Statistica Sinica: Supplement

# Supplementary Material for "A Bernstein-type Inequality for High Dimensional Linear Processes with Applications to Robust Estimation of Time Series Regressions"

Linbo Liu and Danna Zhang

*Department of mathematics, University of California, San Diego,*

In this supplementary material, we provide part of the simulation results and the technical proofs for all the results presented in the main body of the paper.

## S1. Simulation on Time Series Regression

In this section, we evaluate the finite sample performance of the modified $\ell_1$-regularized Huber estimator proposed in Section 3.1 compared with the regular Huber estimator. We generate the linear process $\{X_i\}_{i=1}^n$ from a VAR model,

$$X_{i+1} = AX_i + \epsilon_i,$$

where we consider a Toeplitz transition matrix $A = (\lambda^{|i-j|})$ with $\lambda = 0.5$ and further scaled by $2\lambda_{\max}(A)$ to ensure the stationarity of $(X_i)$. The innovation vectors $\epsilon_i$ have i.i.d. coordinates drawn from Student's $t$-distribution with $\mathrm{df} = 5$. We construct $\xi_i$ in (3.1) as

$$\xi_i = \sum_{k=0}^{\infty} b_k \eta_{i-k},$$

where $b_k = \rho^k$ with a $\rho$ drawn from $Unif(-0.8, 0.8)$ and $\eta_i$ follows a Student's $t$-distribution with $\mathrm{df} = 5$ and is independent of $\epsilon_i$. Recall the linear model

$$Y_i = X_i^\top \beta^* + \xi_i.$$

We choose $\beta^* = (1, 1, \dots, 1, 0, 0, \dots, 0)^\top$ with $s$ elements of value 1 and $p - s$ elements of value 0 for $s = 2\lfloor \log(p) \rfloor$. In weight function

$$w(x) = \min\left\{1, \frac{b}{|Bx|_2}\right\},$$

we select $b = 5, 15, 50, 100$ and $B = I_p$. The simulation results with different $n, p$ are summarized in Table 1. We observe that neither small $b$ nor large $b$ can be consistently beneficial. The weight function with a small $b$ shrinks the covariates too aggressively, hence discards too much information of the tail behavior of the linear process. Large $b$ makes the shrinkage less effective, hence approaches the Huber estimator.

## S2. Proofs of Results in Section 2

In this section, we provide the proofs of the results presented in Section 2.

Table 1: Experiment results on time series regression.

| $(n, p)$ | $(100, 10)$ | $(100, 100)$ | $(100, 500)$ | $(100, 1000)$ |
|---|---|---|---|---|
| Huber | 0.77 (0.090) | 3.64 (0.120) | 4.37 (0.079) | 5.65 (0.064) |
| Weighted Huber ($b = 5$) | 0.80 (0.043) | 4.25 (0.069) | 4.55 (0.082) | 5.89 (0.041) |
| Weighted Huber ($b = 15$) | 0.68 (0.035) | 3.15 (0.092) | 4.15 (0.035) | 5.28 (0.161) |
| Weighted Huber ($b = 50$) | 0.87 (0.042) | 3.24 (0.155) | 4.00 (0.090) | 5.11 (0.049) |
| Weighted Huber ($b = 100$) | 0.70 (0.086) | 3.70 (0.086) | 4.26 (0.135) | 5.30 (0.077) |

*Proof of Theorem 1.* We first define the filtration $\{\mathcal{F}_i\}$ with the $\sigma$-field $\mathcal{F}_i = \sigma(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_{i-1}, \dots)$, and the projection operator $P_j(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_j) - \mathbb{E}(\cdot|\mathcal{F}_{j-1})$. Conventionally it follows that $P_j(G(X_i)) = 0$ for $j \geq i + 1$. We can write

$$\sum_{i=1}^{n} G(X_i) - \mathbb{E}G(X_i) = \sum_{j=-\infty}^{n} \Big( \sum_{i=1}^{n} P_j(G(X_i)) \Big) =: \sum_{j=-\infty}^{n} L_j,$$

where $L_j = \sum_{i=1}^{n} P_j(G(X_i))$. By the Markov inequality, for any $\lambda > 0$,

$$\begin{aligned} & \mathbb{P}\bigg( \sum_{i=1}^{n} G(X_i) - \mathbb{E}G(X_i) \geq 2x \bigg) \leq \mathbb{P}\bigg( \sum_{j=-\infty}^{0} L_j \geq x \bigg) + \mathbb{P}\bigg( \sum_{j=1}^{n} L_j \geq x \bigg) \\ \leq \;& \mathrm{e}^{-\lambda x} \mathbb{E}\bigg[ \exp\bigg\{ \lambda \sum_{j=-\infty}^{0} L_j \bigg\} \bigg] + \mathrm{e}^{-\lambda x} \mathbb{E}\bigg[ \exp\bigg\{ \lambda \sum_{j=1}^{n} L_j \bigg\} \bigg]. \end{aligned} \tag{S2.1}$$

We shall bound the right-hand side of (S2.1) with a suitable choice of $\lambda > 0$. Observing that $\{L_j\}_{j \leq n}$ is a sequence of martingale differences with respect to $\{\mathcal{F}_j\}$, we firstly seek an upper bound on $\mathbb{E}[\mathrm{e}^{\lambda L_j} \big| \mathcal{F}_{j-1}]$. By the Lipschitz condition (2.6) and the boundedness of $G$, it follows that

$$
\begin{aligned}
|L_j| &\le \sum_{i=1\vee j}^{n} \min\left\{\left|\mathbb{E}\left[G(X_i)\middle|\mathcal{F}_j\right] - \mathbb{E}\left[G(X_i)|\mathcal{F}_{j-1}\right]\right|, 2M\right\} \\
&\le \sum_{i=1\vee j}^{n} \min\left\{g^\top |A_{i-j}| \mathbb{E}\left[|\boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_j'|\middle|\mathcal{F}_j\right], 2M\right\}, \qquad \text{(S2.2)}
\end{aligned}
$$

where $\varepsilon_j'$ is an i.i.d. copy of $\boldsymbol{\varepsilon}_j$. For notational convenience, we denote $b_i^\top = g^\top |A_i|$ and $\eta_j = \mathbb{E}(|\boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_j'|\big|\mathcal{F}_j)$. Then we have

$$
|L_j| \le 2M \sum_{i=1\vee j}^{n} \mathbb{I}(b_{i-j}^\top \eta_j \ge 2M) + \sum_{i=1\vee j}^{n} b_{i-j}^\top \eta_j \mathbb{I}(b_{i-j}^\top \eta_j \le 2M) =: I_j + II_j.
$$

For $j \le 0$ and $k \ge 2$, by the triangle inequality, it holds that

$$
\begin{aligned}
\mathbb{E}[|L_j|^k \big| \mathcal{F}_{j-1}] &\le \left[\left(\mathbb{E}[|I_j|^k \big|\mathcal{F}_{j-1}]\right)^{1/k} + \left(\mathbb{E}[|II_j|^k \big|\mathcal{F}_{j-1}]\right)^{1/k}\right]^k \\
&\le \left(\|I_j\|_k + \|II_j\|_k\right)^k. \qquad \text{(S2.3)}
\end{aligned}
$$

Moreover,

$$
\|I_j\|_k \le 2M \sum_{i=-j}^{\infty} \left\|\mathbb{I}(b_i^\top \eta_j \ge 2M)\right\|_k \le 2M \sum_{i=-j}^{\infty} \left[\mathbb{P}\left((b_i^\top \eta_j)^2 \ge (2M)^2\right)\right]^{1/k}. \qquad \text{(S2.4)}
$$

Recall the definitions of $\gamma$ and $\tau$. We have $|b_i|_1 \le \gamma \rho_0^{i/\tau}$, which implies

$$
\mathbb{E}[(b_i^\top \eta_j)^2] \le 2\sigma^2 |b_i|_1^2 \le 2\gamma^2 \sigma^2 \rho_0^{2i/\tau}, \text{ for all } j.
$$

By the Markov inequality, we obtain from (S2.4) that for $k \ge 2$,

$$
\|I_j\|_k \le 2M \left(\frac{\gamma\sigma}{\sqrt{2}M}\right)^{2/k} \frac{\rho_0^{-2j/k\tau}}{1 - \rho_0^{2/k\tau}}. \qquad \text{(S2.5)}
$$

In view of the fact $1-x \geq -x\log x$ for $x \in (0,1)$, we can further relax the bound in (S2.5). Applying the Stirling formula, for $k \geq 2$, we can obtain

$$\begin{aligned} \|I_j\|_k^k &\leq k^k \tau^k \rho_0^{-2/\tau} \left(\frac{M}{\log(1/\rho_0)}\right)^k \left(\frac{\gamma\sigma}{\sqrt{2}M}\right)^2 \rho_0^{-2j/\tau} \\ &\leq \frac{1}{2\sqrt{2\pi}} \left(\frac{\gamma\sigma}{\rho_0 M}\right)^2 k!\tau^k \left(\frac{\mathrm{e}M}{\log(1/\rho_0)}\right)^k \rho_0^{-2j/\tau}. \end{aligned}$$

Define the constants

$$C_1 = \frac{1}{2\sqrt{2\pi}}\rho_0^{-2}, \quad \text{and} \quad C_2 = \frac{\mathrm{e}}{\log(1/\rho_0)}.$$

Then we can simply write

$$\|I_j\|_k^k \leq C_1 k! \tau^k C_2^k M^{k-2} \gamma^2 \sigma^2 \rho_0^{-2j/\tau}. \tag{S2.6}$$

Analogously, for $k \geq 2$, we can also get

$$\|II_j\|_k^k \leq \Big[\sum_{i=-j}^{\infty} \big\{\mathbb{E}\big[(b_i^\top \eta_j)^2\,(2M)^{k-2}\big]\big\}^{1/k}\Big]^k \leq C_1 k! \tau^k C_2^k M^{k-2} \gamma^2 \sigma^2 \rho_0^{-2j/\tau}. \tag{S2.7}$$

By (S2.3), (S2.6) and (S2.7), we have

$$\mathbb{E}[|L_j|^k \big| \mathcal{F}_{j-1}] \leq C_1 k! \tau^k (C_2')^k M^{k-2} \gamma^2 \sigma^2 \rho_0^{-2j/\tau}, \tag{S2.8}$$

where $C_2' = 2C_2 = 2\mathrm{e}/\log(1/\rho_0)$. Now we are ready to derive an upper bound for $\mathbb{E}[\mathrm{e}^{\lambda L_j} \big| \mathcal{F}_{j-1}]$. By the Taylor expansion, we have

$$\mathbb{E}[\mathrm{e}^{\lambda L_j} | \mathcal{F}_{j-1}] = 1 + \mathbb{E}[\lambda L_j | \mathcal{F}_{j-1}] + \sum_{k=2}^{\infty} \frac{1}{k!} \mathbb{E}[\lambda^k L_j^k | \mathcal{F}_{j-1}].$$

Notice that $\mathbb{E}[L_j \big| \mathcal{F}_{j-1}] = 0$. For $0 < \lambda < (C_2' M\tau)^{-1}$, we have

$$\begin{aligned} \mathbb{E}[\mathrm{e}^{\lambda L_j} \big| \mathcal{F}_{j-1}] &\leq 1 + C_1 M^{-2} \gamma^2 \sigma^2 \rho_0^{-2j/\tau} \sum_{k=2}^{\infty} \Big( C_2' M \tau \lambda \Big)^k \\ &\leq \exp\left\{ \frac{C_1' \gamma^2 \sigma^2 \tau^2 \rho_0^{-2j/\tau} \lambda^2}{1 - C_2' M \tau \lambda} \right\}, \end{aligned}$$

where the constant

$$C_1' = C_1 (C_2')^2 = \frac{1}{2\sqrt{2\pi}} \left( \frac{2\mathrm{e}}{\rho_0 \log(1/\rho_0)} \right)^2,$$

Thus, recursively conditioning on $\mathcal{F}_0, \mathcal{F}_{-1}, \ldots$, we have for $0 < \lambda < (C_2'\tau)^{-1}$,

$$\begin{aligned} \mathbb{P}\bigg( \sum_{j=-\infty}^{0} L_j \geq x \bigg) &\leq \mathrm{e}^{-\lambda x} \mathbb{E}\left[ \exp\left\{ \lambda \sum_{j=-\infty}^{0} L_j \right\} \right] \\ &\leq \mathrm{e}^{-\lambda x} \exp\left\{ \frac{C_1' \gamma^2 \sigma^2 \tau^2 (1 - \rho_0^{2/\tau})^{-1} \lambda^2}{1 - C_2' M \tau \lambda} \right\}. \end{aligned}$$

Specifically, choosing $\lambda = x[C_2' M\tau x + 2C_1' \gamma^2 \sigma^2 \tau^2 (1 - \rho_0^{2/\tau})^{-1}]^{-1}$ yields

$$\begin{aligned} \mathbb{P}\bigg( \sum_{j=-\infty}^{0} L_j \geq x \bigg) &\leq \exp\left\{ -\frac{x^2}{4C_1' \gamma^2 \sigma^2 \tau^2 (1 - \rho_0^{2/\tau})^{-1} + 2C_2' M \tau x} \right\} \\ &\leq \exp\left\{ -\frac{x^2}{2C_1' \gamma^2 \sigma^2 \rho_0^{-2} (\log(1/\rho_0))^{-1} \tau^3 + 2C_2' M \tau x} \right\} \\ &= \exp\left\{ -\frac{x^2}{C_1'' \tau^3 \gamma^2 \sigma^2 + 2C_2' M \tau x} \right\}, \qquad \text{(S2.9)} \end{aligned}$$

where $C_1'' = 2C_1' \rho_0^{-2} (\log(1/\rho_0))^{-1}$. We can deal with $L_j$ for $j \geq 1$ by similar arguments and obtain

$$\mathbb{E}[\mathrm{e}^{\lambda L_j} \big| \mathcal{F}_{j-1}] \leq \exp\left\{ \frac{C_1' \gamma^2 \sigma^2 \tau^2 \lambda^2}{1 - C_2' M \tau \lambda} \right\} \text{ for } j \geq 1.$$

In a similar way as deriving (S2.9), it follows that

$$\mathbb{P}\left(\sum_{j=1}^{n} L_j \geq x\right) \leq \exp\left\{-\frac{x^2}{C_1''\gamma^2\sigma^2\tau^2 n + 2C_2' M\tau x}\right\}. \tag{S2.10}$$

Combining (S2.1), (S2.9) and (S2.10), we have

$$\mathbb{P}\Big(\sum_{i=1}^{n} G(X_i) - \mathbb{E}[G(X_i)] \geq x\Big) \leq 2\exp\left\{-\frac{x^2}{4C_1''\tau^2(\tau \vee n) + 4C_2' M\tau x}\right\},$$

which implies (2.7) for $\tau \leq n$. □

*Proof of Theorem 2.* We follow the starting steps when proving Theorem 1. Without assuming $G$ bounded, we have

$$|L_j| \leq \sum_{i=1\vee j}^{n} g^\top |A_{i-j}| \mathbb{E}\left[|\boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_j'||\mathcal{F}_j\right] = \sum_{i=1\vee j}^{n} b_{i-j}^\top \eta_j =: d_j^\top \eta_j.$$

For $j \leq -\tau$, we have

$$|d_j|_1 \leq \sum_{i=1}^{n} |b_{i-j}|_1 \leq \gamma \frac{\rho_0^{1/\tau}}{1-\rho_0^{1/\tau}} \cdot \rho_0^{-j/\tau} \leq (\log(1/\rho_0))^{-1}\gamma\tau\rho_0^{-j/\tau}. \tag{S2.11}$$

Note that

$$\mathbb{E}[e^{\lambda|L_j|}|\mathcal{F}_{j-1}] \leq \mathbb{E}[e^{\lambda d_j^\top \eta_j}|\mathcal{F}_{j-1}] = \mathbb{E}[e^{\lambda d_j^\top \eta_j}] \leq \mathbb{E}[e^{\lambda d_j^\top(|\boldsymbol{\varepsilon}_j| + |\boldsymbol{\varepsilon}_j'|)}]. \tag{S2.12}$$

Let $\lambda^* = c_0(\log(1/\rho_0))(\gamma\tau)^{-1}$ and $Y_j = \lambda^* d_j^\top(|\boldsymbol{\varepsilon}_j| + |\boldsymbol{\varepsilon}_j'|)\rho_0^{j/\tau}$. By (S2.11) and (S2.12), it follows that for any $j \leq -\tau$, $\mathbb{E}e^{Y_j} \leq \theta^2$ and

$$\mathbb{E}[e^{\lambda^*|L_j|} - 1|\mathcal{F}_{j-1}] \;\leq\; \mathbb{E}e^{Y_j\rho_0^{-j/\tau}} - 1 = \int_0^\infty \rho_0^{-j/\tau} e^{x\rho_0^{-j/\tau}} \mathbb{P}(Y_j \geq x)dx$$

$$
\begin{aligned}
&\leq \int_0^\infty \rho_0^{-j/\tau} e^{x\rho_0^{-j/\tau}} e^{-x}\theta^2 dx \\
&\leq \frac{\rho_0^{-j/\tau}\theta^2}{1-\rho_0^{-j/\tau}} \leq \frac{\rho_0^{-j/\tau}\theta^2}{1-\rho_0}.
\end{aligned}
$$

Since $\mathbb{E}[L_j|\mathcal{F}_j] = 0$, for any $0 < \lambda \leq \lambda^*$,

$$
\begin{aligned}
\mathbb{E}[e^{\lambda L_j} - 1|\mathcal{F}_{j-1}] &= \mathbb{E}[e^{\lambda L_j} - \lambda L_j - 1|\mathcal{F}_{j-1}] \\
&\leq \mathbb{E}[e^{\lambda |L_j|} - \lambda |L_j| - 1|\mathcal{F}_{j-1}] \\
&\leq \mathbb{E}[e^{\lambda^* |L_j|} - \lambda^* |L_j| - 1|\mathcal{F}_{j-1}] \cdot \lambda^2/(\lambda^*)^2 \\
&\leq \mathbb{E}[e^{\lambda^* |L_j|} - 1|\mathcal{F}_{j-1}] \cdot \lambda^2/(\lambda^*)^2,
\end{aligned}
$$

in view of $e^x - x \leq e^{|x|} - |x|$ for any $x$ and when $x > 0$, $(e^{\lambda x} - \lambda x - 1)/\lambda^2$ is increasing with $\lambda \in (0, \infty)$. Using $1 + x \leq e^x$, we have

$$
\begin{aligned}
\mathbb{E}[e^{\lambda L_j}|\mathcal{F}_{j-1}] &\leq 1 + \mathbb{E}[e^{\lambda^* |L_j|} - 1|\mathcal{F}_{j-1}] \cdot \lambda^2/(\lambda^*)^2 \\
&\leq 1 + C_1 \rho_0^{-j/\tau} \gamma^2 \tau^2 \theta^2 \lambda^2 \leq \exp\left\{ C_1 \rho_0^{-j/\tau} \gamma^2 \tau^2 \theta^2 \lambda^2 \right\}.
\end{aligned}
$$

where $C = c_0^{-2}(\log(1/\rho_0))^{-2}/(1-\rho)$, which implies that

$$
\mathbb{P}\left( \sum_{j=-\infty}^{-\tau} L_j \geq x \right) \leq \mathrm{e}^{-\lambda x} \mathbb{E}\left[ \exp\left\{ \lambda \sum_{j=-\infty}^{-1} L_j \right\} \right] \leq \mathrm{e}^{-\lambda x} \exp\left\{ C_1 \gamma^2 \tau^3 \theta^2 \lambda^2 \right\}.
$$

with $C_1 = C(\log(1/\rho_0))^{-1}(\rho_0)^{-2}$. For the cases when $j > -\tau$, we use the bound $|d_j|_1 \leq (\rho_0 \log(1/\rho_0))^{-1}\gamma\tau$ and obtain $\mathbb{E}[e^{\lambda L_j}|\mathcal{F}_{j-1}] \leq 1 + C_2 \gamma^2 \tau^2 \theta^2 \lambda^2$ for $C_2 = C/\rho_0^2$ and

$$
\mathbb{P}\left( \sum_{j=-\tau+1}^{n} L_j \geq x \right) \leq \exp\left\{ -\lambda x + C_2 (n+\tau) \gamma^2 \tau^2 \theta^2 \lambda^2 \right\}. \tag{S2.13}
$$

Therefore (2.9) follows by choosing

$$\lambda = \min\left\{\lambda^*,\ \frac{x}{2C_1\gamma^2\tau^3\theta^2},\ \frac{x}{2C_2(n+\tau)\gamma^2\tau^2\theta^2},\right\}.$$

□

By a slight modification of the Lipschitz condition (S2.2), we can develop some ancillary results in Corollar 1 and Corollary 2, that can be useful in estimating time series regression models. The proof follows similarly from that of Theorem 1 without extra technical difficulty.

**Corollary 1.** *Consider the same setting of the model as in Theorem 1. Let* $G : \mathbb{R}^{2p} \to \mathbb{R}$ *be a function with* $|G(u)| \leq M$ *for all* $u \in \mathbb{R}^{2p}$*. Suppose there exists a vector* $g = (g_1, \ldots, g_{2p})^\top$ *with* $g_i \geq 0$ *for* $1 \leq i \leq 2p$ *and* $\sum_{i=1}^{2p} g_i = 1$ *such that*

$$|G(u) - G(v)| \leq \sum_{i=1}^{2p} g_i |u_i - v_i|,\ \textit{for all } u, v \in \mathbb{R}^{2p}.$$

*Then for any* $x > 0$*, we have*

$$\mathbb{P}\Big(\sum_{i=1}^{n} G(X_i, X_{i-1}) - \mathbb{E}G(X_i, X_{i-1}) \geq x\Big) \leq 2\exp\left\{-\frac{x^2}{C_1' n\sigma^2\gamma^2\tau^2 + C_2'\tau M x}\right\}. \tag{S2.14}$$

*Proof of Corollary 1.* It follows from the fact that the $(2p)$-dimensional process $(X_i^\top, X_{i-1}^\top)^\top$ is also linear and satisfies the condition (2.3) with $\gamma$ multiplied by a constant depending on $\rho_0$ only. □

**Corollary 2.** *Consider the same setting of the model as in Theorem 1. Let $G : \mathbb{R}^p \to \mathbb{R}$ be a function with $|G(u)| \le M$ for all $u \in \mathbb{R}^p$. Assume that*

$$|G(u) - G(v)| \le |u - v|_2, \text{ for all } u, v \in \mathbb{R}^p.$$

*Assume that $\log p > 1$ and $\tau \log p \le n$. Then for any $x > 0$, we have*

$$\begin{aligned} & \mathbb{P}\Big(\sum_{i=1}^{n} G(X_i) - \mathbb{E}G(X_i) \ge x\Big) \\ \le \; & 2\exp\Big\{ - \frac{x^2}{C_1'' n(\sigma^2\gamma^2 + M^2)\tau^2(\log p)^2 + C_2''\tau M(\log p)x}\Big\} \end{aligned} \quad \text{(S2.15)}$$

*Proof of Corollary 2.* With a different Lipschitz condition on $G$, the step (S2.2) becomes

$$|L_j| \le \sum_{i=1\vee j}^{n} \min\{|A^{i-j}\eta_j|_2, 2M\} \le \sum_{i=1\vee j}^{n} \min\{\gamma\rho_0^{(i-j)/\tau}|\eta_j|_2, 2M\}.$$

Note that $\mathbb{E}|\eta_j|_2^2 \le 2p\sigma^2$. For $j \le -n_0$ where $n_0 = \lceil \tau \log p / \log(1/\rho_0) \rceil$, by similar arguments in deriving (S2.9), it can be obtained that

$$\mathbb{P}\Big(\sum_{j=-\infty}^{-n_0} L_j \ge x\Big) \le \exp\Big\{ - \frac{x^2}{C_1\tau^3 + C_2 M\tau x}\Big\}. \quad \text{(S2.16)}$$

For $j > -n_0$, we have

$$|L_j| \le 2n_0 M + \sum_{i=j+n_0}^{\infty} \min\{\gamma\rho_0^{(i-j)/\tau}|\eta_j|_2, 2M\}.$$

Similarly as (S2.8), we can get

$$\mathbb{E}[|L_j|^k|\mathcal{F}_{j-1}] \;\; \le \;\; 2^k[(2n_0M)^k + C_1' k!\tau^k (C_2')^k M^{k-2}\gamma^2\sigma^2]$$

$$\leq \quad C_3(C_4 n_0 M)^k k!(1 + M^{-2}\gamma^2\sigma^2),$$

which further implies

$$\mathbb{E}\Big[\exp\Big\{\lambda \sum_{j=-s+1}^{n} L_j\Big\}\Big] \leq \exp\Big\{\frac{C_3 C_4^2 (M^2 + \gamma^2\sigma^2) n_0^2 (n_0 + n)\lambda^2}{1 - C_4 n_0 M \lambda}\Big\},$$

and

$$\mathbb{P}\left(\sum_{j=-n_0+1}^{n} L_j \geq x\right) \leq \exp\Big\{-\frac{x^2}{C_3'(M^2 + \gamma^2\sigma^2) n_0^2 (n_0 + n) + C_4' M\tau(\log p)x}\Big\}.$$

Then (S2.15) follows in view of $n_0 \leq C_{\rho_0} n$. □

*Proof of Theorem 3.* Let $\hat{\mu}_j$ be the Huber estimator of $\mu_j$. Following similar arguments of proving Theorem 3.1 in Zhang (2021), for

$$R_{nj}(a) = \sum_{i=1}^{n}[\phi_\nu(X_{ij} - a) - \mathbb{E}\phi_\nu(X_{ij} - a)],$$

it can be obtained that for any $\delta > 0$ with $\nu^{-1}\delta \leq 1/2$,

$$\mathbb{P}(\hat{\mu}_j - \mu_j \geq \delta) \leq \mathbb{P}(R_{nj}(\mu_j + \delta) \geq n(\delta - 4\nu^{-1}\mu_2^2)).$$

By the Lipschitz continuity of the function $\phi_\nu$ and the uniform bound $|\phi_\nu(x)| \leq \nu$, applying Theorem 1 to $R_{nj}(\mu_j + \delta)$, it follows that

$$\mathbb{P}(R_{nj}(\mu_j + \delta) \geq y) \leq 2\exp\Big\{-\frac{y^2}{2C_1 n\tau^2\gamma^2 + C_2\tau\nu y}\Big\}.$$

Then it follows that

$$\mathbb{P}(\hat{\mu}_j - \mu_j \geq \delta) \leq 2x$$

by letting $n(\delta - 4\nu^{-1}\mu_2^2) = y = \tau\gamma\sqrt{2C_1 n \log(1/x)} + C_2\tau\nu\log(1/x)$ for $0 < x < 1/\mathrm{e}$. The requirement $\nu^{-1}\delta \le 1/2$ is met if we choose $\nu = \frac{2\mu^*}{\sqrt{C_2}}\sqrt{\frac{n}{\log(1/x)}}$ for any $\mu^* \ge \mu_2$ and impose the condition

$$(\sqrt{2C_1C_2}\gamma/\mu_2 + 4C_2)\tau\log(1/x) \le n.$$

For $\delta \le \delta_n = (\sqrt{2C_1}\gamma + 4\sqrt{C_2}\mu^*)\tau\sqrt{\frac{\log(1/x)}{n}}$, we have $\mathbb{P}(\hat{\mu}_j - \mu_j \ge \delta_n) \le 2x$. It can also be obtained that $\mathbb{P}(\hat{\mu}_j - \mu_j \le -\delta_n) \le 2x$ similarly. By letting $x = p^{-c-1}$, for some $c > 0$, it follows that

$$\mathbb{P}\Big(\max_{1\le j\le p} |\hat{\mu}_j - \mu_j| \ge \sqrt{c+1}(\sqrt{2C_1}\gamma + 4\sqrt{C_2}\mu^*)\tau\sqrt{\frac{\log p}{n}}\Big) \le 4p^{-c}.$$

which further implies (2.10). □

## S3. Proofs of Results in Section 3

This section includes all the proofs for the results on robust estimation of time series regressions presented in Section 3.

### S3.1 Proofs of Results in Section 3

Denote $L_n(\beta) = \frac{1}{n}\sum_{i=1}^n \Phi_\nu((Y_i - X_i^\top\beta)w(X_i))$ and $\phi_\nu(\cdot) = \Phi_\nu'(\cdot)$. Recall $b_0 = b/\lambda_{\min}(B)$ and $\kappa(B) = \lambda_{\max}(B)/\lambda_{\min}(B)$.

**Lemma 1** (Deviation bound)**.** *Let Assumptions (A1) (A2) (A3) in Section 3.1 be satisfied. Let* $\nu = c\sigma_\eta(n/\log p)^{1/2}$ *and* $\lambda = Cb_0\sigma_\eta(\log p/n)^{1/2}$ *for a* *sufficiently large* $C$*, with probability at least* $1 - 4p^{-c_1}$ *for some* $c_1 > 0$*, it holds that* $|\nabla L_n(\beta^*)|_\infty \leq \lambda$.

*Proof.* Consider the first component $\nabla L_{n1}(\beta^*)$ of $\nabla L_n(\beta^*)$. We have

$$\nabla L_{n1}(\beta^*) = \frac{1}{n}\sum_{i=1}^{n} \phi_\nu(\xi_i w(X_i))X_{i1}w(X_i).$$

Note that $|\phi_\nu(x) - \phi_\nu(y)| \leq |x - y|$ and $|\phi_\nu(\xi_i w(X_i))X_{i1}w(X_i)| \leq \nu b_0$. Conditioned on $\{X_i\}_{i=1}^n$, by Theorem 1, we have

$$\mathbb{P}\big(|\nabla L_{n1}(\beta^*) - \mathbb{E}[\nabla L_{n1}(\beta^*)]| \geq C' b_0 x \ |(X_i)_i\big) \leq 4p^{-c},$$

for $x = \sigma_\eta\sqrt{\log p/n} + \nu \log p/n$ and some constant $c > 1$. Hence by a union bound, with probability at least $1 - 4p^{-c_1}$ for $c_1 > 0$, it holds that

$$|\nabla L_n(\beta^*) - \mathbb{E}[\nabla L_n(\beta^*)]|_\infty \leq C' b_0 x.$$

As $\mathbb{E}|\phi_\nu(\xi_i w(X_i))| = \mathbb{E}[|\xi_i w(X_i)|\mathbf{1}(|\xi_i w(X_i)| > \nu)] \leq C_\rho \sigma_\eta^2 \nu^{-1}$, we have

$$|\mathbb{E}[\nabla L_{n1}(\beta^*)]| \leq \mathbb{E}|\nabla L_{n1}(\beta^*)| \leq C_\rho b_0 \sigma_\eta^2 \nu^{-1}. \tag{S3.1}$$

Therefore, choosing $\nu = c\sigma_\eta(n/\log p)^{1/2}$ and $\lambda = Cb_0\sigma_\eta\sqrt{\log p/n}$ ensures that $|\nabla L_n(\beta^*)|_\infty \leq \lambda$ with high probability.

□

**Lemma 2** (RSC condition)**.** *Let Assumptions (A1) (A2) (A3) be satisfied. Assume*

$$b_0(b_0 + \kappa(B)\gamma\sigma_\varepsilon)\tau\sqrt{s}\sqrt{(\log p)^3/n} \to 0.$$

*We have the following holds uniformly for all $\beta$, such that $|\Delta|_2 \le \nu/(2b_0)$ and $|\Delta_{S^c}|_1 \le 3|\Delta_S|_1$ with probability no less than $1-4p^{-c_2}$ that*

$$L_n(\beta) - L_n(\beta^*) - \nabla L_n(\beta^*)^\top(\beta-\beta^*) \ge \frac{1}{2}\lambda_{\min}(\mathbb{E}[\frac{w^2(X_i)}{2}X_iX_i^\top])|\beta-\beta^*|_2^2. \tag{S3.2}$$

*Proof.* Denote $S = \mathrm{supp}(\beta^*)$. We will show that with high probability, (S3.2) holds uniformly over the set

$$\mathcal{C} := \{\beta : |\beta-\beta^*| \le \frac{\nu}{2b_0}, |\beta_{S^c} - \beta^*_{S^c}|_1 \le 3|\beta_S - \beta^*_S|_1\},$$

Let $\mathcal{T}(\beta,\beta^*) = L_n(\beta) - L_n(\beta^*) - \nabla L_n(\beta^*)^\top(\beta-\beta^*)$, then it follows the same argument as Appendix B.3 in Loh (2021) that

$$\mathcal{T}(\beta,\beta^*) \ge \frac{1}{n}\sum_{i=1}^n \frac{1}{2}(w(X_i)X_i^\top(\beta-\beta^*))^2\mathbf{1}_{A_i},$$

where $A_i = \{\xi_i \le \nu/2\}$. Denote $\Gamma = \frac{1}{n}\sum_{i=1}^n \frac{w(X_i)^2}{2}X_iX_i^\top\mathbf{1}_{A_i}$. For any $u$ such that $|u|_2 \le 1$, we have

$$u^\top\Gamma u = \frac{1}{n}\sum_{i=1}^n \frac{1}{2}(u^\top X_i w(X_i))^2\mathbf{1}_{A_i}.$$

Notice that $\frac{1}{2}|(u^\top xw(x))^2 - (u^\top yw(y))^2| \le b_0(\kappa(B)+1)|x-y|_2$ and $|(u^\top xw(x))^2| \le b_0^2$. Conditioned on $\xi_i$, by Corollary 2 we have

$$\mathbb{P}(|u^\top\Gamma u - \mathbb{E}[u^\top\Gamma u]| \ge t|(\xi_i)_i) \le 4\exp\{-c_3 s\log p\},$$

where $t = Cb_0(b_0+\kappa(B)\gamma\sigma_\varepsilon)\tau\sqrt{s}\sqrt{(\log p)^3/n}$ for a sufficiently large $C$ such that $c_3 > 4$. Note that $t \to 0$ by assumption. Following the same spirit of the $\varepsilon$-net argument in lemma 15 of Loh and Wainwright (2012), we can obtain that

$$\left|v^\top\bigl(\Gamma-\mathbb{E}\Gamma\bigr)v\right|\le t,\ \forall\, v\in\mathbb{R}^p,\ |v|_0\le 2s,\ |v|_2\le 1,$$

holds with probability at least

$$1-4\exp\left\{2s\log 9+2s\log p-c_3 s\log p\right\}\ge 1-4p^{-c_2},$$

provided that $p\to\infty$ and a sufficiently large $c_3$. By Lemma 12 in Loh and Wainwright (2012), it further implies that

$$|v^\top(\Gamma-\mathbb{E}\Gamma)v|\le 27t\biggl(|v|_2^2+\frac{|v|_1^2}{s}\biggr),\ \forall v\in\mathbb{R}^p. \tag{S3.3}$$

Denote $\Delta=\beta-\beta^*$, then we have

$$\mathcal{T}(\beta,\beta^*)\ge\Delta^\top\Gamma\Delta\ge\mathbb{E}[\Delta^\top\Gamma\Delta]-27t(|\Delta|_2^2+\frac{|\Delta|_1^2}{s}). \tag{S3.4}$$

Moreover, as $\mathbb{E}|\xi_i|^2\le C_\rho\sigma_\eta^2$ and $\nu\to\infty$,

$$\begin{aligned}\mathbb{E}[\Delta^\top\Gamma\Delta]&=\mathbb{E}[\frac{w^2(X_i)}{2}(\Delta^\top X_i)^2]\cdot\mathbb{P}(|\xi_i|\le\frac{\nu}{2})\\&\ge\lambda_{\min}(\mathbb{E}[\frac{w^2(X_i)}{2}X_iX_i^\top])|\Delta|_2^2\cdot\Bigl(1-\frac{4\mathbb{E}|\xi_i|^2}{\nu^2}\Bigr)\\&\ge\frac{3}{4}\lambda_{\min}(\mathbb{E}[\frac{w^2(X_i)}{2}X_iX_i^\top])|\Delta|_2^2,\end{aligned}$$

Also, for $\beta\in\mathcal{C}$, $|\Delta|_2^2+\frac{|\Delta|_1^2}{s}\le 17|\Delta|_2^2$. By (S3.4), we conclude that

$$\begin{aligned}\mathcal{T}(\beta,\beta^*)&\ge\Bigl(\frac{3}{4}\lambda_{\min}(\mathbb{E}[\frac{w^2(X_i)}{2}X_iX_i^\top])-459t\Bigr)|\Delta|_2^2\\&\ge\frac{1}{2}\lambda_{\min}(\mathbb{E}[\frac{w^2(X_i)}{2}X_iX_i^\top])|\Delta|_2^2\end{aligned}$$

□

*Proof of Theorem 4.* With Lemma 1 and Lemma 2, the proof follows the same spirit as Appendix B.1 of Loh (2021) without extra technical difficulty.

□

## S3.2 Proofs of Results in Section 3.2

We shall first prove Proposition 1.

*Proof of Proposition 1.* If $\lambda_{\max}(A) < 1$, for any $\epsilon > 0$, the matrix $B = A/[\lambda_{\max}(A)+\epsilon]$ has spectral radius strictly less than 1. By Theorem 5.6.12 of Golub and Van Loan (2013), $B$ is convergent in the sense that $\lim_{k\to\infty} B^k = 0$. Thus, $\|B^k\| \to 0$ as $k \to \infty$ and there exists some $N = N(\varepsilon, A)$ such that $\|B^k\| < 1$ for all $k \geq N$, which implies $\|A^k\| \leq [\lambda_{\max}(A)+\epsilon]^k$ for all $k \geq N$. Therefore, given the constant $0 < \rho_0 < 1$ and with an arbitrarily small $\epsilon$ with $\lambda_{\max}(A) + \epsilon < 1$, there must exist some finite $k$ such that $\|A^k\| \leq \rho_0$. The proof of the converse is easier by the fact that $[\lambda_{\max}(A)]^k = \lambda_{\max}(A^k) \leq \|A^k\|$ for any $k$. □

To prove Theorem 5, we introduce some preparatory lemmas. Define $\widetilde{L}_j(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n(\widetilde{X}_{ij} - \boldsymbol{b}^\top\widetilde{X}_{i-1})^2$ for $1 \leq j \leq p$.

**Lemma 3.** *Let Assumption (B1) be satisfied. For* $\nu \asymp \mu_q(n/\log p)^{1/2(q-1)}$ *and* $\lambda \asymp \tau\gamma\mu_q(\|A\|_\infty+1)[(\log p)/n]^{1/2-1/2(q-1)}$, *with probability at least* $1-4p^{-c_1}$ *for some* $c_1>0$, *it holds that*

$$\left|\widetilde{L}_j(\boldsymbol{a}_{j\cdot})\right|_\infty \le \lambda, \text{ for all } 1\le j\le p. \tag{S3.5}$$

*Proof of Lemma 3.* We consider the first component of $\nabla\widetilde{L}_j(\boldsymbol{a}_{j\cdot})$, denoted by $\nabla\widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})$. Other components can be manipulated analogously. Let $G(X_i,X_{i-1}) = 2(\widetilde{X}_{i1}-\widetilde{X}_{i-1}^\top\boldsymbol{a}_{j\cdot})\widetilde{X}_{(i-1)1}$, where $\tilde{X}_{(i-1)1}$ is the first element of $\tilde{X}_{(i-1)}$. Then we can write

$$\nabla\widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot}) = \frac{1}{n}\sum_{i=1}^{n} G(X_i,X_{i-1}).$$

Notice that $|G|\le 2(\|A\|_\infty+1)\nu^2$ and $|G(u)-G(v)|\le g^\top|u-v|$, where $|g|_1\le 4(\|A\|_\infty+1)\nu$. By Corollary 1, for $x=c'\gamma\tau\sqrt{(\log p)/n}$, we have

$$\mathbb{P}\left(\left|\nabla\widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})-\mathbb{E}\big[\nabla\widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})\big]\right|\ge 4\nu(\|A\|_\infty+1)x\right)\le 4\exp\Big\{-\frac{(c')^2\log p}{2C_1}\Big\}. \tag{S3.6}$$

In view of $\mathbb{E}[\nabla L_n(\boldsymbol{a}_{j\cdot})]=0$, the triangle inequality and $|\widetilde{X}_{ij}|\le|X_{ij}|$,

$$\begin{aligned}
\big|\mathbb{E}\big[\nabla\widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})\big]\big| &= \big|\mathbb{E}\big[\nabla\widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})\big]-\mathbb{E}\big[\nabla L_{j1}(\boldsymbol{a}_{j\cdot})\big]\big|\\
&= 2\mathbb{E}\big[\big|(\widetilde{X}_{ij}-\boldsymbol{a}_{j\cdot}^\top\widetilde{X}_{i-1})\widetilde{X}_{(i-1)1}-(X_{ij}-\boldsymbol{a}_{j\cdot}^\top X_{i-1})X_{(i-1)1}\big|\big]\\
&\lesssim \mathbb{E}\big[\big|X_{(i-1)1}(\widetilde{X}_{ij}-X_{ij})\big|\big]+\mathbb{E}\big[\big|X_{ij}(X_{(i-1)1}-\widetilde{X}_{(i-1)1})\big|\big]\\
&\quad +|\boldsymbol{a}_{j\cdot}|^\top\mathbb{E}\big[\big|X_{(i-1)1}(\widetilde{X}_{i-1}-X_{i-1})\big|\big]
\end{aligned}$$

$$+|\boldsymbol{a}_{j\cdot}|^\top \mathbb{E}\big[\big|X_{i-1}(\widetilde{X}_{(i-1)1} - X_{(i-1)1})\big|\big]. \tag{S3.7}$$

Since $|\widetilde{X}_{ij} - X_{ij}| \le |X_{ij}|\mathbf{1}\{|X_{ij}| \ge \nu\}$, by Hölder's inequality, we have

$$\begin{aligned}\mathbb{E}\big[\big|X_{(i-1)1}(X_{ij} - \widetilde{X}_{ij})\big|\big] &\le \|\tilde{X}_{(i-1)1}\|_q \cdot \|\tilde{X}_{ij} - X_{ij}\|_{q/(q-1)} \\ &\le \mu_q \|\tilde{X}_{ij} - X_{ij}\|_{q/(q-1)},\end{aligned}$$

where

$$\|\tilde{X}_{ij} - X_{ij}\|_{q/(q-1)}^{q/(q-1)} \le \mathbb{E}|X_{ij}|^{q/(q-1)}\mathbf{1}\{|X_{ij}| \ge \nu\} \le \mu_q^q \nu^{-q(q-2)/(q-1)}.$$

It then follows that $\mathbb{E}\big[\big|X_{(i-1)1}(X_{ij} - \widetilde{X}_{ij})\big|\big] \le \mu_q^q \nu^{2-q}$. Other terms in (S3.7) can be dealt with similarly. With the choice of $\nu$, we can get $\big|\mathbb{E}\big[\nabla \widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})\big]\big| \le c\nu(\|A\|_\infty + 1)x$. Letting $\lambda = C\nu(\|A\|_\infty + 1)x$ for a sufficiently large $C$ and $c' > 2\sqrt{C_1}$, it follows from (S3.6) that

$$\mathbb{P}\left(\left|\nabla \widetilde{L}_{j1}(\boldsymbol{a}_{j\cdot})\right| \ge \lambda\right) \le 4\exp\Big\{-\frac{(c')^2 \log p}{2C_1}\Big\}.$$

By the Bonferroni inequality, we have

$$\mathbb{P}\left(\left|\nabla \widetilde{L}_{j}(\boldsymbol{a}_{j\cdot})\right|_\infty \ge \lambda, \text{ for all } 1 \le j \le p\right) \le 4p^{-c_1}$$

where $c_1 = 2^{-1}C_1^{-1}(c')^2 - 2 > 0$. □

Define a cone $C(S) = \{\Delta \in \mathbb{R}^p : |\Delta_{S^c}|_1 \le 3|\Delta_S|_1\}$ for a subset $S \subseteq \{1, 2, \ldots, p\}$. We shall verify a restricted eigenvalue (RE) condition on the set $C(S)$ in the lemma below.

**Lemma 4.** *Let Assumptions (B1) and (B2) be satisfied. Choose* $\nu \asymp \mu_q(n/\log p)^{1/(2q-2)}$*. Then for all* $\Delta \in C(S)$*,*

$$\Delta^\top \nabla^2 \widetilde{L}_j(\boldsymbol{a}_{j\cdot})\Delta \geq \frac{1}{2}|\Delta|_2^2 \tag{S3.8}$$

*holds with probability at least* $1 - 4p^{-c_2}$ *for some constant* $c_2 > 0$*.*

*Proof of Lemma 4.* Denote $\widetilde{X} = (\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_{n-1})^\top$. Then $\nabla^2 \widetilde{L}_j(\boldsymbol{a}_{j\cdot}) = 2\widetilde{X}^\top \widetilde{X}/n =: \Gamma$. We shall first show that with probability at least $1 - 4p^{-c_2}$ for some positive constant $c_2$, it holds that

$$\left|v^\top \big(\Gamma - \mathbb{E}\Gamma\big) v\right| \leq t,\ \forall\, v \in \mathbb{R}^p,\ |v|_0 \leq 2s,\ |v|_2 \leq 1, \tag{S3.9}$$

where $t = c_1 \mu_q \gamma \tau s^2 (\log p/n)^{1/2 - 1/[2(q-1)]}$. For any $u \in \mathbb{R}^p$ such that $|u|_2 \leq 1$ and $|u|_0 \leq s$ hence $|u|_1 \leq \sqrt{s}$, write

$$u^\top(\Gamma - \mathbb{E}\Gamma)u = 2n^{-1}\sum_{i=0}^{n-1}(u^\top \widetilde{X}_i)^2 - \mathbb{E}(u^\top \widetilde{X}_i)^2 =: n^{-1}\sum_{i=0}^{n-1} G(X_i) - \mathbb{E}[G(X_i)].$$

Thus, for $G(X_i) = (u^\top \widetilde{X}_i)^2$, we have

$$|G(x) - G(y)| \leq 2|u^\top(x+y)\cdot u^\top(x-y)| \leq 4s\nu g^\top |x-y|,$$

where $|g|_1 \leq 1$. Apply Theorem 1 to function $G(X_i)/(4s\nu)$ and we have for any fixed $u$ such that $|u|_2 \leq 1$ and $|u|_0 \leq s$,

$$\mathbb{P}\Big(\left|u^\top\big(\Gamma - \mathbb{E}\Gamma\big)u\right| \geq t\Big) \leq 4\exp\big\{-c_3 s^2 \log p\big\}.$$

Following the same spirit of the $\varepsilon$-net argument in lemma 15 of Loh and Wainwright (2012), we can obtain that (S3.9) holds with probability at least

$$1 - 4\exp\left\{2s\log 9 + 2s\log p - c_3 s^2 \log p\right\} \geq 1 - 4p^{-c_2},$$

provided that $p \to \infty$ and a sufficiently large $c_3$ (or equivalently $c_1$). By Lemma 12 in Loh and Wainwright (2012) and (S3.9), it further implies that with probability greater than $1 - 4p^{-c_2}$,

$$|v^\top(\Gamma - \mathbb{E}\Gamma)v| \leq 27t\left(|v|_2^2 + \frac{|v|_1^2}{s}\right), \ \forall v \in \mathbb{R}^p. \tag{S3.10}$$

Note that when $\Delta \in C(S)$,

$$|\Delta|_1 = |\Delta_S|_1 + |\Delta_{S^c}|_1 \leq 4|\Delta_S|_1 \leq 4\sqrt{s}|\Delta_S|_2 \leq 4\sqrt{s}|\Delta|_2. \tag{S3.11}$$

Furthermore, some algebra delivers that

$$\begin{aligned} \Delta^\top \mathbb{E}\big[\Gamma\big]\Delta = 2\mathbb{E}[(\widetilde{X}_1^\top \Delta)^2] \ &\geq \ 2\Big(\Delta^\top \mathbb{E}[X_1 X_1^\top]\Delta - \Delta^\top \mathbb{E}[X_1 X_1^\top - \widetilde{X}_1 \widetilde{X}_1^\top]\Delta\Big) \\ &\geq \ 2|\Delta|_2^2 - 2|\Delta|_1^2 \big|\mathbb{E}[X_1 X_1^\top - \widetilde{X}_1 \widetilde{X}_1^\top]\big|_\infty. \end{aligned} \tag{S3.12}$$

For any $1 \leq j, k \leq p$, by the triangle and Hölder's inequality,

$$|\mathbb{E}\tilde{X}_{ij}\tilde{X}_{ik} - \mathbb{E}X_{ij}X_{ik}| \leq |\mathbb{E}(\tilde{X}_{ij} - X_{ij})\tilde{X}_{ik})| + |\mathbb{E}(\tilde{X}_{ik} - X_{ik})X_{ij})|.$$

We have

$$|\mathbb{E}(\tilde{X}_{ij} - X_{ij})\tilde{X}_{ik})| \ \leq \ \|\tilde{X}_{ik}\|_q \cdot \|\tilde{X}_{ij} - X_{ij}\|_{q/(q-1)}$$

$$\leq \mu_q \|\tilde{X}_{ij} - X_{ij}\|_{q/(q-1)},$$

where

$$\|\tilde{X}_{ij} - X_{ij}\|_{q/(q-1)}^{q/(q-1)} \leq \mathbb{E}|X_{ij}|^{q/(q-1)} \mathbf{1}\{|X_{ij}| \geq \nu\} \leq \mu_q^q \nu^{-q(q-2)/(q-1)}.$$

It then follows that $|\mathbb{E}(\tilde{X}_{ij} - X_{ij})\tilde{X}_{ik}| \leq \mu_q^q \nu^{2-q}$. We can also deal with $|\mathbb{E}(\tilde{X}_{ik} - X_{ik})X_{ij})|$ similarly. As a result, we have the bias

$$|\mathbb{E}[\tilde{X}_{ij}\tilde{X}_{ik} - X_{ij}X_{ik}]| \leq 2\mu_q^q \nu^{2-q} \leq C\mu_q^2 \big(\frac{\log p}{n}\big)^{\frac{1}{2} - \frac{1}{2q-2}}. \tag{S3.13}$$

By (S3.11), (S3.12) and (S3.13), it follows that

$$\Delta^\top \mathbb{E}\big[\Gamma\big]\Delta \geq 2|\Delta|_2^2 - 16Cs\mu_q^2 \big(\frac{\log p}{n}\big)^{\frac{1}{2} - \frac{1}{2q-2}} |\Delta|_2^2 \geq |\Delta|_2^2. \tag{S3.14}$$

Recall that $t = c_1 \mu_q \gamma \tau s^2 (\log p/n)^{1/2 - 1/[2(q-1)]} \to 0$ by Assumption (B2). Combining (S3.10) and (S3.14), we can establish the following RE condition

$$\nabla^2 L_j(\boldsymbol{a}_{j\cdot}) \geq |\Delta|_2^2 - 27t(|\Delta|_2^2 + |\Delta|_1^2/s) \geq |\Delta|_2^2 - 459t|\Delta|_2^2 \geq \frac{1}{2}|\Delta|_2^2,$$

for all $\Delta \in C(S)$ with probability no less than $1 - 4p^{-c_2}$. □

*Proof of Theorem 5.* Let $\widehat{\Delta}_j = \widehat{\boldsymbol{a}}_{j\cdot} - \boldsymbol{a}_{j\cdot}$ for $j = 1, \ldots, p$. As the solution of (3.5), $\widehat{\boldsymbol{a}}_{j\cdot}$ satisfies

$$\widetilde{L}_j(\widehat{\boldsymbol{a}}_{j\cdot}) + \lambda|\widehat{\boldsymbol{a}}_{j\cdot}|_1 \leq \widetilde{L}_j(\boldsymbol{a}_{j\cdot}) + \lambda|\boldsymbol{a}_{j\cdot}|_1,$$

which together with convexity implies,

$$0 \le \widetilde{L}_j(\widehat{\boldsymbol{a}}_{j\cdot}) - \widetilde{L}_j(\boldsymbol{a}_{j\cdot}) - \langle \nabla \widetilde{L}_j(\boldsymbol{a}_{j\cdot}), \widehat{\Delta}_j \rangle \le \lambda(|\boldsymbol{a}_{j\cdot}|_1 - |\widehat{\boldsymbol{a}}_{j\cdot}|_1) + \big|\nabla \widetilde{L}_j(\boldsymbol{a}_{j\cdot})\big|_\infty |\widehat{\Delta}_j|_1. \tag{S3.15}$$

Denote by $A$ and $B$ the events in Lemma 3 and Lemma 4 respectively. Then $\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) \ge 1 - 8p^{-c}$ for $c = \min\{c_1, c_2\}$. Conditioned on the event $A$, (S3.15) implies

$$\begin{aligned} 0 &\le |\boldsymbol{a}_{j\cdot,S}|_1 - |\widehat{\boldsymbol{a}}_{j\cdot,S}|_1 - |\widehat{\boldsymbol{a}}_{j\cdot,S^c}|_1 + \frac{1}{2}|\widehat{\Delta}_j|_1 \\ &\le |\widehat{\Delta}_{j,S}|_1 - |\widehat{\Delta}_{j,S^c}|_1 + \frac{1}{2}|\widehat{\Delta}_j|_1 = \frac{3}{2}|\widehat{\Delta}_{j,S}|_1 - \frac{1}{2}|\widehat{\Delta}_{j,S^c}|_1, \end{aligned}$$

which further implies $\widehat{\Delta}_j \in C(S)$ for all $1 \le j \le p$. Conditioned on the event $B$, by (S3.5) and the second inequality in (S3.15), we have

$$\frac{1}{2}|\widehat{\Delta}_j|_2^2 \le \big(\lambda + \big|\nabla L_n(\boldsymbol{a}_{j\cdot})\big|_\infty\big)|\widehat{\Delta}_j|_1 \le 6\sqrt{s}\lambda|\widehat{\Delta}_j|_2. \tag{S3.16}$$

This immediately shows for all $1 \le j \le p$

$$|\widehat{\Delta}_j|_2 \le 12\sqrt{s}\lambda \asymp \mu_q\gamma\tau(\|A\|_\infty + 1)\sqrt{s}\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\frac{1}{2q-2}} \tag{S3.17}$$

as well as

$$|\widehat{\Delta}_j|_1 \lesssim \mu_q\gamma\tau s(\|A\|_\infty + 1)\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\frac{1}{2q-2}}.$$

Hence, (3.6) follows in view of $\|\widehat{A} - A\|_\infty = \max_j |\widehat{\Delta}_j|_1$. Moreover, if we consider the estimation of $\mathbf{Vec}(A) = (\boldsymbol{a}_{1\cdot}^\top, \boldsymbol{a}_{2\cdot}^\top, \ldots, \boldsymbol{a}_{p\cdot}^\top)^\top \in \mathbb{R}^{p^2}$ with the s-parsity parameter $\mathcal{S} = \sum_{i=j}^p s_j$, by Assumption (B2′) and similar arguments of verifying the RE condition in Lemma 4, (S3.8) becomes

$$2\Delta^\top \left[ I_p \otimes \left( \frac{\widetilde{X}^\top \widetilde{X}}{n} \right) \right] \Delta \geq \frac{1}{2} |\Delta|_2^2, \quad \text{for all } \Delta \in \mathbb{R}^{p^2}.$$

Thus, similarly as (S3.17), (3.7) follows. □

Next we shall concern the robust Dantzig-type estimator.

**Lemma 5.** *Let Assumption (B1) be satisfied. Choose the truncation parameter* $\nu \asymp \mu_q (n/\log p)^{1/(2q-2)}$. *Let* $\lambda \asymp \mu_q \gamma \tau (\|A\|_1 + 1)[(\log p)/n]^{(q-2)/(2q-2)}$. *Then with probability at least* $1 - 8p^{-c'}$ *for some constant* $c' > 0$, *it holds that*

$$\|\widehat{\Sigma}_0 - \Sigma_0\|_{\max} \leq \lambda_0 \quad \textit{and} \quad \|\widehat{\Sigma}_1 - \Sigma_1\|_{\max} \leq \lambda_0.$$

*Proof of Lemma 5.* Let $\lambda_0 = C\mu_q \tau \gamma [(\log p)/n]^{(q-2)/(2q-2)}$ for a sufficiently large constant $C$. Applying Theorem 1 to the $(m, l)$-th entry of $\widehat{\Sigma}_0$, we have

$$\mathbb{P}\Big( \frac{1}{n} \Big| \sum_{i=1}^n \widetilde{X}_{im} \widetilde{X}_{il} - \mathbb{E} \widetilde{X}_{im} \widetilde{X}_{il} \Big| \geq \lambda_0 \Big) \leq 4 \exp\Big\{ - \frac{c^2 \log p}{2C_1} \Big\} = 4p^{-c^2/(2C_1)}.$$

By (S3.13) in the proof of Lemma 4, we see that

$$\Big| \mathbb{E} \widetilde{X}_{im} \widetilde{X}_{il} - \mathbb{E} X_{im} X_{il} \Big| \leq c\mu_q^2 \big( \frac{\log p}{n} \big)^{\frac{1}{2} - \frac{1}{2q-2}} \leq \lambda_0.$$

Therefore,

$$\mathbb{P}\Big( \frac{1}{n} \Big| \sum_{i=1}^n \widetilde{X}_{im} \widetilde{X}_{il} - \mathbb{E}[X_{im} X_{il}] \Big| \geq \lambda_0 \Big)$$

$$\leq \mathbb{P}\Big(\frac{1}{n}\Big|\sum_{i=1}^{n} \widetilde{X}_{im}\widetilde{X}_{il} - \mathbb{E}[\widetilde{X}_{im}\widetilde{X}_{il}]\Big| \geq C_2\lambda_0\Big) \leq 4p^{-C_3}$$

for some $C_3 > 1$. Taking a union bound yields

$$\mathbb{P}(\|\widehat{\Sigma}_0 - \Sigma_0\|_{\max} \geq \lambda_0) \leq 4p^{-c'},$$

where $c' = C_3 - 1 > 0$. By Corollary 1, similar arguments apply to $\widehat{\Sigma}_1$, which delivers $\|\widehat{\Sigma}_1 - \Sigma_1\|_{\max} \leq \lambda_0$ with probability at least $1 - 4p^{-c'}$. In conclusion, it holds simultaneously that $\|\widehat{\Sigma}_0 - \Sigma_0\|_{\max} \leq \lambda_0$ and $\|\widehat{\Sigma}_1 - \Sigma_1\|_{\max} \leq \lambda_0$ with probability at least $1 - 8p^{-c'}$. □

*Proof of Theorem 6.* We first show that $A$ is feasible to the convex programming (3.8) for $\lambda = (\|A\|_1 + 1)\lambda_0$ with high probability. By the Yule-Walker equation and Lemma 5, we have

$$\begin{aligned}\|\widehat{\Sigma}_0 A - \widehat{\Sigma}_1\|_{\max} &\leq \|\widehat{\Sigma}_0 A - \Sigma_1\|_{\max} + \|\Sigma_1 - \widehat{\Sigma}_1\|_{\max} \\ &\leq \|\widehat{\Sigma}_0 - \Sigma_0\|_{\max}\|A\|_1 + \|\Sigma_1 - \widehat{\Sigma}_1\|_{\max} \leq \lambda,\end{aligned}$$

with probability no less than $1 - 8p^{-c'}$. Therefore, conditioned on the event in Lemma 5, we conclude that $|\widehat{\boldsymbol{a}}_{\cdot j}|_1 \leq |\boldsymbol{a}_{\cdot j}|_1$ for all $j = 1, \ldots, p$ and hence $\|\widehat{A}\|_1 \leq \|A\|_1$. Then we have

$$\begin{aligned}\|\widehat{A} - A\|_{\max} &= \|\Sigma_0^{-1}(\Sigma_0\widehat{A} - \widehat{\Sigma}_1 + \widehat{\Sigma}_1 - \Sigma_1)\|_{\max} \\ &\leq \|\Sigma_0^{-1}(\Sigma_0\widehat{A} - \widehat{\Sigma}_0\widehat{A} + \widehat{\Sigma}_0\widehat{A} - \widehat{\Sigma}_1)\|_{\max} + \|\Sigma_0^{-1}(\widehat{\Sigma}_1 - \Sigma_1)\|_{\max}\end{aligned}$$

$$\leq \|\Sigma_0^{-1}\|_1\|\Sigma_0 - \widehat{\Sigma}_0\|_{\max}\|\widehat{A}\|_1 + \|\Sigma_0^{-1}\|_1\|\widehat{\Sigma}_0\widehat{A} - \widehat{\Sigma}_1\|_{\max} + \|\Sigma_0^{-1}\|_1\|\widehat{\Sigma}_1 - \Sigma_1\|_{\max}.$$

By Lemma 5 and the feasibility of $\widehat{A}$, we have

$$\|\widehat{A} - A\|_{\max} \leq \|\Sigma_0^{-1}\|_1(\lambda_0\|A_1\| + \lambda + \lambda_0) = 2\|\Sigma_0^{-1}\|_1\lambda.$$

Now we shall bound $\|\widehat{A} - A\|_1$ from above. Denote by $S_j$ the support of $\boldsymbol{a}_{\cdot j}$ for $j = 1, \ldots, p$. Then for any $1 \leq j \leq p$, we have

$$\begin{aligned} |\widehat{\boldsymbol{a}}_{\cdot j} - \boldsymbol{a}_{\cdot j}|_1 &= \left|\widehat{\boldsymbol{a}}_{\cdot j,S_j} - \boldsymbol{a}_{\cdot j,S_j}\right|_1 + \left|\widehat{\boldsymbol{a}}_{\cdot j}\right|_1 - \left|\widehat{\boldsymbol{a}}_{\cdot j,S_j}\right|_1 \\ &\leq \left|\widehat{\boldsymbol{a}}_{\cdot j,S_j} - \boldsymbol{a}_{\cdot j,S_j}\right|_1 + \left|\boldsymbol{a}_{\cdot j}\right|_1 - \left|\widehat{\boldsymbol{a}}_{\cdot j,S_j}\right|_1 \\ &\leq 2\left|\widehat{\boldsymbol{a}}_{\cdot j,S_j} - \boldsymbol{a}_{\cdot j,S_j}\right|_1 \leq 4s^*\|\Sigma_0^{-1}\|_1\lambda. \end{aligned} \tag{S3.18}$$

Since (S3.18) holds for all $1 \leq j \leq p$, we conclude that

$$\|\widehat{A} - A\|_1 \leq 4s^*\|\Sigma_0^{-1}\|_1\lambda \lesssim \mu_q s^*\gamma\tau\|\Sigma_0^{-1}\|_1(\|A\|_1 + 1)\left(\frac{\log p}{n}\right)^{\frac{1}{2} - \frac{1}{2q-2}}.$$

□

## References

Golub, G. H. and C. F. Van Loan (2013). *Matrix computations, 4th edition.* Johns Hopkins University Press.

Loh, P.-L. (2021). Scale calibration for high-dimensional robust regression. *Electronic Journal of Statistics 15*(2), 5933–5994.

Loh, P.-L. and M. J. Wainwright (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics 40*(3), 1637.

Zhang, D. (2021). Robust estimation of the mean and covariance matrix for high dimensional time series. *Statistica Sinica 31*(2), 797–820.

Department of Mathematics,

University of California, San Diego,

La Jolla, CA 92093

E-mail: linbo@ucsd.edu

E-mail: daz076@ucsd.edu