Biometrika (2016), **00**, 0, *pp.* 1–16
Printed in Great Britain

Supplementary Material: High Dimensional and Banded Vector Autoregressions

BY SHAOJUN GUO

Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, P.R.China
sjguo@ruc.edu.cn

YAZHEN WANG

Department of Statistics, University of Wisconsin, Madison, WI 53706,U.S.A.
yzwang@stat.wisc.edu

AND QIWEI YAO

Department of Statistics, London School of Economics, London WC2A 2AE, U.K.
q.yao@lse.ac.uk

This supplementary material is organized as follows. We provide the detailed proofs of Theorems 1-4, respectively, in Sections 1-4. Section 5 presents Proposition 1 and its proof, showing the consistency of generalized BIC stated in Remark 1 in the paper. In Section 6, we present the consistency of the BIC selector $\widehat{k}$ in a more general setting when $k_0 \to \infty$. Some technical lemmas and their proofs are collected in Section 7. Section 8 presents some additional simulation results.

1. PROOF OF THEOREM 1

Proof. Without loss of generality, we consider the VAR(1) model with $\|A\|_1 \leq \delta < 1$. Our goal is to prove that $\text{pr}(\widehat{k} = k_0) \to 1$, i.e., $\text{pr}(\widehat{k} \neq k_0) \to 0$. If $\widehat{k} \neq k_0$, then either $\widehat{k} > k_0$ or $\widehat{k} < k_0$ holds. Hence it suffices to show that $\text{pr}(\widehat{k} < k_0) \to 0$ and $\text{pr}(\widehat{k} > k_0) \to 0$. Our proof follows the arguments in Wang, et al. (2009).

Consider the first case. Observe that $\text{pr}(\widehat{k} < k_0) \leq \text{pr}(\widehat{k}_i < k_0)$ for some $i \in \{1, \ldots, p\}$ and the event $(\widehat{k}_i < k_0)$ imply $\{\min_{k<k_0} \text{BIC}_i(k) < \text{BIC}_i(k_0)\}$. To prove $\text{pr}(\widehat{k} < k_0) \to 0$, we only need to show that

$$\text{pr}\{\min_{k<k_0} \text{BIC}_i(k) < \text{BIC}_i(k_0)\} \to 0$$

for some i. Suppose that we have shown that there exists a constant $\eta > 0$ and an event $\mathcal{A}_n$ such that $\text{pr}(\mathcal{A}_n) \to 1$ as $n \to \infty$ and on the event $\mathcal{A}_n$,

$$\text{RSS}_i(k) - \text{RSS}_i(k_0) \geq \eta \text{RSS}_i(k_0)(a_{i,i-k_0}^2 + a_{i,i+k_0}^2), \tag{1}$$

for sufficiently large n, where $a_{j,k}$ is the (j,k)-element of A_1. On the event $\mathcal{A}_n$ with large n, $\log \text{RSS}_i(k) - \log \text{RSS}_i(k_0) \geq \log\{1 + \eta(a_{i,i-k_0}^2 + a_{i,i+k_0}^2)\}$. Note that $\log(1+x) \geq \min(0.5x, \log 2)$ for any $x > 0$. Consequently, with probability tending to one, $\log \text{RSS}_i(k) - \log \text{RSS}_i(k_0)$ can be further bounded below by $\min\{0.5\eta(a_{i,i-k_0}^2 + a_{i,i+k_0}^2), \log 2\}$. Condition

3 implies that for some $i^* \in \{1,\dots,p\}$, $a^2_{i^*,i^*-k_0} + a^2_{i^*,i^*+k_0} \gg C_n \log p/n$ as $n \to \infty$. Hence, it follows that, with probability tending to 1,

$$\min_{k<k_0} \mathrm{BIC}_{i^*}(k) - \mathrm{BIC}_{i^*}(k_0) > \min\{0.5\eta(a^2_{i^*,i^*-k_0} + a^2_{i^*,i^*+k_0}), \log 2\} - C_n k_0 n^{-1} \log(p \vee n) > 0,$$

where $p \vee n = \max(p,n)$. Hence, $\mathrm{pr}\{\min_{k<k_0} \mathrm{BIC}_{i^*}(k) < \mathrm{BIC}_{i^*}(k_0)\} \to 0$ and thus $\mathrm{pr}(\widehat{k} < k_0) \to 0$.

Let us prove (1). For $k < k_0$, denote $H_{i,k} = X_{i,k}\big(X^{\mathrm{T}}_{i,k}X_{i,k}\big)^{-1}X^{\mathrm{T}}_{i,k}$, $X_{i,k_0} = (S^{(1)}_{i,k}, X_{i,k}, S^{(2)}_{i,k})$ and $\beta_{i,k_0} = (b^{\mathrm{T}}_{i,1}, \beta^{\mathrm{T}}_{i,k}, b^{\mathrm{T}}_{i,2})^{\mathrm{T}}$, where $X_{i,k}$ is defined similar to (4) in Section 2.2 except that k_0 is replaced by k. Then $\mathrm{RSS}_i(k) = y^{\mathrm{T}}_{(i)}\big(I_{n-1} - H_{i,k}\big)y_{(i)}$, and by Lemma 5 (ii) or Lemma 6 (ii), we have

$$\mathrm{RSS}_i(k) - \mathrm{RSS}_i(k_0) = (b^{\mathrm{T}}_{i,1}, b^{\mathrm{T}}_{i,2})(S^{(1)}_{i,k}, S^{(2)}_{i,k})^{\mathrm{T}}(I_{n-1} - H_{i,k})(S^{(1)}_{i,k}, S^{(2)}_{i,k}) \begin{pmatrix} b_{i,1} \\ b_{i,2} \end{pmatrix} + o_P(1).$$

From Lemma 5 (i) or Lemma 6 (i) and Lemma 7, there exists a small constant $\eta > 0$ such that, with probability tending to one,

$$\lambda_{\min}\big\{(S^{(1)}_{i,k}, S^{(2)}_{i,k})^{\mathrm{T}}(I - H_{i,k})(S^{(1)}_{i,k}, S^{(2)}_{i,k})\big\} > \eta(1+\eta)n\sigma_i^2,$$

and $\mathrm{RSS}_i(k_0) \le n\sigma_i^2(1+\eta)$. Therefore, (1) follows.

Now we turn to the overfitting case, i.e., $\mathrm{pr}(\widehat{k} > k_0) \to 0$. For $k > k_0$, set $X_{i,k} = (S^{(1)}_{i,k}, X_{i,k_0}, S^{(2)}_{i,k})$, $\beta_{i,k} = (0^{\mathrm{T}}, \beta^{\mathrm{T}}_{i,k_0}, 0^{\mathrm{T}})^{\mathrm{T}}$, $S_{i,k} = (S^{(1)}_{i,k}, S^{(2)}_{i,k})$, and $\widetilde{S}_{i,k} = \big(I_{n-1} - H_{i,k_0}\big)S_{i,k}$. Let η be an arbitrary but fixed positive constant and define

$$\mathcal{B}_n = \Big\{ \inf_{k_0 \le k \le K} \inf_{1 \le i \le p} \frac{\mathrm{RSS}_i(k)}{n\sigma_i^2} > (1-\eta) \Big\},$$

$$\mathcal{C}_n = \bigcup_{\substack{1\le i \le p \\ k_0 \le k \le K}} \Big\{ \lambda^{-1}_{\min}(n^{-1}\widetilde{S}^{\mathrm{T}}_{i,k}\widetilde{S}_{i,k}) < \kappa_1^{-1}(1+\eta), \sup_{1 \le j \le k-k_0} \Big|\big(n^{-1}S^T_{i,k}S_{i,k}\big)_{jj}\Big| < \kappa_2(1+\eta)\Big\}.$$

We first give an upper bound on $\mathrm{RSS}_i(k_0) - \mathrm{RSS}_i(k)$ for $k > k_0$. For each i, $\mathrm{RSS}_i(k)$ can be rewritten as

$$\mathrm{RSS}_i(k) = \inf_b \|y_{(i)} - X_{i,k}b\|^2 = \inf_{b_1,b_2} \|y_{(i)} - X_{i,k_0}b_1 - S_i b_2\|^2.$$

It can be verified that $\mathrm{RSS}_i(k_0) = \|(I_{n-1} - H_{i,k_0})y_{(i)}\|^2$ and

$$\mathrm{RSS}_i(k) = \mathrm{RSS}_i(k_0) - \|\widetilde{S}^{(k)}_i \widehat{b}_2\|^2,$$

where $\widehat{b}_2 = \Big(\widetilde{S}^T_{i,k}\widetilde{S}_{i,k}\Big)^{-1}\widetilde{S}^T_{i,k}e_{(i)}$. Then on the event $\mathcal{C}_n$ we have

$$\begin{aligned} \mathrm{RSS}_i(k_0) - \mathrm{RSS}_i(k) &= e^{\mathrm{T}}_{(i)}\widetilde{S}_{i,k}(\widetilde{S}^{\mathrm{T}}_{i,k}\widetilde{S}_{i,k})^{-1}\widetilde{S}^{\mathrm{T}}_{i,k}e_{(i)} \\ &\le \kappa_1^{-1}(1+\eta)|\tau_i(k) - \tau_i(k_0)| \sup_{j,k\le p} \big|n^{-1/2}e^{\mathrm{T}}_{(j)}(I_{n-1} - H_{i,k_0})x_{(k)}\big|^2. \end{aligned}$$

Define

$$\mathcal{D}_n = \Big\{ \sup_{j,k \le p} \big|n^{-1/2}e^{\mathrm{T}}_{(j)}(I_{n-1} - H_{i,k_0})x_{(k)}\big|^2 \sigma_i^{-2} < \frac{\kappa_1(1-\eta)}{(1+\eta)} C_n \log(p \vee n) \Big\}.$$

On the set $\mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_n$, for all k with $k_0 \le k \le K$,

$$\begin{aligned}\mathrm{RSS}_i(k_0) - \mathrm{RSS}_i(k) &< \sigma_i^2(1-\eta)|\tau_i(k) - \tau_i(k_0)|C_n \log(p \vee n) \\ &< \mathrm{RSS}_i(k) C_n |\tau_i(k) - \tau_i(k_0)| n^{-1} \log(p \vee n).\end{aligned}$$

Note that $\log(1+x) \le x$ for any $x > 0$. Hence, for all k with $k_0 < k \le K$, on the set $\mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_n$,

$$\begin{aligned}\mathrm{BIC}_i(k) - \mathrm{BIC}_i(k_0) &= \log \mathrm{RSS}_i(k) - \log \mathrm{RSS}_i(k_0) + C_n |\tau_i(k) - \tau_i(k_0)| n^{-1} \log(p \vee n) \\ &\ge -\{\mathrm{RSS}_i(k_0) - \mathrm{RSS}_i(k)\}\{\mathrm{RSS}_i(k)\}^{-1} \\ &\quad + C_n |\tau_i(k) - \tau_i(k_0)| n^{-1} \log(p \vee n) > 0,\end{aligned}$$

which indicates that over the set $\mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_n$, we have that $\widehat{k} \le k_0$. To prove that $\mathrm{pr}(\widehat{k} > k_0) \to 0$, it suffices to show that $\mathrm{pr}\left\{(\mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_n)^c\right\} \to 0$. In fact, it follows from Lemma 7 and Lemma 5 or 6 (i) that $\mathrm{pr}\left(\mathcal{B}_n^c\right) \to 0$ and $\mathrm{pr}\left(\mathcal{C}_n^c\right) \to 0$. It remains to show that $\mathrm{pr}\left(\mathcal{D}_n^c\right) \to 0$. Let $\Sigma_{i,k} = n^{-1} E\left(X_{i,k}^{\mathrm{T}} X_{i,k}\right)$, $\widehat{\Sigma}_{i,k} = n^{-1} X_{i,k}^{\mathrm{T}} X_{i,k}$, where $E(X)$ denotes the expectation of X. Set $\widetilde{H}_{i,k} = n^{-1} X_{i,k} \Sigma_{i,k}^{-1} X_{i,k}^{\mathrm{T}}$, and $\widetilde{x}_{(k)} = (I_{n-1} - \widetilde{H}_{i,k}) x_{(k)}$. On the event $\mathcal{C}_n$, we obtain that

$$\begin{aligned}&\sup_{j,k \le p} \left| e_{(j)}^{\mathrm{T}} (I_{n-1} - H_{i,k_0}) x_{(k)} \right| \\ &\le \sup_{j,k \le p} \left| e_{(j)}^{\mathrm{T}} \widetilde{x}_{(k)} \right| + \sup_{j,k \le p} \left| e_{(j)}^{\mathrm{T}} (H_{i,k_0} - \widetilde{H}_{i,k_0}) x_{(k)} \right| \\ &\le \sup_{j,k \le p} \left| e_{(j)}^{\mathrm{T}} \widetilde{x}_{(k)} \right| + n^{-1} \sup_{j,k \le p} \| e_{(j)}^{\mathrm{T}} X_{i,k_0} \|_2 \| \Sigma_{i,k_0}^{-1} \|_2 \| \widehat{\Sigma}_{i,k_0}^{-1} \|_2 \| \widehat{\Sigma}_{i,k_0} - \Sigma_{i,k_0} \|_2 \| X_{i,k_0}^{\mathrm{T}} x_{(k)} \|_2 \\ &\le \sup_{j,k \le p} \left| e_{(j)}^{\mathrm{T}} \widetilde{x}_{(k)} \right| + k_0 \kappa_1^{-2} k_2 (1+\eta)^2 \sup_{j,k \le p} \left| e_{(j)}^{\mathrm{T}} x_{(k)} \right| \cdot \| \widehat{\Sigma}_{i,k_0} - \Sigma_{i,k_0} \|_2,\end{aligned}$$

where $\sup_{1 \le k \le p} (n^{-1} x_{(k)} x_{(k)}^T) \le \kappa_2 (1+\eta)$ is used in the above inequality. Hence, it follows from Lemmas 5 and 6, together with Condition 3, that $\mathrm{pr}\left(\mathcal{D}_n^c\right) \to 0$ as $n \to \infty$. This completes the proof. □

2. PROOF OF THEOREM 2

Proof. Since the VAR(d) model can be formulated as a VAR(1) model, without loss of generality, we consider the VAR(1) case only. With probability tending to one, $\widehat{k} = k_0$, and thus it suffices to consider the set $\mathcal{A}_n = \{\widehat{k} = k_0\}$. Over the set $\mathcal{A}_n$, for each i,

$$\widehat{\beta}_i - \beta_i = \left(X_i^{\mathrm{T}} X_i\right)^{-1} X_i^{\mathrm{T}} e_{(i)}. \tag{2}$$

For each i, the law of large numbers for the stationary process case yields that $n^{-1} X_i^{\mathrm{T}} X_i$ converges to a positive matrix almost surely, and furthermore, with probability tending to one, $\lambda_{\min}\left(n^{-1} X_i^{\mathrm{T}} X_i\right)$ is bounded away from zero. As a matter of fact, if we define

$$\mathcal{B}_n = \bigcap_{1 \le i \le p} \left\{ \lambda_{\min}\left(n^{-1} X_i^{\mathrm{T}} X_i\right) > \kappa_1 (1-\eta) \right\}$$

with a small constant $\eta \in (0,1)$, then it follows from by Lemma 5 or Lemma 6 under different moment conditions that $P\{\mathcal{B}_n\} \to 1$ as $n \to \infty$. Hence, over the event $\mathcal{A}_n \cap \mathcal{B}_n$,

$$\left\| \widehat{\beta}_i - \beta_i \right\|_2^2 \le \kappa_1^{-2} (1-\eta)^{-2} n^{-2} \left\| e_{(i)}^{\mathrm{T}} X_i \right\|_2^2 = C_1 n^{-2} \left\| e_{(i)}^{\mathrm{T}} X_i \right\|_2^2,$$

where $C_1 = \kappa_1^{-2}(1-\eta)^{-2} > 0$. It is not hard to see from Lemma 5(ii) or Lemma 6(ii) that, for all $1 \le i \le p$, $n^{-1}E\|X_i^{\mathrm{T}} e_{(i)}\|_2^2 \le C_2$ with some constant $C_2 > 0$. Therefore, for a large positive constant C, we obtain that

$$\begin{aligned}\mathrm{pr}\left(\left\|\widehat{A}_1 - A_1\right\|_F^2 > Cn^{-1}p\right) &= \mathrm{pr}\left(\left\|\widehat{A}_1 - A_1\right\|_F^2 > Cn^{-1}p, \mathcal{A}_n \cap \mathcal{B}_n\right) + \mathrm{pr}\left(\mathcal{A}_n \cap \mathcal{B}_n\right) \\ &= (Cp)^{-1} n (C_1 n^{-2}) E\left(\sum_{i=1}^{p} \|X_i^{\mathrm{T}} e_{(i)}\|_2^2\right) + \mathrm{pr}\left(\mathcal{A}_n \cap \mathcal{B}_n\right) \\ &= C_1 C_2 C^{-1} + o(1).\end{aligned}$$

We establish the convergence rate of $\|\widehat{A}_1 - A_1\|_F$ by taking a sufficiently large C.

Now we derive the convergence rate of $\|\widehat{A}_1 - A_1\|_2$. For any matrix B, $\|B\|_2^2 \le \|B\|_1 \|B\|_\infty$. Hence, on the event $\mathcal{A}_n$,

$$\|\widehat{A}_1 - A_1\|_2 \le \sqrt{\|\widehat{A}_1 - A_1\|_1}\sqrt{\|\widehat{A}_1 - A_1\|_\infty} \le (2k_0+1) \sup_{i \le p, j \le \tau_i} \left|\widehat{\beta}_{ij} - \beta_{ij}\right|,$$

where $\widehat{\beta}_{ij}$ and β_{ij} are the j-th element of $\widehat{\beta}_i$ and β_i, respectively. Observe from (2) that

$$\sup_{i \le p, j \le \tau_i} \left|\widehat{\beta}_{ij} - \beta_{ij}\right| = \kappa_1^{-1}(1-\eta)^{-1}(2k_0+1)\big(\sup_{i \le p, j \le \tau_i} |e_{(i)}^{\mathrm{T}} x_{(j)}|\big), i = 1, \ldots, p.$$

Hence, using Lemma 5(ii) or Lemma 6(ii), we have

$$\sup_{i \le p, j \le \tau_i} \left|\widehat{\beta}_{ij} - \beta_{ij}\right| = O_P\left\{\left(n^{-1}\log p\right)^{1/2}\right\},$$

which shows that

$$\|\widehat{A}_1 - A_1\|_2 = O_P\left\{\left(n^{-1}\log p\right)^{1/2}\right\}.$$

The proof is completed. □

3. PROOF OF THEOREM 3

Proof. The covariance matrix Σ_0 can be expressed as

$$\Sigma_0 = \Sigma_\varepsilon + \sum_{j=1}^{\infty} B_j, \;\; B_j = J\widetilde{A}^j J^{\mathrm{T}} \Sigma_\varepsilon J (\widetilde{A}^{\mathrm{T}})^j J^{\mathrm{T}}, \;\; j \ge 1,$$

where $J = \left(I_{p\times p}, 0_{p\times(d-1)p}\right)$. Let $\Phi_j = J\widetilde{A}^j J^{\mathrm{T}}$, $j \ge 1$. By the companion matrix $\widetilde{A}$, we can show that $\Phi_0 = I_p$ and $\Phi_j = \sum_{k=1}^{\min(j,d)} \Phi_{j-k} A_k$, $j \ge 1$. It is easy to see that for two banded matrices F and G with bandwidths $2r_1+1$ and $2r_2+1$, respectively, the product matrix FG is also banded and its bandwidth is at most $2(r_1+r_2)+1$. Therefore, it can be verified that Φ_j is banded with bandwidth at most $2jk_0+1$ and then B_j is also banded with its bandwidth at most $2(2jk_0+s_0)+1$ for $j \ge 1$. Take $\Sigma_0^{(r)} = \Sigma_\varepsilon + \sum_{j=1}^{r} B_j$, which is banded with the bandwidth at most $2(2rk_0+s_0)+1$, and $\Sigma_0 - \Sigma_0^{(r)} = \sum_{j=r+1}^{\infty} B_j$. Note that for any $j \ge 1$, $\|B_j\|_2 \le \|\Sigma_\varepsilon\|_2 \|\widetilde{A}^{2j}\|_2 \le C\delta^{2j}$ for some $C > 0$. Write $C_1 = C\|\Sigma_\varepsilon\|_2\left(1-\delta^2\right)^{-1}$. It follows that

$$\|\Sigma_0 - \Sigma_0^{(r)}\|_2 \leq \sum_{j=r+1}^{\infty} \|B_j\|_2 \leq C\|\Sigma_\varepsilon\|_2 \left(1-\delta^2\right)^{-1} \delta^{2(r+1)} = C_1 \delta^{2(r+1)}.$$

By using the inequality $\|B_j\|_1 \leq \{2(2jk_0 + s_0) + 1\}\|B_j\|_2 \leq C(2j+1)\delta^{2j}$ for some $C > 0$, we obtain

$$\|\Sigma_0 - \Sigma_0^{(r)}\|_1 \leq C_2 r \delta^{2(r+1)}.$$

Other inequalities can be proved analogously. The proof is complete. □

4. PROOF OF THEOREM 4

Proof. Now we prove the convergence rate of $\|\widehat{\Sigma}_{n,0}^{(r_n)} - \Sigma_0\|_2$. First, $\|\widehat{\Sigma}_{n,0}^{(r_n)} - \Sigma_0\|_2$ can be bounded above by

$$\|\widehat{\Sigma}_{n,0}^{(r_n)} - \Sigma_0^{(r_n)}\|_2 + \|\Sigma_0^{(r_n)} - \Sigma_0\|_2 = R_{n1} + R_{n2}.$$

Similar to Theorem 2, $R_{n1} \leq (4r_n k_0 + 2s_0 + 1)\sup_{j,k\leq p} |\widehat{\Sigma}_{jk} - \Sigma_{jk}|$. From Lemma 5(i) or Lemma 6(i), we obtain that

$$R_{n1} = O_P\{r_n \left(n^{-1}\log p\right)^{1/2}\}.$$

From Theorem 3, $R_{n2} \leq O(\delta^{2(r_n+1)})$. Note that $r_n = C\log\{n\log^{-1}(p)\}$ with $C > (-4\log\delta)^{-1}$. Combining these results, it follows that

$$\|\widehat{\Sigma}_{n,0}^{(r_n)} - \Sigma_0\|_2 = O_P\Big\{r_n \left(n^{-1}\log p\right)^{1/2} + \delta^{2(r_n+1)}\Big\} = O_P\Big[\log\{n\log^{-1}(p)\}\left(n^{-1}\log p\right)^{1/2}\Big].$$

The proofs of other results are similar and omitted. □

5. PROPOSITION 1 AND ITS PROOF

PROPOSITION 1. *Under Conditions 1-4, we prove that* $\mathrm{pr}(\widehat{k} = k_0, \widehat{d} = d) \to 1$ *as* $n, p \to \infty$.

Proof of Proposition 1. Our primary goal is to prove that $\mathrm{pr}(\widehat{k} = k_0, \widehat{d} = d) \to 1$, i.e., $\mathrm{pr}\{(\widehat{k} \neq k_0) \cup (\widehat{d} \neq d)\} \to 0$. Note that

$$\mathrm{pr}\{(\widehat{k} \neq k_0) \cup (\widehat{d} \neq d)\} \leq \mathrm{pr}(\widehat{k} < k_0) + \mathrm{pr}(\widehat{d} < d) + \mathrm{pr}(\widehat{k} > k_0, \widehat{d} > d).$$

We observe that both events $\{\widehat{k} < k_0\}$ and $\{\widehat{d} < d\}$ correspond to the underfitting case, where some important variables are missed in the estimated model. Hence, following the proofs of Theorem 1, we can show $\mathrm{pr}(\widehat{k} < k_0) + \mathrm{pr}(\widehat{d} < d) \to 0$.

It remains to prove that $\mathrm{pr}(\widehat{k} > k_0, \widehat{d} > d) \to 0$. First look at the event $\mathcal{A} = \big\{\widehat{k} > k_0, \widehat{d} > d\big\}$. Define $\mathcal{A}_1 = \underset{i\leq p}{\cup}\big\{\widehat{k}_i \geq k_0, \widehat{d}_i > d\big\}$, $\mathcal{A}_2 = \underset{i\leq p}{\cup}\big\{\widehat{k}_i < k_0, \widehat{d}_i > d\big\}$, $\mathcal{A}_3 = \underset{i\leq p}{\cup}\big\{\widehat{k}_i > k_0, \widehat{d}_i \geq d\big\}$, and $\mathcal{A}_4 = \underset{i\leq p}{\cup}\big\{\widehat{k}_i > k_0, \widehat{d}_i < d\big\}$. Then $\mathcal{A} \subset \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, which implies that it suffices to show $\mathrm{pr}(\mathcal{A}_k) \to 0$ for each $k = 1, \ldots, 4$. Observe that both events $\mathcal{A}_1$ and $\mathcal{A}_3$ correspond to the overfitting case, where all important variables as well as some unimportant variables are selected by the estimated model. Hence, following the proofs of Theorem 1, we can show $\mathrm{pr}(\mathcal{A}_1) + \mathrm{pr}(\mathcal{A}_3) \to 0$.

Now we are going to prove pr$(\mathcal{A}_2) \to 0$ as $n, p \to \infty$. For each i, $\{\widehat{k}_i < k_0, \widehat{d}_i > d\}$ means $\min_{k<k_0,\ell>d} \widetilde{\text{BIC}}_i(k,\ell) < \widetilde{\text{BIC}}_i(k_0,d)$. Hence, we only need to show, with probablity tending to one,

$$\min_{i\le p} \min_{k<k_0, d<\ell\le L} \Big\{\widetilde{\text{BIC}}_i(k,\ell) - \widetilde{\text{BIC}}_i(k_0,d)\Big\} > 0. \tag{3}$$

Suppose that we have shown that there exists a constant $\eta > 0$ and an event $\mathcal{G}_n$ such that pr$(\mathcal{G}_n) \to 1$ as $n \to \infty$ and on the event $\mathcal{G}_n$,

$$\min_{i\le p} \big\{\text{RSS}_i(k,\ell) - \text{RSS}_i(k_0,d) - \eta \text{RSS}_i(k_0,d)\Delta_i \ge 0, \tag{4}$$

for each $k < k_0$, $d < \ell < L$ and sufficiently large n, where $\Delta_i = \sum_{j=1}^d \big\{(a_{i,i-k_0}^{(j)})^2 + (a_{i,i+k_0}^{(j)})^2\big\}$. As a result, on the event $\mathcal{G}_n$ with large n, $\min_{i\le p}\big\{\log \text{RSS}_i(k,\ell) - \log \text{RSS}_i(k_0,d) - \log(1+\eta\Delta_i)\big\} \ge 0$. Note that $\log(1+x) \ge \min\{0.5x, \log 2\}$ for any $x > 0$. Then, with probability tending to one, $\log \text{RSS}_i(k) - \log \text{RSS}_i(k_0)$ can be further bounded below by $\min(0.5\Delta_i, \log 2)$. Condition 3 implies that $\min_{i\le p} \Delta_i \gg C_n n^{-1}\log(p\vee n)$ as $n\to\infty$. Hence, it follows that, with probability tending to 1,

$$\widetilde{\text{BIC}}_i(k,\ell) - \widetilde{\text{BIC}}_i(k_0,d) \ge \min(0.5\eta\Delta_i, \log 2) - C_n\tau_i(k_0,d)n^{-1}\log(p\vee n) > 0 \tag{5}$$

uniformly for all $k < k_0, d < \ell \le L$ and $i = 1,\ldots,p$. Hence, pr$(\mathcal{A}_2) \to 0$ as $n,p\to\infty$.

Let us turn to prove (4). For $k < k_0$ and $d < \ell \le L$, denote $H_{i,k,\ell} = X_{i,k,\ell}\big(X_{i,k,\ell}^{\mathrm{T}} X_{i,k,\ell}\big)^{-1} X_{i,k,\ell}^{\mathrm{T}}$, where $X_{i,k,\ell}$ is defined as in section 2.2 but replaced k_0 and d by k and ℓ. Then $\text{RSS}_i(k,\ell) = y_{(i)}^{\mathrm{T}}\big(I_{n-1} - H_{i,k,\ell}\big)y_{(i)}$. In fact, $X_{i,k_0,\ell}$ can be rewritten as $X_{i,k_0,\ell} = (S_{i,k,1}^{(1)}, X_{i,k}^{(1)}, S_{i,k,2}^{(2)}, \ldots, S_{i,k,1}^{(\ell)}, X_{i,k}^{(\ell)}, S_{i,k,2}^{(\ell)})$ and, similarly, $\beta_{i,k_0,\ell}^T = \big(b_{i,1}^{(1)}, \beta_{i,k}^{(1)}, b_{i,2}^{(1)}, \ldots, b_{i,1}^{(\ell)}, \beta_{i,k}^{(\ell)}, b_{i,2}^{(\ell)}\big)$. Let $S_{i,\ell} = (S_{i,k,1}^{(1)}, S_{i,k,2}^{(2)}, \ldots, S_{i,k,1}^{(\ell)}, S_{i,k,2}^{(\ell)})$ and $b_{i,\ell}^T = \big(b_{i,1}^{(1)}, b_{i,2}^{(1)}, \ldots, b_{i,1}^{(\ell)}, b_{i,2}^{(\ell)}\big)$. As a result, by Lemma 5 (ii) or Lemma 6 (ii), we have

$$\max_{i\le p} \big|\text{RSS}_i(k,\ell) - \text{RSS}_i(k_0,d) - b_{i,\ell}^{\mathrm{T}} S_{i,\ell}^{\mathrm{T}}(I_{n-1} - H_{i,k,\ell})S_{i,\ell}b_{i,\ell}\big| = o_P(1).$$

From Lemma 5 (i) or Lemma 6 (i) and Lemma 7, there exists a small constant $\eta > 0$ such that, with probability tending to one,

$$\lambda_{\min}\big\{S_{i,\ell}^{\mathrm{T}}(I - H_{i,k,\ell})S_{i,\ell}\big\} > \eta(1+\eta)n\sigma_i^2,$$

and $\text{RSS}_i(k_0,d) \le n\sigma_i^2(1+\eta)$. Note that $b_{i,\ell}^T b_{i,\ell} \ge \sum_{j=1}^d \big\{(a_{i,i-k_0}^{(j)})^2 + (a_{i,i+k_0}^{(j)})^2\big\}$. Therefore, (4) follows.

In a similar manner, pr$(\mathcal{A}_4) \to 0$ can be proved. The proof is completed. □

6. PROPOSITION 2 AND ITS PROOF

PROPOSITION 2. *Under Conditions 1' and 2–4, pr*$(\widehat{k} = k_0) \to 1$ *as* $n, p \to \infty$*, provided* $k_0 \ll C_n^{-1} n/\log(p\vee n)$.

Proof of Proposition 2. First, we can prove the conclusions of Lemma 5, 6 and 7 under the Conditions 1' and (2)–(4). For instance, in the proof of Lemma 5, we bound $\|A_1^l\|_\infty$ in (10) by $\|A_1^l\|_\infty \le C\delta^l$ under Condition 1'. Similarly, the inequalities (11) and (12) in the proof of Lemma 6 can be bounded in a similar way. Then, following the proof of Theorem 1, we can prove the consistency of BIC selector $\widehat{k}$ in the general setting $k_0 \to \infty$. □

7. Seven technical lemmas and their proofs

We first adopt the asymptotic theories using the functional dependent measure of Wu (2005). Assume that z_i is a stationary process of the form $z_i = g(\mathcal{F}_i)$, where $g(\cdot)$ is a measurable function and $\mathcal{F}_i = (\ldots, e_{-1}, e_0, \ldots, e_i)$ with independent and identically distributed random variables $\{e_i; i = 0, \pm 1, \ldots\}$. Wu (2005) defined the functional dependent measure in terms of how the outputs are affected by the inputs. To be specific, denote $\|z\|_q = \{E(|z|^q)\}^{1/q}$ with $q \geq 1$ for a random variable z. The physical or functional dependent measure is defined as

$$\theta_{i,q} = \|z_i - z_i^*\|_q = \|g(\mathcal{F}_i) - g(\mathcal{F}_i^*)\|_q,$$

where $z_i^* = g(\mathcal{F}_i^*)$ is the coupled process of z_i, $\mathcal{F}_i^* = (\ldots, e_{-1}, e_0^*, \ldots, e_i)$ with $\{e_0^*, e_0\}$ being independent and identically distributed. Intuitively, $\theta_{i,q}$ measures the dependency of z_i on e_0 while keeping all other innovations unchanged.

Lemma 1. *(Theorem 2 (ii) of Liu, Xiao and Wu (2013)). Let* $S_n = n^{-1/2}\sum_{i=1}^n z_i$ *and* $\Theta_{m,q} = \sum_{i=m}^\infty \theta_{i,q}$. *Assume that for each* m, $\Theta_{m,q} = O(m^{-\alpha})$ *with* $\alpha > 1/2 - 1/q$ *and* $q > 2$. *Then there exist positive constants* C_1, C_2 *and* C_3 *which only depend on q such that for all* $x > 0$,

$$pr\big(|S_n| \geq x\big) \leq \frac{C_1 \Theta_{0,q}^q n}{(n^{1/2}x)^q} + C_3 \exp\big(C_2 \Theta_{0,q}^{-1} x^2\big).$$

To prove the limit theory for the sub-exponential tail case under Condition 4(ii), we shall use Lemmas 2–4.

Lemma 2. *Suppose that* X *is a random variable. Then,* $E\{\exp(t_0|X|^v)\} < \infty$ *for some* $0 < v \leq 2$ *and* $t_0 > 0$ *if and only if*

$$\limsup_{q\to\infty} q^{-1/v}\|X\|_q < \infty.$$

Proof of Lemma 2. Assume that $\zeta = E\{\exp(t_0|X|^v)\} < \infty$. Then, for any $q \geq 2$,

$$\begin{aligned} E(|X|^q) &= q\int_0^\infty x^{q-1}\mathrm{pr}(|X| > x)dx \\ &\leq \zeta q v^{-1} t_0^{-q/v} \int_0^\infty x^{q/v-1}\exp\big(-x\big)dx = \zeta q v^{-1} t_0^{-q/v}\Gamma\left(\frac{q}{v}\right), \end{aligned} \qquad \square$$

where $\Gamma(\cdot)$ is the Gamma function. By Stirling's formula,

$$\lim_{x\to\infty} \Gamma(x+1)\Big\{ (2\pi x)^{1/2}\left(\frac{x}{e}\right)^x \Big\}^{-1} = 1,$$

we obtain that for all sufficiently large q,

$$\|X\|_q \leq \left(\zeta q v^{-1} t_0^{-q/v}\right)^{1/q} q^{1/(2q)}\left(\frac{q}{v} - 1\right)^{1/v - 1/q} \leq C q^{1/v},$$

where C is a constant depending on ζ, v and t_0 only. This implies that

$$\limsup_{q\to\infty} q^{-1/v}\|X\|_q < \infty.$$

Conversely, assume that $\limsup_{q\to\infty} q^{-1/v}\|X\|_q < \infty$. Then, there exists a positive constant $\phi_0 > 0$ such that, $\|X\|_q \leq \phi_0 q^{1/v}$ for all $q \geq 2$. Note that $\exp(x) = 1 + \sum_{k\geq 1}(k!)^{-1}x^k$. To prove that $E\{\exp(t_0|X|^v)\} < \infty$ for some $t_0 > 0$, we only need to show that there exist positive constants t_0 and k_0 such that

$$\sum_{k \geq k_0} \frac{t_0^k \|X\|_{vk}^{vk}}{k!} < \infty.$$

By Stirling's formula, there exists a large integer k_0 such that for $k \geq k_0$,

$$\Gamma(k+1) = k! \geq (\pi k)^{1/2} \left(\frac{k}{e}\right)^k.$$

With such k_0 and $t_0 = (2\phi_0^v ve)^{-1}$, we have

$$\sum_{k \geq k_0} \frac{t_0^k \|X\|_{vk}^{vk}}{k!} \leq \sum_{k \geq k_0} \frac{(t_0 \phi_0^v ve)^k k^k}{(\pi k)^{1/2} k^k} \leq \sum_{k \geq k_0} 2^{-k} < \infty.$$

LEMMA 3. *Suppose that* $\{X_1, \ldots, X_n\}$ *are independent random variables and* $\sup_{i \leq n} E\{\exp(t_0|X_i|^\alpha)\} \leq \zeta$ *for some positive constants* α, t_0 *and* ζ *with* $0 < \alpha \leq 1$. *Then there exist positive constants* $C_j > 0$ $(j = 1, \ldots, 4)$ *which depend only on* α, t_0 *and* ζ *such that for any* $x > 0$ *and all* n*, the following concentration inequality holds:*

$$pr\Big[\Big|\sum_{i=1}^n \{X_i - E(X_i)\}\Big| > 3x\Big] \leq C_1 \exp\left(-\frac{x^2}{C_2 n + C_3 n^{\frac{1-\alpha}{2-\alpha}} x}\right) + C_1 \exp\left(-\frac{x^{2\alpha}}{C_2 n^{\frac{\alpha}{2-\alpha}} + C_3 x^\alpha}\right) + nC_1 \exp\left(-C_4 x^\alpha\right). \quad (6)$$

In particular, if $\alpha = 1$*, then*

$$pr\Big[\Big|\sum_{i=1}^n \{X_i - E(X_i)\}\Big| > 3x\Big] \leq C_1 \exp\left(-\frac{x^2}{C_2 n + C_3 x}\right) + C_1 n \exp\left(-C_4 x\right)$$

for any $x > 0$ *and* n.

Proof of Lemma 3. For the case of $\alpha = 1$, (6) can be proved by Bernstein's inequality directly. So here we consider the case of $0 < \alpha < 1$ only. Let ξ_{n1} and ξ_{n2} be two constants with $0 < \xi_{n1} < \xi_{n2}$, which depend on n and will be defined below. Let $\widetilde{X}_{i1} = X_i I(|X_i| \leq \xi_{n1})$, $\widetilde{X}_{i2} = X_i I(\xi_{n1} \leq |X_i| \leq \xi_{n2})$ and $\widetilde{X}_{i3} = X_i I(|X_i| > \xi_{n2})$. Then $X_i = \widetilde{X}_{i1} - E(\widetilde{X}_{i1}) + \widetilde{X}_{i2} - E(\widetilde{X}_{i2}) + \widetilde{X}_{i3} - E(\widetilde{X}_{i3})$, and hence

$$\text{pr}\Big[\Big|\sum_{i=1}^n \{X_i - E(X_i)\}\Big| > 3x\Big] \leq \sum_{k=1}^3 \text{pr}\Big[\Big|\sum_{i=1}^n \{\widetilde{X}_{ik} - E(\widetilde{X}_{ik})\}\Big| > x\Big].$$

In the following, we will give an upper bound on each term separately.

Now consider the first term. Let σ^2 be a finite constant such that $\sup_{i \leq n} E|X_i|^2 \leq \sigma^2$. Note that $|\widetilde{X}_{i1}| \leq \xi_{n1}$ and $E\widetilde{X}_{i1}^2 \leq \sigma^2$ for all i. By Bernstein's inequality for bounded variables, we get that

$$\text{pr}\Big[\Big|\sum_{i=1}^n \{\widetilde{X}_{i1} - E(\widetilde{X}_{i1})\}\Big| > x\Big] \leq 2\exp\Big(-\frac{x^2}{2n\sigma^2 + 2\xi_{n1} x/3}\Big). \quad (7)$$

Let us handle the second term. To use Bernstein's equality, we only require an appropriate control of moments. Using integration by parts, we observe that

$$E\big(|\widetilde{X}_{i2}|^q\big) \le q\int_{\xi_{n1}}^{\xi_{n2}} u^{q-1}\mathrm{pr}(|X_i|>u)du + \xi_{n1}^q\mathrm{pr}(|X_i|>\xi_{n1})$$

for $q \ge 2$. For integer $q \ge 2$,

$$\begin{aligned} q\int_{\xi_{n1}}^{\xi_{n2}} u^{q-1}\mathrm{pr}(|X_i|>u)du &\le q\zeta\int_{\xi_{n1}}^{\xi_{n2}} u^{q-1}\exp(-t_0u^\alpha)du \\ &\le q\alpha^{-1}\zeta(2t_0^{-1})^{q/\alpha}\int_{t_0\xi_{n1}^\alpha/2}^{t_0\xi_{n2}^\alpha/2} u^{q/\alpha-1}\exp(-2u)du \\ &\le q\alpha^{-1}\zeta\big(2t_0^{-1}\xi_{n2}^{1-\alpha}\big)^q\exp(-2^{-1}t_0\xi_{n1}^\alpha)\int_{t_0\xi_{n1}^\alpha/2}^{t_0\xi_{n2}^\alpha/2} u^{q-1}\exp(-u)du \\ &\le q!4\alpha^{-1}\zeta\big(t_0^{-1}\xi_{n2}^{1-\alpha}\big)^2\exp(-2^{-1}t_0\xi_{n1}^\alpha)\big(2t_0^{-1}\xi_{n2}^{1-\alpha}\big)^{q-2}. \end{aligned}$$

Choose $\xi_{n1} = \{4t_0^{-1}(1-\alpha)/(2-\alpha)\log n\}^{1/\alpha}$ and $\xi_{n2} = n^{1/(2-\alpha)} \vee x$. Write $\xi_n = n^{(1-\alpha)/(2-\alpha)}$ and $\nu = \max(16\zeta\alpha^{-1}t_0^{-2}, \sigma^2)$. Then

$$q\int_{\xi_{n1}}^{\xi_{n2}} u^{q-1}\mathrm{pr}(|X_i|>u)du \le \frac{1}{2}q!\nu\big\{1\vee x^{2(1-\alpha)}\xi_n^{-2}\big\}\big\{2t_0^{-1}(\xi_n\vee x^{1-\alpha})\big\}^{q-2}.$$

We also have that $\xi_{n1}^q\mathrm{pr}(|X_i|>\xi_{n1}) \le \xi_{n1}^q\exp(-t_0\xi_{n1}^\alpha) = \xi_{n1}^2\exp(-t_0\xi_{n1}^\alpha)\xi_{n1}^{q-2}$. A simple manipulation yields that there exists a positive integer N_{α,t_0} which depends only on α and t_0 such that

$$\xi_{n1}<\xi_{n2},\;\; \xi_{n1}^2\exp(-t_0\xi_{n1}^\alpha)\le 4\alpha^{-1}\zeta t_0^{-2},\;\; 2t_0^{-1}\xi_{n2}^{1-\alpha}\ge\xi_{n1},\;\; \text{and}\;\; 4\log n\le t_0\xi_{n2}^\alpha,$$

if $n > N_{\alpha,t_0}$. Then, if $x \le n^{1/(2-\alpha)}$,

$$E(|\widetilde{X}_{i2}|^q) \le \frac{1}{2}q!\nu\left(2t_0^{-1}\xi_n\right)^{q-2}$$

for $q \ge 2$; otherwise,

$$E(|\widetilde{X}_{i2}|^q) \le \frac{1}{2}q!\nu\big\{x^{2(1-\alpha)}\xi_n^{-2}\big\}\big(2t_0^{-1}x^{1-\alpha}\big)^{q-2}$$

for $q \ge 2$. By Bernstein's inequality, we obtain that

$$\mathrm{pr}\Big[\big|\sum_{i=1}^n\big\{\widetilde{X}_{i2}-E(\widetilde{X}_{i2})\big\}\big|>x\Big] \le 2\exp\Big(-\frac{x^2}{2n\nu+4t_0^{-1}n^{\frac{1-\alpha}{2-\alpha}}x}\Big)+2\exp\Big(-\frac{x^{2\alpha}}{2\nu n^{\frac{\alpha}{2-\alpha}}+4t_0^{-1}x^\alpha}\Big) \quad (8)$$

For the last term, we note that

$$\Big[\big|\sum_{i=1}^n\{\widetilde{X}_{i3}-E(\widetilde{X}_{i3})\}\big|>x\Big] \subset \Big\{\sup_i|X_i|>\xi_{n2}\Big\}\cup\Big\{\sup_i|X_i|\le\xi_{n2},\sum_{i=1}^n|E(X_iI(|X_i|>\xi_{n2}))|>x\Big\}.$$

Therefore, we have

$$\text{pr}\Big[\Big|\sum_{i=1}^{n}(\widetilde{X}_{i3}-E(\widetilde{X}_{i3})\}\Big|>x\Big]$$
$$\leq \text{pr}\Big(\sup_i |X_i|>\xi_{n2}\Big)+\text{pr}\Big[\sup_i |X_i|\leq \xi_{n2}, \sum_{i=1}^{n}|E\{X_iI(|X_i|>\xi_{n2})\}|>x\Big].$$

Note that $\zeta=\sup_{i\leq n}E\{\exp(t_0|X_i|^{\alpha})\}<\infty$. We observe that

$$\text{pr}\Big(\sup_i |X_i|>\xi_{n2}\Big)\leq \zeta n\exp\Big(-t_0\xi_{n2}^{\alpha}\Big)\leq \zeta n\exp\Big(-t_0x^{\alpha}\Big).$$

In a similar fashion, we obtain that

$$\sum_{i=1}^{n}|E\{X_iI(|X_i|>\xi_{n2})\}|\leq n\sigma\left\{\text{pr}(|X_i|>\xi_{n2})\right\}^{1/2}\leq n^{-1}\sigma\zeta\exp\Big(2\log n-2^{-1}t_0\xi_{n2}^{\alpha}\Big).$$

As a result, for $x>\sigma\zeta n^{-1}$ and $n>N_{\alpha,t_0}$,

$$\text{pr}\Big[|\sum_{i=1}^{n}\{\widetilde{X}_{i3}-E(\widetilde{X}_{i3})\}|>x\Big]\leq \zeta n\exp\Big(-t_0x^{\alpha}\Big). \tag{9}$$

Combing the three inequalities (7)-(9), we conclude that, for $x>\sigma\zeta n^{-1}$ and $n>N_{\alpha,t_0}$,

$$\text{pr}\Big[\Big|\sum_{i=1}^{n}\{X_i-E(X_i)\}\Big|>3x\Big]\leq 4\exp\left(-\frac{x^2}{2n\nu+4t_0^{-1}n^{\frac{1-\alpha}{2-\alpha}}x}\right)$$
$$+2\exp\left(-\frac{x^{2\alpha}}{2\nu n^{\frac{\alpha}{2-\alpha}}+4t_0^{-1}x^{\alpha}}\right)+n\zeta\exp\left(-t_0x^{\alpha}\right).$$

If $x\leq\sigma\zeta n^{-1}$ or $n\leq N_{\alpha,t_0}$, we can always multiply a large positive constant C on the right hand side to make the inequality hold. The proof is completed. □

LEMMA 4. *Suppose that* $\{X_1=(X_{1,1},X_{1,2})^{\top},X_2=(X_{2,1},X_{2,2})^{\top},\ldots\}$ *are independent random vectors and* $\sup_{i\leq n,j=1,2}E\{\exp(t_0|X_{i,j}|^{2\alpha})\}\leq\zeta$ *for some positive constants* α, t_0 *and* ζ *with* $0<\alpha\leq 1$. *Denote by* l_n *a sequence that may depend on* n, *and* $1\leq l_n\leq O(n^{\epsilon})$ *with* $0\leq\epsilon<1$. *Then, for each* m *and* m' *with* $m,m'=1,2$, *there exist positive constants* $C_j (j=1,\ldots,4)$ *such that for any* $x>0$, *the following concentration inequality holds:*

$$pr\Big[\Big|\sum_{i=1}^{n}\{X_{i,m}X_{i+l_n,m'}-E(X_{i,m}X_{i+l_n,m'})\}\Big|>3(l_n+1)x\Big]$$
$$\leq (l_n+1)C_1\exp\left(-\frac{x^2}{C_2n+C_3n^{\frac{1-\alpha}{2-\alpha}}x}\right)+C_1(l_n+1)\exp\left(-\frac{x^{2\alpha}}{C_2n^{\frac{\alpha}{2-\alpha}}+C_3x^{\alpha}}\right)$$
$$+C_1(l_n+1)n\exp\left(-C_4x^{\alpha}\right).$$

Proof of Lemma 4. Without loss of generality, we assume that $n/(l_n+1)$ is a positive integer. Here we prove the inequality for $m=1$ and $m'=2$ only. Similar techniques can be applied to other cases. Let $Y_{ji}=X_{(i-1)(l_n+1)+j,1}X_{i(l_n+1)+j-1,2}$. Then, for each j, $\{Y_{ji},i=1,\ldots,n/(l_n+1)\}$ are independent with $\sup_{i,j}E\{\exp(t_0|Y_{ji}|^{\alpha})\}\leq\zeta<\infty$. With the help of

Y_{ji}, $\sum_{i=1}^{n}\big\{X_{i,1}X_{i+l_n,2}-E(X_{i,1}X_{i+l_n,2})\big\}$ can be re-expressed as

$$\sum_{i=1}^{n}\big\{X_{i,1}X_{i+l_n,2}-E(X_{i,1}X_{i+l_n,2})\big\}=\sum_{j=1}^{l_n+1}\sum_{i=1}^{n/(l_n+1)}\big\{Y_{ji}-E(Y_{ji})\big\}.$$

By Lemma 3, we obtain that there exist positive constants $C_j(j=1,\ldots,4)$ such that

$$\text{pr}\Big[\Big|\sum_{i=1}^{n/(l_n+1)}\big\{Y_{ji}-E(Y_{ji})\big\}\Big|>3x\Big]\le C_1\exp\left(-\frac{x^2}{C_2n+C_3n^{\frac{1-\alpha}{2-\alpha}}x}\right)$$
$$+C_1\exp\left(-\frac{x^{2\alpha}}{C_2n^{\frac{\alpha}{2-\alpha}}+C_3x^{\alpha}}\right)+C_1n\exp\left(-C_4x^{\alpha}\right),$$

for each $j=1,\ldots,l_n+1$. Note that

$$\Big|\sum_{i=1}^{n}\big\{X_{i,1}X_{i+l_n,2}-E(X_{i,1}X_{i+l_n,2})\big\}\Big|\le(l_n+1)\sup_{j\le l_n+1}\Big|\sum_{i=1}^{n/(l_n+1)}\big\{Y_{ji}-E(Y_{ji})\big\}\Big|.$$

Therefore,

$$\text{pr}\Big[\Big|\sum_{i=1}^{n}\big\{X_{i,1}X_{i+l_n,2}-E(X_{i,1}X_{i+l_n,2})\big\}\Big|>3(l_n+1)x\Big]$$
$$\le(l_n+1)\sup_{j\le l_n+1}\text{pr}\Big[\Big|\sum_{i=1}^{n/(l_n+1)}\big\{Y_{ji}-E(Y_{ji})\big\}\Big|>3x\Big].$$

The lemma is proved. □

Lemmas 5, 6 and 7 below are based on the VAR(1) model with $\|A_1\|_2\le\delta<1$. Similar techniques can be applied to the general VAR(d) model. For $j,k=1,\ldots,p$, define $\widehat{\Sigma}_{jk}=n^{-1}\sum_{t=1}^{n}y_{j,t}y_{k,t}$ and $\Sigma_{jk}=E(\widehat{\Sigma}_{jk})$. For $i=1,\ldots,p$, let $e_{(i)}=(\varepsilon_{i,2},\ldots,\varepsilon_{i,n})^{\mathrm{T}}$ and $x_{(i)}=(y_{i,1},\ldots,y_{i,n-1})^{\mathrm{T}}$. We should note that Lemmas 5 and 6 have the same rate expressions but the actual rates are different, since they are under Conditions 4(i) and 4(ii), respectively.

LEMMA 5. *Suppose that Conditions (1)–(3) and 4(i) hold.*

(i) For $j,k=1,\ldots,p$, there exist positive constants C_1,C_2 and C_3 free of (j,k,n,p) such that

$$pr\left(\left|\widehat{\Sigma}_{jk}-\Sigma_{jk}\right|>x\right)\le\frac{C_1n}{(nx)^q}+C_2\exp\big(-C_3nx^2\big)$$

holds for $x>0$; consequently, this leads to the following uniform convergence rate:

$$\sup_{1\le j,k\le p}\left|\widehat{\Sigma}_{jk}-\Sigma_{jk}\right|=O_P\Big\{(n^{-1}\log p)^{1/2}\Big\}.$$

(ii) For $j,k=1,\ldots,p$, there exist positive constants C_1,C_2 and C_3 free of (j,k,n,p) such that

$$pr\big\{|e_{(j)}^{\mathrm{T}}x_{(k)}|\ge x\big\}\le\frac{C_1n}{x^{2q}}+C_2\exp\big(-C_3x^2\big)$$

holds for $x>0$; in particular, we have

$$\sup_{1\le j,k\le p}|e_{(j)}^{\mathrm{T}}x_{(k)}|=O_P\Big\{(n\log p)^{1/2}\Big\}.$$

Proof of Lemma 5. Here we prove part (i) only. Part (ii) can be proved analogously. Let $\mu_q = \sup_{j\le p}\|\varepsilon_{j0}\|_q$ for $q\ge 2$. To use the results of Lemma 1, we just need to bound the physical dependent measure of $y_{j,t}y_{k,t}$ for each j and k, denoted by $\widetilde{\theta}_{i,q,j,k} = \|y_{j,i}y_{k,i} - y^*_{j,i}y^*_{k,i}\|_q$ with $y^*_{j,i}$ being the coupled process of $y_{j,i}$. Denote the physical dependent measure of $y_{j,i}$ by $\theta_{i,2q,j} = \|y_{j,i} - y^*_{j,i}\|_{2q}$ with $y^*_{j,i}$ being the coupled process of $y_{j,i}$.

We will show (a) $\sup_{j\le p}\|y_{j,i}\|_{2q}\le C\mu_{2q}$; (b) $\sup_{j\le p}\theta_{i,2q,j}\le C\mu_{2q}(i+1)\delta^i$, where C is some positive constant and depends only on the spectral norm of A_1 rather than q. Observe that $\|y_{j,i}y_{k,i} - y^*_{j,i}y^*_{k,i}\|_q \le \|y_{j,i}y_{k,i} - y^*_{j,i}y_{k,i}\|_q + \|y_{j,i}y_{k,i} - y_{j,i}y_{k,i}\|_q$ and hence

$$\|y_{j,i}y_{k,i} - y^*_{j,i}y^*_{k,i}\|_q \le \sup_{j\le p}\|y_{j,i}\|_{2q}\big(\theta_{i,2q,j} + \theta_{i,2q,k}\big).$$

If both bounds (a) and (b) are obtained, then,

$$\widetilde{\Theta}_{m,q} = \sup_{j,k\le p}\sum_{i=m}^{\infty}\widetilde{\theta}_{i,2q,j,k} \le C\mu_{2q}^2\sum_{i=m}^{\infty}(i+1)\delta^i \le C\mu_{2q}^2(1-\delta)^{-2}(m+1)\delta^m = o(m^{-\alpha})$$

for any $\alpha > 1$. Applying Lemma 1 we prove part (i).

Let us turn to bound $\sup_{j\le p}\|y_{j,i}\|_{2q}$. Let A_1^l be $(a_{l,jk})_{j,k\le p}$ with $l\ge 1$. Since A_1^l is a banded matrix with the bandwidth $\min(2lk_0+1,p)$, we can bound $\|A_1^l\|_\infty$ by

$$\|A_1^l\|_\infty = \max_{j\le p}\sum_{k=1}^{p}|a_{l,jk}| \le \{\min(2lk_0+1,p)\}^{1/2}\|A_1^l\|_2 \le C(2lk_0+1)\delta^l, l\ge 1, \quad (10)$$

which implies that $\|A_1^l\|_\infty \le C(2k_0+1)(l+1)\delta^l, l\ge 0$. Using the innovation representation $y_t = \sum_{l=0}^{\infty}A_1^l\varepsilon_{t-l}$, we get

$$\|y_{j,i}\|_{2q} \le \sum_{l=0}^{\infty}\|\sum_{k=1}^{p}a_{l,jk}\varepsilon_{k,i-l}\|_{2q} \le \sum_{l=0}^{\infty}\sum_{k=1}^{p}|a_{l,jk}|\|\varepsilon_{k,i-l}\|_{2q}.$$

As a result, $\sup_{j\le p}\|y_{j,i}\|_{2q} \le C(2k_0+1)\mu_{2q}\sum_{l=0}^{\infty}(l+1)\delta^l = C(2k_0+1)(1-\delta)^{-2}\mu_{2q} < \infty$. Similarly, we can bound $\sup_{j\le p}\theta_{i,2q,j}$ above by $C(i+1)\delta^i$ with some positive constant C since we have a nice inequality

$$\|y_{j,i} - y^*_{j,i}\|_{2q} = \|\sum_{k=1}^{p}a_{i,jk}\big(\varepsilon_{k,0} - \varepsilon^*_{k,0}\big)\|_{2q} \le 2\mu_{2q}\|A_1^i\|_\infty.$$

The proof is complete. □

LEMMA 6. *Suppose that Conditions (1)–(3) and 4(ii) hold. Then we have*

(i) $\sup_{1\le j,k\le p}|\widehat{\Sigma}_{jk} - \Sigma_{jk}| = O_P\Big\{(n^{-1}\log p)^{1/2}\Big\}$; *(ii)* $\sup_{1\le j,k\le p}|e_{(j)}^{\mathrm{T}}x_{(k)}| = O_P\Big\{(n\log p)^{1/2}\Big\}$.

Proof of Lemma 6. Here we prove part (i) only. The proof of part (ii) can be derived similarly.

Note that $y_t = A_1y_{t-1} + \varepsilon_{t-l}$ and $\|A_1\|_2 \le \delta < 1$. Let A_1^l be $(a_{l,jk})_{j,k\le p}$. For each j, $y_{j,t} = \sum_{l=0}^{\infty}\sum_{m=1}^{p}a_{l,jm}\varepsilon_{m,t-l}$ converges almost surely. Write $\eta_{j,lt} = \sum_{m=1}^{p}a_{l,jm}\varepsilon_{m,t-l}$ for $l\ge 0$. We divide $y_{j,t}$ into two terms $y_{jt} = \sum_{l=0}^{N_n}\eta_{j,lt} + \sum_{l=N_n+1}^{\infty}\eta_{j,lt}$. Here we choose N_n to be

$N_\delta \log(n)$ with $N_\delta > (1+\alpha)\alpha^{-1}(-\log\delta)^{-1}$. Hence, $n\widehat{\Sigma}_{jk}$ can be expressed as

$$\begin{aligned} n\widehat{\Sigma}_{jk} &= \sum_{l,l'=0}^{N_n}\left(\sum_{t=1}^{n}\eta_{j,lt}\eta_{k,l't}\right) + \sum_{l,l'=N_n+1}^{\infty}\left(\sum_{t=1}^{n}\eta_{j,lt}\eta_{k,l't}\right) \\ &+ \sum_{l=0}^{N_n}\sum_{l'=N_n+1}^{\infty}\left(\sum_{t=1}^{n}\eta_{j,lt}\eta_{k,l't}\right) + \sum_{l=N_n+1}^{\infty}\sum_{l'=0}^{N_n}\left(\sum_{t=1}^{n}\eta_{j,lt}\eta_{k,l't}\right) \\ &= S_{jk,1} + S_{jk,2} + S_{jk,3} + S_{jk,4}, \end{aligned}$$

and $n\Big(\widehat{\Sigma}_{jk} - \Sigma_{jk}\Big) = \sum_{m=1}^{4}\big\{S_{jk,m} - E(S_{jk,m})\big\}$. Let us handle the first term $S_{jk,1} - E(S_{jk,1})$. Note that if $\sup_{m,l} E\{\exp(|t_0\varepsilon_{m,l}|^{2\alpha})\} < \infty$,

$$\zeta_\varepsilon = \sup_{m,l,m',l'} E\big\{\exp\big(t_0|\varepsilon_{m,l}\varepsilon_{m',l'}|^{\alpha}\big)\big\} < \infty.$$

By Lemma 4, we obtain the following equality,

$$\begin{aligned} &\text{pr}\Big[\Big|\sum_{t=1}^{n}\big\{\varepsilon_{m,t-l}\varepsilon_{m',t-l'} - E(\varepsilon_{m,t-l}\varepsilon_{m',t-l'})\big\}\Big| > 3(l_n+1)x\Big] \\ &\le (l_n+1)C_1\exp\left(-\frac{x^2}{C_2 n + C_3 n^{\frac{1-\alpha}{2-\alpha}}x}\right) + C_1(l_n+1)\exp\left(-\frac{x^{2\alpha}}{C_2 n^{\frac{\alpha}{2-\alpha}} + C_3 x^{\alpha}}\right) \\ &+ C_1(l_n+1)n\exp\left(-C_4 x^{\alpha}\right) \end{aligned}$$

for some positive constants $C_j (j = 1,\ldots,4)$, where $l_n = |l - l'|$. Taking $x = C(n\log p)^{1/2}$ for some large constant $C > 0$, we derive the following convergence rate,

$$\widetilde{\eta}_n = \sup_{m,m'\le p, l,l'\le N_n}(l'+1)^{-2}(l+1)^{-2}\left|\sum_{t=1}^{n}\big\{\varepsilon_{m,t-l}\varepsilon_{m',t-l'} - E(\varepsilon_{m,t-l}\varepsilon_{m',t-l'})\big\}\right| = O_P\big\{(n\log p)^{1/2}\big\}.$$

Observe that

$$\begin{aligned} \Big|\sum_{t=1}^{n}\big\{\eta_{j,lt}\eta_{k,l't} - E(\eta_{j,lt}\eta_{k,l't})\big\}\Big| &\le \sum_{m=1}^{p}|a_{l,jm}|\sum_{m'=1}^{p}|a_{l',km'}|\Big|\sum_{t=1}^{n}\big\{\varepsilon_{m,t-l}\varepsilon_{m',t-l'} - E(\varepsilon_{m,t-l}\varepsilon_{m',t-l'})\big\}\Big| \\ &\le C(2k_0+1)^2(l+1)^3(l'+1)^3\delta^{l+l'}\widetilde{\eta}_n, \end{aligned} \tag{11}$$

and $\sum_{l=0}^{N_\delta}(l+1)^3\delta^l < \infty$. Therefore,

$$\sup_{j,k\le p}\Big|S_{jk,1} - E(S_{jk,1})\Big| \le C(2k_0+1)^2\widetilde{\eta}_n = O_P\big\{(n\log p)^{1/2}\big\}.$$

Consider the second term. Since $\sup_{m,l} E\big\{\exp(|t_0\varepsilon_{m,l}|^{2\alpha})\big\} < \infty$, $\widetilde{\zeta}_{q,\varepsilon} = \sup_{m,l,m',l'}\Big\|\varepsilon_{m,l}\varepsilon_{m',l'}\Big\|_q \le Cq^{1/\alpha}$ for any $q > 2$. Now we bound $\Big\|S_{jk,2} - ES_{jk,2}\Big\|_q$. To be specific,

$$\begin{aligned} \Big\|\sum_{t=1}^{n}\big\{\eta_{j,lt}\eta_{k,l't} - E(\eta_{j,lt}\eta_{k,l't})\big\}\Big\|_q &\le n\sum_{m,m'=1}^{p}|a_{l,jm}||a_{l,km'}|\sup_{m,m',l,l'}\Big\|\varepsilon_{m,l}\varepsilon_{m,l'}\Big\|_q \\ &\le n(2k_0+1)^2(l+1)(l'+1)\delta^{l+l'}\widetilde{\zeta}_{q,\varepsilon}. \end{aligned} \tag{12}$$

Hence,

$$\left\|S_{jk,2}-E(S_{jk,2})\right\|_q \le Cnq^{1/\alpha}\sum_{l,l'=N_n+1}^{\infty}(l+1)(l'+1)\delta^{l+l'} \le C\cdot nN_n^2\delta^{2N_n}q^{1/\alpha}.$$

Write $\eta_{n2}=\big(nN_n^2\delta^{2N_n}\big)^{-1}\big\{S_{jk,2}-E(S_{jk,2})\big\}$. It follows from Lemma 2 that there exists a constant $\lambda>0$ such that $E\big\{\exp(\lambda|\eta_{n2}|^{\alpha})\big\}<\infty$. Consequently, for a large constant $C>0$, we have that

$$\mathrm{pr}\Big\{\sup_{j,k\le p}\Big|S_{jk,2}-E(S_{jk,2})\Big|>C(\log n)^2\Big\}\le O(1)p^2\exp\Big\{-tC^{\alpha}\cdot n(\log n)^{-2\alpha}(\log n)^{2\alpha}\Big\}\to 0,$$

as $n\to\infty$, which implies that $\sup_{j,k\le p}\Big|S_{jk,2}-E(S_{jk,2})\Big|=O_P\Big\{(\log n)^2\Big\}=o_P\big\{(n\log p)^{1/2}\big\}$. Similarly, we can prove that $\sup_{j,k\le p}\Big|S_{jk,m}-E(S_{jk,m})\Big|=o_P\big\{(n\log p)^{1/2}\big\}, m=3,4$.

Finally putting together the convergence rate results for the four terms we conclude

$$\sup_{j,k\le p}\Big|\widehat{\Sigma}_{jk}-\Sigma_{jk}\Big|=O_P\big\{(n^{-1}\log p)^{1/2}\big\}.$$

The proof is complete. □

LEMMA 7. *Suppose that Conditions (1)–(3) and 4(i) or 4(ii) hold. Then, for each finite k with $k\ge k_0$,*

$$\sup_{1\le i\le p}\Big|\frac{\mathrm{RSS}_i(k)}{n\sigma_i^2}-1\Big|=O_P\Big\{(n^{-1}\log p)^{1/2}\Big\},$$

as $n\to\infty$, where $\mathrm{RSS}_i(k)$ is defined in (2.6) and σ_i^2 is the (i,i)-th element of Σ_ε.

Proof of Lemma 7. For $k>k_0$, the term $\mathrm{RSS}_i(k)$ can be decomposed as

$$\mathrm{RSS}_i(k)=e_i^{\mathrm T}e_i-e_i^{\mathrm T}X_i\big(X_i^{\mathrm T}X_i\big)^{-1}X_i^{\mathrm T}e_i=R_{i1}-R_{i2},$$

where $e_{(i)}=(\varepsilon_{i,2},\dots,\varepsilon_{i,n})^T$, and X_i is a $(n-1)\times\tau_i(k)$ matrix with $x_{i,1+j}$ as its j-th row. We will show below that, under Assumptions (1)–(3) and 4(i) or 4(ii),

$$\text{(a)}\quad \sup_{i\le p}\Big|R_{i1}-n\sigma_i^2\Big|=O_P\{(n\log p)^{1/2}\};\quad \text{(b)}\quad \sup_{i\le p}|R_{i2}|=O_P\big(\log p\big).$$

With results in (a) and (b), it follows that

$$\sup_{i\le p}\Big|\frac{\mathrm{RSS}_i(k)}{n\sigma_i^2}-1\Big|\le\sup_{i\le p}\Big|\frac{R_{i1}}{n\sigma_i^2}-1\Big|+\sup_{i\le p}\Big|\frac{R_{i2}}{n\sigma_i^2}\Big|=O_P\big\{(n^{-1}\log p)^{1/2}\big\}.$$

Suppose first that Condition 4(i) holds. Consider the term $R_{i1}-n\sigma_i^2$. Lemma 1 shows that

$$\sup_{i\le p}\Big|e_i^{\mathrm T}e_i-n\sigma_i^2\Big|=O_P\big\{(n\log p)^{1/2}\big\}.$$

Let us handle the term $\sup_{i\le p}|R_{i2}|$. Define

$$\mathcal{A}_n=\Big\{\inf_{i\le p}\lambda_{\min}\big(n^{-1}X_i^{\mathrm T}X_i\big)>\kappa_1(1-\eta)\Big\}$$

with $0 < \eta < 1$. It follows from Lemma 5(i) and Condition 3 that $P(\mathcal{A}_n) \to 1$ as $n \to \infty$. On the event $\mathcal{A}_n$, the term $\sup_{i \le p} |R_{i2}|$ can be bounded above by $\big(\kappa_1(1-\eta)\big)^{-1} k_0 \sup_{j,k \le p} n^{-1} |e_{(j)}^{\mathrm{T}} x_{(k)}|^2$. By Lemma 5 (ii), we obtain that, $\sup_{j,k \le p} |e_{(j)}^{\mathrm{T}} x_{(k)}| = O_P\big\{(n \log p)^{1/2}\big\}$, which implies that (b) holds.

Suppose that Condition 4(ii) holds. Consider the term $R_{i1} - n\sigma_i^2$. By Lemma 3 and taking $x = C(n \log p)^{1/2}$ with large constant $C > 0$, we have that

$$\sup_{i \le p} \Big| e_i^{\mathrm{T}} e_i - n\sigma_i^2 \Big| = O_P\big\{(n \log p)^{1/2}\big\}.$$

Similarly, we can establish (b) from Lemma 6. The proof is complete. □

8. An additional Simulation Study

We conduct an additional Monte Carlo experiment to examine the proposed methodology. We consider an $\mathrm{VAR}(1)$ model with the sample size $n + 2$, where the last two observations are used to calculate the one-step and two-step ahead post-sample prediction errors.

The data were generated from the $\mathrm{VAR}(1)$ model with the banded coefficient matrix A specified in scenario (1) in the paper. We set $n = 200$, $p = 100, 200$, $k_0 = 2$, and each setting was repeated 100 times. To mimic the real world with the true ordering unknown, we considered three other orderings through random permutation. The first ordering was generated through local permutation, where we partitioned the components of y_t into [p/5] groups with each group containing 5 components. We then performed a random permutation within each group. The other two orderings were generated through permutating the whole compoments of y_t together. Also included in the comparison is the sparse AR model determined by lasso. Table 1 below reports simulation results of BIC and prediction errors. It indicates that the model with the true ordering offers the best post-sample prediction, followed by the model with the local permutation only, and then the lasso-based model, while the two models with arbitrary permutations perform the worst.

Table 1. *Average BIC, estimated bandwidth parameter and one-step-ahead and two-step-ahead post-sample predictive errors over* 100 *replications, with their corresponding standard errors in parentheses.*

Ordering	BIC	Bandwidth	One-step ahead	Two-step ahead
Case 1: $n = 200, p = 100$				
True ordering	546(3.5)	1.78(0.52)	0.787(0.06)	0.837(0.07)
Local permutation	549(5.6)	2.14(0.83)	0.788(0.06)	0.837(0.07)
Random permutation	546(11)	0.71(0.90)	0.828(0.07)	0.848(0.08)
Random permutation	547(13)	0.70(1.06)	0.827(0.06)	0.848(0.08)
Lasso	–	–	0.823(0.06)	0.846(0.08)
Case 2: $n = 200, p = 200$				
True ordering	1093(4.8)	1.87(0.36)	0.786(0.04)	0.830(0.05)
Local permutation	1102(10)	2.39(0.75)	0.787(0.04)	0.829(0.05)
Random permutation	1098(22)	0.89(0.80)	0.829(0.05)	0.839(0.05)
Random permutation	1096(20)	0.78(0.70)	0.831(0.05)	0.838(0.05)
Lasso	–	–	0.827(0.05)	0.838(0.05)

REFERENCES

LIU, W., XIAO, H. AND WU, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinca* **23**, 1257-1272.

WANG, H., LI, B. AND LENG, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *J. R. Statist. Soc. B* **71**, 671-683.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Nation. Acad. Sci* **102**, 14150-14154.