Supplementary Material to "High-dimensional Linear Regression for Dependent Data with Applications to Nowcasting"

Yuefeng Han and Ruey S. Tsay

University of Chicago

In this supplementary document, we provide the proofs of main results in the paper.

A Lemmas

We start with some lemmas that are useful in deriving the main results of the paper.

Lemma 1. *Assume that $\|e_.\|_{q,\alpha} < \infty$, where $q > 2$ and $\alpha > 0$, $\sum_{i=1}^n w_i^2 = n$. Let $w = (w_1, ..., w_n)$, $\varsigma_n = 1$ (resp. $(\log n)^{1+2q}$ or $n^{q/2-1-\alpha q}$) if $\alpha > 1/2 - 1/q$ (resp. $\alpha = 0$ or $\alpha < 1/2 - 1/q$). Then for all $x > 0$, $S_n = \sum_{i=1}^n w_i e_i$,*

$$P(|S_n| \geq x) \leq K_1 \frac{\varsigma_n |w|_q^q \|e_.\|_{q,\alpha}^q}{x^q} + K_2 \exp\left(-\frac{K_3 x^2}{n\|e_.\|_{2,\alpha}^2}\right)$$

where K_1, K_2, K_3 are constants that depend only on q and α.

Proof. See Wu and Wu (2016) Theorem 2. □

Lemma 2. *Assume $\||\mathbf{x}_.|_\infty\|_{q,\alpha} < \infty$, where $q > 2$ and $\alpha > 0$, and $\Psi_{2,\alpha} < \infty$, $\sum_{i=1}^n w_i^2 = n$. Let $w = (w_1, ..., w_n)$ and $T_n = \sum_{i=1}^n w_i \mathbf{x}_i$. (i) If $\alpha > 1/2 - 1/q$, then for $x \gtrsim \sqrt{n \log p}\Psi_{2,\alpha} + |w|_q (\log p)^{3/2} \||\mathbf{x}_.|_\infty\|_{q,\alpha}$,*

$$P(|T_n|_\infty \geq x) \leq \frac{K_{q,\alpha} |w|_q^q (\log p)^{q/2} \||\mathbf{x}_.|_\infty\|_{q,\alpha}^q}{x^q} + K_{q,\alpha} \exp\left(-\frac{K_{q,\alpha} x^2}{n\Psi_{2,\alpha}^2}\right). \tag{A.1}$$

(ii) If $0 < \alpha < 1/2 - 1/q$, then for $x \gtrsim \sqrt{n \log p}\Psi_{2,\alpha} + n^{1/2-\alpha-1/q} |w|_q (\log p)^{3/2} \||\mathbf{x}_.|_\infty\|_{q,\alpha}$,

$$P(|T_n|_\infty \geq x) \leq \frac{K_{q,\alpha} n^{q/2-1-\alpha q} |w|_q^q (\log p)^{q/2} \||\mathbf{x}_.|_\infty\|_{q,\alpha}^q}{x^q} + K_{q,\alpha} \exp\left(-\frac{K_{q,\alpha} x^2}{n\Psi_{2,\alpha}^2}\right), \tag{A.2}$$

where $K_{q,\alpha}$ is a constant that depends on q and α only.

Proof. The lemma can be shown following similar arguments as those in the proof of Zhang and Wu (2017) Theorem 6.2. Details are omitted. □

Lemma 3. *Let A and B denote two positive semi-definite, s-dimensional square matrices. If $\max_{1 \leq j,k \leq s} |A_{jk} - B_{jk}| \leq \delta$, then $\inf_{|\zeta|_2=1} \zeta' B \zeta > \inf_{|\zeta|_2=1} \zeta' A \zeta - s\delta$.*

Proof. See Lemma 3 of Medeiros and Mendes (2016). □

Lemma 4. *For linear model $Y = X\beta + e$, assume that the matrix $X_{(1)}^T X_{(1)}$ is invertible. Then for any given $\lambda > 0$, and any noise term $e \in \mathbb{R}^n$, there exists a Lasso estimator $\hat{\beta}(\lambda)$ which satisfies $\hat{\beta}(\lambda) =_s \beta$, if and only if the following two conditions hold*

$$sign\left(\beta_{(1)} + (\frac{1}{n}X_{(1)}^T X_{(1)})^{-1}\left[\frac{1}{n}X_{(1)}^T e - \lambda sign(\beta_{(1)})\right]\right) = sign(\beta_{(1)}),$$
$$\left|X_{(2)}^T X_{(1)}(X_{(1)}^T X_{(1)})^{-1}\left[\frac{1}{n}X_{(1)}^T e - \lambda sign(\beta_{(1)})\right] - \frac{1}{n}X_{(2)}^T e\right| \leq \lambda,$$

where the vector inequality and equality are taken elementwise, $\beta_{(1)}$ and $\beta_{(2)}$ denote the first s and last $p - s$ entries of β respectively.

Proof. See Wainwright (2009). □

B A general theorem of estimation error for weak sparsity

Lemma 5. *Define $\hat{\Delta} = \hat{\beta} - \beta$, where β satisfies weakly sparsity condition (Assumption 1),* i.e., $\sum_{j=1}^p |\beta_j|^\theta \leq K_\theta$ *for $0 \leq \theta < 1$. Suppose $\hat{\Delta}\hat{\Sigma}\hat{\Delta} \geq \kappa|\hat{\Delta}|_2^2$, where κ is a positive constant that does not depend on $\hat{\Delta}$. Choose $\lambda \geq 2|n^{-1}\sum_{i=1}^n \mathbf{x}_i e_i|_\infty$. Then we have for some constants C_1, C_2,*

$$|\hat{\Delta}|_2^2 \leq C_1 K_\theta \left(\frac{\lambda}{\kappa}\right)^{2-\theta}, \tag{B.1}$$

$$|\hat{\Delta}|_1 \leq C_2 K_\theta \left(\frac{\lambda}{\kappa}\right)^{1-\theta}. \tag{B.2}$$

This result is deterministic and non-asymptotic. The statistical performance of $\hat{\beta}$ relies on the restricted eigenvalue condition properties of sample covariance $\hat{\Sigma}$.

Proof. This result is just a simple application of the theoretical framework established in Negahban et al. (2012), for the sake of brevity, we omitted the detailed proof here. □

C Proof of Theorem 1

Proof. Recall $\hat{\Sigma} = (\hat{\sigma}_{jk})_{1\leq j,k\leq p} = 1/n\sum_{i=1}^n x_i x_i^T = n^{-1}X^T X$, $\Sigma = (\sigma_{jk})_{1\leq j,k\leq p}$. Define the events

$$\mathcal{A} = \{|\hat{\Sigma} - \Sigma|_\infty \leq a\} = \{\max_{j,k} |\hat{\sigma}_{jk} - \sigma_{jk}| \leq a\}, \tag{C.1}$$

$$\mathcal{B} = \{n^{-1}\left|X^T e\right|_\infty \leq \lambda/2\}. \tag{C.2}$$

The first step is to control the probability $\mathsf{P}(\mathcal{A}^c)$ and $\mathsf{P}(\mathcal{B}^c)$. By Hölder's inequality, we have for $m \geq 0$ that

$$\begin{aligned}\sum_{l=m}^\infty \|x_{lj}e_l - x_{lj}^* e_l^*\|_\tau &\leq \sum_{l=m}^\infty \left(\|x_{lj}(e_l - e_l^*)\|_\tau + \|(x_{lj} - x_{lj}^*)e_l^*\|_\tau\right) \\ &= \sum_{l=m}^\infty \left(\|x_{lj}\|_\gamma \|e_l - e_l^*\|_q + \|x_{lj} - x_{lj}^*\|_\gamma \|e_l^*\|_q\right).\end{aligned}$$

Since $\alpha = \min(\alpha_X, \alpha_e)$, the dependence adjusted norm satisfies

$$\|x_{.j}e_.\|_{\tau,\alpha} \leq \|x_{.j}\|_{\gamma,0}\|e_.\|_{q,\alpha_e} + \|x_{.j}\|_{\gamma,\alpha_X}\|e_.\|_{q,0} \leq 2\|x_{.j}\|_{\gamma,\alpha_X}\|e_.\|_{q,\alpha_e}. \tag{C.3}$$

Similarly, we have

$$\|x_{.j}x_{.k} - \sigma_{jk}\|_{\gamma/2,\alpha_X/2} \leq 2\|x_{.j}\|_{\gamma,\alpha_X}\|x_{.k}\|_{\gamma,\alpha_X}, \tag{C.4}$$

Hence,

$$\max_{1\leq j\leq p} \|x_{.j}e_.\|_{\tau,\alpha} \leq 2M_eM_X, \tag{C.5}$$

$$\max_{1\leq j,k\leq p} \|x_{.j}x_{.k} - \sigma_{jk}\|_{\gamma/2,\alpha_X/2} \leq 2M_X^2. \tag{C.6}$$

Employing a similar derivation, we can show that,

$$\|\max_{1\leq j\leq p} |x_{.j}e_.|\|_{\tau,\alpha} \leq 2\||\mathbf{x}_.|_\infty\|_{\gamma,\alpha_X}M_e, \tag{C.7}$$

$$\|\max_{1\leq j,k\leq p} |x_{.j}x_{.k} - \sigma_{jk}|\|_{\gamma/2,\alpha_X/2} \leq 2\||\mathbf{x}_.|_\infty\|^2_{\gamma,\alpha_X}. \tag{C.8}$$

Note that $M_X \leq \||\mathbf{x}_.|_\infty\|_{\gamma,\alpha_X} \leq \Upsilon_{\gamma,\alpha_X}$.

If $\tau > 2$, for $\lambda \gtrsim \sqrt{\log p/n}M_eM_X + n^{\rho/\tau-1}(\log p)^{3/2}M_e\||\mathbf{x}_.|_\infty\|_{\gamma,\alpha_X}$, adopting (C.5), (C.7) and Lemma 2, we have,

$$\mathsf{P}(\mathcal{B}^c) = C_4\frac{n^\rho(\log p)^{\tau/2}\||\mathbf{x}_.|_\infty\|^\tau_{\gamma,\alpha_X}M_e^\tau}{(n\lambda)^\tau} + C_5e^{-C_6n\lambda^2/(M_X^2M_e^2)}.$$

Under our choice of λ, if $\tau > 2$, $\mathsf{P}(\mathcal{B}^c) = C_4(\log p)^{-\tau} + C_5p^{-C_6}$. Similarly, we can prove, if $na \gtrsim \sqrt{n\log p}M_X^2 + n^{2\nu/\gamma}(\log p)^{3/2}\||\mathbf{x}_.|_\infty\|^2_{\gamma,\alpha_X}$, $\mathsf{P}(\mathcal{A}^c) = C_1(\log p)^{-\gamma/2} + C_2p^{-C_3}$.

Denote $\omega = \sqrt{\log p/n}M_X^2 + n^{2\nu/\gamma-1}(\log p)^{3/2}\||\mathbf{x}_.|_\infty\|^2_{\gamma,\alpha_X}$. Then for some constant $\eta_1 > 0$, we have

$$\mathsf{P}\left(\forall\Delta \in \mathbb{R}^p, \Delta'\hat{\Sigma}\Delta \geq \Delta'\Sigma\Delta - \eta_1\omega|\Delta|_1^2\right) \geq 1 - C_1(\log p)^{-\gamma/2} - C_2p^{-C_3}. \tag{C.9}$$

In other words, with high probability $1-\mathsf{P}(\mathcal{A}^c)$, the Restricted Strong Convexity condition $\Delta'\hat{\Sigma}\Delta \geq \kappa|\Delta|_2^2 - \eta_1\omega|\Delta|_1^2$ holds.

Denote $\hat{\Delta} = \hat{\beta} - \beta$. For a threshold $\delta > 0$, we choose

$$d = \#\{j \in \{1, 2, ..., p\}||\beta_j| \geq \delta\}.$$

Let $S = \{j : |\beta_j| \geq \delta\}$ and $S^c = \{j : |\beta_j| < \delta\}$. Applying Lemma 1 in Negahban et al. (2012), if $\lambda \geq 2|n^{-1}\sum_{i=1}^n x_ie_i|_\infty$, it holds that,

$$|\hat{\Delta}_{S^c}|_1 \leq 3|\hat{\Delta}_S|_1 + 4\sum_{j\in S^c}|\beta_j|.$$

We thus have

$$|\hat{\Delta}|_1 \leq |\hat{\Delta}_S|_1 + |\hat{\Delta}_{S^c}|_1 \leq 4|\hat{\Delta}_S|_1 + 4\sum_{j\in S^c}|\beta_j| \leq 4\sqrt{d}|\hat{\Delta}_S|_2 + 4\sum_{j\in S^c}|\beta_j|.$$

If follows that

$$\sum_{j \in S^c} |\beta_j| \le \delta \sum_{j \in S^c} \left(\frac{|\beta_j|}{\delta} \right)^{\theta} \le \delta^{1-\theta} K_\theta. \tag{C.10}$$

Thus

$$|\hat{\Delta}|_1 \le 4\sqrt{d} |\hat{\Delta}_S|_2 + 4\delta^{1-\theta} K_\theta.$$

On the other hand, we have

$$d \le \sum_{j \in S^c} \left(\frac{|\beta_j|}{\delta} \right)^{\theta} \le \delta^{-\theta} K_\theta. \tag{C.11}$$

Suppose $|\hat{\Delta}|_2 \ge c_1 \sqrt{K_\theta} (\lambda/\kappa)^{1-\theta/2}$ for some constant $c_1 > 0$. Then by (C.10) and (C.11), setting $\delta = \lambda/\kappa$,

$$\begin{aligned} |\hat{\Delta}|_1 &\le 4\sqrt{d} |\hat{\Delta}_S|_2 + 4\delta^{1-\theta} K_\theta \\ &\le 4\sqrt{K_\theta} \left(\frac{\lambda}{\kappa} \right)^{-\theta/2} |\hat{\Delta}|_2 + 4 \left(\frac{\lambda}{\kappa} \right)^{1-\theta} K_\theta \\ &\le 4(1 + c_1^{-1}) \sqrt{K_\theta} \left(\frac{\lambda}{\kappa} \right)^{-\theta/2} |\hat{\Delta}|_2. \end{aligned}$$

Recall $\lambda_{\min}(\Sigma) \ge \kappa > 0$. If $32(1 + c_1^{-1})^2 \eta_1 K_\theta \omega \lambda^{-\theta} \le \kappa^{1-\theta}$, we will have,

$$\mathsf{P} \left(\hat{\Delta}' \hat{\Sigma} \hat{\Delta} \ge \frac{1}{2} \kappa |\hat{\Delta}|_2^2 \right) \ge 1 - C_1 (\log p)^{-\gamma/2} - C_2 p^{-C_3}.$$

An application of Lemma 5 shows that for constants $c_2, c_3 > 0$, if $\lambda \ge 2|n^{-1} \sum_{i=1}^n x_i e_i|_\infty$, with probability at least $1 - C_1 (\log p)^{-\gamma/2} - C_2 p^{-C_3}$,

$$\begin{aligned} |\hat{\Delta}|_2 &\le c_2 \sqrt{K_\theta} \left(\frac{\lambda}{\kappa} \right)^{1-\theta/2}, \\ |\hat{\Delta}|_1 &\le c_3 K_\theta \left(\frac{\lambda}{\kappa} \right)^{1-\theta}. \end{aligned}$$

When $|\hat{\Delta}|_2 \le c_1 \sqrt{K_\theta} (\lambda/\kappa)^{1-\theta/2}$ for some constant $c_1 > 0$. Then by (C.10) and (C.11), setting $\delta = \lambda/\kappa$, we can still obtain

$$\begin{aligned} |\hat{\Delta}|_1 &\le 4\sqrt{d} |\hat{\Delta}_S|_2 + 4\delta^{1-\theta} K_\theta \\ &\le 4\sqrt{K_\theta} \left(\frac{\lambda}{\kappa} \right)^{-\theta/2} |\hat{\Delta}|_2 + 4 \left(\frac{\lambda}{\kappa} \right)^{1-\theta} K_\theta \\ &\le 4(1 + c_1) K_\theta \left(\frac{\lambda}{\kappa} \right)^{1-\theta}. \end{aligned}$$

Therefore, with probability at least $1 - C_1 (\log p)^{-\gamma/2} - C_2 p^{-C_3} - C_4 (\log p)^{-\tau}$, we have bounds (18) and (19). $\square$

D Proof of Theorem 2

Proof. Applying Theorem 1 with $\theta = 0$, with probability at least $1 - C_1(\log p)^{-\gamma/2} - C_2 p^{-C_3} - C_4(\log p)^{-\tau}$, we have

$$\begin{aligned} |\hat{\beta} - \beta|_2 &\lesssim \sqrt{s}\lambda/\kappa, \\ |\hat{\beta} - \beta|_1 &\lesssim s\lambda/\kappa. \end{aligned}$$

Since $s = K_\theta$, $s\omega \lesssim 1$ implies that

$$n \gtrsim M_X^4 s^2 \log p + s^{1/(1-2\nu/\gamma)} (\log p)^{3/(2-4\nu/\gamma)} \| |\mathbf{x}_{\cdot}|_\infty \|_{\gamma,\alpha_X}^{2/(1-2\nu/\gamma)}.$$

Recall the events

$$\begin{aligned} \mathcal{A} &= \{|\hat{\Sigma} - \Sigma|_\infty \leq a\} = \{\max_{j,k} |\hat{\sigma}_{jk} - \sigma_{jk}| \leq a\}, \\ \mathcal{B} &= \{n^{-1} \left|X^T e\right|_\infty \leq \lambda/2\}. \end{aligned}$$

Since $\hat{\beta}$ minimizes equation (2), we have

$$\frac{1}{2}|Y - X\hat{\beta}|_2^2 + \lambda|\hat{\beta}|_1 \leq \frac{1}{2}|Y - X\beta|_2^2 + \lambda|\beta|_1. \tag{D.1}$$

After some algebra, this reduces to

$$(\hat{\beta} - \beta)\hat{\Sigma}(\hat{\beta} - \beta) + \lambda|\hat{\beta}|_1 \leq 2e^T X(\hat{\beta} - \beta)/n + \lambda|\beta|_1 \tag{D.2}$$

On the event $\mathcal{B}$, the above inequality implies that

$$0 \leq (\hat{\beta} - \beta)\hat{\Sigma}(\hat{\beta} - \beta) \leq \frac{3}{2}\lambda|\hat{\beta}_J - \beta_J|_1 - \frac{1}{2}\lambda|\hat{\beta}_{J^c}|_1 \tag{D.3}$$

Then inequality (D.3) implies that

$$\frac{1}{2}\lambda|\hat{\beta} - \beta|_1 + (\hat{\beta} - \beta)\hat{\Sigma}(\hat{\beta} - \beta) \leq 2\lambda|\hat{\beta}_J - \beta_J|_1 \leq 2\lambda\sqrt{s}|\hat{\beta}_J - \beta_J|_2 \tag{D.4}$$

So (22) follow on the event $\mathcal{A} \cap \mathcal{B}$. □

E Proof of Theorem 3

Proof. Reall $|\Sigma_{11}^{-1}|_2 = 1/N_1$ and let $|\hat{\Sigma}_{11}^{-1}|_2 = 1/N_2$. Without loss of generality, let $J = \mathit{support}(\beta) = \{1, ..., s\}$. Let $X = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)'$ and denote by $X_{(1)}$ and $X_{(2)}$ the first s and last $p - s$ columns of X. Denote $W_n = \sum_{i=1}^n x_i e_i$ and $W_n(1)$, $x_{i,(1)}$, $\beta_{(1)}$ and $W_n(2)$, $x_{i,(2)}$, $\beta_{(2)}$ the first s and last $p - s$ entries of W_n, x_i and β, respectively. Define b =sign$(\beta_{(1)})$. Let

$$\begin{aligned} B &= (\frac{1}{n}X_{(1)}^T X_{(1)})^{-1} \left[\frac{1}{n}X_{(1)}^T e - \lambda b\right], \\ D_k &= X_{(2),k}^T \left\{X_{(1)}(X_{(1)}^T X_{(1)})^{-1}\lambda b - \left[X_{(1)}(X_{(1)}^T X_{(1)})^{-1}X_{(1)}^T - I\right]\frac{e}{n}\right\}, \end{aligned}$$

where $X_{(2),k} = (x_{1k}, ..., x_{nk})^T$ denote the k-th columns of X and $s+1 \le k \le p$. Denote the j-th element of B as B_j.

By rearranging terms, it is easy to see that the events

$$\mathcal{B} = \{\max_{1\le j\le s} |B_j| < L\}, \tag{E.1}$$

$$\mathcal{D} = \{\max_{s+1\le k\le p} |D_k| < \lambda\}, \tag{E.2}$$

are sufficient to guarantee that conditions in Lemma 4 hold. Then $\mathsf{P}(\hat{\beta} \neq_s \beta) \le \mathsf{P}(\mathcal{B}^c) + \mathsf{P}(\mathcal{D}^c)$.

We first analyze the event $\mathcal{D}$. Recall $\mathsf{E}(x_{ik}|X_{(1)}, e) = [\Sigma_{21}\Sigma_{11}^{-1} x_{i,(1)}]_k$ and $z_{ik} = x_{ik} - \mathsf{E}(x_{ik}|X_{(1)}, e)$ for $s+1 \le k \le p$. Let $\omega_1 = X_{(1)}(X_{(1)}^T X_{(1)})^{-1}\lambda b$, $\omega_2 = [I - X_{(1)}(X_{(1)}^T X_{(1)})^{-1} X_{(1)}^T]e/n$ and $\omega = \omega_1 + \omega_2$. Denote $Z_k = (z_{1k}, ..., z_{nk})^T$, $U_k = Z_k^T \omega$ and $\mu_k = \mathsf{E}(X_{(2),k}^T \omega | X_{(1)}, e)$. Note that $\mathsf{E} Z_k = 0$ and $\omega_1^T \omega_2 = 0$. Then by the irrepresentable condition,

$$\begin{aligned} \max_{s+1\le k\le p} |D_k| &= \max_{s+1\le k\le p} |\mu_k + U_k| \\ &\le \max_{s+1\le k\le p} [|\mu_k| + |U_k|] \\ &\le (1-\eta)\lambda + \max_{s+1\le k\le p} |U_k|. \end{aligned}$$

From this inequality, we have

$$\{\max_{s+1\le k\le p} |U_k| < \eta\lambda\} \subset \{\max_{s+1\le k\le p} |D_k| < \lambda\}.$$

Define the events

$$\mathcal{A}_1 = \{|\hat{\Sigma}_{11} - \Sigma_{11}|_\infty \le a\} = \{\max_{1\le j,k\le s} |\hat{\sigma}_{jk} - \sigma_{jk}| \le a\}, \tag{E.3}$$

$$\mathcal{A}_2 = \{n^{-1} e_i^2 \le 2\sigma\}, \tag{E.4}$$

$$\mathcal{T} = \{|\omega|_2^2 \le \delta_*\}. \tag{E.5}$$

By Lemma 3, on the event $\mathcal{A}_1$ with $a = N_1/(2s)$,

$$N_2 = \inf_{|\zeta|_2 = 1} \zeta^T \hat{\Sigma}_{11} \zeta > \inf_{|\zeta|_2=1} \zeta^T \Sigma_{11} \zeta - sa = \frac{N_1}{2}.$$

By Lemma 1,

$$\mathsf{P}\left(\left|\sum_{i=1}^n (e_i^2 - \sigma)\right| \ge n\sigma\right) \le \frac{n\|e.\|_{q,\alpha_e}^q}{n^q \sigma^q} + \exp\left(-\frac{n\sigma^2}{\|e.\|_{2,\alpha_e}^2}\right) := P_2$$

Denote $P_1 = \mathsf{P}(\mathcal{A}^c)$ with $a = N_1/(2s)$. We know

$$\omega_1^T \omega_1 = \lambda^2 b^T (X_{(1)}^T X_{(1)})^{-1} b \le \frac{\lambda^2 s}{n N_2},$$

and

$$\omega_2^T \omega_2 \le \frac{e^T e}{n^2}.$$

Thus, we have

$$\mathsf{P}(\mathcal{T}^c) \le \mathsf{P}\left(\omega_1^T\omega_1 \ge \frac{2\lambda^2 s}{nN_1}\right) + \mathsf{P}\left(\omega_2^T\omega_2 \ge 2n\sigma\right) \le P_1 + P_2.$$

By Lemma 2, if $\eta\lambda \gtrsim \sqrt{\delta_* \log p}\Psi_{2,\alpha_X,(2)} + n^{(\iota-1)/\gamma}\delta_*^{1/2}(\log p)^{3/2}\| |Z_\cdot|_\infty \|_{\gamma,\alpha_X}$,

$$\mathsf{P}\left(\max_{s+1\le k\le p}|U_k| \ge \eta\lambda \mid \mathcal{T}\right) \le C_1(\log(p-s))^{-\gamma} + C_2(p-s)^{-C_3} := P_3.$$

By the total probability rule, we have

$$\mathsf{P}(\mathcal{D}^c) \le \mathsf{P}\left(\max_{s+1\le k\le p}|U_k| \ge \eta\lambda \mid \mathcal{T}\right) + \mathsf{P}(\mathcal{T}^c) \le P_1 + P_2 + P_3.$$

Now we analyze the event $\mathcal{B}$. Note that $|\hat{\Sigma}_{11}^{-1}b|_\infty \le \sqrt{s}|\hat{\Sigma}_{11}^{-1}|_2 = \sqrt{s}/N_2$. Recall $\lambda \le nN_1L/(4\sqrt{s})$. On the event $\mathcal{A}$, $nL - \lambda|[\hat{\Sigma}_{11}^{-1}b]_j| \ge nL(1-N_1/(4N_2)) \ge \sqrt{n}L/2$ for all $1 \le j \le s$. Simple application of the Cauchy inequality shows that

$$\sup_{|\zeta|_2=1}\zeta^T\hat{\Sigma}_{11}^{-1}W_n(1) \le \frac{1}{N_2}\sqrt{\sum_{j=1}^s(\sum_{i=1}^n x_{ij}e_i)^2}.$$

This yields

$$\begin{aligned}
\mathcal{B} &= \bigcap_{j=1}^s\{|[\hat{\Sigma}_{11}^{-1}W_n(1)]_j| < \frac{1}{2}nL\} \\
&= \{\sup_{|\zeta|_2=1}\zeta^T\hat{\Sigma}_{11}^{-1}W_n(1) < \frac{1}{2}nL\} \\
&\supseteq \left\{\sqrt{\sum_{j=1}^s(\sum_{i=1}^n x_{ij}e_i)^2} < \frac{1}{2}nLN_2\right\} \\
&\supseteq \left\{\max_{1\le j\le s}\left|\sum_{i=1}^n x_{ij}e_i\right| < \lambda\right\}\bigcap\left\{|\hat{\Sigma}_{11} - \Sigma_{11}|_\infty \le \frac{N_1}{2s}\right\}.
\end{aligned}$$

Thus,

$$\mathsf{P}(\mathcal{B}^c) \le \mathsf{P}(|W_n(1)|_\infty \ge \lambda) + P_1.$$

By carrying out similar procedures as those in the proof of Theorem 1, we can control the probability P_1 and $\mathsf{P}(|W_n(1)|_\infty \ge \lambda)$. Then (31) follows. □

F Proof of Proposition 1

Proof. Let $\gamma_l = \mathsf{E}y_iy_{i-l}$. Set the candidate lags of this AR(2) model as d. Since $\gamma_0 = 1$, we have

$$\Sigma_{11} = \begin{pmatrix} 1 & \gamma_1 \\ \gamma_1 & 1 \end{pmatrix},$$

and

$$\Sigma_{21} = \begin{pmatrix} \gamma_2 & \gamma_1 \\ \cdots & \cdots \\ \gamma_{d-1} & \gamma_{d-2} \end{pmatrix}.$$

Basic calculation shows that

$$\Sigma_{11}^{-1} = \begin{pmatrix} \frac{1}{1-\gamma_1^2} & -\frac{\gamma_1}{1-\gamma_1^2} \\ -\frac{\gamma_1}{1-\gamma_1^2} & \frac{1}{1-\gamma_1^2} \end{pmatrix},$$

and

$$\begin{aligned} \gamma_1 &= \frac{\phi_1}{1-\phi_2}, \\ \gamma_l &= \phi_1\gamma_{l-1} + \phi_2\gamma_{l-2}, \end{aligned}$$

for $2 \leq l \leq d$.

We first consider the case $\phi_1 > 0$ and $\phi_2 > 0$. Then the Strong Irrepresentable Condition

$$|\Sigma_{21}\Sigma_{11}^{-1}\text{sign}(\beta_{(1)})|_\infty = \max_{2\leq j\leq d-1} \frac{\gamma_j}{1-\gamma_1^2} - \frac{\gamma_{j-1}\gamma_1}{1-\gamma_1^2} - \frac{\gamma_j\gamma_1}{1-\gamma_1^2} + \frac{\gamma_{j-1}}{1-\gamma_1^2} < 1$$

For $j = 2$, it can be shown that

$$\frac{\gamma_j}{1-\gamma_1^2} - \frac{\gamma_{j-1}\gamma_1}{1-\gamma_1^2} - \frac{\gamma_j\gamma_1}{1-\gamma_1^2} + \frac{\gamma_{j-1}}{1-\gamma_1^2} < 1$$

is equivalent to $\phi_1 + \phi_2 < 1$. Then $\gamma_1 < 1$ and $\gamma_j < \gamma_{j-1}$ for all $j \geq 1$. Thus, we have, $|\Sigma_{21}\Sigma_{11}^{-1}\text{sign}(\beta_{(1)})|_\infty < 1$ is equivalent to $\phi_1 + \phi_2 < 1$.

Similarly, we can prove the cases $\phi_1 > 0, \phi_2 \leq 0$ and $\phi_1 \leq 0, \phi_2 > 0$ and $\phi_1 \leq 0, \phi_2 \leq 0$.

□

References

Marcelo C. Medeiros and Eduardo F. Mendes. L1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *J Econom*, 191(1):255 – 271, 2016. 2

Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statist. Sci.*, 27 (4):538–557, 11 2012. 2, 3

Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso). *IEEE Trans. Inf. Theory*, 55(5):2183–2202, 2009. 2

Wei-Biao Wu and Ying Nian Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Statist.*, 10(1):352–379, 2016. 1

Danna Zhang and Wei Biao Wu. Gaussian approximation for high dimensional time series. *Ann. Statist.*, 45(5):1895–1919, 2017. 1