Estimation of autocovariance matrices for high dimensional linear processes

Konrad Furmańczyk¹

Received: 25 January 2019 / Published online: 26 August 2020
© The Author(s) 2020

Abstract

In this paper under some mild restrictions upper bounds on the rate of convergence for estimators of $p \times p$ autocovariance and precision matrices for high dimensional linear processes are given. We show that these estimators are consistent in the operator norm in the sub-Gaussian case when $p = \mathcal{O}(n^{\gamma/2})$ for some $\gamma > 1$, and in the general case when $p^{2/\beta}(n^{-1} \log p)^{1/2} \rightarrow 0$ for some $\beta > 2$ as $p = p(n) \rightarrow \infty$ and the sample size $n \rightarrow \infty$. In particular our results hold for multivariate AR processes. We compare our results with those previously obtained in the literature for independent and dependent data. We also present non-asymptotic bounds for the error probability of these estimators.

Keywords High dimensional data · Linear process · Autocovariance matrix

1 Introduction

Estimation of covariance matrices in a high dimensional setting has been one of the fundamental statistical issues in the last decade. Some statistical applications of estimation of covariance matrices have been presented for ridge regression in Hoerl and Kennard (1970), in regularized discriminant analysis—(Friedman 1989) and in principal component analysis—(Johnstone and Lu 2009). For an overview of this topic and its applications see (Bickel and Levina 2008b; Birnbaum and Nadler 2012; Chen et al. 2013; Fan et al. 2006; Rothman et al. 2009). The problem of estimating covariance matrices for dependent data has been recently investigated by (Chen et al. 2013; Bhattacharjee and Bose 2014a, b; Guo et al. 2016; Jentsch and Politis 2015; McMurry and Politis 2010), and Wu and Pourahmadi (2009). The estimation of the inverse covariance matrix is used in the recovery of the true unknown structure of undirected

✉ Konrad Furmańczyk
konrad_furmanczyk@sggw.edu.pl

¹ Institute of Information Technology, Warsaw University of Life Sciences (SGGW), Nowoursynowska 159, 02-776 Warsaw, Poland

graphical models, especially in Gaussian graphical models, where a zero entry of the inverse covariance matrix is associated with a missing edge between two vertices in the graph. The recovery of undirected graphs on the basis of the estimation of the precision matrices for a general class of nonstationary time series is considered in Xu et al. (2020).

Consider a p -dimensional linear process

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} \text{ (almost surely),} \quad (1)$$

where the Ψ_j are $p \times p$ matrices, $\varepsilon_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,p})'$, (ε_t) are i.i.d. vectors in $\mathbb{R}^p$ with mean $\mathbf{0}$ and variance-covariance matrix Σ . Under a causality condition, a vector $ARMA$ process which is a basic model in econometrics and finance is a linear process (Brockwell and Davis 2002). We assume that (ε_t) satisfies one of the following conditions:

(Gauss) ε_i is Gaussian with mean $\mathbf{0}$ and variance-covariance matrix Σ .

(SGauss) $(\varepsilon_{i,l} \varepsilon_{j,s})$ is sub-Gaussian with constant σ^2 , that is,

$$E \exp(u \varepsilon_{i,l} \varepsilon_{j,s}) \leq \exp(\sigma^2 u^2 / 2)$$

for all $u \in \mathbb{R}$, $i, j = 1, 2, \dots$ and $l, s = 1, \dots, p$,

(NGa $_{\beta}$) $E |\varepsilon_{i,j}|^{\beta} < \infty$ for some $\beta > 2$ for $i = 1, 2, \dots$ and $j = 1, \dots, p$.

For example, condition (SGauss) is satisfied for bounded sequences $(\varepsilon_{i,l})$ i.e. $\sup_{i,l} |\varepsilon_{i,l}| \leq M$ for some $M > 0$. We can observe that (SGauss) is implied by sub-Gaussian condition for the vectorization $\text{vec}(\varepsilon'_i \otimes \varepsilon_j)$ of the Kronecker product $\varepsilon'_i \otimes \varepsilon_j$ for all i, j i.e. for all $u \in \mathbb{R}^{p^2}$

$$E \exp(u' \text{vec}(\varepsilon'_i \otimes \varepsilon_j)) \leq \exp(u' \sigma^2 u / 2)$$

for some $\sigma^2 > 0$. Condition (NGa $_{\beta}$) is a moment condition for the innovation process without assumptions on the dependency of the coordinates of the innovation process (ε_i) .

Let Γ_k be the k th order autocovariance matrix,

$$\Gamma_k = \text{Cov}(\mathbf{X}_t, \mathbf{X}'_{t-k}) = \sum_{j=k}^{\infty} \Psi_j \Sigma \Psi'_{j-k}. \quad (2)$$

The matrix Γ_k will be estimated from the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ because in practice we do not know the matrices Ψ_j and Σ . Define the sample autocovariance matrix of order k

as

$$\hat{\mathbf{\Gamma}}_k = \frac{1}{n-k} \sum_{t=k+1}^n \mathbf{X}_t \mathbf{X}'_{t-k} =: (\hat{\gamma}_{ij}^k)$$

for $0 \leq k \leq n-1$, and the banded version of $\hat{\mathbf{\Gamma}}_k$ (as in Bhattacharjee and Bose (2014b)) is given by

$$\mathbf{B}_{l_n}(\hat{\mathbf{\Gamma}}_k) = (\hat{\gamma}_{ij}^k \mathbf{1}(|i-j| \leq l_n)) \quad (3)$$

for some sequence of thresholds $l_n \rightarrow \infty$ as $n \rightarrow \infty$, where $\mathbf{1}(\cdot)$ is the indicator function. We will assume that $p = p(n) \rightarrow \infty$ as $n \rightarrow \infty$.

The main contribution of our paper is that in Theorem 1 we obtain the rate in the operator norm of $\mathbf{B}_{l_n}(\hat{\mathbf{\Gamma}}_k) - \mathbf{\Gamma}_k$ for high dimensional linear process

$$\|\mathbf{B}_{l_n}(\hat{\mathbf{\Gamma}}_k) - \mathbf{\Gamma}_k\|_2 = \mathcal{O}_P(\|\mathbf{\Sigma}\|_{(1,1)}(c_n)^{\alpha/(\alpha+1)}),$$

for some $c_n \rightarrow 0$ and some $\alpha > 0$. Under the sub-Gaussian condition ((Gauss) and (SGauss)) we obtain the same rate $\mathcal{O}_P((n^{-1} \log p)^{\alpha/2(\alpha+1)})$ as Bhattacharjee and Bose (2014b) but under weaker assumptions for matrices of coefficients Ψ_j and $\mathbf{\Sigma}$. In particular, under causality, our results include vector autoregressive $AR(r)$ processes. Similar results (Corollary 1) we obtain for the precision matrix

$$\|\mathbf{B}_{l_n}^{-1}(\hat{\mathbf{\Gamma}}_k) - \mathbf{\Gamma}_k^{-1}\|_2 = \mathcal{O}_P(\|\mathbf{\Sigma}\|_{(1,1)}(c_n)^{\alpha/(\alpha+1)}).$$

An interesting problem is to obtain lower bounds and the optimal rate of convergence. Cai et al. (2010) obtained the minimax bound for i.i.d. observations for tapering estimators of $\mathbf{\Gamma}_0$. For dependent data this problem is still open. Below we briefly present the state of research related to the estimation of the covariance matrix for independent and dependent observations.

The sample covariance matrix $\hat{\mathbf{\Gamma}}_0 = (\hat{\gamma}_{ij}^0)$ performs poorly in a high dimensional setting. In the Gaussian case when $\mathbf{\Gamma}_0 = \mathbf{I}$ is the identity matrix and $p/n \rightarrow c \in (0, 1)$, the empirical distribution of the eigenvalues of the sample covariance matrix $\hat{\mathbf{\Gamma}}_0$ follows the Marchenko and Pastur (1967) law, which is supported on the interval $((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$.

For the i.i.d. case Bickel and Levina (2008a) proposed thresholding of the sample covariance matrix and obtained rates of convergence for the thresholding estimators for a proper choice of the threshold λ_n , where the estimator is given by

$$\hat{\gamma}_{ij}^\lambda = \hat{\gamma}_{ij}^0 \mathbf{1}(\hat{\gamma}_{ij}^0 \geq \lambda_n).$$

Rothman et al. (2009) considered a class of universal thresholding rules with more general thresholding functions than hard thresholding. An interesting generalization of this method can be found in Cai and Liu (2011) for sparse covariance matrices,

where an adaptive thresholding estimator is given by

$$\hat{\gamma}_{ij}^* = S_{\lambda_{ij}}(\hat{\gamma}_{ij}^0),$$

where $S_{\lambda_{ij}}(\cdot)$ is a general thresholding function with data-driven thresholds λ_{ij} . For other interesting results in this area, see (Birnbaum and Nadler 2012; Cai et al. 2010; Fan et al. 2006; Furrer and Bengtsson 2007; Huang et al. 2006).

There are few results for high-dimensional dependent data: (Bhattacharjee and Bose 2014a,b; Chen et al. 2013; Jentsch and Politis 2015) and Guo et al. (2016). Bhattacharjee and Bose (2014a) considered the estimation of the high dimensional variance-covariance matrix under a general Gaussian model with weak dependence in both rows and columns of the data matrix. They showed that the bounded and tapered sample variance-covariance matrices are consistent under a suitable column dependence model. But their conditions do not allow control of the first few autocovariances, only they control higher order autocovariances. Bhattacharjee and Bose (2014b) showed that under suitable assumptions for the linear process, the banded sample autocovariance matrices are consistent in the high dimensional setting. Chen et al. (2013) obtained the rate of convergence for a banded autocovariance estimator in operator norm for a general dependent model. A similar result under more restricted assumptions was obtained by (Jentsch and Politis 2015; Guo et al. 2016) established similar results for multivariate sparse autoregressive AR processes.

The rest of the paper is organized as follows. In “The rate of convergence of autocovariance estimation” section we deal with the problem of estimating the k th order autocovariance matrix Γ_k by (3) in high dimensions (Theorem 1). The rate of convergence for the estimator of the precision matrix Γ_k^{-1} is given in Corollary 1.

In Proposition 1 we obtain bounds on the error probability of the estimation of the k th order autocovariance. In “Comparison of our results with previous studies” section we compare our results with the results obtained by (Bickel and Levina 2008a,b) for independent normal and nonnormal data, with the minimax upper bound for tapering estimator in Cai et al. (2010) and with the results for dependent data obtained by (Chen et al. 2013; Bhattacharjee and Bose 2014b; Guo et al. 2016) and Jentsch and Politis (2015). In the special case of a multi-dimensional linear process we obtain a sharper bound (Theorem 1) than Chen et al. (2013) and Jentsch and Politis (2015). We also obtain a better rate for the estimation error for multivariate sparse AR processes than Guo et al. (2016).

Finally, the conclusions are presented in Sect. 2.5. All the proofs and auxiliary lemmas are given in the “Appendix”.

2 The rate of convergence of autocovariance estimation

For any $p \times p$ matrix $\mathbf{M} = (m_{ij})$ we define the following matrix norms which are convenient for comparison with other results of autocovariance estimation:

$$\|\mathbf{M}\|_2 = \sqrt{\lambda_{\max}(\mathbf{M}'\mathbf{M})} \text{ (operator norm),}$$

where $\lambda_{\max}(\mathbf{M}'\mathbf{M})$ is the maximum eigenvalue of the matrix $\mathbf{M}'\mathbf{M}$,

$$T(\mathbf{M}, t) = \max_{1 \leq j \leq p} \sum_{i: |i-j| > t} |m_{ij}|$$

for some threshold $t > 0$,

$$\|\mathbf{M}\|_{(1,1)} = \max_{1 \leq j \leq p} \sum_{i=1}^p |m_{ij}|,$$

$$\|\mathbf{M}\|_{\infty} = \max_{1 \leq i, j \leq p} |m_{ij}|.$$

We consider the following conditions on the matrices Ψ_j (see (1)), Γ_k and Σ for all $t > 0$ and $j = 0, 1, \dots$:

(A1) there exists a sequence d_j with $\sum_{j=1}^{\infty} d_j^2 < \infty$ such that

$$\max(T(\Psi_j, t), T(\Psi_j', t)) \leq C_1 d_j t^{-\alpha}$$

for some constants $C_1 > 0, \alpha > 0$,

(A2) $T(\Sigma, t) \leq C_2 t^{-\alpha}$ for some constants $C_2 > 0, \alpha > 0$,

(A3) $\sum_{j=1}^{\infty} r_j^2 < \infty$, where $r_j = \max(\|\Psi_j\|_{(1,1)}, \|\Psi_j'\|_{(1,1)})$,

(A4) $\sum_{j=1}^{\infty} r_j < \infty$, where $r_j = \max(\|\Psi_j\|_{(1,1)}, \|\Psi_j'\|_{(1,1)})$,

(A5) $\lambda_{\max}(\Sigma) \leq C_3$ for some constant $C_3 > 0$.

These conditions are restrictions on the parameter space. It is obvious that (A3) implies (A4) but the converse is not true. If the covariance matrix $\Sigma = (\gamma_{ij})$ is such that $|\gamma_{ij}| \leq C|i-j|^{-\alpha-1}$ for all i, j for some $\alpha > 0$, then $T(\Sigma, t) \leq C_2 t^{-\alpha}$ and (A2) holds. Conditions (A1)-(A2) are tapering conditions and specify the rate of decay for the matrices Ψ_j and Σ away from the diagonal.

Of course, if (A3) holds then $\|\Gamma_k\|_{(1,1)} < \infty$ (see (2)), and $\sum_{j=0}^{\infty} \|\Psi_j\|_{(1,1)}^2 < \infty$ implies that the series in (1) converges almost surely.

Next, in “Remarks on the condition (A1) for $AR(1)$ processes” and “Remarks on the condition (A1) for $AR(r)$ processes” sections in a few remarks we will give a discussion about condition (A1) is fulfilled for vector $AR(1)$ and $AR(r)$ processes. In “Comparison of our results with previous studies” section we compare our main result Theorem 1 with the previous studies. In “Conclusions” section we summarize results related to Theorem 1 available in the literature.

2.1 The main results

The main results in this section concern the rate of convergence in operator norm for the k th order autocovariance matrix Γ_k and for the precision matrix Γ_k^{-1} .

Theorem 1 Suppose (A1)-(A2) hold. Then

$$\|\mathbf{B}_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = \mathcal{O}_P(\|\Sigma\|_{(1,1)}(c_n)^{\alpha/(\alpha+1)}) \quad (4)$$

for $l_n = (c_n)^{-1/(\alpha+1)}$. Here $c_n = \sqrt{n^{-1} \log p}$ when ((Gauss) and (A5)) or (SGauss)) and (A3) hold, and $p = \mathcal{O}(n^{\gamma/2})$ for some $\gamma > 1$ as $n \rightarrow \infty$; and $c_n = p^{2/\beta} \sqrt{n^{-1} \log p}$ when (NGa $_{\beta}$) and (A4) hold, and $p^{2/\beta} \sqrt{n^{-1} \log p} \rightarrow 0$ as $n \rightarrow \infty$.

The proof of Theorem 1 is similar to the proofs of Theorems 1-2 in Bhattacharjee and Bose (2014b), but instead of their Lemmas 3, 5 we use a maximal inequality for the (SGauss) case and Pisier's inequality (see (Pisier 1983)) for the (NGa $_{\beta}$) case.

Corollary 1 Under the assumptions of Theorem 1 and when Γ_k^{-1} exists and $\|\Gamma_k^{-1}\|_2 = \mathcal{O}(1)$, we have

$$\|\mathbf{B}_{l_n}^{-1}(\hat{\Gamma}_k) - \Gamma_k^{-1}\|_2 = \mathcal{O}_P(\|\Sigma\|_{(1,1)}(c_n)^{\alpha/(\alpha+1)}), \quad (5)$$

where l_n and c_n are defined in Theorem 1.

Corollary 1 is a simple consequence of Theorem 1. We also obtain the rate of convergence of $\hat{\Gamma}_k$ in supremum norm, which will be used in the proof of Theorem 1.

Lemma 1 We have

$$\|\hat{\Gamma}_k - \Gamma_k\|_{\infty} = \mathcal{O}_P(c_n), \quad (6)$$

where $c_n = \sqrt{n^{-1} \log p}$ if (SGauss) and (A3) hold and $p = \mathcal{O}(n^{\gamma/2})$ for some $\gamma > 1$ as $n \rightarrow \infty$, and $c_n = p^{1/\beta} \sqrt{n^{-1} \log p}$ when (NGa $_{\beta}$) and (A4) hold and $p^{2/\beta} \sqrt{n^{-1} \log p} \rightarrow 0$ as $n \rightarrow \infty$.

Next, we obtain non-asymptotic bounds on the error probability of the estimation of the k th order autocovariance.

Proposition 1 Suppose (A1)-(A2) hold. Then for any $\eta > \tilde{C}_1 \|\Sigma\|_{(1,1)} + \tilde{C}_2$,

$$\begin{aligned} &P\left(\|\mathbf{B}_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 > \eta\right) \\ &\leq 2p^2 \sum_{i,j=0}^{\infty} \exp\left(-\frac{(n-k)^2(\eta - (\tilde{C}_1 \|\Sigma\|_{(1,1)} + \tilde{C}_2)l_n^{-\alpha})^2}{2\sigma^2(2l_n+1)^2 r_i^2 r_j^2}\right) \end{aligned} \quad (7)$$

for some constants $\tilde{C}_1, \tilde{C}_2$ where (SGauss) and (A3) hold, and

$$P\left(\|\mathbf{B}_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 > \eta\right) \leq C_{\beta} C_* \frac{p^{2/\beta} n^{1/2} (2l_n+1)}{(n-k)(\eta - (\tilde{C}_1 \|\Sigma\|_{(1,1)} + \tilde{C}_2)l_n^{-\alpha})} \quad (8)$$

for some constants $\tilde{C}_1, \tilde{C}_2, C_*$ ($\sum_{i=0}^{\infty} r_i$)² and C_{β} depending on β when (NGa $_{\beta}$) and (A4) hold.

2.2 Remarks on the condition (A1) for AR(1) processes

As a special case, we consider a multivariate AR(1) process

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{A}^j \varepsilon_{t-j} \quad (9)$$

for some $p \times p$ matrix $\mathbf{A}$. Then $\Psi_j = \mathbf{A}^j$ for $j = 1, 2, \dots$. Let $t > 0$. We impose one of two conditions on the matrix $\mathbf{A}$:

- (w1) there exists $\alpha > 0$, a constant $C > 0$, and a sequence b_j such that $\sum_{j=1}^{\infty} j b_j \max(\|\mathbf{A}\|_{(1,1)}^{j-1}, \|\mathbf{A}'\|_{(1,1)}^{j-1}) < \infty$ and $\max(\|\mathbf{A}\|_{(1,1)}, \|\mathbf{A}'\|_{(1,1)}) < 1$ and $\max(T(\mathbf{A}, t/2^{j-1}), T(\mathbf{A}', t/2^{j-1})) \leq C b_j t^{-\alpha}$.
- (w2) $\max(\|\mathbf{A}\|_{(1,1)}, \|\mathbf{A}'\|_{(1,1)}) < 2^{-\alpha}$ and $\max(T(\mathbf{A}, t/2^{j-1}), T(\mathbf{A}', t/2^{j-1})) \leq C t^{-\alpha}$ for some $\alpha > 0$ and some constant $C > 0$.

Remark A Condition (A1) is fulfilled for $d_j = j b_j \max(\|\mathbf{A}\|_{(1,1)}^{j-1}, \|\mathbf{A}'\|_{(1,1)}^{j-1})$ when (w1) holds, because from Lemma 5 (see “Appendix”)

$$\begin{aligned} \max(T(\Psi_j, t), T(\Psi_j', t)) &= \max(T(\mathbf{A}^j, t), T((\mathbf{A}')^j, t)) \\ &\leq j \max(\|\mathbf{A}\|_{(1,1)}^{j-1} T(\mathbf{A}, t/2^{j-1}), \|\mathbf{A}'\|_{(1,1)}^{j-1} T(\mathbf{A}', t/2^{j-1})). \end{aligned}$$

Similarly, condition (A1) is fulfilled for $d_j = j 2^{(j-1)\alpha} \max(\|\mathbf{A}\|_{(1,1)}^{j-1}, \|\mathbf{A}'\|_{(1,1)}^{j-1})$ when (w2) holds. Condition (A3) holds when $\max(\|\mathbf{A}\|_{(1,1)}, \|\mathbf{A}'\|_{(1,1)}) < 1$, because

$$\sum_{j=1}^{\infty} r_j = \sum_{j=1}^{\infty} \max(\|\Psi_j\|_{(1,1)}, \|\Psi_j'\|_{(1,1)}) \leq \sum_{j=1}^{\infty} \max(\|\mathbf{A}\|_{(1,1)}^j, \|\mathbf{A}'\|_{(1,1)}^j).$$

2.3 Remarks on the condition (A1) for AR(r) processes

Now, we consider a multivariate AR(r) process of order r given by

$$\mathbf{X}_t = \sum_{i=1}^r \mathbf{A}_i \mathbf{X}_{t-i} + \varepsilon_t,$$

$t \geq 1$, where the $p \times p$ matrices $\mathbf{A}_i, i = 1, \dots, r$, are called parameter matrices. Under some regularity condition (for more details see ((Bhattacharjee and Bose 2014b) (4), p. 264) and Brockwell and Davis (2002)) this process has the representation (1), where $\Psi_0 = \mathbf{I}$ and

$$\Psi_j = \sum_{i=1}^{\min(j,r)} \mathbf{A}_i \Psi_{j-i} \text{ for } j = 1, 2, \dots \quad (10)$$

We consider the class $\mathcal{A}$ of r -sequences of matrices, defined by

$$\mathcal{A} = \left\{ \begin{array}{l} \max(\|\mathbf{A}_i\|_{(1,1)}, \|\mathbf{A}'_i\|_{(1,1)}) \leq C_1 \delta^i \text{ and for all } t > 0, \\ \max(T(\mathbf{A}_i, t), T(\mathbf{A}'_i, t)) \leq C_2 \delta^i t^{-\alpha} \text{ for some } 0 < \delta < 1, \\ \alpha > 0, C_1 = 1/(2^\alpha r), i = 1, 2, \dots, r \text{ for some constant } C_2 > 0. \end{array} \right.$$

Observe that if $T(\mathbf{A}_i, t) \leq C_2 \delta^i t^{-\alpha}$ for some $0 < \delta < 1$ and

$$\max_{1 \leq k \leq p} |a_i^{k,k}| \leq C \delta^i,$$

where $\mathbf{A}_i = (a_i^{s,k})$, then $\|\mathbf{A}_i\|_{(1,1)} \leq (2^\alpha C_2 + C) \delta^i$. Indeed, for any $t > 0$,

$$\begin{aligned} \|\mathbf{A}_i\|_{(1,1)} &\leq T(\mathbf{A}_i, t) + \max_{1 \leq k \leq p} \sum_{s: |s-k| \leq t} |a_i^{s,k}| \\ &\leq C_2 \delta^i t^{-\alpha} + \max_{1 \leq k \leq p} \sum_{s: |s-k| \leq t} |a_i^{s,k}|. \end{aligned}$$

Therefore, putting $t = 1/2$, we get

$$\|\mathbf{A}_i\|_{(1,1)} \leq C_2 2^\alpha \delta^i + \max_{1 \leq k \leq p} |a_i^{k,k}| \leq (2^\alpha C_2 + C) \delta^i.$$

Similarly, we may obtain $\|\mathbf{A}'_i\|_{(1,1)} \leq (2^\alpha C_2 + C) \delta^i$.

Proposition 2 *Under the condition of class $\mathcal{A}$, we have*

$$\max(\|\Psi_j\|_{(1,1)}, \|\Psi'_j\|_{(1,1)}) \leq C_1 \delta^j \quad (11)$$

and

$$\max(T(\Psi_j, t), T(\Psi'_j, t)) \leq C_2 j \delta^j t^{-\alpha} \quad (12)$$

for $j = 1, 2, \dots$

Remark B Immediately from Proposition 2, we have that under the condition of class $\mathcal{A}$ condition (A1) holds for $d_j = j \delta^j$ for multivariate $AR(r)$ processes.

2.4 Comparison of our results with previous studies

Similar results to those given in Theorem 1 have been presented in Bickel and Levina (2008b) for i.i.d. Gaussian observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ and in Bhattacharjee and Bose (2014b) for a p -dimensional linear process with $n^{-1} \log p \rightarrow 0$. In Bhattacharjee and Bose (2014b) it is assumed that the matrix Σ belongs to the class

$$\mathcal{U} = \{ \Sigma : 0 < \varepsilon < \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) < 1/\varepsilon, T(\Sigma, t) < C t^{-\alpha} \text{ for all } t > 0 \},$$

where $\varepsilon, \alpha, C > 0$ and $\lambda_{\min}(\mathbf{\Sigma}), \lambda_{\max}(\mathbf{\Sigma})$ are respectively the minimum and maximum eigenvalues of $\mathbf{\Sigma}$, the coefficient matrices (Ψ_j) are in $\mathcal{T}_{\beta, \lambda} \cap \mathcal{G}_{\alpha, \eta, \nu}$ for some $0 < \beta < 1, \lambda \geq 0, \alpha, \nu > 0, 0 < \eta < 1$, where

$$\mathcal{T}_{\beta, \lambda} = \left\{ (\Psi_j) : \sum_{j=0}^{\infty} r_j^{\beta} < \infty, \sum_{j=0}^{\infty} r_j^{2(1-\beta)} j^{\lambda} < \infty \right\},$$

$$\mathcal{G}_{\alpha, \eta, \nu} = \left\{ (\Psi_j) : T(\Psi_j, t \sum_{u=0}^j \eta^u) < C t^{-\alpha} r_j j^{\nu} \sum_{u=0}^j \eta^{-u\alpha}, \text{ and } \sum_{j=k}^{\infty} \frac{r_j r_{j-k} j^{\nu}}{\eta^{\alpha j}} < \infty \right\},$$

with $r_j = \max(\|\Psi_j\|_{(1,1)}, \|\Psi_j'\|_{(1,1)})$. Additionally they assumed that for some $\lambda_0 > 0$,

$$\sup_{j \geq 1} E(e^{\lambda \varepsilon_{1,j}}) < \infty \text{ for all } |\lambda| < \lambda_0. \quad (13)$$

In our Theorem 1 we consider (Gauss), (SGauss) or (NGa $_{\beta}$) for (ε_i) . In particular, condition (NGa $_{\beta}$) is much weaker than (13). The class $\mathcal{U}$ implies our conditions (A2) and (A5). Also conditions (A1) and (A3) are weaker than $\mathcal{T}_{\beta, \lambda} \cap \mathcal{G}_{\alpha, \eta, \nu}$. For example if $r_j \sim j^{-a}$ for $a > 1$, then (A4) holds but the conditions in $\mathcal{T}_{\beta, \lambda}$ are satisfied for $a > \max(1/\beta, (\lambda + 1)/2(1 - \beta))$, and those in $\mathcal{G}_{\alpha, \eta, \nu}$ are satisfied for $r_j \sim b^j$ for some $0 < b < 1$.

Bickel and Levina (2008a) considered asymptotic behavior of the threshold estimator of the covariance matrix $\mathbf{\Gamma}_0$ of the form

$$T_u(\hat{\mathbf{\Gamma}}_0) = (\hat{\gamma}_{ij}^0 \mathbf{1}(|\hat{\gamma}_{ij}^0| \geq u)). \quad (14)$$

For Gaussian data, where $\mathbf{\Gamma}_0 \in \mathcal{G}_r(M)$ and

$$\mathcal{G}_r(M) := \left\{ \mathbf{\Gamma}_0 = (\gamma_{ij}) : \max_{1 \leq i \leq p} \gamma_{ii} \leq 1, \max_{1 \leq i \leq p} \sum_{j=1}^p |\gamma_{ij}|^r \leq M \right\}, 0 \leq r < 1, \quad (15)$$

they obtained

$$\|T_u(\hat{\mathbf{\Gamma}}_0) - \mathbf{\Gamma}_0\|_2 = \mathcal{O}_P(M u^{1-r}) \quad (16)$$

where $u = C(\log p)^{1/2} n^{-1/2}$ for a sufficiently large constant C . For nonnormal data where $\mathbf{\Gamma}_0 \in \mathcal{G}_r(M)$, they obtained (16) for $u = C p^{2/q} n^{-1/2}$, for a sufficiently large constant C (see (Chen et al. 2013), (48)).

Cai et al. (2010) obtained the minimax upper bound for a special class of tapering estimators of Γ_0 for i.i.d. observations under (A2) and (A5). Its upper rate bound equals $\mathcal{O}_P(\min\{n^{-2\alpha/(2\alpha+1)} + n^{-1} \log p, p/n\})$ and is better than our result for (Gauss) and (SGauss) from Theorem 1. It means that for i.i.d. observations our result is suboptimal.

A sharper rate than in (16) for a nonnormal case was obtained by Chen et al. (2013), where data come from a general weak dependence model. In particular, when data come from a linear process (1) with (NGa_β) for $\beta = 2q$ for some $q > 2$, and the coefficient matrices $\Psi_j = (\psi_{k,s}(j))_{1 \leq k,s \leq p}$ satisfy

$$\max_{1 \leq k \leq p} \sum_{s=1}^p (\psi_{k,s}(j))^2 = \mathcal{O}(j^{-(2+2\gamma)})$$

for some $\gamma > 1/2 - 1/q$, $\Gamma_0 \in \mathcal{G}_r(M)$ and $M < p$, then one can obtain (16) for $u = u_*$, where $u_* = \max(u_1, u_2, u_3)$, $u_1 = M^{\frac{1}{q+r}} p^{\frac{1}{q+r}} n^{\frac{1-q}{q+r}}$, $u_2 = \sqrt{n^{-1} \log p}$, $u_3 = M^{-\frac{2}{q-2r}} p^{\frac{2}{q-2r}} n^{\frac{(1-q)}{q-2r}}$ (for more details see discussion preceding Corollary 2.7 in Chen et al. (2013)).

Under condition (Gauss) or (SGauss), we obtain the same rate of convergence as in Bickel and Levina (2008a) for the covariance matrix Γ_0 for normal data and as in Bhattacharjee and Bose (2014b) for Γ_k . In particular, if $\Gamma_0 \in \mathcal{G}_r(M)$ then $\|\Sigma\|_{(1,1)} \leq M$, and from Theorem 1 we have

$$\|\mathbf{B}_{l_n}(\hat{\Gamma}_0) - \Gamma_0\|_2 = \mathcal{O}_P(M(n^{-1} \log p)^{\frac{\alpha}{2(\alpha+1)}}). \quad (17)$$

If $M \sim p^\eta$ for some $\eta \in [0, 1)$, then the r.h.s. of (16) for $u = u_*$ equals

$$\max\left(p^{\frac{(1-r)(2-2\eta)}{q+r} + \eta} n^{\frac{(1-q)(1-r)}{q+r}}, p^\eta (n^{-1} \log p)^{\frac{1-r}{2}}, p^{\eta + \frac{(1-r)(2-2\eta)}{q-2r}} n^{\frac{(1-q)(1-r)}{q-2r}}\right), \quad (18)$$

which is greater than the r.h.s. of (17), $p^\eta (n^{-1} \log p)^{\frac{\alpha}{2(\alpha+1)}}$. Thus, our rate of convergence is better than the rate (16) for $u = u_*$ in Chen et al. (2013) for a linear process.

Guo et al. (2016) worked with a bounded covariance estimator for $l_n = C \log(n/\log p)$ for some $C > 0$ for a sparse multivariate AR model. The rate of convergence for operator norm was $\mathcal{O}_P(\log(n/\log p) \sqrt{n^{-1} \log p})$. In a special case, from our Theorem 1 we obtain the better rate $\mathcal{O}_P(\log^\alpha(n^{-1} \log p))$.

Jentsch and Politis (2015) dealt with so called flat-top tapered covariance matrix estimation for a multivariate strictly stationary time series process. In the special case when the observations come from a p dimensional linear process $\mathbf{X}_t = \sum_{j=-\infty}^{\infty} \Psi_j \varepsilon_{t-j}$, where the sequence (Ψ_j) of coefficient matrices is component-wise absolutely summable (then $\sum_{h=-\infty}^{\infty} \sum_{i,j=1}^p |\gamma_{i,j}(h)| < \infty$, where $\Gamma_k = (\gamma_{i,j}(h))$)

with i.i.d. noise (ε_t) with fourth moments, they obtained

$$\left(E\|\hat{\Gamma}_{k,l} - \Gamma_k\|_2^2\right)^{1/2} = \mathcal{O}\left(\frac{lp^2}{\sqrt{n}} + \sum_{h=l+1}^{n-1} \sum_{j=1}^p |\gamma_{i,j}(h)|\right), \quad (19)$$

where $\hat{\Gamma}_{k,l}$ is a flat-top tapered covariance matrix estimator of Γ_k , where $l = o(\sqrt{n})$ is the banding parameter. If $\psi_{i,j}(h) \leq Ch^{-(1+\alpha)}$ for some $C > 0$ and $\alpha > 0$, where $\Psi_h = (\psi_{i,j}(h))$, then (A1) holds and under (A2) and (SGauss) we deduce from Theorem 1 that the r.h.s. of (4) is of order $\mathcal{O}_P((n^{-1} \log p)^{\alpha/2(\alpha+1)})$. Therefore $|\gamma_{i,j}(h)| = \mathcal{O}(h^{-\alpha-1})$ and for $l = n^{1/2-\kappa}$ for some $\kappa \in (0, 1/2)$ we have $\sum_{h=l+1}^{n-1} \sum_{j=1}^p |\gamma_{i,j}(h)| = \mathcal{O}(l^{-\alpha})$, so the r.h.s. of (19) is of order $\mathcal{O}(p^2 n^{-\kappa} + l^{-\alpha}) = \mathcal{O}(p^2 n^{-\kappa} + n^{-(\frac{1}{2}-\kappa)\alpha})$, which is worse rate than in Theorem 1. Similarly when (NGa β) and $p = \mathcal{O}(n^{\gamma/2})$ for some $\gamma > 1$, then from Theorem 1 the r.h.s. of (4) is of order $\mathcal{O}_P((p^{2/\beta} \sqrt{n^{-1} \log p})^{\alpha/(\alpha+1)})$ and this rate is sharper than the r.h.s. of (19).

2.5 Conclusions

Our main result (Theorem 1) was compared with related results available in the literature. In the special case of a multidimensional linear process we obtained a better rate of convergence of our covariance estimator in operator norm than (Chen et al. 2013; Jentsch and Politis 2015) and Guo et al. (2016). Our result is similar to that of Bhattacharjee and Bose (2014b), but under milder assumptions for the noise process (ε_t) and for the admissible class matrices Γ_k .

Comparing the results of estimating the covariance matrix is difficult because those results are for different assumptions on the class of covariance matrices and for independent or dependent data. For independent data Cai et al. (2010) obtained the optimal rate for tapering estimators of Γ_0 . In contrast such results do not exist for dependent data and the problem of finding the optimal rate of convergence in Theorem 1 is still open.

Appendix

Lemma 2 (Bhattacharjee and Bose 2014b). *For any matrices $\mathbf{A}$, $\mathbf{B}$ and for all α , β , $t > 0$,*

- (i) $\|\mathbf{AB}\|_{(1,1)} \leq \|\mathbf{A}\|_{(1,1)} \|\mathbf{B}\|_{(1,1)}$,
- (ii) $T(\mathbf{AB}, (\alpha + \beta)t) \leq \|\mathbf{A}\|_{(1,1)} T(\mathbf{B}, \alpha t) + \|\mathbf{B}\|_{(1,1)} T(\mathbf{A}, \beta t)$.

Lemma 3 *Suppose (A1)–(A2) and (A3) hold. Then, for all $t > 0$,*

$$T(\Gamma_k, t) \leq (\tilde{C}_1 \|\Sigma\|_{(1,1)} + \tilde{C}_2) t^{-\alpha}$$

for some $\alpha > 0$, where $\tilde{C}_1$, $\tilde{C}_2$ are some constants.

Proof Observe

$$T(\Gamma_k, t) = T\left(\sum_{j=k}^{\infty} \Psi_j \Sigma \Psi'_{j-k}, t\right) \leq \sum_{j=k}^{\infty} T(\Psi_j \Sigma \Psi'_{j-k}, t).$$

From Lemma 2(ii), we have

$$T(\Gamma_k, t) \leq \sum_{j=k}^{\infty} \|\Psi_j\|_{(1,1)} T(\Sigma \Psi'_{j-k}, t/2) + \sum_{j=k}^{\infty} \|\Sigma \Psi'_{j-k}\|_{(1,1)} T(\Psi_j, t/2).$$

It follows from Lemma 2(i) that

$$\|\Sigma \Psi'_{j-k}\|_{(1,1)} \leq \|\Sigma\|_{(1,1)} \|\Psi'_{j-k}\|_{(1,1)}.$$

Hence

$$T(\Gamma_k, t) \leq \sum_{j=k}^{\infty} r_j T(\Sigma \Psi'_{j-k}, t/2) + \sum_{j=k}^{\infty} r_{j-k} \|\Sigma\|_{(1,1)} T(\Psi_j, t/2).$$

Again using Lemma 2(ii), we obtain

$$T(\Sigma \Psi'_{j-k}, t/2) \leq \|\Sigma\|_{(1,1)} T(\Psi'_{j-k}, t/4) + \|\Psi'_{j-k}\|_{(1,1)} T(\Sigma, t/4)$$

and

$$\begin{aligned} T(\Gamma_k, t) &\leq \sum_{j=k}^{\infty} r_j (\|\Sigma\|_{(1,1)} T(\Psi'_{j-k}, t/4) + r_{j-k} T(\Sigma, t/4)) \\ &\quad + \sum_{j=k}^{\infty} r_{j-k} \|\Sigma\|_{(1,1)} T(\Psi_j, t/2). \end{aligned}$$

By assumptions (A1)-(A2), we see that

$$\begin{aligned} T(\Gamma_k, t) &\leq \sum_{j=k}^{\infty} r_j (\|\Sigma\|_{(1,1)} 4^\alpha C_1 d_{j-k} t^{-\alpha} + r_{j-k} 4^\alpha C_2 t^{-\alpha}) \\ &\quad + \sum_{j=k}^{\infty} r_{j-k} \|\Sigma\|_{(1,1)} 2^\alpha C_1 d_j t^{-\alpha} \\ &\leq \|\Sigma\|_{(1,1)} 4^\alpha C_1 t^{-\alpha} \sum_{j=k}^{\infty} r_j d_{j-k} + 4^\alpha C_2 t^{-\alpha} \sum_{j=k}^{\infty} r_j r_{j-k} \\ &\quad + \|\Sigma\|_{(1,1)} 2^\alpha C_1 t^{-\alpha} \sum_{j=k}^{\infty} r_{j-k} d_j. \end{aligned} \tag{20}$$

From the Schwarz inequality, (A1) and (A3), we have

$$\sum_{j=k}^{\infty} r_j d_{j-k} \leq \sqrt{\sum_{j=k}^{\infty} r_j^2} \sqrt{\sum_{j=k}^{\infty} d_{j-k}^2} < \infty$$

and similarly $\sum_{j=k}^{\infty} r_{j-k} d_j < \infty$, $\sum_{j=k}^{\infty} r_j r_{j-k} < \infty$. Therefore, from (20), we get

$$T(\Gamma_k, t) \leq (\tilde{C}_1 \|\Sigma\|_{(1,1)} + \tilde{C}_2) t^{-\alpha} \quad (21)$$

for some constants $\tilde{C}_1 > 0$ and $\tilde{C}_2 > 0$. $\square$

Lemma 4 Suppose (SGauss) and (A3) hold. Then for any $\eta > 0$,

$$P\left(\|\hat{\Gamma}_k - \Gamma_k\|_{\infty} > \eta\right) \leq 2p^2 \sum_{i,j=0}^{\infty} \exp\left(-\frac{(n-k)^2 \eta^2}{2\sigma^2 r_i^2 r_j^2}\right). \quad (22)$$

Suppose (NGa $_{\beta}$) and (A4) hold. Then for any $\eta > 0$,

$$P\left(\|\hat{\Gamma}_k - \Gamma_k\|_{\infty} > \eta\right) \leq C_{\beta} C^* \frac{p^{2/\beta} n^{1/2}}{(n-k)\eta}, \quad (23)$$

where $C^* := (\sum_{i=0}^{\infty} r_i)^2$ and C_{β} is some constant depending on β .

Proof By a simple calculation,

$$\begin{aligned} \|\hat{\Gamma}_k - \Gamma_k\|_{\infty} &= \left\| \frac{1}{n-k} \sum_{m=k+1}^n \mathbf{X}_m \mathbf{X}'_{m-k} - E(\mathbf{X}_m \mathbf{X}'_{m-k}) \right\|_{\infty} \\ &= \left\| \frac{1}{n-k} \sum_{m=k+1}^n \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \Psi_j \varepsilon_{m-j} \varepsilon'_{m-i-k} \Psi'_i - \right. \\ &\quad \left. \frac{1}{n-k} \sum_{m=k+1}^n \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \Psi_j E(\varepsilon_{m-j} \varepsilon'_{m-i-k}) \Psi'_i \right\|_{\infty} \\ &= \left\| \frac{1}{n-k} \sum_{m=k+1}^n \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \Psi_j (\varepsilon_{m-j} \varepsilon'_{m-i-k} - E(\varepsilon_{m-j} \varepsilon'_{m-i-k})) \Psi'_i \right\|_{\infty} \\ &\leq \frac{1}{n-k} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \left\| \Psi_j \Psi'_i \right\|_{(1,1)} \left\| \sum_{m=k+1}^n \varepsilon_{m-j} \varepsilon'_{m-i-k} - E(\varepsilon_{m-j} \varepsilon'_{m-i-k}) \right\|_{\infty} \\ &\leq \frac{1}{n-k} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} r_j r_i \left\| \sum_{m=k+1}^n \varepsilon_{m-j} \varepsilon'_{m-i-k} - E(\varepsilon_{m-j} \varepsilon'_{m-i-k}) \right\|_{\infty}. \end{aligned}$$

Hence

$$\begin{aligned}
 & P\left(\left\|\hat{\mathbf{r}}_k - \mathbf{r}_k\right\|_{\infty} > \eta\right) \\
 & \leq P\left(\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} r_j r_i \left\|\sum_{m=k+1}^n \varepsilon_{m-j} \varepsilon'_{m-i-k} - E(\varepsilon_{m-j} \varepsilon'_{m-i-k})\right\|_{\infty} > (n-k)\eta\right) \\
 & = P\left(\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} r_j r_i \max_{1 \leq l, s \leq p} \left|\sum_{m=k+1}^n \varepsilon_{m-j, l} \varepsilon_{m-i-k, s} - E(\varepsilon_{m-j, l} \varepsilon_{m-i, s})\right| \right. \\
 & \quad \left. > (n-k)\eta\right).
 \end{aligned}$$

Put $\zeta_{m, i, j}^{l, s} := \varepsilon_{m-j, l} \varepsilon_{m-i-k, s}$. Then

$$P\left(\left\|\hat{\mathbf{r}}_k - \mathbf{r}_k\right\|_{\infty} > \eta\right) \leq \sum_{j, i=0}^{\infty} P\left(\max_{1 \leq l, s \leq p} \left|\sum_{m=k+1}^n \zeta_{m, i, j}^{l, s} - E(\zeta_{m, i, j}^{l, s})\right| > \frac{(n-k)\eta}{r_j r_i}\right). \quad (24)$$

From (SGauss) and (A3) we conclude that

$$\begin{aligned}
 & P\left(\max_{1 \leq l, s \leq p} \left|\sum_{m=k+1}^n \zeta_{m, i, j}^{l, s} - E(\zeta_{m, i, j}^{l, s})\right| > \frac{(n-k)\eta}{r_j r_i}\right) \\
 & \leq \sum_{s=1}^p P\left(\max_{1 \leq l \leq p} \left|\sum_{m=k+1}^n \zeta_{m, i, j}^{l, s} - E(\zeta_{m, i, j}^{l, s})\right| > \frac{(n-k)\eta}{r_j r_i}\right).
 \end{aligned}$$

Since for fixed s , $(\zeta_{m, i, j}^{l, s})$ is sub-Gaussian with constant σ^2 , we have

$$E \exp(u \zeta_{m, i, j}^{l, s}) \leq \exp(u^2 \sigma^2 / 2).$$

Hence, from the maximal inequality for sub-Gaussian r.v.'s we obtain

$$P\left(\max_{1 \leq l \leq p} \left|\sum_{m=k+1}^n \zeta_{m, i, j}^{l, s} - E(\zeta_{m, i, j}^{l, s})\right| > \frac{(n-k)\eta}{r_j r_i}\right) \leq 2p \exp\left(-\frac{(n-k)^2 \eta^2}{2\sigma^2 r_j^2 r_i^2}\right).$$

Hence, we have

$$P\left(\left\|\hat{\mathbf{r}}_k - \mathbf{r}_k\right\|_{\infty} > \eta\right) \leq 2p^2 \sum_{i, j=0}^{\infty} \exp\left(-\frac{(n-k)^2 \eta^2}{2\sigma^2 r_j^2 r_i^2}\right).$$

Suppose (NGa $_{\beta}$) and (A4) hold. Applying to the r.h.s. of (24) Pisier's maximal inequality (Pisier 1983)

$$E \max_{1 \leq i \leq N} |Z_i| \leq N^{1/Q} \max_{1 \leq i \leq N} \|Z_i\|_Q,$$

which holds for any random variables (Z_i) with $\|Z_i\|_Q = E^{1/Q} |Z_i|^Q < \infty$ for $Q > 1$, and putting $Q = \beta$, we obtain

$$\begin{aligned} P\left(\max_{1 \leq l, s \leq p} \left| \sum_{m=k+1}^n \zeta_{m,i,j}^{l,s} - E(\zeta_{m,i,j}^{l,s}) \right| > \frac{(n-k)\eta}{r_j r_i}\right) \\ \leq \frac{r_j r_i}{(n-k)\eta} E\left(\max_{1 \leq l, s \leq p} \left| \sum_{m=k+1}^n \zeta_{m,i,j}^{l,s} - E(\zeta_{m,i,j}^{l,s}) \right|\right) \\ \leq \frac{p^{2/\beta} r_j r_i}{(n-k)\eta} \max_{1 \leq l, s \leq p} \left\| \sum_{m=k+1}^n \zeta_{m,i,j}^{l,s} - E(\zeta_{m,i,j}^{l,s}) \right\|_{\beta} \\ \leq \frac{p^{2/\beta}}{(n-k)\eta} r_j r_i \max_{1 \leq l, s \leq p} \left\| \sum_{m=k+1}^n \zeta_{m,i,j}^{l,s} - E(\zeta_{m,i,j}^{l,s}) \right\|_{\beta}. \end{aligned}$$

For fixed i, j, l, s the sequence $(\zeta_{m,i,j}^{l,s})_m$ is i.i.d. and using the moment bound (Petrov 1995) and (NGa $_{\beta}$), we obtain

$$\left\| \sum_{m=k+1}^n \zeta_{m,i,j}^{l,s} - E(\zeta_{m,i,j}^{l,s}) \right\|_{\beta} \leq C_{\beta} n^{1/2}.$$

Hence, we get (23), as desired. $\square$

Proof of Lemma 1 From Lemma 4 under (SGauss), it follows that

$$P\left(\left\| \hat{\mathbf{\Gamma}}_k - \mathbf{\Gamma}_k \right\|_{\infty} > \eta\right) \leq 2p^2 \sum_{i,j=0}^{\infty} \exp\left(-\frac{(n-k)^2 \eta^2}{2\sigma^2 r_j^2 r_i^2}\right).$$

Since for all $x > 0$, we have $\exp(-x) < x^{-\gamma}$ for some $\gamma > 1$, it follows that

$$\sum_{i,j=0}^{\infty} \exp\left(-\frac{(n-k)^2 \eta^2}{2\sigma^2 r_j^2 r_i^2}\right) < \frac{(2\sigma^2)^{\gamma}}{(n-k)^{2\gamma} \eta^{2\gamma}} \left(\sum_{j=0}^{\infty} r_j^{2\gamma}\right)^2.$$

Putting $\eta = \sqrt{n^{-1} \log p}$, we obtain

$$p^2 \sum_{i,j=0}^{\infty} \exp\left(-\frac{(n-k)^2 \eta^2}{2\sigma^2 r_j^2 r_i^2}\right) = \mathcal{O}\left(\frac{p^2}{(n-k)^{2\gamma} \eta^{2\gamma}}\right)$$

$$= \mathcal{O}\left(\frac{p^2}{n^\gamma \log^\gamma p}\right) = \mathcal{O}\left(\frac{1}{\log^\gamma p}\right) = o(1)$$

as $n \rightarrow \infty$. Therefore, (6) holds for $c_n = \sqrt{n^{-1} \log p}$. Under (NGa $_\beta$) and from Lemma 4, we have

$$P\left(\|\hat{\Gamma}_k - \Gamma_k\|_\infty > \eta\right) \leq C_\beta C^* \frac{p^{2/\beta} n^{1/2}}{(n-k)\eta}.$$

Putting $\eta = p^{2/\beta} \sqrt{n^{-1} \log p}$, we obtain

$$\frac{p^{2/\beta} n^{1/2}}{(n-k)\eta} \sim \frac{p^{2/\beta}}{n^{1/2} \eta} = \frac{1}{\sqrt{\log p}} \rightarrow 0$$

as $n \rightarrow \infty$. □

Remark C From Lemma A.3 (Bickel and Levina 2008b) under the assumption that (ε_t) is Gaussian and (A5), we may deduce that for some $\delta > 0$ and any $|\eta| \leq \delta$,

$$P\left(\|\hat{\Gamma}_k - \Gamma_k\|_\infty > \eta\right) \leq C_1^* \exp(-C_2^*(n-k)\eta^2)$$

for some constants $C_1^*, C_2^* > 0$. Reasoning as in the proof of Lemma 1, we may obtain (6) for $c_n = \sqrt{n^{-1} \log p}$.

Proof of Theorem 1 From the inequality in (Bhattacharjee and Bose (2014b), p. 280), we find that

$$\|\mathbf{B}_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 \leq (2l_n + 1) \|\hat{\Gamma}_k - \Gamma_k\|_\infty + T(\Gamma_k, l_n). \quad (25)$$

From Lemma 3, we have

$$T(\mathbf{0}_k, l_n) = \mathcal{O}(\|\Sigma\|_{(1,1)} l_n^{-\alpha}) \quad (26)$$

for any $l_n \rightarrow \infty$ as $n \rightarrow \infty$. From Lemma 1 (see also Remark C when (Gauss) holds), we get

$$\|\hat{\Gamma}_k - \Gamma_k\|_\infty = \mathcal{O}_P(c_n) \quad (27)$$

for c_n as in Lemma 1. Consequently, due to (25)-(27), we have

$$\|\mathbf{B}_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = \mathcal{O}_P(l_n c_n + \|\Sigma\|_{(1,1)} l_n^{-\alpha}).$$

Putting $l_n = c_n^{-\frac{1}{\alpha+1}}$, we obtain

$$\|\mathbf{B}_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = \mathcal{O}_P(\|\Sigma\|_{(1,1)} l_n^{-\alpha}) = \mathcal{O}_P(\|\Sigma\|_{(1,1)} (c_n)^{\frac{\alpha}{\alpha+1}}) \text{ and (4) holds.}$$

Proof of Corollary 1 Reasoning as in (Bhattacharjee and Bose (2014b), Section 6.2), we have: if A^{-1} exists and $\|A - B\| \leq \|A^{-1}\|^{-1}$, then $\|B^{-1} - A^{-1}\| \leq \frac{\|A^{-1}\|^2 \|A - B\|}{1 - \|A^{-1}\| \|A - B\|}$. Put $A = \Gamma_k$ and $B = B_{l_n}(\hat{\Gamma}_k)$. Since $\|\Gamma_k^{-1}\|_2 = \mathcal{O}(1)$ and from Theorem 1, $\|B_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 = \mathcal{O}_P(a_n)$ as $a_n \rightarrow 0$, it follows that for large n we have $\|A^{-1}\| \|A - B\| \leq 1$. Therefore for some $C > 0$ and large n , we find that

$$\|B_{l_n}^{-1}(\hat{\Gamma}_k) - \Gamma_k^{-1}\|_2 \leq C \|B_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2.$$

Hence, directly from (4), we have (5).

Proof of Proposition 1 From (25), we have for any $\eta > 0$,

$$P\left(\|B_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 > \eta\right) \leq P\left((2l_n + 1)\|\hat{\Gamma}_k - \Gamma_k\|_\infty + T(\Gamma_k, l_n) > \eta\right).$$

Using (21), we obtain

$$P\left(\|B_{l_n}(\hat{\Gamma}_k) - \Gamma_k\|_2 > \eta\right) \leq P\left(\|\hat{\Gamma}_k - \Gamma_k\|_\infty > \frac{\eta - (\tilde{C}_1 \|\Sigma\|_{(1,1)} + \tilde{C}_2) l_n^{-\alpha}}{2l_n + 1}\right).$$

Therefore, from Lemma 4, we deduce (7)-(8).

Proof of Proposition 2 First, we will show (11) by induction. For $j = 1$, under the condition of class $\mathcal{A}$, $\|\Psi_j\|_{(1,1)} = \|\Psi_1\|_{(1,1)} = \|\mathbf{A}_1\|_{(1,1)} \leq C_1 \delta$ and (11) holds for $j = 1$. From (10), $C_1 r < 1$ and the induction assumption that $\|\Psi_i\|_{(1,1)} \leq C_1 \delta^i$ for all $i \leq j$, we have

$$\begin{aligned} \|\Psi_{j+1}\|_{(1,1)} &\leq \sum_{i=1}^{\min(j+1,r)} \|\mathbf{A}_i \Psi_{j+1-i}\|_{(1,1)} \leq \sum_{i=1}^{\min(j+1,r)} \|\mathbf{A}_i\|_{(1,1)} \|\Psi_{j+1-i}\|_{(1,1)} \\ &\leq \sum_{i=1}^{\min(j+1,r)} C_1^2 \delta^i \delta^{j+1-i} = C_1^2 \sum_{i=1}^{\min(j+1,r)} \delta^{j+1} \leq C_1^2 r \delta^{j+1} \leq C_1 \delta^{j+1}. \end{aligned}$$

In a similar manner, we have $\|\Psi'_j\|_{(1,1)} \leq C_1 \delta^j$. Hence, (11) is proved. Next, by induction we will show (12). For $j = 1$, $T(\Psi_j, t) = T(\Psi_1, t) = T(\mathbf{A}_1, t) \leq C_2 \delta t^{-\alpha}$ under the condition of class $\mathcal{A}$. Hence, (12) is satisfied for $j = 1$. By (10) and Lemma 2(ii),

$$\begin{aligned} T(\Psi_{j+1}, t) &\leq \sum_{i=1}^{\min(j+1,r)} T(\mathbf{A}_i \Psi_{j+1-i}, t) \\ &\leq \sum_{i=1}^{\min(j+1,r)} \{\|\mathbf{A}_i\|_{(1,1)} T(\Psi_{j+1-i}, t/2) + \|\Psi_{j+1-i}\|_{(1,1)} T(\mathbf{A}_i, t/2)\}. \end{aligned}$$

Then, by the induction assumption that $T(\Psi_i, t) \leq C_2 i \delta^i t^{-\alpha}$ for all $i \leq j$ and under the conditions of class $\mathcal{A}$ ($C_1 = 1/(2^\alpha r)$) it follows that

$$\begin{aligned} T(\Psi_{j+1}, t) &\leq \sum_{i=1}^{\min(j+1, r)} \{C_1 \delta^i C_2 (j+1-i) \delta^{j+1-i} 2^\alpha t^{-\alpha} + C_1 \delta^{j+1-i} C_2 \delta^i 2^\alpha t^{-\alpha}\} \\ &\leq \sum_{i=1}^{\min(j+1, r)} \{C_1 C_2 j \delta^{j+1} 2^\alpha t^{-\alpha} + C_1 C_2 \delta^{j+1} 2^\alpha t^{-\alpha}\} \\ &\leq r C_1 C_2 2^\alpha \delta^{j+1} t^{-\alpha} (j+1) = C_2 (j+1) \delta^{j+1} t^{-\alpha}. \end{aligned}$$

By a similar reasoning, we obtain $T(\Psi'_j, t) \leq C_2 j \delta^j t^{-\alpha}$. Hence, by induction the proof of (12) is complete.

Lemma 5 For any matrix $\mathbf{A}$,

$$T(\mathbf{A}^j, t) \leq j \|\mathbf{A}\|_{(1,1)}^{j-1} T(\mathbf{A}, t/2^{j-1}) \quad (28)$$

for all $t > 0$ and $j = 1, 2, \dots$

Proof Clearly for $j = 1$ inequality (28) holds. Now, we assume that (28) is true for some j . From the induction assumption and Lemma 2 (ii),

$$\begin{aligned} T(\mathbf{A}^{j+1}, t) &\leq \|\mathbf{A}^j\|_{(1,1)} T(\mathbf{A}, t/2) + \|\mathbf{A}\|_{(1,1)} T(\mathbf{A}^j, t/2) \\ &\leq \|\mathbf{A}\|_{(1,1)}^j T(\mathbf{A}, t/2) + j \|\mathbf{A}\|_{(1,1)}^j T(\mathbf{A}, t/2^j) \\ &\leq \|\mathbf{A}\|_{(1,1)}^j T(\mathbf{A}, t/2^j) + j \|\mathbf{A}\|_{(1,1)}^j T(\mathbf{A}, t/2^j) \\ &\leq (j+1) \|\mathbf{A}\|_{(1,1)}^j T(\mathbf{A}, t/2^j). \end{aligned}$$

By induction, we see that (28) holds for all positive integers j .

Acknowledgements The author would like to thank the editor and the reviewers for useful comments and suggestions which have improved the presentation of the paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Bhattacharjee M, Bose A (2014a) Consistency of large dimensional sample covariance matrix under dependence. Stat Methodol 20:11–26

- Bhattacharjee M, Bose A (2014b) Estimation of autocovariance matrices for infinite dimensional vector linear process. *J Time Ser Anal* 35:262–281
- Bickel PJ, Levina E (2008a) Covariance regularization by thresholding. *Ann Stat* 36:2577–2604
- Bickel PJ, Levina E (2008b) Regularized estimation of large covariance matrices. *Ann Stat* 36:199–227
- Birnbaum A, Nadler B (2012) High dimensional sparse covariance estimation: accurate thresholds for the maximal diagonal entry and for the largest correlation coefficient, Technical Report
- Brockwell P, Davis R (2002) Introduction to time series and forecasting, 2nd edn. Springer Texts in Statistics, New York
- Cai T, Liu WD (2011) Adaptive thresholding for sparse covariance matrix estimation. *J Am Statist Assoc* 106:672–684
- Cai T, Zhang C, Zhou H (2010) Optimal rates of convergence for covariance matrix estimation. *Ann Stat* 38:2118–2144
- Chen X, Xu M, Wu WB (2013) Covariance and precision matrix estimation for high-dimensional time series. *Ann Stat* 41:2994–3021
- Fan J, Fan Y, Lv J (2006) High dimensional covariance matrix estimation using a factor model, Technical report, Princeton Univ
- Friedman J (1989) Regularized discriminant analysis. *J Am Statist Assoc* 84:165–175
- Furrer R, Bengtsson T (2007) Estimation of high-dimensional prior and posteriori covariance matrices in Kalman filter variants. *J Multivar Anal* 98:227–255
- Guo S, Wang Y, Yao Q (2016) High-dimensional and banded vector autoregressions. *Biometrika* 103:889–903
- Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12:55–67
- Huang J, Liu N, Pourahmadi M, Liu L (2006) Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* 98:85–98
- Jentsch C, Politis D (2015) Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Ann Stat* 43:1117–1140
- Johnstone IM, Lu AY (2009) On consistency and sparsity for principal components analysis in high-dimensions. *J Am Statist Assoc* 104:682–693
- Marchenko VA, Pastur LA (1967) Distributions of eigenvalues of some sets of random matrices. *Math USSR-Sb* 1:507–536
- McMurry T, Politis D (2010) Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J Time Ser Anal* 31:471–482
- Petrov VV (1995) Limit theorems of probability theory. Oxford University Press, Oxford
- Pisier G (1983) Some applications of the metric entropy to harmonic analysis, Banach Spaces, Harmonic Ann. Prob. Theory. Lecture Notes in Math. **995**, Springer, New York
- Rothman AJ, Levina E, Zhu J (2009) Generalized thresholding of large covariance matrices. *J Am Statist Assoc* 104:177–186
- Wu WB, Pourahmadi M (2009) Banding sample autocovariance matrices of stationary processes. *Statist Sin* 19:1755–1768
- Xu M, Chen X, Wu WB (2020) Estimation of dynamic networks for high-dimensional nonstationary time series. *Entropy* 22, 55