Inverse moment bounds for sample autocovariance matrices based on detrended time series and their applications

Tzu-Chang F. Cheng, Ching-Kang Ing,
Shu-Hui Yu
University of Illinois at Urbana-Champaign, Academia Sinica,
National University of Kaohsiung

Abstract

In this paper, we assume that observations are generated by a linear regression model with short- or long-memory dependent errors. We establish inverse moment bounds for k_n-dimensional sample autocovariance matrices based on the least squares residuals (also known as the detrended time series), where $k_n \ll n$, $k_n \to \infty$ and n is the sample size. These results are then used to derive the mean-square error bounds for the finite predictor coefficients of the underlying error process. Based on the detrended time series, we further estimate the inverse of the n-dimensional autocovariance matrix, R_n^{-1}, of the error process using the banded Cholesky factorization. By making use of the aforementioned inverse moment bounds, we obtain the convergence of moments of the difference between the proposed estimator and R_n^{-1} under spectral norm.

AMS 2000 subject classifications: Primary 15A09 ; secondary 15A52 , 62M10

KEY WORDS: Regression model with time series errors; Inverse moment bounds; Sample autocovariance matrix; Detrended time series; Moment convergence; Banded Cholesky factorization.

1 INTRODUCTION

Consider a linear regression model with serially correlated errors,

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + z_t = \sum_{i=1}^{p} x_{ti}\beta_i + z_t, \qquad t = 1, \cdots, n, \tag{1.1}$$

where $\mathbf{x}_t$'s are p-dimensional nonrandom input vectors, $\boldsymbol{\beta}$ is a p-dimensional unknown coefficient vector, and $\{z_t\}$ is an unobserved stationary process with mean zero. Let $\check{\mathbf{x}}_j = (x_{1j}, \cdots, x_{nj})'$, $1 \le j \le p$, and $\mathbf{y} = (y_1, \cdots, y_n)'$. It is natural to estimate $\mathbf{z} = (z_1, \cdots, z_n)'$ via the least squares residuals

$$\hat{\mathbf{z}} = (\hat{z}_1, \cdots, \hat{z}_n)' = (I - M_p)\mathbf{y} = (I - M_p)\mathbf{z}, \tag{1.2}$$

where M_p is the orthogonal projection matrix of $\overline{\mathrm{sp}}\{\check{\mathbf{x}}_1, \cdots, \check{\mathbf{x}}_p\}$, the closed span of $\{\check{\mathbf{x}}_1, \cdots, \check{\mathbf{x}}_p\}$. Note that $\hat{\mathbf{z}}$ is also known as the detrended time series; see Chapter 1 of Brockwell and Davis (1991). Since many time series data are only available after being detrended, investigations for some commonly used statistics based on $\hat{\mathbf{z}}$ are quite relevant. Let $\{\mathbf{o}_i = (o_{1i}, \cdots, o_{ni})', i = 1, \cdots, r\}$,

$1 \le r \le p$, be an orthonormal basis of $\overline{\text{sp}}\{\check{\mathbf{x}}_1, \cdots, \check{\mathbf{x}}_p\}$. It is well-known that $M_p = \sum_{i=1}^{r} \mathbf{o}_i \mathbf{o}_i'$, and hence

$$\hat{\mathbf{z}} = \mathbf{z} - \sum_{i=1}^{r} \nu_{in} \mathbf{o}_i, \tag{1.3}$$

where $\nu_{in} = \mathbf{o}_i' \mathbf{z}$. Denote by $\lambda_{\min}(A)$ the minimum eigenvalue of the symmetric matrix A. The main objective of this paper is to establish

$$E\left\{\lambda_{\min}^{-q}\big(\hat{R}(k)\big)\right\} := E\left\{\lambda_{\min}^{-q}\big(\frac{1}{n}\sum_{t=k}^{n} \hat{\mathbf{z}}_t(k)\hat{\mathbf{z}}_t'(k)\big)\right\} = O(1), \tag{1.4}$$

where q is some positive number, $\hat{\mathbf{z}}_t(k) = (\hat{z}_t, \cdots, \hat{z}_{t-k+1})'$, $\{z_t\}$ can be a short- or long-memory process, and $k = k_n$ is allowed to tend to ∞ as $n \to \infty$. By (1.3),

$$\hat{\mathbf{z}}_t(k) = \mathbf{z}_t(k) - \sum_{i=1}^{r} \nu_{in} \mathbf{o}_t^{(i)}(k), \tag{1.5}$$

where $\mathbf{z}_t(k) = (z_t, \cdots, z_{t-k+1})'$ and $\mathbf{o}_t^{(i)}(k) = (o_{t,i}, \cdots, o_{t-k+1,i})'$. When the time trend vectors $\mathbf{x}_t$ are known to be zero for all $1 \le t \le n$ (which yields $\mathbf{y} = \mathbf{z} = \hat{\mathbf{z}}$ and $\hat{\mathbf{z}}_t(k) = \mathbf{z}_t(k)$ for all $k \le t \le n$), (1.4) plays a major role in developing estimation, prediction and model selection theories of $\{z_t\}$. For example, assuming that $\{z_t\}$ is a stationary gaussian autoregressive (AR) model of finite order, Fuller and Hasza (1981) established (1.4) with k being a fixed positive integer. They further applied this result to analyze the biases and mean squared errors of the least squares estimators of the AR coefficients, and to provide an asymptotic expression for the mean squared prediction error of the corresponding least squares predictor. In order to establish some rigorous prediction and model selection theories, Bhansali and Papangelou (1991), Findley and Wei (2002), and Chan and Ing (2011), respectively, extended Fuller and Hasza's (1981) result to non-Gaussian AR models of finite order, multivariate time series models, and nonlinear stochastic regression models. All these results, however, require that k is fixed with n. On the other hand, Ing and Wei (2003) allowed that $k = k_n$ approaches ∞ at a suitable rate and established (1.4) under a class of short-memory processes. Their result was then used by a number of authors to deal with prediction and model selection/averaging problems in some misspecified time series models; see, e.g., Ing and Wei (2005), Zhang, Wan and Zou (2013) and Greenaway-McGrevy (2013).

If $\mathbf{x}_t$ are nonzero for some $1 \le t \le n$, then $\mathbf{z}$ is in general not observable and $\hat{\mathbf{z}}$ can be used in place of $\mathbf{z}$ to conduct statistical inferences for $\{z_t\}$. As shown in Section 3, some of these inferences can also be justified theoretically using (1.4) with $\hat{\mathbf{z}}_t(k) \ne \mathbf{z}_t(k)$, whose proof, however, is still lacking due to technical difficulties. In this article, we shall fill this gap by extending Ing and Wei's (2003) argument to the case of $\hat{\mathbf{z}}_t(k) \ne \mathbf{z}_t(k)$; see Theorems 1 and 2 of Section 2. Note that one major step in Ing and Wei's (2003) proof is to show that for any $\mathbf{v} \in R^k$ with $\|\mathbf{v}\| = 1$, the conditional distribution of $\mathbf{v}'\mathbf{z}_t(k)$ given $\sigma(\varepsilon_s, \varepsilon_{s-1}, \ldots)$ is sufficiently smooth when $t - s$ is sufficiently large and the distribution of ε_t obeys a type of Lipschitz condition, where $\|\cdot\|$ denotes the Euclidean norm, $\{\varepsilon_t\}$ is the white noise process driving $\{z_t\}$ and $\sigma(\varepsilon_s, \varepsilon_{s-1}, \ldots)$ is the σ-field generated by $\{\varepsilon_l, -\infty < l \le s\}$. In view of (1.5), the corresponding property in the present case is that the conditional distribution of $\mathbf{v}'\hat{\mathbf{z}}_t(k)$ given $\sigma(z_s, \ldots z_1, \nu_{1n}, \ldots \nu_{rn})$ is sufficiently smooth when $t - s$ is sufficiently large. However, since $\{\nu_{1n}, \ldots \nu_{rn}\}$ are determined by all $z_1, \ldots z_n$, we need a new distributional assumption on $\{z_t\}$, (C1), to establish this property. In fact, even if (C1) is imposed, this property may still fail to hold once $|\mathbf{v}'\mathbf{o}_t^{(j)}(k)|$ or $\sum_{i=1}^{s} o_{ij}^2$ is large for some $1 \leq j \leq r$; see Appendix. Fortunately, we argue in the proof of Theorem 1 that the number of such (t, s) pairs is small, and hence Ing and Wei's (2003) approach can be carried over to the present case after a suitable adjustment. Moreover, by making use of Theorem 2.1 of Ing and Wei (2006), we derive sharp upper bounds for $E(\nu_{in}^2), 1 \leq i \leq r$ under a mild dependence assumption, (2.1), on $\{z_t\}$. This enables us to relax the short-memory assumption made by Ing and Wei (2003).

The rest of the paper is organized as follows. In Section 2, we begin with introducing (2.1), (C1) and the other two conditions, (C2) and (C3), which place mild restrictions on the design matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ and the spectral density of $\{z_t\}$, respectively. Under these assumptions, Theorem 1 provides an upper bound for $E(\lambda_{\min}^{-q}(\hat{R}(k)))$, with $k = k_n = o(n)$ and $\hat{\mathbf{z}}_t(k) \neq \mathbf{z}_t(k)$. While this bound still approaches ∞ as k_n does, it is a stepping stone on the path to proving (1.4) when z_t is assumed to be a linear process driven by a white noise with finite moments up to a certain order. For more details, see Theorem 2. In Section 3, we assume that z_t admits an AR(∞) representation and give two interesting applications of Theorem 2. The first application is devoted to the mean-square error bounds for the finite predictor coefficients obtained from $\hat{\mathbf{z}}$ and an AR(k_n) model; see Theorem 3. It is worth mentioning that although this kind of results has been pursued by many authors (e.g., Bhansali and Papangelou (1991), Chan and Ing (2011), Findley and Wei (2002), Masuda (2013), Jeganathan (1989) and Sieders and Dzhaparidze (1987)), no results have been established for detrended time series so far, even in the short-memory case. The second application focuses on the problem of estimating the inverse of the covariance matrix of $\mathbf{z}$, $R_n := E(\mathbf{z}\mathbf{z}')$. When $\hat{\mathbf{z}}_t(k) = \mathbf{z}_t(k)$, Wu and Pourahmadi (2009) established the consistency of the banded covariance matrix estimators of R_n under short-memory time series. McMurry and Politis (2010) subsequently generalized their result to the tapered covariance matrix estimators. Based on these developments, it is easy to construct a consistent estimator of R_n^{-1} by inverting a consistent banded (or tapered) estimator of R_n; see Corollary 1 of Wu and Pourahmadi (2009). Although these estimators perform well in the short-memory case, they are not necessarily suitable for long-memory time series whose autocovariance functions are not absolutely summable. In particular, the banded and tapered estimators of R_n may incur large truncation errors in the long-memory case, thereby failing to achieve consistency. To rectify this deficiency, we estimate R_n^{-1} directly using $\hat{\mathbf{z}}$ and the banded Cholesky decomposition. We further apply Theorem 2 to develop the moment convergence of the difference between the proposed estimator and R_n^{-1} under spectral norm; see Theorem 4.

2 Inverse moment bounds

Assume that in model (1.1), $\{z_t\}$ is a stationary time series whose autocovariance function, $\gamma_l = E(z_t z_{t+l})$, satisfies

$$\gamma_l = O\left(|l|^{-1+2d}\right), \tag{2.1}$$

for some $0 < d < 1/2$. As mentioned in Section 1, the main purpose of this section is to establish (1.4) with $\hat{\mathbf{z}}_t(k) \neq \mathbf{z}_t(k)$. To this end, we need the following assumptions. Define

$$\mathcal{F}_n = \left\{ \sum_{m=1}^{n} s_m z_m : \sum_{m=1}^{n} s_m^2 = 1 \right\}.$$

(C1). There exist positive numbers α, δ and M_0 such that for any $0 < u - v \leq \delta$, any $f, f_1, \cdots, f_{k_1} \in \mathcal{F}_n$ and any $n \geq 1$,

$$P\left(v < f/\eta^{1/2} \leq u \,|f_1, \cdots, f_{k_1}\right) \leq M_0(u-v)^{\alpha} \qquad \text{a.s.}, \tag{2.2}$$

provided $\eta = \eta(f, f_1, \cdots, f_{k_1}) = \min_{(c_1,\cdots,c_{k_1})' \in R^{k_1}} E(f - \sum_{i=1}^{k_1} c_i f_i)^2 > 0$.

(C2). There exist an orthonormal basis, $\{\mathbf{o}_1, \cdots, \mathbf{o}_r\}$, of $\overline{\text{sp}}\{\check{\mathbf{x}}_1, \cdots, \check{\mathbf{x}}_p\}$, and constants $0 < \delta_1 < 1$ and $1 < \delta_2 < \infty$ such that for all large n,

$$\max_{1 \leq j \leq r} \sum_{i=1}^{\lfloor n\delta_1 \rfloor} o_{ij}^2 < \frac{1}{\delta_2 r}, \tag{2.3}$$

where $\lfloor a \rfloor$ denotes the largest integer $\leq a$.

(C3). The spectral density of $\{z_t\}$, $f_z(\lambda)$, satisfies

$$f_z(\lambda) \neq 0, -\pi \leq \lambda \leq \pi. \tag{2.4}$$

Remark 1. Conditions (2.1) and (C3) are fulfilled by many short- and long-memory time series models encountered in general practice. For example, they are satisfied by the linear process,

$$z_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \tag{2.5}$$

where ε_t's are independent random variables with $E(\varepsilon_t) = 0$ and $0 < E(\varepsilon_t^2) = \sigma^2 < \infty$, and b_j's are real numbers obeying

$$b_0 = 1, b_l = O\left(l^{-1+d}\right), \text{ for some } 0 < d < 1/2, \tag{2.6}$$

and

$$\frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} b_j e^{-ij\lambda} \right|^2 \neq 0, \text{ for any } -\pi \leq \lambda \leq \pi. \tag{2.7}$$

A well-known special case of (2.5)-(2.7) is the autoregressive fractionally integrated moving average (ARFIMA) process,

$$\phi(B)(1-B)^s z_t = \theta(B)\varepsilon_t,$$

where $0 \leq s \leq d$, B is the backshift operator, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ are polynomials of orders p and q, respectively, $|\phi(z)\theta(z)| \neq 0$ for $|z| \leq 1$, and $|\phi(z)|$ and $|\theta(z)|$ have no common zeros. In the sequel, the process is denoted by ARFIMA(p, s, q), which has short memory when $s = 0$.

Remark 2. In the case of $\hat{\mathbf{z}}_t(k) = \mathbf{z}_t(k)$, conditions like (C1) have been frequently used to establish results similar to (1.4); see, e.g., Bhansali and Papangelou (1991), Papangelou (1994) and Katayama (2008). In the case of $\hat{\mathbf{z}}_t(k) \neq \mathbf{z}_t(k)$, (C1) can be used in conjunction with (C2) and

(C3) to show that the conditional distribution of $\mathbf{v}'\hat{\mathbf{z}}_t(k)$ given $\sigma(z_s, \dots z_1, \nu_{1n}, \dots \nu_{rn})$ is sufficiently smooth (in the sense of (2.15)), provided $t-s$ is sufficiently large, $|\mathbf{v}'\mathbf{o}_t^{(j)}(k)|$ is sufficiently small and $s \leq \lfloor n\delta_1 \rfloor$. As will become clear later, this is a key step toward proving (1.4) in the latter case. When x_t is a linear process, (C1) is usually more restrictive than (55) of Findley and Wei (2002) and (K.2) of Ing and Wei (2003), which impose Lipschitz-type conditions on the distribution functions of ε_t. However, (C1) enables us to handle the conditional distribution of $\mathbf{v}'\hat{\mathbf{z}}_t(k)$ given $\sigma(z_s, \dots z_1, \nu_{1n}, \dots \nu_{rn})$ in a more mathematically tractable way, noting that ν_{in} can be arbitrary linear combinations of $z_1, \dots, z_n$. Condition (C1) is easily satisfied by any Gaussian process with non-degenerate finite-dimensional distributions. While it is possible to verify (C1) under non-Gaussian processes or linear processes with errors satisfying some smoothness conditions, the details need to be treated separately and are not pursued here.

Remark 3. Condition (C2) is satisfied by most commonly used design matrices. One typical example is $\mathbf{x}_t = (1, t, \dots, t^{p-1})', p \geq 1$, which implies $E(\hat{z}_s\hat{z}_t) \to 0$ as $|t-s| \to \infty$. Condition (C2) can even accommodate design matrices yielding that $E(\hat{z}_s\hat{z}_t)$ is not negligible for large $|t-s|$. To see this, let $\mathbf{X} = (1, 0, \dots, 0, 1)'$, and hence $r = 1$, $\mathbf{o}_1 = (1/\sqrt{2}, 0, \dots, 0, 1/\sqrt{2})'$, and (C2) is satisfied by any $0 < \delta_1 < 1$ and $1 < \delta_2 < 2$. In addition, since $\hat{z}_1 = (z_1 - z_n)/2 = -\hat{z}_n$, $\lim_{n\to\infty} E(\hat{z}_1\hat{z}_n) = -\gamma_0/2 \neq 0$. This example also points out a fundamental difference between $\hat{z}_t$ and z_t because by (2.1), $E(z_s z_t) = \gamma_{t-s}$ always converges 0 as $|t-s| \to \infty$.

The next theorem generalizes Lemma 1 of Ing and Wei (2003) to situations where $\hat{\mathbf{z}}_t(k_n) \neq \mathbf{z}_t(k_n)$ and $\sum_{j=0}^{\infty} |\gamma_j|$ is allowed to be unbounded.

Theorem 1. *Assume* (1.1), (2.1) and (C1)-(C3). *Suppose* $k_n \asymp n^{\theta_1}$ *with* $0 < \theta_1 < 1$. *Then for any* $q, \iota > 0$,

$$E\left(\lambda_{\min}^{-q}\left(\hat{R}(k_n)\right)\right) = O\left(k_n^{q+\iota}\right).$$

Proof. Let $w_t = \hat{\mathbf{z}}_t(k_n)\hat{\mathbf{z}}_t'(k_n)$, M be a sufficiently large constant whose value will be specified later, and $g_n = \lfloor \frac{n-k_n+1}{Mk_n} \rfloor$. Some algebraic manipulations imply

$$\begin{aligned}
\lambda_{\min}^{-q}\left(\hat{R}(k_n)\right) &\leq n^q \lambda_{\min}^{-q}\left(\sum_{j=0}^{g_n-1}\sum_{l=0}^{Mk_n-1} w_{g_n l+j+k_n}\right) \\
&\leq Ck_n^q \frac{1}{g_n}\sum_{j=0}^{g_n-1}\lambda_{\min}^{-q}\left(\sum_{l=0}^{Mk_n-1} w_{g_n l+j+k_n}\right), \qquad (2.8)
\end{aligned}$$

where C, here and hereafter, denotes a generic positive constant independent of n. Let $\iota > 0$ be arbitrarily chosen. If for all $j = 1, \cdots, g_n - 1$ and all large n,

$$E\left(\lambda_{\min}^{-q}\left(\sum_{l=0}^{Mk_n-1} w_{g_n l+k_n+j}\right)\right) \leq Ck_n^{\iota} \qquad (2.9)$$

holds true, then this, in conjunction with (2.8), yields the desired conclusion.

The rest of proof is only devoted to proving (2.9) with $j = 0$ since (2.9) with $j > 0$ can be proved similarly. Let $b_n^* = k_n^\iota$ and $\xi > \max\{[(3/\iota)+1]q/2, [(2d/\theta_1)+1](q/2\iota)+q/2\}$. Then, it follows that,

$$\begin{aligned} E\left(\lambda_{\min}^{-q}\left(\sum_{l=0}^{Mk_n-1} w_{g_n l+k_n}\right)\right) &\leq b_n^* + \int_{b_n^*}^{\infty} P\left(\sum_{l=0}^{Mk_n-1} \|\hat{\mathbf{z}}_{g_n l+k_n}(k_n)\|^2 > \mu^{2\xi/q}/k_n\right) d\mu \\ &+ \int_{b_n^*}^{\infty} P\left(G_n(\mu)\right) d\mu := b_n^* + D_{1n} + D_{2n}, \end{aligned} \tag{2.10}$$

where

$$G_n(\mu) = \left\{\inf_{\|\mathbf{v}\|=1} \sum_{l=0}^{Mk_n-1} \left(\mathbf{v}'\hat{\mathbf{z}}_{g_n l+k_n}(k_n)\right)^2 < \mu^{-1/q}, \sum_{l=0}^{Mk_n-1} \|\hat{\mathbf{z}}_{g_n l+k_n}(k_n)\|^2 \leq \mu^{2\xi/q}/k_n\right\}.$$

By (2.1), Theorem 2.1 of Ing and Wei (2006), Hölder's inequality and $\|\mathbf{o}_i\|^2 = 1$, $1 \leq i \leq r$,

$$E(\nu_{in}^2) \leq C\left(\sum_{j=1}^{n} |o_{ji}|^{\frac{2}{1+2d}}\right)^{1+2d} \leq Cn^{2d}\sum_{j=1}^{n} o_{ji}^2 \leq Cn^{2d}, \qquad 1 \leq i \leq r. \tag{2.11}$$

In view of (2.11), Chebyshev's inequality and the definition of ξ, we get

$$D_{1n} \leq C\frac{k_n^3 + k_n n^{2d}}{k_n^{(2\xi-q)\iota/q}} = o(1). \tag{2.12}$$

We next give a bound for D_{2n}. Following the argument given in page 137 of Ing and Wei (2003), it can be shown that for any $\mu > 1$, there exists a set of k_n-dimensional unit vectors $\mathbf{v}_1, \cdots, \mathbf{v}_{m^*}$, with $m^* = m^*(\mu) \leq (\lfloor \mu^{(\xi+1/2)/q} \rfloor + 1)^{k_n}$ such that

$$\begin{aligned} P(G_n(\mu)) &\leq \sum_{j=1}^{m^*} P\left(\bigcap_{l=0}^{Mk_n-1} \left|\mathbf{v}_j'\hat{\mathbf{z}}_{g_n l+k_n}(k_n)\right| \leq 3\mu^{-1/2q}\right) \\ &= \sum_{j=1}^{m^*} E\left\{\prod_{l=0}^{Mk_n-1} I_{E_{j,l}(\mu)}\right\}, \end{aligned} \tag{2.13}$$

where $E_{j,l}(\mu) = \{|\mathbf{v}_j'\hat{\mathbf{z}}_{g_n l+k_n}(k_n)| \leq 3\mu^{-1/2q}\}$ and the dependence of $\mathbf{v}_j$ on μ is suppressed for simplicity.

In what follows, we shall show that for any $1 \leq j \leq m^*$ and $\mu > 1$, there exist positive integers $r_j < Mk_n - 1$ and $l_{1,j} < \cdots < l_{r_j,j} < Mk_n - 1$ such that the mean-square error of the best linear predictor of $\mathbf{v}_j'\mathbf{z}_{g_n l_{p+1,j}+k_n}(k_n)$ based on $\nu_{1n}, \ldots, \nu_{rn}$ and $z_1, \ldots z_{g_n l_{p,j}+k_n}$ is bounded away from 0 for all large n and all $1 \leq p \leq r_j - 1$, $1 \leq j \leq m^*$ and $\mu > 1$. To achieve this goal, first express $\mathbf{v}_j$ as $(v_{j,1}, \cdots, v_{j,k_n})'$. For $1 \leq j \leq m^*$ and $0 \leq l \leq Mk_n - 1$, define $\mathbf{v}_j(l) = (w_1, \cdots, w_n)'$, where $w_i = v_{j,s}$ if $i = g_n l + k_n - s + 1$ for some $1 \leq s \leq k_n$, and 0 otherwise. Let $0 < \theta_2 < 1$,

$$L_h(j) = \left\{l : \left|\mathbf{v}_j'(l)\mathbf{o}_h\right| = |\mathbf{v}_j'\mathbf{o}_{g_n l+k_n}^{(h)}(k_n)| \geq \frac{\theta_2(\delta_2 - 1)}{2r^{1/2}\delta_2}, 0 \leq l \leq Mk_n - 1\right\}, \text{ and } L(j) = \bigcup_{h=1}^{r} L_h(j),$$

noting that δ_2 is defined in (C2). Since $\|\mathbf{o}_h\|^2 = 1$, $1 \le h \le r$, it holds that for all $1 \le j \le m^*$ and $\mu > 1$, $\sharp L(j) \le r\{(2r^{1/2}\delta_2)/[\theta_2(\delta_2 - 1)]\}^2 := D$. Choose $0 < \delta_1' < \delta_1$ and set

$$\bar{L}(j) = \left\{ l : 0 \le l \le \lfloor \delta_1' M k_n \rfloor - 1, l \notin L(j) \right\}.$$

Now define $\{l_{1,j}, \cdots, l_{r_j,j}\} = \bar{L}(j)$, where $l_{s,j} < l_{t,j}$ if $s < t$. It is readily seen that for all $1 \le j \le m^*$ and $\mu > 1$, $r_j \ge \lfloor \delta_1' M k_n \rfloor - D$. By making use of (C2) and (C3), we show in Appendix that for all large n and all $1 \le p \le r_j - 1$, $1 \le j \le m^*$ and $\mu > 1$,

$$\min_{\substack{(c_1,\cdots,c_r)\in R^r,\\ (d_1,\cdots,d_{g_n l_{p,j}+k_n})\in R^{g_n l_{p,j}+k_n}}} E\left(\mathbf{v}_j'(l_{p+1,j})\mathbf{z} - \sum_{i=1}^{r} c_j \nu_{in} - \sum_{j=1}^{g_n l_{p,j}+k_n} d_j z_j\right)^2 \ge \underline{\eta}, \tag{2.14}$$

where $\underline{\eta}$ is some positive number independent of n, p, j and μ. Therefore, the property mentioned at the beginning of this paragraph holds true.

By (2.14), (C1) and $\mathbf{v}_j'\hat{\mathbf{z}}_{g_n l_{p+1,j}+k_n}(k_n) = \mathbf{v}_j'(l_{p+1,j})\mathbf{z} - \sum_{i=1}^{r} \nu_{in}\mathbf{v}_j'(l_{p+1,j})\mathbf{o}_i$, it follows that the conditional distribution of $\mathbf{v}_j'\hat{\mathbf{z}}_{g_n l_{p+1,j}+k_n}(k_n)$ given $\mathcal{D}_j(p) = \sigma(\nu_{1n}, \cdots, \nu_{rn}, z_1, \cdots, z_{g_n l_{p,j}+k_n})$ is *uniformly* smooth in the sense that for all large n and all $1 \le p \le r_{j-1}$, $1 \le j \le m^*$, and $\mu > 1$,

$$\begin{aligned}
& P\left(\left| \mathbf{v}_j'\hat{\mathbf{z}}_{g_n l_{p+1,j}+k_n}(k_n) \right| < 3\mu^{-1/2q} \middle| \mathcal{D}_j(p) \right) \\
= \;& P\left(-3\mu^{-1/2q} + \sum_{i=1}^{r} \nu_{in}\mathbf{v}_j'(l_{p+1,j})\mathbf{o}_i < \mathbf{v}_j'(l_{p+1,j})\mathbf{z} < 3\mu^{-1/2q} + \sum_{i=1}^{r} \nu_{in}\mathbf{v}_j'(l_{p+1,j})\mathbf{o}_i \middle| \mathcal{D}_j(p) \right) \\
\le \;& M_0 \left(6\underline{\eta}^{-1/2}\mu^{-1/2q}\right)^{\alpha} \quad \text{a.s.}
\end{aligned} \tag{2.15}$$

Equipped with (2.15), we are now ready to provide an upper bound for $E(\prod_{l=0}^{Mk_n-1} I_{E_{j,l}(\mu)})$, which further leads to an upper bound for D_{2n}. Let $\bar{L}^{(p)}(j) = \{l_{1,j}, \cdots, l_{p,j}\}$, $1 \le p \le r_j$. By (2.15) and $r_j \ge \lfloor \delta_1' M k_n \rfloor - D$, one has for all large n and all $1 \le j \le m^*$ and $\mu > 1$,

$$\begin{aligned}
E\left(\prod_{l=0}^{Mk_n-1} I_{E_{j,l}(\mu)}\right) &\le E\left(\prod_{l\in\bar{L}(j)} I_{E_{j,l}(\mu)}\right) \\
\le \;& E\left[\prod_{l\in\bar{L}^{(r_j-1)}(j)} I_{E_{j,l}(\mu)} E\left(I_{E_{j,l_{r_j,j}}(\mu)} \middle| \mathcal{D}_j(r_j - 1)\right)\right] \\
\le \;& M_0\left(6\underline{\eta}^{-1/2}\mu^{-1/2q}\right)^{\alpha} E\left(\prod_{l\in\bar{L}^{(r_j-1)}(j)} I_{E_{j,l}(\mu)}\right) \\
\le \;& \left\{M_0\left(6\underline{\eta}^{-1/2}\mu^{-1/2q}\right)^{\alpha}\right\}^{\lfloor \delta_1' M k_n \rfloor - D}.
\end{aligned} \tag{2.16}$$

Choose $M > (2\xi + 1)/(\alpha\delta_1')$. Combining (2.16) and (2.13) yields that for all large n, there is a positive constant, $\bar{M}$, independent of n such that

$$D_{2n} \le \bar{M}^{k_n} \int_{b_n^*}^{\infty} \mu^{-\dfrac{\alpha\delta_1' M k_n - \alpha D - (2\xi+1)k_n}{2q}} d\mu \le C. \tag{2.17}$$

Consequently, (2.9) with $j = 0$ follows from (2.10), (2.12) and (2.17). □

Remark 4. When $1 \leq k_n = k < \infty$ is fixed with n, Theorem 2.1 of Chan and Ing (2011) generalizes Lemma 1 of Ing and Wei (2003) in another direction. In particular, their theorem shows that for any $q > 1$,

$$E\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \lambda_{\min}^{-q}\left(n^{-1}\sum_{t=k}^{n}\mathbf{z}_t(k,\boldsymbol{\theta})\mathbf{z}_t^{'}(k,\boldsymbol{\theta})\right)\right\} = O(1), \tag{2.18}$$

where $\mathbf{z}_t^{'}(k,\boldsymbol{\theta}) = (z_t(\boldsymbol{\theta}), \ldots, z_{t-k+1}(\boldsymbol{\theta}))$ and $\{z_t(\boldsymbol{\theta})\}$ is a stationary time series indexed by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with $\boldsymbol{\Theta}$ being a compact set in R^s for some $1 \leq s < \infty$. Equation (2.18) can be used to establish moment bounds for the conditional sum of squares estimators in ARMA models; see Theorem 3.3 of Chan and Ing (2011). Clearly, (2.18) reduces to (1.4) with $\hat{\mathbf{z}}_t(k) = \mathbf{z}_t(k)$ if $\boldsymbol{\Theta}$ only contains one point. On the other hand, since $z_t(\boldsymbol{\theta})$, like z_t, depends only on the information up to time t, the proof of (2.18) is more closely related to that of Lemma 1 of Ing and Wei (2003) (or Property A of Findley and Wei (2002)) than Theorem 1. Indeed, the proof of Theorem 1 follows similar arguments put forth in Findley and Wei (2002) and Ing and Wei (2003) up until equation (2.13), and the arguments between (2.13) and (2.17), focusing exclusively on $\hat{z}_t$, are new to the literature.

For an $n \times m$ matrix A, define its spectral norm $\|A\|_2 = (\sup_{\|\mathbf{x}\|=1,\mathbf{x}\in R^m} \mathbf{x}'A'A\mathbf{x})^{1/2}$. Then, Theorem 1 implies that for any $q, \iota > 0$,

$$E\left\|\hat{R}^{-1}(k_n)\right\|_2^q = O(k_n^{q+\iota}). \tag{2.19}$$

This result serves as a stepping stone to proving (1.4), or equivalently, for some $q > 0$,

$$E\left\|\hat{R}^{-1}(k_n)\right\|_2^q = O(1), \tag{2.20}$$

under additional assumptions. To introduce the details, set $R(k) = E(\mathbf{z}_k(k)\mathbf{z}_k'(k))$. According to (2.1) and Proposition 5.1.1 of Brockwell and Davis (1991), $R^{-1}(k)$ exists for any $k \geq 1$.

Theorem 2. *Assume* (1.1), (2.5)-(2.7), (C1) and (C2). *Suppose*

$$\sup_{-\infty<t<\infty} E\left(|\varepsilon_t|^{2q_1}\right) < \infty, \text{ for some } q_1 \geq 2, \tag{2.21}$$

and

$$k_n \asymp n^{\theta_1}, \text{where } 0 < \theta_1 < \frac{1}{4} \text{ if } 0 < d \leq \frac{1}{4}, \text{and } 0 < \theta_1 < \frac{1}{2} - d \text{ if } \frac{1}{4} < d < \frac{1}{2}. \tag{2.22}$$

Then, for $0 < q < q_1$,

$$E\left\|\hat{R}^{-1}(k_n) - R^{-1}(k_n)\right\|_2^q = o(1), \tag{2.23}$$

and (2.20) *follows. Moreover, if* (2.21) *and* (2.22) *are replaced by*

$$\sup_{-\infty<t<\infty} E\left(|\varepsilon_t|^{q}\right) < \infty, \quad \text{for any } q > 0, \tag{2.24}$$

and

$$k_n \asymp n^{\theta_1}, \text{where } 0 < \theta_1 < \frac{1}{2} \text{ if } 0 < d \leq \frac{1}{4}, \text{ and } 0 < \theta_1 < 1 - 2d \text{ if } \frac{1}{4} < d < \frac{1}{2}, \tag{2.25}$$

respectively, then, (2.20) *and* (2.23) *hold for any* $q > 0$.

Proof. We first assume that (2.21) and (2.22) hold true. By

$$\left\|\hat{R}^{-1}(k_n) - R^{-1}(k_n)\right\|_2^q \le \left\|\hat{R}^{-1}(k_n)\right\|_2^q \left\|\hat{R}(k_n) - R(k_n)\right\|_2^q \left\|R^{-1}(k_n)\right\|_2^q, \tag{2.26}$$

(2.1) (which is implied by (2.5) and (2.6)), (2.21) and an argument used in Lemma 2 of Ing and Wei (2003), it follows that

$$E\left\|\hat{R}(k_n) - R(k_n)\right\|_2^{q_1} \le \begin{cases} C\dfrac{k_n^{q_1}}{n^{q_1/2}}, & \text{if} \quad 0 < d < \dfrac{1}{4}, \\ C\left(\dfrac{\log n}{n}\right)^{q_1/2} k_n^{q_1}, & \text{if} \quad d = \dfrac{1}{4}, \\ C\left(\dfrac{k_n}{n^{1-2d}}\right)^{q_1}, & \text{if} \quad \dfrac{1}{4} < d < \dfrac{1}{2}. \end{cases} \tag{2.27}$$

Moreover, (2.7) yields

$$\sup_{k \ge 1} \left\|R^{-1}(k)\right\|_2 < \infty. \tag{2.28}$$

Consequently, (2.20) and (2.23) follow from (2.19), (2.22), (2.26)-(2.28) and Hölder's inequality.

If in place (2.21), we assume (2.24), then the q_1 in (2.27) can be replaced by any $q > 0$. This modification, together with (2.19), (2.25), (2.26), (2.28) and an argument similar to that used in the proof of Theorem 2 of Ing and Wei (2003), yields that (2.20) and (2.23) hold for any $q > 0$, which leads to the second conclusion of the theorem.

Remark 5. When (2.25) is restricted to $0 < d < 1/4$ and $k_n \asymp n^{\theta_1}$ for some $0 < \theta_1 < 1/2$, the second conclusion of Theorem 2 can be modified as follows: (2.20) and

$$E\left\|\hat{R}^{-1}(k_n) - R^{-1}(k_n)\right\|_2^q = O\left(\frac{k_n^q}{n^{q/2}}\right) \tag{2.29}$$

hold for any $q > 0$. Equation (2.29) is a strengthened version of (2.21) of Ing and Wei (2003), which gives the same rate of convergence of $E\|\hat{R}^{-1}(k_n) - R^{-1}(k_n)\|_2^q$ in the case that $\hat{\mathbf{z}}_t(k_n) = \mathbf{z}_t(k_n)$ and $\{z_t\}$ is a short-memory process.

3 Applications

Throughout this section, we assume that z_t admits the following AR(∞) representation:

$$z_{t+1} + \sum_{i=1}^{\infty} a_i z_{t+1-i} = \varepsilon_{t+1}, \tag{3.1}$$

where

$$a_i = O(i^{-1-d}), \text{ with } 0 < d < 1/2. \tag{3.2}$$

Note that (3.1) and (3.2) include the ARFIMA(p, s, q) model, with $0 \le s \le d$, as a special case. Denote the coefficient vector $(a_1, a_2, \ldots)'$ by $\mathbf{a}$. This section aims at estimating $\mathbf{a}$ and $R_n^{-1} = R^{-1}(n)$ based on the detrended series $\hat{z}_1, \ldots \hat{z}_n$, and providing the corresponding mean convergence results using Theorem 2.

3.1 Estimation of a

We first consider a finite-order approximation model corresponding to (3.1),

$$z_{t+1} + \sum_{i=1}^{k} a_i(k) z_{t+1-i} = \varepsilon_{t+1,k},$$

where $\mathbf{a}(k) = (a_1(k), \cdots, a_k(k))'$ minimizes $E(z_{t+1} + c_1 z_t + \cdots + c_k z_{t+1-k})^2$ over $(c_1, \cdots, c_k)' \in R^k$. The least squares estimator of $\mathbf{a}(k)$ based on $\hat{z}_1, \cdots, \hat{z}_n$ is given by

$$\hat{\mathbf{a}}(k) = \arg \min_{(c_1,\cdots,c_k)' \in R^k} \sum_{t=k}^{n-1} (\hat{z}_{t+1} + c_1 \hat{z}_t + \cdots + c_k \hat{z}_{t+1-k})^2.$$

Estimating $\mathbf{a}$ by the finite predictor coefficients $\hat{\mathbf{a}}^*(k) = (\hat{\mathbf{a}}'(k), 0, \cdots)'$, the objective of Section 3.1 is to establish the moment convergence of $\|\hat{\mathbf{a}}^*(k) - \mathbf{a}\|$ to zero. As shown in the next theorem, this goal is achievable if $k = k_n$ approaches ∞ at a suitable rate.

Theorem 3. *Assume* (1.1), (2.5)-(2.7), (C1), (C2), (3.1), (3.2) *and* (2.24). *Suppose k_n satisfies* (2.25). *Then for any $q > 0$,*

$$E\,\|\hat{\mathbf{a}}^*(k_n) - \mathbf{a}\|^q = o(1). \tag{3.3}$$

Proof. Clearly,

$$\|\hat{\mathbf{a}}^*(k_n) - \mathbf{a}\|^q \le C \left\{ \|\hat{\mathbf{a}}(k_n) - \mathbf{a}(k_n)\|^q + \|\mathbf{a}^*(k_n) - \mathbf{a}\|^q \right\}, \tag{3.4}$$

where $\mathbf{a}^*(k) = (\mathbf{a}'(k), 0, \cdots)'$. It follows from (C3) that

$$\begin{aligned} \|\mathbf{a}^*(k_n) - \mathbf{a}\|^q &\le C \left\{ E \left(\sum_{i=k_n+1}^{\infty} a_i z_{t+1-i} \right)^2 \right\}^{q/2} \\ &\le C \left\{ E \left(\sum_{i=k_n+1}^{\lfloor k_n^{1/d} \rfloor} a_i z_{t+1-i} \right)^2 + E \left(\sum_{\lfloor k_n^{1/d} \rfloor + 1}^{\infty} a_i z_{t+1-i} \right)^2 \right\}^{q/2}. \end{aligned} \tag{3.5}$$

By Theorem 2.1 of Ing and Wei (2006),

$$E \left(\sum_{i=k_n+1}^{\lfloor k_n^{1/d} \rfloor} a_i z_{t+1-i} \right)^2 \le C k_n^{-1},$$

and by (3.2) and Minkowski's inequality,

$$E \left(\sum_{\lfloor k_n^{1/d} \rfloor + 1}^{\infty} a_i z_{t+1-i} \right)^2 \le C k_n^{-1},$$

The above two equations, together with (3.5), give

$$\|\mathbf{a}^*(k_n) - \mathbf{a}\|^q \leq C k_n^{-q/2}. \tag{3.6}$$

On the other hand, by Theorem 2 and some algebraic manipulations, one gets

$$\begin{aligned} E\,\|\hat{\mathbf{a}}(k_n) - \mathbf{a}(k_n)\|^q &\leq CE\left\{\left\|\hat{R}^{-1}(k_n)\right\|_2^q \left\|\frac{1}{n}\sum_{t=k_n}^{n-1} \hat{\mathbf{z}}_t(k_n)\left(\hat{z}_{t+1} + \hat{\mathbf{z}}_t'(k_n)\mathbf{a}(k_n)\right)\right\|^q\right\} \\ &\leq C\left\{E\left\|\frac{1}{n}\sum_{t=k_n}^{n-1} \hat{\mathbf{z}}_t(k_n)\left(\hat{z}_{t+1} + \hat{\mathbf{z}}_t'(k_n)\mathbf{a}(k_n)\right)\right\|^{2q}\right\}^{1/2}. \end{aligned} \tag{3.7}$$

Moreover, it follows from (3.4) and the proof of Lemma 4.2 of Ing, Chiou, and Guo (2013) that

$$E\left\|\frac{1}{n}\sum_{t=k_n}^{n-1} \hat{\mathbf{z}}_t(k_n)\left(\hat{z}_{t+1} + \hat{\mathbf{z}}_t'(k_n)\mathbf{a}(k_n)\right)\right\|^{2q} \leq C\left\{\left(\frac{k_n}{n}\right)^q + \left(\frac{1}{n^{1-2d}}\right)^q + \left(\frac{k_n}{n^{2-4d}}\right)^q\right\}. \tag{3.8}$$

Consequently, (3.3) is ensured by (3.4), (3.6), (3.7), (3.8) and (2.25).

Remark 6. In many applications, trend estimation can also be performed by first estimating the coefficients, $\boldsymbol{\xi} = (\boldsymbol{\tau}', \phi_1, \ldots, \phi_k)'$, in an AR model around a deterministic time trend,

$$y_t = \mathbf{x}_t'\boldsymbol{\tau} + \sum_{s=1}^{k} \phi_s y_{t-s} + \varepsilon_t, \tag{3.9}$$

and then plugging the estimated coefficients into the formula, $u_t = E(y_t) = \phi^{-1}(B)w_t$, where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_k z^k \neq 0$ for all $|z| \leq 1$ and $w_t = \mathbf{x}_t'\boldsymbol{\tau}$. Here we assume $|u_t| < \infty$ for any $-\infty < t < \infty$. Denote by $\hat{\boldsymbol{\xi}}$ the least squares estimator of $\boldsymbol{\xi}$, where $\hat{\boldsymbol{\xi}}$ satisfies

$$(\sum_{t=k+1}^{n} \mathbf{w}_t\mathbf{w}_t')\hat{\boldsymbol{\xi}} = \sum_{t=k+1}^{n} \mathbf{w}_t y_t,$$

with $\mathbf{w}_t = (\mathbf{x}_t', \mathbf{y}_{t-1}'(k))'$ and $\mathbf{y}_t(k) = (y_t, \ldots, y_{t-k+1})'$. In the following, we shall illustrate the moment convergence of $\|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|$ via an approach somewhat different from the one used in Theorem 3. To simplify the exposition, we only consider the case where (3.9) is correctly specified and $1 \leq k < \infty$ is fixed with n. Assume also that there is a $k \times p$ matrix $\mathbf{T}$ (independent of t) such that

$$\mathbf{T}\mathbf{x}_t = -(u_{t-1}, \ldots, u_{t-k})'. \tag{3.10}$$

Note that the $\mathbf{T}$ obeying (3.10) in the case of $\mathbf{x}_t = (1, t, \ldots, t^{p-1}), p \geq 1$ is given in Appendix B of Ing (2003). Equation (3.10) yields $\mathbf{w}_t^* = \mathbf{H}\mathbf{w}_t = (\mathbf{x}_t', \mathbf{z}_{t-1}'(k))'$, where with I_m and $\mathbf{0}_{m_1 \times m_2}$ denoting the m-dimensional identity matrix and the $m_1 \times m_2$ zero matrix, respectively,

$$\mathbf{H} = \left(\begin{array}{c|c} I_p & \mathbf{0}_{p \times k} \\ \hline \mathbf{T} & I_k \end{array}\right),$$

and $\mathbf{z}_t^{'}(k) = (z_t, \ldots, z_{t-k+1})$ with $z_t = y_t - u_t = \phi^{-1}(B)\varepsilon_t$. Let $\tilde{\mathbf{x}}_i = (x_{k+1,i}, \ldots, x_{ni})^{'}$ and define $\mathbf{E} = \mathrm{diag}(\|\tilde{\mathbf{x}}_1\|, \ldots, \|\tilde{\mathbf{x}}_p\|)$,

$$\mathbf{V} \quad = \quad \left(\begin{array}{c|c} \mathbf{E}^{-1} & \mathbf{0}_{p\times k} \\ \hline \mathbf{0}_{k\times p} & n^{-1/2} I_k \end{array} \right),$$

and $\tilde{R}(p+k) = \mathbf{V}(\sum_{t=k+1}^{n} \mathbf{w}_t^{*}\mathbf{w}_t^{*^{'}})\mathbf{V}^{'}$. When $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \cdots, \tilde{\mathbf{x}}_p)$ is nonsingular and the distribution functions of ε_t satisfy some smoothness conditions, it can be shown that for any $q > 0$,

$$E\|\tilde{R}^{-1}(p+k)\|_2^q = O(1), \tag{3.11}$$

provided $\sup_{-\infty<t<\infty} E(|\varepsilon_t|^{c_q}) < \infty$ for c_q large enough (depending on q). Moreover, Lemma 4 of Ing and Wei (2003) implies

$$E\| \sum_{t=k+1}^{n} \mathbf{V}\mathbf{w}_t^{*}\varepsilon_t\|^q = O(1), \tag{3.12}$$

under $\sup_{-\infty<t<\infty} E(|\varepsilon_t|^{s_1}) < \infty$ for $s_1 \geq \max\{2, q\}$. Now, the intended moment bound,

$$E\|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^q = o(1), \tag{3.13}$$

follows from (3.11), (3.12), $\|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\| \leq \|\mathbf{V}\|_2\|\mathbf{H}\|_2\|\tilde{R}^{-1}(p+k)\|_2\|\sum_{t=k+1}^{n}\mathbf{V}\mathbf{w}_t^{*}\varepsilon_t\|$, Höler's inequality and an additional assumption, $\min_{1\leq i\leq p}\|\tilde{\mathbf{x}}_i\| \to \infty$ as $n \to \infty$. For $\mathbf{x}_t = (1, t, \ldots, t^{p-1})^{'}$, (3.11) has been reported in Lemma B.1 of Ing (2003), which is also closely related to Lemma 1 of Yu, Lin and Cheng (2012). In fact, because $\mathbf{w}_t^{*}$ contains deterministic components and is $\sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$-measurable, the proof of (3.11) is different from that of (2.20) (or (1.4)) not only in the case of $\hat{\mathbf{z}}_t(k) \neq \mathbf{z}_t(k)$, but also in the case of $\hat{\mathbf{z}}_t(k) = \mathbf{z}_t(k)$ (see Remark 3 of Yu, Lin and Cheng (2012) for some discussion). Another important difference between (2.20) with $\hat{\mathbf{z}}_t(k) \neq \mathbf{z}_t(k)$ and (3.11) is that whereas the latter is obtained under the non-singularity of $\tilde{\mathbf{X}}$ and $\min_{1\leq i\leq p}\|\tilde{\mathbf{x}}_i\| \to \infty$ as $n \to \infty$, these restrictions are not necessary for the former. Finally, we remark that the proof of (3.13) is more involved when (3.10) fails to hold. It is expected that an argument used in the proof of Lemma 2 of Yu, Lin and Cheng (2012) can be generalized to establish the desired property. The details, however, are beyond the scope of the present article.

Remark 7. When $\mathbf{x}_t = 0$ for all t and $\{z_t\}$ is a short-memory process, several research studies related to Theorem 3 have been conducted and reported in the literature. For example, it is shown in Corollary 2 of Wu and Pourahmadi (2009) that

$$\|\hat{\mathbf{a}}_{n,l} - \mathbf{a}(n)\| = o_p(1), \tag{3.14}$$

where $l \to \infty$ at a rate much slower than n and $\hat{\mathbf{a}}_{n,l} = \hat{\Sigma}_{n,l}^{-1}\tilde{\boldsymbol{\gamma}}_n$, with $\hat{\Sigma}_{n,l} = (\check{\gamma}_{i-j}I_{|i-j|\leq l})_{1\leq i,j\leq n}$, $\check{\gamma}_k = n^{-1}\sum_{i=1}^{n-|k|} z_i z_{i+|k|}, k = 0, \pm 1, \ldots, \pm(n-1)$, $\tilde{\boldsymbol{\gamma}}_n = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_n)^{'}$ and $\tilde{\gamma}_i = \check{\gamma}_i I_{|i|\leq l}$. In addition, Corollary 1 of Bickel and Gel (2011) shows that

$$\|\hat{\boldsymbol{\tau}}_{p,n}^{b} - \mathbf{a}(p)\| = o_p(1), \tag{3.15}$$

where $p/n \to 0$ as $p \to \infty$ and $n \to \infty$ and $\hat{\boldsymbol{\tau}}_{p,n}^{b} = B_k(\hat{R}_{p,n})^{-1}\check{\boldsymbol{\gamma}}_p$, with $B_k(\hat{R}_{p,n}) = (\check{\gamma}_{i-j}I_{|i-j|\leq k})_{1\leq i,j\leq p}$, $\check{\boldsymbol{\gamma}}_p = (\check{\gamma}_1, \ldots, \check{\gamma}_p)^{'}$, and $k \asymp (n/p)^{\delta_2}$ for some $0 < \delta_2 < 1$. Note first that since (3.14) and (3.15) focus on convergence in probability instead of convergence of moments, assumptions like (C1) (or (K.2) of Ing and wei (2003)) are not needed for these two equations. In addition, the moment conditions used to derive (3.14) and (3.15) are much weaker than that of Theorem 3. On the other hand, the proofs of (3.14) and (3.15), depending heavily on $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$, are difficult to be extended to long-memory time series. Moreover, while Theorem 3 leads immediately to $\|\hat{\mathbf{a}}^*(k_n) - \mathbf{a}\| = o_p(1)$, (3.14) and (3.15) cannot guarantee the moment convergence of $\|\hat{\mathbf{a}}_{n,l} - \mathbf{a}(n)\|$ and $\|\hat{\boldsymbol{\tau}}^b_{p,n} - \mathbf{a}(p)\|$. In fact, the latter results are remain unestablished because of the lack of moment bounds for $\|\hat{\Sigma}^{-1}_{n,l}\|_2$ and $\|B_k(\hat{R}_{p,n})^{-1}\|_2$.

3.2 Estimation of R_n^{-1}

Since R_n is symmetric and positive definite, it has a modified Cholesky decomposition (see, for example, Wu and Pourahmadi (2003)),

$$T_n R_n T_n' = D_n, \tag{3.16}$$

where $T_n = (t_{ij})_{1\le i,j\le n}$ is a unit lower triangular matrix with

$$t_{ij} = \begin{cases} 0, & \text{if } i < j; \\ 1, & \text{if } i = j; \\ a_{i-j}(i-1), & \text{if } 2 \le i \le n \text{ and } 1 \le j \le i-1, \end{cases}$$

and

$$D_n = diag(\gamma_0, \sigma^2(1), \cdots, \sigma^2(n-1)),$$

with $\sigma^2(k) = E(z_{t+1} + \mathbf{a}_t'(k)\mathbf{z}_t(k))^2$. In view of (3.16), R_n^{-1} can be expressed as

$$R_n^{-1} = T_n' D_n^{-1} T_n. \tag{3.17}$$

We further assume that $a_i(m)$'s obey (C4) and (C5):

(C4) There exists $C_1 > 0$ such that for $1 \le i \le m$ and all $m \ge 1$,

$$|a_i(m)| \le C_1 |a_i| \left(\frac{m}{m-i+1} \right)^d .$$

(C5) There exist $0 < \delta < 1$ and $C_2 > 0$ such that for $1 \le i \le \delta m$ and all $m \ge 1$,

$$|a_i(m) - a_i| \le C_2 \frac{i|a_i|}{m}.$$

Conditions (C4) and (C5) assert that the finite-past predictor coefficients $a_i(m), i = 1, \ldots, m$ approach the corresponding infinite-past predictor coefficients $a_1, a_2, \ldots$ in a nonuniform way. More specifically, they require that $|a_i(m)/a_i|$ is very close to 1 when $i = o(m)$, but has order of magnitude $m^{(1-\theta)d}$ when $m - i \asymp m^\theta$ with $0 \le \theta < 1$. This does not seem to be counterintuitive because for a long-memory process, the finite order truncation tends to create severer upward distortions in those a_i's with i near the truncation lag $m+1$. When $\{z_t\}$ is an $I(d)$ process, (C4) and (C5) follow directly from the proof of Theorem 13.2.1 of Brockwell and Davis (1991). Moreover, it is shown in Theorem 2.1 of Ing, Chiou and Guo (2013) that (C4) and (C5) are also satisfied by ARFIMA(p, d, q)

model. For a more detailed discussion on these two conditions, see Section 2 of Ing, Chiou and Guo (2013).

To consistently estimate R_n^{-1} based on $\hat{\mathbf{z}}$, we begin with a truncated version of (3.17),

$$T_n^{'}(k)D_n^{-1}(k)T_n(k), k \geq 1, \tag{3.18}$$

where $T_n(k) = (t_{ij}(k))_{1\leq i,j\leq n}$, with

$$t_{ij}(k) = \begin{cases} 0, & \text{if } i < j, \text{ or } \; k+1 < i \leq n \text{ and } 1 \leq j \leq i-k-1; \\ 1, & \text{if } i = j; \\ a_{i-j}(i-1), & \text{if } 2 \leq i \leq k \text{ and } 1 \leq j \leq i-1; \\ a_{i-j}(k), & \text{if } k+1 \leq i \leq n \text{ and } i-k \leq j < i-1, \end{cases}$$

and

$$D_n(k) = diag(\gamma_0, \sigma^2(1), \cdots, \sigma^2(k), \cdots, \sigma^2(k)).$$

We then estimate R_n^{-1} using a sample counterpart of (3.18),

$$\hat{C}_n^{(k)} := \hat{T}_n^{'}(k)\hat{D}_n^{-1}(k)\hat{T}_n(k), \tag{3.19}$$

where $\hat{T}_n(k)$ is $T_n(k)$ with $\mathbf{a}(l)$ replaced by $\hat{\mathbf{a}}(l)$ for $1 \leq l \leq k$ and $\hat{D}_n(k)$ is $D_n(k)$ with γ_0 replaced by $\hat{\gamma}_0 = n^{-1}\sum_{t=1}^n \hat{z}_t^2$ and $\sigma^2(l)$ replaced by $\hat{\sigma}^2(l) = (n-l)^{-1}\sum_{t=l}^{n-1}(\hat{z}_{t+1} + \hat{\mathbf{a}}_t^{'}(l)\hat{\mathbf{z}}_t(l))^2$ for $1 \leq l \leq k$. To obtain moment convergence results for $\|\hat{C}_n^{(k)} - R_n^{-1}\|_2$, we need some auxiliary lemmas.

Lemma 1. *Assume* (2.5)-(2.7), (3.1), (3.2), (C4) *and* (C5). *Then for* $2 \leq k \leq n < \infty$,

$$\left\|R_n^{-1} - T_n^{'}(k)D_n^{-1}(k)T_n(k)\right\|_2 \leq C^* \left(\frac{\log n \log k}{k^d}\right)^{1/2}, \tag{3.20}$$

where $C^* > 0$ *is independent of* n *and* k.

Proof. See Theorem 2.2 of Ing, Chiou and Guo (2013).

Lemma 2. *Under the assumptions of Theorem 3, we have for any* $q > 0$,

$$E\left\{(k_n \log k_n)^{q/2} \max_{1\leq k\leq k_n} \|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^q\right\} = o(1). \tag{3.21}$$

Proof. By Theorem 2, it holds that

$$\max_{1\leq k\leq k_n} E\left\|\hat{R}^{-1}(k)\right\|_2^q \leq E\left\|\hat{R}^{-1}(k_n)\right\|_2^q = O(1). \tag{3.22}$$

Using (3.22) and the same argument as in (3.7) and (3.8), we obtain for $1 \leq k \leq k_n$,

$$E\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^q \leq C\left\{\left(\frac{k}{n}\right)^{q/2} + \left(\frac{1}{n^{1-2d}}\right)^{q/2} + \left(\frac{k^{1/2}}{n^{1-2d}}\right)^q\right\},$$

which in conjunction with (2.24) yields that for any $\xi > 0$

$$\begin{aligned} & (k_n \log k_n)^{q/2} E \max_{1\leq k\leq k_n} \|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^q \\ \leq \;& C\left\{\left(\frac{(\log k_n)k_n^{2+\xi}}{n}\right)^{q/2} + \left(\frac{(\log k_n)k_n^{1+\xi}}{n^{1-2d}}\right)^{q/2} + \left(\frac{(\log k_n)^{1/2}k_n^{1+\xi}}{n^{1-2d}}\right)^q\right\}. \end{aligned} \tag{3.23}$$

Now, the desired conclusion (3.21) follows immediately from (2.25) and (3.23).

Lemma 3. *Under the assumptions of Theorem 2, we have for any $q > 0$,*

$$E\left\|\hat{D}_n^{-1}(k_n)\right\|_2^q = O(1). \tag{3.24}$$

Proof. Note first that

$$\left\|\hat{D}_n^{-1}(k_n)\right\|_2 = \max\{\hat{\gamma}_0^{-1}, \hat{\sigma}^{-2}(1), \cdots, \hat{\sigma}^{-2}(k_n)\}. \tag{3.25}$$

It is straightforward to see that

$$\hat{\gamma}_0^{-1} \leq \lambda_{\min}^{-1}(\hat{R}(k_n+1))$$

and for $1 \leq k \leq k_n$,

$$\hat{\sigma}^{-2}(k) \leq \lambda_{\min}^{-1}(\hat{R}(k+1)) \leq \lambda_{\min}^{-1}(\hat{R}(k_n+1)).$$

These equations, together with (3.25) and Theorem 2, yield (3.24).

Lemma 4. *Under the assumptions of Theorem 3, we have for any $q > 0$,*

$$(\log k_n)^q E\left\|\hat{D}_n^{-1}(k_n) - D^{-1}(k_n)\right\|_2^q = o(1). \tag{3.26}$$

Proof. By Lemma 3 and Hölder's inequality,

$$\begin{aligned} E\left(\left\|\hat{D}_n^{-1}(k_n) - D^{-1}(k_n)\right\|_2^q\right) &\leq E\left\{\left\|\hat{D}_n^{-1}(k_n)\right\|_2^q \left\|\hat{D}_n(k_n) - D_n(k_n)\right\|_2^q \left\|D^{-1}(k_n)\right\|_2^q\right\} \\ &\leq C\left(E\left\|\hat{D}_n(k_n) - D_n(k_n)\right\|_2^{3q}\right)^{1/3}. \end{aligned} \tag{3.27}$$

It is easy to see that

$$\left\|\hat{D}_n(k_n) - D(k_n)\right\|_2 \leq \max\left\{|\hat{\gamma}_0 - \gamma_0|, \max_{1\leq k\leq k_n} |\hat{\sigma}^2(k) - \sigma^2(k)|\right\}. \tag{3.28}$$

By (2.1), Lemma 2 of Ing and Wei (2003), Theorem 2.1 of Ing and Wei (2006) and some algebraic manipulations, it can be shown that

$$(\log k_n)^q E|\hat{\gamma}_0 - \gamma_0|^q = O\left(\left(\frac{(\log k_n)^2 \log n}{n}\right)^{q/2} + \left(\frac{\log k_n}{n^{1-2d}}\right)^q\right) = o(1). \tag{3.29}$$

An argument similar to that used to prove (3.23) further yields that for any $\xi > 0$,

$$\begin{aligned} E\left\{(\log k_n)^q \max_{1\leq k\leq k_n} |\hat{\sigma}^2(k) - \sigma^2(k)|^q\right\} &\leq \left\{\left(\frac{(\log k_n)k_n^{1+\xi}}{n}\right)^q \right. \\ + \left(\frac{(\log k_n)k_n^{\xi}}{n^{1-2d}}\right)^q &+ \left.\left(\frac{(\log k_n)k_n^{1+\xi}}{n^{2-4d}}\right)^q + \left(\frac{(\log k_n)(\log n)^{1/2}k_n^{\xi}}{n^{1/2}}\right)^q\right\} = o(1), \end{aligned} \tag{3.30}$$

where the last inequality is ensured by (2.25). In view of (3.27)-(3.30), the desired conclusion (3.26) follows.

The main result of this section is given as follows.

Theorem 4. *Under the assumptions of Theorem 3, one has for any $q > 0$,*

$$E\left\|\hat{C}_n^{(k)} - R_n^{-1}\right\|_2^q = o(1). \tag{3.31}$$

Proof. Clearly,

$$\begin{aligned}\left\|\hat{C}_n^{(k)} - R_n^{-1}\right\|_2 &\leq C\left\{\left\|R_n^{-1} - T_n'(k_n)D_n^{-1}(k_n)T_n(k_n)\right\|_2\right.\\ &+ \left.\left\|\hat{C}_n^{(k)} - T_n'(k_n)D_n^{-1}(k_n)T_n(k_n)\right\|_2\right\}. \end{aligned} \tag{3.32}$$

Note first that Lemma 1 and (2.25) implies

$$\left\|R_n^{-1} - T_n'(k_n)D_n^{-1}(k_n)T_n(k_n)\right\|_2 = o(1). \tag{3.33}$$

In addition, by Proposition 2.1 of Ing, Chiou and Guo (2013), $\|T_n(k_n)\|_2 = O((\log k_n)^{1/2})$. Moreover, an argument similar to that used in the proof of Proposition 3.1 of Ing, Chiou and Guo (2013) yields $\|\hat{T}_n(k_n) - T_n(k_n)\|_2^2 \leq Ck_n \max_{1\leq k\leq k_n} \|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_2^2$. These latter two inequalities, together with Lemmas 2-4 and

$$\begin{aligned}&\left\|\hat{C}_n^{(k)} - T_n'(k_n)D_n^{-1}(k_n)T_n(k_n)\right\|_2\\ \leq\ &\|\hat{T}_n(k_n) - T_n(k_n)\|_2\|\hat{D}_n^{-1}(k_n)\|_2(\|\hat{T}_n(k_n) - T_n(k_n)\|_2 + \|T_n(k_n)\|_2)\\ &+\|T_n(k_n)\|_2\|\hat{D}_n^{-1}(k_n) - D_n^{-1}(k_n)\|_2(\|\hat{T}_n(k_n) - T_n(k_n)\|_2 + \|T_n(k_n)\|_2)\\ &+\|T_n(k_n)\|_2\|D_n^{-1}(k_n)\|_2\|\hat{T}_n(k_n) - T_n(k_n)\|_2,\end{aligned}$$

imply $\mathrm{E}\|\hat{C}_n^{(k)} - T_n'(k_n)D_n^{-1}(k_n)T_n(k_n)\|_2^q = o(1)$ for any $q > 0$. Combining this with (3.32) and (3.33) leads to (3.31).

APPENDIX

Proof of (2.14). Note first that for any $1 \leq j \leq m^*(\mu)$ and $1 \leq k \leq r_j - 1$, the left-hand side of (2.14) is bounded below by $\lambda_{\min}(E(\mathbf{w}(j,k+1)\mathbf{w}'(j,k+1)))$, where

$$\mathbf{w}(j,k+1) = (\mathbf{v}_j'(l_{k+1,j})\mathbf{z}, \nu_{1n}, \cdots, \nu_{rn}, z_{s(j,k)}, z_{s(j,k)-1}, \cdots, z_1)'$$

with $s(j,k) = g_n l_{k,j} + k_n$. Moreover, we have

$$\begin{aligned}&\lambda_{\min}\left(E(\mathbf{w}(j,k+1)\mathbf{w}'(j,k+1))\right) \geq \lambda_{\min}(R_n)\lambda_{\min}(G(j,k+1))\\ \geq\ &\lambda_{\min}(R_n)\lambda_{\min}\left(B'(j,k+1)B(j,k+1)\right)\lambda_{\min}(F(j,k+1)),\end{aligned} \tag{A.1}$$

where

$$\begin{aligned}G(j,k+1) &= \left(\begin{array}{c|c} G^*(j,k+1) & A'(j,k+1) \\ \hline A(j,k+1) & I_{s(j,k)} \end{array}\right)\\ &= \left(\begin{array}{cccc|ccc} 1 & \mathbf{o}_1'\mathbf{v}_j(l_{k+1,j}) & \cdots & \mathbf{o}_r'\mathbf{v}_j(l_{k+1,j}) & 0 & \cdots & 0 \\ \mathbf{o}_1'\mathbf{v}_j(l_{k+1,j}) & 1 & \cdots & 0 & o_{s(j,k),1} & \cdots & o_{11} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{o}_r'\mathbf{v}_j(l_{k+1,j}) & 0 & \cdots & 1 & o_{s(j,k),r} & \cdots & o_{1r} \\ \hline 0 & o_{s(j,k),1} & \cdots & o_{s(j,k),r} & & & \\ \vdots & \vdots & \ddots & \vdots & & I_{s(j,k)} & \\ 0 & o_{11} & \cdots & o_{1r} & & & \end{array}\right),\end{aligned}$$

$$B(j,k+1)=\left(\begin{array}{c|c} I_{r+1} & \mathbf{0}_{(r+1)\times s(j,k)} \\ \hline A(j,k+1) & I_{s(j,k)} \end{array}\right),$$

and

$$\begin{aligned} F(j,k+1) &= (B^{-1}(j,k+1))'G(j,k+1)B^{-1}(j,k+1) \\ &= \left(\begin{array}{c|c} G^*(j,k+1)-A'(j,k+1)A(j,k+1) & \mathbf{0}_{(r+1)\times s(j,k)} \\ \hline \mathbf{0}_{s(j,k)\times(r+1)} & I_{s(j,k)} \end{array}\right). \end{aligned}$$

By (2.28), it holds that for any $n \geq 1$, there exists $\underline{l}_1 > 0$ such that

$$\lambda_{\min}(R_n) > \underline{l}_1 > 0. \tag{A.2}$$

Since

$$B'(j,k+1)B(j,k+1)=\left(\begin{array}{c|c} I_{r+1} & \mathbf{0}_{(r+1)\times s(j,k)} \\ \hline \mathbf{0}_{s(j,k)\times(r+1)} & \mathbf{0}_{s(j,k)\times s(j,k)} \end{array}\right)+(A(j,k+1),I_{s(j,k)})'(A(j,k+1),I_{s(j,k)}),$$

one obtains from straightforward calculations that for any $\boldsymbol{v}'=(\boldsymbol{v}_1',\boldsymbol{v}_2')\in R^{r+1+s(j,k)}$ with $\|\boldsymbol{v}\|=1$, $\boldsymbol{v}_1\in R^{r+1}$ and $\boldsymbol{v}_2\in R^{s(j,k)}$,

$$\begin{aligned} & \boldsymbol{v}'B'(j,k+1)B(j,k+1)\boldsymbol{v} \\ = \ & \|\boldsymbol{v}_1\|^2+\|\boldsymbol{v}_2\|^2+\boldsymbol{v}_1'A'(j,k+1)A(j,k+1)\boldsymbol{v}_1+2\boldsymbol{v}_2'A(j,k+1)\boldsymbol{v}_1 \\ \geq \ & \begin{cases} \|\boldsymbol{v}_1\|^2, & \text{if} \quad \|\boldsymbol{v}_1\|\geq 1/2, \\ 1-2\|\boldsymbol{v}_2\|\|A(j,k+1)\|_2\|\boldsymbol{v}_1\|, & \text{if} \quad \|\boldsymbol{v}_1\|<1/2. \end{cases} \end{aligned}$$

In addition, (C2) implies that for all large n and all $1\leq k\leq r_j-1$ and $1\leq j\leq m^*(\mu)$,

$$\|A(j,k+1)\|_2^2 \leq \operatorname{tr}(A'(j,k+1)A(j,k+1)) \leq r\max_{1\leq t\leq r}\sum_{i=1}^{\lfloor n\delta_1\rfloor} o_{it}^2 \leq \delta_2^{-1},$$

which, together with the above equation, yields that for all large n and all $1\leq k\leq r_j-1$ and $1\leq j\leq m^*(\mu)$,

$$\lambda_{\min}(B'(j,k+1)B(j,k+1)) \geq \min\{1/4, 1-\delta_2^{-1/2}\}. \tag{A.3}$$

Moreover, it follows from (C2) and the definition of $l_{k,j}$ that for all large n and all $1\leq k\leq r_j-1$ and $1\leq j\leq m^*(\mu)$,

$$\begin{aligned} & \lambda_{\min}(F(j,k+1)) \geq \min\{1, \lambda_{\min}(G^*(j,k+1)-A'(j,k+1)A(j,k+1))\} \\ \geq \ & \min\{1, \lambda_{\min}(G^*(j,k+1))-\|A(j,k+1)\|_2^2\} \\ \geq \ & 1-2r^{1/2}\max_{1\leq j\leq m^*(\mu)}\max_{1\leq k\leq r_j-1}\max_{1\leq t\leq r}\left|\mathbf{o}_t'\mathbf{v}_j(l_{k+1,j})\right| - r\max_{1\leq t\leq r}\sum_{i=1}^{\lfloor n\delta_1\rfloor} o_{it}^2 \\ \geq \ & 1-\frac{\theta_2(\delta_2-1)}{\delta_2}-\frac{1}{\delta_2}>0. \end{aligned} \tag{A.4}$$

The desired conclusion (2.14) now is ensured by (A.1)-(A.4).

4 Acknowledgments

We would like to thank the co-editor and three anonymous referees for their insightful and constructive comments, which greatly improve the presentation of this paper. The research of Ching-Kang Ing was supported in part by the Academia Sinica Investigator Award, and that of Shu-Hui Yu was partially supported by the National Science Council of Taiwan under grant NSC 99-2118-M-390-002.

REFERENCES

P. J. Bickel and Y. R. Gel (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *J. R. Statist. Soc. B*, **73** 711–728.

P. J. Brockwell and R. A. Davis (1991). *Time series: theory and methods*, 2nd edn. New York: Springer.

R. J. Bhansali and F. Papangelou (1991). Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order. *Ann. Statist.* **19** 1155–1162.

N. H. Chan and C.-K. Ing (2011). Uniform moment bounds of Fisher's information with applications to time series. *Ann. Statist.* **39** 1526–1550.

D. F. Findley and C. Z. Wei (2002). AIC, overfitting principles, and boundedness of moments of inverse matrices for vector autoregressive models fit to weakly stationary time series. *J. Multivariate Anal.* **83** 415–450.

W. A. Fuller and D. P. Hasza (1981). Properties of predictors for autoregressive time series. *J. Amer. Statist. Assoc.* **76** 155–161.

Ryan Greenaway-McGrevy (2013). Multistep Prediction Of Panel Vector Autoregressive Processes. *Economet. Theory* **29** 699–734.

C.-K. Ing (2003). Multistep prediction in autoregressive processes. *Economet. Theory* **19** 254–279.

C.-K. Ing, H. T. Chiou and M. H. Guo (2013). Estimation of inverse autocovariance matrices for long memory processes. Technical Report.

C.-K. Ing and C.-Z. Wei (2003). On same-realization prediction in an infinite-order autoregressive process. *J. Multivariate Anal.* **85** 130–155.

C.-K. Ing and C.-Z. Wei (2005). Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.* **33** 2423–2474.

C.-K. Ing and C.-Z. Wei (2006). A maximal moment inequality for long range dependent time series with applications to estimation and model selection. *Statist. Sinica* **16**, 721–740.

P. Jeganathan (1989). A note on inequalities for probabilities of large deviations of estimators in nonlinear regression models. *J. Multivariate Anal.* **30** 227–240

N. Katayama (2008). Asymptotic prediction of mean squared error for long-memory processes with estimated parameters. *J. Forecast.* **27** 690–720.

H. Masuda (2013). Convergence of gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *Ann. Statist.* **41** 1593–1641.

T. L. McMurry, and D. N. Politis (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482.

F. Papangelou (1994). On a distributional bound arising in autoregressive model fitting. *J. Appl. Prob.* **31** 401–408.

A. Sieders and K. Dzhaparidze (1987). A large deviation result for parameter estimators and its application to nonlinear regression analysis. *Ann. Statist.* **15** 1031–1049.

S.-H. Yu, C.-C. Lin and H.-W. Cheng (2012). A note on mean squared prediction error under the unit root model with deterministic trend. *J. Time Series Anal.* **33** 276–286.

W. B. Wu and M. Pourahmadi (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844.

W. B. Wu and M. Pourahmadi (2009). Banding sample covariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768.

X. Zhang, A. T. K. Wan and G. Zou (2013). Model averaging by jackknife criterion in models with dependent data. *J. Economet.* **174** 82–94