Bernoulli **22**(3), 2016, 1301–1330
DOI: 10.3150/14-BEJ692

Estimation of inverse autocovariance matrices for long memory processes

CHING-KANG ING[1,*], HAI-TANG CHIOU[2,**] and MEIHUI GUO[2,†]

[1]*Institute of Statistical Science, Academia Sinica, Taipei, Taiwan. E-mail:* [*]*cking@stat.sinica.edu.tw*
[2]*Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan.*
E-mail: [**]*d002040001@student.nsysu.edu.tw;* [†]*guomh@math.nsysu.edu.tw*

This work aims at estimating inverse autocovariance matrices of long memory processes admitting a linear representation. A modified Cholesky decomposition is used in conjunction with an increasing order autoregressive model to achieve this goal. The spectral norm consistency of the proposed estimate is established. We then extend this result to linear regression models with long-memory time series errors. In particular, we show that when the objective is to consistently estimate the inverse autocovariance matrix of the error process, the same approach still works well if the estimated (by least squares) errors are used in place of the unobservable ones. Applications of this result to estimating unknown parameters in the aforementioned regression model are also given. Finally, a simulation study is performed to illustrate our theoretical findings.

Keywords: inverse autocovariance matrix; linear regression model; long memory process; modified Cholesky decomposition

1. Introduction

Statistical inference for dependent data often involves consistent estimates of the inverse autocovariance matrix of a stationary time series. For example, by making use of a consistent estimate of the inverse autocovariance matrix of a short-memory time series (in the sense that its autocovariance function is absolutely summable), Wu and Pourahmadi [19] constructed estimates of the finite-past predictor coefficients of the time series and derived their error bounds. Moreover, in regression models with short-memory errors, Cheng, Ing and Yu [6] proposed feasible generalized least squares estimates (FGLSE) of the regression coefficients using a consistent estimate of the inverse autocovariance matrix of the error process. They then established an asymptotically efficient model averaging result based on the FGLSEs.

Having observed a realization $u_1, \ldots, u_n$ of a zero-mean stationary time series $\{u_t\}$, a natural estimate of its autocovariance function $\gamma_k = \operatorname{cov}(u_0, u_k)$ is the sample autocovariance function $\hat{\gamma}_k = n^{-1}\sum_{i=1}^{n-|k|} u_i u_{i+|k|}$, $k = 0, \pm 1, \ldots, \pm(n-1)$. Moreover, it is known that the k_n-dimensional sample autocovariance matrix $\breve{\boldsymbol{\Omega}}_{k_n} = (\hat{\gamma}_{i-j})_{1\le i,j\le k_n}$ and its inverse $\breve{\boldsymbol{\Omega}}_{k_n}^{-1}$ are consistent estimates of their population counterparts $\boldsymbol{\Omega}_{k_n} = (\gamma_{i-j})_{1\le i,j\le k_n}$ and $\boldsymbol{\Omega}_{k_n}^{-1}$, provided $k_n \ll n$ and $\sum_{k=1}^{\infty}|\gamma_k| < \infty$. See, for example, Berk [1], Shibata [15], Ing and Wei [9] and Wu and Pourahmadi [19]. However, when the objective is to estimate the n-dimensional autocovariance matrix $\boldsymbol{\Omega}_n$, Wu and Pourahmadi [19] showed that $\breve{\boldsymbol{\Omega}}_n$ is no longer consistent in the short-memory case. In addition, Palma and Pourahmadi [14] pointed out that this dilemma carries over to the long-memory case, assuming $\sum_{k=1}^{\infty}|\gamma_k| = \infty$. To circumvent this difficulty, Wu and Pourahmadi [19] proposed a banded covariance matrix estimate $\breve{\mathbf{\Omega}}_{n,l} = (\hat{\gamma}_{i-j}\mathbf{1}_{|i-j|\le l})_{1\le i,j\le n}$ of $\mathbf{\Omega}_n$, where $l \ge 0$ is an integer and called the banding parameter. When $\{u_t\}$ is a short-memory time series satisfying some mild conditions and $l = l_n$ grows to infinity with n at a suitable rate, they established consistency of $\breve{\mathbf{\Omega}}_{n,l}$ and $\breve{\mathbf{\Omega}}_{n,l}^{-1}$ under spectral norm. The result of Wu and Pourahmadi [19] was subsequently improved by Xiao and Wu [20] to a better convergence rate, and extended by McMurry and Politis [13] to tapered covariance matrix estimates. Alternatively, Bickel and Gel [2] considered a banded covariance matrix estimate $\breve{\mathbf{\Omega}}_{p_n,l}$ of $\mathbf{\Omega}_{p_n}$, with $p_n = o(n)$. Assuming that $\{u_t\}$ is a stationary short-memory AR(∞) process, they obtained $\breve{\mathbf{\Omega}}_{p_n,l}$'s consistency under the Frobenius norm, provided $l = l_n$ tends to infinity sufficiently slowly.

Although the banded and tapered covariance matrix estimates work well for the short-memory time series, they are not necessarily suitable for the long-memory case because the autocovariance function of the latter is not absolutely summable. As a result, the banded and tapered matrix estimates may incur large truncation errors, which prevent them from achieving consistency. A major repercussion of this inconsistency property is that a consistent estimate of $\mathbf{\Omega}_n^{-1}$ can no longer be obtained by inverting $\breve{\mathbf{\Omega}}_{n,l}$ or its tapered version. On the other hand, since the spectral densities of most long-memory time series encountered in common practice are bounded away from zero, it follows from Proposition 4.5.3 of Brockwell and Davis [4] that

$$\sup_{k\ge 1}\left\|\mathbf{\Omega}_k^{-1}\right\|_2 < \infty, \tag{1.1}$$

where for a k-dimensional matrix A, $\|A\|_2 = \sup_{\{\mathbf{x}\in R^k:\ \mathbf{x}'\mathbf{x}=1\}}(\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x})^{1/2}$ denotes its spectral norm. Motivated by (1.1), this paper aims to propose a direct estimate of $\mathbf{\Omega}_n^{-1}$ and establish its consistency in the spectral norm sense, which is particularly relevant under the long-memory setup.

To fix ideas, assume

$$u_t = \sum_{j=0}^{\infty}\psi_j w_{t-j}, \tag{1.2}$$

where $\psi_0 = 1$ and $\{w_t\}$ is a martingale difference sequence with $E(w_t) = 0$ and $E(w_t^2) = \sigma^2$ for all t, and

$$\psi_j = \mathrm{O}\big(j^{-1+d}\big), \tag{1.3}$$

with d satisfying $0 < d < 1/2$. We shall also assume that $\{u_t\}$ admits an AR(∞) representation,

$$u_t = \sum_{i=1}^{\infty}a_i u_{t-i} + w_t, \tag{1.4}$$

where

$$a_i = \mathrm{O}\big(i^{-1-d}\big). \tag{1.5}$$

In view of (1.3), the autocovariance function of $\{u_t\}$ obeys

$$\gamma_k = \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} \sigma^2 = \mathrm{O}\big(|k|^{-1+2d}\big), \tag{1.6}$$

which may not be absolutely summable. A well-known model satisfying (1.2)–(1.5) is the FARIMA(p, d, q) processes,

$$\phi(B)(1-B)^d u_t = \theta(B) w_t, \tag{1.7}$$

where B is the backshift operator, $\phi(z)$ and $\theta(z)$ 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. Note that when (1.7) is assumed, the spectral density of $\{u_t\}$, $f_u(\lambda)$, satisfies

$$\inf_{\lambda \in [-\pi, \pi]} f_u(\lambda) > 0, \tag{1.8}$$

from which (1.1) follows.

Let

$$\sigma_k^2 = E(u_t - a_{k,1} u_{t-1} - \cdots - a_{k,k} u_{t-k})^2, \tag{1.9}$$

where $k \geq 1$ and

$$(a_{k,1}, \ldots, a_{k,k}) = \operatorname*{arg\,min}_{(\alpha_1, \ldots, \alpha_k) \in R^k} E(u_t - \alpha_1 u_{t-1} - \cdots - \alpha_k u_{t-k})^2. \tag{1.10}$$

To directly estimate $\mathbf{\Omega}_n^{-1}$, we start by defining the modified Cholesky decomposition (see, e.g., Berk [1] and Wu and Pourahmadi [18]) of $\mathbf{\Omega}_n$:

$$\mathbf{T}_n \mathbf{\Omega}_n \mathbf{T}_n' = \mathbf{D}_n,$$

where

$$\mathbf{D}_n = \operatorname{diag}\big(\gamma_0, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{n-1}^2\big),$$

and $\mathbf{T}_n = (t_{ij})_{1 \leq i, j \leq n}$ is a lower triangular matrix satisfying

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

Hence,

$$\mathbf{\Omega}_n^{-1} = \mathbf{T}_n' \mathbf{D}_n^{-1} \mathbf{T}_n. \tag{1.11}$$

Because there are too many parameters in $\mathbf{T}_n$ and $\mathbf{D}_n$, we are led to consider a banded Cholesky decomposition of $\mathbf{\Omega}_n^{-1}$,

$$\mathbf{\Omega}_n^{-1}(k) = \mathbf{T}_n'(k) \mathbf{D}_n^{-1}(k) \mathbf{T}_n(k), \tag{1.12}$$

where $1 \le k \ll n$ is referred to as the banding parameter and allowed to grow to infinity with n,

$$\mathbf{D}_n(k) = \operatorname{diag}\bigl(\gamma_0, \sigma_1^2, \ldots, \sigma_k^2, \ldots, \sigma_k^2\bigr),$$

and $\mathbf{T}_n(k) = (t_{ij}(k))_{1 \le i,j \le n}$ with

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

We propose estimating $\mathbf{\Omega}_n^{-1}$ using the sample counterpart of (1.12),

$$\hat{\mathbf{\Omega}}_n^{-1}(k) := \hat{\mathbf{T}}_n'(k)\hat{\mathbf{D}}_n^{-1}(k)\hat{\mathbf{T}}_n(k), \tag{1.13}$$

where $\hat{\mathbf{T}}_n(k)$ and $\hat{\mathbf{D}}_n(k)$ are obtained by plugging in the least squares estimates of the coefficients in $\mathbf{T}_n(k)$ and the corresponding residual variances in $\mathbf{D}_n(k)$; see Section 3 for more details.

Under (1.2)–(1.5), this paper establishes

$$\bigl\|\hat{\mathbf{\Omega}}_n^{-1}(k) - \mathbf{\Omega}_n^{-1}\bigr\|_2 = o_p(1), \tag{1.14}$$

with $k = K_n \to \infty$ satisfying (3.16). To appreciate the subtlety of (1.14), note that if m independent realizations $\mathbf{U}^{(1)} = (u_1^{(1)}, \ldots, u_n^{(1)})', \ldots, \mathbf{U}^{(m)} = (u_1^{(m)}, \ldots, u_n^{(m)})'$ of $\{u_t\}$ are available, Bickel and Levina [3] introduced alternative estimates $\check{\mathbf{T}}_{n,m}(k)$ and $\check{\mathbf{D}}_{n,m}(k)$ of $\mathbf{T}_n(k)$ and $\mathbf{D}_n(k)$ through a multivariate analysis approach, where $k < m < n$. More specifically, set $\tilde{\mathbf{U}}_j = (u_j^{(1)}, \ldots, u_j^{(m)})'$ and denote the regression coefficients of $\tilde{\mathbf{U}}_j$ on $\tilde{\mathbf{U}}_{j-1}, \ldots, \tilde{\mathbf{U}}_{\max\{j-k,1\}}$ by $\check{\mathbf{a}}_j$. Then $\check{\mathbf{T}}_{n,m}(k)$ and $\check{\mathbf{D}}_{n,m}(k)$, respectively, are obtained by replacing the coefficients in the ith row of $\mathbf{T}_n(k)$ with $-\check{\mathbf{a}}_i$, and ith diagonal element of $\mathbf{D}_n(k)$ with the corresponding residual variance, where $i = 2, \ldots, n$. Bickel and Levina [3] also showed that the resultant estimate $\check{\mathbf{\Omega}}_{n,m}^{-1}(k) = \check{\mathbf{T}}_{n,m}'(k)\check{\mathbf{D}}_{n,m}^{-1}(k)\check{\mathbf{T}}_{n,m}(k)$ of $\mathbf{\Omega}_n^{-1}$ has the property

$$\bigl\|\check{\mathbf{\Omega}}_{n,m}^{-1}(k) - \mathbf{\Omega}_n^{-1}\bigr\|_2 = o_p(1), \tag{1.15}$$

under $m \to \infty$, $m^{-1}\log n \to 0$, $k = K_{n,m} \asymp (m/\log n)^{\theta}$ for some $0 < \theta < 1/2$, (1.1), and

$$\sup_{k \ge 1} \|\mathbf{\Omega}_k\|_2 < \infty, \tag{1.16}$$

where $g(x) \asymp h(x)$ means that there exists a constant $0 < C < \infty$ such that $C \le \liminf_{x\to\infty} h(x)/g(x) \le \limsup_{x\to\infty} h(x)/g(x) \le C^{-1}$. Since (1.16) fails to hold for long-memory processes like (1.7) and $m \to \infty$ is needed in (1.15), the most distinctive feature of (1.14) is that it holds for one ($m = 1$) realization, without imposing (1.16). It is also noteworthy that Cai, Ren and Zhou [5] have recently established the optimal rate of convergence for estimating the inverse of a Toeplitz covariance matrix under the spectral norm. However, the covariance matrix associated with (1.7) is still precluded by the class of matrices considered in their paper, which needs to obey assumptions like (1.16) and (1.1).

The rest of the paper is organized as follows. In Section 2, we analyze the difference between $\mathbf{\Omega}_n^{-1}(k)$ and $\mathbf{\Omega}_n^{-1}$. In particular, by deriving convergence rates of $\|\mathbf{T}_n(k) - \mathbf{T}_n\|_2$ and $\|\mathbf{D}_n(k) - \mathbf{D}_n\|_2$, we obtain a convergence rate of $\|\mathbf{\Omega}_n^{-1}(k) - \mathbf{\Omega}_n^{-1}\|_2$, which plays an indispensable role in the proof of (1.14). Section 3 is devoted to proving (1.14). By establishing a number of sharp bounds for the higher moments of the quadratic forms in u_t, we obtain a convergence rate of $\|\hat{\mathbf{\Omega}}_n^{-1}(k) - \mathbf{\Omega}_n^{-1}(k)\|_2$, which, in conjunction with the results in Section 2, leads to a convergence rate of $\|\hat{\mathbf{\Omega}}_n^{-1}(k) - \mathbf{\Omega}_n^{-1}\|_2$, and hence (1.14). In Section 4, the results in Section 3 are extended to regression models with long-memory errors satisfying (1.2)–(1.5). Specifically, we show that when the unobservable long-memory errors are replaced by the corresponding least squares residuals, our estimate of $\mathbf{\Omega}_n^{-1}$ still has the same convergence rate, without imposing *any* assumptions on the design matrices. Moreover, the estimated matrix is applied to construct an estimate of the finite-past predictor coefficient vector of the error process, and an FGLSE of the regression coefficient vector. Rates of convergence of the latter two estimates are also derived in a somewhat intricate way. In Section 5, we present a Monte Carlo study of the finite-sample performance of the proposed inverse matrix estimates.

2. Bias analysis of banded Cholesky factors

Our analysis of $\|\mathbf{\Omega}_n^{-1} - \mathbf{\Omega}_n^{-1}(k)\|_2$ is reliant on the following two conditions on $a_{m,i}$'s defined in (1.10).

(i) There exists $C_1 > 0$ such that for any $1 \le i \le m < \infty$,

$$\left|\frac{a_{m,i}}{a_i}\right| \le C_1\left(\frac{m}{m-i+1}\right)^d. \tag{2.1}$$

(ii) There exist $C_2 > 0$ and $0 < \delta < 1$ such that for any $1 \le i \le \delta m$ and $1 \le m < \infty$,

$$\left|\frac{a_{m,i}}{a_i} - 1\right| \le C_2\frac{i}{m}. \tag{2.2}$$

Some comments on (2.1) and (2.2) are in order. Note first that (2.1) and (2.2), together with (1.5), immediately imply that there exists $C > 0$ such that for any $k = 1, 2, \ldots,$

$$\sum_{i=1}^{k} |a_{k,i}| \le C, \tag{2.3}$$

which will be used frequently in the sequel. Throughout the rest of the paper, C denotes a generic positive constant independent of any unbounded index sets of positive integers. These two conditions assert that the finite-past predictor coefficients $a_{m,i}, i = 1, \ldots, m$ approach to the corresponding infinite-past predictor coefficients $a_1, a_2, \ldots$ in a nonuniform way. More specifically, they require that $a_{m,i}/a_i$ is very close to 1 when $i = \mathrm{o}(m)$, but has order of magnitude $m^{(1-\theta)d}$

when $m - i \asymp m^{\theta}$ with $0 \leq \theta < 1$. 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 $m + 1$. In fact, when $\{u_t\}$ is an I(d) process with $0 < d < 1/2$, (2.1) and (2.2) follow directly from the proof of Theorem 13.2.1 of Brockwell and Davis [4]. In the following, we shall show that (2.1) and (2.2) are satisfied by model (1.7). To this end, we need an auxiliary lemma.

Lemma 2.1. *Assume* (1.2), (1.4),

$$\psi_i \sim i^{-1+d}\mathcal{L}(i) \tag{2.4}$$

and

$$a_i \sim \frac{i^{-1-d} d \sin(\pi d)}{\pi \mathcal{L}(i)}, \tag{2.5}$$

where $g(x) \sim h(x)$ *if* $\lim_{x\to\infty} g(x)/h(x) = 1$ *and* $\mathcal{L}(x)$ *is a positive slowly varying function, namely*, $\lim_{x\to\infty} \mathcal{L}(\lambda x)/\mathcal{L}(x) = 1$ *for all* $\lambda > 0$. *Then for all large* m,

$$\max_{1\leq i\leq m} \left| \frac{m(a_{m,i} - a_i)}{\sum_{j=i\wedge(m+1-i)}^{\infty} |a_j|} \right| \leq C,$$

where $u \wedge v = \min\{u, v\}$.

Proof. For $h, j \in \mathbf{N} \cup \{0\}$, we define

$$d_s(h, j) = \begin{cases} \xi_{h+j}, & \text{if } s = 1; \\ \displaystyle\sum_{v=0}^{\infty} \xi_{h+j+v} d_{s-1}(h, v), & \text{if } s = 2, 3, \ldots, \end{cases}$$

where $\xi_t = \sum_{v=0}^{\infty} \psi_v a_{v+t}$ for $t = 0, 1, \ldots$. By Theorem 2.9 of Inoue and Kasahara [11], we obtain

$$\begin{aligned} &m\left|(a_{m,i} - a_i)\right| \\ &\quad = m\left|\sum_{s=1}^{\infty}\sum_{j=0}^{\infty} a_{i+j} d_{2s}(m+1, j) + \sum_{s=1}^{\infty}\sum_{j=0}^{\infty} a_{m+1-i+j} d_{2s-1}(m+1, j)\right|. \end{aligned} \tag{2.6}$$

Let $\kappa > 1$ satisfy $0 < \kappa \sin(\pi d) < 1$. According to Proposition 3.2(i) of Inoue and Kasahara [11], there exists $N \in \mathbf{N}$ such that

$$0 < d_s(h, j) \leq \frac{g_s(0)\{\kappa \sin(\pi d)\}^s}{h}, \qquad j \in \mathbf{N} \cup \{0\}, s \in \mathbf{N}, h \geq N, \tag{2.7}$$

where

$$g_s(x) = \begin{cases} \dfrac{1}{\pi(1+x)}, & \text{if } s = 1; \\ \dfrac{1}{\pi^2}\displaystyle\int_0^\infty \frac{dv_1}{(v_1+1)(v_1+1+x)}, & \text{if } s = 2; \\ \dfrac{1}{\pi^s}\displaystyle\int_0^\infty \cdots \int_0^\infty \frac{1}{v_{s-1}+1} & \\ \qquad \times \left\{\displaystyle\prod_{j=1}^{s-2} \frac{1}{v_{j+1}+v_j+1}\right\} \dfrac{1}{v_1+1+x}\, dv_{s-1}\cdots dv_1, & \text{if } s = 3, 4, \ldots. \end{cases}$$

Thus, for $m \geq N$ and $i = 1, 2, \ldots, m$,

$$\begin{aligned} & m\left|\sum_{s=1}^\infty \sum_{j=0}^\infty a_{i+j} d_{2s}(m+1, j)\right| \\ & \quad \leq m \sum_{j=0}^\infty |a_{i+j}| \left(\sum_{s=1}^\infty \frac{g_{2s}(0)\{\kappa \sin(\pi d)\}^{2s}}{m+1}\right) \\ & \quad = \frac{m}{m+1} \sum_{j=i}^\infty |a_j| \sum_{s=1}^\infty g_{2s}(0)\big\{\kappa \sin(\pi d)\big\}^{2s} \leq C \sum_{j=i}^\infty |a_j|, \end{aligned} \tag{2.8}$$

where the first inequality is by (2.7) and the last one is by Lemma 3.1(i) of Inoue and Kasahara [11]. Similarly, (2.7) and Lemma 3.1(ii) of Inoue and Kasahara [11] imply

$$m\left|\sum_{s=1}^\infty \sum_{j=0}^\infty a_{m+1-i+j} d_{2s-1}(m+1, j)\right| \leq C \sum_{j=m+1-i}^\infty |a_j|. \tag{2.9}$$

Combining (2.6), (2.8) and (2.9) yields the desired conclusion. □

Remark 2.1. Theorem 3.3 of Inoue and Kasahara [11] shows that for any fixed integer i,

$$\lim_{m\to\infty} m(a_{m,i} - a_i) = d^2 \sum_{j=i}^\infty a_j. \tag{2.10}$$

Therefore, Lemma 2.1 can be viewed as a *uniform* extension of (2.10).

Remark 2.2. Note that (1.3) and (1.5) are fulfilled by (2.4) and (2.5) if

$$0 < \inf_i \mathcal{L}(i) \leq \sup_i \mathcal{L}(i) < \infty. \tag{2.11}$$

By making use of Lemma 2.1, the next theorem shows that (2.1) and (2.2) are met by (2.4) and (2.5) with $\mathcal{L}(i)$ obeying (2.11). Since the coefficients of the MA and AR representations of (1.7) take the form of (2.4) and (2.5), respectively, for which $\mathcal{L}(i)$ is a constant function (see Corollary 3.1 of Kokoszka and Taqqu [12] and Example 2.6 of Inoue and Kasahara [11]), this theorem guarantees that (1.7) satisfies (2.1) and (2.2), confirming the flexibility of these two conditions.

Theorem 2.1. *Under the same assumptions as in Lemma 2.1 with $\mathcal{L}(i)$ satisfying (2.11), we have (2.1) and (2.2).*

Proof. It suffices to show that (2.1) and (2.2) hold for all sufficiently large m. By Lemma 2.1 and (2.11), it follows that for all $1 \le i \le m$ and all large m,

$$\left|m(a_{m,i} - a_i)\right| \le C \max\left\{i^{-d}, (m-i+1)^{-d}\right\},$$

yielding

$$\left|\frac{a_{m,i}}{a_i}\right| = \left|\frac{m(a_{m,i} - a_i)}{ma_i} + 1\right| \le C \frac{\max\{i^{-d}, (m-i+1)^{-d}\}}{mi^{-1-d}} + 1 \le C\left(\frac{m}{m-i+1}\right)^d.$$

Therefore, (2.1) follows. Similarly, for all $1 \le i \le \delta m$ with $0 < \delta < 1$ and all large m,

$$\left|\frac{a_{m,i} - a_i}{a_i}\right| \le C \frac{\max\{i^{-d}, (m-i+1)^{-d}\}}{mi^{-1-d}} \le C\frac{i}{m},$$

which leads to (2.2). Thus the proof is complete. □

Throughout the rest of this paper, let K_n denote a sequence of numbers satisfying $K_n \to \infty$ and $K_n/n \to 0$ as $n \to \infty$. We are now ready to provide upper bounds for $\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_2$ and $\|\mathbf{D}_n^{-1} - \mathbf{D}_n(K_n)^{-1}\|_2$ in Propositions 2.1 and 2.2, which in turn lead to a rate of convergence of $\|\boldsymbol{\Omega}_n^{-1} - \boldsymbol{\Omega}_n^{-1}(K_n)\|_2$ in Theorem 2.2. Before proceeding, we need two technical lemmas.

Lemma 2.2. *Assume (1.5), (2.1) and (2.2). Then*

(i) $\sum_{j=K_n+1}^{k} |a_{k,j}| \le CK_n^{-d}$ *for any* $K_n + 1 \le k \le n-1$.
(ii) $\sum_{j=1}^{K_n} |a_{k,j} - a_{K_n,j}| \le CK_n^{-d}$ *for any* $K_n + 1 \le k \le n-1$.
(iii) $\sum_{j=\max(1,K_n+1-k)}^{K_n} |a_{j+k,j} - a_{K_n,j}| \le CK_n^{-d}$ *for any* $1 \le k \le n - K_n - 1$.
(iv) $\sum_{j=1}^{n-k-1} |a_{j+k,j} - a_{K_n,j}| \le CK_n^{-d}$ *for any* $n - K_n \le k \le n-2$.

Proof. The proof is straightforward, and thus omitted. □

Lemma 2.3. *Assume (1.2)–(1.5). Then for any $k \ge 1$, $\sigma_k^2 - \sigma^2 \le Ck^{-1}$, where σ_k^2 is defined in (1.9).*

Proof. In view of (1.4) and (1.10), it follows that for any $k \geq 1$, $\sigma_k^2 - \sigma^2 \leq E(\sum_{j=k+1}^{\infty} a_j u_{t-j})^2$. In addition, by (1.6) (which is ensured by (1.2) and (1.3)), (1.5), and Theorem 2.1 of Ing and Wei [10], one has for any $k \geq 1$ and $m \geq k+1$, $E(\sum_{j=k+1}^{m} a_j u_{t-j})^2 \leq Ck^{-1}$, which, together with the previous inequality, gives the desired conclusion. □

Proposition 2.1. *Under the same assumptions as in Lemma 2.2,*

(i) $\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_2 = \mathrm{O}((K_n^{-d} \log n)^{1/2})$.
(ii) $\|\mathbf{T}_n(K_n)\|_2 = \mathrm{O}((\log K_n)^{1/2})$.

Proof. Let $\|\mathbf{B}\|_k = \max_{\|\mathbf{z}\|_k=1} \|\mathbf{Bz}\|_k$ denote the k-norm of an $h \times h$ matrix $\mathbf{B}$, where $\|\mathbf{z}\|_k = (\sum_{i=1}^{h} |z_i|^k)^{1/k}$ is the k-norm of the vector $\mathbf{z} = (z_1, \ldots, z_h)'$. Then, by Lemma 2.2(i) and (ii),

$$\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_\infty = \max_{K_n+1 \leq i \leq n-1} \sum_{j=K_n+1}^{i} |a_{i,j}| + \sum_{j=1}^{K_n} |a_{K_n,j} - a_{i,j}| = \mathrm{O}\big(K_n^{-d}\big).$$

Moreover, $\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_1$ is the maximum of

$$\max_{0 \leq k \leq n-K_n-1} \left\{ \sum_{i=0}^{n-K_n-k-2} |a_{K_n+1+i+k, K_n+1+i}| + \sum_{j=\max(1, K_n+1-k)}^{K_n} |a_{j+k,j} - a_{K_n,j}| \right\}$$

and $\max_{n-K_n \leq k \leq n-2} \sum_{j=1}^{n-k-1} |a_{j+k,j} - a_{K_n,j}|$. By (2.1) and Lemma 2.2(iii) and (iv),

$$\max_{0 \leq k \leq n-K_n-1} \sum_{i=0}^{n-K_n-k-2} |a_{K_n+1+i+k, K_n+1+i}| = \mathrm{O}(\log n),$$

$$\max_{0 \leq k \leq n-K_n-1} \sum_{j=\max(1, K_n+1-k)}^{K_n} |a_{j+k,j} - a_{K_n,j}| = \mathrm{O}\big(K_n^{-d}\big)$$

and

$$\max_{n-K_n \leq k \leq n-2} \sum_{j=1}^{n-k-1} |a_{j+k,j} - a_{K_n,j}| = \mathrm{O}\big(K_n^{-d}\big).$$

Hence, $\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_1 = \mathrm{O}(\log n)$. The proof of (i) is completed by

$$\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_2 \leq \big(\|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_1 \|\mathbf{T}_n - \mathbf{T}_n(K_n)\|_\infty\big)^{1/2}.$$

Similarly, it can be shown that $\|\mathbf{T}_n(K_n)\|_\infty = \mathrm{O}(1)$ and $\|\mathbf{T}_n(K_n)\|_1 = \mathrm{O}(\log K_n)$, yielding (ii). □

Proposition 2.2. *Under the same assumptions as in Lemma 2.3,*

(i) $\|\mathbf{D}_n^{-1} - \mathbf{D}_n^{-1}(K_n)\|_2 = \mathrm{O}(K_n^{-1})$.

(ii) $\|\mathbf{D}_n^{-1}(K_n)\|_2 = \mathrm{O}(1)$.

Proof. Equation (i) is an immediate consequence of Lemma 2.3. Equation (ii) follows from $\|\mathbf{D}_n^{-1}(K_n)\|_2 = \max_{0\le k\le K_n} \sigma_k^{-2} \le \sigma^{-2}$, where $\sigma_0^2 = \gamma_0$. □

Theorem 2.2. *Assume* (1.2)–(1.5), (2.1) *and* (2.2). *Suppose*

$$\frac{\log n \log K_n}{K_n^d} = \mathrm{o}(1). \tag{2.12}$$

Then

$$\left\|\mathbf{\Omega}_n^{-1} - \mathbf{\Omega}_n^{-1}(K_n)\right\|_2 = \mathrm{O}\left(\log n K_n^{-d} \log K_n\right)^{1/2} = \mathrm{o}(1). \tag{2.13}$$

Moreover, if (1.8) *is assumed*,

$$\left\|\mathbf{\Omega}_n^{-1}(K_n)\right\|_2 = \mathrm{O}(1). \tag{2.14}$$

Proof. Equation (2.13) follows directly from Propositions 2.1 and 2.2,

$$\begin{aligned}
\left\|\mathbf{\Omega}_n^{-1} - \mathbf{\Omega}_n^{-1}(K_n)\right\|_2 \le{} & \left\|\mathbf{T}_n - \mathbf{T}_n(K_n)\right\|_2 \left\|\mathbf{D}_n^{-1}\right\|_2 \left(\left\|\mathbf{T}_n - \mathbf{T}_n(K_n)\right\|_2 + \left\|\mathbf{T}_n(K_n)\right\|_2\right) \\
& + \left\|\mathbf{T}_n(K_n)\right\|_2 \left\|\mathbf{D}_n^{-1} - \mathbf{D}_n^{-1}(K_n)\right\|_2 \left(\left\|\mathbf{T}_n - \mathbf{T}_n(K_n)\right\|_2 + \left\|\mathbf{T}_n(K_n)\right\|_2\right) \\
& + \left\|\mathbf{T}_n(K_n)\right\|_2 \left\|\mathbf{D}_n^{-1}(K_n)\right\|_2 \left\|\mathbf{T}_n - \mathbf{T}_n(K_n)\right\|_2,
\end{aligned}$$

and (2.12). Equations (2.13) and (1.1) (which is ensured by (1.8)) further lead to (2.14). □

3. Main results

In the sequel, the following assumptions on the innovation process $\{w_t\}$ of (1.2) are frequently used:

(M1) $\{w_t, \mathcal{F}_t\}$ is a martingale difference sequence, where $\mathcal{F}_t$ is an increasing sequence of σ-field generated by $w_s, s \le t$.

(M2) $E(w_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s.

(M3) For some $q \ge 1$, there is a constant $C_q > 0$ such that

$$\sup_{-\infty<t<\infty} E\left(|w_t|^{4q}|\mathcal{F}_{t-1}\right) \le C_q \qquad \text{a.s.}$$

As mentioned in Section 1, $\hat{\mathbf{T}}_n(K_n)$ is obtained by replacing $\mathbf{a}(k) = (a_{k,1}, \dots, a_{k,k})'$ in $\mathbf{T}_n(K_n)$ with the corresponding the least squares estimates $\hat{\mathbf{a}}(k) = (\hat{a}_{k,1}, \dots, \hat{a}_{k,k})'$, where $k = 1, \dots, K_n$ and

$$(\hat{a}_{k,1}, \dots, \hat{a}_{k,k}) = \operatorname*{arg\,min}_{(\alpha_1,\dots,\alpha_k)\in R^k} \sum_{t=k+1}^{n} (u_t - \alpha_1 u_{t-1} - \alpha_2 u_{t-2} - \dots - \alpha_k u_{t-k})^2.$$

Similarly, $\hat{\mathbf{D}}_n(K_n)$ is obtained by replacing σ_k^2 in $\mathbf{D}_n(K_n)$ with $\hat{\sigma}_k^2$, where $k = 0, \dots, K_n$ and

$$\hat{\sigma}_0^2 = (n-1)^{-1} \sum_{t=1}^{n} (u_t - \bar{u})^2, \qquad \bar{u} = n^{-1} \sum_{t=1}^{n} u_i,$$
$$\hat{\sigma}_k^2 = (n-k)^{-1} \sum_{t=k+1}^{n} \left(u_t - \sum_{j=1}^{k} \hat{a}_{k,j} u_{t-j} \right)^2.$$

Recall $\hat{\boldsymbol{\Omega}}_n^{-1}(K_n) = \hat{\mathbf{T}}_n'(K_n) \hat{\mathbf{D}}_n^{-1}(K_n) \hat{\mathbf{T}}_n(K_n)$. The objective of this section is to show that $\|\hat{\boldsymbol{\Omega}}_n^{-1}(K_n) - \boldsymbol{\Omega}_n^{-1}\|_2 = \mathrm{o}_p(1)$ in Theorem 3.1. To this end, we develop rates of convergence of $\|\hat{\mathbf{T}}_n(K_n) - \mathbf{T}_n(K_n)\|_2$ and $\|\hat{\mathbf{D}}_n^{-1}(K_n) - \mathbf{D}_n^{-1}(K_n)\|_2$ in Propositions 3.1 and 3.2, respectively, whose proofs are heavily reliant on the following four lemmas, Lemmas 3.1–3.4.

Lemma 3.1. *Assume* (1.2)–(1.5) *and* (M1)–(M3). *Let* $\mathbf{U}_t(k) = (u_t, u_{t-1}, \dots, u_{t-k+1})'$ *and* $w_{k,t+1} = u_{t+1} - \mathbf{a}(k)'\mathbf{U}_t(k)$. *Then for any* $1 \le k \le n-1$,

$$E \left\| \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k)(w_{k,t+1} - w_{t+1}) \right\|_2^{2q} \le C \left(\frac{1}{n-k} \right)^{q(1-2d)} \tag{3.1}$$

and

$$E \left\| \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k) w_{k,t+1} \right\|_2^{2q} \le C \left\{ \left(\frac{1}{n-k} \right)^{q(1-2d)} + \left(\frac{k}{n-k} \right)^q \right\}. \tag{3.2}$$

Moreover, for $\theta > 1/q$,

$$\max_{1 \le k \le K_n} \left\| \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k) w_{k,t+1} \right\|_2^2 = \mathrm{O}_p \left(\frac{K_n^\theta}{n^{1-2d}} + \frac{K_n^{1+\theta}}{n} \right). \tag{3.3}$$

Proof. By (1.6), Lemma 2.3 and an argument similar to that used in Lemma 3 of Ing and Wei [9], one has for any $1 \le k \le n-1$,

$$\begin{aligned} &E \left\| \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k)(w_{k,t+1} - w_{t+1}) \right\|_2^{2q} \\ &\quad \le C(n-k)^{-q} k^q \left\{ \left(\sigma_k^2 - \sigma^2 \right) \sum_{i=-(n-k)+1}^{n-k-1} |\gamma_i| \right\}^q \le C \left(\frac{1}{n-k} \right)^{q(1-2d)}, \end{aligned}$$

which gives (3.1). Equation (3.2) follows from (3.1) and for any $1 \le k \le n-1$,

$$E \left\| \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k) w_{t+1} \right\|_2^{2q} \le C k^q (n-k)^{-q}, \tag{3.4}$$

whose proof is exactly same as that of Lemma 4 of Ing and Wei [9]. To show (3.3), note that by (3.2) and $K_n = \mathrm{o}(n)$,

$$E \max_{1 \le k \le K_n} \left\| \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k) w_{k,t+1} \right\|_2^{2q} \le C \sum_{k=1}^{K_n} \left\{ n^{-q(1-2d)} + k^q n^{-q} \right\} \le C \left\{ K_n n^{-q(1-2d)} + K_n^{q+1} n^{-q} \right\}.$$

This, together with $\theta > 1/q$, gives the desired conclusion (3.3). □

Remark 3.1. Lemma A.1 of Godet [8] establishes an inequality closely related to (3.1). In particular, the inequality yields

$$E \left\| \frac{1}{\sqrt{n-k}} \sum_{t=k}^{n-1} \mathbf{U}_t(k)(w_{k,t+1} - w_{t+1}) \right\|_2^{2q} \le C \left\{ k(n-k)^{2d} \right\}^q \left(\sigma_k^2 - \sigma^2 \right)^q.$$

This bound together with Lemma 2.3 also leads to (3.1).

Lemma 3.2. *Let*

$$\hat{\mathbf{\Gamma}}_{k,n} = \frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k) \mathbf{U}_t(k)'.$$

Assume (1.2), (1.3), (1.8) *and* (M1)–(M3) *with* $q = 1$. *Suppose*

$$K_n = \begin{cases} \mathrm{o}\left(n^{1/2}\right), & \textit{if } 0 < d < 1/4; \\ \mathrm{o}\left((n/\log n)^{1/2}\right), & \textit{if } d = 1/4; \\ \mathrm{o}\left(n^{1-2d}\right), & \textit{if } 1/4 < d < 1/2. \end{cases} \tag{3.5}$$

Then

$$\left\| \hat{\mathbf{\Gamma}}_{K_n,n}^{-1} \right\|_2 = \mathrm{O}_p(1). \tag{3.6}$$

Proof. By the first moment bound theorem of Findley and Wei [7], (1.6) and an argument similar to that used in Lemma 2 of Ing and Wei [9], it follows that

$$E \| \hat{\mathbf{\Gamma}}_{K_n,n} - \mathbf{\Omega}_{K_n} \|_2^2 = \begin{cases} \mathrm{O}\left(K_n^2 (n-K_n)^{-1}\right), & \text{if } 0 < d < 1/4; \\ \mathrm{O}\left(K_n^2 (n-K_n)^{-1} \log(n-K_n)\right), & \text{if } d = 1/4; \\ \mathrm{O}\left(K_n^2 (n-K_n)^{-2+4d}\right), & \text{if } 1/4 < d < 1/2. \end{cases} \tag{3.7}$$

Combining this, (3.5) and (1.8) leads to (3.6). □

Lemma 3.3. *Under the same assumptions as in Theorem 2.2, one has for any $k \geq 1$ and $m = 0, \pm 1, \pm 2, \dots$, $\gamma_{\tau_k}(m) = C(|m|+1)^{-1+2d}$, where with $\tau_{k,t} = u_{t+1} - w_{t+1} - \mathbf{a}'(k)\mathbf{U}_t(k) = w_{k,t+1} - w_{t+1}$, $\gamma_{\tau_k}(m) = E(\tau_{k,1}\tau_{k,m+1})$.*

Proof. This result follows by a tedious but direct calculation. The details are omitted. □

Lemma 3.4. *Assume that (2.1), (2.2), and the assumptions of Lemma 3.1 hold. Then, for any $1 \leq k \leq n-1$,*

$$E\left|\frac{1}{n-k}\sum_{t=k}^{n-1} w_{k,t+1}^2 - \sigma_k^2\right|^{2q} \leq \begin{cases} C(n-k)^{-q}, & \text{if } 0 < d < 1/4, \\ C\left(\left\{(n-k)^{-1}\log(n-k)\right\}^q\right), & \text{if } d = 1/4, \\ C(n-k)^{-2q+4qd}, & \text{if } 1/4 < d < 1/2. \end{cases} \tag{3.8}$$

Moreover, for $\theta > 1/(2q)$,

$$\max_{1\leq k\leq K_n}\left|\frac{1}{n-k}\sum_{t=k}^{n-1} w_{k,t+1}^2 - \sigma_k^2\right| = \begin{cases} \mathrm{O}_p\left(K_n^\theta n^{-1/2}\right), & \text{if } 0 < d < 1/4, \\ \mathrm{O}_p\left(K_n^\theta (\log n)^{1/2} n^{-1/2}\right), & \text{if } d = 1/4, \\ \mathrm{O}_p\left(K_n^\theta n^{-1+2d}\right), & \text{if } 1/4 < d < 1/2. \end{cases} \tag{3.9}$$

Proof. To show (3.8), note first that

$$E\left|\frac{1}{n-k}\sum_{t=k}^{n-1} w_{k,t+1}^2 - \sigma_k^2\right|^{2q} \leq C\left(E\left|(A1)\right|^{2q} + E\left|(A2)\right|^{2q} + E\left|(A3)\right|^{2q}\right), \tag{3.10}$$

where $E|(A1)|^{2q} = E|\frac{1}{n-k}\sum_{t=k}^{n-1} w_{t+1}^2 - \sigma^2|^{2q}$, $E|(A2)|^{2q} = E|\frac{2}{n-k}\sum_{t=k}^{n-1} w_{t+1}\tau_{k,t}|^{2q}$, and $E|(A3)|^{2q} = E|\frac{1}{n-k}\sum_{t=k}^{n-1}\tau_{k,t}^2 - (\sigma_k^2 - \sigma^2)|^{2q}$. It is clear that for any $1 \leq k \leq n-1$,

$$E\left|(A1)\right|^{2q} \leq C(n-k)^{-q}. \tag{3.11}$$

In addition, the first moment bound theorem of Findley and Wei [7] implies that for any $1 \leq k \leq n-1$,

$$E\left|(A2)\right|^{2q} \leq C\left((n-k)^{-1}\gamma_{\tau_k}(0)\right)^q,$$
$$E\left|(A3)\right|^{2q} \leq C\left\{(n-k)^{-1}\sum_{j=0}^{n-k-1}\gamma_{\tau_k}^2(j)\right\}^q,$$

which, together with Lemmas 2.3 and 3.3, (3.10) and (3.11), yield (3.8). Equation (3.9) follows immediately from (3.8) and an argument similar to that used to prove (3.3). The details are omitted. □

We are now ready to establish rates of convergence of $\|\hat{\mathbf{T}}_n(K_n) - \mathbf{T}_n(K_n)\|_2$ and $\|\hat{\mathbf{D}}_n^{-1}(K_n) - \mathbf{D}_n^{-1}(K_n)\|_2$.

Proposition 3.1. *Assume* (1.2)–(1.5), (1.8) *and* (M1)–(M3). *Suppose* (3.5). *Then for any* $\theta > 1/q$,

$$\left\|\hat{\mathbf{T}}_n(K_n) - \mathbf{T}_n(K_n)\right\|_2^2 = \mathrm{O}_p\left(\frac{K_n^{1+\theta}}{n^{1-2d}} + \frac{K_n^{2+\theta}}{n}\right). \tag{3.12}$$

Proof. Let $\mathbf{S}_n = (s_{ij})_{1\le i,j\le n} = \hat{\mathbf{T}}_n(K_n) - \mathbf{T}_n(K_n)$. Then

$$\max_{1\le i\le n} \sum_{t=1}^{n} s_{it}^2 \le \max_{1\le k\le K_n} \left\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\right\|_2^2,$$

and for each $1 \le j \le n$, $\sharp B_j \le 2K_n - 1$, where $B_j = \{i\colon \sum_{t=1}^{n} s_{it}s_{jt} \ne 0\}$. These and some algebraic manipulations yield

$$\begin{aligned}
\left\|\mathbf{S}_n\mathbf{S}_n'\right\|_1 &= \max_{1\le j\le n} \sum_{i\in B_j} \left|\sum_{t=1}^{n} s_{it}s_{jt}\right| \\
&\le \max_{1\le j\le n} \sum_{i\in B_j} \left(\sum_{t=1}^{n} s_{it}^2\right)^{1/2} \left(\sum_{h=1}^{n} s_{jh}^2\right)^{1/2} \le CK_n \max_{1\le k\le K_n} \left\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\right\|_2^2 \\
&\le CK_n \left\|\hat{\boldsymbol{\Gamma}}_{K_n,n}^{-1}\right\|_2^2 \max_{1\le k\le K_n} \left\|\frac{1}{n-k} \sum_{t=k}^{n-1} \mathbf{U}_t(k) w_{k,t+1}\right\|_2^2.
\end{aligned}$$

Now, the desired conclusion (3.12) follows from (3.3), (3.6) and $\|\mathbf{S}_n\|_2^2 \le \|\mathbf{S}_n\mathbf{S}_n'\|_1$. □

Proposition 3.2. *Assume* (2.1), (2.2), *and the same assumptions as in Proposition* 3.1. *Suppose* (3.5). *Then for any* $\theta > 1/q$,

$$\left\|\hat{\mathbf{D}}_n^{-1}(K_n) - \mathbf{D}_n^{-1}(K_n)\right\|_2 = \begin{cases} \mathrm{O}_p\left(n^{-1/2}K_n^{\theta}\right), & \textit{if } 0 < d < 1/4, \\ \mathrm{O}_p\left((\log n/n)^{1/2}K_n^{\theta}\right), & \textit{if } d = 1/4, \\ \mathrm{O}_p\left(n^{-1+2d}K_n^{\theta}\right), & \textit{if } 1/4 < d < 1/2. \end{cases} \tag{3.13}$$

Proof. Note first that

$$\left\|\hat{\mathbf{D}}_n(K_n) - \mathbf{D}_n(K_n)\right\|_2 = \max_{0\le k\le K_n} \left|\hat{\sigma}_k^2 - \sigma_k^2\right|, \tag{3.14}$$

recalling $\sigma_0^2 = \gamma_0$. By (1.6) and an argument similar to that used to prove (3.7), it holds that

$$E\left(\hat{\sigma}_0^2 - \sigma_0^2\right)^2 = \begin{cases} \mathrm{O}\left(n^{-1}\right), & \text{if } 0 < d < 1/4, \\ \mathrm{O}(\log n/n), & \text{if } d = 1/4, \\ \mathrm{O}\left(n^{-2+4d}\right), & \text{if } 1/4 < d < 1/2. \end{cases} \tag{3.15}$$

Straightforward calculations show

$$\max_{1\le k\le K_n}\left|\hat{\sigma}_k^2-\sigma_k^2\right| \le \max_{1\le k\le K_n}\left|\frac{1}{n-k}\sum_{t=k}^{n-1} w_{k,t+1}^2-\sigma_k^2\right| + C\left\|\hat{\boldsymbol{\Gamma}}_{K_n,n}^{-1}\right\|_2 \max_{1\le k\le K_n}\left\|\frac{1}{n-k}\sum_{t=k}^{n-1}\mathbf{U}_t(k)w_{k,t+1}\right\|_2^2,$$

which, in conjunction with (3.15), (3.14), (3.3), (3.6) and (3.9), results in (3.13). □

The main results of this section is given as follows.

Theorem 3.1. *Assume the same assumptions as in Proposition 3.2. Suppose*

$$\frac{\log n\log K_n}{K_n^d}+\frac{K_n^{1+\theta}\log K_n}{n^{1-2d}}+\frac{K_n^{2+\theta}\log K_n}{n}=\mathrm{o}(1), \tag{3.16}$$

for some $\theta>1/q$. Then

$$\begin{aligned}\left\|\hat{\boldsymbol{\Omega}}_n^{-1}(K_n)-\boldsymbol{\Omega}_n^{-1}\right\|_2 &= \mathrm{O}_p\left(\left(\frac{\log n\log K_n}{K_n^d}\right)^{1/2}+\left(\frac{K_n^{1+\theta}\log K_n}{n^{1-2d}}+\frac{K_n^{2+\theta}\log K_n}{n}\right)^{1/2}\right)\\ &= \mathrm{o}_p(1)\end{aligned} \tag{3.17}$$

and

$$\left\|\hat{\boldsymbol{\Omega}}_n^{-1}(K_n)\right\|_2=\mathrm{O}_p(1). \tag{3.18}$$

Proof. By Propositions 3.1 and 3.2, (1.8), (ii) of Proposition 2.1, (ii) of Proposition 2.2 and

$$\begin{aligned}&\left\|\hat{\boldsymbol{\Omega}}_n^{-1}(K_n)-\boldsymbol{\Omega}_n^{-1}(K_n)\right\|_2\\ &\quad\le \left\|\hat{\mathbf{T}}_n(K_n)-\mathbf{T}_n(K_n)\right\|_2\left\|\hat{\mathbf{D}}_n^{-1}(K_n)\right\|_2\left(\left\|\hat{\mathbf{T}}_n(K_n)-\mathbf{T}_n(K_n)\right\|_2+\left\|\mathbf{T}_n(K_n)\right\|_2\right)\\ &\qquad+\left\|\mathbf{T}_n(K_n)\right\|_2\left\|\hat{\mathbf{D}}_n^{-1}(K_n)-\mathbf{D}_n^{-1}(K_n)\right\|_2\left(\left\|\hat{\mathbf{T}}_n(K_n)-\mathbf{T}_n(K_n)\right\|_2+\left\|\mathbf{T}_n(K_n)\right\|_2\right)\\ &\qquad+\left\|\mathbf{T}_n(K_n)\right\|_2\left\|\mathbf{D}_n^{-1}(K_n)\right\|_2\left\|\hat{\mathbf{T}}_n(K_n)-\mathbf{T}_n(K_n)\right\|_2,\end{aligned}$$

one obtains

$$\left\|\hat{\boldsymbol{\Omega}}_n^{-1}(K_n)-\boldsymbol{\Omega}_n^{-1}(K_n)\right\|_2=\mathrm{O}_p\left(\left(\frac{K_n^{1+\theta}\log K_n}{n^{1-2d}}+\frac{K_n^{2+\theta}\log K_n}{n}\right)^{1/2}\right).$$

This, together with (3.16) and Theorem 2.2, leads to the desired conclusions (3.17) and (3.18). □

Remark 3.2. It would be interesting to compare Theorem 3.1 with the moment bounds for $\hat{\boldsymbol{\Gamma}}^{-1}_{K_n,n}$ given by Godet [8]. If $\{u_t\}$ is a Gaussian process satisfying (1.2)–(1.5) and (1.8), then Theorem 2.1 of Godet [8] yields that for

$$K_n = \mathrm{O}(n^{\lambda}) \qquad \text{with } 0 < \lambda < \min\{1/2, 1-2d\}, \tag{3.19}$$

$$E\|\hat{\boldsymbol{\Gamma}}^{-1}_{K_n,n} - \boldsymbol{\Omega}^{-1}_{K_n}\|_2 = \begin{cases} \mathrm{O}(n^{-1/2}K_n), & \text{if } 0 < d < 1/4, \\ \mathrm{O}((\log n/n)^{1/2}K_n), & \text{if } d = 1/4, \\ \mathrm{O}(n^{-1+2d}K_n), & \text{if } 1/4 < d < 1/2. \end{cases} \tag{3.20}$$

One major difference between $\hat{\boldsymbol{\Omega}}^{-1}_n(K_n)$ and $\hat{\boldsymbol{\Gamma}}^{-1}_{K_n,n}$ is that the former aims at estimating the inverse autocovariance matrix of all n observations, $\boldsymbol{\Omega}^{-1}_n$, but the latter only focuses on that of K_n consecutive observations, $\boldsymbol{\Omega}^{-1}_{K_n}$, with $K_n \ll n$. While (3.20) plays an important role in analyzing the mean squared prediction error of the least squares predictor of u_{n+1} based on the AR(K_n) model, $\hat{\boldsymbol{\Gamma}}^{-1}_{K_n,n}$ cannot be used in situations where consistent estimates of $\boldsymbol{\Omega}^{-1}_n$ are indispensable. See Section 4.2 for some examples. Moreover, the convergence rate of $\hat{\boldsymbol{\Omega}}^{-1}_n(K_n)$ is determined by not only the estimation error $\|\hat{\boldsymbol{\Omega}}^{-1}_n(K_n) - \boldsymbol{\Omega}^{-1}_n(K_n)\|_2$, but also the approximation error $\|\boldsymbol{\Omega}^{-1}_n(K_n) - \boldsymbol{\Omega}^{-1}_n\|_2$. This latter type of error, however, is irrelevant to the convergence rate of $\hat{\boldsymbol{\Gamma}}^{-1}_{K_n,n}$. Finally, we note that (3.20) gives a stronger mode of convergence than (3.17), but at the expense of more stringent assumptions on moments and distributions.

4. Some extensions

Consider a linear regression model with serially correlated errors,

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t = \sum_{i=1}^{p} x_{ti}\beta_i + u_t, \tag{4.1}$$

where $\boldsymbol{\beta}$ is an unknown coefficient vector, $\mathbf{x}_t$'s are p-dimensional nonrandom input vectors and u_t's are unobservable random disturbances satisfying the long-memory conditions described previously. Having observed $\mathbf{y}_n = (y_1, \ldots, y_n)'$ and $\check{\mathbf{x}}_{nj} = (x_{1j}, \ldots, x_{nj})'$, $1 \le j \le p$, it is natural to estimate $\mathbf{u}_n = (u_1, \ldots, u_n)'$ via the least squares residuals

$$\tilde{\mathbf{u}}_n = (\tilde{u}_1, \ldots, \tilde{u}_n)' = (\mathbf{I}_n - M_{np})\mathbf{y}_n = (\mathbf{I}_n - M_{np})\mathbf{u}_n,$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and M_{np} is the orthogonal projection matrix of $\overline{\mathrm{sp}}\{\check{\mathbf{x}}_{n1}, \ldots, \check{\mathbf{x}}_{np}\}$, the closed span of $\{\check{\mathbf{x}}_{n1}, \ldots, \check{\mathbf{x}}_{np}\}$. Note that $\tilde{\mathbf{u}}_n$ is also known as a detrended time series, in particular when $\mathbf{x}_t$ represents the trend or seasonal component of y_t. Let $\{\check{\mathbf{q}}_{ni} = (\mathsf{q}_{1i}, \ldots, \mathsf{q}_{ni})', i = 1, \ldots, r\}$, $1 \le r \le p$, be an orthonormal basis of $\overline{\mathrm{sp}}\{\check{\mathbf{x}}_{n1}, \ldots, \check{\mathbf{x}}_{np}\}$. It is well known that $M_{np} = \sum_{i=1}^{r} \check{\mathbf{q}}_{ni}\check{\mathbf{q}}_{ni}'$, and hence with $v_i = \check{\mathbf{q}}_{ni}'\mathbf{u}_n$,

$$\tilde{\mathbf{u}}_n = \mathbf{u}_n - \sum_{i=1}^{r} v_i \check{\mathbf{q}}_{ni}. \tag{4.2}$$

In Section 4.1, we shall show that the inverse autocovariance matrix, $\mathbf{\Omega}_n^{-1}$, of $\mathbf{u}_n$ can still be consistently estimated by the modified Cholesky decomposition method proposed in Section 3 with $\mathbf{u}_n$ replaced by $\tilde{\mathbf{u}}_n$, which is denoted by $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$. We also show that $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ and $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$ share the same rate of convergence. Moreover, we propose an estimate of $\mathbf{a}(n) = (a_{n,1}, \dots, a_{n,n})'$, the n-dimensional finite predictor coefficient vector of $\{u_t\}$, based on $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$, and derive its convergence rate. These asymptotic results are obtained without imposing *any* assumptions on the design matrix $\mathbf{X}_n = (\check{\mathbf{x}}_{n1}, \dots, \check{\mathbf{x}}_{np})$. On the other hand, we assume that $\mathbf{X}_n$ has a full rank in Section 4.2, and propose an FGLSE of $\boldsymbol{\beta}$ based on $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$. The rate of convergence of the proposed FGLSE is also established in Section 4.2.

4.1. Consistent estimates of $\mathbf{\Omega}_n^{-1}$ and $\mathbf{a}(n)$ based on $\tilde{\mathbf{u}}_n$

Define

$$\tilde{\mathbf{\Omega}}_n^{-1}(K_n) := \tilde{\mathbf{T}}_n(K_n)'\tilde{\mathbf{D}}_n^{-1}(K_n)\tilde{\mathbf{T}}_n(K_n)$$

where $\tilde{\mathbf{T}}_n(K_n)$ and $\tilde{\mathbf{D}}_n(K_n)$ are $\hat{\mathbf{T}}_n(K_n)$ and $\hat{\mathbf{D}}_n(K_n)$ with $\hat{a}_{ij}$ and $\hat{\sigma}_i^2$, respectively, replaced by $\tilde{a}_{ij}$ and $\tilde{\sigma}_i^2$ defined as follows:

$$(\tilde{a}_{k,1}, \dots, \tilde{a}_{k,k}) = \operatorname*{arg\,min}_{(\alpha_1,\dots,\alpha_k)\in R^k} \sum_{t=k+1}^{n} (\tilde{u}_t - \alpha_1\tilde{u}_{t-1} - \alpha_2\tilde{u}_{t-2} - \cdots - \alpha_k\tilde{u}_{t-k})^2,$$

$$\tilde{\sigma}_0^2 = (n-1)^{-1}\sum_{t=1}^{n}(\tilde{u}_t - \bar{\tilde{u}})^2, \qquad \bar{\tilde{u}} = n^{-1}\sum_{t=1}^{n}\tilde{u}_i,$$

$$\tilde{\sigma}_k^2 = (n-k)^{-1}\sum_{t=k+1}^{n}\left(\tilde{u}_t - \sum_{j=1}^{k}\tilde{a}_{k,j}\tilde{u}_{t-j}\right)^2.$$

By establishing probability bounds for $\|\tilde{\mathbf{T}}_n(K_n) - \mathbf{T}_n(K_n)\|_2$ and $\|\tilde{\mathbf{D}}_n^{-1}(K_n) - \mathbf{D}_n^{-1}(K_n)\|_2$ in Proposition 4.1, we obtain the convergence rate of $\|\tilde{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2$ in Theorem 4.1. According to (4.2), $\mathbf{u}_n$ and $\tilde{\mathbf{u}}_n$ differ by the vector $\sum_{i=1}^{r} v_i\check{\mathbf{q}}_{ni}$, whose entries are weighted sums of $u_1, u_2, \dots, u_n$ with weights $\mathsf{q}_{t_1,i}\mathsf{q}_{t_2,j}$ for some $1 \le t_1, t_2 \le n$ and $1 \le i, j \le r$. To explore the contributions of $\sum_{i=1}^{r} v_i\check{\mathbf{q}}_{ni}$ to $\|\tilde{\mathbf{T}}_n(k) - \mathbf{T}_n(k)\|_2$ and $\|\tilde{\mathbf{D}}_n^{-1}(k) - \mathbf{D}_n^{-1}(k)\|_2$, we need moment bounds for the linear combinations of u_i's and $\tau_{k,i}$'s, which are introduced in the following lemma.

Lemma 4.1. *Let $c_1, \dots, c_m$ be any real numbers. Under the same assumptions as in Lemma 3.1,*

$$E\left(\sum_{i=1}^{m} c_i u_i\right)^{4q} \le C\left(\sum_{i=1}^{m} c_i^2\right)^{2q} m^{4qd}. \tag{4.3}$$

Moreover, if (2.1) *and* (2.2) *also hold true, then*

$$E\left(\sum_{i=1}^{m} c_i \tau_{k,i}\right)^{4q} \le C\left(\sum_{i=1}^{m} c_i^2\right)^{2q} m^{4qd}. \tag{4.4}$$

Proof. By Lemma 2 of Wei [16], we have $E(\sum_{i=1}^m c_i u_i)^{4q} \le C\{E(\sum_{i=1}^m c_i u_i)^2\}^{2q}$. Theorem 2.1 of Ing and Wei [10] and Jensen's inequality further yield $E(\sum_{i=1}^m c_i u_i)^2 \le C(\sum_{i=1}^m |c_i|^{2/(1+2d)})^{1+2d} \le C m^{2d}(\sum_{i=1}^m c_i^2)$. Hence, (4.3) follows. Equation (4.4) is ensured by Lemma 3.3 and an argument similar to that used to prove (4.3). □

Equipped with Lemma 4.1, we can prove another auxiliary lemma, which plays a key role in establishing Proposition 4.1. First, some notation: $\tilde{w}_{k,t+1} = \tilde{u}_{t+1} - \mathbf{a}(k)'\tilde{\mathbf{U}}_t(k)$, $\tilde{\mathbf{U}}_t(k) = (\tilde{u}_t, \tilde{u}_{t-1}, \ldots, \tilde{u}_{t-k+1})'$, $\tilde{\boldsymbol{\Gamma}}_{k,n} = \frac{1}{n-k}\sum_{t=k}^{n-1} \tilde{\mathbf{U}}_t(k)\tilde{\mathbf{U}}_t(k)'$, $\mathbf{q}_t = (\mathsf{q}_{t,1}, \mathsf{q}_{t,2}, \ldots, \mathsf{q}_{t,r})'$, $\mathbf{Q}_t(k) = (\mathbf{q}_t, \mathbf{q}_{t-1}, \ldots, \mathbf{q}_{t-k+1})'$ and $\mathbf{V}_n = (v_1, \ldots, v_r)'$.

Lemma 4.2.

(i) *Assume that the same assumptions as in Lemma* 3.4 *hold. Then for* $K_n = \mathrm{o}(n)$ *and* $\theta > 1/q$,

$$\max_{1\le k\le K_n}\left\|\frac{1}{n-k}\sum_{t=k}^{n-1}\tilde{\mathbf{U}}_t(k)\tilde{w}_{k,t+1}\right\|_2^2 = \mathrm{O}_p\left(\frac{K_n^{\theta}}{n^{1-2d}} + \frac{K_n^{1+\theta}}{n}\right).$$

(ii) *Assume that the same assumptions as in Lemma* 3.2 *hold. Then for* K_n *satisfying* (3.5), $\|\tilde{\boldsymbol{\Gamma}}_{K_n,n}^{-1}\|_2 = \mathrm{O}_p(1)$.

(iii) *Assume that the same assumptions as in Lemma* 3.4 *hold. Then for* $K_n = \mathrm{o}(n)$ *and* $\theta > 1/(2q)$,

$$\max_{1\le k\le K_n}\left|\frac{1}{n-k}\sum_{t=k}^{n-1}\tilde{w}_{k,t+1}^2 - \sigma_k^2\right| = \begin{cases} \mathrm{O}_p\bigl(K_n^{\theta} n^{-1/2}\bigr), & \text{if } 0 < d < 1/4;\\ \mathrm{O}_p\bigl(K_n^{\theta}(\log n)^{1/2} n^{-1/2}\bigr), & \text{if } d = 1/4;\\ \mathrm{O}_p\bigl(K_n^{\theta} n^{-1+2d}\bigr), & \text{if } 1/4 < d < 1/2.\end{cases}$$

Proof. We begin by proving (i). Define $(B1) = \|\frac{1}{n-k}\sum_{t=k}^{n-1}\tilde{\mathbf{U}}_t(k)(\tilde{w}_{k,t+1} - w_{t+1})\|_2^{2q}$ and $(B2) = \|\frac{1}{n-k}\sum_{t=k}^{n-1}\tilde{\mathbf{U}}_t(k) w_{t+1}\|_2^{2q}$. Straightforward calculations yield

$$\left\|\frac{1}{n-k}\sum_{t=k}^{n-1}\tilde{\mathbf{U}}_t(k)\tilde{w}_{k,t+1}\right\|_2^{2q} \le C\bigl\{(B1) + (B2)\bigr\}, \tag{4.5}$$

$$E(B1) \le C n^{-2q} k^{q-1}\sum_{j=1}^{k}\bigl\{E\bigl|(B3)\bigr|^{2q} + E\bigl|(B4)\bigr|^{2q} + E\bigl|(B5)\bigr|^{2q} + E\bigl|(B6)\bigr|^{2q}\bigr\} \tag{4.6}$$

and

$$E(B2) \leq C\left\{E\left\|\frac{1}{n-k}\sum_{t=k}^{n-1}\mathbf{U}_t(k)w_{t+1}\right\|_2^{2q} + E(B7)\right\}, \tag{4.7}$$

where

$$(B3) = \sum_{t=k}^{n-1} u_{t+1-j}\tau_{k,t}, \qquad (B4) = \mathbf{V}_n' \sum_{t=k}^{n-1} \mathbf{q}_{t+1-j}\tau_{k,t},$$

$$(B5) = \mathbf{V}_n' \sum_{t=k}^{n-1}\big(\mathbf{q}_{t+1} - \mathbf{Q}_t'(k)\mathbf{a}(k)\big)u_{t+1-j},$$

$$(B6) = \mathbf{V}_n' \left\{\sum_{t=k}^{n-1} \mathbf{q}_{t+1-j}\big(\mathbf{q}_{t+1}' - \mathbf{a}(k)'\mathbf{Q}_t(k)\big)\right\}\mathbf{V}_n,$$

$$(B7) = \left\|\left\{\frac{1}{n-k}\sum_{t=k}^{n-1}\mathbf{Q}_t(k)w_{t+1}\right\}\mathbf{V}_n\right\|_2^{2q}.$$

An argument similar to that used to prove (3.1) implies $E|(B3)|^{2q} = \mathrm{O}(k^{-q}(n-k)^{q+2qd})$. In addition, by (4.3), (4.4), (2.3) and $\sum_{t=1}^{n} \mathsf{q}_{t,i}^2 = 1$ for $i = 1, 2, \dots, r$, one obtains

$$E\big|(B4)\big|^{2q} \leq r^{2q-1}\sum_{i=1}^{r}\big[E\big(v_i^{4q}\big)\big]^{1/2}\left[E\left(\sum_{t=k}^{n-1}\mathsf{q}_{t+1-j,i}\tau_{k,t}\right)^{4q}\right]^{1/2} \leq Cn^{4qd},$$

$E|(B5)|^{2q} \leq Cn^{4qd}$, $E|(B6)|^{2q} \leq Cn^{4qd}$, and $E(B7) \leq Ck^q n^{2qd}/n^{2q}$. With the help of these moment inequalities, (3.4) and (4.5)–(4.7), the proof of (i) can be completed in the same way as the proof of (3.3). Moreover, by modifying the proofs of (3.6) and (3.9) accordingly, we can establish (ii) and (iii). The details, however, are not presented here. □

Proposition 4.1. *Assume the same assumptions as in Proposition* 3.2. *Suppose* (3.5). *Then for any* $\theta > 1/q$,

(i) $\|\tilde{\mathbf{T}}_n(K_n) - \mathbf{T}_n(K_n)\|_2^2 = \mathrm{O}_p((K_n^{1+\theta}/n^{1-2d}) + (K_n^{2+\theta}/n))$.

(ii)

$$\big\|\tilde{\mathbf{D}}_n^{-1}(K_n) - \mathbf{D}_n^{-1}(K_n)\big\|_2 = \begin{cases} \mathrm{O}_p\big(n^{-1/2}K_n^{\theta}\big), & \textit{if } 0 < d < 1/4; \\ \mathrm{O}_p\big((\log n/n)^{1/2}K_n^{\theta}\big), & \textit{if } d = 1/4; \\ \mathrm{O}_p\big(n^{-1+2d}K_n^{\theta}\big), & \textit{if } 1/4 < d < 1/2. \end{cases}$$

Proof. In view of the proof of Proposition 3.1, (i) follows directly from (i) and (ii) of Lemma 4.1. To show (ii), note first that (3.15) still holds with $\hat{\sigma}_0^2$ replaced by $\tilde{\sigma}_0^2$. This, in conjunction with (i)–(iii) of Lemma 4.1 and the argument used in the proof of Proposition 3.2, yields (ii). □

We are now in a position to introduce Theorem 4.1.

Theorem 4.1. *Consider the regression model* (4.1). *With the same assumptions as in Proposition* 3.2, *suppose that* (3.16) *holds for some* $\theta > 1/q$. *Then*

$$\begin{aligned} &\|\tilde{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2 \\ &\quad = \mathrm{O}_p\left(\left(\frac{\log n \log K_n}{K_n^d}\right)^{1/2} + \left(\frac{K_n^{1+\theta}\log K_n}{n^{1-2d}} + \frac{K_n^{2+\theta}\log K_n}{n}\right)^{1/2}\right) = \mathrm{o}_p(1) \end{aligned} \tag{4.8}$$

and

$$\|\tilde{\mathbf{\Omega}}_n^{-1}(K_n)\|_2 = \mathrm{O}_p(1). \tag{4.9}$$

Proof. In view of the proof of Theorem 3.1, (4.8) and (4.9) are immediate consequences of Proposition 4.1 and Theorem 2.2. □

Remark 4.1. Since no assumptions are imposed on the design matrix $\mathbf{X}_n$, one of the most intriguing implications of Theorem 4.1 is that $\mathbf{\Omega}_n^{-1}$ can be consistently estimated by $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ even when $\mathbf{X}_n$ is singular. Moreover, according to (4.8) and (3.17), it is interesting to point out that $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ and $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$ share the same rate of convergence.

Next, we consider the problem of estimating $\mathbf{a}(n)$ under model (4.1). Recall Yule–Walker equations $\mathbf{a}(n) = \mathbf{\Omega}_n^{-1}\boldsymbol{\gamma}_n$, where $\boldsymbol{\gamma}_n = (\gamma_1, \dots, \gamma_n)'$. A truncated version of $\mathbf{\Omega}_n^{-1}\boldsymbol{\gamma}_n$ is given by $\check{\mathbf{a}}(n) = \mathbf{\Omega}_n^{-1}(K_n)\check{\boldsymbol{\gamma}}_n$, where $\check{\boldsymbol{\gamma}}_n = (\gamma_1, \dots, \gamma_{K_n}, 0, \dots, 0)'$ is an n-dimensional vector. A natural estimate of $\check{\mathbf{a}}(n)$ is $\mathbf{a}^*(n) = \tilde{\mathbf{\Omega}}_n^{-1}(K_n)\tilde{\boldsymbol{\gamma}}_n$, where $\tilde{\boldsymbol{\gamma}}_n = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_{K_n}, 0, \dots, 0)'$ is an n-dimensional vector with $\tilde{\gamma}_j$ denoting the $(1, j+1)$th entry of $\tilde{\mathbf{\Gamma}}_{K_n+1,n}$. We shall show that when K_n is suitably chosen, $\mathbf{a}^*(n)$ is a consistent estimate of $\mathbf{a}(n)$.

Corollary 4.1. *Assume the same assumptions as in Theorem* 4.1. *Suppose that* (3.16) *holds and* $\frac{K_n^{1+2d/3}}{n^{1-2d}} = \mathrm{o}(1)$. *Then for any* $\theta > 1/q$,

$$\begin{aligned} \|\mathbf{a}^*(n) - \mathbf{a}(n)\|_2 &= \begin{cases} \mathrm{O}_p\left(\left(\frac{1}{K_n^{1-2d}} + \frac{\log n}{K_n^d} + \frac{K_n^{1+\theta}}{n^{1-2d}} + \frac{K_n^{2+\theta}}{n}\right)^{1/2}\right), & \textit{if } 0 < d \le 1/4, \\ \mathrm{O}_p\left(\left(\frac{1}{K_n^{1-2d}} + \frac{\log n}{K_n^d} + \frac{K_n^{1+\theta}}{n^{1-2d}} + \frac{K_n^{3+2d}}{n^{3-6d}}\right)^{1/2}\right), & \textit{if } 1/4 < d < 1/2 \end{cases} \\ &= \mathrm{o}_p(1). \end{aligned}$$

Proof. Note first that

$$\|\mathbf{a}^*(n) - \mathbf{a}(n)\|_2 \le \|\check{\mathbf{a}}(n) - \mathbf{a}(n)\|_2 + \|\mathbf{a}^*(n) - \check{\mathbf{a}}(n)\|_2, \tag{4.10}$$

$$\|\check{\mathbf{a}}(n) - \mathbf{a}(n)\|_2 \le \|\mathbf{\Omega}_n^{-1}(\check{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_n)\|_2 + \|(\mathbf{\Omega}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1})\check{\boldsymbol{\gamma}}_n\|_2, \tag{4.11}$$

$$\|\mathbf{a}^*(n) - \check{\mathbf{a}}(n)\|_2 \le \|\tilde{\mathbf{\Omega}}_n^{-1}(K_n)(\tilde{\boldsymbol{\gamma}}_n - \check{\boldsymbol{\gamma}}_n)\|_2 + \|(\tilde{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}(K_n))\check{\boldsymbol{\gamma}}_n\|_2. \tag{4.12}$$

Moreover,

$$\left\|\boldsymbol{\Omega}_n^{-1}(\check{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_n)\right\|_2 \le \left\|\mathbf{T}_n'\mathbf{D}_n^{-1}\right\|_2\left\|\mathbf{T}_n(\check{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_n)\right\|_2 \le C\left\|\mathbf{T}_n(\check{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_n)\right\|_2 \tag{4.13}$$

and

$$\begin{aligned}\boldsymbol{\Omega}_n^{-1}-\boldsymbol{\Omega}_n^{-1}(K_n) = {}&\big(\mathbf{T}_n-\mathbf{T}_n(K_n)\big)'\mathbf{D}_n^{-1}\big(\big(\mathbf{T}_n-\mathbf{T}_n(K_n)\big)+\mathbf{T}_n(K_n)\big)\\ &+\mathbf{T}_n'(K_n)\big(\mathbf{D}_n^{-1}-\mathbf{D}_n^{-1}(K_n)\big)\big(\big(\mathbf{T}_n-\mathbf{T}_n(K_n)\big)+\mathbf{T}_n(K_n)\big)\\ &+\mathbf{T}_n'(K_n)\mathbf{D}_n^{-1}(K_n)\big(\mathbf{T}_n-\mathbf{T}_n(K_n)\big).\end{aligned} \tag{4.14}$$

By (1.5), (1.6), (2.1), (2.2), it follows that $\|\mathbf{T}_n(\check{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_n)\|_2 = \mathrm{O}(K_n^{-1/2+d})$, $\|\mathbf{T}_n(K_n)\check{\boldsymbol{\gamma}}_n\|_2 = \mathrm{O}(1)$, and $\|(\mathbf{T}_n-\mathbf{T}_n(K_n))\check{\boldsymbol{\gamma}}_n\|_2 = \mathrm{O}(K_n^{-1/2+d})$. These bounds, together with (4.11), (4.13) and (4.14), yield

$$\left\|\check{\mathbf{a}}(n)-\mathbf{a}(n)\right\|_2 = \mathrm{O}\big(K_n^{-1/2+d}+\big(K_n^{-d}\log n\big)^{1/2}\big). \tag{4.15}$$

By the first moment bound theorem of Findley and Wei [7], Cauchy–Schwarz inequality, Proposition 4.1, (4.7), Lemma 4.2(ii), (4.9) and (4.12), it can be shown that

$$\left\|\mathbf{a}^*(n)-\check{\mathbf{a}}(n)\right\|_2 = \begin{cases}\mathrm{O}_p\big(\big(n^{-1+2d}K_n^{1+\theta}+n^{-1}K_n^{2+\theta}\big)^{1/2}\big), & \text{if } 0<d\le 1/4,\\ \mathrm{O}_p\big(\big(n^{-1+2d}K_n^{1+\theta}+n^{-3+6d}K_n^{3+2d}\big)^{1/2}\big), & \text{if } 1/4<d<1/2,\end{cases} \tag{4.16}$$

for any $\theta > 1/q$. Now, the desired conclusion follows from (4.10), (4.15) and (4.16). □

Remark 4.2. When $u_1,\dots,u_n$ are observable, Wu and Pourahmadi [19] constructed an estimate, $\check{\boldsymbol{\Omega}}_{n,K_n}^{-1}\check{\gamma}_n$, of $\mathbf{a}(n)$, where $\check{\boldsymbol{\Omega}}_{n,K_n} = (\hat{\gamma}_{i-j}\mathbf{1}_{|i-j|\le K_n})_{1\le i,j\le n}$ and $\check{\boldsymbol{\gamma}}_n = (\check{\gamma}_1,\dots,\check{\gamma}_n)'$ with $\check{\gamma}_i = \hat{\gamma}_i\mathbf{1}_{\{i\le K_n\}}$. By assuming $\sum_{j=1}^{\infty}|\gamma_j|<\infty$, they obtained a convergence rate of the proposed estimate in terms of K_n, the moment restriction on w_t, and $\sum_{j=K_n}^{\infty}|\gamma_k|$; see Corollary 2 of Wu and Pourahmadi [19]. However, their proof, relying heavily on $\sum_{j=1}^{\infty}|\gamma_j|<\infty$, is no longer applicable here.

4.2. The rate of convergence of the FGLSE

In this section, we assume that $\mathbf{X}_n$ is nonsingular, and hence $\boldsymbol{\beta}$ is uniquely defined. We estimate $\boldsymbol{\beta}$ using the FGLSE,

$$\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}} = \big(\mathbf{X}_n'\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n)\mathbf{X}_n\big)^{-1}\mathbf{X}_n'\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n)\mathbf{y}_n.$$

The main objective of this section is to investigate the convergence rate of $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$. To simplify the exposition, we shall focus on polynomial regression models and impose the following conditions on a_i:

$$a_j \sim C_0 j^{-1-d} \quad\text{and}\quad \sum_{j=0}^{\infty}a_j\mathrm{e}^{ij\lambda}=0 \qquad \text{if and only if } \lambda=0, \tag{4.17}$$

where $a_0 = -1$ and $C_0 \neq 0$. As mentioned in Section 2, (4.17) is fulfilled by the FARIMA model defined in (1.7). When K_n diverges to infinity at a suitable rate, we derive the rate of convergence of $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$ in the next corollary. It is important to be aware that our proof is not a direct application of Theorem 4.1. Instead, it relies on a very careful analysis of the joint effects between the Cholesky factors and the regressors.

Corollary 4.2. *Consider the regression model* (4.1) *with* $x_{ti} = t^{i-1}$ *for* $i = 1, \ldots, p$. *Assume the same assumptions as in Theorem* 4.1 *with* (1.5) *replaced by* (4.17). *Suppose that* (3.5) *holds and* $n^{-1+2d} K_n^{1+2d} + n^{-1} K_n^{2+2d} = \mathrm{o}(1)$. *Then*

(i) $\|\mathbf{L}_n(\boldsymbol{\beta}_{\mathrm{FGLS}} - \boldsymbol{\beta})\|_2 = \mathrm{O}_p(1)$,

(ii) $\|\mathbf{L}_n(\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}} - \boldsymbol{\beta})\|_2 = \mathrm{O}_p(1)$,

where $\mathbf{L}_n = n^{-d} \operatorname{diag}(n^{1/2}, n^{3/2}, \ldots, n^{p-1/2})$ *and* $\boldsymbol{\beta}_{\mathrm{FGLS}}$ *is* $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$ *with* $\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n)$ *replaced by* $\boldsymbol{\Omega}_n^{-1}(K_n)$.

Proof. We only prove Corollary 4.2 for $p = 2$. The proof for $p \neq 2$ is analogous. We begin by showing (i). Let $\tilde{\mathbf{L}}_n = K_n^{-d} \operatorname{diag}(n^{1/2}, n^{3/2})$. Then straightforward calculations yield

$$\begin{aligned} &\left\|\mathbf{L}_n(\boldsymbol{\beta}_{\mathrm{FGLS}} - \boldsymbol{\beta})\right\|_2 \\ &\quad \leq n^{-d} K_n^d \left\|\tilde{\mathbf{L}}_n \left(\mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{X}_n\right)^{-1} \tilde{\mathbf{L}}_n\right\|_2 \left\|\tilde{\mathbf{L}}_n^{-1} \mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{u}_n\right\|_2. \end{aligned} \tag{4.18}$$

Moreover, by (4.17),

$$n^{-d} K_n^d \left\|\tilde{\mathbf{L}}_n^{-1} \mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{u}_n\right\|_2 = \mathrm{O}_p(1), \tag{4.19}$$

$$\left\|\tilde{\mathbf{L}}_n \left(\mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{X}_n\right)^{-1} \tilde{\mathbf{L}}_n\right\|_2 \leq \left(n^{-1} \sum_{t=0}^{\kappa n} \lambda_{\min}(A_{\lfloor \kappa n \rfloor + t} + A_{n-t})\right)^{-1} = \mathrm{O}(1), \tag{4.20}$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of matrix A, $0 < \kappa < 1$ and $A_t = \eta_t' \eta_t$, with η_t denoting the tth row of $n^{1/2} \mathbf{D}_n^{-1/2}(K_n) \mathbf{T}_n(K_n) \mathbf{X}_n \tilde{\mathbf{L}}_n^{-1}$. Combining (4.18)–(4.20) yields (i). To show (ii), note first that

$$\left\|\mathbf{L}_n(\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}} - \boldsymbol{\beta})\right\|_2 \leq \left\|\mathbf{L}_n(\boldsymbol{\beta}_{\mathrm{FGLS}} - \boldsymbol{\beta})\right\|_2 + \left\|\mathbf{L}_n(\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}} - \boldsymbol{\beta}_{\mathrm{FGLS}})\right\|_2, \tag{4.21}$$

$$\left\|\mathbf{L}_n(\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}} - \boldsymbol{\beta}_{\mathrm{FGLS}})\right\|_2 \leq \left\|(D1)\right\|_2 + \left\|(D2)\right\|_2, \tag{4.22}$$

where $(D1) = \mathbf{L}_n((\mathbf{X}_n' \tilde{\boldsymbol{\Omega}}_n^{-1}(K_n) \mathbf{X}_n)^{-1} - (\mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{X}_n)^{-1}) \mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{u}_n$, and $(D2) = \mathbf{L}_n(\mathbf{X}_n' \tilde{\boldsymbol{\Omega}}_n^{-1}(K_n) \mathbf{X}_n)^{-1} \mathbf{X}_n' (\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n) - \boldsymbol{\Omega}_n^{-1}(K_n)) \mathbf{u}_n$. In addition,

$$\left\|(D5)\right\|_2 \leq \left(\left\|(D5)\right\|_2 + \left\|(D3)\right\|_2\right) \left\|(D4)\right\|_2 \left\|(D3)\right\|_2 \tag{4.23}$$

where $(D3) = \tilde{\mathbf{L}}_n(\mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{X}_n^{-1}) \tilde{\mathbf{L}}_n$, $(D4) = \tilde{\mathbf{L}}_n^{-1} \mathbf{X}_n' (\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n) - \boldsymbol{\Omega}_n^{-1}(K_n)) \mathbf{X}_n \tilde{\mathbf{L}}_n^{-1}$, and $(D5) = \tilde{\mathbf{L}}_n((\mathbf{X}_n' \tilde{\boldsymbol{\Omega}}_n^{-1}(K_n) \mathbf{X}_n^{-1}) - (\mathbf{X}_n' \boldsymbol{\Omega}_n^{-1}(K_n) \mathbf{X}_n^{-1})) \tilde{\mathbf{L}}_n$. By (4.17) and some algebraic manipulations, one obtains $\|(D3)\|_2 = \mathrm{O}(1)$ and $\|(D4)\|_2 = \mathrm{o}_p(1)$. Thus, by (4.23), $\|(D5)\|_2 = \mathrm{o}_p(1)$.

The bounds for $\|(D3)\|_2$ and $\|(D5)\|_2$, together with (4.17) and (4.19), imply

$$\|(D1)\|_2 \le n^{-d} K_n^d \|(D5)\|_2 \|\tilde{\mathbf{L}}_n^{-1} \mathbf{X}_n' \boldsymbol{\Omega}^{-1} n(K_n)\mathbf{u}_n\|_2 = \mathrm{o}_p(1), \tag{4.24}$$

$$\begin{aligned}\|(D2)\|_2 &\le n^{-d} K_n^d \big(\|(D5)\|_2 + \|(D3)\|_2\big) \|\tilde{\mathbf{L}}_n^{-1} \mathbf{X}_n' \big(\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n) - \boldsymbol{\Omega}_n^{-1}(K_n)\big)\mathbf{u}_n\|_2 \\ &= \mathrm{o}_p(1).\end{aligned} \tag{4.25}$$

Now, the desired conclusion (ii) follows from (4.24), (4.25), (4.21) and (4.22) and (i). □

Remark 4.3. Under assumptions similar to those of Corollary 4.2, Theorems 2.2 and 2.3 of Yajima [21] show that the best linear unbiased estimate (BLUE) $\hat{\boldsymbol{\beta}}_{\mathrm{BLUE}} = (\mathbf{X}_n' \boldsymbol{\Omega}_n^{-1} \mathbf{X}_n)^{-1} \mathbf{X}_n' \times \boldsymbol{\Omega}_n^{-1} \mathbf{y}_n$, and the LSE, $\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{X}_n' \mathbf{y}_n$, of $\boldsymbol{\beta}$ have the same rate of convergence, and this rate is, in turn, the same as that of $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$.

We close this section with a subtle example showing that the convergence rate of $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$ is faster than that of $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$, but slower than that of $\hat{\boldsymbol{\beta}}_{\mathrm{BLUE}}$. Consider model (4.1), with $p = 1$, $x_{t1} = 1 + \cos(\theta t)$, and $\theta \neq 0$. Assume the same assumptions as in Corollary 4.2. Then, by an argument similar to that used in the proof of Corollary 4.2, it can be shown that the rate of convergence of $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$ is $n^{-1/2+d} K_n^{-d}$. On the other hand, Theorems 2.1 and 2.2 and Example 2.1(ii) of Yajima [22] yield that the convergence rates of $\hat{\boldsymbol{\beta}}_{\mathrm{BLUE}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ are $n^{-1/2}$ and $n^{-1/2+d}$, respectively. This example gives a warning that the convergence rate of $\hat{\boldsymbol{\beta}}_{\mathrm{BLUE}}$ is not necessarily maintained by its feasible counterpart, even if the consistency of $\tilde{\boldsymbol{\Omega}}_n^{-1}(K_n)$ holds true.

5. Simulation study

In Section 5.1, we introduce a data-driven method for choosing the banding parameter K_n. With this K_n, we demonstrate the finite sample performance of the inverse autocovariance estimator proposed in Section 3 under FARIMA(p, d, q) processes, and that proposed in Section 4 under polynomial regression models with I(d) errors. The details are given in Sections 5.2 and 5.3, respectively.

5.1. Selection of K_n

Our approach for choosing K_n is based on the idea of subsampling and risk-minimization (SAR) introduced by Bickel and Levina [3] and Wu and Pourahmadi [19]. We first split the time series data $\{u_i\}_{i=1}^n$ into $\lfloor n/b \rfloor$ nonoverlapping subseries $\{u_j\}_{j=(v-1)b+1}^{vb}$ of equal length b, where b is a prescribed integer and $v = 1, 2, \dots, \lfloor n/b \rfloor$ with $\lfloor a \rfloor$ denoting the largest integer $\le a$. Let $1 \le L < H < b$ be another prescribed integers. For a given banding parameter $L \le k < H$, let $\hat{\boldsymbol{\Omega}}_{H,k,v}^{-1}$ represent our inverse autocovariance matrix estimator of $\boldsymbol{\Omega}_H{}^{-1}$ based on the vth subseries

$\{u_j\}_{j=(v-1)b+1}^{vb}$. Define the average risk

$$\hat{R}^{(O)}(k) = \frac{1}{\lfloor n/b \rfloor} \sum_{v=1}^{\lfloor n/b \rfloor} \|\hat{\mathbf{\Omega}}_{H,k,v}^{-1} - \mathbf{\Omega}_H^{-1}\|_2.$$

Our goal is to find a banding parameter such that $\hat{R}^{(O)}(k)$ is minimized. However, since $\mathbf{\Omega}_H^{-1}$ is unknown, we use $\hat{\mathbf{\Gamma}}_{H,n}^{-1}$, the H-dimensional inverse sample autocovariance matrix, as its surrogate, and replace $\hat{R}^{(O)}(k)$ by

$$\hat{R}(k) = \frac{1}{\lfloor n/b \rfloor} \sum_{v=1}^{\lfloor n/b \rfloor} \|\hat{\mathbf{\Omega}}_{H,k,v}^{-1} - \hat{\mathbf{\Gamma}}_{H,n}^{-1}\|_2,$$

noting that when $H \ll n$, $\hat{\mathbf{\Gamma}}_{H,n}^{-1}$ is a consistent estimator of $\mathbf{\Omega}_H^{-1}$. Now the banding parameter K_n is chosen to minimize $\hat{R}(k)$ over the interval $[L, H)$. In our simulation study, b is set to $\lfloor n/5 \rfloor$. In addition, inspired by Theorem 3.1, we choose $L = \lfloor \log n \rfloor$ and $H = \lceil n^{0.4} \rceil$, where $\lceil a \rceil$ denotes the smallest integer $\geq a$. The banding parameter for the detrended time series is also chosen in the same manner.

5.2. Finite sample performance of $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$

We explore the finite sample performance of $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$, with K_n determined by the SAR method, under the following four data generating processes (DGPs):

DGP 1: $(1-B)^d u_t = w_t$; DGP 2: $(1-0.7B)(1-B)^d u_t = w_t$;

DGP 3: $(1-B)^d u_t = (1-0.4B)w_t$; DGP 4: $(1+0.4B)(1-B)^d u_t = (1-0.3B)w_t$,

where the w_t's are i.i.d. $N(0,1)$ innovations. To improve the speed and accuracy, we adopt the method of Wu, Michailidis and Zhang [17] to generate the long memory data $\{u_1, \dots, u_n\}$. The performance of $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$ is evaluated by $\hat{l}_2(d)$, the average value of $\|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2$ over 1000 replications, with $n = 250, 500, 1000, 2000, 4000$. The results are summarized in Table 1. Note first that for each combination of d and DGP, $\hat{l}_2(d)$ shows an obvious downward trend as n increases. Moreover, when $n = 4000$, all $\hat{l}_2(d)$ are less than 0.65 except for $d = 0.1$ and DGP = DGP 3 or DGP 4. In the latter two cases, $\hat{l}_2(d)$, lying between 0.93 and 1.34, are still reasonably small. These findings suggest that $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$ is a reliable estimate of $\mathbf{\Omega}_n^{-1}$, particularly when n is large enough.

On the other hand, the decreasing rate of $\hat{l}_2(d)$ apparently changes over d and DGP. To provide a better understanding of this phenomenon, we first consider the fastest possible convergence rate that can be derived from Theorem 3.1:

$$\begin{aligned} &\|\hat{\mathbf{\Omega}}_n^{-1}(K_n^*) - \mathbf{\Omega}_n^{-1}\|_2 \\ &\quad = \begin{cases} \mathrm{O}_p\bigl(n^{-d/(4+2d+2\theta)}(\log n)^{(4+d+2\theta)/(4+2d+2\theta)}\bigr), & \text{if } 0 < d \leq \tilde{d}, \\ \mathrm{O}_p\bigl(n^{-d(1-2d)/(2+2d+2\theta)}(\log n)^{(2+d+2\theta)/(2+2d+2\theta)}\bigr), & \text{if } \tilde{d} < d < 1/2, \end{cases} \end{aligned} \tag{5.1}$$

Table 1. The values of $\hat{l}_2(d)$ under DGPs 1–4

$n \setminus d$	0.01	0.1	0.25	0.4	0.49	0.01	0.1	0.25	0.4	0.49
			DGP 1					DGP 2		
250	0.501	0.546	0.603	0.699	0.758	0.936	1.040	1.250	1.512	1.676
500	0.389	0.443	0.455	0.527	0.595	0.759	0.837	0.981	1.192	1.309
1000	0.276	0.366	0.335	0.396	0.444	0.537	0.595	0.734	0.867	0.977
2000	0.217	0.344	0.274	0.334	0.367	0.441	0.498	0.597	0.732	0.814
4000	0.173	0.344	0.216	0.257	0.298	0.345	0.389	0.481	0.573	0.647
			DGP 3					DGP 4		
250	0.767	1.007	0.775	0.642	0.660	1.141	1.495	1.129	0.836	0.839
500	0.642	0.952	0.652	0.514	0.529	0.923	1.373	0.942	0.725	0.688
1000	0.512	0.953	0.579	0.420	0.443	0.724	1.366	0.839	0.604	0.594
2000	0.435	0.928	0.495	0.358	0.376	0.625	1.339	0.714	0.518	0.497
4000	0.373	0.931	0.430	0.299	0.320	0.550	1.337	0.614	0.434	0.416

where $\tilde{d} = \{(3+2\theta)/2 + \theta^2/4\}^{1/2} - (1+\theta/2)$, and

$$K_n^* = \begin{cases} (n \log n)^{1/(2+d+\theta)}, & \text{if } 0 < d \le \tilde{d}, \\ (\log n)^{1/(1+d+\theta)} n^{(1-2d)/(1+\theta+d)}, & \text{if } \tilde{d} < d < 1/2. \end{cases}$$

Because w_t's are normally distributed, in view of Theorem 3.1, θ can be any positive number, and hence $\tilde{d}$ is arbitrarily close to $\sqrt{1.5} - 1$ (which, rounded to the nearest thousandth, is 0.225). We then measure the relative performance of $\|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2$ and $\|\hat{\mathbf{\Omega}}_n^{-1}(K_n^*) - \mathbf{\Omega}_n^{-1}\|_2$ using the ratio $\hat{l}_2(d)/\operatorname{OP}(d)$, where

$$\operatorname{OP}(d) = \begin{cases} 0.05 n^{-d/(4+2d)} (\log n)^{(4+d)/(4+2d)}, & \text{if } 0 < d \le 0.225, \\ 0.05 n^{-d(1-2d)/(2+2d)} (\log n)^{(2+d)/(2+2d)}, & \text{if } 0.225 < d < 1/2, \end{cases} \tag{5.2}$$

which is obtained from the bound in (5.1) with θ set to 0 and constants set to 0.05. The values of $\hat{l}_2(d)/\operatorname{OP}(d)$ under DGPs 1–4 are summarized in Table 2. Note that while the exact constants are not reported in (5.1), setting them to 0.05 helps us to better interpret some numerical results in Table 1 through Table 2.

For $n \ge 1000$, all values of $\hat{l}_2(d)/\operatorname{OP}(d)$ fall in a reasonable range of $(0.4, 5.0)$, suggesting that the rate of convergence of $\|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2$ is comparable to the optimal rate obtained from Theorem 3.1. Moreover, the asymptotic behaviors of $\hat{l}_2(d)$ can be well explained by $\operatorname{OP}(d)$ when DGP = DGP 1 and $d \ge 0.1$. In particular, when $n = 4000$, the rankings of $\{\hat{l}_2(0.1), \hat{l}_2(0.25), \hat{l}_2(0.4), \hat{l}_2(0.49)\}$ coincide exactly with those of $\{\operatorname{OP}(0.1), \operatorname{OP}(0.25), \operatorname{OP}(0.4), \operatorname{OP}(0.49)\}$, and $\operatorname{OP}(0.25) = \min_{d \in \{0.1, 0.25, 0.4, 0.49\}} \operatorname{OP}(d)$. This gives reasons for explaining why $d = 0.25$ often provides better results than $d = 0.1, 0.4$ or 0.49. The behavior of $\hat{l}_2(0.01)$, however, is apparently inconsistent with that of $\operatorname{OP}(0.01)$. Specifically, for $n \ge 250$, $\hat{l}_2(0.01) < \min_{d \in \{0.1, 0.25, 0.4, 0.49\}} \hat{l}_2(d)$, whereas $\operatorname{OP}(0.01) > \max_{d \in \{0.1, 0.25, 0.4, 0.49\}} \operatorname{OP}(d)$. One possible explanation of this discrepancy is that when d is extremely small, the constant associated with the

Table 2. The values of $\hat{l}_2(d)/\,\mathrm{OP}(d)$ under the DGPs 1–4

$n \setminus d$	0.01	0.1	0.25	0.4	0.49	0.01	0.1	0.25	0.4	0.49
			DGP 1					DGP 2		
250	1.849	2.349	3.414	3.782	3.704	3.453	4.476	7.081	8.187	8.187
500	1.278	1.728	2.399	2.629	2.640	2.493	3.264	5.172	5.950	5.804
1000	0.817	1.309	1.661	1.842	1.805	1.589	2.127	3.644	4.030	3.976
2000	0.585	1.138	1.291	1.460	1.380	1.188	1.647	2.814	3.198	3.064
4000	0.428	1.062	0.973	1.062	1.045	0.854	1.200	2.171	2.370	2.272
			DGP 3					DGP 4		
250	2.829	4.333	4.388	3.476	3.226	4.207	6.433	6.392	4.528	4.096
500	2.108	3.710	3.434	2.565	2.345	3.032	5.352	4.965	3.618	3.052
1000	1.514	3.405	2.873	1.951	1.802	2.143	4.880	4.162	2.807	2.418
2000	1.172	3.072	2.333	1.564	1.415	1.685	4.431	3.366	2.263	1.873
4000	0.922	2.875	1.939	1.237	1.122	1.361	4.130	2.772	1.795	1.461

convergence rate of $\|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2$ can also be very small, and the constant, 0.05, assigned to $\mathrm{OP}(d)$ fails to do a good job in this extremal case.

It is relatively difficult to understand the behaviors of $\hat{l}_2(d)$ through $\mathrm{OP}(d)$ when short-memory AR or MA components are added into the I(d) model. However, using the $\hat{l}_2(d)$ in DGP 1 as the basis for comparison, it seems fair to comment that the AR component tends to increase $\hat{l}_2(d)$ with $d \geq 0.25$ and $d = 0.01$, whereas the MA component tends to increase $\hat{l}_2(d)$ with $d \leq 0.25$. When both components are included, the values of $\hat{l}_2(d)$ are uniformly larger than those in the I(d) case. We leave a further investigation of the impact of the AR and MA components on the finite sample performance of $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$ as a future work.

In the following, we shall perform a sensitivity analysis of the SAR method by perturbing the parameter c in cK_n. We define the sensitivity function

$$\mathrm{SF}(c) = \frac{\|\hat{\mathbf{\Omega}}_n^{-1}(cK_n) - \mathbf{\Omega}_n^{-1}\|_2 - \|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2}{\|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2}.$$

For each $c = 0.8, 1.2$, $d = 0.1, 0.25, 0.45$, DGP = DGP 1–4, and $n = 250, 500, 1000, 2000, 4000$, we compute the average of $\mathrm{SF}(c)$, denoted by $\overline{\mathrm{SF}}(c)$, based on 1000 replications, and the five-number summaries of $\overline{\mathrm{SF}}(c)$ for each n are presented in Table 3. Table 3 shows that the maximum values of $\overline{\mathrm{SF}}(c)$ are all positive and decrease as n increases. In contrast, the minimum values of $\overline{\mathrm{SF}}(c)$ are all negative and start to increase when $n \geq 1000$. When $n = 4000$, the maximum $\overline{\mathrm{SF}}(c)$ and minimum $\overline{\mathrm{SF}}(c)$ are 0.152 and -0.194, respectively, yielding that the average of $\|\hat{\mathbf{\Omega}}_n^{-1}(cK_n) - \mathbf{\Omega}_n^{-1}\|_2$ falls between 0.806–1.152 times the average of $\|\hat{\mathbf{\Omega}}_n^{-1}(K_n) - \mathbf{\Omega}_n^{-1}\|_2$, for all c's, d's and DGPs under consideration. Our analysis reveals that a small perturbation of K_n will not lead to a drastic change on estimation errors.

Table 3. 5-number summaries of $\overline{\mathrm{SF}}(c)$

n	Minimum	1st quartile	Median	3rd quartile	Maximum
250	−0.173	−0.109	0.050	0.087	0.289
500	−0.229	−0.074	0.044	0.150	0.228
1000	−0.250	−0.022	0.029	0.160	0.210
2000	−0.210	−0.025	−0.005	0.134	0.156
4000	−0.194	−0.029	−0.014	0.112	0.152

5.3. Finite sample performance of $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$

We consider three polynomial regression models:

Model 1: $y_t = 1 + u_t, t = 1, 2, \ldots, n$,
Model 2: $y_t = 1 + 2t + u_t, t = 1, 2, \ldots, n$,
Model 3: $y_t = 5 + t + 2t^4 + u_t, t = 1, 2, \ldots, n$,

where u_t's are generated by DGP 1. The performance of $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ (with K_n determined by the SAR method) is investigated with polynomial degree known or unknown. In the latter situation, we perform best subset selection in the following fifth-order model,

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \beta_5 t^5 + u_t, \qquad t = 1, 2, \ldots, n,$$

using the selection criterion,

$$L_n(M) = \log \hat{\sigma}_n^2(M) + \#M/\log(n), \tag{5.3}$$

suggested by Ing and Wei [10], where $\mathcal{M} = \{M\colon\ M \subseteq \{1, t, t^2, t^3, t^4, t^5\}\}$ and $\hat{\sigma}_n^2(M)$ is the residual mean square error of model M. Note that according to Theorem 4.1 of Ing and Wei [10], $L_n(M)$ is a consistent criterion in regression models with long-memory errors. The performance of $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ is evaluated by $\tilde{l}_2(d)$, which is $\hat{l}_2(d)$ with u_t's replaced by the corresponding detrended series. The values of $\tilde{l}_2(d)$ are documented in Table 4, in which $d \in \{0.1, 0.25, 0.4\}$, $n \in \{250, 500, 1000, 2000, 4000\}$ and models are known or selected by $L_n(M)$. Table 4 also reports the correct selection frequencies (in 1000 simulations), which is denoted by $\hat{q}_i(d)$ for model i and long-memory parameter d.

All $\hat{q}_3(d)$'s are larger than 0.9. However, $\hat{q}_1(0.45)$ and $\hat{q}_2(0.45)$ only fall in the interval (0.44, 0.63) and the intercept (constant time trend) is often excluded by $L_n(M)$ in these cases. In fact, identifying the intercept is a notoriously challenging problem when d is large and the intercept parameter is not far enough away from 0. Fortunately, Table 4 shows that the $\tilde{l}_2(d)$ values obtained with or without model selection procedure are similar, even when $\hat{q}_i(d)$ is much smaller than 1. This result may be due to the fact that under models 1 and 2, the performance of $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ is insensitive to misspecification of the intercept, provided d is large enough. Another interesting finding is that for each regression model considered in this section and each (n, d) combination,

Table 4. $\tilde{l}_2(d)$ with (in parentheses) or without model selection and $\hat{q}_i(d)$

	Model 1					
	$\tilde{l}_2(d)$			$\hat{q}_1(d)$		
$n \setminus d$	0.1	0.25	0.45	0.1	0.25	0.45
250	0.566 (0.566)	0.592 (0.593)	0.702 (0.693)	0.992	0.867	0.558
500	0.461 (0.461)	0.456 (0.454)	0.535 (0.533)	0.999	0.918	0.593
1000	0.385 (0.385)	0.333 (0.332)	0.404 (0.400)	1.000	0.962	0.622
2000	0.358 (0.358)	0.271 (0.271)	0.334 (0.331)	1.000	0.984	0.620
4000	0.353 (0.353)	0.216 (0.216)	0.267 (0.264)	1.000	0.996	0.609
	Model 2					
	$\tilde{l}_2(d)$			$\hat{q}_2(d)$		
$n \setminus d$	0.1	0.25	0.45	0.1	0.25	0.45
250	0.585 (0.574)	0.602 (0.604)	0.690 (0.691)	0.688	0.487	0.455
500	0.477 (0.470)	0.457 (0.455)	0.533 (0.532)	0.839	0.552	0.454
1000	0.399 (0.397)	0.336 (0.336)	0.400 (0.398)	0.947	0.657	0.475
2000	0.368 (0.368)	0.273 (0.275)	0.331 (0.331)	0.995	0.735	0.444
4000	0.359 (0.359)	0.218 (0.218)	0.263 (0.264)	1.000	0.809	0.458
	Model 3					
	$\tilde{l}_2(d)$			$\hat{q}_3(d)$		
$n \setminus d$	0.1	0.25	0.45	0.1	0.25	0.45
250	0.611 (0.611)	0.617 (0.618)	0.687 (0.686)	1.000	0.995	0.904
500	0.493 (0.493)	0.463 (0.463)	0.532 (0.530)	1.000	0.999	0.912
1000	0.411 (0.411)	0.339 (0.339)	0.396 (0.395)	1.000	1.000	0.916
2000	0.376 (0.376)	0.275 (0.275)	0.329 (0.328)	1.000	1.000	0.942
4000	0.318 (0.318)	0.218 (0.218)	0.262 (0.261)	1.000	1.000	0.967

the behavior of $\tilde{l}_2(d)$ coincides with that of $\hat{l}_2(d)$ with DGP = DGP 1. Putting these characteristics together suggests that $\tilde{\mathbf{\Omega}}_n^{-1}(K_n)$ is a reliable surrogate for $\hat{\mathbf{\Omega}}_n^{-1}(K_n)$. This conclusion is particularly relevant in situations where the latter matrix becomes infeasible.

Acknowledgements

The work of Ching-Kang Ing was supported in part by the National Science Council of Taiwan under Grant NSC 97-2628-M-001-022-MY2 and Academia Sinica Investigator Award. The research of Hai-Tang Chiou was supported by Academia Sinica Investigator Award and NSC 102-2118-M-110-002-MY2 from Taiwan's National Science Council. The research of Meihui

Guo was partly supported by Grant number NSC 102-2118-M-110-002-MY2, from Taiwan's National Science Council.

References

[1] Berk, K.N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502. MR0421010

[2] Bickel, P.J. and Gel, Y.R. (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 711–728. MR2867455

[3] Bickel, P.J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

[4] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR1093459

[5] Cai, T.T., Ren, Z. and Zhou, H.H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probab. Theory Related Fields* **156** 101–143. MR3055254

[6] Cheng, T.-C.F., Ing, C.-K. and Yu, S.-H. (2015). Toward optimal model averaging in regression models with time series errors. *J. Econometrics*. To appear.

[7] Findley, D.F. and Wei, C.Z. (1993). Moment bounds for deriving time series CLTs and model selection procedures. *Statist. Sinica* **3** 453–480. MR1243396

[8] Godet, F. (2010). Prediction of long memory processes on same-realisation. *J. Statist. Plann. Inference* **140** 907–926. MR2574654

[9] Ing, C.-K. and Wei, C.-Z. (2003). On same-realization prediction in an infinite-order autoregressive process. *J. Multivariate Anal.* **85** 130–155. MR1978181

[10] Ing, C.-K. and Wei, C.-Z. (2006). A maximal moment inequality for long range dependent time series with applications to estimation and model selection. *Statist. Sinica* **16** 721–740. MR2281299

[11] Inoue, A. and Kasahara, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.* **34** 973–993. MR2283400

[12] Kokoszka, P.S. and Taqqu, M.S. (1995). Fractional ARIMA with stable innovations. *Stochastic Process. Appl.* **60** 19–47. MR1362317

[13] McMurry, T.L. and Politis, D.N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482. MR2732601

[14] Palma, W. and Pourahmadi, M. (2012). Banded regularization and prediction of long-memory time series. Working paper.

[15] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164. MR0557560

[16] Wei, C.Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* **15** 1667–1682. MR0913581

[17] Wu, W.B., Michailidis, G. and Zhang, D. (2004). Simulating sample paths of linear fractional stable motion. *IEEE Trans. Inform. Theory* **50** 1086–1096. MR2094869

[18] Wu, W.B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. MR2024760

[19] Wu, W.B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MR2589209

[20] Xiao, H. and Wu, W.B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. MR3014314

[21] Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791–807. MR0947579

[22] Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann*. *Statist*. **19** 158–177. MR1091844

Received December 2013 and revised November 2014