, Lemmas B 4 and B 5, we immediately obtain the result. □

Let $u_t^{(i)}(\ddot{\delta})$ and $\varepsilon_t^{(i)}(\ddot{\delta}, \mu)$ be the *i*-th derivatives of $u_t(\ddot{\delta})$ and $\varepsilon_t(\ddot{\delta}, \mu)$ with respect to $\ddot{\delta}$. Then, similarly to Lemmas B 4 and B 5, we obtain the following lemmas.

Lemma B 7. $\sum_{t=1}^{T}u_t^{(1)}(\ddot{\delta})\, u_t^{(1)}(\ddot{\delta})'/T - \sum_{t=1}^{T}\varepsilon_t^{(1)}(\ddot{\delta}, \mu)\varepsilon_t^{(1)}(\ddot{\delta}, \mu)'/T \rightarrow 0$, *a.c., and* $\sum_{t=1}^{T}u_t^{(i)}(\ddot{\delta})\, u_t(\ddot{\delta})/T - \sum_{t=1}^{T}\varepsilon_t^{(i)}(\ddot{\delta}, \mu)\, \varepsilon_t(\ddot{\delta}, \mu)/T \rightarrow 0$, *a.c.,* $i=1, 2$, *as* $T \rightarrow \infty$ *uniformly in* $\ddot{\delta} \in \boldsymbol{D}_{1,1}^s$.

Lemma B 8. $\sum_{t=1}^{T}u_t^{(1)}(\ddot{\delta})\, u_t^{(1)}(\ddot{\delta})'/T \rightarrow \mathrm{E}[u_t^{(1)}(\ddot{\delta})\, u_t^{(1)}(\ddot{\delta})'] \equiv \sigma^2 \boldsymbol{I}(\ddot{\delta})$ *a.c., and* $\sum_{t=1}^{T}u_t^{(i)}(\ddot{\delta})\, u_t(\ddot{\delta})/T \rightarrow \mathrm{E}[u_t^{(i)}(\ddot{\delta})\, u_t(\ddot{\delta})]$, *a.c.,* $i=1, 2$, *as* $T \rightarrow \infty$ *uniformly in* $\ddot{\delta} \in \boldsymbol{D}_{1,1}^s$.

We omit the proofs since these results are obtained in the same way as those in Lemmas B 4 and B 5. Note that $\boldsymbol{I}(\ddot{\delta})$ is continuous on $\boldsymbol{D}_{i,j}^s$ and $\boldsymbol{I}(\delta) = \boldsymbol{I}_\delta$.

Lemmas B 4 to B 8 concentrate on the case of $\ddot{\delta}, \delta \in \boldsymbol{D}_{1,1}^s$. However, the next lemma shows that these results hold even if $\ddot{\delta}, \delta \in \boldsymbol{D}_{i,j}^s$ *for* $i, j = 1, 2, 3$.

Lemma B 9. *Lemmas B 4 - B 8 still hold if* $\boldsymbol{D}_{1,1}^s$ *is replaced by* $\boldsymbol{D}_{i,j}^s$ *for* $i, j = 1, 2, 3$.
Proof. For the case of $\ddot{\delta}, \delta \in \boldsymbol{D}_{2,1}^s$, rewrite $\varepsilon_t(\ddot{\delta})$ and $u_t(\ddot{\delta})$ as

$$\varepsilon_t(\ddot{\delta}) = \sum_{j=0}^{t-1}\pi_j(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s)\sum_{k=0}^{t-j-1}\phi_k(d_0 + \frac{1}{4}, d_s)\varepsilon_{t-j-k} = \sum_{j=0}^{t-1}\pi_j(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s)e_{t-j}(d_0 + \frac{1}{4}, d_s),$$

$$u_t(\ddot{\delta}) = \varepsilon_t(\ddot{\delta}) + \sum_{j=0}^{t-1}\pi_j(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s)v'_{t-j}(d_0 + \frac{1}{4}, d_s) + \sum_{j=t}^{\infty}\pi_j(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s)v_{t-j}(d_0 + \frac{1}{4}, d_s),$$

where

$$v_t(a, b) = \sum_{k=0}^{\infty}\phi_k(a, b)\varepsilon_{t-k} = \sum_{k=0}^{t-1}\phi_k(a, b)\varepsilon_{t-k} + \sum_{k=t}^{\infty}\phi_k(a, b)\varepsilon_{t-k} = e_t(a, b) + v'_t(a, b), \quad \text{(say)}$$

and we have used the fact that $\sum_{j=0}^{t-1}\sum_{k=0}^{t-j-1}a_{k,j} = \sum_{j=0}^{t-1}\sum_{k=0}^{j}a_{k,j-k}$ and $\pi_j(a+b, c+d) = \sum_{k=0}^{j}\pi_k(a, c)\, \pi_{j-k}(b, d)$. Using the proof of Lemma B 2, we again establish the absolute summable sequences $\{\pi'_{j,i+k}(\tau)\}$ such that $|\partial^{i+k}\pi_j(\ddot{d}_0 + 1/4, \ddot{d}_s)/(\partial d_0^i\, \partial d_s^k)| \leq \pi'_{j,i+k}(\tau)$ and $\pi'_{j,i+k}(\tau) = O((\log j)^{i+k} j^{-\tau-1})$ for $i+k = 0, ..., 3$ because $\ddot{d}_0 + \ddot{d}_s + 1/4, \ddot{d}_s \in (\tau, 1/2-\tau)$. It follows that the rest of the proof relating to $\boldsymbol{D}_{2,1}^s$ is obtained in the same way as those in Lemmas B 4 to B 8. Since other $\boldsymbol{D}_{i,j}^s$s can be treated similarly, we omit the proof. □

The following lemma implies that strong consistency and order in probability of sample mean, $\bar{x} = \sum_{t=1}^{T}x_t/T$, such as Lemma 9 of Leipus and Viano (2000) are unaffected if $x_t - \mu = \varepsilon_t = 0$, for all $t \leq 0$.

Lemma B 10. *Under the Assumption 1, it holds that, as* $T \rightarrow \infty$,

$$\bar{x} \overset{\text{a.c.}}{\rightarrow} \mu \textit{ and } \mathrm{E}(\bar{x} - \mu)^2 = O(T^{2(d_0+d_s)-1}). \tag{33}$$

Proof. We assume that $\vartheta(z) = 1$ for simplicity. Since $\sum_{t=1}^{T}(x_t - \mu) = \sum_{t=1}^{T}\sum_{j=0}^{t-1}\phi_j(d_0, d_s)\varepsilon_{t-j} = \sum_{t=1}^{T}\sum_{j=0}^{T-t}\phi_j(d_0, d_s)\varepsilon_t$, we have, by Lemma A 1,

$$\mathrm{E}(\bar{x}-\mu)^2=\frac{\sigma^2}{T^2}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\psi_j(d_0, d_s)\right)^2=O\left(\frac{1}{T^2}\sum_{t=1}^{T}t^{2(d_0+d_s)}\right)=O(T^{2(d_0+d_s)-1}),$$

as $T \to \infty$. The general case can be treated similarly because $\vartheta(1)$ converges absolutely by our assumptions. It follows from Lemma B 3 that $\bar{x} \xrightarrow{\text{a.e.}} \mu$. □

Finally, we consider lemmas for the weak uniform law of large numbers relating to $\varepsilon_t(\ddot{\delta}) = \varepsilon_t(\ddot{\delta}, \bar{x})$.

Lemma B 11. *(i)* $\Sigma_{t=1}^{T}\varepsilon_t^{(1)}(\ddot{\delta})\varepsilon_t^{(1)}(\ddot{\delta})'/T - \Sigma_{t=1}^{T}\varepsilon_t^{(1)}(\ddot{\delta}, \mu)\varepsilon_t^{(1)}(\ddot{\delta}, \mu)'/T \xrightarrow{p} 0$, *and* $\Sigma_{t=1}^{T}\varepsilon_t^{(i)}(\ddot{\delta})\varepsilon_t(\ddot{\delta})/T - \Sigma_{t=1}^{T}\varepsilon_t^{(i)}(\ddot{\delta}, \mu)\varepsilon_t(\ddot{\delta}, \mu)/T \xrightarrow{p} 0$, $i=0, 1, 2$, *as* $T \to \infty$ *uniformly in* $\ddot{\delta} \in D_{j,k}^{s}$ *for* $j, k=1, 2, 3$. *(ii) Lemmas B 4, B 6, B 7 and B 9 still hold in probability if* $\varepsilon_t(\ddot{\delta}, \mu)$ *is replaced by* $\varepsilon_t(\ddot{\delta})$.

Proof. *(i)* First we consider $\Sigma_{t=1}^{T}\varepsilon_t^2(\ddot{\delta})/T - \Sigma_{t=1}^{T}\varepsilon_t^2(\ddot{\delta}, \mu)/T$. Since $\varepsilon_t(\ddot{\delta}) = \varepsilon_t(\ddot{\delta}, \mu) - (\bar{x}-\mu)\ \Sigma_{j=0}^{t-1}\pi_j(\ddot{d}_0, \ddot{d}_s)$, it is sufficient to show that

$$(\bar{x}-\mu)^2\sup_{\ddot{\delta}}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\pi_k(\ddot{d}_0, \ddot{d}_s)\right)^2 \xrightarrow{p} 0, \text{ as } T \to \infty. \quad (34)$$

By Lemma A 1, there exists a number $a \geq 0$ such that $|\Sigma_{j=0}^{t-1}\pi_j(\ddot{d}_0+a, \ddot{d}_s)| < \infty$ uniformly in $\ddot{\delta} \in D_{j,k}^{s}$ (i.e., $\ddot{d}_0+\ddot{d}_s+a \in (0, 1/2)$). Similar to the proof of Lemma B 9, the RHS of

$$\left|\sum_{j=0}^{t-1}\pi_j(\ddot{d}_0, \ddot{d}_s)\right| = \left|\sum_{i=0}^{t-1}\pi_i(-a, 0)\sum_{j=0}^{t-i-1}\pi_j(\ddot{d}_0+a, \ddot{d}_s)\right| \leq \sum_{i=0}^{t-1}|\pi_i(-a, 0)|\left|\sum_{j=0}^{t-i-1}\pi_j(\ddot{d}_0+a, \ddot{d}_s)\right|$$

is $O(\Sigma_{i=0}^{t-1}|\pi_i(-a, 0)|) = O(t^a)$, as $t \to \infty$. Therefore, by Lemma B 10,

$$(\bar{x}-\mu)^2\sup_{\ddot{\delta}}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\pi_k(\ddot{d}_0, \ddot{d}_s)\right)^2 = O_p(T^{2(d_0+d_s+a)-1})$$

and (34) holds. Other cases can be treated similarly because each derivative of $\Sigma_{j=0}^{t-1}\pi_j(\ddot{d}_0+a, \ddot{d}_s)$ is bounded by Weierstrass's Double Series Theorem, which establishes (i). Using triangle inequality and (i), we immediately obtain (ii). □

Proof of Theorem 1. For simplicity we focus on the proof of Theorem 1 with $\vartheta(z)=1$.

First, we prove weak consistency of $\hat{\delta}$ by showing (29). Using Lemmas B 6, B 9, B 11, and

$$\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\ddot{\delta})^2 - \mathrm{E}[u_t(\ddot{\delta})]^2\right| \leq \left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\ddot{\delta})^2 - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\ddot{\delta}, \mu)^2\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\ddot{\delta}, \mu)^2 - \mathrm{E}[u_t(\ddot{\delta})]^2\right|,$$

the condition (29) is established.

For $\hat{\sigma}^2 = \Sigma_{t=1}^{T}\varepsilon_t^2(\hat{\delta})/T$. Since $\mathrm{E}[u_t^2(\ddot{\delta})]$ is continuous on $D_{i,j}^{s}$, by Lemma B 11, as $T \to \infty$,

$$|\hat{\sigma}^2-\sigma^2| \leq \left|\sum_{t=1}^{T}\varepsilon_t^2(\hat{\delta})/T - \mathrm{E}[u_t^2(\hat{\delta})]\right| + \left|\mathrm{E}[u_t^2(\hat{\delta})]-\sigma^2\right|$$

$$\leq \sup_{\ddot{\delta}}\left|\sum_{t=1}^{T}\varepsilon_t^2(\ddot{\delta})/T - E[u_t^2(\ddot{\delta})]\right| + \left|\mathrm{E}[u_t^2(\hat{\delta})]-\sigma^2\right| \xrightarrow{p} 0.$$

We now establish the asymptotic normality of the estimates. For δ^* on the line segment joining $\hat{\delta}$ and δ we have

$$\frac{1}{\sqrt{T}}\frac{\partial S(\hat{\delta})}{\partial \delta}=0=\frac{1}{\sqrt{T}}\frac{\partial S(\delta)}{\partial \delta}+\left(\frac{1}{T}\frac{\partial^2 S(\delta^*)}{\partial \delta \partial \delta'}\right)\sqrt{T}\,(\hat{\delta}-\delta), \tag{35}$$

in probability. Since $I(\ddot{\delta})$ is continuous on $D^s_{i,j}$, by Lemma B 11, we have, as $T \to \infty$,

$$\begin{aligned}|S^{(2)}(\delta^*)/T-I_\delta| &\le |S^{(2)}(\delta^*)/T-Q^{(2)}(\delta^*)/T|+|Q^{(2)}(\delta^*)/T-I(\delta^*)|+|I(\delta^*)-I_\delta| \\ &\le \sup_{\ddot{\delta}}|S^{(2)}(\ddot{\delta})/T-Q^{(2)}(\ddot{\delta})/T|+\sup_{\ddot{\delta}}|Q^{(2)}(\ddot{\delta})/T-I(\ddot{\delta})|+|I(\delta^*)-I_\delta| \xrightarrow{p} 0.\end{aligned}$$

Therefore $T^{-1}\partial^2 S(\delta^*)/(\partial\delta\partial\delta') \to I_\delta$ in probability, as $T \to \infty$. Since $\varepsilon_t(\delta)=\varepsilon_t(\delta,\mu)-(\bar{x}-\mu)\sum_{j=0}^{t-1}\pi_j(d_0, d_s)$,

$$(\bar{x}-\mu)^2\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\pi_j(d_0, d_s)\right)^2=O_p(1) \quad \text{and} \quad (\bar{x}-\mu)^2\sum_{t=1}^{T}\left\|\sum_{j=0}^{t-1}\frac{\partial \pi_j(d_0, d_s)}{\partial \delta}\right\|^2=\mathrm{O}_p((\log T)^2) \tag{36}$$

by the same argument as in the proof of Lemma B 11, Lemmas A 1 and B 10, we have, as $T \to \infty$,

$$\frac{1}{\sqrt{T}}\frac{\partial S(\delta)}{\partial \delta}=\frac{1}{\sqrt{T}\sigma^2}\sum_{t=1}^{T}\varepsilon_t(\delta,\mu)\frac{\partial \varepsilon_t(\delta,\mu)}{\partial \delta}+o_p(1). \tag{37}$$

Therefore, as $T \to \infty$, we can rewrite (35) as:

$$\frac{1}{\sqrt{T}\sigma^2}\sum_{t=1}^{T}\varepsilon_t(\delta,\mu)\frac{\partial \varepsilon_t(\delta,\mu)}{\partial \delta}-I_\delta\sqrt{T}\,(\hat{\delta}-\delta)=o_p(1).$$

Since the process $\varepsilon_t(\delta,\mu)\,\{\partial\varepsilon_t(\delta,\mu)/(\partial\delta)\}$ is a martingale difference, the central limit theorem follows from the central limit theorem for martingale differences, which proves the theorem [see, e.g., Fuller (1996, Theorem 5.3.4 and Theorem 5.5.1)]. Now the first derivative of $\varepsilon_t(\delta,\mu)$ with respect to δ is given by (7) and each element of $\{\delta_k\}$ is defined as follows:

$$\frac{\partial \varepsilon_t(\delta,\mu)}{\partial d_0}=\log(1-L)\,\varepsilon_t=-\sum_{k=1}^{\infty}\frac{1}{k}L^k\varepsilon_t,$$

$$\frac{\partial \varepsilon_t(\delta,\mu)}{\partial d_s}=\log(1-L^s)\varepsilon_t=-\sum_{k=1}^{\infty}\frac{1}{k}L^{ks}\varepsilon_t=-\sum_{k=1}^{\infty}s_kL^k\varepsilon_t,$$

$$\frac{\partial \varepsilon_t(\delta,\mu)}{\partial \phi_j}=-\phi^{-1}(L)\,L^j\,\varepsilon_t=-L^j\sum_{k=1}^{\infty}\phi_k^*L^k\varepsilon_t=-\sum_{k=j}^{\infty}\phi_{k-j}^*L^k\varepsilon_t \text{ for } j=1, ..., p,$$

$$\frac{\partial \varepsilon_t(\delta,\mu)}{\partial \theta_j}=-\theta^{-1}(L)\,L^j\,\varepsilon_t=-L^j\sum_{k=0}^{\infty}\theta_k^*L^k\varepsilon_t=-\sum_{k=j}^{\infty}\theta_{k-j}^*L^k\varepsilon_t \text{ for } j=1, ..., q,$$

$$\frac{\partial \varepsilon_t(\delta,\mu)}{\partial \Phi_j}=-\Phi^{-1}(L^s)\,L^{js}\,\varepsilon_t=-L^{js}\sum_{k=0}^{\infty}\Phi_k^*L^k\varepsilon_t=-\sum_{k=js}^{\infty}\phi_{k-js}^*L^k\varepsilon_t \text{ for } j=1, ..., p_s,$$

$$\frac{\partial \varepsilon_t(\delta,\mu)}{\partial \Theta_j}=-\Theta^{-1}(L^s)\,L^{js}\,\varepsilon_t=-L^{js}\sum_{k=0}^{\infty}\Theta_k^*L^k\varepsilon_t=-\sum_{k=js}^{\infty}\Theta_{k-js}^*L^k\varepsilon_t \text{ for } j=1, ..., q_s, \tag{38}$$

where $s_j = s/j$ for $j = s, 2s, \ldots, ; = 0$ otherwise, ϕ_j^*, θ_j^*, Φ_j^* and Θ_j^* are the coefficients in the expansions $\phi^{-1}(z) = \Sigma_{j=0}^{\infty}\phi_j^* z^j$ and $\theta^{-1}(z) = \Sigma_{j=0}^{\infty}\theta_j^* z^j$, $\Phi^{-1}(z^s) = \Sigma_{j=0}^{\infty}\Phi_j^* z^j$ and $\Theta^{-1}(z^s) = \Sigma_{j=0}^{\infty}\Theta_j^* z^j$, respectively and $\phi_j^* = \theta_j^* = \Phi_j^* = \Theta_j^* = 0$ for $j < 0$. The second derivatives can be obtained similarly, which establish I_δ.

Remark 1. If μ is known and $\bar{x}$ of $\varepsilon_t(\ddot{\delta}) = \varepsilon_t(\ddot{\delta}, \bar{x})$ is replaced by μ, then $\hat{\delta}$ is a strongly consistent estimator because Lemmas B 6 and B 9 hold and $E[u_t(\ddot{\delta})]^2$ reaches its minimum at δ simiarly to (29) [see, e.g., Fuller (1996, Lemma 5.5.2)]. It implies strong consistency of $\hat{\sigma}^2$ and asymptotic normality of (6) similarly to the proof of Theorem 1.

Section III considers (un)constrained estimators in order to study the behaviour of test statistics for the testing problems about d_0 and d_s under local alternatives. The following remarks show the proof of strong consistency of estimators under local alternatives.

Remark 2. for the local model, $(1-L)^{d_{T,0}}(1-L^s)^{d_s}x_t = \varepsilon_t$, $t \geq 1$, $d_{T,0} = d_0 + \theta$ and $\theta = c/\sqrt{T}$, if the CSS estimator $\hat{d}_s$ is given by evaluating the residual, $\varepsilon_t'(\ddot{d}_s) = \Sigma_{j=0}^{t-1}\pi_j(d_0, \ddot{d}_s)\, x_{t-j}$ in place of $\varepsilon_t(\ddot{\delta}, \bar{x})$, the property of strong consistency of $\hat{d}_s$ is immediately obtained.

For the case of $D_{1,1}^s$, let $u_t'(\ddot{d}_s) = \Sigma_{j=0}^{\infty}\pi_j(d_0, \ddot{d}_s)v_{t-j}'$ where $\{v_t'\}$ is given by $v_t' = \Sigma_{j=0}^{\infty}\phi_j(d_{T,0}, d_s)\varepsilon_{t-j}$. Then by the proof of Lemma B 4, we have $\Sigma_{t=1}^{T}u_t'(\ddot{d}_s)^2/T - \Sigma_{t=1}^{T}\varepsilon_t'(\ddot{d}_s)^2/T \to 0$ a.c. uniformly in $\ddot{d}_s$. Now rewrite $u_t'(\ddot{d}_s)$ as

$$u_t'(\ddot{d}_s) = (1-L)^{d_0}(1-L^s)^{\ddot{d}_s}v_t' = (1-L)^{d_0}(1-L^s)^{\ddot{d}_s}(1-L)^{-(d_0+\theta)}(1-L^s)^{-d_s}\varepsilon_t$$
$$= (1-L)^{-\theta}(1-L^s)^{-(d_s-\ddot{d}_s)}\varepsilon_t = (1-L)^{-\theta}w_t(\ddot{d}_s)$$

where $w_t(\ddot{d}_s) = (1-L^s)^{-(d_s-\ddot{d}_s)}\varepsilon_t$. By a Taylor expansion around $\theta = 0$, we have

$$u_t'(\ddot{d}_s) = w_t(\ddot{d}_s) + \frac{c}{\sqrt{T}}\sum_{k=1}^{\infty}\frac{w_{t-k}^*(\ddot{d}_s)}{k} = w_t(\ddot{d}_s) + \frac{c}{\sqrt{T}}z_t(\ddot{d}_s), \quad \text{(say)}$$

where $w_t^*(\ddot{d}_s) = (1-L)^{-\theta^*}w_t(\ddot{d}_s)$ and θ^* is on the line segment joining θ and 0. Note that absolute value of coefficients of expanded series of $(1-L^s)^{\ddot{d}_s}$ are dominated by absolute summable sequences $\{\pi_{j,0}(\tau)\}$ as in the proof of Lemma B 2, which do not depend on $\ddot{d}_s$. It follows that there exists a number $T_0 > 0$, $\tau' \in (0, \tau)$, and for all $T > T_0$,

$$|w_t(\ddot{d}_s)| = |(1-L^s)^{\ddot{d}_s}(1-L^s)^{-d_s}\varepsilon_t| \leq \sum_{j=0}^{\infty}\pi_{j,0}(\tau)\,|(1-L^s)^{-d_s}\varepsilon_{t-j}|,$$

$$|z_t(\ddot{d}_s)| = |\log(1-L)(1-L)^{-\theta^*}(1-L^s)^{-(d_s-\ddot{d}_s)}\varepsilon_t|$$
$$= |(1-L)^{\ddot{d}_s-\theta^*}(1+L)^{\ddot{d}_s}\prod_{j=1}^{s/2-1}(1-2\cos(2\pi j/s)\,L+L^2)^{\ddot{d}_s}\{\log(1-L)(1-L^s)^{-d_s}\varepsilon_t\}|$$
$$\leq \sum_{j=0}^{\infty}\pi_{j,0}(\tau')\,|\log(1-L)(1-L^s)^{-d_s}\varepsilon_{t-j}|,$$

where $\tau' < \tau - \theta$ and $1/2 - \tau - \theta < 1/2 - \tau'$. Therefore, the RHS of

$$\left|\frac{1}{T}\sum_{t=1}^{T}u_t'(\ddot{d}_s)^2 - \frac{1}{T}\sum_{t=1}^{T}w_t(\ddot{d}_s)^2\right| = \left|\frac{c^2}{T^2}\sum_{t=1}^{T}z_t(\ddot{d}_s)^2 + \frac{2c}{T\sqrt{T}}\sum_{t=1}^{T}z_t(\ddot{d}_s)\,w_t(\ddot{d}_s)\right|$$

is bounded by some nondegenerate random variable, $z_T(\tau, \tau')$, say, which does not depend on

$\ddot{d}_s$ and θ^*, and as $T \to \infty$, $z_T(\tau, \tau') \to 0$ almost certainly by pointwise ergodic theorem. It follows that, as $T \to \infty$, $\Sigma_{t=1}^T u_t'(\ddot{d}_s)^2/T - \Sigma_{t=1}^T w_t(\ddot{d}_s)^2/T \to 0$ a.c. uniformly in $\ddot{d}_s$, and hence $\Sigma_{t=1}^T \varepsilon_t'(\ddot{d}_s)^2/T - \Sigma_{t=1}^T w_t(\ddot{d}_s)^2/T \to 0$ a.c. uniformly in $\ddot{d}_s$.

An almost certain convergence of $\Sigma_{t=1}^T w_t(\ddot{d}_s)^2/T$ is shown similarly to the proofs of Lemmas B 5 and B 9. Therefore, $\Sigma_{t=1}^T \varepsilon_t'(\ddot{d}_s)^2/T \to \mathrm{E}[w_t(\ddot{d}_s)]^2$ a.c. uniformly in $\ddot{d}_s$, which implies strong consistency of $\hat{d}_s$ similarly to the proof of Theorem 1.

When $\hat{d}_s$ is given by evaluating the residual, $\varepsilon_t((d_0, \ddot{d}_s)', \bar{x})$, a weak consistency of $\hat{d}_s$ is obtained from Lemma B 11.

Remark 3. For the model defined in Remark 2, in order to estimate the true parameter $\delta = (d_{T,0}, d_s)'$, if the CSS estimator $\tilde{\delta} = (\tilde{d}_{T,0}, \tilde{d}_s)'$ is given by evaluating the residual $\varepsilon_t(\ddot{\delta}) = \Sigma_{k=0}^{t-1}\ \pi_k(\ddot{d}_0, \ddot{d}_s)\ x_{t-k}$ similarly to Section III, the property of strong consistency of $\tilde{\delta}$ is obtained by modifying Remark 2.

For the case of $D_{1,1}^s$, let $u_t(\ddot{\delta}) = \Sigma_{k=0}^\infty \pi_k(\ddot{d}_0, \ddot{d}_s) v_{t-k}(\delta)$, $v_t(\delta) = \Sigma_{k=0}^\infty \psi_k(d_{T,0}, d_s)\varepsilon_{t-k}$, and $w_t(\ddot{\delta}) = \Sigma_{j=0}^\infty \psi_j(d_0 - \ddot{d}_0, d_s - \ddot{d}_s)\varepsilon_{t-j}$. Then by a straightforward extension of the method used by Lemmas B 4 and B 5, we have $\Sigma_{t=1}^T \varepsilon_t^2(\ddot{\delta})/T - \Sigma_{t=1}^T u_t^2(\ddot{\delta})/T \to 0$ and $\Sigma_{t=1}^T w_t^2(\ddot{\delta})/T \to \mathrm{E}[w_t^2(\ddot{\delta})]$ a.c. as $T \to \infty$ uniformly in $\ddot{\delta} \in D_{1,1}^s$. Using the argument as in $u_t'(\ddot{d}_s)$, $w_t(\ddot{d}_s)$, and $z_t(\ddot{d}_s)$ of Remark 2, we again establish that $\Sigma_{t=1}^T u_t^2(\ddot{\delta})/T - \Sigma_{t=1}^T w_t^2(\ddot{\delta})/T \to 0$ and $\mathrm{E}[u_t(\ddot{\delta})]^2 - \mathrm{E}[w_t(\ddot{\delta})]^2 \to 0$ a.c. as $T \to \infty$ uniformly in $\ddot{\delta} \in D_{1,1}^s$. It follows that $\Sigma_{t=1}^T \varepsilon_t^2(\ddot{\delta})/T - \mathrm{E}[u_t(\ddot{\delta})]^2 \to 0$ a.c. as $T \to \infty$ uniformly in $\ddot{\delta} \in D_{1,1}^s$. Hence, by Gallant and White (1988, Theorem 3.3), the proof of strong consistency of $\tilde{\delta}$ is obtained easily by the fact that $-\Sigma_{k=1}^\infty \pi_k(d_{T,0}, d_s)\ v_{t-k}(\delta)$ uniquely determines the best linear predictor of $v_t(\delta)$ on the basis of the mean squared error based on the infinite past $v_{t-1}(\delta)$, $v_{t-2}(\delta)$, ...

When $\tilde{\delta}$ is given by evaluating the residual, $\varepsilon_t(\ddot{\delta}, \bar{x})$, a weak consistency of $\hat{d}_s$ is obtained from Lemma B 11.

The asymptotic normality of the estimates is obtained in the same way as those in Theorem 1. Therefore, as $T \to \infty$, $\sqrt{T}\,(\tilde{\delta} - \delta) \xrightarrow{d} N(0, I_\delta^{-1})$,

$$\sqrt{T}\,(\tilde{d}_{T,0} - d_{T,0}) \xrightarrow{d} N(0, \sigma_{d_0}^{-2}), \text{ and } \sqrt{T}\,\sigma_{d_0}(\tilde{d}_{T,0} - d_0) \xrightarrow{d} N(c\sigma_{d_0}, 1), \qquad (39)$$

where $\sigma_{d_0}^2 = (\pi^2/6)(1 - \mathrm{s}^{-2})$. The case of general SARFIMA model (3) can be treated similarly.

C. Asymptotic Results Relating to Residual Autocorrelation Functions

We prove the following lemma that is needed to prove Theorems 2 to 4.

Lemma C1. *Let* $w_{1,T,j}$, $w_{2,T,j}$, *and* $w_{3,T,j}$, $j = 1, 2, ..., T$ *be triangular array of random variables such that* $\Sigma_{j=1}^T w_{1,T,j}^2 = O_p(T^{1/2})$, $\Sigma_{j=1}^T w_{2,T,j}^2 = O_p(1)$, *and* $\Sigma_{j=1}^T w_{3,T,j}^2 = O_p(1)$ as $T \to \infty$ *and let* $\{a_j\}$ *be positive sequences such that* $a_j = O(j^{-1})$ *as* $j \to \infty$. Then

$$\lim_{m \to \infty} \limsup_{T \to \infty} \Pr\left(\left| T^{-1/2} \sum_{k=m+1}^{T-1} a_k \sum_{j=k+1}^{T} v_{1,T,j-k} v_{2,T,j} \right| > \varepsilon \right) = 0$$

for every $\varepsilon > 0$, *where* $(v_{1,T,j}, v_{2,T,j}) = (w_{1,T,j}, w_{2,T,j})$, $(w_{2,T,j}, w_{1,T,j})$, $(w_{2,T,j}, w_{3,T,j})$.

Proof. By using the Cauchy-Schwarz inequality and the fact that $\Sigma_{k=m+1}^\infty k^{-1-a} \le a^{-1} m^{-a}$ for any $a > 0$, there exists a number $T_0 > 0$ and for all $T > T_0$,

$$\frac{1}{\sqrt{T}}\sum_{k=m+1}^{T-1} a_k \sum_{j=k+1}^{T} w_{1,T,j-k}w_{2,T,j} \leq \frac{1}{\sqrt{T}}\sum_{k=m+1}^{T-1} a_k\left(\sum_{j=k+1}^{T} w_{1,T,j-k}^2\right)^{1/2}\left(\sum_{j=k+1}^{T} w_{2,T,j}^2\right)^{1/2}$$

$$= O_p\left(T^{-1/4}\sum_{k=m+1}^{T-1} a_k\right) = O_p\left(\sum_{k=m+1}^{\infty} k^{-1/4}a_k\right) = O_p(m^{-1/4}),$$

$$\frac{1}{\sqrt{T}}\sum_{k=m+1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T} w_{2,T,j-k}w_{3,T,j} \leq \frac{1}{\sqrt{T}}\sum_{k=m+1}^{T-1} a_k\left(\sum_{j=k+1}^{T} w_{2,T,j-k}^2\right)^{1/2}\left(\sum_{j=k+1}^{T} w_{3,T,j}^2\right)^{1/2}$$

$$= O_p\left(\frac{1}{\sqrt{T}}\sum_{k=m+1}^{T-1} a_k\right) = O_p\left(\sum_{k=m+1}^{\infty} k^{-1/2}a_k\right) = O_p(m^{-1/2}),$$

as $m \to \infty$. The case of $(v_{1,T,j}, v_{2,T,j}) = (w_{2,T,j}, w_{1,T,j})$ can be treated similarly. □

For any fixed $m \geq 1$, let $\hat{r} = (\hat{r}(1), ..., \hat{r}(m))'$ be the m-dimensional vector of residual autocorrelations using the CSS estimator and let $r = (r(1), ..., r(m))'$, $r(j) = \sum_{t=1}^{T-j}\varepsilon_t\varepsilon_{t+j}/\sum_{t=1}^{T}\varepsilon_t^2$, $j = 1, ..., m$.

Proof of Theorem 2. An outline of the proof is due to Tanaka (1999, Theorem 3.3).

Strong consistency of $\hat{\xi}$ is given by Remark 2 in Appendix B. First, we consider the limiting distribution of $\hat{\xi} - \xi$. Let $\xi = (d_s, \vartheta')' = (\xi_1, ..., \xi_P)'$, $p + q + p_s + q_s + 1 = P$ and the CSS function be $S(\alpha_0, \xi)$. Then, as $T \to \infty$, we have

$$\frac{1}{\sqrt{T}}\frac{\partial S(0, \hat{\xi})}{\partial \xi_i} = 0 = \frac{1}{\sqrt{T}}\frac{\partial S(\alpha_0, \xi)}{\partial \xi_i} - \frac{c}{T}\frac{\partial^2 S(\alpha_0^*, \xi^*)}{\partial \alpha_0 \partial \xi_i} + \frac{1}{T}\sum_{j=1}^{P}\frac{\partial^2 S(\alpha_0^*, \xi^*)}{\partial \xi_i \partial \xi_j}\sqrt{T}(\hat{\xi}_j - \xi_j)$$

for $i = 1, ..., P$ where $\| (\alpha_0^*, \xi^{*\prime})' - (\alpha_0, \xi')' \| \leq \| (0, \hat{\xi}')' - (\alpha_0, \xi')' \|$. It follows from Lemma B 11 and (37) that, as $T \to \infty$,

$$\sqrt{T}(\hat{\xi} - \xi) = \left(-\frac{1}{T}\frac{\partial^2 S(\alpha_0, \xi)}{\partial \xi \partial \xi'}\right)^{-1}\left(\frac{1}{\sqrt{T}}\frac{\partial S(\alpha_0, \xi)}{\partial \xi} - \frac{c}{T}\frac{\partial^2 S(\alpha_0, \xi)}{\partial \alpha_0 \partial \xi}\right) + o_p(1)$$
$$\xrightarrow{d} N(I_\xi^{-1}I_{\alpha_0\xi}\, c, I_\xi^{-1})$$

where $I_\xi = -\lim_{T\to\infty} T^{-1}E[\partial^2 S(\alpha_0,\xi)/(\partial\xi\partial\xi')]$ and $I_{\alpha_0\xi} = -\lim_{T\to\infty} T^{-1}E[\partial^2 S(\alpha_0,\xi)/(\partial\alpha_0\partial\xi)]$. For $\hat{r}$, since $\hat{r}(i)$ consists of $\hat{\xi}$ and $\alpha_0 = 0$, by a Taylor series expansion, we have, as $T \to \infty$,

$$\sqrt{T}\hat{r} = \left(\frac{\partial r}{\partial \xi'} \quad I_m\right)\begin{bmatrix}\sqrt{T}(\hat{\xi} - \xi) \\ \sqrt{T}r\end{bmatrix} - c\frac{\partial r}{\partial \alpha_0} + o_p(1)$$

$$= (-J_{m\xi} \quad I_m)\left[\begin{pmatrix}\sqrt{T}(\tilde{\xi} - \xi) \\ \sqrt{T}r\end{pmatrix} + \begin{pmatrix} cI_\delta^{-1}I_{\alpha_{0\xi}} \\ 0\end{pmatrix}\right] + J_{m\alpha_0}c + o_p(1),$$

$$J_m = \left(\frac{1}{i} \mid s_i \mid \phi_{i-j}^* \mid \theta_{i-j}^* \mid \Phi_{i-js}^* \mid \Theta_{i-js}^*\right)m$$
$$\quad 1 \quad\ 1 \quad\ p \quad\ q \quad\ p_s \quad\ q_s$$

$$= (J_{m\alpha_0} \quad J_{m\xi}),$$

where J_m is the $m\times(P+1)$ matrix, $J_{m\xi}$ is an $m\times P$ matrix with the first column vector of J_m removed, $J_{m\alpha_0}$ is an m-vector defined by the first column vector of J_m, $\tilde{\xi}$ is the unrestricted CSS estimator of ξ under $H_{A,1}$, and 0 is an $m\times 1$ zero vector. By the argument in Appendix B, it follows that, as $T\to\infty$, $\sqrt{T}\hat{r}\xrightarrow{d}N(0, I_m-J_{m\xi}I_\xi^{-1}J'_{m\xi})+(J_{m\alpha_0}-J_{m\xi}I_\xi^{-1}I_{\alpha_0\xi})\,c$. Hence $\sqrt{T}J'_{m\alpha_0}\hat{r}$ is asymptotically normal with mean $J'_{m\alpha_0}(J_{m\alpha_0}-J_{m\xi}I_\xi^{-1}I_{\alpha_0\xi})\,c$ and variance $J'_{m\alpha_0}(I_m-J_{m\xi}I_\xi^{-1}J'_{m\xi})J_{m\alpha_0}$. Using, as $m\to\infty$, $J'_{m\alpha_0}J_{m\alpha_0}\to\pi^2/6$, $J'_{m\alpha_0}J_{m\xi}\to I'_{\alpha_0\xi}$, we obtain the asymptotic distribution of (13).

Finally, we will show that

$$\lim_{m\to\infty}\limsup_{T\to\infty}\Pr\left(\left|\sqrt{T}\sum_{k=m+1}^{T-1}\frac{1}{k}\hat{r}(\mathrm{k})\right|>\varepsilon\right)=0 \tag{40}$$

by Brockwell and Davis (1991, Proposition 6.3.9). We assume that $\xi=0$ and $\mu=0$ are known for simplicity. Since $\hat{\varepsilon}_j=\varepsilon_j+(c/\sqrt{T})\,\Sigma_{k=1}^{j-1}\varepsilon_{j-k}/k+O_p(1/T)$ and $\Sigma_{j=1}^{T}\hat{\varepsilon}_j^2/T\xrightarrow{p}\sigma^2$, as $T\to\infty$, we have, as $T\to\infty$,

$$\begin{aligned}\sqrt{T}\sum_{k=m+1}^{T-1}\frac{1}{k}\hat{r}(k)&=\sqrt{T}\sum_{k=m+1}^{T-1}\frac{1}{k}\left\{\sum_{j=k+1}^{T}\hat{\varepsilon}_{j-k}\hat{\varepsilon}_j\Big/\sum_{j=1}^{T}\hat{\varepsilon}_j^2\right\}\\&=\frac{\sqrt{T}}{\sigma^2}\sum_{k=m+1}^{T-1}\frac{1}{k}\left\{\sum_{j=k+1}^{T}\hat{\varepsilon}_{j-k}\hat{\varepsilon}_j/T\right\}+o_p(1)\\&=\frac{1}{\sigma^2\sqrt{T}}\sum_{k=m+1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T}\varepsilon_{j-k}\varepsilon_j+\frac{c}{\sigma^2T}\sum_{k=m+1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T}\sum_{l=1}^{j-k-1}\frac{1}{l}\varepsilon_{j-k-l}\varepsilon_j\\&\quad+\frac{c}{\sigma^2T}\sum_{k=m+1}^{T-1}\sum_{j=k+1}^{T}\sum_{l=1}^{j-1}\frac{1}{l}\varepsilon_{j-l}\varepsilon_{j-k}+O_p\left(\frac{1}{T^{3/2}}\sum_{k=m+1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T}1\right)+o_p(1)\\&=A_{T,m}+B_{T,m}+C_{T,m}+D_{T,m}+o_p(1),\quad\text{(say)}.\end{aligned}$$

For $A_{T,m}$,

$$\mathrm{E}[A_{T,m}]^2=\frac{1}{T\sigma^4}\mathrm{E}\left[\sum_{k=m+1}^{T-1}\frac{1}{k}\sum_{j=k+1}^{T}\varepsilon_{j-k}\varepsilon_j\right]^2=\frac{1}{T}\sum_{k=m+1}^{T-1}\sum_{j=k+1}^{T}\frac{1}{k^2}\le C_1\sum_{k=m+1}^{\infty}\frac{1}{k^2}\le\frac{C_2}{m}.$$

It follows from Chebyshev's inequality that there exists a number $T_0>0$ and for all $T>T_0$, $A_{T,m}=O_p(m^{-1/2})$, as $m\to\infty$. Proofs of other cases can be obtained by using Lemma C 1. For a proof of the general case, since $\varepsilon_j((d_0,\hat{\xi}')',\bar{x})=\varepsilon_j((d_0,\hat{\xi}')',\mu)-(\bar{x}-\mu)\,\Sigma_{k=0}^{j-1}\pi_k((d_0,\hat{\xi}')')$,

$$\varepsilon_j((d_0,\hat{\xi}')',\mu)=\varepsilon_j+(-c/\sqrt{T},(\hat{\xi}-\xi)')\sum_{k=0}^{j-1}\delta_k\varepsilon_{j-k}+O_p\left(\frac{1}{T}\right),$$

$$(\bar{x}-\mu)\sum_{k=0}^{j-1}\pi_k((d_0,\hat{\xi}')')=(\bar{x}-\mu)\sum_{k=0}^{j-1}\pi_k((d_0,\xi')')+(\bar{x}-\mu)(\hat{\xi}-\xi)'\sum_{k=0}^{j-1}\frac{\partial\pi_k(\delta^*)}{\partial\xi},$$

where $\|\delta^*-\delta\|\le\|(-c/\sqrt{T},(\hat{\xi}-\xi)')'-\delta\|$ and $(\hat{\xi}-\xi)=O_p(T^{-1/2})$, it can be treated similarly by Lemma B 11, (36), and Lemma C 1. □

Proof of Theorem 4. First, we consider weak consistency of least-squares estimates and CSS estimates. Let $\varepsilon_t(\ddot{\vartheta})=\ddot{\vartheta}(L)^{-1}\{\tilde{y}_t-\tilde{\varphi}_t.\hat{\beta}\}=\ddot{\vartheta}(L)^{-1}\tilde{x}_t(\alpha)-\ddot{\vartheta}(L)^{-1}\tilde{\varphi}_t.(\hat{\beta}-\beta)=\varepsilon_{t,1}(\ddot{\vartheta})+\varepsilon_{t,2}(\ddot{\vartheta})$.

To prove consistency, it is sufficient to check that

$$\sup_{\ddot{\vartheta}} \frac{1}{T}\sum_{t=1}^{T}\{\varepsilon_{t,2}(\ddot{\vartheta})\}^2 = O_p\left(\frac{1}{T}\right) \text{ and } \mathrm{Var}[D_T(\hat{\beta}-\beta)] = O(1), \text{ as } T \to \infty, \tag{41}$$

because the strong uniform law of large numbers of $\Sigma_{t=1}^{T}\{\varepsilon_{t,1}(\ddot{\vartheta})\}^2/T$ is obtained by using the same argument of Remark 2. By (d) in Assumption 2 and $D_T(\hat{\beta}-\beta) = D_T(\Phi'\Phi)^{-1}D_T D_T^{-1}\Phi' x(\alpha)$, if $\mathrm{Var}[D_T^{-1}\Phi' x(\alpha)] = O(1)$, then $\mathrm{Var}[D_T(\hat{\beta}-\beta)] = O(1)$. Let $\vartheta(z) = \Sigma_{j=0}^{\infty}\vartheta_j z^j$, $u_t = \vartheta(L)\varepsilon_t = \Sigma_{j=0}^{\infty}\vartheta_j\varepsilon_{t-j} - \Sigma_{j=t}^{\infty}\vartheta_j\varepsilon_{t-j} = u_{t,1} + u_{t,2}$, then, by a Taylor series expansion of $\tilde{x}_t(\alpha)$ around $\alpha = 0$, we have

$$\tilde{x}_t(\alpha) = (1-L)^{-\alpha_0}(1-L^s)^{-\alpha_s}u_t = u_t - \frac{c'}{\sqrt{T}}\begin{pmatrix}\log(1-L)\\ \log(1-L^s)\end{pmatrix}u_t + O_p\left(\frac{1}{T}\right) = u_t + u_{t,T}^*, \quad \text{(say)},$$

$E[u_{t,T}^*]^2 = O(T^{-1})$ for $t = 1, 2, ..., T$ as $T \to \infty$, and ith element of $D_T^{-1}\Phi' x(\alpha)$ is $d_{Tii}^{-1}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}\tilde{x}_t(\alpha) = d_{Tii}^{-1}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}(u_{t,1} + u_{t,2} + u_{t,T}^*)$. By the proof of Theorem 9.3.1 of Fuller (1996), $E[d_{Tii}^{-1}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}u_{t,1}]^2 = O(1)$. By using the Cauchy-Schwarz inequality, $E[\Sigma_{t=1}^{T}u_{t,2}^2] = O(\Sigma_{t=1}^{T}a^t) = O(1)$ for some $a \in (0, 1)$ and $\Sigma_{t=1}^{T}E(u_{t,T}^*)^2 = O(1)$, $E(d_{Tii}^{-1}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}u_{t,2})^2 \le (d_{Tii}^{-2}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}^2) E(\Sigma_{t=1}^{T}u_{t,2}^2) = O(1)$ and $E(d_{Tii}^{-1}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}u_{t,T}^*)^2 \le (d_{Tii}^{-2}\Sigma_{t=1}^{T}\tilde{\varphi}_{t,i}^2)\{\Sigma_{t=1}^{T}E(u_{t,T}^*)^2\} = O(1)$. It follows that $\mathrm{Var}[D_T^{-1}\Phi' x(\alpha)] = O(1)$. Let $\ddot{\vartheta}(z)^{-1} = \Sigma_{j=0}^{\infty}\ddot{\vartheta}_j z^j$, then

$$\begin{aligned}\sum_{t=1}^{T}\{\varepsilon_{t,2}(\ddot{\vartheta})\}^2 &= \sum_{t=1}^{T}\sum_{i,j=0}^{t-1}\ddot{\vartheta}_i\ddot{\vartheta}_j(\hat{\beta}-\beta)' D_T D_T^{-1}\tilde{\varphi}_{t-i.}{}'\tilde{\varphi}_{t-j.}D_T^{-1}D_T(\hat{\beta}-\beta)\\ &\le \left(\sum_{i=0}^{T}|\ddot{\vartheta}_i|\right)^2 \sup_{0\le i,j\le T}\sum_{t=1}^{T}|(\hat{\beta}-\beta)' D_T D_T^{-1}\tilde{\varphi}_{t-i.}{}'\tilde{\varphi}_{t-j.}D_T^{-1}D_T(\hat{\beta}-\beta)| = O_p(1)\end{aligned} \tag{42}$$

because, by Assumption 2 and $\tilde{\varphi}_t = 0$ for $t \le 0$, $\Sigma_{i=0}^{\infty}|\ddot{\vartheta}_i|$ is uniformly convergent, each element of the $r \times r$ matrix $\Sigma_{t=1}^{T}D_T^{-1}\tilde{\varphi}'_{t-i.}\tilde{\varphi}_{t-j.}D_T^{-1}$ is less than one in absolute value for any $0 \le i, j \le T$, and $\mathrm{Var}[D_T(\hat{\beta}-\beta)] = O(1)$, which establishes (41).

Next, we consider the asymptotic distribution of $\hat{\vartheta} - \vartheta$. Let the CSS function $S((d_0, d_s, \ddot{\vartheta}')', \sigma^2)$ be $S(\alpha, \ddot{\vartheta})$, $\varepsilon_t(\alpha, \ddot{\vartheta}) = (1-L)^{\alpha_0}(1-L^s)^{\alpha_s}\varepsilon_{t,1}(\ddot{\vartheta}) + (1-L)^{\alpha_0}(1-L^s)^{\alpha_s}\varepsilon_{t,2}(\ddot{\vartheta}) = \varepsilon_{t,1}(\alpha, \ddot{\vartheta}) + \varepsilon_{t,2}(\alpha, \ddot{\vartheta})$, and $\varepsilon_t(\ddot{\vartheta}) = \varepsilon_t(0, \ddot{\vartheta})$. Then, as $T \to \infty$, we have

$$\frac{1}{\sqrt{T}}\frac{\partial S(0, \hat{\vartheta})}{\partial\vartheta} = 0 = \frac{1}{\sqrt{T}}\frac{\partial S(\alpha, \vartheta)}{\partial\vartheta} - \frac{\partial^2 S(\alpha^*, \vartheta^*)}{\partial\vartheta\partial\alpha'}\frac{c}{T} + \frac{1}{T}\frac{\partial^2 S(\alpha^*, \vartheta^*)}{\partial\vartheta\partial\vartheta'}\sqrt{T}(\hat{\vartheta}-\vartheta), \tag{43}$$

where $\|\alpha^* - 0\| \le \|\alpha - 0\|$ and $\|\vartheta^* - \vartheta\| \le \|\hat{\vartheta} - \vartheta\|$. Let $\|\alpha^*\| \le \|\alpha\|$ and $|a_j| = O((\log j)^k j^{a-1})$ for some $k > 0$ and $0 < a < 1/2$, then

$$\varepsilon_{t,2}(\alpha, \vartheta) = \varepsilon_{t,2}(\vartheta) + \frac{c'}{\sqrt{T}}\begin{pmatrix}\log(1-L)\\ \log(1-L^s)\end{pmatrix}\varepsilon_{t,2}(\alpha^*, \vartheta),$$

$$\frac{\partial\varepsilon_{t,2}(\alpha, \vartheta)}{\partial\alpha} = \begin{pmatrix}\log(1-L)\\ \log(1-L^s)\end{pmatrix}\varepsilon_{t,2}(\alpha, \vartheta),$$

$$\frac{\partial\varepsilon_{t,2}(\alpha, \vartheta)}{\partial\vartheta} = (1-L)^{\alpha_0}(1-L^s)^{\alpha_s}\frac{\partial\varepsilon_{t,2}(\vartheta)}{\partial\vartheta}$$

$$= \frac{\partial \varepsilon_{t,2}(\vartheta)}{\partial \vartheta} + \frac{c'}{\sqrt{T}} \begin{pmatrix} \log(1-L) \\ \log(1-L^s) \end{pmatrix} \frac{\partial \varepsilon_{t,2}(\alpha^*, \vartheta)}{\partial \vartheta},$$

$$\sum_{t=1}^{T} \left\{ \sum_{j=0}^{t-1} a_j \varepsilon_{t-j,2}(\ddot{\vartheta}) \right\}^2 = O_p\left(\left\{ \sum_{j=0}^{T} |a_j| \right\}^2 \right) = O_p((\log T)^{2k} T^{2a}) = o_p(T), \tag{44}$$

where the last equation follows from (42). It follows from (42) and (44) that

$$\frac{1}{\sqrt{T}} \frac{\partial S(\alpha, \vartheta)}{\partial \vartheta} = -\frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^{T} \varepsilon_t(\alpha, \vartheta) \frac{\partial \varepsilon_t(\alpha, \vartheta)}{\partial \vartheta} = -\frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^{T} \varepsilon_{t,1}(\alpha, \vartheta) \frac{\partial \varepsilon_{t,1}(\alpha, \vartheta)}{\partial \vartheta} + o_p(1)$$

as $T \to \infty$. Using (41) and (44), we find that, as $T \to \infty$, each term of the RHS of (43) divided by T is unaffected by $\varepsilon_{t,2}(\alpha, \ddot{\vartheta})$ in probability uniformly in $\ddot{\vartheta} \in D_\vartheta$. The rest of the proof of asymptotic distributions of $\sqrt{T}(\hat{\vartheta} - \vartheta)$ is obvious from the proof of Theorem 2. Hence, we omit the proof.

Finally, we will prove (40) to derive asymptotic distribution of S_T. Since $\varepsilon_j(\hat{\vartheta}) = \varepsilon_{j,1}(\hat{\vartheta}) + \varepsilon_{j,2}(\hat{\vartheta})$, $\varepsilon_{j,1}(\hat{\vartheta}) = \varepsilon_j + (-c'/\sqrt{T}, (\hat{\vartheta} - \vartheta)') \Sigma_{k=0}^{j-1} \delta_k \varepsilon_{j-k} + O_p(1/T)$, and $\hat{\vartheta} - \vartheta = O_p(T^{-1/2})$, it can be treated similarly to the proof of Theorem 2 by Lemma C 1 and (41).

References

Abramowitz, M. and Stegan, I.A. (1974), *Handbook of Mathematical Functions*, Dover Publications.

Agiakloglou, C. and Newbold, P. (1994), "Lagrange Multiplier Tests for Fractional Difference," *Journal of Time Series Analysis* 15, pp.253-262.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman and Hall, London.

Beran, J. (1995), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models," *Journal of Royal Statistical Society*, B 57, pp.659-672.

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

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

Chung, C. -F. (1996), "A generalized Fractionally Integrated ARMA Process," *Journal of Time Series Analysis*, 17, pp.111-140.

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

Doob, J. L. (1953), *Stochastic Processes*, John Wiley, New York.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, 2nd ed., John Wiley, New York.

Gallant, A. R. and White, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, New York.

Ghysels, E. and Osborn, D. R. (2001), *The Econometric Analysis of Seasonal Time Series*, Cambridge, New York.

Giraitis, L. and Leipus, R. (1995), "A Generalized Fractionally Differencing Approach in

Long-memory Modelling," *Lithuanian Mathematical Journal*, 35, pp.65-81.

Godfrey, L.G. (1979), "Testing the adequacy of a time series model," *Biometrika*, 66, pp.67-72.

Gradshteyn, I.S. and Ryzhik, I.M. (2000), *Table of Integrals, Series and Products*, 6th ed. Academic Press, New York.

Granger, C. W. J. and Joyeux, R. (1980), "An Introduction to Long-range Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, pp.15-30.

Gray, H. L., Zhang, N. -F., and Woodward, W. A. (1989), "On Generalized Fractional Processes," *Journal of Time Series Analysis*, 10, pp.233-257.

Honda, M. (2000), *Keiei no tame no Jyuyou no Bunseki to Yosoku* [*Analysis and Forecasting of Demand for Marketing Research*], Sannoudaigakushuppanbu, Tokyo.

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

Imhof, J.P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, pp.419-426.

Katayama, N. (2004a), "Essays on Seasonally and Fractionally Differenced Time Series," Doctoral dissertation, Hitotsubashi University.

Katayama, N. (2004b), "Seasonally and Fractionally Differenced Time Series," *Discussion Paper Series* No. 11, Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences, A 21st-century COE Program, Institute of Economic Research, Hitotsubashi University.

Leipus, R. and Viano, M. -C. (2000), "Modelling Long-memory Time Series with Finite or Infinite Variance: A General Approach," *Journal of Time Series Analysis*, 21, pp.61-74.

Porter-Hudak, S. (1990), "An Application of Seasonal Fractionally Differenced model to the Monetary Aggregates," *Journal of the American Statistical Association*, 85, pp.338-344.

Robinson, P. M. (1991), "Testing for Strong serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, pp.67-84.

Robinson, P. M. (1994), "Efficient Tests of Nonstationary Hypotheses," *Journal of American Statistical Association*, 89, pp.1420-1437.

Silvapulle, P. (2001), "A Score Test for Seasonal Fractional Integration and Cointegration," *Econometric Reviews*, 20 (1), pp.85-104.

Tanaka, K. (1999), "The nonstationary Fractional Unit Root," *Econometric Theory*, 15, pp. 549-582.

Viano, M. -C., Deniau, C. and Oppenheim, G. (1995), "Long-range Dependence and Mixing for Discrete Time Fractional Processes," *Journal of Time Series Analysis*, 16, pp.323-338.

Woodward, W. A., Cheng, Q. C., and Gray, H. L. (1998), "A k-factor Long-memory Model," *Journal of Time Series Analysis*, 19, pp.485-504.

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

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

Yamamoto, T. (1988), *Keizai no Jikeiretsu Bunseki* [*Time Series Analysis in Economics*], Soubunsha, Tokyo.

Zayed, A.I. (1980), "On the Singularities of Gegenbauer (ultraspherical) Expansions," *Transactions of the American Mathematical Society*, 262, pp.487-503.