Supplemental Materials for "Variable screening via quantile partial" correlation

Shujie Ma, Runze Li and Chih-Ling Tsai

February 16, 2016

In this document, we first provide the proofs for Lemmas A.1-A.5. Then, we present additional simulation results for Examples 1-3. Specifically, We report the results for the moderate correlation coefficient $\rho = 0.5$ in Examples 1 and 2. In addition, we report the results for $p = 2,000$ in Example 1. In Example 2, because $p = 2,000$ yields similar performance to that of $p = 1,000$, we do not report it. Moreover, the comparisons of QPCD, l_1 and SCAD associated with Example 3 are presented. Finally, we list some acronyms used in the manuscript. Detailed illustrations of the supplementary materials are given below.

Proof of Lemma A.1. Denote $\boldsymbol{\beta}^0_{\tau,-j} = (\beta^0_{0\tau}, \cdots, \beta^0_{(j-1)\tau}, \beta^0_{(j+1)\tau}, \cdots, \beta^0_{p\tau})^T$. If $\beta^0_{j\tau} = 0$, then $(\boldsymbol{\beta}^{0T}_{\tau,-j}, \beta^0_{j\tau})^T = (\boldsymbol{\beta}^{0T}_{\tau,-j}, 0)^T$ is the unique solution to

$$E[\psi_\tau(Y - \boldsymbol{\beta}^T_{\tau,-j}\mathbf{X}_{-j} - \beta_{j\tau}X_j)\mathbf{X}] = 0. \quad \text{(S.1)}$$

Hence, $E[\psi_\tau(Y - \boldsymbol{\beta}^{0T}_{\tau,-j}\mathbf{X}_{-j})] = 0$, and $E[\psi_\tau(Y - \boldsymbol{\beta}^{0T}_{\tau,-j}\mathbf{X}_{-j})X_k] = 0$ for all $k = 1, \ldots, p$, which implies $E[\psi_\tau(Y - \boldsymbol{\beta}^{0T}_{\tau,-j}\mathbf{X}_{-j})\mathbf{X}_{-j}] = \mathbf{0}$. By the definition of $\boldsymbol{\alpha}^0_j$, we also have $E[\psi_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}^0_j)\mathbf{X}_{-j}] = \mathbf{0}$. Thus, $\boldsymbol{\alpha}^0_j = \boldsymbol{\beta}^0_{\tau,-j}$. Moreover, by (5), we have $E[\psi_\tau(Y - \boldsymbol{\alpha}^{0T}_j\mathbf{X}_{-j} - \beta^*_{0\tau} -$

Shujie Ma is Assistant Professor, Department of Statistics, University of California-Riverside, Riverside, CA 92521 (Email: shujie.ma@ucr.edu). Runze Li is Verne M. Willaman Professor, Department of Statistics, the Pennsylvania State University, University Park, PA 16802 (Email: rzli@psu.edu). Chih-Ling Tsai is Distinguished Professor and Robert W. Glock Endowed Chair in Management, Graduate School of Management, University of California at Davis, Davis, CA 95616 (Email: cltsai@ucdavis.edu). Ma's research was supported by NSF grant DMS 1306972 and Hellman Fellowship. Li's research was supported by NSF grant DMS 1512422 and NIDA, NIH grants P50 DA036107 and P50 DA039838. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF, NIDA or NIH. The authors are grateful to the Editor, the Associate Editor, and three anonymous reviewers for their constructive comments that helped us improve the article substantially.

$\beta^*_{j\tau}X_j)X_j] = 0$ and $E[\psi_\tau(Y - \boldsymbol{\alpha}_j^{0T}\mathbf{X}_{-j} - \beta^*_{0\tau} - X_j\beta^*_{j\tau})] = 0$, and by (S.1) and the fact that $\boldsymbol{\alpha}_j^0 = \boldsymbol{\beta}^0_{\tau,-j}$, we have $E[\psi_\tau(Y - \boldsymbol{\alpha}_j^{0T}\mathbf{X}_{-j} - 0 - \beta^0_{j\tau}X_j)X_j] = 0$ and $E[\psi_\tau(Y - \boldsymbol{\alpha}_j^{0T}\mathbf{X}_{-j} - 0 - \beta^0_{j\tau}X_j)] = 0$. As a result, $\beta^*_{0\tau} = 0$ and $\beta^*_{j\tau} = \beta^0_{j\tau} = 0$.

On the other hand, if $\beta^*_{j\tau} = 0$, then we have $E[\psi_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}^0_j)] = 0$ and $E[\psi_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}^0_j - \beta^*_{0\tau})] = 0$. These imply that $\beta^*_{0\tau} = 0$. Using the fact that $E[\psi_\tau(Y - \boldsymbol{\alpha}_j^{0T}\mathbf{X}_{-j} - 0 \times X_j)] = 0$, $E[\psi_\tau(Y - \boldsymbol{\alpha}_j^{0T}\mathbf{X}_{-j} - 0 \times X_j)X_k] = \mathbf{0}$ for $k \neq j$, and $E[\psi_\tau(Y - \boldsymbol{\alpha}_j^{0T}\mathbf{X}_{-j} - 0 \times X_j)X_j] = 0$, we further obtain that $(\boldsymbol{\alpha}_j^{0T}, 0)^T$ is a solution to

$$E[\psi_\tau(Y - \ \boldsymbol{\beta}^T_{\tau,-j}\mathbf{X}_{-j} - \beta_{j\tau}X_j)\mathbf{X}] = 0. \tag{S.2}$$

Since $(\boldsymbol{\beta}^{0T}_{\tau,-j}, \beta^0_{j\tau})^T$ is the unique solution to (S.2), we have $(\boldsymbol{\alpha}_j^{0T}, 0)^T = (\boldsymbol{\beta}^{0T}_{\tau,-j}, \beta^0_{j\tau})^T$. Accordingly, $\beta^0_{j\tau} = 0$. $\square$

Proof of Lemma A.2. By the definitions of $\widehat{\boldsymbol{\vartheta}}_j$ and $\boldsymbol{\vartheta}^0_j$, we have

$$\begin{aligned} \widehat{\boldsymbol{\vartheta}}_j &= (n^{-1}\sum\nolimits_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}^T_{i,\mathcal{S}_j})^{-1}(n^{-1}\sum\nolimits_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}X_{ij}), \\ \boldsymbol{\vartheta}^0_j &= \{E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}^T_{i,\mathcal{S}_j})\}^{-1}E(\mathbf{X}_{i,\mathcal{S}_j}X_{ij}). \end{aligned}$$

Then

$$\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}^0_j = \Upsilon_{1j} + \Upsilon_{2j}, \tag{S.3}$$

where

$$\Upsilon_{1j} = \{E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}^T_{i,\mathcal{S}_j})\}^{-1}\{n^{-1}\sum\nolimits_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}X_{ij} - E(\mathbf{X}_{i,\mathcal{S}_j}X_{ij})\} \text{ and}$$
$$\Upsilon_{2j} = [(n^{-1}\sum\nolimits_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}^T_{i,\mathcal{S}_j})^{-1} - \{E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}^T_{i,\mathcal{S}_j})\}^{-1}](n^{-1}\sum\nolimits_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}X_{ij}).$$

Denote $T_{ikj} = X_{ik}X_{ij} - E(X_{ik}X_{ij})$ for $k \in \{0\} \cup \mathcal{S}_j$. By Condition (C2), $|T_{ikj}| \leq 2M_1^2$ and $\text{var}(T_{ikj}) \leq M_1^4$. Employing Bernstein's Inequality (Lemma 2.2.9, van der Vaart and Wellner (1996)), we then have, for any constant $c^*_1 > 0$,

$$\begin{aligned} P\left(|n^{-1}\sum\nolimits_{i=1}^n T_{ikj}| \geq c^*_1 n^{-1}\delta_n\right) &\leq 2\exp\{-c_1^{*2}\delta_n^2/(2(nM_1^4 + 2M_1^2c^*_1\delta_n))\} \\ &\leq 2\exp(-c^*_2\delta^2_n n^{-1}), \end{aligned}$$

for some positive constant c^*_2, when n is large enough. The above results, together with the union bound of probability, imply that

$$P\left(||n^{-1}\sum\nolimits_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}X_{ij} - E(\mathbf{X}_{i,\mathcal{S}_j}X_{ij})|| \geq c^*_1 n^{-1}r_n^{1/2}\delta_n\right) \leq 2r_n\exp(-c^*_2\delta^2_n n^{-1}).$$

By Condition (C2) that $\lambda_{\min}(E(\mathbf{X}_{\mathcal{S}_j}\mathbf{X}_{\mathcal{S}_j}^T)) \geq m$, we thus have

$$P\left(||\Upsilon_{1j}|| \geq c_1^* m^{-1} n^{-1} r_n^{1/2}\delta_n\right) \leq 2r_n \exp(-c_2^*\delta_n^2 n^{-1}). \tag{S.4}$$

Define $D_j = n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T - E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)$ and $D_{j,kk'} = n^{-1}\sum_{i=1}^n X_{ik}X_{ik'} - E(X_{ik}X_{ik'})$ for $k, k' \in \{0\} \cup \mathcal{S}_j$. After algebraic simplification, we obtain

$$|\lambda_{\min}(n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T) - \lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))| \leq r_n|D_j| \text{ and}$$
$$||n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T - E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)|| \leq r_n|D_j|.$$

By Bernstein's Inequality, we have, for any $\delta_1 > 0$ and $c_3^* > 0$, there exists some positive constant c_4^* such that

$$P(|D_{j,kk'}| \geq c_3^* n^{-1}\delta_1) \leq 2\exp(-c_4^*\delta_1^2 n^{-1}).$$

The above results, in conjunction with the union bound of probability, lead to

$$\begin{aligned} & P\left(|\lambda_{\min}(n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T) - \lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))| \geq c_3^* r_n n^{-1}\delta_1^*\right) \\ \leq\ & 2r_n^2\exp(-c_4^*\delta_1^{*2}n^{-1}) \text{ and} \end{aligned}$$

$$P(||n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T - E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)|| \geq c_3^* r_n n^{-1}\delta_1) \leq 2r_n^2\exp(-c_4^*\delta_1^2 n^{-1}), \tag{S.5}$$

for any $\delta_1^* > 0$ and $\delta_1 > 0$. Let $\delta_1^* = c_5^*(c_3^*)^{-1}r_n^{-1}nm$ for some constant $c_5^* \in (0,1)$. Then, we obtain $c_3^* r_n n^{-1}\delta_1^* \leq c_5^*\lambda_{\min}(E(\mathbf{X}_{\mathcal{S}_j}\mathbf{X}_{\mathcal{S}_j}^T))$ and

$$\begin{aligned} & P\left(|\lambda_{\min}(n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T) - \lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))| \geq c_5^*\lambda_{\min}(E(\mathbf{X}_{\mathcal{S}_j}\mathbf{X}_{\mathcal{S}_j}^T))\right) \\ \leq\ & 2r_n^2\exp(-c_4^* c_5^{*2}(c_3^*)^{-2}m^2 n r_n^{-2}). \end{aligned} \tag{S.6}$$

In addition,

$$\begin{aligned} & |\{\lambda_{\min}(n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)\}^{-1} - \{\lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\}^{-1}| \\ \geq\ & (1/(1-c_5^*)-1)\{\lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\}^{-1} \end{aligned}$$

implies

$$|\lambda_{\min}(n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T) - \lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))| \geq c_5^*\lambda_{\min}(E(\mathbf{X}_{\mathcal{S}_j}\mathbf{X}_{\mathcal{S}_j}^T)).$$

This, together with the above result, yields

$$\begin{aligned} & P\left[|\{\lambda_{\min}(n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)\}^{-1} - \{\lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\}^{-1}| \right. \\ & \left. \geq (1/(1-c_5^*)-1)\{\lambda_{\min}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\}^{-1}\right] \\ \leq \; & 2r_n^2\exp(-c_4^*c_5^{*2}(c_3^*)^{-2}m^2nr_n^{-2}). \end{aligned} \tag{S.7}$$

Using the fact that

$$\begin{aligned} ||\Upsilon_{2j}|| \;\leq\; & ||n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)^{-1}|| \times ||\{E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)\}^{-1}|| \times \\ & ||n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T - E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)|| \times ||n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}X_{ij}|| \end{aligned}$$

as well as employing Condition (C2) that $||n^{-1}\sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j}X_{ij}|| \leq M_4$ and $\lambda_{\min}(E(\mathbf{X}_{\mathcal{S}_j}\mathbf{X}_{\mathcal{S}_j}^T)) \geq m$, and equations (S.5) and (S.7), we obtain

$$\begin{aligned} & P\left\{||\Upsilon_{2j}|| \geq ((1/(1-c_5^*)-1)m^2)(c_3^*r_nn^{-1}\delta_1)M_4\right\} \\ \leq \; & 2r_n^2\exp(-c_4^*\delta_1^2n^{-1}) + 2r_n^2\exp(-c_4^*c_5^{*2}(c_3^*)^{-2}m^2nr_n^{-2}). \end{aligned}$$

Accordingly, for any $c_6^* > 0$, by letting $c_3^* = (1/(1-c_5^*)-1)m^2)^{-1}M_4^{-1}c_6^*$, we have

$$P\left\{||\Upsilon_{2j}|| \geq c_6^*r_nn^{-1}\delta_1\right\} \leq 2r_n^2\exp(-c_4^*\delta_1^2n^{-1}) + 2r_n^2\exp(-c_7^*nr_n^{-2}), \tag{S.8}$$

for some positive constant c_7^*. This, in conjunction with (S.3)and (S.4), implies that, for any $c_1^* > 0$ and $c_6^* > 0$, there exist some positive constants c_2^*, c_4^* and c_7^* such that

$$\begin{aligned} & P\left(||\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}_j^0|| \geq c_1^*m^{-1}n^{-1}r_n^{1/2}\delta_n + c_6^*r_nn^{-1}\delta_1\right) \\ \leq \; & 2r_n\exp(-c_2^*\delta_n^2n^{-1}) + 2r_n^2\exp(-c_4^*\delta_1^2n^{-1}) + 2r_n^2\exp(-c_7^*nr_n^{-2}). \end{aligned}$$

Let $\delta_1 = \delta_n$, we consequently have that, for any $c_1 > 0$, there exist some positive constants c_2 and c_3 such that

$$P(||\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}_j^0|| \geq c_1r_nn^{-1}\delta_n) \leq 4r_n^2\exp(-c_2\delta_n^2n^{-1}) + 2r_n^2\exp(-c_3nr_n^{-2}).$$

□

Proof of Lemma A.3. By definition, $\widehat{\boldsymbol{\pi}}_j$ and $\boldsymbol{\pi}_j^0$ are the minimizers of

$$\begin{aligned} \varpi_n(\boldsymbol{\pi}_j) &= n^{-1}\sum_{i=1}^n[\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\pi}_j) - \rho_\tau(Y_i)] \;\text{ and} \\ \varpi(\boldsymbol{\pi}_j) &= E[\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\pi}_j) - \rho_\tau(Y_i)], \end{aligned}$$

respectively. To prove this lemma, we will employ the result given in Lemma 2 of

?, i.e., for any $\xi > 0$,

$$
\begin{aligned}
& P\left(||\widehat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_j^0|| \geq \xi\right) \\
\leq \; & P\left(\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \leq \xi} |\varpi_n(\boldsymbol{\pi}_j) - \varpi(\boldsymbol{\pi}_j)| \geq \frac{1}{2} \inf_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| = \xi} \varpi(\boldsymbol{\pi}_j) - \varpi(\boldsymbol{\pi}_j^0)\right). \qquad \text{(S.9)}
\end{aligned}
$$

We first show that, for some positive constant c_8^*,

$$
\inf_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| = c_4 n^{-\kappa}} \varpi(\boldsymbol{\pi}_j) - \varpi(\boldsymbol{\pi}_j^0) \geq c_8^* n^{-2\kappa}. \qquad \text{(S.10)}
$$

Let $\boldsymbol{\pi}_j = \boldsymbol{\pi}_j^0 + c_4 n^{-\kappa}\mathbf{u}$ with $||\mathbf{u}|| = 1$. Then, using the identity in Knight (1998), we obtain

$$
\begin{aligned}
& \varpi(\boldsymbol{\pi}_j) - \varpi(\boldsymbol{\pi}_j^0) \\
= \; & c_4 n^{-\kappa} E\left[\mathbf{X}_{i,\mathcal{S}_j}^T \mathbf{u}\{I(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 \leq 0) - \tau\}\right] \\
& + E\left[\int_0^{c_4 n^{-\kappa}\mathbf{X}_{i,\mathcal{S}_j}^T \mathbf{u}} \{I(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 \leq s) - I(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 \leq 0)\} ds\right] \\
= \; & E\int_0^{c_4 n^{-\kappa}\mathbf{X}_{i,\mathcal{S}_j}^T \mathbf{u}} f_{Y|\mathbf{X}}(\zeta) s ds,
\end{aligned}
$$

for ζ between $\mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 + s$ and $\mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0$. By Conditions (C1) and (C2), there exists $c_9^* > 0$ such that

$$
\varpi(\boldsymbol{\pi}_j) - \varpi(\boldsymbol{\pi}_j^0) \geq c_9^* E(c_4 n^{-\kappa}\mathbf{X}_{i,\mathcal{S}_j}^T \mathbf{u})^2 \geq c_9^* c_4^2 m n^{-2\kappa},
$$

which yields (S.10). By (S.9) and (S.10), we have

$$
\begin{aligned}
& P\left(||\widehat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_j^0|| \geq c_4 n^{-\kappa}\right) \leq P(\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \leq c_4 n^{-\kappa}} |\varpi_n(\boldsymbol{\pi}_j) - \varpi(\boldsymbol{\pi}_j)| \geq \frac{1}{2} c_8^* n^{-2\kappa}) \\
\leq \; & P\left(|\varpi_n(\boldsymbol{\pi}_j^0) - \varpi(\boldsymbol{\pi}_j^0)| \geq \frac{1}{2} c_8^* n^{-2\kappa}\right) + \\
& P\left(\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \leq c_4 n^{-\kappa}} |\varpi_n(\boldsymbol{\pi}_j) - \varpi_n(\boldsymbol{\pi}_j^0) - \varpi(\boldsymbol{\pi}_j) + \varpi(\boldsymbol{\pi}_j^0)| \geq \frac{1}{2} c_8^* n^{-2\kappa}\right). \qquad \text{(S.11)}
\end{aligned}
$$

We next derive the bounds for the above two probabilities, respectively. By Condition (C2), $|\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0) - \rho_\tau(Y_i)| \leq \tilde{C} \sup |\mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0| \leq c_{10}^* M_3$ and $\text{var}\Big\{\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0) - \rho_\tau(Y_i)\Big\} \leq c_{11}^*$ for some positive constants $\tilde{C}$, c_{10}^*, and c_{11}^*. This, together with Bernstein's Inequality, leads to

$$
\begin{aligned}
P\left(|\varpi_n(\boldsymbol{\pi}_j^0) - \varpi(\boldsymbol{\pi}_j^0)| \geq \frac{1}{2} c_8^* n^{-2\kappa}\right) & \leq 2\exp\{-\frac{c_8^{*2} n^{2-2\kappa}/4}{2(n c_{11}^* + c_{10}^* M_3 \frac{1}{2} c_8^* n^{1-2\kappa}/3}\} \\
& \leq 2\exp\{-c_{12}^* n^{1-2\kappa}\}, \qquad \text{(S.12)}
\end{aligned}
$$

for some positive constant c_{12}^*.

Define

$$V_{ij}(\boldsymbol{\pi}_j) = \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j) - \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0).$$

Under this definition, the second probability in (S.11) is

$$P(\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \le c_4 n^{-\kappa}} |n^{-1} \sum\nolimits_{i=1}^n V_{ij}(\boldsymbol{\pi}_j) - EV_{ij}(\boldsymbol{\pi}_j)| \ge \frac{1}{2} c_8^* n^{-2\kappa}).$$

Employing the identity in Knight (1998), we have

$$\begin{aligned} V_{ij}(\boldsymbol{\pi}_j) &= \mathbf{X}_{i,\mathcal{S}_j}^T (\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0)\{I(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 \le 0) - \tau\} \\ &+ \int_0^{\mathbf{X}_{i,\mathcal{S}_j}^T (\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0)} \{I(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 \le s) - I(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 \le 0)\} ds. \end{aligned}$$

By Condition (C2), we obtain

$$\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \le c_4 n^{-\kappa}} |V_{ij}(\boldsymbol{\pi}_j)| \le 2 \sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \le c_4 n^{-\kappa}} |\mathbf{X}_{i,\mathcal{S}_j}^T (\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0)| \le c_4 M_1 r_n^{1/2} n^{-\kappa}. \tag{S.13}$$

This, in conjunction with the symmetrization theorem in van der Vaart and Wellner (1996) and the contraction theorem in Ledoux and Talagrand (1991), implies that

$$E[\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \le c_4 n^{-\kappa}} |n^{-1} \sum\nolimits_{i=1}^n V_{ij}(\boldsymbol{\pi}_j) - EV_{ij}(\boldsymbol{\pi}_j)|] \le c_{13}^* r_n^{1/2} n^{-\kappa - 1/2},$$

for some positive constant c_{13}^*. Denote $V = \sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \le c_4 n^{-\kappa}} |n^{-1} \sum_{i=1}^n V_{ij}(\boldsymbol{\pi}_j) - EV_{ij}(\boldsymbol{\pi}_j)|$. Accordingly,

$$\begin{aligned} & P\left(\sup_{||\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^0|| \le c_4 n^{-\kappa}} |n^{-1} \sum\nolimits_{i=1}^n V_{ij}(\boldsymbol{\pi}_j) - EV_{ij}(\boldsymbol{\pi}_j)| \ge \frac{1}{2} c_8^* n^{-2\kappa}\right) \\ = \ & P\left(V \ge E(V) + (\frac{1}{2} c_8^* n^{-2\kappa} - E(V))\right) \\ \le \ & P\left(V \ge E(V) + (\frac{1}{2} c_8^* n^{-2\kappa} - c_{13}^* r_n^{1/2} n^{-\kappa - 1/2})\right). \end{aligned}$$

By (S.13) and Massart's concentration theorem (Massart (2000)), we further have

$$\begin{aligned} & P\left(V \ge E(V) + (\frac{1}{2} c_8^* n^{-2\kappa} - c_{13}^* r_n^{1/2} n^{-\kappa - 1/2})\right) \\ \le \ & \exp\{-\frac{n(\frac{1}{2} c_8^* n^{-2\kappa} - c_{13}^* r_n^{1/2} n^{-\kappa - 1/2})^2}{2(2 c_4 M_1 r_n^{1/2} n^{-\kappa})^2}\} \le \exp\{-c_{14}^* r_n^{-1} n^{1 - 2\kappa}\}, \end{aligned}$$

for some positive constant c^*_{14}. As a result,

$$P\left(\sup_{||\boldsymbol{\pi}_j-\boldsymbol{\pi}_j^0||\le c_4 n^{-\kappa}} |n^{-1}\sum_{i=1}^n V_{ij}(\boldsymbol{\pi}_j) - EV_{ij}(\boldsymbol{\pi}_j)| \ge \frac{1}{2}c_8^* n^{-2\kappa}\right) \le \exp\{-c^*_{14} r_n^{-1} n^{1-2\kappa}\}. \tag{S.14}$$

Consequently, (S.11), (S.12), and (S.14) lead to

$$P\left(||\widehat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_j^0|| \ge c_4 n^{-\kappa}\right) \le 2\exp(-c^*_{12} n^{1-2\kappa}) + \exp(-c^*_{14} r_n^{-1} n^{1-2\kappa}) \le 3\exp(-c_5 r_n^{-1} n^{1-2\kappa}),$$

for some positive constant c_5. □

Proof of Lemma A.4. Using the fact that $E\{\psi_\tau(Y - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\pi}_j^0)\mathbf{X}_{\mathcal{S}_j}\} = \mathbf{0}$, we have

$$E\{\psi_\tau(Y - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\pi}_j^0)(X_j - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\vartheta}_j^0)\} = E\{\psi_\tau(Y - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\pi}_j^0)X_j\}.$$

Denote

$$\begin{aligned} & n^{-1}\sum_{i=1}^n\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\pi}}_j)(X_{ij} - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\vartheta}}_j)\} - E\{\psi_\tau(Y - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\pi}_j^0)X_j\} \\ = \; & \Delta_{1j} + \Delta_{2j} + \Delta_{3j}, \end{aligned} \tag{S.15}$$

where

$$\begin{aligned} \Delta_{1j} &= n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0)X_{ij} - E\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0)X_{ij}\}, \\ \Delta_{2j} &= n^{-1}\sum_{i=1}^n\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\pi}}_j) - \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0)\}X_{ij} \quad \text{and} \\ \Delta_{3j} &= -n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\pi}}_j)\mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\vartheta}}_j. \end{aligned}$$

We next find the probability bounds for Δ_{1j}, Δ_{2j}, and Δ_{3j}. It is worth noting that $|\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0)X_{ij}| \le M_1$. We then apply Bernstein's Inequality and obtain that, for any given $c^*_{15} > 0$, there exists some positive constant c^*_{16} such that

$$P(|\Delta_{1j}| \ge c^*_{15} n^{-\kappa}) \le 2\exp(-c^*_{16} n^{1-2\kappa}). \tag{S.16}$$

By Condition (C1), there exists a $\mathbf{u}^*$ with $||\mathbf{u}^*|| \le 1$ such that

$$\begin{aligned} & E\sup_{||\mathbf{u}||\le 1} |\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}(\boldsymbol{\pi}_j^0 + c_4 n^{-\kappa}\mathbf{u})) - \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0)\}X_{ij}| \\ = \; & E|\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}(\boldsymbol{\pi}_j^0 + c_4 n^{-\kappa}\mathbf{u}^*)) - \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0)\}X_{ij}| \\ \le \; & E(|\int_{\mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j^0}^{\mathbf{X}^T_{i,\mathcal{S}_j}(\boldsymbol{\pi}_j^0 + c_4 n^{-\kappa}\mathbf{u}^*)} f_{Y|\mathbf{X}}(y)dy||X_{ij}|) \\ \le \; & c_4 n^{-\kappa} E|\mathbf{X}^T_{i,\mathcal{S}_j}\mathbf{u}^*| \le c^*_{17} r_n^{1/2} n^{-\kappa}, \end{aligned} \tag{S.17}$$

where $c^*_{17} = M_1 c_4$ and c_4 is given in Lemma A.3. Analogously, we have

$$E \sup_{||\mathbf{u}||\le 1} |\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}(\boldsymbol{\pi}^0_j + c_4 n^{-\kappa}\mathbf{u})) - \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}^0_j)\}X_{ij}|^2 \le c^*_{18} r_n^{1/2} n^{-\kappa}$$

for some positive constant c^*_{18}. Denote

$$\Pi_{ij} = \sup_{||\mathbf{u}||\le 1} |\{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}(\boldsymbol{\pi}^0_j + c_4 n^{-\kappa}\mathbf{u})) - \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}^0_j)\}X_{ij}|.$$

By Bernstein's Inequality, we have that, for any given $c^*_{19} > 0$, there exists some positive constant c^*_{20} such that

$$\begin{aligned} & P\left(|n^{-1}\sum\nolimits_{i=1}^n \Pi_{ij} - E\Pi_{ij}| \ge c^*_{19} r_n^{1/2} n^{-\kappa}\right) \\ \le\ & 2\exp\{-\frac{c^{*2}_{19} r_n n^{2-2\kappa}}{2(c^*_{18} r_n^{1/2} n^{1-\kappa} + 2M_1 c^*_{19} r_n^{1/2} n^{1-\kappa}/3}\} \le 2\exp(-c^*_{20} r_n^{1/2} n^{1-\kappa}). \end{aligned}$$

This, together with (S.17), leads to

$$P\left(|n^{-1}\sum\nolimits_{i=1}^n \Pi_{ij}| \ge c^*_{21} r_n^{1/2} n^{-\kappa}\right) \le 2\exp(-c^*_{20} r_n^{1/2} n^{1-\kappa}).$$

where $c^*_{21} = c^*_{17} + c^*_{19}$ and $c^*_{17} = M_1 c_4$ for any $c_4 > 0$ and $c^*_{19} > 0$. By Lemma A.3, we further obtain

$$\begin{aligned} & P\left(|n^{-1}\sum\nolimits_{i=1}^n \{\psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\pi}}_j) - \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}^0_j)\}X_{ij}| \ge c^*_{21} r_n^{1/2} n^{-\kappa}\right) \\ \le\ & P\left(|n^{-1}\sum\nolimits_{i=1}^n \Pi_{ij}| \ge c^*_{21} r_n^{1/2} n^{-\kappa}\right) + P\left(||\widehat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}^0_j|| \ge c_4 n^{-\kappa}\right) \\ \le\ & 2\exp(-c^*_{20} r_n^{1/2} n^{1-\kappa}) + 3\exp(-c_5 r_n^{-1} n^{1-2\kappa}) \le 5\exp(-c^*_{22} r_n^{-1} n^{1-2\kappa}), \end{aligned}$$

for some positive constant c^*_{22}. Accordingly,

$$P(|\Delta_{2j}| \ge c^*_{21} r_n^{1/2} n^{-\kappa}) \le 5\exp(-c^*_{22} r_n^{-1} n^{1-2\kappa}). \tag{S.18}$$

Denote $g(\boldsymbol{\pi}_j) = n^{-1}\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j)$ and define its subdifferential as $\partial g(\boldsymbol{\pi}_j) = \{\partial g_k(\boldsymbol{\pi}_j) : k \in \{0\} \cup \mathcal{S}_j\}^T$ with

$$\partial g_k(\boldsymbol{\pi}_j) = -n^{-1}\sum\nolimits_{i=1}^n \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j)X_{ik} - n^{-1}\sum\nolimits_{i=1}^n I(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\boldsymbol{\pi}_j = 0)v_i X_{ik},$$

and $v_i \in [\tau-1, \tau]$. By the definition of $\widehat{\boldsymbol{\pi}}_j$, there exists $v^*_i \in [\tau-1, \tau]$ such that $\partial g_k(\widehat{\boldsymbol{\pi}}_j) = 0$. Thus,

$$\begin{aligned} \Delta_{3j} &= -n^{-1}\sum\nolimits_{i=1}^n \psi_\tau(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\pi}}_j)\mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\vartheta}}_j \\ &= n^{-1}\sum\nolimits_{i=1}^n I(Y_i - \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\pi}}_j = 0)v^*_i \mathbf{X}^T_{i,\mathcal{S}_j}\widehat{\boldsymbol{\vartheta}}_j. \end{aligned} \tag{S.19}$$

Moreover,

$$
\begin{aligned}
& \left|n^{-1}\sum\nolimits_{i=1}^{n} I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0)v_i^*\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\vartheta}}_j\right| \\
\leq\ & n^{-1}\sum\nolimits_{i=1}^{n} I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0)|\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\vartheta}}_j| \\
\leq\ & n^{-1}\sum\nolimits_{i=1}^{n} I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0)|\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0| + n^{-1}\sum\nolimits_{i=1}^{n} I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0)r_n^{1/2}M_1||\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0|| \\
\leq\ & n^{-1}\sum\nolimits_{i=1}^{n} I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0)(M_2+r_n^{1/2}M_1||\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0||). \qquad \text{(S.20)}
\end{aligned}
$$

By Lemma A.2 with $\delta_n = nr_n^{-1}$, we have

$$P(||\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0|| \geq c_1) \leq 6r_n^2\exp(-c_8nr_n^{-2}),$$

for some positive constant c_8. As a result,

$$P(M_2+r_n^{1/2}M_1||\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0|| \geq M_2+c_1M_1r_n^{1/2}) \leq 6r_n^2\exp(-c_8nr_n^{-2}). \qquad \text{(S.21)}$$

It is worth noting that $E\{I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0)\}=0$ and $P\{n^{-1}\sum_{i=1}^n I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0) > \epsilon\}=0$ for any $\epsilon>0$. Letting $\epsilon = r_n^{-1/2}n^{-1}$, we thus have

$$P\{n^{-1}\sum\nolimits_{i=1}^{n} I(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j=0) > r_n^{-1/2}n^{-1}\}=0. \qquad \text{(S.22)}$$

This, together with (S.19), (S.20), and (S.21) implies that

$$P\left\{|\Delta_{3j}| \geq r_n^{-1/2}n^{-1}(M_2+c_1M_1r_n^{1/2})\right\} \leq 6r_n^2\exp(-c_8nr_n^{-2}). \qquad \text{(S.23)}$$

By (S.15), (S.16), (S.18) and (S.23), we finally obtain that, for any given positive constants c_{15}^*, c_{21}^*, and c_1,

$$
\begin{aligned}
& P\Big[|n^{-1}\sum\nolimits_{i=1}^{n}\{\psi_\tau(Y_i-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\pi}}_j)(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\vartheta}}_j)\} - E\{\psi_\tau(Y-\mathbf{X}_{\mathcal{S}_j}^T\boldsymbol{\pi}_j^0)X_j\}| \\
& \geq c_{15}^*n^{-\kappa}+c_{21}^*r_n^{1/2}n^{-\kappa}+r_n^{-1/2}n^{-1}(M_2+c_1M_1r_n^{1/2})\Big] \\
\leq\ & 2\exp(-c_{16}^*n^{1-2\kappa})+5\exp(-c_{22}^*r_n^{-1}n^{1-2\kappa})+6r_n^2\exp(-c_8nr_n^{-2}).
\end{aligned}
$$

Accordingly, the result in Lemma A.4 follows for any given constant c_6 and some positive constants c_7 and c_8. □

Proof of Lemma A.5. By the definitions of $\widehat{\sigma}_j^2$ and σ_j^2, we have that $|\widehat{\sigma}_j^2-\sigma_j^2| \leq \tilde{\Upsilon}_{1j} + \tilde{\Upsilon}_{2j}(\widehat{\boldsymbol{\vartheta}}_j)$, where

$$
\begin{aligned}
\tilde{\Upsilon}_{1j} &= |n^{-1}\sum\nolimits_{i=1}^{n}(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)^2 - E(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)^2| \ \text{ and} \\
\tilde{\Upsilon}_{2j}(\widehat{\boldsymbol{\vartheta}}_j) &= |n^{-1}\sum\nolimits_{i=1}^{n}(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\vartheta}}_j)^2 - n^{-1}\sum\nolimits_{i=1}^{n}(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)^2|.
\end{aligned}
$$

By Bernstein's Inequality, we have that, for any given $c_{23}^*>0$,

$$P(\tilde{\Upsilon}_{1j}\geq c_{23}^* r_n^{1/2} n^{-\kappa})\leq 2\exp(-c_{24}^* r_n n^{1-2\kappa}), \tag{S.24}$$

for some positive constant c_{24}^*. In addition,

$$\begin{aligned}
\tilde{\Upsilon}_{2j}(\widehat{\boldsymbol{\vartheta}}_j) &= |n^{-1}\sum_{i=1}^n\{(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\widehat{\boldsymbol{\vartheta}}_j)+(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)\}\{\mathbf{X}_{i,\mathcal{S}_j}^T(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)\}| \\
&= |n^{-1}\sum_{i=1}^n\{2(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)+\mathbf{X}_{i,\mathcal{S}_j}^T(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)\}\{\mathbf{X}_{i,\mathcal{S}_j}^T(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)\}| \\
&\leq |n^{-1}\sum_{i=1}^n\{2(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)\{\mathbf{X}_{i,\mathcal{S}_j}^T(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)\}| \\
&\quad +(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)^T(n^{-1}\sum_{i=1}^n\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0).
\end{aligned} \tag{S.25}$$

By (S.5) and the inequality that $|\lambda_{\max}(\mathbf{A})-\lambda_{\max}(\mathbf{B})|\leq||\mathbf{A}-\mathbf{B}||$ for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, we have

$$P(|\lambda_{\max}(n^{-1}\sum_{i=1}^n\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)-\lambda_{\max}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))|\geq c_3^* r_n n^{-1}\delta_1)\leq 2r_n^2\exp(-c_4^*\delta_1^2 n^{-1}),$$

for any $\delta_1>0$. Then applying the same techniques used in obtaining (S.6), we have that for any constant $c\in(0,1)$, there exists some finite positive constant c_{25}^* such that

$$\begin{aligned}
& P\left(|\lambda_{\max}(n^{-1}\sum_{i=1}^n\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)-\lambda_{\max}(E\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))|\geq c\lambda_{\max}(E(\mathbf{X}_{\mathcal{S}_j}\mathbf{X}_{\mathcal{S}_j}^T))\right) \\
\leq\ & 2r_n^2\exp(-c_{25}^* n r_n^{-2}).
\end{aligned}$$

As a result,

$$\begin{aligned}
& P\left(\lambda_{\max}((n^{-1}\sum_{i=1}^n\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\geq(1+c)\lambda_{\max}(E(\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\right) \\
\leq\ & 2r_n^2\exp(-c_{25}^* n r_n^{-2}).
\end{aligned}$$

This, together with Condition (C2), leads to

$$P\left(\lambda_{\max}((n^{-1}\sum_{i=1}^n\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T))\geq(1+c)M\right)\leq 2r_n^2\exp(-c_{25}^* n r_n^{-2}).$$

Furthermore, we apply Lemma A.2 by letting $\delta_n=c_{26}^* r_n^{-1/2} n^{1-\kappa}$ and obtain

$$P(||\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0||\geq c_1 c_{26}^* r_n^{1/2} n^{-\kappa})\leq 4r_n^2\exp(-c_2 c_{26}^{*2} r_n^{-1} n^{1-2\kappa})+2r_n^2\exp(-c_3 n r_n^{-2}). \tag{S.26}$$

This, in conjunction with the above equation, implies that

$$\begin{aligned}
& P\left((\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)^T(n^{-1}\sum_{i=1}^n\mathbf{X}_{i,\mathcal{S}_j}\mathbf{X}_{i,\mathcal{S}_j}^T)(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)\geq c_1^2 c_{26}^{*2}(1+c)M(r_n^{1/2}n^{-\kappa})^2\right) \\
\leq\ & 4r_n^2\exp(-c_2 c_{26}^{*2} r_n^{-1} n^{1-2\kappa})+2r_n^2\exp(-c_3 n r_n^{-2})+2r_n^2\exp(-c_{25}^* n r_n^{-2}).
\end{aligned} \tag{S.27}$$

By Condition (C2), we have $\sup_{i,j}|X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0|\leq M_1+M_2$. Let $\boldsymbol{\vartheta}_j=\boldsymbol{\vartheta}_j^0+$

$c_1 c_{26}^* r_n^{1/2} n^{-\kappa}\mathbf{u}$ with $\mathbf{u} \in R^{|\mathcal{S}_j|}$ and $||\mathbf{u}|| \leq 1$. Define

$$\Phi_j(\mathbf{u}) = n^{-1}\sum\nolimits_{i=1}^n (X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0)\{\mathbf{X}_{i,\mathcal{S}_j}^T(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}_j^0)\}.$$

Accordingly, $E(\Phi_j(\mathbf{u})) = 0$,

$$\mathrm{var}\{(X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0)\{\mathbf{X}_{i,\mathcal{S}_j}^T(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}_j^0)\}\} \leq c_{27}^{**}(r_n^{1/2} n^{-\kappa})^2,$$

where $c_{27}^{**} = (M_1+M_2)^2 M(c_1 c_{26}^*)^2$, and $|\Phi_j(\mathbf{u})| \leq c_{27}^* r_n n^{-\kappa}$, where $c_{27}^* = (M_1+M_2)M_1 c_1 c_{26}^*$. By Bernstein's inequality, we have that, for any $c_{28}^* > 0$,

$$\begin{aligned} P\left(|\Phi_j(\mathbf{u})| \geq c_{28}^* r_n^{1/2} n^{-\kappa}\right) &\leq 2\exp\left(-\frac{c_{28}^{*2} n^2 (r_n^{1/2} n^{-\kappa})^2}{2(c_{27}^{**} n (r_n^{1/2} n^{-\kappa})^2 + c_{27}^* c_{28}^* n r_n^{1/2} (r_n^{1/2} n^{-\kappa})^2/3)}\right) \\ &\leq 2\exp(-c_{29}^* r_n^{-1/2} n), \end{aligned} \tag{S.28}$$

for some finite positive constant c_{29}^*.

We next partition $\Gamma = \{\mathbf{u} : \mathbf{u} \in R^{|\mathcal{S}_j|}, ||\mathbf{u}|| \leq 1\}$ as a union of l_n disjoint subsets $\Gamma_1, \ldots, \Gamma_{l_n}$. with equal spaces in each direction of $\mathbf{u}$. Clearly, $\sup_{\mathbf{u},\mathbf{u}' \in \Gamma_k} ||\mathbf{u} - \mathbf{u}'|| \leq \sqrt{r_n}/l_n^{1/|\mathcal{S}_j|}$ for all $k = 1, \ldots, l_n$. Choose $\mathbf{u}_k \in \Gamma_k$ for $k = 1, \ldots, l_n$, we then have

$$\sup_{\mathbf{u}\in\Gamma} |\Phi_j(\mathbf{u})| \leq \sup_k |\Phi_j(\mathbf{u}_k)| + \sup_k \sup_{\mathbf{u}\in\Gamma_k} |\Phi_j(\mathbf{u}) - \Phi_j(\mathbf{u}_k)|.$$

By (S.28) and the union rule of probability, we obtain

$$P\left(\sup_k |\Phi_j(\mathbf{u}_k)| \geq c_{28}^* r_n^{1/2} n^{-\kappa}\right) \leq 2 l_n \exp(-c_{29}^* r_n^{-1/2} n).$$

In addition,

$$\begin{aligned} & \sup_k \sup_{\mathbf{u}\in\Gamma_k} |\Phi_j(\mathbf{u}) - \Phi_j(\mathbf{u}_k)| \\ = \; & \sup_k \sup_{\mathbf{u}\in\Gamma_k} |n^{-1}\sum\nolimits_{i=1}^n (X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0)\{\mathbf{X}_{i,\mathcal{S}_j}^T (c_1 c_{26}^* r_n^{1/2} n^{-\kappa}(\mathbf{u} - \mathbf{u}_k))\}| \\ \leq \; & (M_1 + M_2) M_1 c_1 c_{26}^* r_n n^{-\kappa} \sup_k \sup_{\mathbf{u}\in\Gamma_k} ||\mathbf{u} - \mathbf{u}_k|| = c_{30}^* r_n^{3/2} n^{-\kappa}/l_n^{1/|\mathcal{S}_j|} \end{aligned}$$

where $c_{30}^* = (M_1 + M_2) M_1 c_1 c_{26}^*$. By letting $l_n^{1/|\mathcal{S}_j|} = r_n$, the above results imply that for any given $c_{28}^* > 0$ and $c_{30}^* > 0$,

$$P(\sup_{\mathbf{u}\in\Gamma} |\Phi_j(\mathbf{u})| \geq c_{28}^* r_n^{1/2} n^{-\kappa} + c_{30}^* r_n^{1/2} n^{-\kappa}) \leq 2 r_n^{r_n} \exp(-c_{29}^* r_n^{-1/2} n) \leq 2\exp(-c_{32}^* r_n^{-1/2} n),$$

for some positive constant c_{32}^*. Thus, on the event that $||\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}_j^0|| \leq c_1 c_{26}^* r_n n^{-\kappa}$, we have that, for any $c_{31}^* > 0$,

$$P(|n^{-1}\sum\nolimits_{i=1}^n \{2(X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0)\{\mathbf{X}_{i,\mathcal{S}_j}^T(\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}_j^0)\}| \geq c_{31}^* r_n^{1/2} n^{-\kappa}) \leq 2\exp(-c_{32}^* r_n^{-1/2} n).$$

This, in conjunction with (S.26), leads to

$$P\left(|n^{-1}\sum_{i=1}^{n}\{2(X_{ij}-\mathbf{X}_{i,\mathcal{S}_j}^T\boldsymbol{\vartheta}_j^0)\{\mathbf{X}_{i,\mathcal{S}_j}^T(\widehat{\boldsymbol{\vartheta}}_j-\boldsymbol{\vartheta}_j^0)\}|\geq c_{31}^* r_n^{1/2}n^{-\kappa}\right)$$
$$\leq \ 2\exp(-c_{32}^* r_n^{-1/2}n)+4r_n^2\exp(-c_2 c_{26}^{*2} r_n^{-1}n^{1-2\kappa})+2r_n^2\exp(-c_3 n r_n^{-2}). \qquad \text{(S.29)}$$

By (S.25), (S.27), (S.29), and the assumption of $r_n^{1/2}n^{-\kappa}=o(1)$, we obtain that, for any $c_{32}^*>0$, there are finite positive constants c_{33}^* and c_{34}^* such that

$$P\left(|\tilde{\Upsilon}_{2j}(\widehat{\boldsymbol{\vartheta}}_j)|\geq c_{32}^* r_n^{1/2}n^{-\kappa}\right)\leq 8r_n^2\exp(-c_{33}^* r_n^{-1}n^{1-2\kappa})+(2+6r_n^2)\exp(-c_{34}^* n r_n^{-2}). \qquad \text{(S.30)}$$

By (S.24) and (S.30), we have

$$P\left(|\tilde{\Upsilon}_{2j}(\widehat{\boldsymbol{\vartheta}}_j)|\geq c_{32}^* r_n^{1/2}n^{-\kappa}\right)\leq 8r_n^2\exp(-c_{33}^* r_n^{-1}n^{1-2\kappa})+(2+6r_n^2)\exp(-c_{34}^* n r_n^{-2})$$

Letting $c_9=c_{32}^*+c_{23}^*$ for any $c_{23}^*>0$ and $c_{32}^*>0$, $c_{10}=\min(c_{24}^*,c_{33}^*)$, and $c_{11}=c_{34}^*$, we then employ (S.24) and (S.30) to complete the proof of (A.1). Moreover, the assumption that $r_n^{1/2}n^{-\kappa}=o(1)$ implies $c_9 r_n^{1/2}n^{-\kappa}\leq a\sigma_j^2$ for large n. Therefore, the result (A.2) follows directly from (A.1). □

Example S1. The data are generated from the model in Example 1. Table S1 reports R_j ($j=1,\ldots,4$) and M for $p=1,000$ and $\rho=0.5$. We find that even under the moderate correlation ($\rho=0.5$), the SIS approach fails to identify the fourth predictor, which is marginally uncorrelated with Y (see the large values of R_4 in Table S1). Table S2 reports TP and FP calculated under three selection methods, EBIC1, EBIC2 and LASSO, for $p=1,000$ and $\rho=0.5$, and Table S3 correspondingly presents the percentages of correct-fitting (C), over-fitting (O), and incorrect-fitting (I) for $p=1,000$ and $\rho=0.5$. Both tables show that QPCS-EBIC2 performs the best. In Table S4, we report R_j ($j=1,\ldots,4$) and M for $p=2,000$. We observe similar patterns as those given in Tables 2 and S1 for $p=1,000$. Furthermore, Table S5 reports TP and FP calculated under three selection methods, EBIC1, EBIC2 and LASSO, and Table S6 presents the percentages of correct-fitting (C), over-fitting (O), and incorrect-fitting (I) for $p=2,000$. Both tables show similar patterns as those given in Tables 3, S2, 4 and S3, respectively, for $p=1,000$.

Example S2. The data are generated from the model in Example 2. Table S7 reports R_j ($j=1,\ldots,4$) and M for $p=1,000$ and $\rho=0.5$. This table shows that SIS gives large values for R_4, R_5 and M even for a moderately large correlation. Hence, SIS is not able to identify variables X_4 and X_5 in this case. In addition, Tables S8 and S9 summarize the results of subset selection, by presenting TP, FP, and the proportions of correct-fitting (C), over-fitting (O) and incorrect-fitting (I) calculated via EBIC1, EBIC2 and LASSO for $p = 1,000$ and $\rho = 0.5$. Both tables show that QPCS-EBIC2 performs the best, as in Example 1.

Example S3. Case 1. In this example, we compare QPCS with SCAD (Wang et al., 2012) directly so that SCAD is not used as the second-stage variable selection of ISIS-SCAD in Example 3. For the sake of completeness, we also include l_1 (Belloni and Chernozhukov, 2011) for comparison. The data are generated from the model in Example 1, except that the random errors are simulated from the t-distribution with three degrees of freedom. In addition, we use $p = 300$, instead of $p = 1,000$, since SCAD requires a large amount of computation time. Table S10 reports TP and FP, and Table S11 represents the percentages of correct-fitting (C), over-fitting (O), and incorrect-fitting (I). Both tables indicate that QPCS is superior to SCAD and l_1 across all cases, $\rho = 0.95$, 0.5, and 0.05 and $\tau = 0.2$, 0.5 and 0.8.

Case 2. In this example, the data are generated from the model in Example 3, except that the covariates are simulated from the block covariance matrix. Specifically, we generate the first 50 covariates from $N(0, \Sigma_1)$, where $\Sigma_1 = \{\sigma_{ij}\}$ is a 50×50 covariance matrix satisfying $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$, $j \neq i$, except that $\sigma_{4j} = \sigma_{i4} = \sqrt{\rho}$. We next generate the remaining $p - 50$ covariates from $N(0, \Sigma_2)$, where $\Sigma_2 = \{\vartheta_{ij}\}$ is a $(p-50) \times (p-50)$ covariance matrix satisfying $\vartheta_{ij} = 0.5^{|i-j|}$. As given in Example 3, we set $\rho = 0.95$, $n = 200$ and $p = 1,000$, and then compare our proposed QPCS-EBIC2 method with the l_1 penalization and ISIS-SCAD methods. Table S12 reports TP, FP, and the percentages of correct-fitting (C), over-fitting (O), and incorrect-fitting (I). We find that QPCS exhibits patterns similar to those in Tables 8 and 9 obtained via the exchangeable covariance matrix with $\rho = 0.95$. In addition, QPCS outperforms l_1 and ISIS. It is of interest to note that the small number of stronger correlations of covariates induced by Σ_1 and the larger number of weaker correlations of covariates induced by Σ_2 together lead to a small number of true positives in both l_1 and ISIS. In addition, we have conducted simulation studies with $\rho = 0.05$ and 0.5. Tables S13 and S14 show that QPCS is still superior to l_1 and ISIS, although l_1 and ISIS perform better than in the case with $\rho = 0.95$.

Table S1: The average rank of the relevant predictors R_j and the average number of the minimum size of the selected model M with $n = 200$, $p = 1,000$ and $\rho = 0.5$ in Example S1.

		Standard Normal					Laplace Distribution				
τ	Method	R_1	R_2	R_3	R_4	M	R_1	R_2	R_3	R_4	M
						$\rho = 0.5$					
	QPCS	2.020	1.975	2.005	4.000	4.000	1.975	1.985	2.040	4.000	4.000
0.2	QTCS	3.010	2.920	3.135	3.215	4.740	2.995	3.080	3.085	3.290	4.795
	QFS	3.240	3.315	3.375	4.595	5.530	3.390	3.400	3.440	5.205	5.935
	SIS	15.630	17.530	9.890	479.775	488.905	9.520	11.685	9.135	478.410	483.320
	QPCS	1.970	2.090	1.940	4.000	4.00	2.010	2.090	1.900	4.000	4.000
0.5	QTCS	2.885	3.070	2.825	2.960	4.570	3.070	2.920	2.860	3.090	4.635
	QFS	3.060	3.245	2.985	3.845	5.035	3.405	3.085	3.030	4.215	5.250
	SIS	3.770	12.460	3.665	502.095	508.070	15.130	19.245	12.995	508.765	521.350
	QPCS	1.980	2.055	1.965	4.000	4.000	1.955	2.055	1.990	4.000	4.000
0.8	QTCS	2.895	3.210	2.845	3.120	4.675	2.970	3.215	3.045	3.390	4.860
	QFS	3.145	3.565	3.180	4.600	5.535	3.285	3.560	3.355	5.025	5.820
	SIS	12.980	15.485	13.880	508.930	516.920	10.595	11.760	12.015	494.145	502.295

Table S2: Variable selection results of TP and FP for the extended BIC and LASSO with $n = 200$, $p = 1,000$ and $\rho = 0.5$ in Example S1.

			QPCS			TPCS			QFS		
ρ	τ		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
			Standard Normal								
	0.2	TP	4.000	4.000	2.000	4.000	4.000	2.015	4.000	4.000	2.000
		FP	10.270	0.380	8.465	10.455	1.230	9.030	11.030	2.040	8.530
0.5	0.5	TP	4.000	4.000	2.465	4.000	4.000	2.465	4.000	4.000	2.485
		FP	9.560	0.145	12.850	10.080	0.740	12.045	10.310	1.220	11.710
	0.8	TP	4.000	4.000	3.000	4.000	4.000	3.000	4.000	4.000	3.000
		FP	10.465	0.345	9.850	10.420	1.205	9.335	10.855	2.065	8.965
			Laplace Distribution								
	0.2	TP	4.000	4.000	2.000	4.000	4.000	2.015	4.000	4.000	2.005
		FP	9.760	0.270	10.035	10.090	1.255	8.995	10.090	2.350	8.515
0.5	0.5	TP	4.000	4.000	2.535	4.000	4.000	2.435	4.000	4.000	2.510
		FP	6.520	0.015	12.770	7.690	0.675	11.645	8.170	1.285	11.415
	0.8	TP	4.000	4.000	3.000	4.000	4.000	3.000	4.000	4.000	3.000
		FP	10.025	0.370	9.730	10.195	1.410	9.315	10.860	2.350	8.555

Table S3: Variable selection results of C, O, and I for the extended BIC and LASSO with $n = 200$, $p = 1,000$ and $\rho = 0.5$ in Example S1.

			QPCS			QTCS			QFS		
ρ	τ		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
			Standard Normal								
		C	0.000	0.730	0.000	0.000	0.315	0.000	0.000	0.050	0.000
	0.2	O	1.000	0.270	0.000	1.000	0.685	0.000	1.000	0.950	0.000
		U	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.875	0.000	0.000	0.485	0.000	0.000	0.175	0.000
0.5	0.5	O	1.000	0.125	0.000	1.000	0.515	0.000	1.000	0.825	0.000
		U	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.735	0.000	0.000	0.335	0.000	0.000	0.050	0.000
	0.8	O	1.000	0.265	0.000	1.000	0.665	0.000	1.000	0.950	0.000
		U	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
			Laplace Distribution								
		C	0.000	0.795	0.000	0.000	0.325	0.000	0.000	0.035	0.000
	0.2	O	1.000	0.205	0.000	1.000	0.675	0.000	1.000	0.965	0.000
		U	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.050	0.985	0.000	0.000	0.530	0.000	0.000	0.135	0.000
0.5	0.5	O	0.950	0.015	0.000	1.000	0.470	0.000	1.000	0.865	0.000
		U	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.785	0.000	0.000	0.260	0.000	0.000	0.035	0.000
	0.8	O	1.000	0.215	0.000	1.000	0.740	0.000	1.000	0.965	0.000
		U	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000

Table S4: The average rank of the relevant predictors R_j and the average number of the minimum size of the selected model M with $n = 200$ and $p = 2,000$ in Example S1.

		Standard Normal					Laplace Distribution				
τ	Method	R_1	R_2	R_3	R_4	M	R_1	R_2	R_3	R_4	M
						$\rho = 0.5$					
0.2	QPCS	2.035	1.905	2.060	4.000	4.000	1.990	2.000	2.010	4.000	4.000
	QTCS	3.165	3.075	3.155	3.485	4.935	3.080	3.160	3.060	3.465	4.905
	QFS	3.535	3.400	3.470	5.435	6.040	3.375	3.540	3.520	5.585	6.195
	SIS	25.920	17.245	23.870	1003.665	1017.36	22.950	32.175	26.215	1003.965	1021.29
0.5	QPCS	2.010	2.020	1.970	4.000	4.000	2.035	1.970	1.995	4.000	4.000
	QTCS	3.140	3.000	2.775	3.160	4.680	3.045	3.125	2.930	3.345	4.800
	QFS	3.400	3.205	2.980	4.250	5.290	3.335	3.395	3.095	4.565	5.500
	SIS	9.795	18.610	7.820	1054.395	1066.505	9.575	20.115	11.255	1001.745	1012.950
0.8	QPCS	2.075	2.010	1.915	4.000	4.000	2.065	1.965	1.970	4.000	4.000
	QTCS	2.995	3.105	3.140	3.415	4.865	2.995	3.270	2.930	3.385	4.850
	QFS	3.245	3.405	3.555	5.110	5.860	3.360	3.660	3.450	5.615	6.215
	SIS	14.590	20.050	14.905	1006.765	1016.415	19.695	20.400	27.790	1034.740	1049.395
						$\rho = 0.95$					
0.2	QPCS	2.035	2.110	2.175	3.895	4.125	10.420	19.160	6.270	33.445	55.655
	QTCS	4.150	4.575	4.255	9.650	11.220	30.840	29.810	7.745	37.595	81.770
	QFS	7.250	5.410	4.560	660.760	660.830	25.385	18.910	22.760	981.605	994.365
	SIS	677.930	653.695	675.665	995.105	1355.240	681.825	702.625	714.190	986.870	1370.82
0.5	QPCS	4.125	2.010	2.200	12.490	14.725	4.070	7.580	4.470	22.450	28.550
	QTCS	8.095	4.345	7.250	27.700	33.465	15.985	25.365	15.110	72.465	97.975
	QFS	11.755	4.435	6.645	899.075	903.865	17.285	21.755	12.395	976.545	980.015
	SIS	531.075	558.745	538.965	1026.580	1332.385	729.425	744.340	721.420	1008.755	1418.55
0.8	QPCS	2.055	2.030	2.115	4.970	5.085	20.030	6.375	11.655	29.015	52.110
	QTCS	4.105	8.920	5.045	9.590	15.520	19.170	25.025	31.305	31.500	81.320
	QFS	4.725	4.390	8.300	683.525	683.570	38.175	33.615	25.365	987.695	1014.255
	SIS	614.225	601.260	608.625	1015.980	1345.145	623.695	613.790	620.420	1006.725	1352.010

Table S5: Variable selection results of TP and FP for the extended BIC and LASSO with $n = 200$ and $p = 2,000$ in Example S1.

			QPCS			TPCS			QFS		
ρ	τ		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
			Standard Normal								
	0.2	TP	4.000	4.000	2.000	4.000	4.000	2.070	4.000	4.000	2.000
		FP	11.150	0.770	9.970	11.010	1.845	9.930	11.260	3.025	9.395
0.5	0.5	TP	4.000	4.000	2.460	4.000	4.000	2.295	4.000	4.000	2.170
		FP	10.600	0.260	12.245	10.730	0.970	12.535	11.150	1.575	12.335
	0.8	TP	4.000	4.000	3.000	4.000	4.000	3.000	4.000	4.000	2.535
		FP	11.310	0.555	9.995	11.215	1.645	10.775	11.490	2.845	10.820
	0.2	TP	4.000	3.990	1.830	3.890	3.730	1.860	3.220	3.010	1.280
		FP	11.090	0.740	1.470	11.190	3.040	1.610	12.210	5.115	2.190
0.95	0.5	TP	3.980	3.955	2.210	3.670	3.035	1.895	3.065	2.330	1.240
		FP	10.730	0.325	3.010	11.175	2.165	4.910	12.085	3.265	5.885
	0.8	TP	3.995	3.985	2.835	3.915	3.785	2.180	3.195	3.000	1.295
		FP	11.225	0.725	1.450	11.480	3.055	2.225	12.360	5.375	3.190
			Laplace Distribution								
	0.2	TP	4.000	4.000	2.000	4.000	4.000	2.000	4.000	4.000	2.000
		FP	10.975	0.470	10.360	10.750	1.670	9.415	11.250	2.925	8.645
0.5	0.5	TP	4.000	4.000	2.465	4.000	4.000	2.530	4.000	4.000	2.505
		FP	8.115	0.030	12.760	9.530	0.845	12.045	10.295	1.545	11.710
	0.8	TP	4.000	4.000	3.000	4.000	4.000	2.555	4.000	4.000	2.120
		FP	10.940	0.565	10.200	10.895	1.530	11.125	11.615	3.020	10.740
	0.2	TP	3.830	3.575	1.440	3.405	2.575	1.520	2.735	1.965	1.080
		FP	11.290	1.105	1.370	11.420	2.640	2.500	12.335	3.615	3.065
0.95	0.5	TP	3.900	3.735	2.205	3.320	2.025	1.705	2.810	1.405	1.210
		FP	8.625	0.260	3.800	10.380	1.525	5.520	11.110	1.920	6.365
	0.8	TP	3.855	3.505	2.290	3.400	2.565	1.805	2.760	1.930	1.170
		FP	10.875	1.045	1.600	11.695	2.635	3.260	12.490	3.335	3.870

Table S6: Variable selection results of C, O, and I for the extended BIC and LASSO with $n = 200$ and $p = 2,000$ in Example S1.

			QPCS			QTCS			QFS		
ρ	τ		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
			Standard normal								
		C	0.000	0.530	0.000	0.000	0.200	0.000	0.000	0.015	0.000
	0.2	O	1.000	0.470	0.000	1.000	0.800	0.000	1.000	0.985	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.805	0.000	0.000	0.405	0.000	0.000	0.110	0.000
0.5	0.5	O	1.000	0.195	0.000	1.000	0.595	0.000	1.000	0.890	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.630	0.000	0.000	0.205	0.000	0.000	0.010	0.000
	0.8	O	1.000	0.370	0.000	1.000	0.795	0.000	1.000	0.990	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.490	0.000	0.000	0.095	0.000	0.000	0.005	0.000
	0.2	O	1.000	0.500	0.000	0.900	0.745	0.000	0.255	0.250	0.000
		I	0.000	0.010	1.000	0.100	0.160	1.000	0.745	0.745	1.000
		C	0.000	0.735	0.000	0.000	0.125	0.000	0.000	0.005	0.000
0.95	0.5	O	0.980	0.230	0.000	0.720	0.440	0.000	0.120	0.105	0.000
		I	0.020	0.035	1.000	0.280	0.435	1.000	0.880	0.890	1.000
		C	0.000	0.555	0.000	0.000	0.100	0.000	0.000	0.000	0.000
	0.8	O	0.995	0.440	0.000	0.925	0.770	0.000	0.220	0.220	0.000
		I	0.005	0.005	1.000	0.075	0.230	1.000	0.780	0.780	1.000
			Laplace Distribution								
		C	0.000	0.690	0.000	0.000	0.260	0.000	0.000	0.025	0.000
	0.2	O	1.000	0.310	0.000	1.000	0.740	0.000	1.000	0.975	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.015	0.970	0.000	0.000	0.420	0.000	0.000	0.080	0.000
0.5	0.5	O	0.985	0.030	0.000	1.000	0.580	0.000	1.000	0.920	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.595	0.000	0.000	0.300	0.000	0.000	0.025	0.000
	0.8	O	1.000	0.405	0.000	1.000	0.700	0.000	1.000	0.975	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	1.000
		C	0.000	0.365	0.000	0.000	0.060	0.000	0.000	0.000	0.000
	0.2	O	0.860	0.405	0.000	0.565	0.315	0.000	0.005	0.000	0.000
		I	0.140	0.230	1.000	0.435	0.625	1.000	0.995	1.000	1.000
		C	0.005	0.760	0.000	0.000	0.095	0.000	0.000	0.000	0.000
0.95	0.5	O	0.930	0.115	0.000	0.495	0.180	0.000	0.035	0.030	0.000
		I	0.065	0.125	1.000	0.505	0.725	1.000	0.965	0.970	1.000
		C	0.000	0.375	0.005	0.000	0.040	0.000	0.000	0.000	0.000
	0.8	O	0.885	0.365	0.005	0.570	0.295	0.000	0.015	0.015	0.000
		I	0.115	0.260	0.990	0.430	0.665	1.000	0.985	0.985	1.000

Table S7: The average rank of the relevant predictors R_j and the average minimum size of the selected model M with $n = 200$, $p = 1,000$ and $\rho = 0.5$ in Example S2.

		Standard normal						Laplace Distribution					
τ	Method	R_1	R_2	R_3	R_4	R_5	M	R_1	R_2	R_3	R_4	R_5	M
							$\rho = 0.5$						
0.2	QPCS	2.065	1.890	2.045	4.000	5.000	5.000	1.990	1.980	2.035	3.995	5.165	5.165
	QTCS	3.080	3.065	3.010	3.240	5.670	5.750	3.125	3.180	2.960	3.410	5.810	5.920
	QFS	3.390	3.345	3.390	4.905	6.630	6.735	3.375	3.685	3.165	5.260	6.940	7.110
	SIS	33.120	34.420	21.135	476.265	164.745	536.495	8.760	12.885	15.680	496.275	143.240	540.005
0.5	QPCS	1.915	2.055	2.030	4.000	5.000	5.000	2.040	2.080	1.880	4.000	5.020	5.020
	QTCS	2.900	2.955	2.935	3.010	5.440	5.570	3.115	2.965	2.820	3.160	5.630	5.685
	QFS	3.105	3.235	3.225	4.150	6.035	6.210	3.370	3.145	2.980	4.295	6.185	6.275
	SIS	7.490	14.315	10.620	509.160	104.955	546.700	8.840	14.130	5.470	511.065	127.065	547.09
0.8	QPCS	2.005	2.100	1.910	4.005	4.980	5.000	1.980	2.060	1.960	4.000	8.490	8.490
	QTCS	3.040	3.440	2.985	3.430	5.660	5.880	3.180	3.095	3.005	3.465	5.955	6.050
	QFS	3.355	3.815	3.180	5.095	6.525	6.825	3.480	3.480	3.375	5.370	7.235	7.370
	SIS	14.555	19.735	12.790	503.950	152.775	569.655	19.550	10.745	14.795	514.765	184.595	577.355

Table S8: Variable selection results of TP and FP for the extended BIC and LASSO with $n = 200$, $p = 1,000$ and $\rho = 0.5$ in Example S2.

			QPCS			QTCS			QFS		
ρ	τ		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
						Standard Normal					
	0.2	TP	5.000	5.000	3.000	5.000	5.000	3.005	5.000	5.000	3.000
		FP	9.435	0.335	9.320	9.775	1.290	8.700	10.080	2.300	8.445
0.5	0.5	TP	5.000	5.000	3.455	5.000	5.000	3.450	5.000	4.995	3.465
		FP	8.455	0.115	12.105	9.165	0.740	11.700	9.625	1.390	11.260
	0.8	TP	5.000	5.000	4.000	5.000	5.000	4.000	5.000	5.000	4.000
		FP	9.785	0.305	9.655	9.970	1.355	8.640	10.215	2.350	8.200
						Laplace Distribution					
	0.2	TP	5.000	4.975	3.000	5.000	4.980	3.005	4.995	4.985	3.005
		FP	9.785	0.360	9.290	9.480	1.310	8.785	9.980	2.395	8.170
0.5	0.5	TP	5.000	4.985	3.530	5.000	4.990	3.460	5.000	4.995	3.450
		FP	6.400	0.015	12.285	7.345	0.705	11.430	7.885	1.305	11.075
	0.8	TP	4.990	4.975	3.990	4.995	4.985	4.000	4.985	4.970	4.000
		FP	9.360	0.380	9.555	9.415	1.390	8.975	10.085	2.630	8.420

Table S9: Variable selection results of C, O, and I for the extended BIC and LASSO with $n = 200$, $p = 1,000$ and $\rho = 0.5$ in Example S2.

			QPCS			QTCS			QFS		
ρ	τ		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
			Standard Normal								
		C	0.000	0.750	0.000	0.000	0.310	0.000	0.000	0.045	0.000
	0.2	O	1.000	0.250	0.000	1.000	0.690	0.000	1.000	0.955	0.000
		I	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		C	0.000	0.895	0.000	0.000	0.485	0.000	0.000	0.140	0.000
0.5	0.5	O	1.000	0.105	0.000	1.000	0.515	0.005	1.000	0.855	0.005
		I	0.000	0.000	1.000	0.000	0.000	0.995	0.000	0.005	0.995
		C	0.000	0.770	0.000	0.000	0.300	0.000	0.000	0.020	0.000
	0.8	O	1.000	0.230	0.005	1.000	0.700	0.000	1.000	0.980	0.000
		I	0.000	0.000	0.995	0.000	0.000	1.000	0.000	0.000	1.000
			Laplace Distribution								
		C	0.000	0.720	0.000	0.000	0.005	0.000	0.000	0.055	0.000
	0.2	O	1.000	0.255	0.000	1.000	0.995	0.000	1.000	0.930	0.000
		I	0.000	0.025	1.000	0.000	0.000	1.000	0.000	0.015	1.000
		C	0.000	0.970	0.000	0.000	0.005	0.000	0.000	0.135	0.000
0.5	0.5	O	1.000	0.015	0.000	1.000	0.995	0.000	1.000	0.860	0.000
		I	0.000	0.015	1.000	0.000	0.000	1.000	0.000	0.005	1.000
		C	0.000	0.690	0.000	0.000	0.005	0.000	0.000	0.030	0.000
	0.8	O	1.000	0.285	0.000	1.000	0.995	0.000	1.000	0.940	0.000
		I	0.000	0.025	1.000	0.000	0.000	1.000	0.000	0.030	1.000

Table S10: Variable selection results of TP and FP for the QPCS, l_1 and SCAD methods with $p = 300$ in Case 1 of Example S3.

		$\rho = 0.95$			$\rho = 0.50$			$\rho = 0.05$		
τ		QPCS	l_1	SCAD	QPCS	l_1	SCAD	QPCS	l_1	SCAD
0.2	TP	3.830	0.625	0.500	4.000	3.505	2.505	4.000	4.000	4.000
	FP	0.165	0.570	0.580	0.000	60.020	3.505	0.000	24.045	0.000
0.5	TP	3.995	0.395	0.385	4.000	3.000	3.495	4.000	4.000	4.000
	FP	0.000	0.320	0.425	0.000	42.945	3.540	0.000	15.025	1.005
0.8	TP	4.000	0.645	0.515	4.000	3.500	2.995	4.000	4.000	4.000
	FP	0.005	0.590	0.505	0.000	60.020	3.965	0.000	24.045	0.000

Table S11: Variable selection results of C, O, and I for the QPCS, l_1 and SCAD methods with $p = 300$ in Case 1 of Example S3.

		$\rho = 0.95$			$\rho = 0.50$			$\rho = 0.05$		
τ		QPCS	l_1	SCAD	QPCS	l_1	SCAD	QPCS	l_1	SCAD
	C	0.830	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000
0.2	O	0.080	0.000	0.000	0.000	0.500	0.000	0.000	1.000	0.000
	I	0.090	1.000	1.000	0.000	0.500	1.000	0.000	0.000	0.000
	C	0.995	0.000	0.000	1.000	0.000	0.495	1.000	0.000	0.000
0.5	O	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000
	I	0.005	1.000	1.000	0.000	1.000	0.505	0.000	0.000	0.000
	C	0.995	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000
0.8	O	0.005	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	I	0.000	1.000	1.000	0.000	1.000	1.000	0.000	0.000	0.000

Table S12: Variable selection results for the QPCS, l_1, and ISIS methods with $n = 200$, $p = 1,000$, $\rho = 0.95$ and $\tau = 0.5$ in Case 2 of Example S3.

	QPCS	l_1	ISIS	QPCS	l_1	ISIS
	Standard Normal			Laplace Distribution		
TP	3.945	0.195	0.185	3.780	0.160	0.160
FP	0.110	24.910	2.085	0.100	25.550	2.190
C	0.860	0.000	0.000	0.835	0.000	0.000
O	0.095	0.000	0.005	0.070	0.000	0.010
I	0.045	1.000	0.995	0.095	1.000	0.990

Table S13: Variable selection results for the QPCS, l_1, and ISIS methods with $n = 200$, $p = 1,000$, $\rho = 0.5$ and $\tau = 0.5$ in Case 2 of Example S3.

	QPCS	l_1	ISIS	QPCS	l_1	ISIS
	Standard Normal			Laplace Distribution		
TP	4.000	3.970	4.000	4.000	3.795	3.980
FP	0.055	31.600	1.005	0.010	31.715	1.660
C	0.950	0.000	0.450	0.990	0.000	0.370
O	0.050	0.970	0.550	0.010	0.800	0.625
I	0.000	0.030	0.000	0.000	0.200	0.005

Table S14: Variable selection results for the QPCS, l_1, and ISIS methods with $n = 200$, $p = 1,000$, $\rho = 0.05$ and $\tau = 0.5$ in Case 2 of Example S3.

	QPCS	l_1	ISIS	QPCS	l_1	ISIS
	Standard Normal			Laplace Distribution		
TP	4.000	4.000	4.000	4.000	4.000	4.000
FP	0.065	23.715	1.005	0.025	24.515	2.145
C	0.940	0.000	0.475	0.975	0.000	0.200
O	0.060	1.000	0.525	0.025	1.000	0.800
I	0.000	0.000	0.000	0.000	0.000	0.000

References

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse model. *Annals of Statistics*, 95, 759-771.

Fan, J. and Lv, J. (2008). Sure independence screening for ultra-high dimensional feature space (with discussion). *Journal of the Royal Statistical Society: Series B*, 70, 849-911.

Knight, K. (1998). Limiting distribution for L_1 regression estimators under general conditions. *Annals of Statistics,* 26, 755-770.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete (3).* Berlin: Springer.

Massart, P. (2000). Some applications of concentration inequalities to statistics. *Annales de la faculté des sciences de Toulouse*. 9, 245-303.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, New York: Springer.

Wang, L., Wu, Y., and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107, 214-222.

Table S15: A list of some acronyms used in the manuscript.

Acronym	**Definition**
FR	forward regression
QPCOR	quantile partial correlation
EBIC	extended Bayesian information criterion
ISIS	iterative sure independent screening
MCP	minimax concave penalty
QCOR	quantile correlation
QPCS	quantile partial correlation screening
QTCS	quantile tilted correlation screening
QFR	quantile forward regression
SCAD	smoothly clipped absolute deviation
SIS	sure independent screening
TCS2	tilted correlation screening