Discussion Paper Series

No.11

Seasonally and Fractionally Differenced Time Series

Naoya Katayama

January 2004
revised August 2006

Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences
A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

SEASONALLY AND FRACTIONALLY DIFFERENCED TIME SERIES*

NAOYA KATAYAMA†

Department of Economics
Hitotsubashi University

Key Words and Phrases : Fractional differencing; Lagrange multiplier test; Long memory; Seasonal differencing; Seasonal persistence.

JEL classifications : C22, C51

* This paper is based on a portion of Chapters 1 and 2 of the author's Ph.D. thesis. The author thanks Professor Katsuto Tanaka and Professor Yoshihiro Yajima for their valuable comments and suggestions. This work was supported by the Japan Society for the Promotion of Science.

† Corresponding author; Naoya Katayama. Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, JAPAN. E-mail: pg01207@srv.cc.hit-u.ac.jp

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 tend towards normality. The paper provides a modelling application using data on total power consumption in Japan.

1 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 $\phi(z)=1-\sum_{i=1}^{p}\phi_i z^i$ and $\theta(z)=1+\sum_{i=1}^{q}\theta_i z^i$ 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 $i=1,\ldots,k$.

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 $\phi(z)\Phi(z^s)$, $\theta(z)\Theta(z^s)$ 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 2, 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 3, the large-sample distribution of the residual autocorrelation is derived and 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 2 and 3 impose the condition $\{x_t=\mu,\ t\le 0\}$ 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 4. Section 5 illustrates the use of the SARFIMA model. Section 6 concludes.

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

2 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

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

where $\vartheta(L) = \theta(L)\Theta(L^s)/[\phi(L)\Phi(L^s)]$. We make the assumption that $\{x_t = \mu,\ t \le 0\}$ 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) $(d_0 + d_s, d_s) \in D_i^s \times D_j^s = D_{i,j}^s$ for some $i, j = 1, 2, 3$ where $D_1^s = [\tau, 1/2 - \tau]$, $D_2^s = [-\tau, \tau]$, $D_3^s = [-1/2 + \tau, -\tau]$, and $\tau \in (0, 1/4)$. (d) Let $\boldsymbol{\vartheta}$ be $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \Phi_1, \ldots, \Phi_{p_s}, \Theta_1, \ldots, \Theta_{q_s})'$ and D_ϑ be a compact space such that, for any $\boldsymbol{\vartheta} \in D_\vartheta$, $\phi(z)$, $\Phi(z^s)$, $\theta(z)$, and $\Theta(z^s)$ satisfy conditions given in Section 1. 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 $t \le 0$, 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 $|d_0 + d_s|, |d_s| \le 1/2$, 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 $\boldsymbol{\delta}$ be a true parameter vector $(d_0, d_s, \boldsymbol{\vartheta}')'$, let $\ddot{\boldsymbol{\delta}}$ be a corresponding parameter vector, and assume that $\ddot{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}$ are in the same compact parameter space defined by Assumption 1. Let $\bar{x} = \sum_{t=1}^{T} x_t/T$, and let $\pi_k(\ddot{\boldsymbol{\delta}})$ be defined by $\sum_{k=0}^{\infty} \pi_k(\ddot{\boldsymbol{\delta}}) z^k = (1 - z)^{\ddot{d}_0}(1 - z^s)^{\ddot{d}_s}\ddot{\vartheta}(z)^{-1}$, where $\ddot{\vartheta}(z)$ be given by replacing $\boldsymbol{\vartheta}$ in $\vartheta(z)$ by $\ddot{\boldsymbol{\vartheta}} \in D_\vartheta$ in Assumption 1. Then the CSS estimator $(\widehat{\boldsymbol{\delta}}', \widehat{\sigma}^2)'$ of $(\boldsymbol{\delta}', \sigma^2)'$ is obtained by maximizing the CSS function:

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

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

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 $(d_0 + d_s, d_s) \in D_{1,1}^s$ (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 $\boldsymbol{\delta}$'s CSS estimator $\widehat{\boldsymbol{\delta}}$ is independent on that of σ^2, $\widehat{\sigma}^2 = \sum_{t=1}^{T}\varepsilon_t^2(\widehat{\boldsymbol{\delta}})/T$, 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 $\widehat{\boldsymbol{\delta}}$ and $\widehat{\sigma}^2$ be the CSS estimator of the parameter vector $(\boldsymbol{\delta}', \sigma^2)'$ based on a sample $\{x_t\}_{t=1}^{T}$ given by (3) and Assumption 1. Then it follows that, as $T \to \infty$,*

$$\widehat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}, \qquad \widehat{\sigma}^2 \xrightarrow{p} \sigma^2, \tag{5}$$

and

$$\sqrt{T}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{I}_{\boldsymbol{\delta}}{}^{-1}),\ \sqrt{T}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0,\ 2\sigma^4 + \kappa_4) \tag{6}$$

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

$$\boldsymbol{I}_{\boldsymbol{\delta}} = \sum_{k=1}^{\infty}\boldsymbol{\delta}_k\boldsymbol{\delta}_k', \quad \text{and} \quad \frac{\partial\varepsilon_t(\boldsymbol{\delta}, \mu)}{\partial\boldsymbol{\delta}} = \sum_{k=1}^{t-1}\boldsymbol{\delta}_k L^k\varepsilon_t, \tag{7}$$

and each element of $\{\boldsymbol{\delta}_k\}$ 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 $\varepsilon_t(\ddot{\boldsymbol{\delta}})$ is replaced by μ, then $\widehat{\boldsymbol{\delta}}$ and $\widehat{\sigma}^2$ 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 $\Phi(L^s) = 1 - \Phi L^s$, $\boldsymbol{I_\delta}$ can be written as

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

3 Tests Based on Residual Autocorrelation

This section discusses testing for the integration order, namely, the Portmanteau test, and 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.

Box and Jenkins (1976, Chapter 8) pointed out that it is important to check the assumption of independence of $\{\varepsilon_t\}$ by using the residual autocorrelation function. If the fitted model is appropriate, then the residuals should behave in a manner that is consistent with the model. A well-known method is the (modified) Portmanteau test, which evaluates the sum of the squared residual sequences.

Under the Assumption 1, let $\{\widehat{\varepsilon}_t\}$ be the residual sequence of the CSS estimator $\widehat{\boldsymbol{\delta}}$ in Theorem 1 such that $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\delta}})$ and let $\widehat{r}(j) = \sum_{t=1}^{T-j} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t+j} / \sum_{t=1}^{T} \widehat{\varepsilon}_t^2$. Then we have the following theorem.

Lemma 1. *For any fixed $m \geq 1$, let $\widehat{\boldsymbol{r}} = (\widehat{r}(1), \ldots, \widehat{r}(m))'$ be the m-dimensional vector of residual autocorrelations using the CSS estimator under the same conditions as in Theorem 1. Then $\sqrt{T}\widehat{\boldsymbol{r}}$ is asymptotically normal with mean zero and covariance matrix $\boldsymbol{I}_m - \boldsymbol{J}_m \boldsymbol{I_\delta}^{-1} \boldsymbol{J}'_m$, where $\boldsymbol{J}_m$ is the $m \times (2 + p + q + p_s + q_s)$ matrix with each (i, j) element of the partitioned matrix given by*

$$\boldsymbol{J}_m = \Big(\underbrace{\tfrac{1}{i}}_{1} \;\Big|\; \underbrace{s_i}_{1} \;\Big|\; \underbrace{\phi^*_{i-j}}_{p} \;\Big|\; \underbrace{\theta^*_{i-j}}_{q} \;\Big|\; \underbrace{\Phi^*_{i-js}}_{p_s} \;\Big|\; \underbrace{\Theta^*_{i-js}}_{q_s} \Big) \, m,$$

$s_j = s/j$ for $j = s, 2s, \ldots$; $= 0$ otherwise, ϕ^*_j, θ^*_j, Φ^*_j and Θ^*_j are the coefficients in the expansions $\phi^{-1}(z) = \sum_{j=0}^{\infty} \phi^*_j z^j$ and $\theta^{-1}(z) = \sum_{j=0}^{\infty} \theta^*_j z^j$, $\Phi^{-1}(z^s) = \sum_{j=0}^{\infty} \Phi^*_j z^j$ and $\Theta^{-1}(z^s) = \sum_{j=0}^{\infty} \Theta^*_j z^j$, respectively and $\phi^*_j = \theta^*_j = \Phi^*_j = \Theta^*_j = 0$ for $j < 0$.

An application of the Portmanteau test should be examined at this stage. Owing to the seasonal components of $\boldsymbol{J}_m$, corresponding to s_i, Φ^*_{i-js}, and Θ^*_{i-js}, some elements of $\boldsymbol{I_\delta}$ are difficult to approximate using $\boldsymbol{J}'_m \boldsymbol{J}_m$ for a finite sample of $\{x_t\}_{t=1}^{T}$. For the case in (8) with $m = hs$ where h is an appropriate positive integer, we have

$$\boldsymbol{J}'_m \boldsymbol{J}_m = \begin{pmatrix} \sum_{i=1}^{hs} 1/i^2 & \sum_{i=1}^{h} 1/(i^2 s) & \sum_{i=1}^{hs} \phi^{i-1}/i & \sum_{i=1}^{h} \Phi^{i-1}/(is) \\ . & \sum_{i=1}^{h} 1/i^2 & \sum_{i=1}^{h} \phi^{is-1}/i & \sum_{i=1}^{h} \Phi^{i-1}/i \\ . & . & \sum_{i=1}^{hs} \phi^{2(i-1)} & \sum_{i=1}^{h} \phi^{is-1}\Phi^{i-1} \\ . & . & . & \sum_{i=1}^{h} \Phi^{2(i-1)} \end{pmatrix}.$$

From $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$, $\sum_{i=0}^{\infty} \rho^i = 1/(1-\rho)$, $\sum_{i=1}^{\infty} \rho^i/i = -\log(1-\rho)$ for $|\rho| < 1$ and (8), we found that some elements of $\boldsymbol{J}'_m \boldsymbol{J}_m$ with the finite sum with i running from 1 to h may fail to approximate elements corresponding to $\boldsymbol{I_\delta}$ (e.g., when $T = 100$, $s = 12$, and $m = 36$, h is only 3). It follows that the approximate covariance matrix of $\sqrt{T}\widehat{\boldsymbol{r}}$ is not idempotent unless T and m are sufficiently large. Hence, and particularly in a small sample, the (modified) Portmanteau test statistics are difficult to interpret. Specifically, Ansley and Newbold (1979) investigated a simulation study of individual residual autocorrelations and the modified Portmanteau test statistics for the SARMA model and concluded that agreement between asymptotic and empirical significance levels is very poor, even for samples of

100 observations. Furthermore, these test statistics have very low power in numerous examples used to illustrate the finite sample properties of tests of model adequacy based on the (modified) Portmanteau test for the ARMA model.

For this reason, we propose an alternative test procedure based on the LM test.

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

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

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

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 $S(\alpha_0, \boldsymbol{\xi}, \sigma^2)$, where $\boldsymbol{\xi} = (d_s, \boldsymbol{\vartheta}')'$ is unknown vector, whereas d_0 is any preassigned value. Then the score-like test statistic is given by

$$\begin{aligned} S_T(\alpha_0|H_{A,1}) &= \left.\frac{\partial S(\alpha_0, \boldsymbol{\xi}, \sigma^2)}{\partial \alpha_0}\right|_{H_0:\alpha_0=0,\ \boldsymbol{\xi}=\widehat{\boldsymbol{\xi}},\ \sigma^2=\widehat{\sigma}^2} = \frac{1}{\widehat{\sigma}^2}\sum_{i=2}^{T}\left(\sum_{j=1}^{i-1}\frac{\varepsilon_{i-j}\left((d_0, \widehat{\boldsymbol{\xi}}')', \bar{x}\right)}{j}\right)\varepsilon_i\left((d_0, \widehat{\boldsymbol{\xi}}')', \bar{x}\right) \\ &= T\sum_{i=1}^{T-1}\frac{\widehat{r}(i)}{i} \end{aligned} \tag{11}$$

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

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

where $\boldsymbol{\xi} = (d_0, \boldsymbol{\vartheta}')'$ is unknown vector, whereas d_s is any preassigned value. This implies that the residuals $\{\widehat{\varepsilon}_t\}$ 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 $\boldsymbol{I_\delta}^{-1}$ 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 $\boldsymbol{I_\delta}^{-1}$ 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} ($\widehat{\sigma}_{d_0}$ or $\widehat{\sigma}_{d_s}$) can be obtained by inserting the CSS estimator $\widehat{\boldsymbol{\delta}}$ into $\boldsymbol{\delta}$ in $\boldsymbol{I}_{\boldsymbol{\delta}}$. In addition, using a $T \times (2+p+q+p_s+q_s)$ matrix $\boldsymbol{X} = (\partial \boldsymbol{\varepsilon}/\partial \boldsymbol{\delta}')|_{H_0}$ with each (i,j) element of the partitioned matrix:

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

where $(1,j)$ element is zero and $\widehat{\varepsilon}_t = 0$ for $t \leq 0$, we can also obtain a consistent estimator of $\boldsymbol{I}_{\boldsymbol{\delta}}$, $\boldsymbol{X}'\boldsymbol{X}/(T\widehat{\sigma}^2)$ where $\widehat{\sigma}^2 = \sum_{t=1}^{T} \widehat{\varepsilon}_t^2/T$.

Hence, we suggest the following test statistics:

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

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 $100a\,\%$ of $N(0,1)$ for a test of asymptotic size a.

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

$$y_t = \boldsymbol{\varphi}_{t\cdot}\boldsymbol{\beta} + x_t, \qquad (1-L)^{d_0+\alpha_0}(1-L^s)^{d_s+\alpha_s} x_t = \vartheta(L)\varepsilon_t, \qquad t \geq 1, \tag{17}$$

where $\{\boldsymbol{\varphi}_{t\cdot}\}$ is a $1 \times r$ sequences of fixed, nonstochastic variables, $\boldsymbol{\beta}$ 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 $\{(y_t, \boldsymbol{\varphi}_{t\cdot})\}_{t=1}^{T}$.

The assumed null model H_0 is obtained by imposing the restrictions $\boldsymbol{\alpha} \equiv (\alpha_0, \alpha_s)' = \mathbf{0}$ and the alternative, $H_{A,3}$, is $\boldsymbol{\alpha} \neq \mathbf{0}$.

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

Let

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

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

Assumption 2. For the model in (17), (a) $\{x_t = y_t = 0,\ \boldsymbol{\varphi}_{t\cdot} = \mathbf{0},\ t \leq 0\}$. (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) $\lim_{T\to\infty} d_{Tii} = \infty$, $i = 1, 2, \ldots, r$. (d) $\lim_{T\to\infty} \boldsymbol{D}_T^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{D}_T^{-1} = \boldsymbol{A}$, where $\boldsymbol{A}$ is nonsingular.

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 $\boldsymbol{S}_T$ defined in (18) with $\boldsymbol{\alpha} = \boldsymbol{c}/\sqrt{T}$ where $\boldsymbol{c}$ is a 2×1 constant vector, as $T \to \infty$,*

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

where $\boldsymbol{\Sigma}^{-1}$ *is a* 2×2 *partitioned matrix in the north-west corner of* $\boldsymbol{I_\delta}^{-1}$ *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* $\chi^2(m, \tau^2) = (Z_1 + \tau)^2 + \sum_{i=2}^m Z_i^2$*, 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 $\boldsymbol{\Sigma}^{-1}$, $\widehat{\boldsymbol{\Sigma}}^{-1}$ can be obtained, the test statistic,

$$\lambda_T(\boldsymbol{\alpha}|\, H_{A,3}) = \boldsymbol{S}_T' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{S}_T / T \tag{20}$$

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 $\lambda_T(\boldsymbol{\alpha}|\, H_{A,3})$ exceeds the upper $100a\,\%$ of $\chi^2(2)$ for a test of asymptotic size a.

Furthermore, for practical implementation, we can calculate $\lambda_T(\boldsymbol{\alpha}|\, H_{A,3})$ 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 $\widehat{\varepsilon} = (\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)'$ as the dependent variable. Next, substitute $\widehat{\varepsilon}$ and the CSS estimates for the regressor $\boldsymbol{X}$ as in (15). Then conduct OLS regression and calculate the corresponding T times R^2 statistic, $T\,\widehat{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\varepsilon}/\,\widehat{\varepsilon}'\widehat{\varepsilon}$, as $\lambda_T(\boldsymbol{\alpha}|\, H_{A,3})$.

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* $\widehat{\varepsilon}_t = x_t$ *and* $\{x_t = 0, t \le 0\}$*. 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* $\sum_{i=1}^m i^{-2} \sim \pi^2/6$*,*

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

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

$$\textit{(C):}\quad \Pr\left(\frac{1}{T}\boldsymbol{S}_T'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{S}_T > \chi^2_{2,a}\right) \longrightarrow \Pr\left(\chi^2\left(2, \frac{c^2\pi^2}{6}\right) > \chi^2_{2,a}\right),$$

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

where $Z_1 \sim N(0,1)$*,* $\boldsymbol{S}_T$ *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* $\boldsymbol{\Sigma}_2$ *is a* 2×2 *partitioned matrix in the north-west corner of* $\boldsymbol{I_\delta}$*.*

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* $\sum_{i=1}^{[m/s]} i^{-2} \sim \pi^2/6$,

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

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

$$\text{(C'):} \quad \Pr\left(\frac{1}{T} \boldsymbol{S}_T' \boldsymbol{\Sigma}_2^{-1} \boldsymbol{S}_T > \chi^2_{2,a} \right) \longrightarrow \Pr\left(\chi^2\left(2, \frac{c^2\pi^2}{6}\right) > \chi^2_{2,a} \right),$$

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

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 $S_T'(\alpha|H_{A,1})$, $S_T'(\alpha|H_{A,2})$, $\lambda_T(\boldsymbol{\alpha}|\, H_{A,3})$, 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 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.

——Figures 1 and 2——

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 $(\widetilde{d}_0, \widetilde{d}_s)'$ be the unrestricted CSS estimators of $(d_0, d_s)'$ in (3) by maximizing the CSS function (4). Then it follows that, as $T \to \infty$,

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

where σ_{d_0}, σ_{d_s}, and $\boldsymbol{\Sigma}$ 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.

4 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).

4.1 Results on CSS estimates

The models employed here are

$$\begin{aligned} \text{DGP 1:} \quad & (1-\phi L)(1-L)^{d_0}(1-L^{12})^{d_s}(x_t - 1) = \varepsilon_t, \\ \text{and DGP 2:} \quad & (1-\Phi L^{12})(1-L)^{d_0}(1-L^{12})^{d_s}(x_t - 1) = \varepsilon_t. \end{aligned}$$

Tables 1 and 2 examine the finite sample performance of the estimates discussed in Section 2. 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 $(\widehat{d_0} - d_0, \widehat{\phi} - \phi)'$ and $(\widehat{d_s} - d_s, \widehat{\Phi} - \Phi)'$ 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 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.

——Tables 1 and 2——

4.2 Testing for the integration order

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

$$\begin{aligned} &\text{DGP 3: } (1-\vartheta L)(1-L)^{\alpha} x_t = \varepsilon_t, \quad \text{DGP 4: } (1-\vartheta L)(1-L^{12})^{\alpha} x_t = \varepsilon_t, \\ &\text{DGP 5: } (1-\vartheta L^{12})(1-L)^{\alpha} x_t = \varepsilon_t, \quad \text{DGP 6: } (1-\vartheta L^{12})(1-L^{12})^{\alpha} x_t = \varepsilon_t, \end{aligned}$$

where we fixed $\vartheta = 0.8$ or -0.8 and assumed $\mathrm{E}[x_t] = 0$ 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):

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

where $\widehat{\sigma}_{d_0}$ and $\widehat{\sigma}_{d_s}$ are computed from (15). These have the same asymptotic results in Theorems 2 and 3 by (41). The statistics $\lambda_{T,0S}$ are also LM test statistics, obtained using Godfrey's TR^2 statistics,

[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 $\sum_{j=1}^{1000}(a_j - \bar{a})^2/1000$. The SRMSE is the square root of $\sum_{j=1}^{1000}(a_j - a)^2/1000$, which is equal to the square root of $(\text{bias})^2 + (\text{MASE})^2$.

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 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 3. 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 $\nu_j = (j-1)\pi/6$, $j = 1, \dots, 7$, 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 $\mathbf{\Sigma}$, we used a Hessian (the second-order 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.

——Tables 3 and 4——

5 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 1990 to December 2001 (sum of the ten electric power companies, unit: MWh, sample size: 144) [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.

—— Figures 3 and 4 ——

To search for the best representation of this data, we first fitted differenced data $y_t = (1 - L)(1 - L^{12})x_t$ by the CSS method where we used a sample mean of $\{y_t\}$, $\bar{y}$ as an estimator of $\mathrm{E}[y_t] = \mu$ and set $s = 12$. BIC and AIC criterion are also used under the assumption of normality [see, e.g., Brockwell and Davis (1991, Section 9.3)]. Calculations of BIC and AIC are given by $-2S(\widehat{\boldsymbol{\delta}}, \widehat{\sigma}^2) + \log(\text{sample size used for CSS estimation}) \times (\text{number of estimated parameters})$ and $-2S(\widehat{\boldsymbol{\delta}}, \widehat{\sigma}^2) + 2(\text{number of estimated parameters})$, respectively. Fitting SARFIMA models or SARIMA

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

models are limited to having SARMA parameters with $0 \leq p, q, p_s, q_s \leq 4$ and the total number of estimated SARFIMA parameters (d_0, d_s, SARMA parameters, and σ^2) is less than 8. The total number of models is 772. From among these estimation results, we selected four models in terms of BIC and AIC that satisfy the following conditions: (i) Modified portmanteau tests are not rejected with the significance level 5% until 36 degrees of freedom. (ii) The estimated SARFIMA parameters all converged and satisfy Assumption 1 (c) and (d). Condition (i) uses results of Section 3. All calculations were made using S-PLUS [3].

Table 5 shows the best four models in terms of BIC model selection with estimators. ID denotes the model identification within 772 models. NE indicates the corresponding parameter is not estimated and set to be 0. The numbers in parenthesis in the columns of BIC (AIC) denote the ranking of models in terms of BIC (AIC). These four models have the same number of parameters and show that similar models are selected. Our main concern is whether the $\{x_t\}$ is seasonally overdifferenced (ID 518 and ID 521) or not seasonally overdifferenced (ID 148 and ID 151) because the estimator of d_s in ID 518 (521) appears to relate to the estimator of Θ_1 in ID 148 (151).

—— Table 5 ——

Table 6 shows p-values of testing for the integration order corresponding to those four models using the LM test statistics in Section 3. In this table, models ID 518 and ID 521 correspond to some models in alternative hypotheses of the first and second rows of SARFIMA models, and models ID 148 and ID 151 correspond to null hypotheses of the third and fourth rows of SARFIMA models. Our findings are as follows: (i) Results for SARFIMA$(1, \alpha_0, 0)(3, \alpha_s, 0)_s$ and SARFIMA$(0, \alpha_0, 1)(3, \alpha_s, 0)_s$ support estimation of d_s and restriction of $d_0 = 0$ for models ID 518 and ID 521. (ii) Results for SARFIMA$(1, \alpha_0, 0)(3, \alpha_s, 1)_s$ and SARFIMA$(0, \alpha_0, 1)(3, \alpha_s, 1)_s$ show relatively small p-values. Therefore, we cannot conclude that $\{x_t\}$ is not overdifferenced and d_s should set to be zero.

—— Table 6 ——

The model ID 518 is the best model in terms of BIC and AIC among the 772 model candidates. The estimated model of ID 518 is

$$(1 + 0.221L)(1 + 0.162L^{12} + 0.292L^{24} + 0.393L^{36})(1 - L^{12})^{-0.384}(y_t + 18770.34) = \varepsilon_t,$$
$$y_t = (1 - L)(1 - L^{12})x_t, \qquad \text{and} \qquad \hat{\sigma} = 298563.4.$$

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.

—— Figure 5 ——

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 the above four models in terms of BIC and AIC. In place of the sample mean, we specified the sample median because the electric energy demand can be affected by excessive changes in air temperature and the sample median is robust to additive outliers. Nonetheless, model ID 518 is still selected as the best model in terms of BIC and AIC among the 772 candidates and the rankings and estimates are almost the same.

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.

[3]These programs are available on request.

6 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 (2004)]. Sections 2 and 3 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, $\nu_1, \nu_2, \ldots, \nu_k$ in (1), are known.

Section 2 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 1 to the following linear regression model:

$$y_t = \widetilde{\varphi}_{t\cdot}\boldsymbol{\beta} + x_t, \qquad (1-L)^{d_0}(1-L^s)^{d_s}x_t = \vartheta(L)\varepsilon_t, \qquad (t = 1, 2, \ldots, T).$$

In this case, consistency of the least-squares estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, depends on differencing parameters, i.e., $\mathrm{Var}[\boldsymbol{D}_T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})]$ 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 $\mathrm{Var}[\boldsymbol{D}_T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})] = O(\sum_{j=0}^{T} |\gamma(j)|)$, as $T \to \infty$ [see, e.g., Section 9.1 in Fuller (1966) and Section 2 in Yajima (1988)]. But we cannot prove consistency and asymptotic normality of CSS estimates $(\widehat{\boldsymbol{\delta}}', \widehat{\sigma}^2)$. 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, $(d_0 + d_s, d_s) \in D^s_{1,3}$, which is different from that of the ARFIMA model. In Section 3, 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.

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

$$F(z) = (1-z)^{-d_1}(1+z)^{-d_k}\prod_{i=2}^{k-1}(1-2\eta_i z + z^2)^{-d_i} = \prod_{i=1}^{k}(1-2\eta_i z+z^2)^{-D_i} = \sum_{j=0}^{\infty}\psi_j z^j, \tag{22}$$

$|z| < 1$, 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 $i = 2, \ldots, k-1$. Then

1. *$\psi_0 = 1$, and*

$$\psi_j = \frac{2}{j}\sum_{i=0}^{j-1}\sum_{m=1}^{k} D_m \cos[(j-i)\nu_m]\,\psi_i, \text{ for } j \geq 1, \tag{23}$$

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

$$\psi_j \sim \sum_{i=1}^{k}\frac{\kappa_i(j)}{\Gamma(d_i)}\, j^{d_i-1}, \text{ as } j \to \infty, \tag{24}$$

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

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

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

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

$$\sum_{j=0}^{n}\psi_{1,j}(d)=\frac{n+1}{d}\psi_{1,n+1}(d)=\psi_{1,n}(d+1)\sim\frac{n^d}{\Gamma(d+1)},\tag{25}$$

$$\sum_{j=0}^{n}|\psi_{1,j}(d)|=\begin{cases}\displaystyle\sum_{j=0}^{n}\psi_{1,j}(d)\sim\frac{n^d}{\Gamma(d+1)}, & \text{if } d\in(0,1),\\ \displaystyle 2-\psi_{1,n}(d+1)\sim 2-\frac{n^d}{\Gamma(d+1)}, & \text{if } d\in(-1,0),\end{cases}\tag{26}$$

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

4. [The summability of Gegenbauer porlynomials by Theorem (2.1) in Zayed (1980)][4]. *Let Gegenbauer polynomials be* $C_j^d(\eta)$, $j=0,1,2,\dots$, *which are defined by the generating relation* $(1-2\eta z+z^2)^{-d}=\sum_{j=0}^{\infty}C_j^d(\eta)z^j$, $|\eta|<1,|z|<1$. *If* $d\in(-1,0)\cup(0,1)$, $A=\sum_{j=0}^{\infty}a_j$ *is convergent,* $B=\sum_{j=0}^{\infty}b_j$, *and* $b_j=\sum_{k=0}^{j}a_{j-k}C_k^d(\eta)$, *then* $C=\sum_{j=0}^{\infty}C_j^d(\eta)$ *is convergent, and* $B=AC$.

5. *Let* $\psi_{k,j}(d)$ *be defined by* $(1+z)^{-d}=\sum_{j=0}^{\infty}\psi_{k,j}(d)z^j$, $\psi_{k,j}(d)=(-1)^j\psi_{1,j}(d)$, *and* $d\in(-1,0)\cup(0,1)$, *where* $\psi_{1,j}(d)$ *is given by 3. Then* $\sum_{j=0}^{\infty}\psi_{k,j}(d)$ *is convergent.*

6. *Let, in (22),* $d_i\in(-1,0)\cup(0,1)$ *for* $i=1,2,\dots,k$. *If* $d_1\in(-1,0)$, *then* $\sum_{j=0}^{\infty}\psi_j$ *is convergent, and*

$$\sum_{j=0}^{n}\psi_j\sim\kappa_1\psi_{1,n}(d+1)\sim\frac{\kappa_1}{\Gamma(d_1+1)}n^{d_1},\quad \text{as } n\to\infty,\tag{27}$$

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

Proof of Lemma A 1. 1. By the nth derivative, we have $F^{(n)}(0)=n!\,\psi_n$, while for the middle term of (22), the first derivative is given by

$$F^{(1)}(z)=F(z)\left(\frac{-d_1}{z-1}+\frac{-d_k}{z+1}+\sum_{j=2}^{k-1}\frac{-2d_j(z-\eta_j)}{(z-e^{i\nu_j})(z-e^{-i\nu_j})}\right)\equiv F(z)G(z),\quad \text{(say)}.$$

Using the nth derivative, we have

$$G^{(n)}(0)=n!\,d_1+(-1)^{n+1}n!\,d_k+\sum_{j=2}^{k-1}2d_jn!\cos[(1+n)\nu_j].$$

Hence, by $F^{(n)}(0)=\{F(0)G(0)\}^{(n-1)}=n!\,\psi_n$ and the nth derivative of a product formula [Leibniz's rule, see 0.42 in Gradshteyn and Ryzhik (2000)], we obtain the result.

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

3. By (23) and $\psi_{1,n}(d) \sim n^{d-1}/\Gamma(d)$ as $n \to \infty$, we obtain (25). It is also obtained from properties of Cesàro numbers and the Pochhammer symbol [see, e.g., Section 1 of Chapter 3 in Zygmund (1968) and Proposition 1.2.3 in Dunkl and Xu (2001)]. To demonstrate (26), we note that $\Gamma(\nu) > 0$ for $\nu > 0$ and $\Gamma(\nu) < 0$ for $\nu \in (-1, 0)$. For the case of $d \in (-1, 0)$, since

$$\sum_{j=0}^{n} |\psi_{1,j}(d)| = 1 + \sum_{j=1}^{n} \frac{\Gamma(j+d)}{|\Gamma(d)|\Gamma(j+1)} = 1 - \sum_{j=1}^{n} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} = 2 - \sum_{j=0}^{n} \psi_{1,j}(d),$$

we obtain the result by (25). The case of $d \in (0,1)$ is obvious because $|\psi_{1,j}(d)| = \psi_{1,j}(d)$, $j \geq 0$.

4. The proof of the case of $d \in (0,1)$ draws heavily on Theorem (2.1) in Zayed (1980). Consider the function

$$G(\eta) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{(-1)^l}{(j+d)^{2l}} C_j^d(\eta) a_{i-j} = \sum_{i=0}^{\infty} \sum_{j=0}^{i} d_j(\eta) a_{i-j}$$

where $d_j(\eta) = (-1)^l C_j^d(\eta)/\{(j+d)^{2l}\}$ and $2l \geq 2d+1$. Since $C_j^d(\cos\nu)$ is expressed as $C_j^d(\cos\nu) = \sum_{i=0}^{j} \psi_{1,i}(d)\psi_{1,j-i}(d)\cos[(2i-j)\nu]$, we have $|C_j^d(\eta)| \leq C_1 j^{2d-1}$, $|d_j(\eta)| \leq C_1 j^{2d-1}/\{(j+d)^{2l}\}$, and $\sum_{j=0}^{\infty} d_j(\eta)$ converges absolutely by our choice of l. Using the same argument as in the proof of Mertens' Theorem [e.g., see Chapter X of Hardy (1991)], the RHS of

$$\left| \sum_{i=0}^{n} \sum_{j=0}^{i} d_j(\eta) a_{i-j} \right| = \left| \sum_{j=0}^{n} d_j(\eta) \sum_{i=0}^{n-j} a_i \right| \leq \sum_{j=0}^{n} \frac{C_1 j^{2d-1}}{(j+d)^{2l}} \left| \sum_{i=0}^{n-j} a_i \right|$$

is convergent as $n \to \infty$ and $G(\eta)$ is uniformly convergent. Borrowing the differential operator in (2.5) of Zayed (1980), $D = (1-\eta^2)\partial^2/\partial\eta^2 - (2d+1)\eta\partial/\partial\eta - d^2$, and Weierstrass's Double Series Theorem, we have $D^l G(\eta) = B$. Since A, B, and C are all convergent, $B = AC$ by Abel's Theorem [see Chapter X of Hardy (1991)].

The case of $d \in (-1,0)$ can be treated similarly because $\sum_{j=0}^{\infty} C_j^d(\eta)$ is absolutely convergent.

5. To prove the case of $d \in (0,1)$, it is sufficient to check the conditions of Leibniz's Theorem: (i) $\psi_{1,j}(d) > 0$, for all j, (ii) $\psi_{1,j}(d) \to 0$ as $j \to \infty$, and (iii) $\psi_{1,j}(d) \geq \psi_{1,j+1}(d)$, for all j because $\sum_{j=0}^{\infty} \psi_{k,j}(d) = \sum_{j=0}^{\infty} (-1)^j \psi_{1,j}(d)$ is the alternating series. Since $\Gamma(\nu) > 0$ for $\nu > 0$, $\psi_{1,j}(d)$ is positive. Furthermore, equation (7) of Yajima (1985):

$$(n+1)^{t-1} \leq \frac{\Gamma(n+t)}{\Gamma(n+1)} \leq n^{t-1}, \quad \text{for } 0 \leq t \leq 1, \text{ and } n = 1, 2, \ldots, \tag{28}$$

implies (ii) and (iii).

The proof of the case of $d \in (-1,0)$ is obtained easily because $\sum_{j=0}^{\infty} \psi_{k,j}(d)$ is absolutely convergent.

6. We first rewrite (22) as $F(z) = G_1(z)G_2(z)$, where $G_1(z) = (1-z)^{-d_1} = \sum_{j=0}^{\infty} \psi_{1,j}(d_1) z^j$ and $G_2(z) = (1+z)^{-d_k} \prod_{i=2}^{k-1} (1 - 2\eta_i z + z^2)^{-d_i} = \sum_{j=0}^{\infty} g_{2,j} z^j$.

For the case of $d_1 \in (-1, 0)$, by 4 and 5, $G_2(1) = \sum_{j=0}^{\infty} g_{2,j}$ is convergent, and, by 3, $\sum_{j=0}^{\infty} \psi_{1,j}(d_1)$ is absolutely convergent. Then $\sum_{j=0}^{\infty} \psi_j$ is convergent and

$$F(1) = \sum_{j=0}^{\infty} \psi_j = \left(\sum_{j=0}^{\infty} \psi_{1,j}(d_1) \right) \left(\sum_{j=0}^{\infty} g_{2,j} \right) = G_1(1)G_2(1) = 0 \tag{29}$$

where we have used Mertens' Theorem and Abel's Theorem again. It follows from (29) and (25) that, as $n \to \infty$,

$$\sum_{j=0}^{n} \psi_j \sim \left(\sum_{j=0}^{n} \psi_{1,j}(d_1) \right) \left(\sum_{j=0}^{n} g_{2,j} \right) \sim \left(\sum_{j=0}^{n} \psi_{1,j}(d_1) \right) G_2(1) \sim \frac{n^{d_1}}{\Gamma(1+d_1)} G_2(1).$$

For the case of $d_1 \in (0,1)$, since $\sum_{j=0}^{\infty} g_{2,j}$ is convergent and $\sum_{j=0}^{n} |\psi_{1,j}(d_1)| \sim n^{d_1}/\Gamma(1+d_1)$ by (26), we have, as $n \to \infty$,

$$\left|\sum_{j=0}^{n} \psi_j\right| = \left|\sum_{j=0}^{n} \psi_{1,j}(d_1) \sum_{k=0}^{n-j} g_{2,k}\right| \leq \sum_{j=0}^{n} |\psi_{1,j}(d_1)| \left|\sum_{k=0}^{n-j} g_{2,k}\right| = O\left(\sum_{j=0}^{n} |\psi_{1,j}(d_1)|\right) = O(n^{d_1}).$$

□

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

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

be the residual process for evaluating the CSS function,

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

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

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

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

We show that $\widehat{\boldsymbol{\delta}}$ is a consistent estimator of $\boldsymbol{\delta}$ by showing that

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(\ddot{\boldsymbol{\delta}})^2 \xrightarrow{p} \mathrm{E}[u_t(\ddot{\boldsymbol{\delta}})]^2, \quad \text{as } T \to \infty \text{ uniformly in } \ddot{\boldsymbol{\delta}} \in D_{i,j}^s \tag{30}$$

because $\widehat{\boldsymbol{\delta}}$ is the estimator of $\boldsymbol{\delta}$ that minimizes the objective function $\sum_{t=1}^{T} \varepsilon_t^2(\ddot{\boldsymbol{\delta}})/T$. This is sufficient condition for weak consistency by Fuller (1996, Lemma 5.5.1 and Lemma 5.5.2) because $\mathrm{E}[u_t^2(\ddot{\boldsymbol{\delta}})]$ reaches its minimum at $\boldsymbol{\delta}$ by the fact that $-\sum_{k=1}^{\infty} \pi_k(d_0, d_s)v_{t-k}(\boldsymbol{\delta})$ uniquely determines the best linear predictor of $v_t(\boldsymbol{\delta})$ on the basis of the mean squared error based on the infinite past $v_{t-1}(\boldsymbol{\delta}), v_{t-2}(\boldsymbol{\delta}), \ldots$ (i.e., $\varepsilon_t = u_t(\boldsymbol{\delta}) = v_t(\boldsymbol{\delta}) + \sum_{k=1}^{\infty} \pi_k(d_0, d_s)v_{t-k}(\boldsymbol{\delta})$), which establishes the condtion (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 $c_j = \sum_{k=0}^{j} a_k b_{j-k}$, $j \geq 0$. Then $|c_j| \leq C j^{-(\tau+1)}$, 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

$$\begin{aligned}
|c_j| &\leq \sum_{k=0}^{[j/2]} \frac{C_1|a_k|}{(j-k+1)^{\tau+1}} + \sum_{k=[j/2]+1}^{j} \frac{C_1|b_{j-k}|}{(k+1)^{\tau+1}} \\
&\leq \frac{C_1}{(j-[j/2]+1)^{\tau+1}}\sum_{k=0}^{[j/2]} |a_k| + \frac{C_1}{([j/2]+2)^{\tau+1}}\sum_{k=[j/2]+1}^{j} |b_{j-k}| \\
&\leq C_2\,(j/2+1)^{-\tau-1} \leq C\,j^{-\tau-1}
\end{aligned}$$

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 $\ddot{\boldsymbol{\delta}} \in D^s_{1,1}$. Then (i) there exist absolutely summable sequences $\{\pi_{j,0}(\tau)\}$, which do not depend on $\ddot{\boldsymbol{\delta}}$, 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 $\ddot{\boldsymbol{\delta}}$, 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 $(1-z)^a(1-z^s)^b = (1-z)^{a+b}(1+z)^b \prod_{j=1}^{s/2-1}(1 - 2\cos(2\pi j/s)z + z^2)^b$ 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 $(1-z)^d = \sum_{j=0}^{\infty} a_j(d) z^j$. Then, equation (28) 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 $(1 - 2\cos(\theta)z + z^2)^{-\nu} = \sum_{j=0}^{\infty} C_j^{\nu}(\cos\theta) z^j$,

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

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

Lemma B 4. $\sum_{t=1}^T \varepsilon_t^2(\ddot{\boldsymbol{\delta}}, \mu)/T - \sum_{t=1}^T u_t^2(\ddot{\boldsymbol{\delta}})/T \to 0$ *a.c. as $T \to \infty$ uniformly in $\ddot{\boldsymbol{\delta}} \in D^s_{1,1}$.*

Proof. Rewriting $\varepsilon_t(\ddot{\boldsymbol{\delta}}, \mu)$ and $u_t(\ddot{\boldsymbol{\delta}})$ as

$$\begin{aligned}
\varepsilon_t(\ddot{\boldsymbol{\delta}}, \mu) &= \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) \sum_{k=0}^{t-j-1} \psi_k(d_0, d_s)\varepsilon_{t-j-k} = \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) e_{t-j}(\boldsymbol{\delta}), \\
u_t(\ddot{\boldsymbol{\delta}}) &= \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\boldsymbol{\delta}) + \sum_{j=t}^{\infty} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\boldsymbol{\delta}) \\
&= \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) e_{t-j}(\boldsymbol{\delta}) + \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) v'_{t-j}(\boldsymbol{\delta}) + \sum_{j=t}^{\infty} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\boldsymbol{\delta}) \\
&= \varepsilon_t(\ddot{\boldsymbol{\delta}}, \mu) + \sum_{j=0}^{t-1} \pi_j(\ddot{d}_0, \ddot{d}_s) v'_{t-j}(\boldsymbol{\delta}) + \sum_{j=t}^{\infty} \pi_j(\ddot{d}_0, \ddot{d}_s) v_{t-j}(\boldsymbol{\delta}) \\
&= \varepsilon_t(\ddot{\boldsymbol{\delta}}, \mu) + w_{t,1} + w_{t,2}, \quad \text{(say)},
\end{aligned}$$

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

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

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

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

and $\mathrm{E}\, z_{t,0}^4 = O(1)$. 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. $\sum_{t=1}^T u_t^2(\ddot{\boldsymbol{\delta}})/T \to \mathrm{E}[u_t^2(\ddot{\boldsymbol{\delta}})]$ *a.c. as* $T \to \infty$ *uniformly in* $\ddot{\boldsymbol{\delta}} \in D_{1,1}^s$.

Proof. For fixed $\ddot{\boldsymbol{\delta}}$, we have $\sum_{t=1}^T u_t^2(\ddot{\boldsymbol{\delta}})/T \to \mathrm{E}[u_t^2(\ddot{\boldsymbol{\delta}})]$. Therefore, the rest of the proof is devoted to showing uniformity in $\ddot{\boldsymbol{\delta}} \in D_{1,1}^s$. By Lemma B 2, we have $|u_t(\ddot{\boldsymbol{\delta}})| \le w_{t,1}$ and $|\partial u_t(\ddot{\boldsymbol{\delta}})/\partial \ddot{\boldsymbol{\delta}}| \le w_{t,2}$ where $w_{t,1} = \sum_{j=0}^\infty \pi_{j,0}(\tau)|v_{t-j}(\boldsymbol{\delta})|$ and $w_{t,2} = \sum_{j=1}^\infty \pi_{j,1}(\tau)|v_{t-j}(\boldsymbol{\delta})|$. Let $\Omega_1 = \{\omega | \lim_{T\to\infty} \sum_{t=1}^T w_{t,i}^2/T = \mathrm{E}\, w_{t,i}^2,\ i = 1,2\}$ and $D_0 = \{\boldsymbol{\delta}_i | i = 1,2,\ldots\}$ be a countable dense subset of $D_{1,1}^s$. Put $\Omega_{\delta(i)} = \{\omega | \lim_{T\to\infty} \sum_{t=1}^T u_t^2(\boldsymbol{\delta}_i)/T = \mathrm{E}[u_t(\boldsymbol{\delta}_i)]^2\}$ and $\Omega = \Omega_1 \cap_i \Omega_{\delta(i)}$, we have $\Pr(\Omega) = 1$ since $u_t(\boldsymbol{\delta}_i)$ 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. $\sum_{t=1}^T \varepsilon_t^2(\ddot{\boldsymbol{\delta}}, \mu)/T \to \mathrm{E}[u_t^2(\ddot{\boldsymbol{\delta}})]$ *a.c. as* $T \to \infty$ *uniformly in* $\ddot{\boldsymbol{\delta}} \in D_{1,1}^s$.

Proof. Using the triangle inequality, Lemmas B 4 and B 5, we immediately obtain the result. □

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

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

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

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

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

Lemma B 9. *Lemmas B 4 - B 8 still hold if* $D_{1,1}^s$ *is replaced by* $D_{i,j}^s$ *for* $i, j = 1, 2, 3$.

Proof. For the case of $\ddot{\boldsymbol{\delta}}, \boldsymbol{\delta} \in D_{2,1}^s$, rewrite $\varepsilon_t(\ddot{\boldsymbol{\delta}})$ and $u_t(\ddot{\boldsymbol{\delta}})$ as

$$\begin{aligned}
\varepsilon_t(\ddot{\boldsymbol{\delta}}) &= \sum_{j=0}^{t-1} \pi_j\left(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s\right) \sum_{k=0}^{t-j-1} \psi_k\left(d_0 + \frac{1}{4}, d_s\right)\varepsilon_{t-j-k} = \sum_{j=0}^{t-1} \pi_j\left(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s\right) e_{t-j}\left(d_0 + \frac{1}{4}, d_s\right), \\
u_t(\ddot{\boldsymbol{\delta}}) &= \varepsilon_t(\ddot{\boldsymbol{\delta}}) + \sum_{j=0}^{t-1} \pi_j\left(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s\right) v'_{t-j}\left(d_0 + \frac{1}{4}, d_s\right) + \sum_{j=t}^{\infty} \pi_j\left(\ddot{d}_0 + \frac{1}{4}, \ddot{d}_s\right) v_{t-j}\left(d_0 + \frac{1}{4}, d_s\right),
\end{aligned}$$

where

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

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 $|\partial^{i+k}\pi_j(\ddot{d}_0+1/4,\ddot{d}_s)/(\partial d_0^i\partial d_s^k)|\le\pi'_{j,i+k}(\tau)$ and $\pi'_{j,i+k}(\tau)=O((\log j)^{i+k}(j^{-\tau-1}+j^{\tau-5/4}))$ for $i+k=0,\dots,3$ because $\ddot{d}_0+\ddot{d}_s+1/4\in(1/4-\tau,1/4+\tau)$ and $\ddot{d}_s\in(\tau,1/2-\tau)$. It follows that the rest of the proof relating to $D_{2,1}^s$ is obtained in the same way as those in Lemmas B 4 to B 8. Since other $D_{i,j}^s$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 $t\le 0$.

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

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

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

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

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 $\bar{x}\xrightarrow{a.c.}\mu$. □

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

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

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

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

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

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

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

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

and (34) holds. Other cases can be treated similarly because each derivative of $\sum_{j=0}^{t-1}\pi_j(\ddot{d}_0+a,\ddot{d}_s)$ 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 $\widehat{\boldsymbol{\delta}}$ by showing (30). Using Lemmas B 6, B 9, B 11, and

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

the condition (30) is established.

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

$$\begin{aligned}\left|\widehat{\sigma}^2-\sigma^2\right| &\le \left|\sum_{t=1}^{T}\varepsilon_t^2(\widehat{\boldsymbol{\delta}})/T-\mathrm{E}[u_t^2(\widehat{\boldsymbol{\delta}})]\right|+\left|\mathrm{E}[u_t^2(\widehat{\boldsymbol{\delta}})]-\sigma^2\right|\\ &\le \sup_{\ddot{\boldsymbol{\delta}}}\left|\sum_{t=1}^{T}\varepsilon_t^2(\ddot{\boldsymbol{\delta}})/T-\mathrm{E}[u_t^2(\ddot{\boldsymbol{\delta}})]\right|+\left|\mathrm{E}[u_t^2(\widehat{\boldsymbol{\delta}})]-\sigma^2\right|\xrightarrow{p}0.\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

Since the process $\varepsilon_t(\boldsymbol{\delta},\mu)\{\partial\varepsilon_t(\boldsymbol{\delta},\mu)/(\partial\boldsymbol{\delta})\}$ 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 $\varepsilon_t(\boldsymbol{\delta},\mu)$ with respect to $\boldsymbol{\delta}$ is given by (7) and each element of $\{\boldsymbol{\delta}_k\}$ is defined as follows:

$$\frac{\partial \varepsilon_t(\boldsymbol{\delta},\mu)}{\partial d_0} = \log(1-L)\varepsilon_t = -\sum_{k=1}^{\infty}\frac{1}{k}L^k\varepsilon_t, \quad \frac{\partial \varepsilon_t(\boldsymbol{\delta},\mu)}{\partial d_s} = \log(1-L^s)\varepsilon_t = -\sum_{k=1}^{\infty}\frac{1}{k}L^{ks}\varepsilon_t = -\sum_{k=1}^{\infty}s_k L^k\varepsilon_t,$$
$$\frac{\partial \varepsilon_t(\boldsymbol{\delta},\mu)}{\partial \phi_j} = -\phi^{-1}(L)L^j\varepsilon_t = -L^j\sum_{k=0}^{\infty}\phi_k^* L^k\varepsilon_t = -\sum_{k=j}^{\infty}\phi_{k-j}^* L^k\varepsilon_t \qquad \text{for } j=1,\ldots,p,$$
$$\frac{\partial \varepsilon_t(\boldsymbol{\delta},\mu)}{\partial \theta_j} = -\theta^{-1}(L)L^j\varepsilon_t = -L^j\sum_{k=0}^{\infty}\theta_k^* L^k\varepsilon_t = -\sum_{k=j}^{\infty}\theta_{k-j}^* L^k\varepsilon_t \qquad \text{for } j=1,\ldots,q,$$
$$\frac{\partial \varepsilon_t(\boldsymbol{\delta},\mu)}{\partial \Phi_j} = -\Phi^{-1}(L^s)L^{js}\varepsilon_t = -L^{js}\sum_{k=0}^{\infty}\Phi_k^* L^k\varepsilon_t = -\sum_{k=js}^{\infty}\phi_{k-js}^* L^k\varepsilon_t \qquad \text{for } j=1,\ldots,p_s,$$
$$\frac{\partial \varepsilon_t(\boldsymbol{\delta},\mu)}{\partial \Theta_j} = -\Theta^{-1}(L^s)L^{js}\varepsilon_t = -L^{js}\sum_{k=0}^{\infty}\Theta_k^* L^k\varepsilon_t = -\sum_{k=js}^{\infty}\Theta_{k-js}^* L^k\varepsilon_t \qquad \text{for } j=1,\ldots,q_s, \tag{38}$$

where $s_j = s/j$ for $j = s, 2s, \ldots,$ $:= 0$ otherwise, ϕ_j^*, θ_j^*, Φ_j^* and Θ_j^* are defined by Lemma 1. The second derivatives can be obtained similarly, which establish $\boldsymbol{I}_{\boldsymbol{\delta}}$. □

Remark 1. If μ is known and $\bar{x}$ of $\varepsilon_t(\ddot{\boldsymbol{\delta}}) = \varepsilon_t(\ddot{\boldsymbol{\delta}}, \bar{x})$ is replaced by μ, then $\widehat{\boldsymbol{\delta}}$ is a strongly consistent estimator because Lemmas B 6 and B 9 hold and $\mathrm{E}[u_t(\ddot{\boldsymbol{\delta}})]^2$ reaches its minimum at $\boldsymbol{\delta}$ simiarly to (30) [see, e.g., Fuller (1996, Lemma 5.5.2)]. It implies strong consistency of $\widehat{\sigma}^2$ and asymptotic normality of (6) similarly to the proof of Theorem 1.

Section 3 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 $\widehat{d}_s$ is given by evaluating the residual, $\varepsilon'_t(\ddot{d}_s) = \sum_{j=0}^{t-1}\pi_j(d_0, \ddot{d}_s)x_{t-j}$ in place of $\varepsilon_t(\ddot{\boldsymbol{\delta}}, \bar{x})$, the property of strong consistency of $\widehat{d}_s$ is immediately obtained.

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

$$\begin{aligned} u'_t(\ddot{d}_s) &= (1-L)^{d_0}(1-L^s)^{\ddot{d}_s}v'_t = (1-L)^{d_0}(1-L^s)^{\ddot{d}_s}(1-L)^{-(d_0+\theta)}(1-L^s)^{-d_s}\varepsilon_t \\ &= (1-L)^{-\theta}(1-L^s)^{-(d_s-\ddot{d}_s)}\varepsilon_t = (1-L)^{-\theta}w_t(\ddot{d}_s) \end{aligned}$$

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 $T > T_0$,

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

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 $T \to \infty$, $\sum_{t=1}^{T} u_t'(\ddot{d}_s)^2/T - \sum_{t=1}^{T} w_t(\ddot{d}_s)^2/T \to 0$ a.c. uniformly in $\ddot{d}_s$, and hence $\sum_{t=1}^{T} \varepsilon_t'(\ddot{d}_s)^2/T - \sum_{t=1}^{T} w_t(\ddot{d}_s)^2/T \to 0$ a.c. uniformly in $\ddot{d}_s$.

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

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

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

For the case of $D_{1,1}^s$, let $u_t(\ddot{\boldsymbol{\delta}}) = \sum_{k=0}^{\infty} \pi_k(\ddot{d}_0, \ddot{d}_s) v_{t-k}(\boldsymbol{\delta})$, $v_t(\boldsymbol{\delta}) = \sum_{k=0}^{\infty} \psi_k(d_{T,0}, d_s)\varepsilon_{t-k}$, and $w_t(\ddot{\boldsymbol{\delta}}) = \sum_{j=0}^{\infty} \psi_j(d_0 - \ddot{d}_0, d_s - \ddot{d}_s)\varepsilon_{t-j}$. Then by a straightforward extension of the method used by Lemmas B 4 and B 5, we have $\sum_{t=1}^{T} \varepsilon_t^2(\ddot{\boldsymbol{\delta}})/T - \sum_{t=1}^{T} u_t^2(\ddot{\boldsymbol{\delta}})/T \to 0$ and $\sum_{t=1}^{T} w_t^2(\ddot{\boldsymbol{\delta}})/T \to \mathrm{E}[w_t^2(\ddot{\boldsymbol{\delta}})]$ a.c. as $T \to \infty$ uniformly in $\ddot{\boldsymbol{\delta}} \in D_{1,1}^s$. 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 $\sum_{t=1}^{T} u_t^2(\ddot{\boldsymbol{\delta}})/T - \sum_{t=1}^{T} w_t^2(\ddot{\boldsymbol{\delta}})/T \to 0$ and $\mathrm{E}[u_t(\ddot{\boldsymbol{\delta}})]^2 - \mathrm{E}[w_t(\ddot{\boldsymbol{\delta}})]^2 \to 0$ a.c. as $T \to \infty$ uniformly in $\ddot{\boldsymbol{\delta}} \in D_{1,1}^s$. It follows that $\sum_{t=1}^{T} \varepsilon_t^2(\ddot{\boldsymbol{\delta}})/T - \mathrm{E}[u_t(\ddot{\boldsymbol{\delta}})]^2 \to 0$ a.c. as $T \to \infty$ uniformly in $\ddot{\boldsymbol{\delta}} \in D_{1,1}^s$. Hence, by Gallant and White (1988, Theorem 3.3), the proof of strong consistency of $\widetilde{\boldsymbol{\delta}}$ is obtained easily by the fact that $-\sum_{k=1}^{\infty} \pi_k(d_{T,0}, d_s) v_{t-k}(\boldsymbol{\delta})$ uniquely determines the best linear predictor of $v_t(\boldsymbol{\delta})$ on the basis of the mean squared error based on the infinite past $v_{t-1}(\boldsymbol{\delta}), v_{t-2}(\boldsymbol{\delta}), \ldots$.

When $\widetilde{\boldsymbol{\delta}}$ is given by evaluating the residual, $\varepsilon_t(\ddot{\boldsymbol{\delta}}, \bar{x})$, a weak consistency of $\widehat{d}_s$ 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 $T \to \infty$, $\sqrt{T}(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{I}_{\boldsymbol{\delta}}^{-1})$,

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

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

C Asymptotic Results Relating to Residual Autocorrelation Functions

Proof of Lemma 1. The proof of Lemma 1 is obtained similarly to the proof of Theorem 1 of McLeod (1978).

First we assume that μ is known and $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\delta}}, \mu)$. By a Taylor series expansion of $\widehat{\boldsymbol{r}}$ around $\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta}$, we have, as $T \to \infty$,

$$\sqrt{T}\widehat{\boldsymbol{r}} = \sqrt{T}\boldsymbol{r} + \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\delta}'}\sqrt{T}\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) + o_p(1) = \begin{pmatrix} -\boldsymbol{J}_m & \boldsymbol{I}_m \end{pmatrix} \begin{pmatrix} \sqrt{T}\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) \\ \sqrt{T}\,\boldsymbol{r} \end{pmatrix} + o_p(1), \tag{40}$$

where $\boldsymbol{r} = (r(1), \ldots, r(m))'$, $r(j) = \sum_{t=1}^{T-j} \varepsilon_t \varepsilon_{t+j} / \sum_{t=1}^{T} \varepsilon_t^2$, $j = 1, \ldots, m$, and the last equality follows from equations (32) and (33) of McLeod (1978). Note that, as $T \to \infty$, $\mathrm{Var}(r(j)) \sim (T-j)/\{T(T+2)\}$ and $\mathrm{Cov}(r(j), r(k)) \sim 0$ for $j \neq k$. As $T \to \infty$, $\sqrt{T}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{I}_{\boldsymbol{\delta}}^{-1})$ by Theorem 1 and it is known that $\sqrt{T}\boldsymbol{r} \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{I}_m)$. Using the same argument as in the proof of Lemma 2 and Lemma 3 of McLeod (1978), we obtain

$$\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\delta}} - \boldsymbol{\delta} \\ \boldsymbol{r} \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{I_\delta}^{-1} & \boldsymbol{I_\delta}^{-1}\boldsymbol{J}'_m \\ \boldsymbol{J}_m\boldsymbol{I_\delta}^{-1} & \boldsymbol{I}_m \end{pmatrix}\right), \text{ as } T \to \infty,$$

which yields Lemma 1 by (40). When μ is unknown and $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\delta}}, \bar{x})$, it can be shown that by Lemma B 11, (36), and (37), existence of $\bar{x}$ does not affect the limiting distribution of $\sqrt{T}\widehat{\boldsymbol{r}}$, which establishes the lemma [see also the proofs of Lemma 2 and Theorem 1 of McLeod (1978)]. □

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

Lemma C 1. *Let $w_{1,T,j}$, $w_{2,T,j}$, and $w_{3,T,j}$, $j = 1, 2, \ldots, T$ be triangular array of random variables such that $\sum_{j=1}^T w_{1,T,j}^2 = O_p(T^{1/2})$, $\sum_{j=1}^T w_{2,T,j}^2 = O_p(1)$, and $\sum_{j=1}^T w_{3,T,j}^2 = O_p(1)$ 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–Schuwarz inequality and the fact that $\sum_{k=m+1}^{\infty} k^{-1-a} \le a^{-1}m^{-a}$ for any $a > 0$, there exists a number $T_0 > 0$ and for all $T > T_0$,

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{k=m+1}^{T-1} a_k \sum_{j=k+1}^{T} w_{1,T,j-k} w_{2,T,j} &\le \frac{1}{\sqrt{T}} \sum_{k=m+1}^{T-1} a_k \left(\sum_{j=k+1}^{T} w_{1,T,j-k}^2\right)^{1/2} \left(\sum_{j=k+1}^{T} w_{2,T,j}^2\right)^{1/2} \\
&= O_p\left(T^{-1/4} \sum_{k=m+1}^{T-1} a_k\right) = O_p\left(\sum_{k=m+1}^{\infty} k^{-1/4} a_k\right) = O_p\Big(m^{-1/4}\Big), \\
\frac{1}{\sqrt{T}} \sum_{k=m+1}^{T-1} \frac{1}{k} \sum_{j=k+1}^{T} w_{2,T,j-k} w_{3,T,j} &\le \frac{1}{\sqrt{T}} \sum_{k=m+1}^{T-1} a_k \left(\sum_{j=k+1}^{T} w_{2,T,j-k}^2\right)^{1/2} \left(\sum_{j=k+1}^{T} w_{3,T,j}^2\right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{T}} \sum_{k=m+1}^{T-1} a_k\right) = O_p\left(\sum_{k=m+1}^{\infty} k^{-1/2} a_k\right) = O_p\Big(m^{-1/2}\Big),
\end{aligned}$$

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. □

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

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

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

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

$$\begin{aligned}
\sqrt{T}\Big(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\Big) &= \left(-\frac{1}{T} \frac{\partial^2 S(\alpha_0, \boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'}\right)^{-1} \left(\frac{1}{\sqrt{T}} \frac{\partial S(\alpha_0, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} - \frac{c}{T} \frac{\partial^2 S(\alpha_0, \boldsymbol{\xi})}{\partial \alpha_0 \partial \boldsymbol{\xi}}\right) + o_p(1) \\
&\xrightarrow{d} N(\, \boldsymbol{I}_\xi^{-1} \boldsymbol{I}_{\alpha_0 \xi}\, c,\ \boldsymbol{I}_\xi^{-1}\,)
\end{aligned}$$

where $\boldsymbol{I}_\xi = -\lim_{T\to\infty} T^{-1}E[\partial^2 S(\alpha_0,\boldsymbol{\xi})/(\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}')]$ and $\boldsymbol{I}_{\alpha_0\xi} = -\lim_{T\to\infty} T^{-1}E[\partial^2 S(\alpha_0,\boldsymbol{\xi})/(\partial\alpha_0\partial\boldsymbol{\xi})]$. For $\widehat{\boldsymbol{r}}$, since $\widehat{r}(i)$ consists of $\widehat{\boldsymbol{\xi}}$ and $\alpha_0 = 0$, by a Taylor series expansion as in (40), we have, as $T\to\infty$,

$$
\begin{aligned}
\sqrt{T}\,\widehat{\boldsymbol{r}} &= \left(\frac{\partial \boldsymbol{r}}{\partial\boldsymbol{\xi}'}\ \boldsymbol{I}_m\right)\begin{bmatrix}\sqrt{T}\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)\\ \sqrt{T}\,\boldsymbol{r}\end{bmatrix} - c\,\frac{\partial\,\boldsymbol{r}}{\partial\alpha_0} + o_p(1)\\
&= \left(-\boldsymbol{J}_{m\xi}\ \boldsymbol{I}_m\right)\left[\begin{pmatrix}\sqrt{T}\left(\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)\\ \sqrt{T}\,\boldsymbol{r}\end{pmatrix} + \begin{pmatrix} c\,\boldsymbol{I}_{\boldsymbol{\delta}}{}^{-1}\boldsymbol{I}_{\alpha_0\xi}\\ \boldsymbol{0}\end{pmatrix}\right] + \boldsymbol{J}_{m\alpha_0}\,c + o_p(1)
\end{aligned}
$$

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

Finally, we will show that

$$
\lim_{m\to\infty}\limsup_{T\to\infty}\Pr\left(\left|\sqrt{T}\sum_{k=m+1}^{T-1}\frac{1}{k}\,\widehat{r}(k)\right| > \varepsilon\right) = 0 \tag{41}
$$

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

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

For $A_{T,m}$,

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

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 $\varepsilon_j((d_0,\widehat{\boldsymbol{\xi}}')',\bar{x}) = \varepsilon_j((d_0,\widehat{\boldsymbol{\xi}}')',\mu) - (\bar{x}-\mu)\sum_{k=0}^{j-1}\pi_k((d_0,\widehat{\boldsymbol{\xi}}')')$,

$$
\begin{aligned}
&\varepsilon_j((d_0,\widehat{\boldsymbol{\xi}}')',\mu) = \varepsilon_j + \left(-c/\sqrt{T},\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)'\right)\sum_{k=0}^{j-1}\boldsymbol{\delta}_k\varepsilon_{j-k} + O_p\left(\frac{1}{T}\right),\\
&(\bar{x}-\mu)\sum_{k=0}^{j-1}\pi_k((d_0,\widehat{\boldsymbol{\xi}}')') = (\bar{x}-\mu)\sum_{k=0}^{j-1}\pi_k((d_0,\boldsymbol{\xi}')') + (\bar{x}-\mu)\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)'\sum_{k=0}^{j-1}\frac{\partial\pi_k(\boldsymbol{\delta}^*)}{\partial\boldsymbol{\xi}},
\end{aligned}
$$

where $\|\boldsymbol{\delta}^*-\boldsymbol{\delta}\| \le \|(-c/\sqrt{T},(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})')'-\boldsymbol{\delta}\|$ and $(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) = O_p(T^{-1/2})$, 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 $\varepsilon_t(\ddot{\boldsymbol{\vartheta}}) = \ddot{\vartheta}(L)^{-1}\{\widetilde{y}_t - \widetilde{\boldsymbol{\varphi}}_{t\cdot}\widehat{\boldsymbol{\beta}}\} = \ddot{\vartheta}(L)^{-1}\widetilde{x}_t(\boldsymbol{\alpha}) - \ddot{\vartheta}(L)^{-1}\widetilde{\boldsymbol{\varphi}}_{t\cdot}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \varepsilon_{t,1}(\ddot{\boldsymbol{\vartheta}}) + \varepsilon_{t,2}(\ddot{\boldsymbol{\vartheta}})$. To prove consistency, it is sufficient to check that

$$\sup_{\ddot{\boldsymbol{\vartheta}}} \frac{1}{T}\sum_{t=1}^{T}\left\{\varepsilon_{t,2}(\ddot{\boldsymbol{\vartheta}})\right\}^2 = O_p\left(\frac{1}{T}\right) \quad \text{and} \quad \mathrm{Var}\left[\boldsymbol{D}_T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] = O(1), \quad \text{as } T \to \infty, \tag{42}$$

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

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

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

$$\begin{aligned}\sum_{t=1}^{T}\left\{\varepsilon_{t,2}(\ddot{\boldsymbol{\vartheta}})\right\}^2 &= \sum_{t=1}^{T}\sum_{i,j=0}^{t-1}\ddot{\vartheta}_i\ddot{\vartheta}_j\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\boldsymbol{D}_T\boldsymbol{D}_T^{-1}\widetilde{\boldsymbol{\varphi}}_{t-i\cdot}'\widetilde{\boldsymbol{\varphi}}_{t-j\cdot}\boldsymbol{D}_T^{-1}\boldsymbol{D}_T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\\ &\le \left(\sum_{i=0}^{T}|\ddot{\vartheta}_i|\right)^2 \sup_{0\le i,j\le T}\sum_{t=1}^{T}\left|\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\boldsymbol{D}_T\boldsymbol{D}_T^{-1}\widetilde{\boldsymbol{\varphi}}_{t-i\cdot}'\widetilde{\boldsymbol{\varphi}}_{t-j\cdot}\boldsymbol{D}_T^{-1}\boldsymbol{D}_T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right| = O_p(1)\end{aligned} \tag{43}$$

because, by Assumption 2 and $\widetilde{\boldsymbol{\varphi}}_{t\cdot} = \boldsymbol{0}$ for $t \le 0$, $\sum_{i=0}^{\infty}|\ddot{\vartheta}_i|$ is uniformly convergent, each element of the $r \times r$ matrix $\sum_{t=1}^{T}\boldsymbol{D}_T^{-1}\widetilde{\boldsymbol{\varphi}}_{t-i\cdot}'\widetilde{\boldsymbol{\varphi}}_{t-j\cdot}\boldsymbol{D}_T^{-1}$ is less than one in absolute value for any $0 \le i, j \le T$, and $\mathrm{Var}[\boldsymbol{D}_T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})] = O(1)$, which establishes (42).

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

$$\frac{1}{\sqrt{T}}\frac{\partial S(\boldsymbol{0}, \widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}} = 0 = \frac{1}{\sqrt{T}}\frac{\partial S(\boldsymbol{\alpha}, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} - \frac{\partial^2 S(\boldsymbol{\alpha}^*, \boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\alpha}'}\frac{\boldsymbol{c}}{T} + \frac{1}{T}\frac{\partial^2 S(\boldsymbol{\alpha}^*, \boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right), \tag{44}$$

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

$$\begin{aligned}\varepsilon_{t,2}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}) &= \varepsilon_{t,2}(\boldsymbol{\vartheta}) + \frac{\boldsymbol{c}'}{\sqrt{T}}\begin{pmatrix}\log(1-L)\\ \log(1-L^s)\end{pmatrix}\varepsilon_{t,2}(\boldsymbol{\alpha}^*, \boldsymbol{\vartheta}),\\ \frac{\partial \varepsilon_{t,2}(\boldsymbol{\alpha}, \boldsymbol{\vartheta})}{\partial \boldsymbol{\alpha}} &= \begin{pmatrix}\log(1-L)\\ \log(1-L^s)\end{pmatrix}\varepsilon_{t,2}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}),\\ \frac{\partial \varepsilon_{t,2}(\boldsymbol{\alpha}, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} &= (1-L)^{\alpha_0}(1-L^s)^{\alpha_s}\frac{\partial \varepsilon_{t,2}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \frac{\partial \varepsilon_{t,2}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} + \frac{\boldsymbol{c}'}{\sqrt{T}}\begin{pmatrix}\log(1-L)\\ \log(1-L^s)\end{pmatrix}\frac{\partial \varepsilon_{t,2}(\boldsymbol{\alpha}^*, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}},\\ \sum_{t=1}^{T}\left\{\sum_{j=0}^{t-1}a_j\varepsilon_{t-j,2}(\ddot{\boldsymbol{\vartheta}})\right\}^2 &= O_p\left(\left\{\sum_{j=0}^{T}|a_j|\right\}^2\right) = O_p\left((\log T)^{2k}T^{2a}\right) = o_p(T),\end{aligned} \tag{45}$$

where the last equation follows from (43). It follows from (43) and (45) that

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

as $T \to \infty$. Using (42) and (45), we find that, as $T \to \infty$, each term of the RHS of (44) divided by T is unaffected by $\varepsilon_{t,2}(\boldsymbol{\alpha},\ddot{\boldsymbol{\vartheta}})$ in probability uniformly in $\ddot{\boldsymbol{\vartheta}} \in D_\vartheta$. The rest of the proof of asymptotic distributions of $\sqrt{T}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})$ is obvious from the proof of Theorem 2. Hence, we omit the proof.

Finally, we will prove (41) to derive asymptotic distribution of $\boldsymbol{S}_T$. Since $\varepsilon_j(\widehat{\boldsymbol{\vartheta}}) = \varepsilon_{j,1}(\widehat{\boldsymbol{\vartheta}}) + \varepsilon_{j,2}(\widehat{\boldsymbol{\vartheta}})$, $\varepsilon_{j,1}(\widehat{\boldsymbol{\vartheta}}) = \varepsilon_j + (-\boldsymbol{c}'/\sqrt{T}, (\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})')\sum_{k=0}^{j-1}\boldsymbol{\delta}_k\varepsilon_{j-k} + O_p(1/T)$, and $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} = O_p(T^{-1/2})$, it can be treated similarly to the proof of Theorem 2 by Lemma C 1 and (42). □

Acknowledgements

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.

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.

Ansley, C.F. and Newbold, P. (1979), On the finite sample distribution of residual autocorrelations in autoregressive-moving average models. *Biometrika* 66, pp.547–553.

Arteche, J. and Robinson, P. M. (2000), Semiparametric inference in seasonal and cyclical long memory process. *Journal of Time Series Analysis* 21, pp.1–25.

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.

Dunkl, C. F. and Xu, Y. (2001), *Orthogonal Polynomials of Several Variables.* Cambridge, 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.

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.

Hardy, G. H. (1991), *Divergent Series,* 2nd ed. Chelsea, New York.

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. (2004), *Essays on seasonally and fractionally differenced time series.* Doctoral dissertation, 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.

McLeod, A. I. (1978), On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of Royal Statistical Society* B 40, pp.296–302.

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.

Zygmund, A. (1968), *Trigonometric Series.* Vol. I. Cambridge, London.

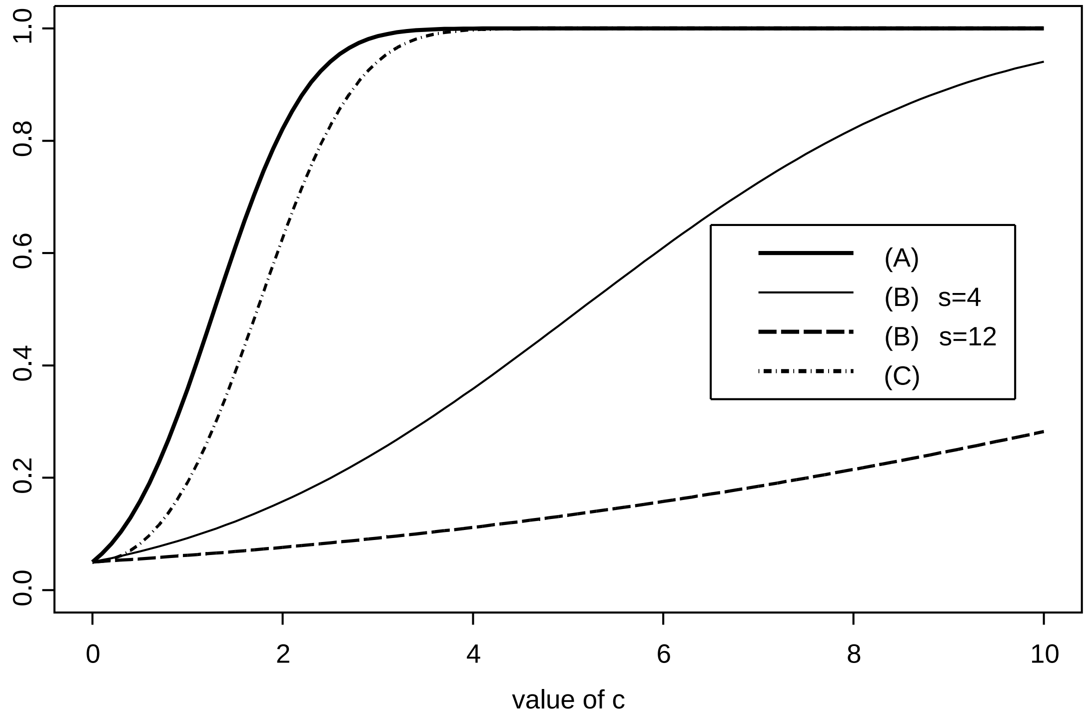

Figure 1: RHS of (A) through (C) in Corollary 1 changing s and c with $a = 0.95$.

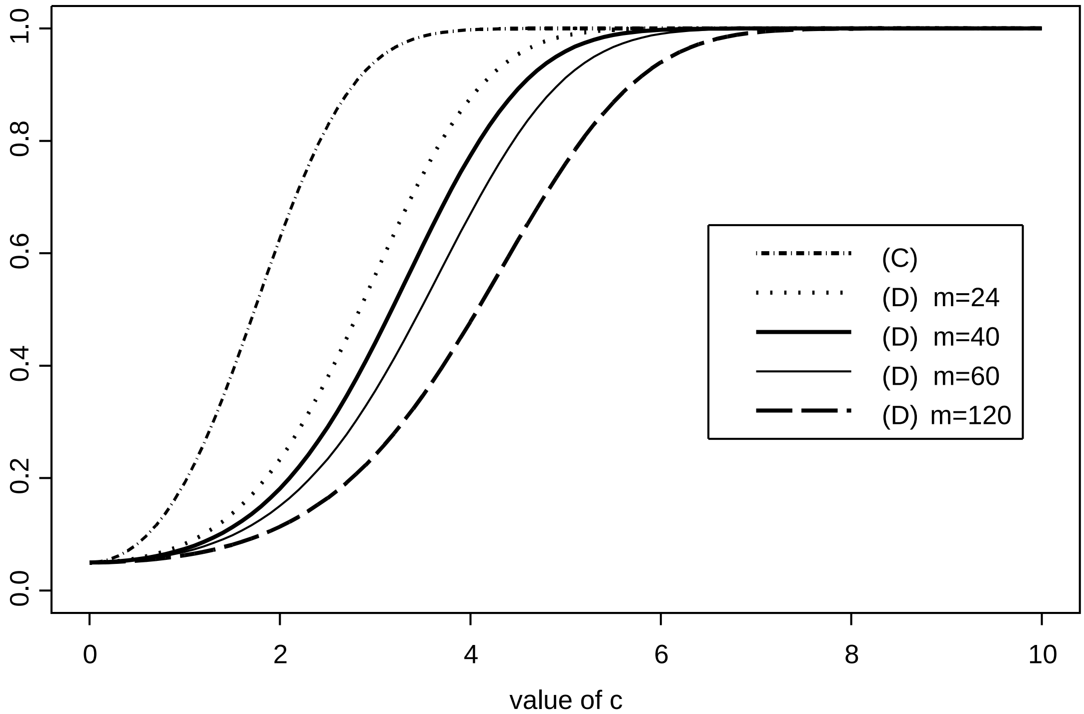

Figure 2: RHS of (C) and (D) in Corollary 1 changing m and c with $a = 0.95$.

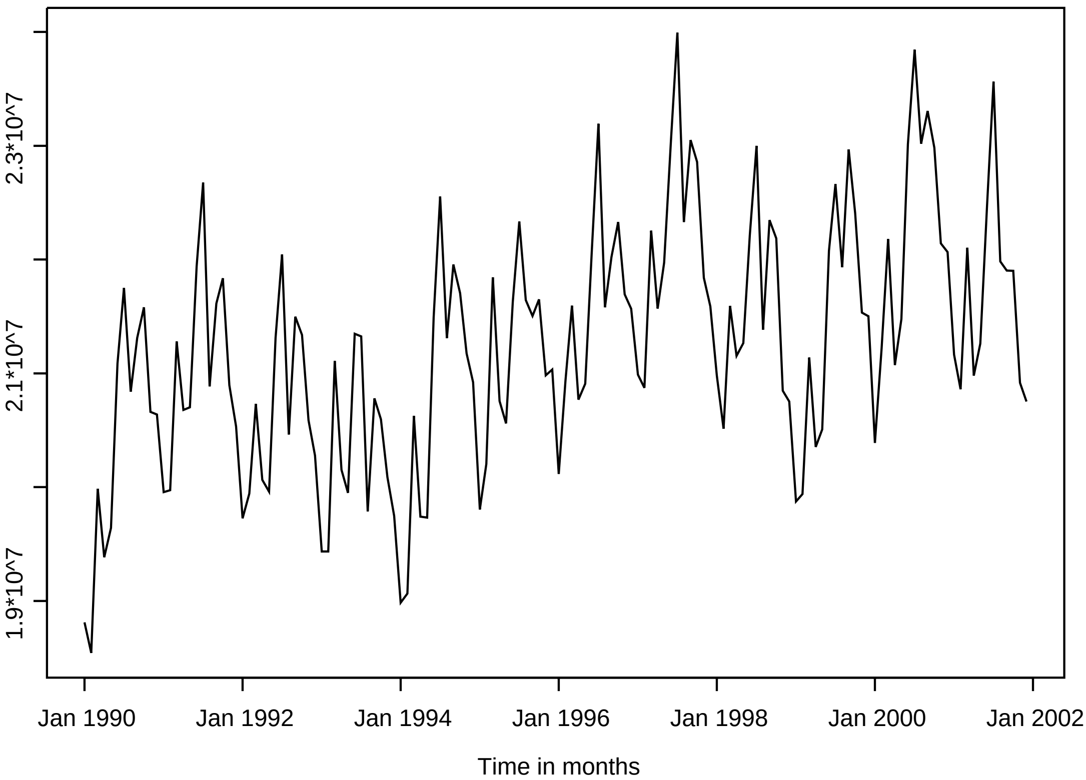

Figure 3: Japanese total power consumption data $\{x_t\}$, January 1990 to December 2001 (sum of the ten electric power companies, unit: MWh, sample size: 144).

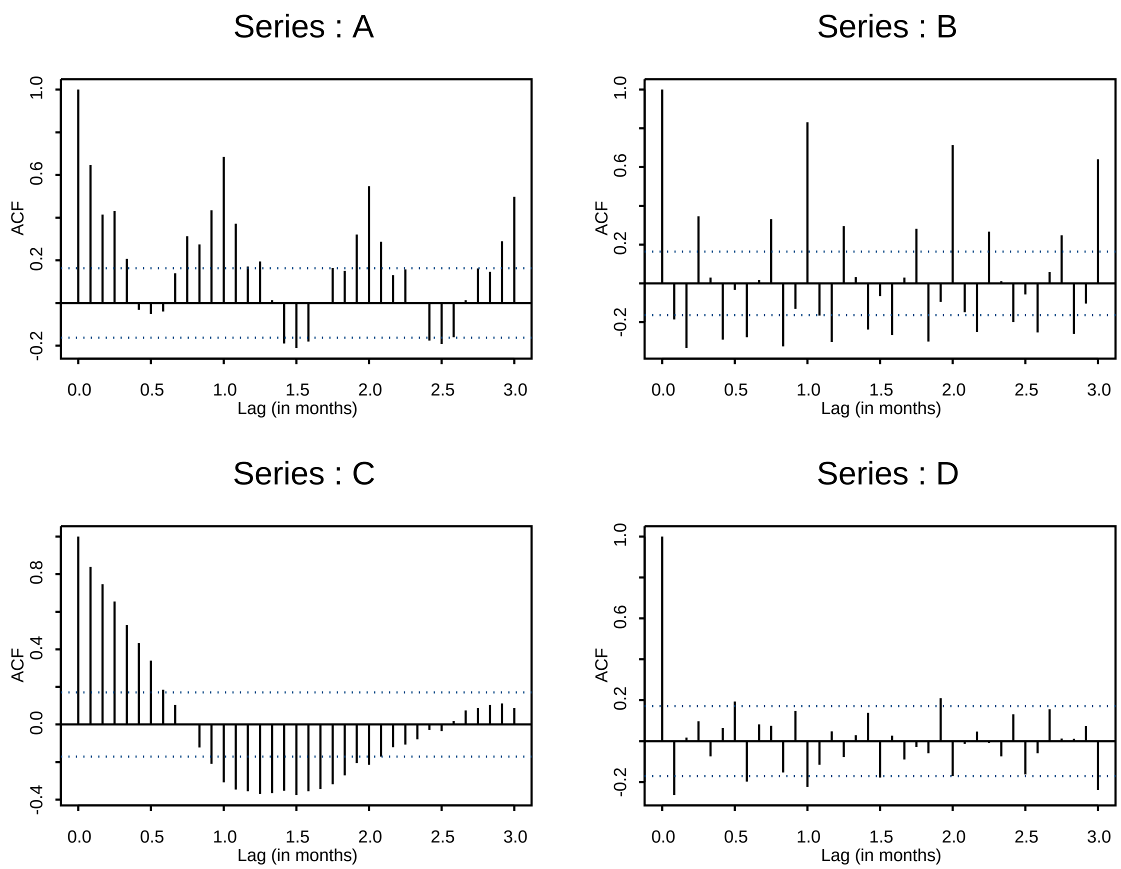

Figure 4: The sample autocorrelation function (ACF) of the transformed series, where A is $\{x_t\}$, B is $\{(1-L)x_t\}$, C is $\{(1-L^{12})x_t\}$, and D is $\{(1-L)(1-L^{12})x_t\}$. Dotted lines are approximate 95 % confidence limits of the ACF of the white noise random variable.

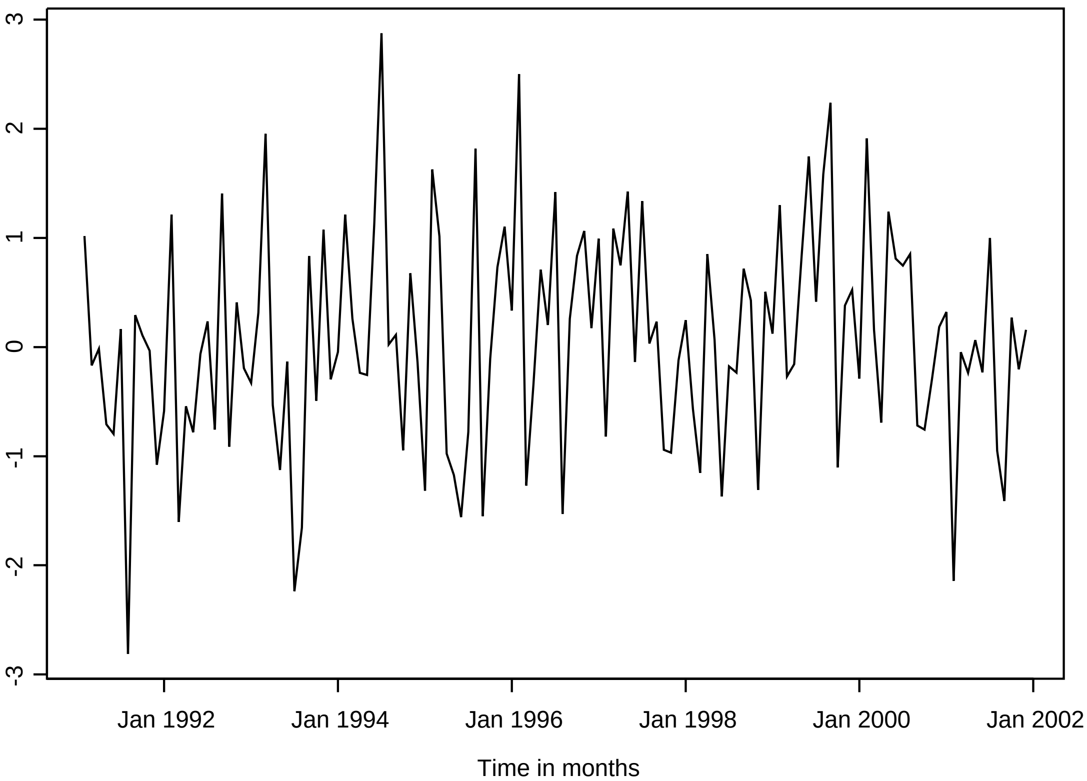

Figure 5: Standardized residuals from the SARFIMA$(1, 1, 0)(3, 1 + d_s, 0)_s$ model (model ID: 518) based on Japanese total power consumption data.

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

Table 2: Simulation on the estimation of $\text{SARFIMA}(0, d_0, 0)(1, d_s, 0)_s$ 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: $(\mu, \sigma^2, s) = (1.00, 1.00, 12)$

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)

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	0.20
			DGP 3					DGP 5		
$W_{T,0}$	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)
			DGP 4					DGP 6		
$W_{T,S}$	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)

Table 5: Summary of BIC and AIC model selection and estimates

ID	BIC	AIC	d_0	d_s	ϕ_1	θ_1	Φ_1	Φ_2	Φ_3	Θ_1	σ^2 ($\times 10^7$)
518	(1) 3573.0	(1) 3555.7	NE	-0.384	-0.221	NE	-0.162	-0.292	-0.393	NE	8914
521	(2) 3573.3	(2) 3556.0	NE	-0.376	NE	-0.208	-0.174	-0.297	-0.389	NE	8935
148	(3) 3574.1	(3) 3556.9	NE	NE	-0.211	NE	0.019	-0.282	-0.371	-0.571	8993
151	(4) 3574.4	(5) 3557.1	NE	NE	NE	-0.200	0.007	-0.286	-0.369	-0.563	9008

$\bar{y} = -18770.34$

Table 6: P-values of testing for $\alpha_0 = \alpha_s = 0$ of the SARFIMA models

	Alternative hypotheses		
Model	$\alpha_0 > 0$, $\alpha_s = 0$	$\alpha_0 = 0$, $\alpha_s < 0$	$\alpha_0 \neq 0$, $\alpha_s \neq 0$
SARFIMA$(1, \alpha_0, 0)(3, \alpha_s, 0)_s$	0.4016	0.0084	0.0062
SARFIMA$(0, \alpha_0, 1)(3, \alpha_s, 0)_s$	0.3207	0.0089	0.0041
SARFIMA$(1, \alpha_0, 0)(3, \alpha_s, 1)_s$	0.2838	0.2876	0.2052
SARFIMA$(0, \alpha_0, 1)(3, \alpha_s, 1)_s$	0.2287	0.2866	0.1673