Discussion Paper Series

No.10

Asymptotic Prediction Mean Squared Error for Strongly Dependent Processes with Estimated Parameters

Naoya Katayama

January 2004
revised September 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/

ASYMPTOTIC PREDICTION MEAN SQUARED ERROR FOR STRONGLY DEPENDENT PROCESSES WITH ESTIMATED PARAMETERS*

NAOYA KATAYAMA†

Department of Economics
Hitotsubashi University

Key Words and Phrases : Prediction mean-squared errors (PMSE); long memory; Akaike information criterion (AIC); model selection; convergence of moments.

JEL classifications : C22, C53

* This paper is based on a portion of Chapter 3 of the author's Ph.D. thesis. The author thanks Professor Katsuto Tanaka and Professor Taku Yamamoto 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

In this paper we deal with the prediction theory of long memory processes. After investigating the general theory relating to convergence of moments of the nonlinear least squares estimators, we evaluate the asymptotic prediction mean squared error of two predictors. One is defined by using the estimator of the differencing parameter and the other is defined by using a fixed, known differencing parameter, which is, in other words, one parametric predictor of the seasonally integrated autoregressive moving average (SARIMA) models. In this paper, results do not impose the normality assumption and deal not only with stationary time series but also with nonstationary ones. The finite sample behavior is examined by simulations using the computer program S-PLUS in terms of the asymptotic theory.

1 Introduction

This paper considers prediction theory for long-memory processes. For the stationary process, $\{y_t\}$, when the system of autocovariances is known, it is well known that the mean squared error of the best linear predictor (BLP) of y_{n+h} based on a finite past $\{y_t\}_{t=1}^n$ is easily obtained and, as $n \to \infty$, it decreases monotonically and converges to that of the BLP based on an infinite past $\{y_t\}_{t=1}^{\infty}$. Because autocovariances are usually unknown, we have to use estimated linear predictors. International Journal of Forecasting (volume 18, issue 2) provides a good survey of this issue. However, few papers address the theoretical optimality of the practical predictors. One is Hidalgo and Yajima (2002), which introduces a semiparametric predictor and demonstrates consistency on the basis of the mean squared error of the BLP based on an infinite past.

The purpose of this paper is to discuss the asymptotic prediction mean squared error, denoted PMSE, of an alternative linear predictor for the long-memory processes with estimated coefficients in parametric models.

First, independently to the other sections, Section 2 discusses the convergence of moments of the estimators. Some authors have considered convergence of moments of the linear least squares estimators to evaluate the PMSE, for example, Fuller and Hasza (1981), Bhansali and Papangelou (1991), Papangelou (1994), and Ing (2001). Independently, under an iid assumption, the asymptotic expansion of the moments of the normalized maximum likelihood estimator has been studied by, for example, McCullagh (1987, Section 7.3). However, none so far have dealt with the case of the non-linear least squares (NLS) estimators with standard properties. We prove the convergence of moments of the scaled estimators without assuming normality. Corollary 2.1 gives sufficient conditions of moments of the estimators of long-memory processes, which are used demonstrating asymptotic PMSE in later sections.

Section 3 considers a scalar process:

$$\phi(L)(1-L^s)^d y_t = \theta(L)\varepsilon_t, \tag{1.1}$$

where $\{\varepsilon_t\}$ is iid $(0, \sigma_0^2)$ and $\mathrm{E}[\varepsilon_t]^r < \infty$ for all positive integers r, s is known and unity or an even integer, $\phi(z) = 1 - \sum_{i=1}^p \phi_i z^i$, $\theta(z) = 1 + \sum_{i=1}^q \theta_i z^i$, and $\phi(z) = 0$ and $\theta(z) = 0$ have no roots in common and all roots are outside the unit circle. In this case, the autoregressive moving average, ARMA(p,q), model $\beta(L)\varepsilon_t \equiv [\theta(L)/\phi(L)]\varepsilon_t$ is stationary and invertible. Hosking (1981) and Hassler (1994) introduced this model and showed that when $d < 1/2$, $\{y_t\}$ is stationary and when $d > -1/2$, $\{y_t\}$ is invertible and demonstrates properties of long or intermediate memory processes. The model is known as the autoregressive fractionally integrated seasonal moving average, or ARFISMA(p,d,q), model when s is an even integer, and as an autoregressive fractionally integrated moving average, ARFIMA(p,d,q), model when $s=1$. However, our main concern is the following non-stationary ARFISMA$(p, d+m, q)$ process:

$$(1-L^s)^m x_t = y_t, \qquad t \geq 1, \tag{1.2}$$

where m is zero or a positive integer. Imposing the restrictions $\{x_{t-ms} = y_t = 0, t \leq 0\}$ and $d \in (-1/2, 1/2)$, we assume that we observe $x_{1-ms}, x_{2-ms}, \ldots, x_n$ and the differenced series $y_1, \ldots, y_n$. Section 3 adopts a NLS method called the conditional sum of squares (CSS) method and proves convergence of moments of the scaled estimators by using Corollary 2.1. Chung and Baillie (1996) considers the small sample properties of CSS estimators of ARFIMA models and provides a survey of this method. Section 3 also introduces a linear predictor with estimated coefficients and proves consistency on the basis of the mean squared error of the BLP for the differenced series $\{y_t\}$. Section 4 extends this result to the non-stationary linear predictor for $\{x_t\}$ and considers the effects of the misspecification of a predictor of the seasonal autoregressive integrated moving average (SARIMA) model with differencing parameter m, when the true model is the ARFISMA model in (1.2). The effects of misspecification of (non)stationary ARMA models are well documented, see for example, Bhansali (1981), Fuller and Hasza (1981), Tanaka and Maekawa (1984), and Kunitomo and Yamamoto (1985). We focus on the misspecification of the differencing parameter d in models (1.1) and (1.2).

Section 5 examines these finite sample properties of the PMSE of predictors. It also reports the rate to select an appropriate predictor with Wald test statistics and Akaike Information Criterion (AIC).

Throughout this paper, let L be the lag operator, $\|\boldsymbol{x}\| = (\boldsymbol{x}'\boldsymbol{x})^{1/2}$ be the Euclidian norm of $\boldsymbol{x}$, $\|\boldsymbol{A}\|_S = \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|$ be the matrix norm of $\boldsymbol{A}$ called the spectral norm of $\boldsymbol{A}$, $\partial f(\boldsymbol{x})/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{y}} = \partial f(\boldsymbol{y})/\partial \boldsymbol{x}$, $f^{(0)}(\boldsymbol{y}) = f(\boldsymbol{y})$, $f^{(1)}(\boldsymbol{y}) = \partial f(\boldsymbol{y})/\partial \boldsymbol{x}$, and $f^{(2)}(\boldsymbol{y}) = \partial^2 f(\boldsymbol{y})/\partial \boldsymbol{x}\partial \boldsymbol{x}'$. In addition, 'RHS' abbreviates 'right-hand side', 'LHS' abbreviates 'left-hand side', and const is used to denote universal appropriate positive constants to economize on notation. All proofs are given in the Appendix.

2 Convergence of moments of the non-linear least squares estimators

In this section we consider the convergence of moments of NLS estimators for some time series models. After assuming the well-known sufficient conditions of strong consistency and asymptotic normality, we provide a set of sufficient conditions for the convergence of moments of NLS estimators. Corollary 2.1 gives simple sufficient conditions to prove the convergence of moments of NLS estimators for a truncated time series model, which is applicable to the long-memory processes in Section 3.

2.1 Case I: stationary processes

We consider the following scalar stationary process, $\{y_t\}$, defined by:

$$y_t = \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}^0) y_{t-j} + \varepsilon_t = \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta}^0)\varepsilon_{t-j} + \varepsilon_t, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{2.1}$$

where $\{\varepsilon_t\}_{t=1}^{\infty}$ is iid $(0, \sigma_0^2)$ and $\mathrm{E}[\varepsilon_t]^r < \infty$ for all positive integers r, $\boldsymbol{\theta}^0$ is a vector of true parameters contained by the compact and convex parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^p$, $\sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})^2$ and $\sum_{j=1}^{\infty} b_j(\boldsymbol{\theta})^2$ are finite for any $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)' \in \boldsymbol{\Theta}$, and, for simplicity, σ_0^2 is assumed to be known.

Let $e_1(\boldsymbol{\theta}) = y_1$, $e_t(\boldsymbol{\theta}) = y_t - \sum_{j=1}^{t-1} a_j(\boldsymbol{\theta}) y_{t-j}$ $(t \geq 2)$, and $\widehat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ be the estimator of $\boldsymbol{\theta}^0$ that minimizes the objective function:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{2n\sigma_0^2} \sum_{t=1}^{n} e_t^2(\boldsymbol{\theta}), \tag{2.2}$$

which is almost certainly twice continuously differentiable on $\boldsymbol{\Theta}$, for $n = 1, 2, \ldots$.

First, we provide the sets of sufficient conditions for strong consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ [see, e.g., Gallant and White (1998, Section 3) and Gourieroux and Monfort (1989, Chapter 24)]. Suppose there exists a function $Q_\infty(\boldsymbol{\theta})$:

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| \xrightarrow{a.c.} \boldsymbol{0} \tag{2.3a}$$

$$\text{The minimum } \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_\infty(\boldsymbol{\theta}) \text{ is attained at a unique value of } \boldsymbol{\theta}^0. \tag{2.3b}$$

These conditions are sufficient to ensure strong consistency. Since Q_n is differentiable with respect to $\boldsymbol{\theta}$, we can expand the partial derivative of $Q_n(\widehat{\boldsymbol{\theta}}_n)$ with respect to $\boldsymbol{\theta}$ in a Taylor series about $\boldsymbol{\theta}^0$, for large n by, for example, Gourieroux and Monfort (1989, Property 24.9):

$$\sqrt{n}\, Q_n^{(1)}(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0} = \sqrt{n}\, Q_n^{(1)}(\boldsymbol{\theta}^0) + Q_n^{(2)}(\boldsymbol{\theta}^*)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0), \tag{2.4}$$

where $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\| \leq \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0\|$. On the Hessian matrix of the above equation, we assume that there exists a $p \times p$ fixed matrix function $\boldsymbol{I}(\boldsymbol{\theta})$ such that $\boldsymbol{I}(\boldsymbol{\theta})$ is continuous on $\boldsymbol{\Theta}$, $\boldsymbol{I}(\boldsymbol{\theta}^0)$ is a symmetric, positive definite matrix, and:

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| Q_n^{(2)}(\boldsymbol{\theta}) - \boldsymbol{I}(\boldsymbol{\theta}) \right| \xrightarrow{a.c.} \boldsymbol{0}. \tag{2.5}$$

Putting $Q_n^{(2)}(\boldsymbol{\theta}^*) = \boldsymbol{I}_n(\boldsymbol{\theta}^0) + \boldsymbol{R}_n(\boldsymbol{\theta}^*)$:

$$\boldsymbol{I}_n(\boldsymbol{\theta}^0) = \frac{1}{n\sigma_0^2}\sum_{t=1}^{n} e_t^{(1)}(\boldsymbol{\theta}^0)e_t^{(1)}(\boldsymbol{\theta}^0)',$$
$$\boldsymbol{R}_n(\boldsymbol{\theta}^*) = \frac{1}{n\sigma_0^2}\sum_{t=1}^{n} e_t(\boldsymbol{\theta}^0)e_t^{(2)}(\boldsymbol{\theta}^0) + \Big\{Q_n^{(2)}(\boldsymbol{\theta}^*) - Q_n^{(2)}(\boldsymbol{\theta}^0)\Big\},$$

we further assume that:

$$\boldsymbol{I}_n(\boldsymbol{\theta}^0) \xrightarrow{a.c.} \boldsymbol{I}(\boldsymbol{\theta}^0), \tag{2.6}$$
$$\text{and} \quad \sqrt{n}\, Q_n^{(1)}(\boldsymbol{\theta}^0) \xrightarrow{d} N\big(\mathbf{0},\ \boldsymbol{I}(\boldsymbol{\theta}^0)\big), \tag{2.7}$$

as $n \to \infty$. Conditions of (2.6) and (2.7) are based on the fact that series $e_t(\boldsymbol{\theta}^0)$ is obtained by truncating the infinite autoregressive representations for y_t. See, for example, Theorem 5.5.1 and Theorem 8.4.1 in Fuller (1996). Then, $Q_n^{(2)}(\boldsymbol{\theta}^*) \xrightarrow{a.c.} \boldsymbol{I}(\boldsymbol{\theta}^0)$ and $\boldsymbol{R}_n(\boldsymbol{\theta}^*) \xrightarrow{a.c.} \mathbf{0}$ from $\widehat{\boldsymbol{\theta}}_n \xrightarrow{a.c.} \boldsymbol{\theta}^0$, $\boldsymbol{I}(\boldsymbol{\theta})$ is continuous, (2.5), (2.6),

$$\begin{aligned}
\Big|Q_n^{(2)}(\boldsymbol{\theta}^*) - \boldsymbol{I}(\boldsymbol{\theta}^0)\Big| &\le \Big|Q_n^{(2)}(\boldsymbol{\theta}^*) - \boldsymbol{I}(\boldsymbol{\theta}^*)\Big| + \big|\boldsymbol{I}(\boldsymbol{\theta}^*) - \boldsymbol{I}(\boldsymbol{\theta}^0)\big| \\
&\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big|Q_n^{(2)}(\boldsymbol{\theta}) - \boldsymbol{I}(\boldsymbol{\theta})\Big| + \big|\boldsymbol{I}(\boldsymbol{\theta}^*) - \boldsymbol{I}(\boldsymbol{\theta}^0)\big| \xrightarrow{a.c.} \mathbf{0}, \\
\text{and} \qquad |\boldsymbol{R}_n(\boldsymbol{\theta}^*)| &= \Big|Q_n^{(2)}(\boldsymbol{\theta}^*) - \boldsymbol{I}_n(\boldsymbol{\theta}^0)\Big| \\
&\le \Big|Q_n^{(2)}(\boldsymbol{\theta}^*) - \boldsymbol{I}(\boldsymbol{\theta}^0)\Big| + \big|\boldsymbol{I}(\boldsymbol{\theta}^0) - \boldsymbol{I}_n(\boldsymbol{\theta}^0)\big| \xrightarrow{a.c.} \mathbf{0}.
\end{aligned}$$

It follows that, as $n \to \infty$:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) = -\Big\{Q_n^{(2)}(\boldsymbol{\theta}^*)\Big\}^{-1}\sqrt{n}\, Q_n^{(1)}(\boldsymbol{\theta}^0) \xrightarrow{d} N\big(\mathbf{0},\ \boldsymbol{I}(\boldsymbol{\theta}^0)^{-1}\big).$$

These are well-known results for NLS estimators. However, the aim of this section is not to derive strong consistency and asymptotic normality of the estimators. Instead, our aim is to derive the moments of the (scaled) estimator converge, as $n \to \infty$, to the moments of the asymptotic distribution. Typically:

$$n\,\mathrm{E}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0\big)\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0\big)' \longrightarrow \boldsymbol{I}(\boldsymbol{\theta}^0)^{-1}, \qquad \text{as } n \to \infty. \tag{2.8}$$

To proceed, put:

$$\mathbf{a}_i(m) = \frac{\partial}{\partial\theta_i}\big(a_1(\boldsymbol{\theta}), a_2(\boldsymbol{\theta}), \ldots, a_m(\boldsymbol{\theta})\big)'\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0},$$

for $i = 1, \ldots, p$. We assume the following.

Assumption 1 For the process, $\{y_t\}$, in (2.1), we assume the following.

(a) The assumptions of (2.3a), (2.3b), (2.5), (2.6), and (2.7) hold.

(b) $\mathbf{a}_1(m), \mathbf{a}_2(m), \ldots, \mathbf{a}_p(m)$ are linearly independent for some $m \ge p$.

(c) $\|e_t^{(1)}(\boldsymbol{\theta}^0)\|$ has finite moments of all orders.

(d) For any finite set $\{j, l_1, l_2, \ldots l_k\}$ of distinct integers, the joint distributions of $y_j, y_{l_1}, y_{l_2}, \ldots, y_{l_k}$ are absolutely continuous and there exists a constant $K > 0$ and a conditional probability density function y_j given $y_{l_1} = y_1^*, y_{l_2} = y_2^*, \ldots, y_{l_k} = y_k^*$, $f_{jl_1l_2\ldots l_k}(\,\cdot\,|y_1^*, y_2^*, \ldots, y_k^*)$, such that

$$f_{jl_1l_2\ldots l_k}(\,y\,|\,y_1^*, y_2^*, \ldots, y_k^*\,) \le K$$

for all $y, y_1^*, y_2^*, \ldots, y_k^*$.

The conditions of (2.3a) and (2.3b) can be replaced by the strong consistency of $\widehat{\boldsymbol{\theta}}_n$. Namely, $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0 \xrightarrow{a.c.} \mathbf{0}$. The condition of (b) is only used in deriving (A.2) in the proof of Theorem 2.1 in Appendix A. For example, when the process $\{y_t\}$ is a pth order autoregressive, AR(p), process, $y_t = \sum_{i=1}^{p} \theta_i y_{t-i} + \varepsilon_t$, $(a_1(\boldsymbol{\theta}), \ldots, a_p(\boldsymbol{\theta})) = (\theta_1, \ldots, \theta_p)$ and $(\mathbf{a}_1(p), \ldots, \mathbf{a}_p(p))$ is a $p \times p$ identity matrix. The condition of (c) is also only used in deriving (A.4) in the proof of Theorem 2.1, Appendix A. The condition of (d) is due to (III) of Bhansali and Papangelou (1991, p.1157). [Due to a printer's error, an incongruous Kt appears in place of the correct bound K in (III). See Papangelou (1994, p.403).]

Then we obtain the following theorem.

Theorem 2.1 *For the model* (2.1)*, let $\widehat{\boldsymbol{\theta}}_n$ be the estimator defined by* (2.2) *and satisfy Assumption 1. Then there exists a number $n_0 > 0$, $r_0 > r \geq 2$, $u, v > 1$, $1/u + 1/v = 1$, and for all $n \geq n_0$:*

$$\mathrm{E} \left\| \left\{ Q_n^{(2)}(\boldsymbol{\theta}^*) \right\}^{-1} \right\|_S^q < \infty, \tag{2.9}$$

for some $q \geq r_0 u$. Furthermore, if:

$$\mathrm{E} \left\| \sqrt{n}\, Q_n^{(1)}(\boldsymbol{\theta}^0) \right\|^{q'} < \infty, \tag{2.10}$$

for some $q' \geq r_0 v$ and all $n \geq n_0$, then:

$$\lim_{n\to\infty} \mathrm{E} \left\| \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \right\|^r = \mathrm{E} \left\| N\big(\mathbf{0},\ \boldsymbol{I}(\boldsymbol{\theta}^0)^{-1}\big) \right\|^r, \tag{2.11}$$

and (2.8) *holds.*

The proof of Theorem 2.1 makes use of arguments by Bhansali and Papangelou (1991) and Ing (2001), which prove the convergence of moments of the normalized linear least squares estimators.

2.2 Case II: non-stationary processes

We consider the following truncated non-stationary process, $\{y_t\}$, defined by:

$$y_t = \sum_{j=1}^{t-1} a_j(\boldsymbol{\theta}^0) y_{t-j} + \varepsilon_t = \sum_{j=1}^{t-1} b_j(\boldsymbol{\theta}^0) \varepsilon_{t-j} + \varepsilon_t, \qquad t = 2, 3 \ldots, \tag{2.12}$$

$y_1 = \varepsilon_1$, and $y_t = 0$ for $t \leq 0$, where $\{\varepsilon_t\}$, $a_j(\boldsymbol{\theta}^0)$s, and $b_j(\boldsymbol{\theta}^0)$s are given by (2.1). Namely, this model is equivalent to the model (2.1) with assumption $\{y_t = \varepsilon_t = 0, t \leq 0\}$.

We derive convergence of moments of NLS estimators of $\boldsymbol{\theta}^0$ when we observe $\{y_t\}_{t=1}^n$ defined by (2.12). Note that we use notations of Section 2.1 to economize notations. Therefore, similarly to Section 2.1, let $e_1(\boldsymbol{\theta}) = y_1$, $e_t(\boldsymbol{\theta}) = y_t - \sum_{j=1}^{t-1} a_j(\boldsymbol{\theta}) y_{t-j}$ $(t \geq 2)$, and $\{y_t\}$ be given by (2.12). Let $\widehat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ be the NLS estimator of $\boldsymbol{\theta}^0$ that minimizes the objective function $Q_n(\boldsymbol{\theta}) = \sum_{t=1}^n e_t^2(\boldsymbol{\theta})/(2n\sigma_0^2)$.

In this model, by using the fact that $\sum_{j=1}^{t-1}\sum_{k=0}^{t-j-1} a_{k,j} = \sum_{j=1}^{t-1}\sum_{k=0}^{j-1} a_{k,j-k}$, we have:

$$e_t^{(1)}(\boldsymbol{\theta}^0) = -\sum_{j=1}^{t-1} a_j^{(1)}(\boldsymbol{\theta}^0) y_{t-j} = -\sum_{j=1}^{t-1} a_j^{(1)}(\boldsymbol{\theta}^0) \sum_{k=0}^{t-j-1} b_k(\boldsymbol{\theta}^0) \varepsilon_{t-j-k} = \sum_{j=1}^{t-1} \boldsymbol{d}_j(\boldsymbol{\theta}^0) \varepsilon_{t-j}, \tag{2.13}$$

for $t \geq 2$, where $\boldsymbol{d}_j(\boldsymbol{\theta}^0) = -\sum_{k=0}^{j-1} b_k(\boldsymbol{\theta}^0) a_{j-k}^{(1)}(\boldsymbol{\theta}^0)$ and $b_0(\boldsymbol{\theta}^0) = 1$. Therefore, if these $\{\boldsymbol{d}_j(\boldsymbol{\theta}^0)\}$ satisfy:

$$\boldsymbol{I}_n(\boldsymbol{\theta}^0) = \frac{1}{n\sigma_0^2} \sum_{t=2}^{n} \sum_{j=1}^{t-1} \sum_{k=1}^{t-1} \boldsymbol{d}_j(\boldsymbol{\theta}^0) \boldsymbol{d}_k(\boldsymbol{\theta}^0)' \varepsilon_{t-j} \varepsilon_{t-k} \xrightarrow{a.c.} \boldsymbol{I}(\boldsymbol{\theta}^0) = \lim_{t\to\infty} \sum_{j=1}^{t-1} \boldsymbol{d}_j(\boldsymbol{\theta}^0) \boldsymbol{d}_j(\boldsymbol{\theta}^0)', \tag{2.14}$$

in (2.6), then we can simplify sufficient conditions of moments of the NLS estimators given by Theorem 2.1.

Assumption 2 For the process, $\{y_t\}$, in (2.12), we assume the following.

(a) The assumptions of (2.3a), (2.3b), (2.5), and (2.6) hold.

(b) There exist p-vector sequences $\{\boldsymbol{d}_j(\boldsymbol{\theta}^0)\}$ given by (2.13) and satisfy (2.14).

(c) ε_1 has a probability density function $f(x)$ such that $f(x) \leq K$ for any $x \in \mathbb{R}$ and some positive constant K.

The conditions of (2.3a) and (2.3b) can be replaced by the strong consistency of $\widehat{\boldsymbol{\theta}}_n$ as noted in Section 2.1. Since $\{a_j(\boldsymbol{\theta})\}$ is a sequence of non-linear functions of $\boldsymbol{\theta}$, it is often difficult to treat $\{\mathbf{a}_i(m)\}$ in (b) of Assumption 1. (b) of Assumption 2 simplifies this condition (see the proof of Corollary 2.1 in Appendix A). This condition not only ensures (c) of Assumption 1 and (2.10), but also satisfies conditions of a central limit theorem for martingale differences because, under the model (2.12), $e_t(\boldsymbol{\theta}^0) = \varepsilon_t$ for $t \geq 1$ and $\{e_t(\boldsymbol{\theta}^0)e_t^{(1)}(\boldsymbol{\theta}^0)\}$ is a sequence of martingale differences. See, for example, Fuller (1996, Theorem 5.3.4 and Theorem 5.5.1). Therefore, (2.7) in Assumption 1 also holds. (d) of Assumption 1 is replaced by (c) of Assumption 2 from Lemma 2.1 below.

Corollary 2.1 *Let the model* (2.12) *hold, and let the conditions of Assumption 2 hold. Then* (2.8), (2.9), *and* (2.11) *hold.*

We establish the following lemma to prove Theorem 2.1 and Corollary 2.1.

Lemma 2.1 *(i) Let $\{X_t\}$ be a strict stationary processes, and let for any finite set $\{j, l_1, l_2, \ldots l_k\}$ of distinct integers, the joint distributions of $X_j, X_{l_1}, X_{l_2}, \ldots, X_{l_k}$ be absolutely continuous and there exist a constant $K > 0$ and a conditional probability density function X_j given $X_{l_1} = x_1, X_{l_2} = x_2, \ldots, X_{l_k} = x_k$, $f_{jl_1l_2\ldots l_k}(\cdot|x_1, x_2, \ldots, x_k)$, such that:*

$$f_{jl_1l_2\ldots l_k}(\,x\,|\,x_1, x_2, \ldots, x_k\,) \leq K, \tag{2.15}$$

for all $x, x_1, x_2, \ldots, x_k$. Put $Y_i = \sum_{j=i}^{n} a_{ij}X_j$ $(i = 1, 2, \ldots, n)$, where $a_{ii} \neq 0$ for $i = 1, \ldots, n$. Then:

$$\begin{aligned}&\Pr(Y_1 \in C_1, Y_{l_1} \in C_{l_1}, Y_{l_2} \in C_{l_2}, \ldots, Y_{l_k} \in C_{l_k})\\ &\qquad \leq K|a_{11}|^{-1}(c_1' - c_1)\Pr(Y_{l_1} \in C_{l_1}, Y_{l_2} \in C_{l_2}, \ldots, Y_{l_k} \in C_{l_k}),\end{aligned} \tag{2.16}$$

where $C_1 = (c_1, c_1')$, $-\infty < c_1 < c_1' < \infty$, $1 < l_1 < l_2 < \cdots < l_k < n$, and $C_{l_1}, C_{l_2}, \ldots, C_{l_k}$ are appropriate intervals.

(ii) When $\{X_t\}$ is an iid process, the condition of (2.15) *can be replaced by:*

$$f_1(x) \leq K, \qquad (\text{for any } x \in \mathbb{R} \text{ and some } K > 0), \tag{2.17}$$

where $f_1(x)$ is the marginal probability density function of X_1.

3 PMSE for long-memory processes

In this section we explore the asymptotic PMSE for the process $\{y_t\}$ in (1.1). For a scalar stationary and invertible process $\{y_t\}$ defined by:

$$\sum_{j=0}^{\infty} \pi_j y_{t-j} = \varepsilon_t, \qquad \text{and} \qquad y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad t = 0, \pm 1, \ldots, \tag{3.1}$$

where $\sum_{j=0}^{\infty} \pi_j^2$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, and $\pi_0 = \psi_0 = 1$, it is well known that, in general, the BLP, denoted by $y_n(h)$, for a future value of y_{n+h} based on data from the infinite past, $y_n, y_{n-1}, \ldots$, is written as:

$$y_n(h) = -\sum_{j=1}^{\infty} \pi_j y_n(h-j) = \sum_{j=1}^{\infty} c_j(h) y_{n+1-j} = \sum_{j=h}^{\infty} \psi_j \varepsilon_{n+h-j}, \tag{3.2}$$

for $h \geq 1$ where $y_n(h-j) = y_{n+h-j}$ for $j \geq h$ and $c_j(h)$s are, given by equation (A5.2.3) in Box and Jenkins (1976):

$$c_j(h) = -\sum_{i=0}^{h-1} \psi_i \pi_{j+h-i-1}, \tag{3.3}$$

for $j \geq 1$. Its PMSE, $\sigma_y^2(h)$, is given by:

$$\sigma_y^2(h) \equiv \mathrm{E}\left[y_{n+h} - y_n(h)\right]^2 = \mathrm{E}\left[\sum_{j=0}^{h-1} \psi_j \varepsilon_{n+h-j}\right]^2 = \sigma_0^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{3.4}$$

However, for simplicity, we impose restrictions on $\{y_t = 0, t \leq 0\}$, or equivalently $\{\varepsilon_t = 0, t \leq 0\}$. In this case, the expression for the BLP $y_n(h)$ based on $\{y_t\}_{t=1}^n$ is simply obtained by imposing $\{y_t = 0, t \leq 0\}$ and $\{\varepsilon_t = 0, t \leq 0\}$ in (3.2) and the PMSE in (3.4) will not change.

Because π_js and ψ_js are usually unknown, Section 3.1 discusses estimated predictors based on data of sample size n for y_t defined by (1.1), whereas Section 3.2 discusses the effects of misspecification when the differencing parameter d is fixed. Throughout this section, we estimate parameters using the CSS method.

3.1 PMSE for ARFISMA processes

We first construct the estimated predictor of y_{n+h}, denoted by $\widehat{y}_n(h)$, when $\{y_t\}_{t=1}^n$ in (1.1) is given. Put $\beta(z) = \theta(z)/\phi(z) = \sum_{i=0}^{\infty} \beta_i z^i$. When $\beta(z) = 1$, under the initial condition, the model (1.1) can be expressed as:

$$y_t = \sum_{j=0}^{t-1} \psi_j(d) \varepsilon_{t-j}, \qquad t \geq 1, \tag{3.5}$$

and $y_t = 0$, $t \leq 0$, where $\psi_j(d) = \Gamma(j+d)/\{\Gamma(d)\Gamma(j+1)\}$, for $j = 0, s, 2s, \ldots = 0$, otherwise, which is the well-known jth MA(∞) coefficient of the ARFIMA$(0,d,0)$ process when $s = 1$. Alternatively, we can rewrite (3.5) as:

$$\sum_{j=0}^{t-1} \pi_j(d) y_{t-j} = \varepsilon_t, \qquad t \geq 1, \tag{3.6}$$

where $\pi_j(d) = \psi_j(-d)$. Similarly, for the ARFISMA(p,d,q) model in (1.1), it can be written as:

$$y_t = \sum_{j=0}^{t-1} \psi_j(\boldsymbol{\delta}^0) \varepsilon_{t-j}, \qquad \sum_{j=0}^{t-1} \pi_j(\boldsymbol{\delta}^0) y_{t-j} = \varepsilon_t, \qquad t \geq 1, \tag{3.7}$$

where $\psi_j(\boldsymbol{\delta}^0)$s are defined in terms of coefficients of β_is and $\psi_j(d)$s, $\pi_j(\boldsymbol{\delta}^0)$s are also similarly defined, $\boldsymbol{\delta}^0$ is the true parameter vector of (1.1) defined by $\boldsymbol{\delta}^0 = \big(d, \boldsymbol{\beta}'\big)'$, and $\boldsymbol{\beta}$ is a $(p+q)$-parameter vector consisting of ARMA(p,q) parameters, i.e., $\boldsymbol{\beta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$.

We now assume the following.

Assumption 3 For the process $\{y_t\}$ in (1.1),

(a) $\{\varepsilon_t\}$ is iid $(0, \sigma_0^2)$ and $\mathrm{E}[\varepsilon_t]^r < \infty$ for all positive integers r and ε_1 has a probability density function $f(x)$ such that $f(x) < K < \infty$ for any $x \in \mathbb{R}$ and some positive constant K.

(b) $\{y_t = 0,\ t \leq 0\}$ or equivalently $\{\varepsilon_t = 0,\ t \leq 0\}$.

(c) $d \in D_l$ for some $l = 1, 2, 3$, where $D_1 = [\alpha,\ 1/2 - \alpha]$, $D_2 = [\alpha - 1/4,\ 1/4 - \alpha]$, $D_3 = [\alpha - 1/2,\ -\alpha]$, and $\alpha \in (0, 1/4)$. D_β be a compact space such that, for any $\boldsymbol{\beta} \in D_\beta$, $\phi(z)$ and $\theta(z)$ satisfy conditions given in Section 1. $\boldsymbol{\delta}^0 = (d, \boldsymbol{\beta}')' \in D_l \times D_\beta = D_\delta$. In addition, σ_0^2 is in the interior of the compact space contained in $\mathbb{R}^+$

Assumption 3 (c) is from Yajima (1985). Yajima (1985) 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 $d \in D_1$ and extend this result to the case of any D_l (see Katayama, 2006, Appendix B).

Given a process $\{y_t\}_{t=1}^n$ defined in (1.1), which satisfies Assumption 3, let $\boldsymbol{\delta}^0$ be a true parameter vector $(d,\boldsymbol{\beta}')'$ and assume that $\boldsymbol{\delta}^0$ and $\boldsymbol{\delta}$ are in the same compact parameter space D_δ defined by Assumption 3. The CSS estimator $(\widehat{\boldsymbol{\delta}}', \widehat{\sigma}^2)'$ of $((\boldsymbol{\delta}^0)', \sigma_0^2)'$ is obtained by maximizing the CSS function:

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

where $\varepsilon_t(\boldsymbol{\delta})$ is defined by $\varepsilon_t(\boldsymbol{\delta}) = \sum_{k=0}^{t-1}\pi_k(\boldsymbol{\delta})y_{t-k}$ for $\boldsymbol{\delta} \in D_\delta$. Note that $\widehat{\boldsymbol{\delta}}$ is given by minimizing the objective function $\sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\delta})$, with respect to $\boldsymbol{\delta}$ and set $\widehat{\sigma}^2 = \sum_{t=1}^n \varepsilon_t(\widehat{\boldsymbol{\delta}})^2/n$. Then, by Katayama (2006, Theorem 1 and Remark 1), the following holds, as $n \to \infty$:

$$\widehat{\boldsymbol{\delta}} \xrightarrow{a.c.} \boldsymbol{\delta}^0 \quad \text{and} \quad \sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}), \tag{3.8}$$

where $\boldsymbol{I}(\boldsymbol{\delta}^0)$ is a positive definite matrix given by:

$$\boldsymbol{I}(\boldsymbol{\delta}^0) = \begin{pmatrix} \pi^2/6 & \boldsymbol{\kappa}' \\ \boldsymbol{\kappa} & \boldsymbol{\Phi} \end{pmatrix} = \sum_{k=1}^{\infty}\boldsymbol{\delta}_k\boldsymbol{\delta}_k', \quad \frac{\partial\varepsilon_t(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\delta}} = \sum_{k=1}^{t-1}\boldsymbol{\delta}_k L^k\varepsilon_t, \tag{3.9}$$

$\boldsymbol{\delta}_k = (s_k, \boldsymbol{\delta}_{k,\beta}')'$ is defined by $\partial\varepsilon_t(\boldsymbol{\delta}^0)/\partial d = \sum_{k=1}^{t-1} s_k L^k \varepsilon_t$ and $\partial\varepsilon_t(\boldsymbol{\delta}^0)/\partial\boldsymbol{\beta} = \sum_{k=1}^{t-1}\boldsymbol{\delta}_{k,\beta}L^k\varepsilon_t$. Each element of $\{\boldsymbol{\delta}_k\}$ is defined as follows:

$$\frac{\partial\varepsilon_t(\boldsymbol{\delta}^0)}{\partial d} = \log(1-L^s)\varepsilon_t = -\sum_{k=1}^{t-1}\frac{1}{k}L^{ks}\varepsilon_t = -\sum_{k=1}^{t-1}s_k L^k\varepsilon_t, \tag{3.10}$$

$$\frac{\partial\varepsilon_t(\boldsymbol{\delta}^0)}{\partial\phi_j} = -\phi^{-1}(L)L^j\varepsilon_t = -\sum_{k=0}^{t-j-1}\phi_k^* L^{k+j}\varepsilon_t = -\sum_{k=j}^{t-1}\phi_{k-j}^* L^k\varepsilon_t, \qquad \text{for } j = 1,\dots,p$$

$$\frac{\partial\varepsilon_t(\boldsymbol{\delta}^0)}{\partial\theta_j} = -\theta^{-1}(L)L^j\varepsilon_t = -\sum_{k=0}^{t-j-1}\theta_k^* L^{k+j}\varepsilon_t = -\sum_{k=j}^{t-1}\theta_{k-j}^* L^k\varepsilon_t, \qquad \text{for } j = 1,\dots,q$$

where $s_j = s/j$ for $j = s, 2s, \dots,$;$= 0$ otherwise, ϕ_j^* and θ_j^* are coefficients in the expansions $\phi^{-1}(z) = \sum_{j=0}^{\infty}\phi_j^* z^j$ and $\theta^{-1}(z) = \sum_{j=0}^{\infty}\theta_j^* z^j$, respectively. Especially, $\sum_{k=1}^{\infty}s_k^2 = \pi^2/6$, $\boldsymbol{\kappa} = \sum_{k=1}^{\infty}\boldsymbol{\delta}_{k,\beta}s_k$ and $\boldsymbol{\Phi} = \sum_{k=1}^{\infty}\boldsymbol{\delta}_{k,\beta}\boldsymbol{\delta}_{k,\beta}'$.

Furthermore, as a consequence of Corollary 2.1, we obtain, as $n \to \infty$:

$$n\,\mathrm{E}\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right)\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right)' \longrightarrow \boldsymbol{I}(\boldsymbol{\delta}^0)^{-1} \qquad \text{and} \qquad \mathrm{E}\left\|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right\|^r = O(n^{-r/2}), \tag{3.11}$$

for $r \geq 1$. The proof of (3.11) is given in the Appendix B.

Let $c_j(h) \equiv c_j(h, \boldsymbol{\delta}^0)$, then the BLP $y_n(h)$ based on $\{y_t\}_{t=1}^n$ is $y_n(h) = \sum_{j=1}^n c_j(h,\boldsymbol{\delta}^0)y_{n+1-j} = \sum_{j=h}^{n+h-1}\psi_j(\boldsymbol{\delta}^0)\varepsilon_{n+h-j}$, whereas $\widehat{y}_n(h)$ is defined by:

$$\widehat{y}_n(h) = \sum_{j=1}^{n}c_j(h,\widehat{\boldsymbol{\delta}})y_{n+1-j} = \sum_{j=h}^{n+h-1}\psi_j(\widehat{\boldsymbol{\delta}})\widehat{\varepsilon}_{n+h-j}, \tag{3.12}$$

where $c_j(h,\widehat{\boldsymbol{\delta}})$ and $\psi_j(\widehat{\boldsymbol{\delta}})$ are given by substituting $\widehat{\boldsymbol{\delta}}$ for $\boldsymbol{\delta}^0$ into $c_j(h,\boldsymbol{\delta}^0)$ and $\psi_j(\boldsymbol{\delta}^0)$, respectively, and $\{\widehat{\varepsilon}_t\}_{t=1}^n$ is the residual sequence given by $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\delta}}) = \sum_{j=0}^{t-1}\pi_j(\widehat{\boldsymbol{\delta}})y_{t-j}$. The final equality of (3.12) is given in the Appendix B.

The prediction error of $\widehat{y}_n(h)$ is given by:

$$
\begin{aligned}
y_{n+h} - \widehat{y}_n(h) &= y_{n+h} - y_n(h) + y_n(h) - \widehat{y}_n(h) \\
&= \sum_{j=0}^{h-1} \psi_j(\boldsymbol{\delta}^0)\varepsilon_{n+h-j} - \sum_{j=1}^{n} \Big\{ c_j(h, \widehat{\boldsymbol{\delta}}) - c_j(h, \boldsymbol{\delta}^0) \Big\} y_{n+1-j}. \qquad (3.13)
\end{aligned}
$$

Since the first and second terms on the RHS of (3.13) are mutually independent, the asymptotic PMSE of $\widehat{y}_n(h)$, denoted by $\widehat{\sigma}_y^2(h)$, is defined as:

$$
\widehat{\sigma}_y^2(h) \equiv \mathrm{E}\Big[y_{n+h} - \widehat{y}_n(h) \Big]^2 = \sigma_y^2(h) + \mathrm{E}\left[\sum_{j=1}^{n} \Big\{ c_j(h, \widehat{\boldsymbol{\delta}}) - c_j(h, \boldsymbol{\delta}^0) \Big\} y_{n+1-j} \right]^2, \qquad (3.14)
$$

where $\sigma_y^2(h)$ is $\sigma_0^2 \sum_{j=0}^{h-1} \psi_j^2$ given by (3.4) with $\psi_j = \psi_j(\boldsymbol{\delta}^0)$.

To derive the asymptotic PMSE of $\widehat{y}_n(h)$, we first establish the following lemmas, which are used repeatedly in what follows.

Lemma 3.1 *Let the random variables be defined by:*

$$
x_{i,n} = \sum_{j=0}^{n-1} \alpha_{i,j}\varepsilon_{n-j}, \quad \text{and} \quad z_{i,n} = \sum_{j=0}^{n-2} \left(\sum_{k=1}^{n-j-1} \beta_{i,k}\varepsilon_{n-j-k} \right) \varepsilon_{n-j}, \quad i = 1, 2,
$$

where $\{\varepsilon_t\} \sim iid\ (0, \sigma_0^2)$, $\mathrm{E}[\varepsilon_t]^4 < \infty$, *as* $j \to \infty$, $\alpha_{i,j} = O(j^{-a})$ *and* $\beta_{i,j} = O(j^{-1})$ *for* $i = 1, 2$, $a \in (1/2, 1]$. *Then it follows that, as* $n \to \infty$*:*

$$
\begin{aligned}
&\mathrm{E}[x_{1,n}x_{2,n}z_{1,n}z_{2,n}]/n = \mathrm{E}[x_{1,n}x_{2,n}]\,\mathrm{E}[z_{1,n}z_{2,n}]/n + o(1), \qquad (3.15a)\\
\text{and} \quad &\mathrm{E}[x_{1,n}x_{2,n}z_{1,n}] = o(\sqrt{n}). \qquad (3.15b)
\end{aligned}
$$

Since the proof can be obtained from the fact that $\sum_{k=1}^{n} k^{-1} \leq 1 + \log n$ and $\sum_{k=1}^{n} k^{c-1} \leq c^{-1}n^c$ $(0 < c < 1)$, we omit it. Note that $\mathrm{E}[x_{1,n}x_{2,n}]\,\mathrm{E}[z_{1,n}z_{2,n}]/n = O(1)$ as $n \to \infty$.

Hereafter, let $c_j^{(i)}(h, \boldsymbol{\delta})$ be the i-th derivatives of $c_j(h, \boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$.

Lemma 3.2 *It holds that* $\pi_j^{(i)}(\boldsymbol{\delta}^0) = O((\log j)^i j^{-d-1})$, $\psi_j^{(i)}(\boldsymbol{\delta}^0) = O((\log j)^i j^{d-1})$, *and* $c_j^{(i)}(h, \boldsymbol{\delta}^0) = O((\log j)^i j^{-d-1})$, $i = 0, 1, 2$, *as* $j \to \infty$.

We omit the proofs since these results are obtained in the same way as those in, for example, Sections 2.11 and (8.8.6) of Fuller (1996).

By a Taylor expansion of $\widehat{y}_n(h)$ around $\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta}^0$, we obtain:

$$
\sum_{j=1}^{n} \Big\{ c_j(h, \widehat{\boldsymbol{\delta}}) - c_j(h, \boldsymbol{\delta}^0) \Big\} y_{n+1-j} = \sum_{j=1}^{n} c_j^{(1)}(h, \boldsymbol{\delta}^0)'(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) y_{n+1-j} + R_{1,n}, \qquad (3.16)
$$

where:

$$
R_{1,n} = \sum_{j=1}^{n} (\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0)' c_j^{(2)}(h, \boldsymbol{\delta}^*)(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) y_{n+1-j},
$$

and $\|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\| \geq \|\boldsymbol{\delta}^* - \boldsymbol{\delta}^0\|$. Using (3.11), we obtain the following theorem.

Theorem 3.1 *Let* $\{y_t\}_{t=1}^n$ *be given by* (1.1) *and Assumption 3. Then it follows that, as* $n \to \infty$*:*

$$
\widehat{\sigma}_y^2(h) \equiv \mathrm{E}\Big[y_{n+h} - \widehat{y}_n(h) \Big]^2 = \sigma_y^2(h) + \frac{\sigma_0^2}{n} \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j(h, \boldsymbol{\delta}^0)' \boldsymbol{I}(\boldsymbol{\delta}^0)^{-1} \boldsymbol{\varphi}_j(h, \boldsymbol{\delta}^0) + o\left(\frac{1}{n}\right), \qquad (3.17)
$$

where $\sigma_y^2(h)$ *and* $\widehat{y}_n(h)$ *are given by* (3.14) *and* (3.12), *respectively:*

$$
\boldsymbol{\varphi}_j(h, \boldsymbol{\delta}^0) \equiv \sum_{k=0}^{j} c_{k+1}^{(1)}(h, \boldsymbol{\delta}^0)\psi_{j-k}(\boldsymbol{\delta}^0) = -\sum_{k=0}^{h-1} \psi_k(\boldsymbol{\delta}^0)\boldsymbol{\delta}_{j+h-k},
$$

and $\{\boldsymbol{\delta}_k\}$ *is given by* (3.9) *and* (3.10), *and* $\boldsymbol{I}(\boldsymbol{\delta}^0)$ *are given by* (3.8).

Note that (3.17) indicates that PMSE of $\widehat{y}_n(h)$ converges to that of BLP with $O(1/n)$. As well, the positive number divided by n in (3.17) consists of information about PMSE of BLP, $\sigma_y^2(h) = \mathrm{E}[\sum_{j=0}^{h-1}\psi_j(\boldsymbol{\delta}^0)\varepsilon_{n+h-j}]^2$ and the asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)$, $\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1} = \{\sum_{k=1}^{\infty}\boldsymbol{\delta}_k\boldsymbol{\delta}_k'\}^{-1}$.

The following results follow from the preceding theorem.

Corollary 3.1 *Under the same conditions as in Theorem 3.1, it follows that, as $n \to \infty$:*

$$\widehat{\sigma}_y^2(1) \equiv \mathrm{E}\Big[y_{n+1} - \widehat{y}_n(1)\Big]^2 = \sigma_0^2\left(1 + \frac{p+q+1}{n}\right) + o\left(\frac{1}{n}\right). \tag{3.18}$$

Example 3.1 Let $\{y_t\}$ be given by the AR(p) process, $y_t = \sum_{i=1}^{p}\phi_i y_{t-i} + \varepsilon_t$. Then $\widehat{y}_n(h) = \sum_{j=1}^{p} c_j(h,\widehat{\boldsymbol{\delta}})y_{n+1-j}$. By Corollary 3.1, $\widehat{\sigma}_y^2(1) \sim \sigma_0^2(1+p/n)$ as $n \to \infty$. Let $\{y_t\}$ be given by the AR(1) process, $y_t = \phi y_{t-1} + \varepsilon_t$ where $0 < |\phi| < 1$ and $\phi = \phi_1$. Then, $\widehat{y}_n(h) = c_1(h,\widehat{\phi})y_n = \widehat{\phi}^h y_n$, $\boldsymbol{I}(\boldsymbol{\delta}^0) = 1/(1-\phi^2)$, $\partial c_1(h,\phi)/\partial\phi = h\phi^{h-1}$, and $\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) = h\phi^{h-1}\phi^j$. It follows that $\sum_{j=0}^{\infty}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)'\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) = h^2\phi^{2(h-1)}$, $h > 0$. Also, $\sigma_y^2(h) = \sigma_0^2\sum_{j=0}^{h-1}\phi^{2j}$. These results are consistent with Yamamoto (1976, p.126).

Furthermore, let $\{y_t\}$ be given by the autoregressive moving average ARMA$(1,1)$ model, $y_t = \phi y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t$, where $0 < |\phi|, |\theta| < 1$ and $(\phi,\theta) = (\phi_1,\theta_1)$. Then, $\boldsymbol{\delta}^0 = (\phi,\theta)'$, $\boldsymbol{\delta}_j = (-\phi^{j-1}, -(-\theta)^{j-1})'$, $\psi_j(\boldsymbol{\delta}^0) = 1$ for $j = 0$; $= (\phi+\theta)\phi^{j-1}$ for $j \geq 1$:

$$\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) = \begin{pmatrix} \phi^j\{(h-1)\phi^{h-2}(\phi+\theta)+\phi^{h-1}\} \\ \phi^{h-1}(-\theta)^j \end{pmatrix}, \quad \boldsymbol{I}(\boldsymbol{\delta}^0) = \begin{pmatrix} 1/(1-\phi^2) & 1/(1+\phi\theta) \\ 1/(1+\phi\theta) & 1/(1-\theta^2) \end{pmatrix},$$

and $\widehat{\sigma}_y^2(h) = \sigma_y^2(h) + \dfrac{\sigma_0^2}{n}\Big\{2\phi^{2(h-1)} + 2(h-1)(\phi+\theta)\phi^{2h-3} + (h-1)^2(1+\phi\theta)^2\phi^{2(h-2)}\Big\} + o\left(\dfrac{1}{n}\right)$,

as $n \to \infty$, where $\sigma_y^2(h) = \sigma_0^2$ for $h = 1$; $= \sigma_0^2\{1 + (\phi+\theta)^2\sum_{j=0}^{h-2}\phi^{2j}\}$ for $h \geq 2$. These results are consistent with Yamamoto (1981, pp.489–490).

3.2 The effects of misspecification of the integration order

We rewrite the model (1.1) as:

$$(1-L^s)^{d_0+\theta}y_t = \beta(L)\varepsilon_t \qquad t \geq 1, \tag{3.19}$$

and $y_t = 0$, $t \leq 0$, where $d = d_0 + \theta$. We assume that the process (3.19) satisfies Assumption 3 and $d_0, d \in D_l$, for some $l = 1,2,3$. The predictors with estimated parameters considered here are $\widehat{y}_n(h)$, which are discussed in Section 3.1, and $\widetilde{y}_n(h)$, which is given by fixed d_0 with estimators of ARMA(p,q) parameters, denoted by $\widetilde{\boldsymbol{\beta}}$, from the process $\{(1-L^s)^{d_0}y_t\}$. That is, let $\widetilde{\boldsymbol{\delta}} = \big(d_0, \widetilde{\boldsymbol{\beta}}'\big)'$, then:

$$\widetilde{y}_n(h) = \sum_{j=1}^{n} c_j(h,\widetilde{\boldsymbol{\delta}})y_{n+1-j} = \sum_{j=h}^{n+h-1}\psi_j(\widetilde{\boldsymbol{\delta}})\widetilde{\varepsilon}_{n+h-j}, \tag{3.20}$$

where $\widetilde{\varepsilon}_t = \varepsilon_t(\widetilde{\boldsymbol{\delta}}) = \sum_{j=0}^{t-1}\pi_j(\widetilde{\boldsymbol{\delta}})y_{t-j}$.

To make a comparison between the two predictors, $\widehat{y}_n(h)$ and $\widetilde{y}_n(h)$, we assume that $\theta = c/\sqrt{n}$ and that c is a fixed constant. The following theorem shows the asymptotic PMSE of $\widetilde{y}_n(h)$.

Theorem 3.2 *Let $\{y_t\}_{t=1}^n$ be given by (3.19) and $\theta = c/\sqrt{n}$ where c is a fixed constant. Then it follows that, as $n \to \infty$:*

$$\widetilde{\sigma}_y^2(h) \equiv \mathrm{E}\Big[y_{n+h} - \widetilde{y}_n(h)\Big]^2 = \sigma_y^2(h) + \frac{\sigma_0^2}{n}\sum_{j=0}^{\infty}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)'\,\boldsymbol{\Gamma}\,\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) + o\left(\frac{1}{n}\right), \tag{3.21}$$

where $\sigma_y^2(h)$ and $\widetilde{y}_n(h)$ are given by (3.14) and (3.20), respectively, $\{\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)\}$ is defined by Theorem 3.1. $\boldsymbol{\Gamma}$ is given by:

$$\boldsymbol{\Gamma} = \begin{pmatrix} c^2 & -c^2\boldsymbol{\kappa}'\boldsymbol{\Phi}^{-1} \\ -c^2\boldsymbol{\Phi}^{-1}\boldsymbol{\kappa} & \boldsymbol{\Phi}^{-1} + c^2\boldsymbol{\Phi}^{-1}\boldsymbol{\kappa}'\boldsymbol{\kappa}\boldsymbol{\Phi}^{-1} \end{pmatrix},$$

where $\boldsymbol{\kappa}$ and $\boldsymbol{\Phi}$ are given by (3.9).

Note that $\boldsymbol{\Gamma}$ is similar to asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0)$, $\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}$. (3.21) indicates that the PMSE of $\widetilde{y}_n(h)$ converges to that of BLP with $O(1/n)$.

Remark 1 Hereafter, to make a comparison between the two different predictors, we define *asymptotic relative efficiency* (ARE) as follows: for a stochastic process $\{y_t\}$, let $\widehat{y}_n(h)$ and $\widetilde{y}_n(h)$ be two predictors of a future value, y_{n+h}, when $\{y_t\}_{t=j}^n$ for some $j < n$ is given. Then the asymptotic relative efficiency between $\widehat{y}_n(h)$ and $\widetilde{y}_n(h)$, denoted by ARE $\left[\widehat{y}_n(h),\ \widetilde{y}_n(h)\right]$, is defined by:

$$\mathrm{ARE}\left[\widehat{y}_n(h),\ \widetilde{y}_n(h)\right] \equiv \lim_{n\to\infty} n\left(\mathrm{E}\left[y_{n+h} - \widehat{y}_n(h)\right]^2 - \mathrm{E}\left[y_{n+h} - \widetilde{y}_n(h)\right]^2\right). \tag{3.22}$$

The asymptotic PMSE of $\widehat{y}_n(h)$ is given by (3.17), which leads to the following result.

Corollary 3.2 *Under the same conditions as in Theorem 3.2, it follows that:*

$$\mathrm{ARE}\left[\widehat{y}_n(h),\ \widetilde{y}_n(h)\right] \begin{cases} \geq 0 & \textit{iff } \omega \geq |\,c\,|; \\ \leq 0 & \textit{iff } \omega \leq |\,c\,|; \end{cases}, \tag{3.23}$$

where $\omega = (\pi^2/6 - \boldsymbol{\kappa}'\boldsymbol{\Phi}^{-1}\boldsymbol{\kappa})^{-1/2}$.

Note that ω is the standard error of the limiting distribution of $\sqrt{n}(\widehat{d} - d)$ given by (3.8). The following result also follows from the preceding theorem.

Corollary 3.3 *Under the same conditions as in Theorem 3.2, it follows that, as* $n \to \infty$*:*

$$\widetilde{\sigma}_y^2(1) \equiv \mathrm{E}\left[y_{n+1} - \widetilde{y}_n(1)\right]^2 = \sigma_0^2\left(1 + \frac{p+q+c^2/\omega^2}{n}\right) + o\left(\frac{1}{n}\right), \tag{3.24}$$

where ω *is given by Corollary 3.2.*

Example 3.2 Let $\{y_t\}_{t=1}^n$ be given by $(1-L)^d y_t = \varepsilon_t$ $(t \geq 1)$ and $y_t = 0$ $(t \leq 0)$ where $d \in (-1/2, 1/2)$. Then $\widehat{y}_n(1) = -\sum_{j=1}^n \pi_j(\widehat{d}) y_{n+1-j}$ and $\widehat{\sigma}_y^2(1)$ is, for large n, $(1+1/n)\sigma_0^2$. If $h \geq 2$, then $\widehat{y}_n(h)$ is given by:

$$\widehat{y}_n(h) = \sum_{j=1}^{n} c_j(h, \widehat{d}) y_{n+1-j} = \sum_{j=h}^{n+h-1} \psi_j(\widehat{d}) \widehat{\varepsilon}_{n+h-j},$$

where $\widehat{\varepsilon}_j = \sum_{k=0}^{j-1} \pi_k(\widehat{d}) y_{j-k}$, $\psi_j(d)$ and $\pi_j(d)$ are given by (3.5) and (3.6), respectively, and:

$$c_j(h,d) = -\sum_{i=0}^{h-1} \psi_i(d) \pi_{j+h-i-1}(d) = -\frac{\Gamma(h+d)}{\Gamma(1+d)\Gamma(h)} \frac{j\pi_j(d)}{j+h-1},$$

by equation (2.2.6) in Miller (1994). The asymptotic PMSE of $\widehat{y}_n(h)$ is given by, as $n \to \infty$:

$$\widehat{\sigma}_y^2(h) \equiv \mathrm{E}\left[y_{n+h} - \widehat{y}_n(h)\right]^2 = \sigma_0^2 \sum_{j=0}^{h-1} \psi_j(d)^2 + \frac{\sigma_0^2}{n} C(h,d) + o\left(\frac{1}{n}\right),$$

where:

$$\begin{aligned} C(h,d) &= \frac{6}{\pi^2} \sum_{j=0}^{\infty} \left\{ \sum_{k=0}^{h-1} \frac{\psi_k(d)}{j+h-k} \right\}^2 \\ &= \sum_{j=0}^{h-1} \psi_j(d)^2 - \frac{6}{\pi^2} \sum_{j=0}^{h-2} \psi_j(d)^2 \sum_{k=1}^{h-j-1} \frac{1}{k^2} + \frac{12}{\pi^2} \sum_{1 \leq j < k \leq h} \frac{\psi_{h-j}(d)\psi_{h-k}(d)}{k-j} \sum_{i=j}^{k} \frac{1}{i}. \end{aligned}$$

Let d be $d_0 + \theta$, where θ and d_0 are fixed constants, then $\widetilde{y}_n(1) = -\sum_{j=1}^{n} \pi_j(d_0) y_{n+1-j}$ and $\widetilde{\sigma}_y^2(1) = \mathrm{E}[(1-L)^{-\theta}\varepsilon_{n+1}]^2$, which is, for large n, the variance of the stationary ARFIMA$(0,\theta,0)$ model, $\sigma_0^2\Gamma(1-2\theta)/\Gamma^2(1-\theta)$.

Let d be $d_0 + \theta$, where d_0 and θ are a fixed constant and $c/\sqrt{n}$, respectively; then, as $n \to \infty$, $\widetilde{\sigma}_y^2(1) \sim \{1 + \pi^2 c^2/(6n)\}\sigma_0^2$. If $h \geq 2$, then $\widetilde{\sigma}_y^2(h) \sim \{\sum_{j=0}^{h-1} \psi_j(d)^2 + \pi^2 c^2 C(h,d)/(6n)\}\sigma_0^2$, as $n \to \infty$. Therefore, under this model the sign of ARE$[\,\widehat{y}_n(h),\ \widetilde{y}_n(h)\,]$ depends on whether $|\,c\,|$ is greater than $\omega = \sqrt{6}/\pi \approx 0.78$.

Example 3.3 Similar results are also obtained if the model of $\{y_t\}$ in Example 3.2 is replaced by $(1-L^s)^d y_t = \varepsilon_t$ where $t \geq 1$, s is even, and $d \in (-1/2, 1/2)$. The expression of the asymptotic PMSE of $\widehat{y}_n(1) = -\sum_{j=1}^{n} \pi_j(\widehat{d}) y_{n+1-j}$ is the same as that of Example 3.2. $\widehat{y}_n(h)$ is given by $\widehat{y}_n(h) = \sum_{j=h}^{n+h-1} \psi_j(\widehat{d})\widehat{\varepsilon}_{n+h-j}$, where $\widehat{\varepsilon}_j = \sum_{k=0}^{j-1} \pi_k(\widehat{d}) y_{j-k}$, $\psi_j(d)$ and $\pi_j(d)$ are given by (3.5) and (3.6), respectively. Letting d be $d_0 + \theta$, where θ and d_0 are fixed constants, then $\widetilde{\sigma}_y^2(1) = \mathrm{E}[(1-L^s)^{-\theta}\varepsilon_{n+1}]^2$, which is, for large n, the variance of the stationary ARFISMA$(0,\theta,0)$ model, $\sigma_0^2\Gamma(1-2\theta)/\Gamma^2(1-\theta)$. Let θ be $c/\sqrt{n}$; then, as $n \to \infty$, $\widetilde{\sigma}_y^2(1) \sim \{1 + \pi^2 c^2/(6n)\}\sigma_0^2$, which implies that we can evaluate the ARE$[\,\widehat{y}_n(1),\ \widetilde{y}_n(1)\,]$ as in Example 3.2.

4 PMSE for integrated long-memory processes

We discuss the asymptotic PMSE of non-stationary ARFISMA processes $\{x_t\}$, which is already given by (1.2) in Section 1:

$$(1-L^s)^m x_t = y_t = (1-L^s)^{-d}\beta(L)\varepsilon_t, \qquad t \geq 1, \tag{4.1}$$

and $x_{t-ms} = y_t = 0$, $t \leq 0$, where $\{y_t\}$ is given by (1.1) with Assumption 3, the vector $(x_{1-ms}, \ldots, x_0)$ is any random vector and is uncorrelated with y_t, $t \geq 1$, and m is known zero or a positive integer.

As in Brockwell and Davis (1991, Section 9.5), given the data $\{x_t\}_{t=1-ms}^{n}$ from (4.1), we can obtain the BLP, denoted by $x_n(h)$, based on $x_{1-ms}, x_{2-ms}, \ldots, x_n$ for a future value x_{n+h}, and define the predictor with estimated parameters using the results of Sections 3.1 and 3.2. The following example shows the asymptotic PMSE of the two predictors with estimated parameters, $\widehat{x}_n(h)$ and $\widetilde{x}_n(h)$.

Example 4.1 ($m = s = 1$ and $\beta(z) = 1$). Let $\{x_t\}_{t=0}^{n}$ be given by $(1-L)x_t = y_t = (1-L)^{-d}\varepsilon_t$ and $d = c/\sqrt{n}$. Then $x_n(h) = x_n + \sum_{j=0}^{h-1} y_n(h-j)$ and $\widehat{x}_n(h) = x_n + \sum_{j=0}^{h-1} \widehat{y}_n(h-j)$ where $y_n(k) = \sum_{l=0}^{n-1} \psi_{l+k}(d)\varepsilon_{n-l}$ and $\widehat{y}_n(k)$ is given similarly to Example 3.2 for $k = 1, \ldots, h$, respectively, and $\widetilde{x}_n(h) = x_n$. By Corollary 3.2 and Example 3.2, the sign of ARE$[\,\widehat{x}_n(h),\ \widetilde{x}_n(h)\,]$ depends on whether $|\,c\,|$ is greater than $\omega = \sqrt{6}/\pi \approx 0.78$.

Example 4.2 ($m = 1$, s is even, and $\beta(z) = 1$). Let $\{x_t\}_{t=1-s}^{n}$ be given by $(1-L^s)x_t = y_t = (1-L^s)^{-d}\varepsilon_t$ with an even integer s and $d = c/\sqrt{n}$. If $h \leq s$, then $x_n(h) = x_{n+h-s} + y_n(h)$ and $\widehat{x}_n(h) = x_{n+h-s} + \widehat{y}_n(h)$ where $y_n(h) = \sum_{l=0}^{n-1} \psi_{l+h}(d)\varepsilon_{n-l}$ and $\widehat{y}_n(h)$ is given similarly to Example 3.3, and $\widetilde{x}_n(h) = x_{n+h-s}$. By Corollary 3.2 and Example 3.3, the sign of ARE$[\,\widehat{x}_n(h),\ \widetilde{x}_n(h)\,]$ depends on whether $|\,c\,|$ is greater than $\omega = \sqrt{6}/\pi$.

Example 4.3 ($m = s = 1$ and $\beta(z) = (1-\phi z)^{-1}$). Let $\{x_t\}_{t=0}^{n}$ be given by $(1-L)x_t = y_t = (1-L)^{-d}(1-\phi L)^{-1}\varepsilon_t$ with $|\phi| < 1$ and $d = c/\sqrt{n}$. Then $x_n(h) = x_n + \sum_{j=0}^{h-1} y_n(h-j)$ and $\widehat{x}_n(h) = x_n + \sum_{j=0}^{h-1} \widehat{y}_n(h-j)$ where $y_n(h) = \sum_{l=0}^{n-1} \psi_{l+h}(d)\varepsilon_{n-l}$, $\boldsymbol{\delta}^0 = (d,\phi)'$, and $\widehat{y}_n(k) = \sum_{j=1}^{n} c_j(k,\widehat{\boldsymbol{\delta}}) y_{n+1-j}$ for $k = 1, \ldots, h$, whereas $\widetilde{x}_n(h) = x_n + \sum_{j=0}^{h-1} \widetilde{y}_n(h-j)$ where $\widetilde{y}_n(k) = \widetilde{\phi}^k y_n$ for $k = 1, \ldots, h$ as in the AR(1) case of Example 3.1. By Corollary 3.2, the sign of ARE$[\,\widehat{x}_n(h),\ \widetilde{x}_n(h)\,]$ depends on c and $\omega = \left[\pi^2/6 - \{\log(1-\phi)\}^2(1-\phi^2)/\phi^2\right]^{-1/2}$.

5 Some simulations for PMSE and model selection

In this section we use simulations to examine the finite sample performance of the predictors discussed in the previous section. The calculations are made using S-PLUS Version 4.5. Observations in the models were generated by Cholesky decomposition of the covariance matrix of the process (see Sections 11.3.1 and 11.3.5 in Beran, 1994). In addition, the Gauss–Newton procedure is used for the maximization of $S(\boldsymbol{\delta}, \sigma^2)$, for which Tanaka (1999, Section 5) follows concrete procedures.

The models used are:

$$\begin{aligned}
&\text{DGP 1: } x_t = x_{t-1} + (1-L)^{-d}\varepsilon_t, && (t = 1, \ldots, n+1),\\
&\text{DGP 2: } x_t = x_{t-12} + (1-L^{12})^{-d}\varepsilon_t, && (t = 1, \ldots, n+12),\\
&\text{DGP 3: } x_t = x_{t-1} + (1-L)^{-d}(1-\phi L)^{-1}\varepsilon_t, && (t = 1, \ldots, n+1),
\end{aligned}$$

where $\{\varepsilon_t\} \sim NID(0,1)$. We consider the predictors $\widetilde{x}_n(h)$ and $\widehat{x}_n(h)$, which are given by Examples 4.1, 4.2, and 4.3. For all simulations, we fix the number of replications at 10,000 and $h = 1, 3, 5$. The first row labelled d denotes the true value of d and the second row labelled c/ω is given by $c = \sqrt{n}d$ where ω is given by Examples 4.1, 4.2, and 4.3. Here, the simulated square root of PMSE, denoted SRPMSE, for h-step ahead is given by:

$$\widetilde{\sigma}_x(h) = \sqrt{\frac{1}{10000}\sum_{j=1}^{10000}\left\{\widetilde{x}_{n,j}(h) - x_{n+h,j}\right\}^2} \text{ and } \widehat{\sigma}_x(h) = \sqrt{\frac{1}{10000}\sum_{j=1}^{10000}\left\{\widehat{x}_{n,j}(h) - x_{n+h,j}\right\}^2},$$

where $x_{n+h,j}$ denotes the actual value of x_{n+h} at the jth simulation of the relevant model above, and $\widetilde{x}_{n,j}(h)$ and $\widehat{x}_{n,j}(h)$ are also predictors of $\widetilde{x}_n(h)$ and $\widehat{x}_n(h)$ from data $\{x_t\}$ and differenced data of sample size n at the jth simulation, respectively. Each table employs underlining to distinguish the smaller of $\widetilde{\sigma}_x(h)$ and $\widehat{\sigma}_x(h)$. Also reported is the SRPMSE of the BLP, $\sigma_x(h)$.

The tables also provide the rate of selection of $\widehat{x}_n(h)$ using the Wald test statistics:

$$W = \sqrt{n}\,\widehat{d}\,/\,\omega. \tag{5.1}$$

This is because, following from Corollary 3.2, $\sqrt{n}|d|/\omega = |c|/\omega \geq 1$ is a necessary and sufficient condition of ARE$[\,\widehat{x}_n(h),\ \widetilde{x}_n(h)\,] \geq 0$, and we can estimate d and ω from CSS methods. Given Tanaka (1999, Section 4), it is known that:

$$\Pr(W > z_\alpha) \longrightarrow \Pr(Z > z_\alpha - c/\omega), \quad \text{as } n \to \infty, \tag{5.2}$$

where $Z \sim N(0,1)$ and z_α is an upper $100\alpha\%$ point of Z. Therefore, when $c \geq 0$, it holds, as $n \to \infty$:

$$\Pr(W > 1 \,|\, c/\omega > 1) \longrightarrow \Pr(Z > 1 - c/\omega \,|\, c/\omega > 1) \geq 0.5. \tag{5.3}$$

It follows that when $\widehat{x}_n(h)$ is more desirable than $\widetilde{x}_n(h)$ in terms of ARE of PMSE, the asymptotic probability of selecting $\widehat{x}_n(h)$ by $W > 1$ is larger than 0.5 and increases as $c \to \infty$. This is a right-sided test, $H_0 : d = 0$ vs $H_1 : d > 0$ using Wald test statistics with a significance level of about 16%. Therefore, Wald test statistics are a candidate for model selection in terms of the ARE of PMSE. Based on these results, we report the rate of selection of predictor $\widehat{x}_n(h)$ as follows. When $d \geq 0$, the rows labelled $W(s)$ denote the percentages of $W > 1$ (selection of $\widehat{x}_n(h)$) and the rows of $W(100\alpha)$ ($\alpha = 0.10, 0.05, 0.01$) denote the percentages of $W > z_\alpha$ (selection of $\widehat{x}_n(h)$) of the right-sided test, $H_0 : d = 0$ vs $H_1 : d > 0$. Similarly, when $d \leq 0$, the rows labelled $W(s)$ denote the percentages of $W < -1$ and the rows of $W(100\alpha)$ denote the percentages of $W < -z_\alpha$ of the left-sided test, $H_0 : d = 0$ vs $H_1 : d < 0$. We also report the percentage of selection of predictor $\widehat{x}_n(h)$ using Akaike information criterion (AIC). These are defined by:

$$\widetilde{AIC} = -2S(\widetilde{\boldsymbol{\delta}}, \widetilde{\sigma}^2) + 2(p+1), \quad \text{and} \quad \widehat{AIC} = -2S(\widehat{\boldsymbol{\delta}}, \widehat{\sigma}^2) + 2(p+2),$$

where p is given by $p = 0$ for DGP 1 and DGP 2 and $p = 1$ for DGP 3, respectively. The rows labelled AIC denote the percentage of $\widehat{AIC} < \widetilde{AIC}$. Namely, the percentage of approximate likelihood ratio test statistics is larger than 2 for the hypothesis test $H_0 : d = 0$ vs $H_1 : d \neq 0$ with significance level about 16%.

These experiments reveal the following.

1. For the results on SRPMSE:
   (a) Both $\widetilde{\sigma}_x(h)$ and $\widehat{\sigma}_x(h)$ approximate $\sigma_x(h)$ well. These results are consistent with the theoretical results given by Sections 3 and 4 because $\widetilde{\sigma}_x^2(h)$ and $\widehat{\sigma}_x^2(h)$ converges to $\sigma_x^2(h)$ with $O(1/n)$.
   (b) It indicates the $\widehat{\sigma}_x^2(h) - \widetilde{\sigma}_x^2(h)$ gets smaller as $|c|/\omega$ gets larger. This is consistent with the theoretical results given by Corollary 3.2 because $\text{ARE}[\,\widehat{x}_n(h),\ \widetilde{x}_n(h)\,] = \text{const}(\omega^2 - c^2)$ (see the proof of the Corollary 3.2 in Appendix B).
   (c) Underlines in the rows of $\widetilde{\sigma}_x(h)$ and $\widehat{\sigma}_x(h)$ show that $\widehat{\sigma}_x(h)$ seems to be smaller than $\widetilde{\sigma}_x(h)$ for $|c|/\omega > 1$, conversely, $\widehat{\sigma}_x(h)$ seems to be larger than $\widetilde{\sigma}_x(h)$ for $|c|/\omega < 1$, which is consistent with theoretical results given by Corollary 3.2 and Examples 4.1, 4.2, and 4.3.

2. For the results on model selections:
   (a) The rows of $W(s)$ and $W(100\alpha)$ show the empirical sizes of the left-hand sided tests have relatively large values, while the rows of AIC present a stable size.
   (b) The rows of $W(s)$ show the percentages of selection of $\widehat{x}_n(h)$ are getting larger as $|c|/\omega$ becomes larger and are larger than 50% for $|c|/\omega > 1$, which is consistent with (5.3).
   (c) The rows of $W(100\alpha)$ are more likely to select $\widetilde{x}_n(h)$ (by a simple model) than $\widehat{x}_n(h)$ (by a complex model) even if $\widehat{x}_n(h)$ is more desirable than $\widetilde{x}_n(h)$ in the sense of ARE of PMSE, which is obvious from the theoretical power function given by (5.2). The rows of AIC also show similar properties.

3. Table 3 (a) shows the case of $\phi = 0.6$, $\widehat{\sigma}_x(h)$ performs poorly and $\widehat{\sigma}_x(h) < \widetilde{\sigma}_x(h)$ for $|d| \geq 0.3$. Correspondingly, the model selection results are likely to choose $\widetilde{x}_n(h)$. In the case of $\phi = -0.8$ given by Table 3 (b), however, it seems that $\widehat{x}_n(h)$ performs relatively well and $\widehat{\sigma}_x(h) \leq \widetilde{\sigma}_x(h)$ for $|d| \geq 0.2$. By Example 4.3 and Tanaka (1999), this is closely related to the fact that the estimators of ϕ and d are negatively correlated and the correlation is much higher for $\phi = 0.6$. This indicates that the finite sample performance of $\widehat{x}_n(h)$ is affected by the true ARMA parameters.

4. As a whole, it indicates that classical test statistics, such as the likelihood ratio, Wald, and Lagrange multiplier test statistics, control the probability of selecting the relatively efficient predictor by the levels of significance. In other words, these can serve as model selection criteria for the ARE of PMSE when a complex model hypothesized is a true model.

TABLE 1 (a) The SRPMSE and the percentage of selection of $\widehat{x}_n(h)$ for DGP 1 ($n = 50$)

d	-0.20	-0.15	-0.10	-0.05	0.00	0.05	0.10	0.15	0.20
c/ω	-1.81	-1.36	-0.91	-0.45	0.00	0.45	0.91	1.36	1.81
$\widetilde{\sigma}_x(1)$	1.0089	0.9981	$\underline{1.0044}$	$\underline{1.0127}$	$\underline{1.0087}$	$\underline{0.9892}$	$\underline{1.0168}$	1.0152	1.0533
$\widetilde{\sigma}_x(3)$	1.5134	1.5766	$\underline{1.6051}$	$\underline{1.6657}$	$\underline{1.7418}$	$\underline{1.8241}$	$\underline{1.9238}$	2.0176	2.1619
$\widetilde{\sigma}_x(5)$	1.8234	1.9034	$\underline{1.9820}$	$\underline{2.1063}$	$\underline{2.2485}$	$\underline{2.3845}$	$\underline{2.6220}$	2.8179	3.0740
$\widehat{\sigma}_x(1)$	$\underline{1.0044}$	$\underline{0.9937}$	1.0096	1.0226	1.0234	0.9973	1.0189	$\underline{1.0058}$	$\underline{1.0139}$
$\widehat{\sigma}_x(3)$	$\underline{1.4937}$	$\underline{1.5630}$	1.6124	1.6856	1.7719	1.8460	1.9293	$\underline{1.9932}$	$\underline{2.0650}$
$\widehat{\sigma}_x(5)$	$\underline{1.7967}$	$\underline{1.8899}$	1.9910	2.1345	2.2933	2.4169	2.6296	$\underline{2.7789}$	$\underline{2.9323}$
$\sigma_x(1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma_x(3)$	1.4691	1.5299	1.5941	1.6615	1.7321	1.8058	1.8826	1.9623	2.0451
$\sigma_x(5)$	1.7371	1.8481	1.9681	2.0974	2.2361	2.3844	2.5426	2.7110	2.8897
AIC	54.9	38.6	24.6	16.9	13.1	16.8	27.0	42.6	59.8
$W(s)$	80.9	67.1	50.1	36.0	21.9/13.9	26.4	42.0	58.7	74.5
$W(10)$	72.5	56.9	39.6	27.6	15.9/8.5	18.2	32.0	48.3	66.2
$W(5)$	59.8	43.1	28.1	18.3	10.1/4.2	9.8	20.3	35.3	53.3
$W(1)$	35.2	22.8	12.8	7.6	3.7/0.9	2.4	6.4	14.3	8.6

DGP 1: $\{x_t\}_{t=1}^{51}$ where $x_t = x_{t-1} + (1-L)^{-d}\varepsilon_t$ and $\{\varepsilon_t\} \sim NID(0,1)$.

TABLE 1 (b) The SRPMSE and the percentage of selection of $\widehat{x}_n(h)$ for DGP 1 ($n = 100$)

d	-0.20	-0.15	-0.10	-0.05	0.00	0.05	0.10	0.15	0.20
c/ω	-2.57	-1.92	-1.28	-0.64	0.00	0.64	1.28	1.92	2.57
$\widetilde{\sigma}_x(1)$	1.0228	1.0135	1.0201	$\underline{0.9941}$	$\underline{1.0047}$	$\underline{1.0117}$	1.0071	1.0310	1.0496
$\widetilde{\sigma}_x(3)$	1.5279	1.5657	1.6308	$\underline{1.6571}$	$\underline{1.7433}$	$\underline{1.8158}$	1.9311	2.0560	2.1733
$\widetilde{\sigma}_x(5)$	1.8136	1.9010	1.9987	$\underline{2.1054}$	$\underline{2.2455}$	$\underline{2.3813}$	2.6101	2.8423	3.0871
$\widehat{\sigma}_x(1)$	$\underline{1.0011}$	$\underline{1.0001}$	$\underline{1.0194}$	0.9990	1.0119	1.0158	$\underline{1.0024}$	$\underline{1.0125}$	$\underline{1.0025}$
$\widehat{\sigma}_x(3)$	$\underline{1.4768}$	$\underline{1.5409}$	$\underline{1.6280}$	1.6663	1.7600	1.8283	$\underline{1.9205}$	$\underline{1.9943}$	$\underline{2.0370}$
$\widehat{\sigma}_x(5)$	$\underline{1.7470}$	$\underline{1.8625}$	$\underline{1.9922}$	2.1151	2.2711	2.3966	$\underline{2.5914}$	$\underline{2.7536}$	$\underline{2.8788}$
$\sigma_x(1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma_x(3)$	1.4691	1.5299	1.5941	1.6615	1.7321	1.8058	1.8826	1.9623	2.0451
$\sigma_x(5)$	1.7371	1.8481	1.9681	2.0974	2.2361	2.3844	2.5426	2.7110	2.8897
AIC	81.8	60.7	37.9	19.0	12.5	20.0	40.0	63.9	82.5
$W(s)$	95.0	84.0	63.2	39.7	20.0/13.6	32.1	56.9	78.2	91.0
$W(10)$	91.5	76.1	52.9	29.8	13.9/8.1	22.9	46.3	70.1	86.8
$W(5)$	83.5	63.2	39.8	19.5	8.5/3.7	13.3	32.6	58.0	79.1
$W(1)$	60.9	37.5	18.8	7.7	2.5/0.7	3.2	12.9	32.8	57.2

DGP 1: $\{x_t\}_{t=1}^{101}$ where $x_t = x_{t-1} + (1-L)^{-d}\varepsilon_t$ and $\{\varepsilon_t\} \sim NID(0,1)$.

TABLE 2 (a) The SRPMSE and the percentage of selection of $\widehat{x}_n(h)$ ($n = 50$)

d	-0.20	-0.15	-0.10	-0.05	0.00	0.05	0.10	0.15	0.20
c/ω	-1.81	-1.36	-0.91	-0.45	0.00	0.45	0.91	1.36	1.81
$\widetilde{\sigma}_x(1)$	1.0381	$\underline{1.0148}$	$\underline{1.0129}$	$\underline{1.0089}$	$\underline{0.9977}$	$\underline{1.0022}$	$\underline{1.0100}$	1.0251	1.0609
$\widetilde{\sigma}_x(3)$	1.0313	1.0157	$\underline{1.0018}$	$\underline{0.9995}$	$\underline{1.0036}$	$\underline{1.0028}$	$\underline{1.0004}$	1.0329	1.0456
$\widetilde{\sigma}_x(5)$	1.0385	$\underline{1.0140}$	$\underline{1.0208}$	$\underline{0.9913}$	$\underline{1.0050}$	$\underline{1.0062}$	$\underline{1.0126}$	1.0263	1.0507
$\widehat{\sigma}_x(1)$	$\underline{1.0348}$	1.0197	1.0207	1.0231	1.0117	1.0213	1.0140	$\underline{1.0195}$	$\underline{1.0301}$
$\widehat{\sigma}_x(3)$	$\underline{1.0306}$	$\underline{1.0156}$	1.0126	1.0147	1.0223	1.0180	1.0123	$\underline{1.0279}$	$\underline{1.0096}$
$\widehat{\sigma}_x(5)$	$\underline{1.0266}$	1.0191	1.0312	1.0056	1.0221	1.0224	1.0130	$\underline{1.0147}$	$\underline{1.0279}$
AIC	36.2	28.0	20.7	15.4	12.7	15.9	23.0	35.3	54.2
$W(s)$	69.6	59.4	48.5	36.0	24.7/20.7	32.4	47.8	62.2	78.6
$W(10)$	61.5	51.0	40.7	28.7	19.1/15.2	25.2	39.4	53.6	72.0
$W(5)$	50.1	40.4	30.6	21.1	13.1/9.6	17.3	29.1	42.8	62.2
$W(1)$	30.7	23.3	16.4	10.6	5.8/3.6	7.1	14.1	25.2	42.2

DGP 2: $\{x_t\}_{t=1}^{62}$ where $x_t = x_{t-12} + (1-L^{12})^{-d}\varepsilon_t$ and $\{\varepsilon_t\} \sim NID(0,1)$. $\sigma_x(h) = 1$ for $h = 1, 3, 5$.

TABLE 2 (b) The SRPMSE and the percentage of selection of $\widehat{x}_n(h)$ ($n = 100$)

d	-0.20	-0.15	-0.10	-0.05	0.00	0.05	0.10	0.15	0.20
c/ω	-2.57	-1.92	-1.28	-0.64	0.00	0.64	1.28	1.92	2.57
$\widetilde{\sigma}_x(1)$	1.0239	1.0271	1.0118	$\underline{1.0015}$	$\underline{1.0047}$	$\underline{1.0122}$	1.0228	1.0193	1.0482
$\widetilde{\sigma}_x(3)$	1.0201	1.0079	1.0100	$\underline{1.0025}$	$\underline{0.9913}$	$\underline{0.9968}$	$\underline{1.0129}$	1.0270	1.0747
$\widetilde{\sigma}_x(5)$	1.0080	1.0119	$\underline{1.0093}$	$\underline{1.0093}$	$\underline{0.9843}$	$\underline{0.9959}$	1.0210	1.0230	1.0414
$\widehat{\sigma}_x(1)$	$\underline{1.0064}$	$\underline{1.0191}$	$\underline{1.0117}$	1.0052	1.0126	1.0156	$\underline{1.0196}$	$\underline{1.0056}$	$\underline{1.0143}$
$\widehat{\sigma}_x(3)$	$\underline{1.0011}$	$\underline{1.0013}$	$\underline{1.0073}$	1.0072	1.0009	1.0019	1.0135	$\underline{1.0078}$	$\underline{1.0269}$
$\widehat{\sigma}_x(5)$	$\underline{0.9929}$	$\underline{1.0024}$	1.0115	1.0128	0.9913	1.0002	$\underline{1.0179}$	$\underline{1.0084}$	$\underline{1.0047}$
AIC	67.4	49.8	32.0	18.8	13.0	18.8	36.6	59.7	81.9
$W(s)$	88.6	77.3	60.5	40.4	22.7/17.5	35.1	59.2	79.4	92.8
$W(10)$	83.1	69.0	50.8	32.0	16.8/11.9	26.3	49.7	72.5	89.0
$W(5)$	74.1	57.7	38.8	22.4	10.6/6.6	17.3	38.1	61.7	83.1
$W(1)$	52.8	35.2	19.9	10.1	3.5/1.8	6.2	19.3	38.5	65.6

DGP 2: $\{x_t\}_{t=1}^{112}$ where $x_t = x_{t-12} + (1-L^{12})^{-d}\varepsilon_t$ and $\{\varepsilon_t\} \sim NID(0,1)$. $\sigma_x(h) = 1$ for $h = 1, 3, 5$.

TABLE 3 (a) The SRPMSE and the percentage of selection of $\widehat{x}_n(h)$ for DGP 3 ($\phi = 0.6$, $n = 100$)

d	-0.30	-0.20	-0.10	0.00	0.10	0.20	0.30
c/ω	-1.17	-0.78	-0.39	0.00	0.39	0.78	1.17
$\widetilde{\sigma}_x(1)$	1.0079	$\underline{1.0087}$	$\underline{0.9971}$	$\underline{0.9882}$	$\underline{1.0007}$	$\underline{1.0050}$	1.0033
$\widetilde{\sigma}_x(3)$	2.1602	$\underline{2.3402}$	$\underline{2.5204}$	$\underline{2.7211}$	$\underline{2.9391}$	$\underline{3.1784}$	3.4517
$\widetilde{\sigma}_x(5)$	2.9281	$\underline{3.3016}$	$\underline{3.6745}$	$\underline{4.1994}$	$\underline{4.6976}$	$\underline{5.3406}$	6.1197
$\widehat{\sigma}_x(1)$	$\underline{1.0056}$	1.0122	1.0002	0.9927	1.0058	1.0075	$\underline{0.9998}$
$\widehat{\sigma}_x(3)$	$\underline{2.1500}$	2.3509	2.5422	2.7492	2.9613	3.1820	$\underline{3.4235}$
$\widehat{\sigma}_x(5)$	$\underline{2.8931}$	3.3176	3.7285	4.2661	4.7459	5.3438	$\underline{6.0502}$
$\sigma_x(1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma_x(3)$	2.1402	2.3224	2.5160	2.7206	2.9361	3.1623	3.3990
$\sigma_x(5)$	2.8552	3.2480	3.6880	4.1776	4.7196	5.3169	5.9725
AIC	36.9	20.1	10.6	9.0	11.1	17.1	30.6
$W(s)$	65.4	45.7	31.2	19.3/14.8	25.9	37.7	55.1
$W(10)$	50.3	33.8	22.8	13.9/10.8	19.6	29.7	45.1
$W(5)$	33.1	22.9	16.1	7.4/9.6	14.2	21.7	34.8
$W(1)$	17.2	12.2	8.9	5.3/3.3	7.6	12.6	21.7

DGP 3: $\{x_t\}_{t=1}^{101}$ where $x_t = x_{t-1} + (1-L)^{-d}(1-\phi L)^{-1}\varepsilon_t$ and $\{\varepsilon_t\} \sim NID(0,1)$.

TABLE 3 (b) The SRPMSE and the percentage of selection of $\widehat{x}_n(h)$ for DGP 3 ($\phi = -0.8$, $n = 100$)

d	-0.30	-0.20	-0.10	0.00	0.10	0.20	0.30
c/ω	-3.61	-2.41	-1.20	0.00	1.20	2.41	3.61
$\widetilde{\sigma}_x(1)$	1.0609	1.0338	$\underline{1.0098}$	$\underline{1.0078}$	1.0132	1.0525	1.1331
$\widetilde{\sigma}_x(3)$	1.2637	1.2740	$\underline{1.2805}$	$\underline{1.3205}$	1.4165	1.5670	1.8353
$\widetilde{\sigma}_x(5)$	1.3755	1.4032	1.4530	$\underline{1.5628}$	1.7332	2.0132	2.5008
$\widehat{\sigma}_x(1)$	$\underline{1.0174}$	$\underline{1.0176}$	1.0100	1.0147	$\underline{1.0093}$	$\underline{1.0155}$	$\underline{1.0097}$
$\widehat{\sigma}_x(3)$	$\underline{1.2124}$	$\underline{1.2500}$	1.2821	1.3309	$\underline{1.4117}$	$\underline{1.4997}$	$\underline{1.5894}$
$\widehat{\sigma}_x(5)$	$\underline{1.3195}$	$\underline{1.3742}$	$\underline{1.4524}$	1.5754	$\underline{1.7270}$	$\underline{1.9152}$	$\underline{2.1286}$
$\sigma_x(1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma_x(3)$	1.2106	1.2322	1.2691	1.3212	1.3882	1.4697	1.5649
$\sigma_x(5)$	1.3094	1.3576	1.4366	1.5483	1.6940	1.8745	2.0902
AIC	95.9	76.1	36.3	12.9	34.7	75.2	95.6
$W(s)$	99.5	92.4	61.9	21.8/12.1	51.4	86.6	98.0
$W(10)$	98.8	87.0	51.3	15.4/7.5	40.2	80.8	97.6
$W(5)$	96.6	77.3	37.6	9.2/3.1	27.5	71.3	94.9
$W(1)$	86.1	53.1	16.9	3.2/0.5	10.2	47.8	85.9

DGP 3: $\{x_t\}_{t=1}^{101}$ where $x_t = x_{t-1} + (1-L)^{-d}(1-\phi L)^{-1}\varepsilon_t$ and $\{\varepsilon_t\} \sim NID(0,1)$.

6 Conclusion

This paper deals with the prediction theory of non-stationary long-memory processes, referred to as the ARFISMA model by Hassler (1994). After investigating the general theory relating to the convergence of the moments of the non-linear least squares estimators, we evaluate the asymptotic prediction mean squared error of two predictors. The first is defined by the estimator of the differencing parameter and the second by a fixed differencing parameter: in other words, a parametric predictor of the SARIMA model. The effects of misspecifying the integration order in the ARFISMA model are clarified by the asymptotic results relating to the prediction mean squared error. The finite sample behaviour of the predictor is investigated using simulation, and the source of differences in behaviour made clear in terms of asymptotic theory. The results also reveal that classical test statistics, like the likelihood ratio and Wald test statistics, can serve as model selection criteria in terms of prediction mean squared error.

References

Beran J. 1994. *Statistics for Long-Memory Processes.* Chapman and Hall: New York.

Bhansali RJ. 1981. Effects of not knowing the order of autoregressive processes on the mean squared error of prediction. *Journal of the American Statistical Association* **76**: 588-597.

Bhansali RJ, Papangelou F. 1991. Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order. *Annals of Statistics* **19**: 1155-1162.

Box GEP, Jenkins GM 1976. *Time Series Analysis: Forecasting and Control.* Holden-Day: San Francisco.

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd edn. Springer: New York.

Chung CF, Baillie RT 1993. Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. *Empirical Economics* **18**: 791-806.

Fuller WA. 1996. *Introduction to Statistical Time Series,* 2nd edn. John Wiley: New York.

Fuller WA, Hasza DP. 1981. Properties of predictors for autoregressive time series. *Journal of the American Statistical Association* **76**: 155-161.

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

Gourieroux C, Monfort A. 1989. *Statistique et Modeles Econometriques*, Volume II, Economica: Paris. [Translated by Vuong Q. 1995. *Statistics and Econometric Models*, Volume II. Cambridge University Press: New-York].

Hassler U. 1994. (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis* **15**: 19-30.

Hidalgo J, Yajima Y. 2002. Prediction and signal extraction of strongly dependent processes in the frequency domain. *Econometric Theory* **18**: 584-624.

Hosking JRM. 1981. Fractional differencing. *Biometrika* **68**: 165-176.

International Journal of Forecasting homepage. 2006. http://www.sciencedirect.com/science/journal/01692070 [1 June 2006]

Ing CK. 2001. A note on mean-squared prediction errors of the least squares predictors in random walk models. *Journal of Time Series Analysis* **22**: 711-724.

Katayama N. 2006. Seasonally and fractionally differenced time series. *Hitotsubashi Journal of Economics,* forthcoming.

Kunitomo N, Yamamoto T. 1985. Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association***80**: 941-950.

Mccullagh P. 1987. *Tensor Methods in Statistics.* Chapman and Hall: New York.

Miller JW. 1994. Forecasting with fractionally differenced time series models. *Ph.D. Dissertation, Southern Methodist University.*

Papangelou F. 1994. On a distributional bound arising in autoregressive model fitting. *Journal of Applied Probability* **31**: 401-408.

Serfling RJ. 1980. *Approximation Theorems of Mathematical Statistics.* John Wiley: New York.

Tanaka K. 1999. The nonstationary fractional unit root. *Econometric Theory* **15**: 549-582.

Tanaka K, Maekawa K. 1984. The sampling distribution of the predictor for an autoregressive model under misspecifications. *Journal of Econometrics* **25**: 327-351.

Yajima Y. 1985. On estimation of long memory time series models. *Australian Journal of Statistics* **27**: 303-320.

Yamamoto T. 1976. Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. *Applied Statistics* **25**: 123-127.

Yamamoto T. 1981. Predictions of multivariate autoregressive-moving average models. *Biometrika* **68**: 485-492.

Appendix A

Proof of Theorem 2.1:

Since $Q_n^{(2)}(\boldsymbol{\theta}^*) = \boldsymbol{I}_n(\boldsymbol{\theta}^0) + \boldsymbol{R}_n(\boldsymbol{\theta}^*)$ and $\boldsymbol{I}(\boldsymbol{\theta}^0)$ are positive definite, the minimum eigenvalues of $Q_n^{(2)}(\boldsymbol{\theta}^*)$ and $\boldsymbol{I}_n(\boldsymbol{\theta}^0)$ are positive for all $n \geq n_0$ by (2.6). It follows that:

$$\boldsymbol{v}'Q_n^{(2)}(\boldsymbol{\theta}^*)\boldsymbol{v} = \frac{1}{n}\sum_{j=0}^{n-1} Z_v^2(j) + \boldsymbol{v}'\boldsymbol{R}_n(\boldsymbol{\theta}^*)\boldsymbol{v}, \ \frac{1-\delta}{n(1-\delta)}\sum_{j=n\delta}^{n-1} Z_v^2(j), \text{ and } \frac{\delta}{n\delta}\sum_{j=0}^{n\delta-1} Z_v^2(j) + \boldsymbol{v}'\boldsymbol{R}_n(\boldsymbol{\theta}^*)\boldsymbol{v}$$

are positive for all $n \geq n_0$ a.c., where $\boldsymbol{v} = (v_1, \ldots, v_p)'$ is a fixed vector such that $\|\boldsymbol{v}\| = 1$, $Z_v(j) = \boldsymbol{v}'e_{j+1}^{(1)}(\boldsymbol{\theta}^0) = \sum_{k=1}^{j} \alpha_v(j,k)y_k$, $\alpha_v(j,k) = -\boldsymbol{v}'a_{j+1-k}^{(1)}(\boldsymbol{\theta}^0)$, $0 < \delta < 1$ and without loss of generality, $n\delta$ is assumed to be a positive integer, and $\sigma_0^2 = 1$. Similarly to equations (3.3) and (3.4) of Ing (2001), rearranging $n^{-1}\sum_{j=n\delta}^{n-1} Z_v^2(j)$, one obtains:

$$\boldsymbol{v}'Q_n^{(2)}(\boldsymbol{\theta}^*)\boldsymbol{v} \geq \frac{1}{n}\sum_{j=n\delta}^{n-1} Z_v^2(j) = \frac{1}{n}\sum_{j=0}^{(1-\delta)n/(lq)-1}\sum_{i=0}^{lq-1} Z_v^2(n(i)+j) > 0 \quad \text{a.c.},$$

where $n(i) = n\delta + (1-\delta)ni/(lq)$, $l > 4(p-1)/q + 2$ and, to simplify the discussion, lq and $(1-\delta)n/(lq)$ are also assumed to be positive integers and $n\delta, (1-\delta)n/(lq) \geq m$. And by the convexity of function x^{-q}, $x > 0$:

$$\left\{\boldsymbol{v}'Q_n^{(2)}(\boldsymbol{\theta}^*)\boldsymbol{v}\right\}^{-q} \leq \left(\frac{1-\delta}{lq}\right)^{-q}\frac{lq}{(1-\delta)n}\sum_{j=0}^{(1-\delta)n/(lq)-1}\left\{\sum_{i=0}^{lq-1} Z_v^2(n(i)+j)\right\}^{-q} \quad \text{a.c..}$$

Let λ_0 be the smallest eigenvalue of $Q_n^{(2)}(\boldsymbol{\theta}^*)$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)'$ be on the unit sphere S of $\mathbb{R}^p$, and $Z_\nu(j) = \boldsymbol{\nu}'e_{j+1}^{(1)}(\boldsymbol{\theta}^0)$. Then $\|\{Q_n^{(2)}(\boldsymbol{\theta}^*)\}^{-1}\|_S = \lambda_0^{-1}$ and $\lambda_0 = \inf_{\|\boldsymbol{\nu}\|=1}\boldsymbol{\nu}'Q_n^{(2)}(\boldsymbol{\theta}^*)\boldsymbol{\nu}$. Therefore, if one can show that there exists some $C > 0$ such that:

$$\mathrm{E}\left[\inf_{\|\boldsymbol{\nu}\|=1}\sum_{i=0}^{lq-1} Z_\nu^2(n(i)+j)\right]^{-q} \leq C < \infty \tag{A.1}$$

holds for all $j = 0, \ldots, (1-\delta)n/(lq) - 1$ and $n \geq n_0$, then (2.9) follows. In this proof we focus on the case of $j = 0$ because the same argument is easily applied to other values of j.

First, we show that:

$$\Pr\left(Q(\boldsymbol{v}) < \beta\right) \leq \text{const}\,\beta^{lq/2}, \tag{A.2}$$

for fixed $\boldsymbol{v}$ such that $\|\boldsymbol{v}\| = 1$ and $\beta > 0$, where $Q(\boldsymbol{v}) = \sum_{i=0}^{lq-1} Z_v^2(n(i))$. Note that, by (b) of Assumption 1, there exists $k(i)$ $(i = 0, 1, \ldots, lq-1)$ such that $k(i) = \max[\,k \,|\, |\alpha_v(n(i),k)| > 0,\ k = n(i-1)+1, \ldots, n(i)\,]$ for $i = 1, \ldots, lq-1$ and $k(0) = \max[\,k \,|\, |\alpha_v(n(0),k)| > 0,\ k = 1, \ldots, n(0)\,]$ because $(1-\delta)n/(lq) \geq m$ and $n(0) = n\delta \geq m$. To prove (A.2), we use the same argument as in the proof of (2.2) of Bhansali and Papangelou (1994, pp.1158–1159), (d) of Assumption 1, and (i) of Lemma 2.1. Let $Y(u)$, $u = 1, 2, \ldots, k(lq-1)$ be defined by:

$$Y(u) = \begin{cases} \sum_{t=1}^{k(i)} \alpha_v(n(i),t)y_t & u = k(i),\ i = 0, 1, \ldots, lq-1, \\ y_u & \text{otherwise.} \end{cases}$$

Then $Y(k(i)) = Z_v(n(i)) = \sum_{t=1}^{k(i)} \alpha_v(n(i),t)y_t$, $i = 0,1,\ldots,lq-1$, and $\{Y(u)\}_{u=1}^{k(lq-1)}$ and $\{y_t\}_{t=1}^{k(lq-1)}$ satisfy conditions of (i) of Lemma 2.1. It follows that:

$$\begin{aligned}
\Pr\left(\sum_{i=0}^{lq-1} Z_v^2(n(i)) < \beta\right) &\leq \Pr\left(\bigcap_{i=0}^{lq-1}\left\{Z_v^2(n(i)) < \beta\right\}\right) = \Pr\left(\bigcap_{i=0}^{lq-1}\left\{-\sqrt{\beta} < Y(k(i)) < \sqrt{\beta}\right\}\right) \\
&\leq \frac{2K\sqrt{\beta}}{|\alpha_v(n(lq-1),k(lq-1))|}\Pr\left(\bigcap_{i=0}^{lq-2}\left\{-\sqrt{\beta} < Y(k(i)) < \sqrt{\beta}\right\}\right) \leq \text{const}\,\beta^{lq/2},
\end{aligned}$$

which proves (A.2).

To demonstrate (A.1) one can follow Bhansali and Papangelou (1991, p.1159). Let $\boldsymbol{v}$ and $\boldsymbol{\nu}$ be on the unit sphere S of $\mathbb{R}^p$, and $Q(\boldsymbol{\nu}) = \sum_{i=0}^{lq-1} Z_\nu^2(n(i))$. Then, using the triangular inequality and the Cauchy–Schwarz inequality, we have:

$$|Z_v(j)| = \left|(\boldsymbol{\nu}+\boldsymbol{v}-\boldsymbol{\nu})' e_{j+1}^{(1)}(\boldsymbol{\theta}^0)\right| \leq |Z_\nu(j)| + \|\boldsymbol{v}-\boldsymbol{\nu}\|\left\|e_{j+1}^{(1)}(\boldsymbol{\theta}^0)\right\|.$$

It implies that if $\|\boldsymbol{v}-\boldsymbol{\nu}\| < \varepsilon$ and $\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\|^2 < 1/\varepsilon$, $i = 0,1,\ldots,lq-1$, then:

$$Q(\boldsymbol{v}) = \sum_{i=0}^{lq-1} Z_v^2(n(i)) \leq 2\sum_{i=0}^{lq-1} Z_\nu^2(n(i)) + 2\|\boldsymbol{v}-\boldsymbol{\nu}\|^2 \sum_{i=0}^{lq-1}\left\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\right\|^2 \leq 2Q(\boldsymbol{\nu}) + 2lq\varepsilon. \tag{A.3}$$

It is easy to see that, given $\varepsilon \in (0,1)$, there is a subset of S' of S, with $(\text{const}/\varepsilon)^{2p-2}$ elements, such that given any $\boldsymbol{\nu} \in S$ there exists $\boldsymbol{v} \in S'$ with $\|\boldsymbol{v}-\boldsymbol{\nu}\| < \varepsilon$. From this, (A.2), and (A.3), we deduce that:

$$\begin{aligned}
&\Pr\left(Q(\boldsymbol{\nu}) < \varepsilon \text{ for some } \boldsymbol{\nu} \in S\right) \qquad\qquad \text{(A.4)}\\
&= \Pr\left(\{Q(\boldsymbol{\nu}) < \varepsilon \text{ for some } \boldsymbol{\nu} \in S\}\bigcap\left\{\left\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\right\|^2 \geq 1/\varepsilon \text{ for some } i = 0,1,\ldots,lq-1\right\}\right) \\
&\quad + \Pr\left(\{Q(\boldsymbol{\nu}) < \varepsilon \text{ for some } \boldsymbol{\nu} \in S\}\bigcap\left\{\left\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\right\|^2 < 1/\varepsilon \text{ for any } i = 0,1,\ldots,lq-1\right\}\right) \\
&\leq \Pr\left(\left\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\right\|^2 \geq 1/\varepsilon \text{ for some } i = 0,1,\ldots,lq-1\right) \\
&\quad + \Pr\left(\{Q(\boldsymbol{\nu}) < \varepsilon \text{ for some } \boldsymbol{\nu} \in S\}\bigcap\{Q(\boldsymbol{v}) \leq 2Q(\boldsymbol{\nu}) + 2lq\varepsilon \text{ for some } \boldsymbol{v} \in S'\}\right) \\
&\leq \sum_{i=0}^{lq-1}\Pr\left(\left\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\right\|^2 \geq 1/\varepsilon\right) + \Pr\left(Q(\boldsymbol{v}) \leq 2\varepsilon + 2lq\varepsilon \text{ for some } \boldsymbol{v} \in S'\right) \\
&\leq \sum_{i=0}^{lq-1}\varepsilon^{lq/2-2p+2}\,\mathrm{E}\left\|e_{n(i)+1}^{(1)}(\boldsymbol{\theta}^0)\right\|^{lq-4p+4} + \left(\frac{\text{const}}{\varepsilon}\right)^{2p-2}(2\varepsilon + 2lq\varepsilon)^{lq/2} \\
&\leq \text{const}\,\varepsilon^{lq/2-2p+2},
\end{aligned}$$

where the last inequality follows from (c) of Assumption 1. By (A.4), $\Pr\left(\inf_{\|\boldsymbol{\nu}\|=1} Q(\boldsymbol{\nu}) < \varepsilon\right) \leq \text{const}\,\varepsilon^{lq/2-2p+2}$, hence:

$$\begin{aligned}
\mathrm{E}\left[\inf_{\|\boldsymbol{\nu}\|=1} Q(\boldsymbol{\nu})\right]^{-q} &= \int_0^\infty \Pr\left(\left\{\inf_{\|\boldsymbol{\nu}\|=1} Q(\boldsymbol{\nu})\right\}^{-q} > t\right)dt \\
&\leq 1 + \int_1^\infty \Pr\left(\inf_{\|\boldsymbol{\nu}\|=1} Q(\boldsymbol{\nu}) < t^{-1/q}\right)dt \\
&\leq 1 + \text{const}\int_1^\infty \left(t^{-1/q}\right)^{lq/2-2p+2} dt,
\end{aligned}$$

which is finite since $l > (4p-4)/q + 2$ and hence demonstrates (A.1) for the case of $j = 0$.

To obtain (2.8) and (2.11), it is sufficient to check the uniform integrability of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)$ (see Serfling, 1980, pp.13–14). By Hölder's inequality, we have, for $n \geq n_0$:

$$\begin{aligned}\mathrm{E}\left\|\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\right\|^{r_0} &\leq \text{const}\,\mathrm{E}\left\|\left\{Q_n^{(2)}(\boldsymbol{\theta}^*)\right\}^{-1}\right\|_S^{r_0}\left\|\sqrt{n}\,Q_n^{(1)}(\boldsymbol{\theta}^0)\right\|^{r_0} \\ &\leq \text{const}\left\{\mathrm{E}\left\|\left\{Q_n^{(2)}(\boldsymbol{\theta}^*)\right\}^{-1}\right\|_S^{r_0 u}\right\}^{1/u}\left\{\mathrm{E}\left\|\sqrt{n}\,Q_n^{(1)}(\boldsymbol{\theta}^0)\right\|^{r_0 v}\right\}^{1/v}.\end{aligned}$$

Using (2.9), (2.10), and Lyapunov's inequality, this uniform integrability is satisfied and the conclusion follows.

Proof of Corollary 2.1:

We borrow the notations in the proof of Theorem 2.1. Since $Q_n^{(2)}(\boldsymbol{\theta}^*) = \boldsymbol{I}_n(\boldsymbol{\theta}^0) + \boldsymbol{R}_n(\boldsymbol{\theta}^*)$ and $\boldsymbol{I}(\boldsymbol{\theta}^0)$ are positive definite, the minimum eigenvalue of $Q_n^{(2)}(\boldsymbol{\theta}^*)$ and $\boldsymbol{I}_n(\boldsymbol{\theta}^0)$ are positive for all $n \geq n_0$ by (2.6). From this and (b) of Assumption 2, $\boldsymbol{v}'Q_n^{(2)}(\boldsymbol{\theta}^*)\boldsymbol{v} = n^{-1}\sum_{j=0}^{n-1} Z_v^2(j) + \boldsymbol{v}'\boldsymbol{R}_n(\boldsymbol{\theta}^*)\boldsymbol{v}$, $(1-\delta)\{n(1-\delta)\}^{-1}\sum_{j=n\delta}^{n-1} Z_v^2(j)$, and $\delta(n\delta)^{-1}\sum_{j=0}^{n\delta-1} Z_v^2(j) + \boldsymbol{v}'\boldsymbol{R}_n(\boldsymbol{\theta}^*)\boldsymbol{v}$ are all positive for all $n \geq n_0$ a.c., where $\boldsymbol{v}$ is a fixed vector such that $\|\boldsymbol{v}\| = 1$, $Z_v(j) = \boldsymbol{v}'e_{j+1}^{(1)}(\boldsymbol{\theta}^0) = \boldsymbol{v}'\sum_{k=1}^{j}\boldsymbol{d}_{j+1-k}(\boldsymbol{\theta}^0)\varepsilon_k = \sum_{k=1}^{j}\alpha_v(j,k)\varepsilon_k$, $e_j^{(1)}(\boldsymbol{\theta}^0) = \sum_{k=1}^{j-1}\boldsymbol{d}_k(\boldsymbol{\theta}^0)\varepsilon_{j-k} = \sum_{k=1}^{j-1}\boldsymbol{d}_{j-k}(\boldsymbol{\theta}^0)\varepsilon_k$, $\alpha_v(j,k) = \boldsymbol{v}'\boldsymbol{d}_{j+1-k}(\boldsymbol{\theta}^0)$, $0 < \delta < 1$ and without loss of generality, $n\delta$ is assumed to be a positive integer, and $\sigma_0^2 = 1$. Let $Z_\nu(j) = \boldsymbol{\nu}'e_{j+1}^{(1)}(\boldsymbol{\theta}^0)$. Similarly to (A.1) in the proof of Theorem 2.1, by equations (3.3) and (3.4) of Ing (2001), if one can show that there exists some $C > 0$ such that:

$$\mathrm{E}\left[\inf_{\|\boldsymbol{\nu}\|=1}\sum_{i=0}^{lq-1} Z_\nu^2(n(i)+j)\right]^{-q} \leq C < \infty \tag{A.5}$$

holds for all $j = 0, \ldots, (1-\delta)n/(lq) - 1$ and $n \geq n_0$, then (2.9) follows. In this proof we focus on the case of $j = 0$, because the same argument is easily applied to other values of j.

First, we show that:

$$\Pr\left(Q(\boldsymbol{v}) < \beta\right) \leq \text{const}\,\beta^{lq/2}, \tag{A.6}$$

for fixed $\boldsymbol{v}$ such that $\|\boldsymbol{v}\| = 1$ and $\beta > 0$, where $Q(\boldsymbol{v}) = \sum_{i=0}^{lq-1} Z_v^2(n(i))$. By (a) and (b) of Assumption 2, we have for large $n > n_0$:

$$\begin{aligned}&\sum_{k=n(i-1)+1}^{n(i)}\left\{\alpha_v(n(i),k)\right\}^2 = \boldsymbol{v}'\sum_{k=1}^{(1-\delta)n/(lq)}\boldsymbol{d}_k(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'\boldsymbol{v} > 0, \\ \text{and}\quad &\sum_{k=1}^{n(0)}\left\{\alpha_v(n(0),k)\right\}^2 = \boldsymbol{v}'\sum_{k=1}^{n\delta}\boldsymbol{d}_k(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'\boldsymbol{v} > 0.\end{aligned}$$

It follows that these have at least one non-zero summand since $\boldsymbol{I}(\boldsymbol{\theta}^0) = \sum_{k=1}^{\infty}\boldsymbol{d}_k(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'$ is positive definite. These prove the existence of $k(i)$ $(i = 0, 1, \ldots, lq-1)$ such that $k(i) = \max[\,k\,|\,|\alpha_v(n(i),k)| > 0,\ k = n(i-1)+1, \ldots, n(i)\,]$ for $i = 1, \ldots, lq-1$ and $k(0) = \max[\,k\,|\,|\alpha_v(n(0),k)| > 0,\ k = 1, \ldots, n(0)\,]$. To prove (A.6), we use the same argument as in the proof of (2.2) of Bhansali and Papangelou (1994, pp.1158–1159), (c) of Assumption 2, and (ii) of Lemma 2.1. Let $Y(u)$, $u = 1, 2, \ldots, k(lq-1)$ be defined by:

$$Y(u) = \begin{cases}\sum_{t=1}^{k(i)}\alpha_v(n(i),t)\varepsilon_t & u = k(i),\ i = 0, 1, \ldots, lq-1, \\ \varepsilon_u & \text{otherwise.}\end{cases}$$

Then $Y(k(i)) = Z_v(n(i)) = \sum_{t=1}^{k(i)}\alpha_v(n(i),t)\varepsilon_t$, $i = 0, 1, \ldots, lq-1$, and $\{Y(u)\}_{u=1}^{k(lq-1)}$ and

$\{\varepsilon_t\}_{t=1}^{k(lq-1)}$ satisfy conditions of (ii) of Lemma 2.1. It follows that:

$$\Pr\left(\sum_{i=0}^{lq-1} Z_v^2(n(i)) < \beta\right) \le \Pr\left(\bigcap_{i=0}^{lq-1} \{Z_v^2(n(i)) < \beta\}\right) = \Pr\left(\bigcap_{i=0}^{lq-1} \left\{-\sqrt{\beta} < Y(k(i)) < \sqrt{\beta}\right\}\right)$$
$$\le \frac{2K\sqrt{\beta}}{|\alpha_v(n(lq-1), k(lq-1))|} \Pr\left(\bigcap_{i=0}^{lq-2} \left\{-\sqrt{\beta} < Y(k(i)) < \sqrt{\beta}\right\}\right) \le \text{const}\, \beta^{lq/2},$$

which proves (A.6). The rest of the proof of (A.5) is obvious from the same argument as in the proof of Theorem 2.1. To obtain (2.8) and (2.11), it is sufficient to check uniform integrability of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)$ (see Serfling, 1980, pp.13–14). We note that $\|e_t^{(1)}(\boldsymbol{\theta}^0)\|$ and $\|\sqrt{n}Q_n^{(1)}(\boldsymbol{\theta}^0)\|$ have finite moments of all orders from (b) of Assumption 2 because, under the model (2.12), $e_t(\boldsymbol{\theta}^0) = \varepsilon_t$ for $t \ge 1$ and $\{e_t(\boldsymbol{\theta}^0)e_t^{(1)}(\boldsymbol{\theta}^0)\}$ is a sequence of martingale differences. Using this and the same argument as in the proof of Theorem 2.1, the uniform integrability is satisfied, which proves (2.8) and (2.11).

Proof of Lemma 2.1:

(i) Let the joint probability distributions of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be $f_{1,\ldots,n}$ and $g_{1,\ldots,n}$, respectively, $\boldsymbol{x} = (X_1, X_2, \ldots, X_n)'$, $\boldsymbol{y} = (Y_1, Y_2, \ldots, Y_n)'$, $\boldsymbol{A}$ is an $n \times n$ upper triangular matrix such that (i,j)th element is a_{ij} and $a_{ij} = 0$, $i > j$, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, and $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y} = (v_1(Y_1, \ldots, Y_n), v_2(Y_2, \ldots, Y_n), \ldots, v_n(Y_n))'$. Then $\det(A) = \prod_{i=1}^n a_{ii}$ and $g_{1,\ldots,n}(y_1, \ldots, y_n) = f_{1,\ldots,n}(v_1(y_1, \ldots, y_n), v_2(y_2, \ldots, y_n), \ldots, v_n(y_n))|\prod_{i=1}^n a_{ii}|^{-1}$. Therefore, let the joint probability distributions of $(X_2, \ldots, X_n)$ and $(Y_2, \ldots, Y_n)$ be $f_{2,\ldots,n}$ and $g_{2,\ldots,n}$, we have:

$$\begin{aligned}
&\Pr(Y_1 \in C_1, Y_{l_1} \in C_{l_1}, Y_{l_2} \in C_{l_2}, \ldots, Y_{l_k} \in C_{l_k}) \\
&= \int_{C_1}\int_{C_{l_1}} \cdots \int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}} g_{1,\ldots,n}(y_1, \ldots, y_n) dy_n \cdots dy_{l_k} \cdots dy_{l_1} dy_1 \\
&= \int_{C_1}\int_{C_{l_1}} \cdots \int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}} \frac{f_{1,\ldots,n}(v_1(y_1, \ldots, y_n), v_2(y_2, \ldots, y_n), \ldots, v_n(y_n))}{f_{2,\ldots,n}(v_2(y_2, \ldots, y_n), \ldots, v_n(y_n))} \left|\prod_{i=1}^n a_{ii}\right|^{-1} \\
&\quad \times f_{2,\ldots,n}(v_2(y_2, \ldots, y_n), \ldots, v_n(y_n)) dy_n \cdots dy_{lk} \cdots dy_{l_1} dy_1 \\
&\le K|a_{11}|^{-1}(c_1' - c_1) \int_{C_{l_1}} \cdots \int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}} f_{2,\ldots,n}(v_2(y_2, \ldots, y_n), \ldots, v_n(y_n)) \left|\prod_{i=2}^n a_{ii}\right|^{-1} \\
&\quad dy_n \cdots dy_{l_k} \cdots dy_{l_1} \\
&= K|a_{11}|^{-1}(c_1' - c_1) \int_{C_{l_1}} \cdots \int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}} g_{2,\ldots,n}(y_2, \ldots, y_n) dy_n \cdots dy_{l_k} \cdots dy_{l_1} \\
&= K|a_{11}|^{-1}(c_1' - c_1) \Pr(Y_{l_1} \in C_{l_1}, Y_{l_2} \in C_{l_2}, \ldots, Y_{l_k} \in C_{l_k}).
\end{aligned}$$

We note that the equalities follow from the fact that:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{pmatrix}, \quad \boldsymbol{A}^{-1} = \begin{pmatrix} a_{11}^{-1} & -a_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{A}_{22}^{-1} \end{pmatrix},$$

and $(X_2, \ldots, X_n)' = \boldsymbol{A}_{22}^{-1}(Y_2, \ldots, Y_n)' = (v_2(Y_2, \ldots, Y_n), \ldots, v_n(Y_n))'$, where $\boldsymbol{A}_{12}$ is $1 \times (n-1)$ matrix and $\boldsymbol{A}_{22}$ is $(n-1) \times (n-1)$ matrix.

(ii) Similarly to the proof of (i), let the joint probability distributions of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be $f_{1,\ldots,n}$ and $g_{1,\ldots,n}$, respectively. Then $g_{1,\ldots,n}(y_1, \ldots, y_n) = f_{1,\ldots,n}(v_1(y_1, \ldots, y_n), v_2(y_2, \ldots, y_n), \ldots, v_n(y_n))|\prod_{i=1}^n a_{ii}|^{-1} =$

$f_1(v_1(y_1,\ldots,y_n))f_1(v_2(y_2,\ldots,y_n))\cdots f_1(v_n(y_n))|\prod_{i=1}^n a_{ii}|^{-1}$. Therefore:

$$
\begin{aligned}
&\Pr(Y_1 \in C_1, Y_{l_1} \in C_{l_1}, Y_{l_2} \in C_{l_2}, \ldots, Y_{l_k} \in C_{l_k}) \\
&= \int_{C_1}\int_{C_{l_1}}\cdots\int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}} g_{1,\ldots,n}(y_1,\ldots,y_n)dy_n\cdots dy_{lk}\cdots dy_{l_1}dy_1 \\
&= \int_{C_1}\int_{C_{l_1}}\cdots\int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}} f_1(v_1(y_1,\ldots,y_n))f_1(v_2(y_2,\ldots,y_n))\cdots f_1(v_n(y_n))\left|\prod_{i=1}^n a_{ii}\right|^{-1} \\
&\qquad\qquad dy_n\cdots dy_{l_k}\cdots dy_{l_1}dy_1 \\
&= \int_{C_{l_1}}\cdots\int_{C_{l_k}}\int_{\mathbb{R}^{n-k-1}}\left\{\int_{C_1} f_1(v_1(y_1,\ldots,y_n))dy_1\right\}f_1(v_2(y_2,\ldots,y_n))\cdots f_1(v_n(y_n))\left|\prod_{i=1}^n a_{ii}\right|^{-1} \\
&\qquad\qquad dy_n\cdots dy_{lk}\cdots dy_{l_1} \\
&\le K|a_{11}|^{-1}(c_1'-c_1)\int_{C_{l_1}}\cdots\int_{C_{lk}}\int_{\mathbb{R}^{n-k-1}} f_1(v_2(y_2,\ldots,y_n))\cdots f_1(v_n(y_n))\left|\prod_{i=2}^n a_{ii}\right|^{-1} \\
&\qquad\qquad dy_n\cdots dy_{l_k}\cdots dy_{l_1} \\
&= K|a_{11}|^{-1}(c_1'-c_1)\Pr(Y_{l_1} \in C_{l_1}, Y_{l_2} \in C_{l_2}, \ldots, Y_{l_k} \in C_{l_k}),
\end{aligned}
$$

where the last equality follows from $f_1(v_2(y_2,\ldots,y_n))\cdots f_1(v_n(y_n))|\prod_{i=2}^n a_{ii}|^{-1} = f_{2,\ldots,n}(v_2(y_2,\ldots,y_n),\ldots,v_n(y_n))|\prod_{i=2}^n a_{ii}|^{-1} = g_{2,\ldots,n}(y_2,\ldots,y_n)$.

Appendix B

To prove the asymptotic results in Section 3, let $S_\varepsilon(\boldsymbol{\delta}) = \sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\delta})/(2n\sigma_0^2)$, $S_u(\boldsymbol{\delta}) = \sum_{t=1}^n u_t^2(\boldsymbol{\delta})/(2n\sigma_0^2)$, $u_t(\boldsymbol{\delta}) = \sum_{k=0}^\infty \pi_k(\boldsymbol{\delta})v_{t-k}$, $v_t = \sum_{k=0}^\infty \psi_k(\boldsymbol{\delta}^0)\varepsilon_{t-k}$, and $\boldsymbol{I}(\boldsymbol{\delta}) = \mathbf{E}[u_t(\boldsymbol{\delta})u_t^{(2)}(\boldsymbol{\delta}) + u_t^{(1)}(\boldsymbol{\delta})u_t^{(1)}(\boldsymbol{\delta})']$, where $\{\varepsilon_t\}_{t=-\infty}^\infty$ is iid $(0,\sigma_0^2)$ and $\mathrm{E}[\varepsilon_t]^r < \infty$ for all positive integers r. Then $\widehat{\boldsymbol{\delta}}$ minimizes $S_\varepsilon(\boldsymbol{\delta})$, $\boldsymbol{I}(\boldsymbol{\delta})$ is continuous on $\boldsymbol{\delta}$, and $\boldsymbol{I}(\boldsymbol{\delta}^0) = \mathrm{E}[u_t^{(1)}(\boldsymbol{\delta}^0)u_t^{(1)}(\boldsymbol{\delta}^0)']$. Furthermore, by Lemmas B 7, B 8, and B 9 of Katayama (2006):

$$
\begin{aligned}
&\sup_{\boldsymbol{\delta}\in D_\delta}\left|\frac{1}{n}\sum_{t=1}^n \varepsilon_t^{(i)}(\boldsymbol{\delta})\varepsilon_t^{(j)}(\boldsymbol{\delta})' - \frac{1}{n}\sum_{t=1}^n u_t^{(i)}(\boldsymbol{\delta})u_t^{(j)}(\boldsymbol{\delta})'\right| \xrightarrow{a.c.} \mathbf{0}, \qquad (\mathrm{A.7})\\
&\sup_{\boldsymbol{\delta}\in D_\delta}\left|\frac{1}{n}\sum_{t=1}^n u_t^{(i)}(\boldsymbol{\delta})u_t^{(j)}(\boldsymbol{\delta})' - \mathrm{E}\left[u_t^{(i)}(\boldsymbol{\delta})u_t^{(j)}(\boldsymbol{\delta})'\right]\right| \xrightarrow{a.c.} \mathbf{0},
\end{aligned}
$$

as $n \to \infty$, where $(i,j) = (1,0)$, $(1,1)$, and $(2,0)$.

Proof of (3.11)**:**

We prove (3.11) by using Corollary 2.1. First, as we noted in Section 2, the conditions of (2.3a) and (2.3b) can be replaced by the strong consistency of CSS (NLS) estimators $\widehat{\boldsymbol{\delta}}$, (3.8) and (3.9) hold, and $\boldsymbol{I}(\boldsymbol{\delta}^0)$ is positive definite. On the conditions of (2.5) and (2.6), let:

$$
\widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0) = \frac{1}{n\sigma_0^2}\sum_{t=1}^n \varepsilon_t^{(1)}(\boldsymbol{\delta}^0)\varepsilon_t^{(1)}(\boldsymbol{\delta}^0)', \quad \text{and} \quad \widehat{\boldsymbol{R}}_n(\boldsymbol{\delta}^*) = S_\varepsilon^{(2)}(\boldsymbol{\delta}^*) - \widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0), \qquad (\mathrm{A.8})
$$

where $\|\boldsymbol{\delta}^* - \boldsymbol{\delta}^0\| \le \|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\|$. Then, by (A.7), we obtain:

$$\sup_{\boldsymbol{\delta} \in D_\delta} \left| S_\varepsilon^{(2)}(\boldsymbol{\delta}) - \boldsymbol{I}(\boldsymbol{\delta}) \right| \le \sup_{\boldsymbol{\delta} \in D_\delta} \left| S_\varepsilon^{(2)}(\boldsymbol{\delta}) - S_u^{(2)}(\boldsymbol{\delta}) \right| + \sup_{\boldsymbol{\delta} \in D_\delta} \left| S_u^{(2)}(\boldsymbol{\delta}) - \boldsymbol{I}(\boldsymbol{\delta}) \right| \xrightarrow{a.c.} \mathbf{0}, \tag{A.9}$$

$$\begin{aligned} \left| \widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0) - \boldsymbol{I}(\boldsymbol{\delta}^0) \right| &\le \sup_{\boldsymbol{\delta} \in D_\delta} \left| \frac{1}{n\sigma_0^2} \sum_{t=1}^{n} \varepsilon_t^{(1)}(\boldsymbol{\delta}) \varepsilon_t^{(1)}(\boldsymbol{\delta})' - \frac{1}{n\sigma_0^2} \sum_{t=1}^{n} u_t^{(1)}(\boldsymbol{\delta}) u_t^{(1)}(\boldsymbol{\delta})' \right| \\ &\quad + \sup_{\boldsymbol{\delta} \in D_\delta} \left| \frac{1}{n\sigma_0^2} \sum_{t=1}^{n} u_t^{(1)}(\boldsymbol{\delta}) u_t^{(1)}(\boldsymbol{\delta})' - \boldsymbol{I}(\boldsymbol{\delta}) \right| \xrightarrow{a.c.} \mathbf{0}, \end{aligned}$$

$$\left| S_\varepsilon^{(2)}(\boldsymbol{\delta}^*) - \boldsymbol{I}(\boldsymbol{\delta}^0) \right| \xrightarrow{a.c.} \mathbf{0}, \quad \text{and} \quad \left| \widehat{\boldsymbol{R}}_n(\boldsymbol{\delta}^*) \right| = \left| S_\varepsilon^{(2)}(\boldsymbol{\delta}^*) - \widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0) \right| \xrightarrow{a.c.} \mathbf{0},$$

as $n \to \infty$. It follows that conditions of (2.5) and (2.6) are satisfied. The conditions of (2.13) and (2.14) are satisfied by (3.9) and $\widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0) \xrightarrow{a.c.} \boldsymbol{I}(\boldsymbol{\delta}^0)$. Finally, (c) of Assumption 2 is satisfied by (a) of Assumption 3. Hence all conditions of Corollary 2.1 are satisfied, which proves (3.11).

Proof of (3.12)**:**

From the RHS of (3.12):

$$\sum_{j=h}^{n+h-1} \psi_j(\widehat{\boldsymbol{\delta}}) \widehat{\varepsilon}_{n+h-j} = \sum_{j=1}^{n} \psi_{j+h-1}(\widehat{\boldsymbol{\delta}}) \widehat{\varepsilon}_{n+1-j} = \sum_{j=1}^{n} \left\{ \sum_{i=0}^{j-1} \psi_{j-i+h-1}(\widehat{\boldsymbol{\delta}}) \pi_i(\widehat{\boldsymbol{\delta}}) \right\} y_{n+1-j}, \tag{A.10}$$

where the last equation follows from $\sum_{j=1}^{n} \sum_{i=0}^{n-j} a_{i,j} = \sum_{j=1}^{n} \sum_{i=0}^{j-1} a_{i,j-i}$. Since $\sum_{i=0}^{j} \psi_{j-i}(\boldsymbol{\delta}) \pi_i(\boldsymbol{\delta}) = 0$ for $j > 0$, we have $\sum_{i=0}^{j} \psi_{j-i}(\widehat{\boldsymbol{\delta}}) \pi_i(\widehat{\boldsymbol{\delta}}) = 0$ for $j > 0$. Hence, putting $\widehat{\pi}_i = \pi_i(\widehat{\boldsymbol{\delta}})$ and $\widehat{\psi}_i = \psi_i(\widehat{\boldsymbol{\delta}})$:

$$\sum_{i=0}^{j-1} \widehat{\psi}_{j-i+h-1} \widehat{\pi}_i = \sum_{i=0}^{j+h-1} \widehat{\psi}_{j-i+h-1} \widehat{\pi}_i - \sum_{i=j}^{j+h-1} \widehat{\psi}_{j-i+h-1} \widehat{\pi}_i = -\sum_{i=0}^{h-1} \widehat{\psi}_{h-i-1} \widehat{\pi}_{i+j} = -\sum_{i=0}^{h-1} \widehat{\psi}_i \widehat{\pi}_{j+h-i-1}.$$

Substituting the RHS of the above equation for the RHS of (A.10), we obtain the result.

Proof of Theorem 3.1:

First, we will show that $\mathrm{E}[R_{1,n}]^2 = o(n^{-1})$ uniformly in $\boldsymbol{\delta}^* \in D_l \times D_\beta$, $l = 1, 2, 3$. Let $\beta(z) = 1$, $\boldsymbol{\delta}^* = d^*$, and $d, \widehat{d} \in D_3$. Then, by the Cauchy–Schwarz inequality, we have:

$$\mathrm{E}[R_{1,n}]^2 \le \left[\mathrm{E} \left| \widehat{d} - d \right|^8 \mathrm{E} \left(\sum_{j=1}^{n} c_j^{(2)}(h, d^*) y_{n+1-j} \right)^4 \right]^{1/2}.$$

By Yajima (1985, (7)):

$$(n+1)^{a-1} \le \frac{\Gamma(n+a)}{\Gamma(n+1)} \le n^{a-1}, \quad \text{for } 0 \le a \le 1, \text{ and } n = 1, 2, \ldots,$$

and Lemma 3.2, there exists $a_j(\alpha)$, which does not depend on d^* and satisfies $\sup_{d^* \in D_3} |c_j^{(2)}(h, d^*)| \le a_j(\alpha)$, for $j \ge 1$ and $a_j(\alpha) = O(\{\log j\}^2 j^{-1/2-\alpha})$. It follows by the Cauchy–Schwarz inequality that:

$$\mathrm{E} \left(\sum_{j=1}^{n} \frac{\partial^2 c_j(h, d^*)}{\partial d^2} y_{n+1-j} \right)^4 \le \mathrm{E} \left(\sum_{j=1}^{n} a_j(\alpha) |y_{n+1-j}| \right)^4 \le \left[\sum_{j=1}^{n} a_j(\alpha) \left(\mathrm{E}[y_{n+1-j}]^4 \right)^{1/4} \right]^4.$$

The RHS of the above equation is $O(\{\log n\}^4 n^{2-4\alpha})$ because $\{y_t\}$ has finite moments of all orders. Since $\mathrm{E}\,|\,\widehat{d} - d\,|^8 = O(n^{-4})$ by (3.11) and $\alpha \in (0, 1/4)$, we have $\mathrm{E}[R_{1,n}]^2 = o(n^{-1})$. The cases of D_1 and D_2 can be obtained similarly because the order of sequences corresponding to $a_j(\alpha)$ is $o(\{\log j\}^2 j^{-1/2-\alpha})$. The case of the ARFISMA(p,d,q) model can be obtained similarly because coefficients of partial derivatives consisting of $\boldsymbol{\beta}$ are absolutely summable and decay exponentially.

For the first term of the RHS of (3.16), we have:

$$\sum_{j=1}^{n} c_j^{(1)}(h,\boldsymbol{\delta}^0) y_{n+1-j} = \sum_{j=0}^{n-1}\sum_{k=0}^{j} c_{k+1}^{(1)}(h,\boldsymbol{\delta}^0)\psi_{j-k}(\boldsymbol{\delta}^0)\varepsilon_{n-j} = \sum_{j=0}^{n-1} \boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)\varepsilon_{n-j},$$

$$\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) = \boldsymbol{\varphi}_{j+1}(h-1,\boldsymbol{\delta}^0) - \psi_{h-1}(\boldsymbol{\delta}^0)\boldsymbol{\delta}_{j+1} = -\boldsymbol{\delta}_{j+h} - \sum_{k=1}^{h-1}\psi_{h-k}(\boldsymbol{\delta}^0)\boldsymbol{\delta}_{j+k} = -\sum_{k=0}^{h-1}\psi_k(\boldsymbol{\delta}^0)\boldsymbol{\delta}_{j+h-k},$$

for $h \ge 2$, and $\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) = O(j^{-1})$, as $j \to \infty$, where we have used equation (A5.2.4) in Box and Jenkins (1976), $\boldsymbol{\delta}_j = O(j^{-1})$, as $j \to \infty$, and the fact that $\sum_{k=0}^{j}\pi_k(\boldsymbol{\delta})\psi_{j-k}(\boldsymbol{\delta}) = 0$ for $j \ge 1$; $= 1$ for $j = 0$, $\boldsymbol{\delta}_j = \sum_{k=0}^{j}\pi_k^{(1)}(\boldsymbol{\delta}^0)\psi_{j-k}(\boldsymbol{\delta}^0)$, $j \ge 1$, similarly to Chung and Baillie (1993, p.804). It follows that, as $n \to \infty$, $\sum_{j=0}^{n-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)\varepsilon_{n-j} = O_p(1)$ and has bounded moments for all orders. Furthermore, from (3.8) and (3.11), we have, as $n \to \infty$, $[(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)'\sum_{j=0}^{n-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)\varepsilon_{n-j}]^2 = O_p(1/n)$, $\mathrm{E}[(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)'\sum_{j=0}^{n-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)\varepsilon_{n-j}]^2 = O(1/n)$, and, by the fact that $\mathrm{E}[R_{1,n}]^2 = o(n^{-1})$:

$$\mathrm{E}\Big[\,\widehat{y}_n(h) - y_n(h)\,\Big]^2 = \mathrm{E}\left[\,\sum_{j=0}^{n-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)'(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)\varepsilon_{n-j}\,\right]^2 + o\Big(\frac{1}{n}\Big). \tag{A.11}$$

Following from (3.8) and the Taylor expansion around $\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta}^0$, there exists $n_0 > 0$ such that, for all $n > n_0$, $S_\varepsilon^{(1)}(\widehat{\boldsymbol{\delta}}) = \mathbf{0} = S_\varepsilon^{(1)}(\boldsymbol{\delta}^0) + \boldsymbol{I}(\boldsymbol{\delta}^0)(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0) + \{S_\varepsilon^{(2)}(\boldsymbol{\delta}^{**}) - \boldsymbol{I}(\boldsymbol{\delta}^0)\}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)$, almost certainly, where $\|\boldsymbol{\delta}^{**}-\boldsymbol{\delta}^0\| \le \|\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0\|$. It follows that:

$$\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0 = -\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}S_\varepsilon^{(1)}(\boldsymbol{\delta}^0) + R_{2,n}, \tag{A.12}$$

where $R_{2,n} = -\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}\{S_\varepsilon^{(2)}(\boldsymbol{\delta}^{**}) - \boldsymbol{I}(\boldsymbol{\delta}^0)\}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)$, and $|R_{2,n}| = o_p(1/\sqrt{n})$ by (A.7). Furthermore, since $S_\varepsilon^{(2)}(\boldsymbol{\delta}^{**})$ and $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)$ have finite moments of all orders, using Lemma A (ii) and Lemma C (i) of Serfling (1980, pp.13–15), $\mathrm{E}\,\|R_{2,n}\|^r = o(n^{-r/2})$ uniformly in $\boldsymbol{\delta}^{**} \in D_\delta$, for $r \ge 2$, as $n \to \infty$.

Combining (A.11) and (A.12), $\mathrm{E}[\,\widehat{y}_n(h) - y_n(h)\,]^2$ is given by, as $n \to \infty$:

$$\mathrm{E}\Big[\,\widehat{y}_n(h) - y_n(h)\,\Big]^2 = \frac{1}{n^2\sigma_0^4}\,\mathrm{E}\left[\sum_{j=0}^{n-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)'\varepsilon_{n-j}\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}\sum_{t=0}^{n-2}\sum_{k=1}^{n-t-1}\boldsymbol{\delta}_k\varepsilon_{n-t-k}\varepsilon_{n-t}\right]^2 + o\Big(\frac{1}{n}\Big), \tag{A.13}$$

where we have used $S_\varepsilon^{(1)}(\boldsymbol{\delta}^0) = \sum_{t=0}^{n-2}\sum_{k=1}^{n-t-1}\boldsymbol{\delta}_k\varepsilon_{n-t-k}\varepsilon_{n-t}/(n\sigma_0^2)$. We note that, as $j \to \infty$, $\boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0) = O(j^{-1})$, $\boldsymbol{\delta}_j = O(j^{-1})$ and, as $n \to \infty$, $\mathrm{E}[S_\varepsilon^{(1)}(\boldsymbol{\delta}^0)S_\varepsilon^{(1)}(\boldsymbol{\delta}^0)'] \sim \boldsymbol{I}(\boldsymbol{\delta}^0)/n$. Hence we obtain the result by (3.15a) of Lemma 3.1 and (A.13).

Proof of Corollary 3.1:

Although the result in (3.18) follows from Theorem 3.1, we present a direct proof since $y_{n+1} - \widehat{y}_n(1)$ is treated as a residual. Similar to the proof of Theorem 3.1 of Tanaka (1999), by a Taylor expansion around $\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta}$, we have:

$$y_{n+1} - \widehat{y}_n(1) = \sum_{j=0}^{n}\pi_j(\widehat{\boldsymbol{\delta}})y_{n+1-j} = \varepsilon_{n+1} + \sum_{j=1}^{n}\boldsymbol{\delta}_j'(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)\varepsilon_{n+1-j} + R_{3,n},$$

where $\mathrm{E}[R_{3,n}]^2 = o(1/n)$ by the same reasoning as $R_{1,n}$ in the proof of Theorem 3.1. Similarly to (A.11)-(A.13), we have, as $n \to \infty$,

$$\mathrm{E}\Big[\,y_{n+1} - \widehat{y}_n(1)\,\Big]^2 \sim \sigma_0^2 + \frac{\sigma_0^2}{n}\sum_{j=1}^{n}\boldsymbol{\delta}_j'\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}\boldsymbol{\delta}_j \sim \sigma_0^2\left\{1 + \frac{\mathrm{tr}(\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}\boldsymbol{I}(\boldsymbol{\delta}^0))}{n}\right\} = \sigma_0^2\left(1 + \frac{1+p+q}{n}\right),$$

where we have used the fact that $\sum_{j=1}^{\infty}\boldsymbol{\delta}_j\boldsymbol{\delta}_j' = \boldsymbol{I}(\boldsymbol{\delta}^0)$.

Proof of convergence of moments of $\widehat{\delta}$ and $\widetilde{\beta}$ for the model (3.19) **with** $\theta = c/\sqrt{n}$**:**

Note that, by (3.9) and (3.10), $\{\boldsymbol{\delta}_k\}$ does not depend on d and hence $\boldsymbol{I}(\boldsymbol{\delta}^0)$ and $\{\boldsymbol{\delta}_k\}$ are fixed matrices. Therefore $\varepsilon_t = u_t(\boldsymbol{\delta}^0) = \varepsilon_t(\boldsymbol{\delta}^0)$, $\varepsilon_t^{(i)}(\boldsymbol{\delta}^0)$, and $u_t^{(i)}(\boldsymbol{\delta}^0)$ for $i = 1, 2$ are the same as those of the model (1.1) with Assumption 3 (θ is fixed), and the results on convergence with probability one, (A.7) and (A.9), still hold if d is replaced by $d = d_0 + c/\sqrt{n}$.

First, we will show that there exists a number $n_0 > 0$, $q > r \geq 2$ and for all and $n \geq n_0$:

$$\mathrm{E}\left\|\left\{\widetilde{Q}_{n,\beta}^{(2)}(\boldsymbol{\delta}^*)\right\}^{-1}\right\|_S^q < \infty \quad \text{and} \quad \mathrm{E}\left\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right\|^r = O(n^{-r/2}), \tag{A.14}$$

where $\|\boldsymbol{\delta}^* - \boldsymbol{\delta}\| \leq \|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}\|$ and $\widetilde{Q}_{n,\beta}^{(2)}(\boldsymbol{\delta}^*) = \partial^2 S_\varepsilon(\boldsymbol{\delta}^*)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}$. We shall consider the limiting distribution of $\sqrt{n}(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0)$. Using the same argument as in Katayama (2006, Remark 2), strong consistency of $\widetilde{\boldsymbol{\beta}}$ is obtained, and for large n:

$$\sqrt{n}\,\frac{\partial S_\varepsilon(\widetilde{\boldsymbol{\delta}})}{\partial\boldsymbol{\beta}} = \mathbf{0} = \sqrt{n}\,\frac{\partial S_\varepsilon(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\beta}} - c\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial d\partial\boldsymbol{\beta}} + \frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\sqrt{n}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right), \tag{A.15}$$

where $\|\boldsymbol{\delta}^* - \boldsymbol{\delta}^0\| \leq \|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\|$. Since (A.7) and (A.9) hold when d is replaced by $d = d_0 + c/\sqrt{n}$:

$$\widetilde{\boldsymbol{I}}_{n,\beta}(\boldsymbol{\delta}^0) \equiv \frac{1}{n\sigma_0^2}\sum_{t=1}^n \frac{\partial\varepsilon_t(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\beta}}\frac{\partial\varepsilon_t(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\beta}'} \xrightarrow{a.c.} \boldsymbol{\Phi}, \quad \frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial d\partial\boldsymbol{\beta}} \xrightarrow{a.c.} \boldsymbol{\kappa}, \tag{A.16}$$

$$\widetilde{\boldsymbol{R}}_{n,\beta}(\boldsymbol{\delta}^*) \equiv \frac{1}{n\sigma_0^2}\sum_{t=1}^n \varepsilon_t(\boldsymbol{\delta}^0)\frac{\partial^2\varepsilon_t(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} + \left\{\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} - \frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\} \xrightarrow{a.c.} \mathbf{0},$$

$$\widetilde{Q}_{n,\beta}^{(2)}(\boldsymbol{\delta}^*) \equiv \frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = \widetilde{\boldsymbol{I}}_{n,\beta}(\boldsymbol{\delta}^0) + \widetilde{\boldsymbol{R}}_{n,\beta}(\boldsymbol{\delta}^*) \xrightarrow{a.c.} \boldsymbol{\Phi},$$

as $n \to \infty$. $\{\boldsymbol{\delta}_{k,\beta}\}$ is given by (3.9) and (3.10) does not depend on d and $\{\varepsilon_t(\boldsymbol{\delta}^0)\partial\varepsilon_t(\boldsymbol{\delta}^0)/\partial\boldsymbol{\beta}\}$ is a sequence of martingale differences. It follows that $n^{1/2}\partial S_\varepsilon(\boldsymbol{\delta}^0)/\partial\boldsymbol{\beta} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Phi})$, as $n \to \infty$, by a central limit theorem for martingale differences. See, for example, Fuller (1996, Theorem 5.3.4). Combining this, (A.15), and (A.16), we obtain

$$\sqrt{n}(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) = [\,-c \quad \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\,]' \xrightarrow{d} [\,-c \quad N(c\boldsymbol{\Phi}^{-1}\boldsymbol{\kappa},\ \boldsymbol{\Phi}^{-1})'\,]'.$$

Now, to show (A.14), we first consider $\mathrm{E}\,\|\{\widetilde{Q}_{n,\beta}^{(2)}(\boldsymbol{\delta}^*)\}^{-1}\|_S^q < \infty$. Let $\widetilde{Z}_{v,\beta}(j) = \sum_{k=1}^j \alpha_{v,\beta}(j,k)\varepsilon_k$, and $\alpha_v(j,k) = \boldsymbol{v}\boldsymbol{\delta}_{j+1-k,\beta}/\sigma_0$. Then, substituting $\widetilde{Q}_{n,\beta}^{(2)}(\boldsymbol{\delta}^*)$, $\widetilde{\boldsymbol{I}}_{n,\beta}(\boldsymbol{\delta}^0)$, $\widetilde{\boldsymbol{R}}_{n,\beta}(\boldsymbol{\delta}^*)$, and $\widetilde{Z}_{v,\beta}(j)$ into $Q_n^{(2)}(\boldsymbol{\theta}^*)$, $\boldsymbol{I}_n(\boldsymbol{\theta}^0)$, $\boldsymbol{R}_n(\boldsymbol{\theta}^*)$, and $Z_v(j)$ and borrowing the other notations in the proof of Theorem 2.1, it is enough to show that:

$$\mathrm{E}\left[\inf_{\|\boldsymbol{\nu}\|=1}\sum_{i=0}^{lq-1}\widetilde{Z}_{\nu,\beta}^2(n(i)+j)\right]^{-q} \leq C < \infty \tag{A.17}$$

holds for all $j = 0, \ldots, (1-\delta)n/(lq) - 1$ and $n \geq n_0$, where $\widetilde{Z}_{\nu,\beta}(j) = \sum_{k=1}^j \alpha_{\nu,\beta}(j,k)\varepsilon_k$, $\alpha_\nu(j,k) = \boldsymbol{\nu}\boldsymbol{\delta}_{j+1-k,\beta}/\sigma_0$. Note that, as in the proof of Corollary 2.1, we obtain $\Pr(\sum_{i=0}^{lq-1}\widetilde{Z}_{v,\beta}^2(n(i)) < \beta_*) \leq \mathrm{const}\,\beta_*^{lq/2}$, $\beta_* > 0$, because $\{\boldsymbol{\delta}_{j,\beta}\}$ does not depend on d and $\boldsymbol{\Phi} = \sum_{j=1}^\infty \boldsymbol{\delta}_{j,\beta}\boldsymbol{\delta}_{j,\beta}'$ is positive definite. The rest of the proof of (A.17) is obtained in the same way as those in the proof of (A.1) of Theorem 2.1. Next we consider $\mathrm{E}\,\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\|^r = O(n^{-r/2})$ in (A.14). By c_r-inequality,

$$\mathrm{E}\left\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right\|^r = \mathrm{E}\left[c^2 + \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right]^{r/2} \leq 2^{r/2-1}\left(|c|^r + \mathrm{E}\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^r\right).$$

Therefore, it is enough to show $\mathrm{E}\,\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 = O(n^{-r/2})$. Now, from (A.15):

$$\sqrt{n}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left\{\widetilde{Q}_{n,\beta}^{(2)}(\boldsymbol{\delta}^*)\right\}^{-1}\left\{\sqrt{n}\,\frac{\partial S_\varepsilon(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\beta}} - c\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial d\partial\boldsymbol{\beta}}\right\}.$$

For any $q' > 1$ and $n \geq n_0$:

$$\mathrm{E}\left\| \sqrt{n}\,\frac{\partial S_\varepsilon(\boldsymbol{\delta}^0)}{\partial \boldsymbol{\beta}} - c\left(\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial d \partial \boldsymbol{\beta}}\right)\right\|^{q'} \tag{A.18}$$
$$\leq 3^{q'-1}\,\mathrm{E}\left\{ \left\| \sqrt{n}\,\frac{\partial S_\varepsilon(\boldsymbol{\delta}^0)}{\partial \boldsymbol{\beta}} \right\|^{q'} + \left\| c\boldsymbol{\kappa} - c\left(\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial d \partial \boldsymbol{\beta}}\right)\right\|^{q'} + \|c\boldsymbol{\kappa}\|^{q'} \right\} < \infty,$$

by Hölder's inequality. We have already shown that $\mathrm{E}\,\|\{\widetilde{Q}^{(2)}_{n,\beta}(\boldsymbol{\delta}^*)\}^{-1}\|_S^q < \infty$ and so the conditions of uniform integrability of $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is satisfied and leads to the conclusion.

Finally, we will show (3.11) for the model (3.19) with $\theta = c/\sqrt{n}$. Namely:

$$n\,\mathrm{E}\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right)\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\right)' \longrightarrow \boldsymbol{I}(\boldsymbol{\delta}^0)^{-1} \qquad \text{and} \qquad \mathrm{E}\left\| \widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0 \right\|^r = O(n^{-r/2}), \tag{A.19}$$

for $r \geq 2$. (3.8), (3.9), and (3.10) still hold by Katayama (2006, Remark 3). Let $\widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0)$ and $\widehat{\boldsymbol{R}}_n(\boldsymbol{\delta}^*)$ be defined by (A.8), $\widehat{Z}_v(j) = \sum_{k=1}^{j} \alpha_{v,\delta}(j,k)\varepsilon_k$, and $\alpha_{v,\delta}(j,k) = \boldsymbol{v}'\boldsymbol{\delta}_{j+1-k}/\sigma_0$. Then, substituting $S^{(2)}_\varepsilon(\boldsymbol{\delta}^*)$, $\widehat{\boldsymbol{I}}_n(\boldsymbol{\delta}^0)$, $\widehat{\boldsymbol{R}}_n(\boldsymbol{\delta}^*)$, and $\widehat{Z}_v(j)$ into $Q^{(2)}_n(\boldsymbol{\theta}^*)$, $\boldsymbol{I}_n(\boldsymbol{\theta}^0)$, $\boldsymbol{R}_n(\boldsymbol{\theta}^*)$, and $Z_v(j)$ and borrowing the other notations in the proof of Theorem 2.1, it is enough to show that:

$$\mathrm{E}\left[\inf_{\|\boldsymbol{\nu}\|=1} \sum_{i=0}^{lq-1} \widehat{Z}^2_\nu(n(i) + j) \right]^{-q} \leq C < \infty \tag{A.20}$$

holds for all $j = 0, \ldots, (1-\delta)n/(lq) - 1$ and $n \geq n_0$, where $\widehat{Z}_\nu(j) = \sum_{k=1}^{j} \alpha_{\nu,\delta}(j,k)\varepsilon_k$ and $\alpha_{\nu,\delta}(j,k) = \boldsymbol{\nu}'\boldsymbol{\delta}_{j+1-k}/\sigma_0$. Since $\{\boldsymbol{\delta}_j\}$ does not depend on d and $\boldsymbol{I}(\boldsymbol{\delta}^0) = \sum_{j=1}^{\infty} \boldsymbol{\delta}_j\boldsymbol{\delta}_j'$ is positive definite, we can use the same lines of the proof of Corollary 2.1. Namely, there exists $k(i)$ $(i = 0, 1, \ldots, lq-1)$ such that $k(i) = \max[\,k\,|\,|\alpha_{v,\delta}(n(i),k)| > 0,\ k = n(i-1) + 1, \ldots, n(i)\,]$ for $i = 1, \ldots, lq-1$ and $k(0) = \max[\,k\,|\,|\alpha_{v,\delta}(n(0),k)| > 0,\ k = 1, \ldots, n(0)\,]$. Then $\widehat{Z}_v(n(i)) = \sum_{t=1}^{k(i)} \alpha_{v,\delta}(n(i),t)\varepsilon_t$. From this, (a) of Assumption 3 and (ii) of Lemma 2.1, we obtain $\Pr(\sum_{i=0}^{lq-1} \widehat{Z}^2_v(n(i)) < \beta_*) \leq \mathrm{const}\,\beta_*^{lq/2}$, $\beta_* > 0$. The rest of the proof of (A.20) for the model (3.19) with $\theta = c/\sqrt{n}$ is obtained by the same way as those in the proof of Theorem 2.1. Since $\boldsymbol{I}(\boldsymbol{\delta}^0) = \sum_{j=1}^{\infty} \boldsymbol{\delta}_j\boldsymbol{\delta}_j'$ is positive definite and $\{\varepsilon_t(\boldsymbol{\delta}^0)\varepsilon^{(1)}_t(\boldsymbol{\delta}^0)\}$ is a sequence of martingale differences, $\|\varepsilon^{(1)}_t(\boldsymbol{\delta}^0)\|$ and $\|\sqrt{n}\,S^{(1)}_\varepsilon(\boldsymbol{\delta}^0)\|$ have finite moments of all orders. It follows that the uniform integrability in the proof of Theorem 2.1 is satisfied, which proves (A.19).

Proof of Theorem 3.2:

Using the same argument as in (A.11) and (A.14), as $n \to \infty$:

$$\mathrm{E}\left[\,\widetilde{y}_n(h) - y_n(h)\,\right]^2 = \mathrm{E}\left[\sum_{j=0}^{n-1} \boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)'(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0)\varepsilon_{n-j} \right]^2 + o\left(\frac{1}{n}\right). \tag{A.21}$$

The first term of the RHS of (A.21) is:

$$\sum_{j=0}^{n-1}\sum_{k=0}^{n-1} \boldsymbol{\varphi}_j(h,\boldsymbol{\delta}^0)'\,\mathrm{E}\left[\begin{pmatrix} c^2/n & -c(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'/\sqrt{n} \\ -c(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sqrt{n} & (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})' \end{pmatrix} \varepsilon_{n-j}\varepsilon_{n-k}\right]\boldsymbol{\varphi}_k(h,\boldsymbol{\delta}^0). \tag{A.22}$$

Similar to (A.12), we can rewrite (A.15) as:

$$\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = -\boldsymbol{\Phi}^{-1}\left\{\frac{\partial S_\varepsilon(\boldsymbol{\delta}^0)}{\partial \boldsymbol{\beta}} - \frac{1}{\sqrt{n}}\boldsymbol{\kappa}c\right\} + R_{4,n}, \tag{A.23}$$

where:

$$R_{4,n} = -\boldsymbol{\Phi}^{-1}\left[\left\{\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} - \boldsymbol{\Phi}\right\}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \left\{\boldsymbol{\kappa} - \frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^*)}{\partial d \partial \boldsymbol{\beta}}\right\}\frac{c}{\sqrt{n}}\right],$$

and $\mathrm{E}\,\|R_{4,n}\|^r = o(n^{-r/2})$ by (A.14) and (A.16). The first term of the RHS of (A.23) is:

$$-\boldsymbol{\Phi}^{-1}\left[\frac{1}{n\sigma_0^2}\sum_{t=0}^{n-2}\sum_{k=1}^{n-t-1}\boldsymbol{\delta}_{k,\beta}\varepsilon_{n-t-k}\varepsilon_{n-t} - \frac{1}{\sqrt{n}}\boldsymbol{\kappa}c\right]. \tag{A.24}$$

Substituting (A.24) for $(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ in (A.22), by Lemma 3.1 and Lemma 3.2, (A.22) demonstrates the second term of the RHS of (3.21) and concludes the proof.

Proof of Corollary 3.2:

Using the same argument as in the proof of Theorem 3.1 and (A.19), we find that Theorem 3.1 still holds if the model (1.1) is replaced by the model (3.19) and $\theta = c/\sqrt{n}$.

To make a comparison between $\widehat{y}_n(h)$ and $\widetilde{y}_n(h)$ from $\mathrm{ARE}[\,\widehat{y}_n(h),\ \widetilde{y}_n(h)\,]$, let $\boldsymbol{v}$ be a fixed $(p+q+1)$-vector. Then we have:

$$\boldsymbol{v}'\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1}\boldsymbol{v} - \boldsymbol{v}'\boldsymbol{\Gamma}\boldsymbol{v} = \boldsymbol{v}'\big(\boldsymbol{I}(\boldsymbol{\delta}^0)^{-1} - \boldsymbol{\Gamma}\big)\boldsymbol{v} = (\omega^2 - c^2)\left[\boldsymbol{v}'\begin{pmatrix} -1 \\ \boldsymbol{\Phi}^{-1}\boldsymbol{\kappa}\end{pmatrix}\right]^2.$$

It follows that the sign of the above equation depends on $\omega^2 - c^2$, which yields the corollary.

Proof of Corollary 3.3:

The result in (3.24) is obtained similarly to the proofs of Corollary 3.1 and Theorem 3.2. Using a Taylor series approximation, we have, as $n \to \infty$:

$$\mathrm{E}\left[y_{n+1} - \widetilde{y}_n(1)\right]^2 \sim \sigma_0^2 + \frac{\sigma_0^2}{n}\sum_{j=1}^{n}\boldsymbol{\delta}_j'\boldsymbol{\Gamma}\boldsymbol{\delta}_j \sim \sigma_0^2\left\{1 + \frac{\mathrm{tr}(\boldsymbol{\Gamma}\boldsymbol{I}(\boldsymbol{\delta}^0))}{n}\right\} = \sigma_0^2\left(1 + \frac{p+q+c^2/\omega^2}{n}\right),$$

where we have used the fact that $\sum_{j=1}^{\infty}\boldsymbol{\delta}_j\boldsymbol{\delta}_j' = \boldsymbol{I}(\boldsymbol{\delta}^0)$.

Hereafter, we present proofs omitted from the paper, which are not to be published.

Omitted proofs A (omitted proofs relating to Section 2)

O.1 On the conditions of Assumption 2 (p.5)

O.1.1 Moments conditions

We establish (c) in Assumption 1 and (2.10) from (b) of Assumption 2.

Let $\boldsymbol{x} = (x_1, x_2, \dots, x_p)'$ be a random vector. Then we have $\mathrm{E}\,\|\boldsymbol{x}\|^{2q} = \mathrm{E}[\boldsymbol{x}'\boldsymbol{x}]^q = \mathrm{E}[\sum_{i=1}^{p} x_i^2]^q \le p^{q-1}\sum_{i=1}^{p}\mathrm{E}[x_i]^{2q}$ for $q > 1$ because $(\sum_{i=1}^{p}|a_i|)^q \le p^{q-1}\sum_{i=1}^{p}|a_i|^q$ for $q > 1$ by Hölder's inequality. Therefore, we consider the case of $p = 1$ for simplicity because the case of $p > 1$ can be obtained similarly.

To show that (c) in Assumption 1 is satisfied, by Lyapunov's inequality, it is sufficient to show that $\mathrm{E}[\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j}]^{2r} < \infty$ where r is any positive integer. We note that when $\sum_{l=1}^{\infty}|a_l|$ is convergent, $\sum_{l=1}^{\infty}|a_l|^{1+\delta}$ is convergent for $\delta \ge 0$. (Proof: Since $\sum_{l=1}^{\infty}|a_l|$ is convergent, there exists m such that $|a_l| < 1$ for all $l > m$. Thus we have $\sum_{l=1}^{\infty}|a_l|^{1+\delta} = \sum_{l=1}^{m}|a_l|^{1+\delta} + \sum_{l=m+1}^{\infty}|a_l|^{1+\delta} \le \sum_{l=1}^{m}|a_l|^{1+\delta} + \sum_{l=m+1}^{\infty}|a_l| < \infty$ because $|a_l|^{1+\delta} \le |a_l|$ for all $l > m$). Hence we obtain $\sum_{j=1}^{\infty}|\boldsymbol{d}_j(\boldsymbol{\theta}^0)|^{2+\delta} < \infty$ by (2.14). Since $\{\varepsilon_t\}$ is iid $(0, \sigma_0^2)$ and $\mathrm{E}[\varepsilon_t]^r < \infty$ for all positive integers r:

$$\mathrm{E}\left[\sum_{j=1}^{t-1}|\boldsymbol{d}_j(\boldsymbol{\theta}^0)|\varepsilon_{t-j}\right]^{2r} \le \mathrm{const}\left(\sum_{k=2}^{2r}\sum_{j=1}^{t-1}|\boldsymbol{d}_j(\boldsymbol{\theta}^0)|^k\right)^r < \infty, \quad \text{for any } t > 2, \tag{O.1}$$

which demonstrates (c) in Assumption 1 for the case of $p = 1$.

Next we establish (2.10) from (b) of Assumption 2. Let $Z_t = \varepsilon_t\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j}$. Then $\mathrm{E}[Z_t] = 0$, $\mathrm{E}[Z_tZ_s] = 0$ for $t \ne s$, $\mathrm{E}[Z_t]^r < \infty$ for all positive integers r by (O.1), and:

$$\mathrm{E}\left[\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j}\right]^{2r} = \frac{1}{n^r}\sum_{2\le t_1,t_2,\dots,t_{2r}\le n}\mathrm{E}\left[Z_{t_1}\cdots Z_{t_{2r}}\right] \le \frac{\mathrm{const}}{n^r}\sum_{2\le t_1,t_2,\dots,t_r\le n} 1 < \infty, \tag{O.2}$$

which establish (2.10) for the case of $p = 1$.

O.1.2 Asymptotic normality

We establish (2.7) from Assumption 2. Since:

$$\begin{aligned}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}} &= \frac{1}{\sqrt{n}\sigma_0^2}\sum_{t=2}^{n}\varepsilon_t\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j} = \frac{1}{\sqrt{n}\sigma_0^2}\sum_{t=3}^{n+2}\varepsilon_t\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j} + o_p(1)\\ &= \frac{1}{\sqrt{n}\sigma_0^2}\sum_{t=1}^{n}\varepsilon_{t+2}\sum_{j=1}^{t+1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t+2-j} + o_p(1) = \frac{1}{\sqrt{n}\sigma_0^2}\sum_{t=1}^{n}v_t\sum_{j=1}^{t+1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)v_{t-j} + o_p(1),\end{aligned}$$

as $n \to \infty$, where $v_t = \varepsilon_{t+2}$ and $\{v_t\}_{t=-1}^{\infty}$ is iid $(0, \sigma_0^2)$, it is sufficient to show that:

$$S_n = \sum_{t=1}^{n} Z_{tn} \xrightarrow{d} N\big(0,\ \boldsymbol{\lambda}'\mathbb{I}_p\boldsymbol{\lambda}\big), \qquad \text{as } n \to \infty, \tag{O.3}$$

by the Cramer–Wold device, where $\mathbb{I}_p$ is a $p \times p$ identity matrix, $\boldsymbol{\lambda}$ is a fixed vector,

$$Z_{tn} = \frac{1}{\sqrt{n}\sigma_0^2}\boldsymbol{\lambda}'\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\left\{\sum_{j=1}^{t+1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)v_{t-j}\right\}v_t,$$

and $\boldsymbol{I}(\boldsymbol{\theta}^0)^{1/2}$ is a $p \times p$ lower triangular matrix defined by $\boldsymbol{I}(\boldsymbol{\theta}^0) = \boldsymbol{I}(\boldsymbol{\theta}^0)^{1/2}\{\boldsymbol{I}(\boldsymbol{\theta}^0)^{1/2}\}'$. Since the case of $\boldsymbol{\lambda} = \mathbf{0}$ is obvious, we consider the case of $\boldsymbol{\lambda} \neq \mathbf{0}$.

To prove (O.3), we use a central limit theorem for martingale differences. Let $\{Z_{tn}| 0 \le t \le n, n \ge 1\}$ denote a triangular array of random variables defined on the probability space $(\Omega, \mathcal{A}, P)$, and let $\mathcal{A}_{tn}$ $(0 \le t \le n,\, n \ge 1)$ be the sigma field generated by $\{Z_{jn}| 0 \le j \le t\}$. Then $\mathcal{A}_{t-1,n}$ is contained in $\mathcal{A}_{tn}$, Z_{tn} is $\mathcal{A}_{tn}$-measurable and $\mathrm{E}[Z_{tn}|\, \mathcal{A}_{t-1,n}] = 0$ a.c. for $1 \le t \le n$. Therefore, we prove (O.3) by showing that conditions (a) of Prakasha Rao (1987, Proposition 1.7.14) hold[1]. Let:

$$V_n^2 = \sum_{t=1}^{n} \mathrm{E}[Z_{tn}^2|\mathcal{A}_{t-1,n}] = \frac{1}{n\sigma_0^2}\sum_{t=1}^{n} \boldsymbol{\lambda}'\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\left\{\sum_{j,k=1}^{t+1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)v_{t-j}v_{t-k}\right\}\left\{\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\right\}'\boldsymbol{\lambda},$$

$$\text{and } s_n^2 = \mathrm{E}[V_n^2] = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\lambda}'\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\sum_{j=1}^{t+1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_j(\boldsymbol{\theta}^0)'\left\{\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\right\}'\boldsymbol{\lambda}.$$

By (b) of Assumption 2, we have, as $n \to \infty$:

$$\begin{aligned}
\boldsymbol{I}_n(\boldsymbol{\theta}^0) &= \frac{1}{n\sigma_0^2}\sum_{t=2}^{n}\sum_{j,k=1}^{t-1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'\varepsilon_{t-j}\varepsilon_{t-k} = \frac{1}{n\sigma_0^2}\sum_{t=2}^{n+1}\sum_{j,k=1}^{t-1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'\varepsilon_{t-j}\varepsilon_{t-k} + o_p(1)\\
&= \frac{1}{n\sigma_0^2}\sum_{t=0}^{n-1}\sum_{j,k=1}^{t+1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'\varepsilon_{t+2-j}\varepsilon_{t+2-k} + o_p(1) = \frac{1}{n\sigma_0^2}\sum_{t=0}^{n-1}\sum_{j,k=1}^{t+1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'v_{t-j}v_{t-k} + o_p(1)\\
&= \frac{1}{n\sigma_0^2}\sum_{t=1}^{n}\sum_{j,k=1}^{t+1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_k(\boldsymbol{\theta}^0)'v_{t-j}v_{t-k} + o_p(1) \xrightarrow{p} \boldsymbol{I}(\boldsymbol{\theta}^0) = \lim_{t\to\infty}\sum_{j=1}^{t+1} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_j(\boldsymbol{\theta}^0)'.
\end{aligned}$$

By this and the Toeplitz Lemma, we obtain, as $n \to \infty$,

$$V_n^2 \xrightarrow{p} \boldsymbol{\lambda}'\boldsymbol{\lambda} > 0, \quad s_n^2 \longrightarrow \boldsymbol{\lambda}'\boldsymbol{\lambda}, \quad \text{and} \quad V_n^2/s_n^2 \xrightarrow{p} 1.$$

Hence condition (ii) of (a) of Prakasha Rao (1987, Proposition 1.7.14) is satisfied. Now for any $\epsilon > 0$ and $\delta > 0$:

$$\begin{aligned}
&s_n^{-2}\sum_{t=1}^{n}\mathrm{E}\left[Z_{tn}^2 I(|Z_{tn}| \ge \epsilon s_n)\right] \le s_n^{-2}\sum_{t=1}^{n}(\epsilon s_n)^{-\delta}\,\mathrm{E}\left[|Z_{tn}|^{2+\delta} I(|Z_{tn}| \ge \epsilon s_n)\right]\\
&= \frac{1}{s_n^{2+\delta}\epsilon^{\delta}n^{1+\delta/2}\sigma_0^{2(2+\delta)}}\sum_{t=1}^{n}\mathrm{E}\left[\left|\boldsymbol{\lambda}'\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\left\{\sum_{j=1}^{t+1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)v_{t-j}\right\}v_t\right|^{2+\delta}\right]\\
&\le \text{const}\, s_n^{-(2+\delta)}\epsilon^{-\delta}n^{-\delta/2},
\end{aligned}$$

by (O.2) [See the proof of Theorem 5.5.1 of Fuller (1996)]. Hence condition (i) of (a) of Prakasha Rao (1987, Proposition 1.7.14) is satisfied and:

$$\frac{S_n}{s_n} = \frac{1}{(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{1/2}}\sum_{t=1}^{n} Z_{tn} + o_p(1) \xrightarrow{d} N(0,1),$$

which establish (O.3).

O.2 Properties of eigenvalues of a non-singular, positive definite, and symmetric matrix (p.17)

In the proof of Theorem 2.1, we argued that "$\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)'$ be on the unit sphere S of $\mathbb{R}^p$, and $Z_\nu(j) = \boldsymbol{\nu}'e_{j+1}^{(1)}(\boldsymbol{\theta}^0)$. Then $\|\{Q_n^{(2)}(\boldsymbol{\theta}^*)\}^{-1}\|_S = \lambda_0^{-1}$ and $\lambda_0 = \inf_{\|\boldsymbol{\nu}\|=1}\boldsymbol{\nu}'Q_n^{(2)}(\boldsymbol{\theta}^*)\boldsymbol{\nu}$. ". This is proven from the following properties:

[1] Prakasha Rao BLS. 1987. *Asymptotic Theory of Statistical Inference*. John Wiley: New York. This proposition is also seen in Fuller (1996, Theorem 5.3.4), however, there is a typographical error in Fuller's text.

1. Let $\boldsymbol{A}$ be a non-singular, positive definite, and symmetric matrix and λ_1 be the smallest eigenvalue of $\boldsymbol{A}$. Then $\sup_{\|\boldsymbol{x}\|=1} \boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x} = \lambda_1^{-1}$. (Proof: Let $\boldsymbol{x}$ be a vector that satisfies $\|\boldsymbol{x}\| = 1$ and λ be an eigenvalue of $\boldsymbol{A}$. Then λ is positive and $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \Rightarrow \boldsymbol{x} = \lambda\boldsymbol{A}^{-1}\boldsymbol{x} \Rightarrow \boldsymbol{x}'\boldsymbol{x} = \lambda\boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x} \Rightarrow \boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x} = \lambda^{-1} = (\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})^{-1}$. It follows that $\sup_{\|\boldsymbol{x}\|=1} \boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x} = \sup_{\|\boldsymbol{x}\|=1}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})^{-1} = (\inf_{\|\boldsymbol{x}\|=1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})^{-1} = \lambda_1^{-1}$.)

2. Let $\boldsymbol{A}$ be a positive definite and symmetric matrix. Then $\|\boldsymbol{A}\|_S$ is the largest eigenvalue of $\boldsymbol{A}$. (Proof: Following from P 11.2.13 of Rao and Rao (1998, p.371) [2] , for a square matrix $\boldsymbol{B}$, $\|\boldsymbol{B}\|_S$ is the largest singular value of $\boldsymbol{B}$. Since $\boldsymbol{A}$ is a positive definite and symmetric matrix, its eigenvalues are all positive real numbers. Therefore, the largest singular value of $\boldsymbol{A}$ is the largest eigenvalue of $\boldsymbol{A}$.).

3. Let $\boldsymbol{A}$ be a non-singular, positive definite, and symmetric matrix and λ_1 be the smallest eigenvalue of $\boldsymbol{A}$. Then $\|\boldsymbol{A}^{-1}\|_S = \lambda_1^{-1}$. (Proof: By assumption, $\boldsymbol{A}^{-1}$ is also non-singular, positive definite, and symmetric. Using 1 and 2, $\|\boldsymbol{A}^{-1}\|_S = \sup_{\|\boldsymbol{x}\|=1} \boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x} = \lambda_1^{-1}$).

O.3 On a subset of S' of S, with $(\mathrm{const}/\varepsilon)^{2p-2}$ elements in the proof of Theorem 2.1 (p.18)

Bhansali and Papangelou (1991, p.1159) showed that "there is a subset S' of S, with fewer than $2[4p^2/\varepsilon^2]^{p-1}$ elements". Though we cannot prove this exactly, we deduce a similar result:

Note that, for any $x, y \geq 0$ and $\epsilon > 0$, if $|x - y| < \epsilon^2$, then $|\sqrt{x} - \sqrt{y}| < \epsilon$. (Proof: When $x > y$, $|x-y| = x - y < \epsilon^2 \Rightarrow x < y + \epsilon^2 \Rightarrow \sqrt{x} < \sqrt{y+\epsilon^2} < \sqrt{y} + \epsilon \Rightarrow \sqrt{x} - \sqrt{y} < \epsilon$ because $\sqrt{a^2+b^2} < a + b$ for any $a, b > 0$. Similarly, when $x < y$, $\sqrt{y} - \sqrt{x} < \epsilon$, it follows that $|\sqrt{x} - \sqrt{y}| < \epsilon$. This also proves the uniform continuity of $f(x) = \sqrt{x}$ for $x > 0$.)

Let $\boldsymbol{v} = (v_1, \ldots, v_p)'$ be on the unit sphere S of $\mathbb{R}^p$ and let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)'$ be on the unit sphere $S' \subset S$ with $\|\boldsymbol{v} - \boldsymbol{\nu}\| < \varepsilon \in (0,1)$. Then, for at least one i, $|v_i| \geq 1/\sqrt{p}$. Therefore, let $v_p \geq 1/\sqrt{p}$, $v_p = \{1 - \sum_{i=1}^{p-1} v_i^2\}^{1/2}$, $\nu_p = \{1 - \sum_{i=1}^{p-1} \nu_i^2\}^{1/2}$, $|v_i - \nu_i| < \varepsilon_i \in (0,1)$ for $i = 1, 2, \ldots, p$, and $\varepsilon_p < v_p$ $(0 < v_p - \varepsilon_p < \nu_p < v_p + \varepsilon_p)$. Since $|v_i|, |\nu_i| \leq 1$ for $i = 1, 2, \ldots, p$, and:

$$\begin{aligned}\left|v_p^2 - \nu_p^2\right| &= \left|\left(1 - \sum_{i=1}^{p-1} v_i^2\right) - \left(1 - \sum_{i=1}^{p-1} \nu_i^2\right)\right| = \left|\sum_{i=1}^{p-1}(\nu_i^2 - v_i^2)\right| \leq \sum_{i=1}^{p-1} |\nu_i - v_i||\nu_i + v_i| \\ &\leq 2\sum_{i=1}^{p-1} |\nu_i - v_i| \leq 2\sum_{i=1}^{p-1} \varepsilon_i,\end{aligned}$$

we have:

$$|v_p - \nu_p| = \left|\sqrt{1 - \sum_{i=1}^{p-1} v_i^2} - \sqrt{1 - \sum_{i=1}^{p-1} \nu_i^2}\right| \leq \sqrt{2\sum_{i=1}^{p-1} \varepsilon_i} = \varepsilon_p \quad \text{(say)},$$

$$\text{and } \|\boldsymbol{v} - \boldsymbol{\nu}\| = \left\{\sum_{i=1}^{p} |v_i - \nu_i|^2\right\}^{1/2} < \left\{\sum_{i=1}^{p-1} \varepsilon_i^2 + \varepsilon_p^2\right\}^{1/2} = \left\{\sum_{i=1}^{p-1} \varepsilon_i^2 + 2\sum_{i=1}^{p-1} \varepsilon_i\right\}^{1/2} < \left\{3\sum_{i=1}^{p-1} \varepsilon_i\right\}^{1/2}.$$

It follows that if we put:

$$\varepsilon_i = \frac{\varepsilon^2}{3p^2} \quad (i = 1, 2, \ldots, p-1) \quad \text{and} \quad \varepsilon_p = \sqrt{2\sum_{i=1}^{p-1} \varepsilon_i},$$

[2] Rao CR, Rao MB. 1998. *Matrix Algebra and Its Applications to Statistics and Econometrics*. World Scientific: Singapore.

then:

$$\|\boldsymbol{v}-\boldsymbol{\nu}\| < \left\{3\sum_{i=1}^{p-1}\varepsilon_i\right\}^{1/2} = \left\{\frac{p-1}{p}\frac{\varepsilon^2}{p}\right\}^{1/2} < \frac{\varepsilon}{\sqrt{p}} < \varepsilon,$$

$$\text{and} \quad \varepsilon_p = \left\{2\sum_{i=1}^{p-1}\varepsilon_i\right\}^{1/2} < \left\{\frac{2}{3}\frac{\varepsilon^2}{p}\right\}^{1/2} < \frac{\varepsilon}{\sqrt{p}} < \frac{1}{\sqrt{p}} \le v_p.$$

The case of $v_p \le -1/\sqrt{p}$ can be deduced similarly by putting $v_p = -\{1-\sum_{i=1}^{p-1} v_i^2\}^{1/2}$, $\nu_p = -\{1-\sum_{i=1}^{p-1}\nu_i^2\}^{1/2}$, and $v_p < -\varepsilon_p$ ($v_p - \varepsilon_p < \nu_p < v_p + \varepsilon_p < 0$).

Since $|\nu_i| < 1$, $\{-1+\varepsilon_i, -1+2\varepsilon_i, -1+3\varepsilon_i, \ldots, 1\}$ have about $2/\varepsilon_i = 6p^2/\varepsilon^2$ elements for $i = 1, 2, \ldots, p-1$. Therefore, for any $\boldsymbol{v} \in S$ and $|v_p| > 1/\sqrt{p}$, S' needs $2(6p^2/\varepsilon^2)^{p-1}$ elements. It follows that, in general, for any $\boldsymbol{v} \in S$, there is a subset S' of S, with $(C/\varepsilon^2)^{p-1}$ elements for some $C > 0$ where $2p(6p^2/\varepsilon^2)^{p-1} < (C/\varepsilon^2)^{p-1}$.

O.4 Properties of the norm in the proof of Theorem 2.1 (p.19)

In the last inequalities of the proof of Theorem 2.1, to prove (2.8) and (2.11), we use the following properties: Let $\boldsymbol{x}$ be $\boldsymbol{x} = (x_1, \ldots, x_p)'$ and $\boldsymbol{A}$ be a $p \times p$ matrix with (i,j)th element, $a_{i,j}$. Then (i) $\max_{1\le i\le p}|x_i| \le \sqrt{p}\|x\|$ and (ii) $\max_{1\le i,j\le p}|a_{i,j}| = \sqrt{p}\|\boldsymbol{A}\|_S$. (Proof: (i) $\max_{1\le i\le p}|x_i| \le \sum_{i=1}^p |x_i| \le \{\sum_{i=1}^p 1 \sum_{i=1}^p |x_i|^2\}^{1/2} = \sqrt{p}\|x\|$. (ii) Let $|a_{i,j}|$ be $|a_{i,j}| = \max_{1\le k,l\le p}|a_{k,l}|$ and $\boldsymbol{e}_j = (0,\ldots,0,1,0,\ldots,0)'$ whose jth element is 1 and zero otherwise. Then $\|\boldsymbol{e}_j\| = 1$ and $|a_{i,j}| \le \sum_{k=1}^p |a_{k,j}| \le \sqrt{p}\|\boldsymbol{A}\boldsymbol{e}_j\| \le \sqrt{p}\sup_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\| = \sqrt{p}\|\boldsymbol{A}\|_S$.)

Omitted proofs B (omitted proofs relating to Sections 3 and 4)

O.5 On the condition of $\{y_t = 0,\ t \le 0\}$ or equivalently $\{\varepsilon_t = 0,\ t \le 0\}$ in Assumption 3 (p.6)

Let $\psi_j(\boldsymbol{\delta}) = \psi_j$ and $\pi_j(\boldsymbol{\delta}) = \pi_j$. Then we have $y_t = \sum_{j=0}^\infty \psi_j\varepsilon_{t-j}$ and $\sum_{j=0}^\infty \pi_j y_{t-j} = \varepsilon_t$. For the case of $\{y_t = 0,\ t \le 0\}$, we have $\sum_{j=0}^{t-1}\pi_j y_{t-j} = \varepsilon_t$ for $t > 0$ and:

$$\begin{aligned} y_t &= \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \sum_{j=0}^{t-1}\psi_j\varepsilon_{t-j} + \sum_{j=t}^{\infty}\psi_j\varepsilon_{t-j} = \sum_{j=0}^{t-1}\psi_j\sum_{k=0}^{t-j-1}\pi_k y_{t-j-k} + \sum_{j=t}^{\infty}\psi_j\varepsilon_{t-j} \\ &= \sum_{j=0}^{t-1}\left(\sum_{k=0}^{j}\psi_{j-k}\pi_k\right)y_{t-j} + \sum_{j=t}^{\infty}\psi_j\varepsilon_{t-j} = y_t + \sum_{j=t}^{\infty}\psi_j\varepsilon_{t-j}, \quad (\text{for } t > 0), \end{aligned}$$

where 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 $\sum_{k=0}^{j}\psi_{j-k}\pi_k = 0$ for $j \ge 1$. It follows that $y_t = \sum_{j=0}^{t-1}\psi_j\varepsilon_{t-j}$ because $\sum_{j=t}^\infty \psi_j\varepsilon_{t-j} = 0$, which implies the condition of $\{\varepsilon_t = 0,\ t \le 0\}$ for $y_t = \sum_{j=0}^\infty \psi_j\varepsilon_{t-j}$.

Conversely, for the case of $\{\varepsilon_t = 0,\ t \le 0\}$, $y_t = \sum_{j=0}^{t-1}\psi_j\varepsilon_{t-j}$ for $t > 0$ and:

$$\begin{aligned} \varepsilon_t &= \sum_{j=0}^{\infty}\pi_j y_{t-j} = \sum_{j=0}^{t-1}\pi_j y_{t-j} + \sum_{j=t}^{\infty}\pi_j y_{t-j} = \sum_{j=0}^{t-1}\pi_j\sum_{k=0}^{t-j-1}\psi_k\varepsilon_{t-j-k} + \sum_{j=t}^{\infty}\psi_j y_{t-j} \\ &= \sum_{j=0}^{t-1}\left(\sum_{k=0}^{j}\pi_{j-k}\psi_k\right)\varepsilon_{t-j} + \sum_{j=t}^{\infty}\pi_j y_{t-j} = \varepsilon_t + \sum_{j=t}^{\infty}\pi_j y_{t-j}, \quad (\text{for } t > 0), \end{aligned}$$

where we have used the fact that $\sum_{k=0}^{j}\pi_{j-k}\psi_k = 0$ for $j \ge 1$. It follows that $\varepsilon_t = \sum_{j=0}^{t-1}\pi_j y_{t-j}$ because $\sum_{j=t}^\infty \pi_j y_{t-j} = 0$, which implies the condition of $\{y_t = 0,\ t \le 0\}$ for $\sum_{j=0}^\infty \pi_j y_{t-j} = \varepsilon_t$.

O.6 Proof of Lemma 3.1 (p.8)

We only prove (3.15b) because the same argument is easily applied to (3.15a).

$$\mathrm{E}[x_{1,n}x_{2,n}z_{1,n}] = \sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\sum_{l=0}^{n-2}\sum_{m=1}^{n-l-1} \alpha_{1,j}\alpha_{2,k}\beta_{1,m}\,\mathrm{E}[\varepsilon_{n-j}\varepsilon_{n-k}\varepsilon_{n-l-m}\varepsilon_{n-l}].$$

By assumption, when $j = l+m \neq k = l$ and $j = l \neq k = l+m$, $\mathrm{E}[\varepsilon_{n-j}\varepsilon_{n-k}\varepsilon_{n-l-m}\varepsilon_{n-l}] = \sigma^4$ and zero, otherwise. Therefore, by using Cauchy–Schwarz inequality, we obtain:

$$\begin{aligned}\mathrm{E}[x_{1,n}x_{2,n}z_{1,n}] &= \sigma^4\sum_{l=0}^{n-2}\alpha_{2,l}\sum_{m=1}^{n-l-1}\alpha_{1,l+m}\beta_{1,m} + \sigma^4\sum_{l=0}^{n-2}\alpha_{1,l}\sum_{m=1}^{n-l-1}\alpha_{2,l+m}\beta_{1,m}\\ &= O\left(\sum_{l=0}^{n-2}\alpha_{2,l}\right) + O\left(\sum_{l=0}^{n-2}\alpha_{1,l}\right) = o(\sqrt{n}), \qquad \text{as } n\to\infty.\end{aligned}$$

O.7 Detailed arguments of Section 4 (p.11)

To obtain the general theory of the asymptotic PMSE of predictors, $x_n(h)$, $\widehat{x}_n(h)$, and $\widetilde{x}_n(h)$ in Section 4, we review the following results. Since technical arguments relating to asymptotic PMSE follow from those of Section 3, we almost omit the proof of the results given in this subsection.

To obtain the BLP, we use vector representation. Rewrite (4.1) as:

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{y}_t, \quad x_t - \boldsymbol{e}'\boldsymbol{A}\boldsymbol{x}_{t-1} = (1-L^s)^m x_t = y_t \qquad t \geq 1, \tag{O.4}$$

where $\boldsymbol{x}_t$, $\boldsymbol{y}_t$, and $\boldsymbol{e}$ are ms-vectors defined by $\boldsymbol{x}_t = (x_t, \ldots, x_{t-ms+1})'$, $\boldsymbol{y}_t = (y_t, 0, \ldots, 0)'$, and $\boldsymbol{e} = (1, 0, \ldots, 0)'$, respectively,

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & \cdots & \cdots & \cdots & a_{ms}\\ 1 & 0 & \cdots & \cdots & \cdots & 0\\ 0 & 1 & \ddots & & & \vdots\\ \vdots & \ddots & \ddots & \ddots & & \vdots\\ \vdots & & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & \cdots & 0 & 1 & 0\end{pmatrix}, \quad (1-z^s)^m = \sum_{j=0}^{m}\binom{m}{j}(-1)^j z^{js} = 1 - \sum_{j=1}^{ms} a_j z^j,$$

and $a_j = 0$ for $j \neq s, 2s, \ldots, ms$. Then we have:

$$\begin{aligned}\boldsymbol{x}_{n+h} &= \boldsymbol{A}\boldsymbol{x}_{n+h-1} + \boldsymbol{y}_{n+h} = \boldsymbol{A}^h\boldsymbol{x}_n + \sum_{j=0}^{h-1}\boldsymbol{A}^j\boldsymbol{y}_{n+h-j},\\ \text{and}\quad x_{n+h} &= \boldsymbol{e}'\boldsymbol{A}^h\boldsymbol{x}_n + \sum_{j=0}^{h-1}\alpha_j y_{n+h-j},\end{aligned}$$

where $\alpha_j = \boldsymbol{e}'\boldsymbol{A}^j\boldsymbol{e}$. As in Brockwell and Davis (1991, Section 9.5), given the data $\{x_t\}_{t=1-ms}^{n}$ from (4.1), the best linear predictor of x_{n+h} based on $x_{1-ms}, x_{2-ms}, \ldots, x_n$ is the projection $P_{S_n}x_{n+h}$ where $S_n = \overline{sp}\{x_{1-ms}, \ldots, x_0, y_1, \ldots, y_n\}$ and $\overline{sp}\{x_{1-ms}, \ldots, x_0\}\perp\overline{sp}\{y_1, \ldots, y_n\}$ by assumption. Hence the best linear predictor, denoted by $x_n(h)$, for a future value, x_{n+h}, is given by:

$$x_n(h) = \boldsymbol{e}'\boldsymbol{A}^h\boldsymbol{x}_n + \sum_{j=0}^{h-1}\alpha_j y_n(h-j), \tag{O.5}$$

where $y_n(k)$ for $k = 1, \ldots, h$ is defined by Section 3. Since the prediction error of $x_n(h)$ can be written in the form:

$$x_{n+h} - x_n(h) = \sum_{j=0}^{h-1}\alpha_j\{y_{n+h-j} - y_n(h-j)\} = \sum_{j=0}^{h-1}\left(\sum_{k=0}^{j}\alpha_{j-k}\psi_k(\boldsymbol{\delta}^0)\right)\varepsilon_{n+h-j},$$

where $\psi_j(\boldsymbol{\delta}^0)$ is given by (3.7), we can express its PMSE as:

$$\sigma_x^2(h) \equiv \mathrm{E}\Big[x_{n+h} - x_n(h)\Big]^2 = \sigma_0^2 \sum_{j=0}^{h-1} \left(\sum_{k=0}^{j} \alpha_{j-k}\psi_k(\boldsymbol{\delta}^0)\right)^2 = \sigma_0^2 \sum_{j=0}^{h-1} \psi_{j,x}^2, \tag{O.6}$$

where $\psi_{j,x} = \sum_{k=0}^{j} \alpha_{j-k}\psi_k(\boldsymbol{\delta}^0)$ is given by $\sum_{j=0}^{\infty} \psi_{j,x} z^j = (1-z^s)^{-m}(1-z^s)^{-d}\beta(z)$, $|z|<1$.

First, we deal with estimated predictors when all parameters are estimated, as in Section 3.1. We next deal with the effects of misspecification in non-stationary ARFISMA models as SARIMA models, as in Section 3.2.

O.7.1 PMSE for $\widehat{x}_n(h)$ when d is fixed

Similarly to Section 3.1, we define the predictor with estimated parameters. The parameters estimated here are $\boldsymbol{\delta}^0 = \big(d, \beta'\big)'$ and σ_0^2. These are estimated by maximizing the CSS function defined in Section 3.1. Let $\widehat{\boldsymbol{\delta}}$ be a corresponding CSS estimator of $\boldsymbol{\delta}^0$, then the predictor, denoted by $\widehat{x}_n(h)$, is defined by:

$$\widehat{x}_n(h) = \boldsymbol{e}'\boldsymbol{A}^h \boldsymbol{x}_n + \sum_{j=0}^{h-1} \alpha_j \widehat{y}_n(h-j), \tag{O.7}$$

where $\widehat{y}_n(k)$, $k = 1, \ldots, h$ is given by (3.12). Its asymptotic PMSE is immediately obtained. Since the error of $\widehat{x}_n(h)$ can be evaluated by $x_{n+h} - \widehat{x}_n(h) = \{x_{n+h} - x_n(h)\} - \{\widehat{x}_n(h) - x_n(h)\}$, we have:

$$\widehat{\sigma}_x^2(h) \equiv \mathrm{E}\Big[x_{n+h} - \widehat{x}_n(h)\Big]^2 = \sigma_x^2(h) + \mathrm{E}\Big[\widehat{x}_n(h) - x_n(h)\Big]^2, \tag{O.8}$$

and the second term on the RHS of the above equation is expressed as follows. We have already shown that, in the proof of Theorem 3.1, $\mathrm{E}[R_{1,n}]^2 = o(1/n)$ uniformly in $\boldsymbol{\delta}^* \in D_\delta$ and $\sum_{j=1}^{n} c_j^{(1)}(h, \boldsymbol{\delta}^0) y_{n+1-j} = \sum_{j=0}^{n-1} \boldsymbol{\varphi}_j(h, \boldsymbol{\delta}^0)\varepsilon_{n-j}$. It follows that:

$$\mathrm{E}\Big[\widehat{x}_n(h) - x_n(h)\Big]^2 = \mathrm{E}\left[\sum_{j=0}^{h-1} \alpha_j \{\widehat{y}_n(h-j) - y_n(h-j)\}\right]^2 \tag{O.9}$$

$$= \sum_{j,k=0}^{h-1} \alpha_j \alpha_k \, \mathrm{E}\left[\sum_{i,l=0}^{n-1} \boldsymbol{\varphi}_i(h-j, \boldsymbol{\delta}^0)'\varepsilon_{n-i}\Big(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\Big)\Big(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\Big)'\boldsymbol{\varphi}_l(h-k, \boldsymbol{\delta}^0)\varepsilon_{n-l}\right] + o\left(\frac{1}{n}\right).$$

Since the above equation follows the similar representation of (A.13) by using (A.12), we obtain the following theorem:

Theorem O.1 *Let $\{x_t\}_{t=1-ms}^{n}$ be given by* (4.1). *Then it follows that, as $n \to \infty$:*

$$\widehat{\sigma}_x^2(h) \equiv \mathrm{E}\Big[x_{n+h} - \widehat{x}_n(h)\Big]^2 \tag{O.10}$$

$$= \sigma_x^2(h) + \frac{\sigma_0^2}{n} \sum_{i=0}^{\infty} \sum_{j=0}^{h-1} \sum_{k=0}^{h-1} \alpha_j \alpha_k \boldsymbol{\varphi}_i(h-j, \boldsymbol{\delta}^0)' \boldsymbol{I}(\boldsymbol{\delta}^0)^{-1} \boldsymbol{\varphi}_i(h-k, \boldsymbol{\delta}^0)' + o\left(\frac{1}{n}\right),$$

where $\boldsymbol{\varphi}_i(k, \boldsymbol{\delta}^0)$ is defined by Theorem 3.1 for $k = 1, \ldots, h$, $\sigma_x^2(h)$ and $\widehat{x}_n(h)$ are given by (O.6) *and* (O.7), *respectively.*

For the case of $h = 1$, we can simplify the preceding result. The following corollary gives the asymptotic PMSE of $\widehat{x}_n(1) = \boldsymbol{e}'\boldsymbol{A}\boldsymbol{x}_n + \widehat{y}_n(1)$, where $\widehat{y}_n(1) = -\sum_{j=1}^{n} \pi_j(\widehat{\boldsymbol{\delta}}) y_{n+1-j}$.

Corollary O.1 *Under the same conditions as in Theorem O.1, it follows that, as $n \to \infty$:*

$$\widehat{\sigma}_x^2(1) \equiv \mathrm{E}\Big[x_{n+1} - \widehat{x}_n(1)\Big]^2 = \sigma_0^2\left(1 + \frac{p+q+1}{n}\right) + o\left(\frac{1}{n}\right). \tag{O.11}$$

The proof is omitted since its error is $x_{n+1} - \widehat{x}_n(1) = \sum_{j=0}^{n} \pi_j(\widehat{\boldsymbol{\delta}}) y_{n+1-j}$, the second moments of which can be obtained by Corollary 3.1.

O.7.2 PMSE for integrated long-memory processes when $d = c/\sqrt{n}$

We next consider the model (4.1) when $d = c/\sqrt{n}$. The predictors considered here are $\widehat{x}_n(h)$ discussed in Section O.7.1 and, similarly to $\widetilde{y}_n(h)$ in Section 3.2, $\widetilde{x}_n(h)$, which is given by the fixed differencing parameter m with $d = 0$ and estimators of ARMA(p, q) parameters, denoted by $\widetilde{\boldsymbol{\beta}}$, from the process $\{(1 - L^s)^m x_t\}$, $t \geq 1$. This is expressed as:

$$\widetilde{x}_n(h) = \boldsymbol{e}'\boldsymbol{A}^h \boldsymbol{x}_n + \sum_{j=0}^{h-1} \alpha_j \widetilde{y}_n(h-j), \tag{O.12}$$

where $\widetilde{y}_n(k)$ for $k = 1, \ldots, h$ is defined by (3.20) and $\widetilde{\boldsymbol{\delta}} = \left(0, \widetilde{\boldsymbol{\beta}}'\right)'$. Note that the expression for $\widetilde{x}_n(h)$ is a predictor of the ARIMA(p, m, q) model with estimated coefficients when $s = 1$, and is a corresponding predictor of the SARIMA model when s is even.

An asymptotic PMSE of $\widetilde{x}_n(h)$ is given by the following theorem.

Theorem O.2 *Let $\{x_t\}_{t=1-ms}^{n}$ be given by* (4.1), *$d = c/\sqrt{n}$, where c is a fixed constant. Then it follows that, as $n \to \infty$:*

$$\begin{aligned} \widetilde{\sigma}_x^2(h) &\equiv \mathrm{E}\left[x_{n+h} - \widetilde{x}_n(h)\right]^2 \\ &= \sigma_x^2(h) + \frac{\sigma_0^2}{n} \sum_{i=0}^{\infty} \sum_{j=0}^{h-1} \sum_{k=0}^{h-1} \alpha_j \alpha_k \boldsymbol{\varphi}_i(h-j, \boldsymbol{\delta}^0)' \boldsymbol{\Gamma} \boldsymbol{\varphi}_i(h-k, \boldsymbol{\delta}^0) + o\left(\frac{1}{n}\right), \end{aligned} \tag{O.13}$$

where $\boldsymbol{\varphi}_i(k, \boldsymbol{\delta}^0)$ is defined by Theorem 3.1 for $k = 1, \ldots, h$, $\widetilde{x}_n(h)$, $\sigma_x^2(h)$, and $\boldsymbol{\Gamma}$ is given by (O.12), (O.6), *and Theorem 3.2, respectively.*

We omit the proof since it can be obtained by Theorem 3.2.

The asymptotic PMSE of $\widehat{x}_n(h)$ is given by (O.10), which leads to the following result.

Corollary O.2 *Under the same conditions as in Theorem O.2, it follows that:*

$$\mathrm{ARE}\left[\widehat{x}_n(h),\ \widetilde{x}_n(h)\right] \begin{cases} \geq 0 & \textit{iff } \omega \geq |\,c\,|; \\ \leq 0 & \textit{iff } \omega \leq |\,c\,|; \end{cases}, \tag{O.14}$$

where ω is given by Corollary 3.2.

We omit the proof since it can be obtained similarly to the proof of Corollary 3.2.

Remark 2 The following arguments are applicable to Example 4.2 and DGP 2 in Section 5: Let s be an even integer and $h \leq s$. Then $x_{n+h} = \boldsymbol{e}'\boldsymbol{A}\boldsymbol{x}_{n+h-1} + y_{n+h}$:

$$\begin{aligned} x_n(h) &= \boldsymbol{e}'\boldsymbol{A}\boldsymbol{x}_{n+h-1} + y_n(h), & \sigma_x^2(h) &= \sigma_y^2(h), \\ \widehat{x}_n(h) &= \boldsymbol{e}'\boldsymbol{A}\boldsymbol{x}_{n+h-1} + \widehat{y}_n(h), & \widehat{\sigma}_x^2(h) &= \widehat{\sigma}_y^2(h), \\ \widetilde{x}_n(h) &= \boldsymbol{e}'\boldsymbol{A}\boldsymbol{x}_{n+h-1} + \widetilde{y}_n(h), & \widetilde{\sigma}_x^2(h) &= \widetilde{\sigma}_y^2(h), \end{aligned}$$

and we can compare $\widehat{x}_n(h)$ and $\widetilde{x}_n(h)$ on the basis of PMSEs by Corollaries 3.2 and 3.3 with $d_0 = 0$ because $a_1 = \cdots = a_{s-1} = 0$. Furthermore, when $\beta(z) = 1$, $\sigma_x^2(h) = \sigma_y^2(h) = \sigma_0^2$ by the definition of $\psi_j(d)$ in (3.5).

O.8 Proof of (A.7) when $d = d_0 + c/\sqrt{n}$ (p.21)

In Appendix B, referring to the results in Katayama (2006), we used the fact that (A.7) holds if fixed d is replaced by $d = d_0 + c/\sqrt{n}$ in the proof of results of Section 3.2. However, since Katayama

(2006) omit this proof, we prove this result. For simplicity, we focus on the case of $\beta(z)=1$. Let $d, d_0, d^*, d^{**} \in D_l$, $u_t(d^*)=\sum_{k=0}^{\infty}\pi_k(d^*)v_{t-k}$, $v_t=\sum_{k=0}^{\infty}\psi_k(d)\varepsilon_{t-k}$:

$$y_t=\sum_{j=0}^{t-1}\psi_j(d)\varepsilon_{t-j}, \quad \varepsilon_t(d^*)=\sum_{j=0}^{t-1}\pi_j(d^*)y_{t-j}, \quad u_t(d^*,d^{**})=\sum_{j=0}^{\infty}\pi_j(d^*)v_{t-j}(d^{**}),$$

$$v_t(d^*)=\sum_{j=0}^{\infty}\psi_j(d^*)\varepsilon_{t-j}=\sum_{j=0}^{t-1}\psi_j(d^*)\varepsilon_{t-j}+\sum_{j=t}^{\infty}\psi_j(d^*)\varepsilon_{t-j}=v_{1,t}(d^*)+v_{2,t}(d^*), \quad \text{(say)}.$$

Then, $y_t=v_{1,t}(d)$, $u_t(d^*)=u_t(d^*,d)$, $v_t=v_t(d)$, $\varepsilon_t=u_t(d)=u_t(d,d)=\varepsilon_t(d)$, and:

$$\begin{aligned}\varepsilon_t(d^*)&=\sum_{j=0}^{t-1}\pi_j(d^*)y_{t-j}=\sum_{j=0}^{t-1}\pi_j(d^*)v_{1,t-j}(d)=\sum_{j=0}^{t-1}\pi_j(d^*)\sum_{k=0}^{t-j-1}\psi_k(d)\varepsilon_{t-j-k}\\&=\sum_{j=0}^{t-1}\pi_j(d^*+c_l)\sum_{k=0}^{t-j-1}\psi_k(d+c_l)\varepsilon_{t-j-k}=\sum_{j=0}^{t-1}\pi_j(d^*+c_l)v_{1,t-j}(d+c_l),\end{aligned}$$

where $(c_1,c_2,c_3)=(0,1/4,1/2)$ and we have used the fact that $\sum_{j=0}^{t-1}\sum_{k=0}^{t-j-1}a_{k,j}=\sum_{j=0}^{t-1}\sum_{k=0}^{j}a_{k,j-k}$, $\pi_j(d)=\psi_j(-d)$, $\pi_j(a+b)=\sum_{k=0}^{j}\pi_k(a)\pi_{j-k}(b)$, and $\psi_j(a+b)=\sum_{k=0}^{j}\psi_k(a)\psi_{j-k}(b)$. [This technique of the proof is also used to prove the strong consistency and the asymptotic normality of the CSS estimators in Lemma B 9 of Katayama (2006)]. It follows that:

$$\varepsilon_t^{(i)}(d^*)=\sum_{j=0}^{t-1}\pi_j^{(i)}(d^*+c_l)v_{1,t-j}(d+c_l), \tag{O.15}$$

$$\begin{aligned}u_t^{(i)}(d^*)&=\left\{\frac{\partial^i}{\partial d^{*i}}(1-L^s)^{d^*}\right\}(1-L^s)^{-d}\varepsilon_t=\{\log(1-L^s)\}^i(1-L^s)^{d^*+c_l}(1-L^s)^{-d-c_l}\varepsilon_t \qquad \text{(O.16)}\\&=\sum_{j=0}^{t-1}\pi_j^{(i)}(d^*+c_l)v_{t-j}(d+c_l)+\sum_{j=t}^{\infty}\pi_j^{(i)}(d^*+c_l)v_{t-j}(d+c_l)\\&=\sum_{j=0}^{t-1}\pi_j^{(i)}(d^*+c_l)v_{1,t-j}(d+c_l)+\sum_{j=0}^{t-1}\pi_j^{(i)}(d^*+c_l)v_{2,t-j}(d+c_l)+\sum_{j=t}^{\infty}\pi_j^{(i)}(d^*+c_l)v_{t-j}(d+c_l)\\&=\varepsilon_t^{(i)}(d^*)+w_{1,t}^{(i)}(d^*)+w_{2,t}^{(i)}(d^*), \quad i=0,1,2, \quad \text{(say)}.\end{aligned}$$

Since $d^*+c_l, d+c_l \in D_1=[\alpha, 1/2-\alpha]$, $\alpha\in(0,1/4)$, it is enough to consider the case of D_1 and $c_1=0$. Note that $\pi_j(d^*)=\psi_j(-d^*)=\Gamma(j-d^*)/\{\Gamma(j+1)\Gamma(-d^*)\}$ for $s=1$:

$$\begin{aligned}\pi_j^{(1)}(d^*)&=-\psi(j-d^*)\pi_j(d^*)+\psi(-d^*)\pi_j(d^*),\\ \pi_j^{(2)}(d^*)&=\pi_j(d^*)\{\psi'(j-d^*)-\psi'(-d^*)\}-\pi_j^{(1)}(d^*)\{\psi(j-d^*)-\psi(-d^*)\},\end{aligned}$$

where $\psi(z)$ and $\psi'(z)$ is defined by $\psi(z)=\Gamma^{(1)}(z)/\Gamma(z)$ and $\psi'(z)=\psi^{(1)}(z)$, respectively. These are known as the Digamma function and Polygamma function, respectively, having the following properties: $\psi(z)=O(\log z)$ and $\psi'(z)=O(z^{-1})$, as $z\to\infty$. For any $d^*\in D_1$ and $s=1$:

$$\begin{aligned}&0<\psi_j(d^*)\le\Gamma(d^*)^{-1}j^{d^*-1}\le\text{const}\, j^{-\alpha-1/2}, \quad j\ge1, \qquad \text{(O.17)}\\&|\pi_j(d^*)|\le|\Gamma(-d^*)|^{-1}(j-1)^{-d^*-1}\le\text{const}(j-1)^{-\alpha-1}, \quad j\ge2,\end{aligned}$$

by Yajima (1985, (7)). It follows that there exist positive sequences $\{c_{\psi,j}(\alpha)\}$ and $\{c_{\pi,j}(\alpha)\}$, which depend on α, such that:

$$\begin{aligned}&\sup_{d^*\in D_1}|\psi_j^{(i)}(d^*)|\le c_{\psi,j}(\alpha)=O(j^{-\alpha-1/2}), \qquad \text{(O.18)}\\&\sup_{d^*\in D_1}|\pi_j^{(i)}(d^*)|\le c_{\pi,j}(\alpha)=O(j^{-\alpha-1}), \quad i=0,1,2,\end{aligned}$$

as $j \to \infty$, where we have used the fact that $\lim_{j\to\infty} \log j/j^{\epsilon} = 0$, $\epsilon > 0$.

If one can show that:

$$\sup_{d^*\in D_1} \left| \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^{(i)}(d^*)\varepsilon_t^{(j)}(d^*) - \frac{1}{n}\sum_{t=1}^{n} u_t^{(i)}(d^*)u_t^{(j)}(d^*) \right| \xrightarrow{a.c.} 0, \tag{O.19}$$

$$\sup_{d^*\in D_1} \left| \frac{1}{n}\sum_{t=1}^{n} u_t^{(i)}(d^*)u_t^{(j)}(d^*) - \mathrm{E}\left[u_t^{(i)}(d^*)u_t^{(j)}(d^*)\right] \right| \xrightarrow{a.c.} 0, \tag{O.20}$$

for $(i,j) = (1,0),(1,1),(2,0)$, then the result follows. We first consider (O.19). For any positive integer r:

$$\mathrm{E}[v_t(d)]^{2r} \leq \text{const}\left(\sum_{k=2}^{2r}\sum_{j=0}^{\infty} |\psi_j(d)|^k\right)^r \leq \text{const}\left(\sum_{k=2}^{2r}\sum_{j=0}^{\infty} c_{\psi,j}(\alpha)^k\right)^r < \infty, \tag{O.21}$$

$$\mathrm{E}[v_{1,t}(d)]^{2r} \leq \text{const}\left(\sum_{k=2}^{2r}\sum_{j=0}^{t-1} |\psi_j(d)|^k\right)^r \leq \text{const}\left(\sum_{k=2}^{2r}\sum_{j=0}^{\infty} c_{\psi,j}(\alpha)^k\right)^r < \infty,$$

$$\mathrm{E}[v_{2,t}(d)]^{2r} \leq \text{const}\left(\sum_{k=2}^{2r}\sum_{j=t}^{\infty} |\psi_j(d)|^k\right)^r \leq \text{const}\left(\sum_{k=2}^{2r}\sum_{j=t}^{\infty} c_{\psi,j}(\alpha)^k\right)^r = O(t^{-2\alpha r}),$$

as $t \to \infty$. Note that these inequalities also hold if d is fixed. Using (O.18), (O.21), and Jensen's Theorem, we obtain:

$$\mathrm{E}\left[\sup_{d^*\in D_1}\left|\varepsilon_t^{(i)}(d^*)\right|\right]^r \leq \mathrm{E}\left[\sum_{j=0}^{t-1} c_{\pi,j}(\alpha)|v_{1,t-j}(d)|\right]^r \leq \text{const}\left(\sum_{j=0}^{\infty} c_{\pi,j}(\alpha)\right)^r < \infty, \tag{O.22}$$

$$\mathrm{E}\left[\sup_{d^*\in D_1}\left|w_{1,t}^{(i)}(d^*)\right|\right]^{2r} \leq \text{const}\sum_{j=0}^{t-1}\sup_{d^*\in D_1}|\pi_j^{(i)}(d^*)|\,\mathrm{E}\left[v_{2,t-j}(d)\right]^{2r}$$
$$\leq \text{const}\sum_{j=0}^{[t/2]} c_{\pi,j}(\alpha)\,\mathrm{E}\left[v_{2,t-j}(d)\right]^{2r} + \text{const}\sum_{j=[t/2]+1}^{t-1} c_{\pi,j}(\alpha)\,\mathrm{E}\left[v_{2,t-j}(d)\right]^{2r}$$
$$= O\left(t^{-2\alpha r}\sum_{j=0}^{[t/2]} c_{\pi,j}(\alpha)\right) + O\left(t^{-\alpha-1}\sum_{j=1}^{t} j^{-2\alpha r}\right) = O\big(t^{-\alpha}\big),$$

$$\mathrm{E}\left[\sup_{d^*\in D_1}\left|w_{2,t}^{(i)}(d^*)\right|\right]^r \leq \mathrm{E}\left[\sum_{j=t}^{\infty} c_{\pi,j}(\alpha)|v_{t-j}(d)|\right]^r \leq \left(\sum_{j=t}^{\infty} j^{-\alpha-1}\right)^r = O\big(t^{-\alpha r}\big),$$

$i = 0,1,2$, as $t \to \infty$. Note that these inequalities also hold if d is fixed. Since:

$$\frac{1}{n}\sum_{t=1}^{n} u_t^{(i)}(d^*)u_t^{(j)}(d^*) - \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^{(i)}(d^*)\varepsilon_t^{(j)}(d^*)$$
$$= \frac{1}{n}\sum_{t=1}^{n}\big\{\varepsilon_t^{(i)}(d^*)w_{1,t}^{(j)}(d^*) + \varepsilon_t^{(i)}(d^*)w_{2,t}^{(j)}(d^*) + w_{1,t}^{(i)}(d^*)\varepsilon_t^{(j)}(d^*) + w_{1,t}^{(i)}(d^*)w_{1,t}^{(j)}(d^*)$$
$$+ w_{1,t}^{(i)}(d^*)w_{2,t}^{(j)}(d^*) + w_{2,t}^{(i)}(d^*)\varepsilon_t^{(j)}(d^*) + w_{2,t}^{(i)}(d^*)w_{1,t}^{(j)}(d^*) + w_{2,t}^{(i)}(d^*)w_{2,t}^{(j)}(d^*)\big\},$$

following from Katayama (2006, Lemma B 3) and (O.22), we obtain (O.19).

To prove (O.20), by a Taylor series expansion:

$$v_t(d) = (1-L^s)^{-d_0-c/\sqrt{n}}\varepsilon_t = (1-L^s)^{-d_0}\varepsilon_t - \frac{c}{\sqrt{n}}\log(1-L^s)(1-L^s)^{-d_0-c^*/\sqrt{n}}\varepsilon_t$$
$$= v_t(d_0) + \omega_{1,t}^*/\sqrt{n},$$

where $\omega^*_{1,t} = -c\log(1-L^s)(1-L^s)^{-d_0-c^*/\sqrt{n}}\varepsilon_t$ and $0<|c^*|<|c|$, we can rewrite $u_t^{(i)}(d^*) = u_t^{(i)}(d^*,d)$ as:

$$u_t^{(i)}(d^*,d) = \sum_{j=0}^{\infty}\pi_j^{(i)}(d^*)v_{t-j}(d_0) + \frac{1}{\sqrt{n}}\sum_{j=0}^{\infty}\pi_j^{(i)}(d^*)\omega^*_{1,t-j} = u_t^{(i)}(d^*,d_0) + \frac{1}{\sqrt{n}}\omega_{2,t}^{(i)}(d^*), \qquad \text{(say)}.$$

Since $-\log(1-L^s)(1-L^s)^{-d} = \sum_{j=0}^{\infty}\{\sum_{k=0}^{j}\psi_{j-k}(d)s_{k+1}\}L^{j+1}$, and:

$$\left|\sum_{k=0}^{j}\psi_{j-k}(d_0+c^*/\sqrt{n})s_{k+1}\right| \le \text{const}\left\{\sum_{k=0}^{[j/2]}c_{\psi,j-k}(\alpha)s_{k+1} + \sum_{k=[j/2]+1}^{j}c_{\psi,j-k}(\alpha)s_{k+1}\right\}$$
$$= O\left(j^{-\alpha-1/2}\sum_{k=0}^{[j/2]}(k+1)^{-1} + j^{-1}\sum_{k=0}^{j}c_{\psi,j}(\alpha)\right) = O\Big(j^{-\alpha-1/2}\log j\Big),$$

as $j\to\infty$, we obtain, $\mathrm{E}[\omega^*_{1,t}]^{2r} \le \text{const}[\sum_{k=2}^{2r}(\sum_{j=1}^{\infty}\{j^{-\alpha-1/2}\log j\}^k+1)]^r < \infty$. It follows that:

$$\mathrm{E}\left[\sup_{d^*\in D_1}\left|\omega_{2,t}^{(i)}(d^*)\right|\right]^r \le \text{const}\left(\sum_{j=0}^{\infty}c_{\pi,j}(\alpha)\right)^r < \infty, \tag{O.23}$$
$$\mathrm{E}\left[\sup_{d^*\in D_1}\left|u_t^{(i)}(d^*,d_0)\right|\right]^r \le \text{const}\left(\sum_{j=0}^{\infty}c_{\pi,j}(\alpha)\right)^r < \infty, \quad i=0,1,2.$$

For the LHS of (O.20), we have:

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^{n}u_t^{(i)}(d^*,d)u_t^{(j)}(d^*,d) - \mathrm{E}\left[u_t^{(i)}(d^*,d)u_t^{(j)}(d^*,d)\right]\right| \qquad\qquad (\text{O.24})\\
&\le \left|\frac{1}{n}\sum_{t=1}^{n}u_t^{(i)}(d^*,d_0)u_t^{(j)}(d^*,d_0) - \mathrm{E}\left[u_t^{(i)}(d^*,d_0)u_t^{(j)}(d^*,d_0)\right]\right|\\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\left|u_t^{(i)}(d^*,d_0)\right|\left|\omega_{2,t}^{(j)}(d^*)t^{-1/2}\right| + \frac{1}{n}\sum_{t=1}^{n}\left|\omega_{2,t}^{(i)}(d^*)t^{-1/2}\right|\left|u_t^{(j)}(d^*,d_0)\right|\\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\left|\omega_{2,t}^{(i)}(d^*)t^{-1/2}\right|\left|\omega_{2,t}^{(j)}(d^*)t^{-1/2}\right| + \left|\mathrm{E}\left[u_t^{(i)}(d^*,d_0)u_t^{(j)}(d^*,d_0)\right] - \mathrm{E}\left[u_t^{(i)}(d^*,d)u_t^{(j)}(d^*,d)\right]\right|.
\end{aligned}$$

The RHS of the above equation converges almost certainly to zero uniformly in $d^*\in D_1$, which proves (O.20). This is because the first term follows from Lemma B 8 of Katayama (2006), the second, third, and fourth terms follow from (O.23) and Lemma B 3 of Katayama (2006), and the last term follows from (O.23) and the fact that:

$$\begin{aligned}
&\sup_{d^*\in D_1}\left|\mathrm{E}\left[u_t^{(i)}(d^*,d_0)u_t^{(j)}(d^*,d_0)\right] - \mathrm{E}\left[u_t^{(i)}(d^*,d)u_t^{(j)}(d^*,d)\right]\right|\\
&\le \sup_{d^*\in D_1}\mathrm{E}\left[\left|u_t^{(i)}(d^*,d_0)\frac{\omega_{2,t}^{(j)}(d^*)}{\sqrt{n}} + \frac{\omega_{2,t}^{(i)}(d^*)}{\sqrt{n}}u_t^{(j)}(d^*,d_0) + \frac{\omega_{2,t}^{(i)}(d^*)}{\sqrt{n}}\frac{\omega_{2,t}^{(j)}(d^*)}{\sqrt{n}}\right|\right] \longrightarrow 0, \quad \text{as } n\to\infty.
\end{aligned}$$

O.9 Proof of convergence with probability one and convergence in the rth mean for $R_{2,n}$ (either fixed d or $d = d_0 + c/\sqrt{n}$) , $R_{4,n}$ ($d = d_0 + c/\sqrt{n}$) and (A.18) ($d = d_0 + c/\sqrt{n}$) [(A.12), p.23; (A.23), p.25; (A.18), p.25]

We prove the following convergence of the rth mean:

$$\lim_{n\to\infty} \mathrm{E}\left|\left[\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^{**})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} - \boldsymbol{I}(\boldsymbol{\delta}^0)\right]_{i,j}\right|^r = 0, \qquad \text{(either fixed } d \text{ or } d = d_0 + c/\sqrt{n}), \tag{O.25}$$

$$\lim_{n\to\infty} \mathrm{E}\left[\left\|\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^{*})}{\partial d\partial\boldsymbol{\beta}} - \boldsymbol{\kappa}\right\|\right]^r = 0, \qquad (d = d_0 + c/\sqrt{n}),$$

$$\lim_{n\to\infty} \mathrm{E}\left|\left[\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^{*})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} - \boldsymbol{\Phi}\right]_{i,j}\right|^r = 0, \qquad (d = d_0 + c/\sqrt{n}),$$

for $r > 1$, where $[\boldsymbol{A}]_{i,j}$ denotes (i,j) element of a matrix $\boldsymbol{A}$, $\|\boldsymbol{\delta}^{**} - \boldsymbol{\delta}^0\| \le \|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\|$, and $\|\boldsymbol{\delta}^{*} - \boldsymbol{\delta}^0\| \le \|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\|$. By Katayama (2006, Theorem 1, Remarks 2 and 3, Lemmas B 7, B 8, and B 9) and Section O.8, $\widehat{\boldsymbol{\delta}} \xrightarrow{a.c.} \boldsymbol{\delta}^0$, $\widetilde{\boldsymbol{\delta}} \xrightarrow{a.c.} \boldsymbol{\delta}^0$, $\boldsymbol{I}(\boldsymbol{\delta})$ is continuous on $\boldsymbol{\delta}$, and (A.7) holds either fixed d or $d = d_0 + c/\sqrt{n}$. It follows that:

$$\begin{aligned}
&\left|S_\varepsilon^{(2)}(\boldsymbol{\delta}^{**}) - \boldsymbol{I}(\boldsymbol{\delta}^0)\right| \\
&\le \left|S_\varepsilon^{(2)}(\boldsymbol{\delta}^{**}) - S_u^{(2)}(\boldsymbol{\delta}^{*})\right| + \left|S_u^{(2)}(\boldsymbol{\delta}^{**}) - \boldsymbol{I}(\boldsymbol{\delta}^{**})\right| + \left|\boldsymbol{I}(\boldsymbol{\delta}^{**}) - \boldsymbol{I}(\boldsymbol{\delta}^0)\right| \\
&\le \sup_{\boldsymbol{\delta}\in D_\delta}\left|S_\varepsilon^{(2)}(\boldsymbol{\delta}) - S_u^{(2)}(\boldsymbol{\delta})\right| + \sup_{\boldsymbol{\delta}\in D_\delta}\left|S_u^{(2)}(\boldsymbol{\delta}) - \boldsymbol{I}(\boldsymbol{\delta})\right| + \left|\boldsymbol{I}(\boldsymbol{\delta}^{**}) - \boldsymbol{I}(\boldsymbol{\delta}^0)\right| \\
&\xrightarrow{a.c.} \mathbf{0} \quad \text{(either fixed } d \text{ or } d = d_0 + c/\sqrt{n}).
\end{aligned}$$

Similarly:

$$\left|\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^{*})}{\partial d\partial\boldsymbol{\beta}} - \boldsymbol{\kappa}\right| \xrightarrow{a.c.} \mathbf{0}, \qquad \text{and} \qquad \left|\frac{\partial^2 S_\varepsilon(\boldsymbol{\delta}^{*})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} - \boldsymbol{\Phi}\right| \xrightarrow{a.c.} \mathbf{0} \qquad (d = d_0 + c/\sqrt{n}).$$

To show (O.25), it is sufficient to check the rth moments of second derivatives of $S_\varepsilon(\boldsymbol{\delta}^*)$ for any $r > 1$ [see Lemma A (ii) and Lemma C (i) of Serfling (1980, pp.13–15)]. We focus on the case of $\beta(z) = 1$ because coefficients of partial derivatives consisting of $\boldsymbol{\beta}$ are absolutely summable and decay exponentially and the same argument is easily applied to the general $\{y_t\}$. For $d, d^* \in D_l$, $l = 1, 2, 3$:

$$S_\varepsilon^{(2)}(d^*) = \frac{1}{n\sigma_0^2}\sum_{t=1}^n \varepsilon_t^{(2)}(d^*)\varepsilon_t(d^*) + \frac{1}{n\sigma_0^2}\sum_{t=1}^n \varepsilon_t^{(1)}(d^*)\varepsilon_t^{(1)}(d^*),$$

where $\varepsilon_t^{(i)}(d^*)$ is given by (O.15). For (O.15), since $d^* + c_l, d + c_l \in D_1 = [\alpha, 1/2 - \alpha]$, $\alpha \in (0, 1/4)$, it is enough to consider the case of D_1 and $c_1 = 0$. To check $\mathrm{E}[S_\varepsilon^{(2)}(d^*)]^r$, from c_r-inequality, it is sufficient to show that:

$$\mathrm{E}\left|\frac{1}{n}\sum_{t=1}^n \varepsilon_t^{(i)}(d^*)\varepsilon_t^{(j)}(d^*)\right|^r < \infty, \tag{O.26}$$

for $(i,j) = (2,0), (1,1)$, uniformly in $d^*, d \in D_1$. By the convexity of function x^r, $x > 0$, $r > 1$, Jensen's Theorem, and Cauchy–Schwarz inequality:

$$\begin{aligned}
\mathrm{E}\left|\frac{1}{n}\sum_{t=1}^n \varepsilon_t^{(i)}(d^*)\varepsilon_t^{(j)}(d^*)\right|^r &\le \mathrm{E}\left[\frac{1}{n}\sum_{t=1}^n \left|\varepsilon_t^{(i)}(d^*)\right|\left|\varepsilon_t^{(j)}(d^*)\right|\right]^r \le \frac{1}{n}\sum_{t=1}^n \mathrm{E}\left[\left|\varepsilon_t^{(i)}(d^*)\right|^r\left|\varepsilon_t^{(j)}(d^*)\right|^r\right] \\
&\le \frac{1}{n}\sum_{t=1}^n \sqrt{\mathrm{E}\left[\left|\varepsilon_t^{(i)}(d^*)\right|\right]^{2r}\mathrm{E}\left[\left|\varepsilon_t^{(j)}(d^*)\right|\right]^{2r}},
\end{aligned}$$

which proves (O.26) from (O.22).

O.10 Proof of (A.21) (p.25)

For simplicity, let $\beta(z) = 1$, $\widetilde{\boldsymbol{\delta}} = \widetilde{d}$. Then, using the similar argument as in (3.13), (3.14), (3.16), and (A.11), we obtain:

$$\begin{aligned}
\mathrm{E}\left[y_{n+h} - \widetilde{y}_n(h)\right]^2 &= \mathrm{E}\left[y_{n+h} - y_n(h) + y_n(h) - \widetilde{y}_n(h)\right]^2 = \mathrm{E}\left[\sum_{j=0}^{h-1} \psi_j(d)\varepsilon_{n+h-j} + y_n(h) - \widetilde{y}_n(h)\right]^2 \\
&= \sigma_y^2(h) + \mathrm{E}\left[y_n(h) - \widetilde{y}_n(h)\right]^2 = \sigma_y^2(h) + \mathrm{E}\left[\sum_{j=1}^{n}\left\{c_j(h,\widetilde{d}) - c_j(h,\boldsymbol{\delta}^0)\right\} y_{n+1-j}\right]^2 \\
&= \sigma_y^2(h) + \mathrm{E}\left[(\widetilde{d} - d)\sum_{j=0}^{n-1} \boldsymbol{\varphi}_j(h,d)\varepsilon_{n-j} + R_{5,n}\right]^2,
\end{aligned}$$

where $R_{5,n} = (\widetilde{d} - d)^2 \sum_{j=1}^{n} c_j^{(2)}(h, d^*) y_{n+1-j}$, and $|d^* - d| \leq |\widetilde{d} - d|$. It follows that it is sufficient to show that $\mathrm{E}[R_{5,n}]^2 = o(n^{-1})$, as $n \to \infty$.

Note that, by Yajima (1985, (7)) and Lemma 3.2, there exists $a_{j,l}(\alpha)$, which does not depend on d and d^*, satisfies:

$$\sup_{d^* \in D_l} \left| c_j^{(2)}(h, d^*) \right| \leq a_{j,l}(\alpha) = O\left(\{\log j\}^2 j^{-\alpha + c_l - 1}\right), \tag{O.27}$$

for $l = 1, 2, 3$, as $j \to \infty$, where $(c_1, c_2, c_3) = (0, 1/4, 1/2)$. Furthermore, by Yajima (1985, (7)), $|\psi_j(d)| \leq Cj^{-c_l - \alpha - 1/2}$ for $j \geq 1$, and:

$$\begin{aligned}
\mathrm{E}[y_t]^4 = \mathrm{E}\left[\sum_{j=0}^{t-1} \psi_j(d)\varepsilon_{t-j}\right]^4 &\leq \mathrm{const}\left[\left\{\sum_{j=0}^{t-1} \psi_j(d)^2\right\}^2 + \sum_{j=0}^{t-1} \psi_j(d)^4\right] \\
&\leq \mathrm{const}\left[\left\{\sum_{j=1}^{t-1} j^{-2\alpha-1}\right\}^2 + \sum_{j=1}^{t-1} j^{-4\alpha-2}\right] + \mathrm{const} < \infty,
\end{aligned} \tag{O.28}$$

for $t \geq 2$, which implies that $\mathrm{E}[y_t]^4$ is finite for any $d \in D_l$, $l = 1, 2, 3$ and $t \geq 1$.

For the case of $d \in D_3$, as in the proof of $\mathrm{E}[R_{1,n}]^2 = o(n^{-1})$ in the proof of Theorem 3.1, using (A.14), (O.27), (O.28), Cauchy–Schwarz inequality, we have:

$$\begin{aligned}
\mathrm{E}\left(\sum_{j=1}^{n} c_j^{(2)}(h, d^*) y_{n+1-j}\right)^4 &\leq \mathrm{E}\left(\sum_{j=1}^{n} a_{j,3}(\alpha) y_{n+1-j}\right)^4 \leq \left[\sum_{j=1}^{n} a_{j,3}(\alpha)\left(\mathrm{E}[y_{n+1-j}]^4\right)^{1/4}\right]^4 \\
&= O\left(\{\log n\}^4 n^{2-4\alpha}\right), \\
\mathrm{E}[R_{5,n}]^2 &\leq \left[\mathrm{E}\left|\widetilde{d} - d\right|^8 \mathrm{E}\left(\sum_{j=1}^{n} c_j^{(2)}(h, d^*) y_{n+1-j}\right)^4\right]^{1/2} = O\left(\sqrt{n^{-4} \times \{\log n\}^4 n^{2-4\alpha}}\right) \\
&= o(n^{-1}), \qquad \text{as } n \to \infty.
\end{aligned}$$

The cases of D_1 and D_2 can be obtained similarly because the order of sequences corresponding to $a_{j,l}(\alpha)$ is $o(\{\log j\}^2 j^{-1/2-\alpha})$. The case of the ARFISMA(p, d, q) model can be obtained similarly because the coefficients of the partial derivatives consisting of $\boldsymbol{\beta}$ are absolutely summable and decay exponentially.

O.11 Proof of Corollary 3.2 (p.10 and p.26)

We prove that Theorem 3.1 still hold if the model (1.1) is replaced by the model (3.19) and $\theta = c/\sqrt{n}$.

We borrow the notations of the proof of Theorem 3.1. Equations (A.19), (O.27) and (O.28) implies $\mathrm{E}[R_{1,n}]^2 = o(n^{-1})$ uniformly in $\boldsymbol{\delta}^* \in D_\delta$. Also, (3.8), (3.9), and (3.10) hold by Katayama (2006, Remark 3), and (A.19) and (O.25) hold which follows (A.11), (A.12), and (A.13). The remainder of the proof is obvious from the fact that $\boldsymbol{\delta}_j$ does not depend on d, $\varphi_j(h, \boldsymbol{\delta}^0) = O(\boldsymbol{\delta}_j)$ as $j \to \infty$, and the last paragraph of the proof of Theorem 3.1.