Copula-based Partial Correlation Screening: a Joint and Robust Approach

Xiaochao Xia[1,2] and Jialiang Li[2,3,4]

[1] *Huazhong Agricultural University*

[2] *National University of Singapore,* [3] *Duke-NUS Graduate Medical School,* [4] *Singapore Eye Research Institute*

Supplementary Material

In this supplementary material, we provide the detailed proofs of the main theorems that are stated in the main manuscript and some additional simulation results for our proposed estimators and variable screening approaches.

Appendix A: Proofs

The following Lemma is a well-known result for the bivariate empirical process theory.

Lemma S1.1. *(Lemma 3.9.28, van der Vaart and Wellner (1996)) Fix* $0 < a < b < 1$*, and suppose that* H *is a distribution function on* $\mathbb{R}^2$ *with marginal distribution functions* F *and* G *that are continuously differentiable on the intervals* $[F^{-1}(a) - \varepsilon, F^{-1}(b) + \varepsilon]$ *and* $[G^{-1}(a) - \varepsilon, G^{-1}(b) + \varepsilon]$ *with positive derivatives* f *and* g*, respectively, for some* $\varepsilon > 0$*. Furthermore, assume that* $\partial H/\partial x$ *and* $\partial H/\partial y$ *exist and are continuous on the product of these intervals. Let* $\phi : D(\bar{\mathbb{R}}) \mapsto \ell^\infty([a, b]^2)$ *is the map from bivariate distribution function* H *on* $\mathbb{R}^2$ *to bivariate distribution on* $[a, b]^2$*, defined by* $\phi(H)(u, v) = H(F^{-1}(u), G^{-1}(v))$*.*

Then the map ϕ is Hadamard-differentiable at H tangentially to $C(\bar{\mathbb{R}})$. The derivative is given by

$$\begin{aligned}\dot{\phi}_H(h)(u,v) &= h(F^{-1}(u),G^{-1}(v)) - \frac{\partial H}{\partial x}(F^{-1}(u),G^{-1}(v))\frac{h(F^{-1}(u),\infty)}{f(F^{-1}(u))}\\ &\quad -\frac{\partial H}{\partial y}(F^{-1}(u),G^{-1}(v))\frac{h(\infty,G^{-1}(v)\}}{g(G^{-1}(v))}.\end{aligned}$$

Proof of Theorem 2.1 By the definition, we have

$$\sqrt{n}\{\widehat{\varrho}_{Y,X}(\tau,\iota)-\varrho_{Y,X}(\tau,\iota)\} = \frac{\sqrt{n}[F_{n,Y,X}(F_{n,Y}^{-1}(\tau),F_{n,X}^{-1}(\iota))-F_{Y,X}(F_Y^{-1}(\tau),F_X^{-1}(\iota))]}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}.$$

Employing Lemma S1.1 and the empirical process theory (Example 3.9.29 in van der Vaart and Wellner (1996)), we are able to show that

$$\sqrt{n}[F_{n,Y,X}(F_{n,Y}^{-1}(\tau),F_{n,X}^{-1}(\iota))-F_{Y,X}(F_Y^{-1}(\tau),F_X^{-1}(\iota))] \stackrel{a}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\varphi(X_i,Y_i;\tau,\iota)-\mathrm{E}\{\varphi(X_i,Y_i;\tau,\iota)\}],$$

where $\stackrel{a}{=}$ means "asymptotically equivalent" and $\varphi(X_i,Y_i;\tau,\iota)$ is given by

$$\begin{aligned}\varphi(Y_i,X_i;\tau,\iota) &= I\{Y_i\le F_Y^{-1}(\tau),X_i\le F_X^{-1}(\iota)\}\\ &\quad -\frac{\partial}{\partial y}F_{Y,X}(F_Y^{-1}(\tau),F_X^{-1}(\iota))\frac{I\{Y_i\le F_Y^{-1}(\tau)\}}{f_Y(F_Y^{-1}(\tau))}\\ &\quad -\frac{\partial}{\partial x}F_{Y,X}(F_Y^{-1}(\tau),F_X^{-1}(\iota))\frac{I\{X_i\le F_X^{-1}(\iota)\}}{f_X(F_X^{-1}(\iota))}.\end{aligned}$$

Furthermore, it is easily verified that

$$\frac{\partial}{\partial y}F_{Y,X}(F_Y^{-1}(\tau),F_X^{-1}(\iota)) = f_Y(F_Y^{-1}(\tau))F_{X|Y=F_Y^{-1}(\tau)}(F_X^{-1}(\iota)), \tag{S1.1}$$

$$\frac{\partial}{\partial x}F_{Y,X}(F_Y^{-1}(\tau),F_X^{-1}(\iota)) = f_X(F_X^{-1}(\iota))F_{Y|X=F_X^{-1}(\iota)}(F_Y^{-1}(\tau)). \tag{S1.2}$$

Hence, by (S1.1) and (S1.2), it follows that

$$\sqrt{n}\{\widehat{\varrho}_{Y,X}(\tau,\iota)-\varrho_{Y,X}(\tau,\iota)\} \stackrel{a}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\xi(Y_i,X_i;\tau,\iota)-\mathrm{E}\{\xi(Y_i,X_i;\tau,\iota)\}],$$

where $\xi(Y, X; \tau, \iota)$ is defined in the line before Theorem 2.1.

Because the class of indicator functions is Donsker, thus $\{\xi(Y_i, X_i; \tau, \iota), (\tau, \iota) \in [a, b]^2\}$ is Donsker. By Donsker's theorem,

$$\sqrt{n}\{\widehat{\varrho}_{Y,X}(\tau, \iota) - \varrho_{Y,X}(\tau, \iota)\} \overset{w}{\rightsquigarrow} \mathbb{G}_{Y,X}(\tau, \iota)$$

in $\ell^\infty([a, b]^2)$, where $\overset{w}{\rightsquigarrow}$ denotes "converge weakly", and the process $\mathbb{G}_{Y,X}(\tau, \iota)$ has mean zero and covariance function $\mathrm{cov}(\mathbb{G}_{Y,X}(\tau_1, \iota_1), \mathbb{G}_{Y,X}(\tau_2, \iota_2)) = \Omega_1(\tau_1, \iota_1; \tau_2, \iota_2)$ given in Theorem 2.1. □

Proof of Theorem 2.2 Following the arguments in the proof of Theorem 2.1, we can show

$$\begin{aligned}
&\sqrt{n}\{[\widehat{\varrho}_{Y,X_1}(\tau, \iota) - \varrho_{Y,X_1}(\tau, \iota)] - [\widehat{\varrho}_{Y,X_2}(\tau, \iota) - \varrho_{Y,X_2}(\tau, \iota)]\} \\
&\overset{a}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\eta(Y_i, X_{i1}, X_{i2}; \tau, \iota) - \mathrm{E}\{\eta(Y_i, X_{i1}, X_{i2}; \tau, \iota)\}]
\end{aligned}$$

where $\eta(Y_i, X_{i1}, X_{i2}; \tau, \iota) = \xi(Y_i, X_{i1}; \tau, \iota) - \xi(Y_i, X_{i2}; \tau, \iota)$ and $\xi(Y, X; \tau, \iota)$ is given in Theorem 2.1. Since $\{\xi(X_i, Y_i; \tau, \iota), (\tau, \iota) \in [a, b]^2\}$ is Donsker, $\{\eta(Y_i, X_{i1}, X_{i2}; \tau, \iota), (\tau, \iota) \in [a, b]^2\}$ is also Donsker. By Donsker's Theorem, we have

$$\sqrt{n}\{[\widehat{\varrho}_{Y,X_1}(\tau, \iota) - \widehat{\varrho}_{Y,X_2}(\tau, \iota)] - [\varrho_{Y,X_1}(\tau, \iota) - \varrho_{Y,X_2}(\tau, \iota)]\} \overset{w}{\rightsquigarrow} \mathbb{G}_{Y,X_1,X_2}(\tau, \iota)$$

in $\ell^\infty([a, b]^2)$, where the process $\mathbb{G}_{Y,X_1,X_2}(\tau, \iota)$ has mean zero and covariance function $\mathrm{cov}(\mathbb{G}_{Y,X_1,X_2}(\tau_1, \iota_1), \mathbb{G}_{Y,X_1,X_2}(\tau_2, \iota_2)) = \Xi(\tau_1, \iota_1; \tau_2, \iota_2)$ given in Theorem 2.2. □

Proof of Theorem 2.3 (i) We first prove the first statement. Observe that

$$F_{n,Y,X_j}(F_{n,Y}^{-1}(\tau), F_{n,X_j}^{-1}(\iota)) - F_{Y,X_j}(F_Y^{-1}(\tau), F_{X_j}^{-1}(\iota)) = II_{nj1} + II_{nj2} \tag{S1.3}$$

where $II_{nj1} = F_{n,Y,X_j}(F^{-1}_{n,Y}(\tau), F^{-1}_{n,X_j}(\iota)) - F_{n,Y,X_j}(F^{-1}_{Y}(\tau), F^{-1}_{X_j}(\iota)) = n^{-1}\sum_{i=1}^n \big[I(Y_i \le F^{-1}_{n,Y}(\tau), X_{ij} \le F^{-1}_{n,X_j}(\iota)) - I(Y_i \le F^{-1}_{Y}(\tau), X_{ij} \le F^{-1}_{X_j}(\iota))\big]$ and $II_{nj2} = F_{n,Y,X_j}(F^{-1}_{Y}(\tau), F^{-1}_{X_j}(\iota)) - F_{Y,X_j}(F^{-1}_{Y}(\tau), F^{-1}_{X_j}(\iota)) = n^{-1}\sum_{i=1}^n \big[I(Y_i \le F^{-1}_{Y}(\tau), X_{ij} \le F^{-1}_{X_j}(\iota)) - F_{Y,X_j}(F^{-1}_{Y}(\tau), F^{-1}_{X_j}(\iota))\big]$. In what follows, we are going to bound the tail probabilities for II_{nj1} and II_{nj2}, respectively.

Since $|I(Y_i \le F^{-1}_{Y}(\tau), X_{ij} \le F^{-1}_{X_j}(\iota)) - F_{Y,X_j}(F^{-1}_{Y}(\tau), F^{-1}_{X_j}(\iota))| \le 2$ and $\mathrm{var}(I(Y_i \le F^{-1}_{Y}(\tau), X_{ij} \le F^{-1}_{X_j}(\iota))) \le 1$, applying Bernstein's inequality (Lemma 2.2.9, van der Vaart and Wellner (1996)) gives that for any $\delta > 0$,

$$\max_{1\le j\le p_n} P(|II_{nj2}| > \delta/n) \le 2\exp(-\frac{\delta^2}{2(n+2\delta/3)}). \tag{S1.4}$$

Next consider II_{nj1}. Note that

$$\begin{aligned} II_{nj1} &= n^{-1}\sum_{i=1}^n I(Y_i \le F^{-1}_{n,Y}(\tau))\big[I(X_{ij} \le F^{-1}_{n,X_j}(\iota)) - I(X_{ij} \le F^{-1}_{X_j}(\iota))\big] \\ &\quad + n^{-1}\sum_{i=1}^n \big[I(Y_i \le F^{-1}_{n,Y}(\tau)) - I(Y_i \le F^{-1}_{Y}(\tau))\big] I(X_{ij} \le F^{-1}_{X_j}(\iota)) \\ &\hat{=}\ II^{(1)}_{nj1} + II^{(2)}_{nj1} \quad \text{(say)}. \end{aligned} \tag{S1.5}$$

On the other hand, it follows from Lemma B of Serfling (1980) (p.96) that $|F^{-1}_{n,X_j}(\iota) - F^{-1}_{X_j}(\iota)| \le 2n^{-1/2}(\log n)^{1/2}/f_{X_j}(F^{-1}_{X_j}(\iota))$ almost surely. Let $M_1^{-1} = \inf_{1\le j\le p_n} f_{X_j}(F^{-1}_{X_j}(\iota))/2$, which is positive under condition (C1). On the event $\{|F^{-1}_{n,X_j}(\iota) - F^{-1}_{X_j}(\iota)| \le M_1 n^{-\varsigma_1}\}$ with any $\varsigma_2 < 1/2$. we have $|II^{(1)}_{nj1}| \le n^{-1}\sum_{i=1}^n |I(X_{ij} \le F^{-1}_{n,X_j}(\iota)) - I(X_{ij} \le F^{-1}_{X_j}(\iota))| \le \frac{1}{n}\sum_{i=1}^n q_\iota(X_{ij}, F^{-1}_{X_j}(\iota))$, where $q_\iota(X_{ij}, F^{-1}_{X_j}(\iota)) = I(F^{-1}_{X_j}(\iota) - M_1 n^{-\varsigma_1} \le X_{ij} \le F^{-1}_{X_j}(\iota) + M_1 n^{-\varsigma_1})$. Because $\mathrm{E}\{q_\iota(X_{ij}, F^{-1}_{X_j}(\iota))\} = \int_{F^{-1}_{X_j}(\iota) - M_1 n^{-\varsigma_1}}^{F^{-1}_{X_j}(\iota) + M_1 n^{-\varsigma_1}} f_{X_j}(s)\mathrm{d}s \le c_1 M_1 n^{-\varsigma_1}$ for some positive constant c_1 by condition (C1). Since $q_\iota(X_{ij}, F^{-1}_{X_j}(\iota)) \le 1$ and $\mathrm{var}(q_\iota(X_{ij}, F^{-1}_{X_j}(\iota))) \le$

COX SCREENING

$\mathrm{E}\{q_\iota(X_{ij}, F_{X_j}^{-1}(\iota))\}$, so applying Bernstein's inequality (Lemma 2.2.9, van der Vaart and Wellner (1996)) yields that

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}[q_\iota(X_{ij}, F_{X_j}^{-1}(\iota)) - \mathrm{E}\{q_\iota(X_{ij}, F_{X_j}^{-1}(\iota))\}]\Big| \geq \frac{\delta}{2n}\Big) \leq 2\exp(-\frac{\delta^2}{8(c_1M_1n^{1-\varsigma_1}+\delta/3)}).$$

This further implies that

$$\begin{aligned}
\max_{1\leq j\leq p_n} P(|II_{nj1}^{(1)}| \geq \delta/n) &\leq P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} q_\iota(X_{ij}, F_{X_j}^{-1}(\iota))\Big| \geq \delta/n\Big) \\
&\leq P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}[q_\iota(X_{ij}, F_{X_j}^{-1}(\iota)) - \mathrm{E}\{q_\iota(X_{ij}, F_{X_j}^{-1}(\iota))\}]\Big| \geq \delta/n - c_1M_1n^{-\varsigma_1}\Big) \\
&\leq P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}[q_\iota(X_{ij}, F_{X_j}^{-1}(\iota)) - \mathrm{E}\{q_\iota(X_{ij}, F_{X_j}^{-1}(\iota))\}]\Big| \geq \delta/(2n)\Big) \\
&\leq 2\exp(-\frac{\delta^2}{8(c_1M_1n^{1-\varsigma_1}+\delta/3)}), \qquad \text{(S1.6)}
\end{aligned}$$

where the third inequality is due to $\delta/(2n) \geq c_1M_1n^{-\varsigma_1}$. Similarly, by condition (C1), we can show that

$$\max_{1\leq j\leq p_n} P(|II_{nj1}^{(2)}| \geq \delta/n) \leq 2\exp(-\frac{\delta^2}{8(c_2M_2n^{1-\varsigma_2}+\delta/3)}), \qquad \text{(S1.7)}$$

for any $\varsigma_2 \in (0, 1/2)$ and for some positive constants c_2 and M_2. Therefore, combining

(S1.4)-(S1.7), we obtain that

$$
\begin{aligned}
\max_{1\le j\le p_n} P(|II_{nj1}+II_{nj2}|\ge 3\delta/n) &\le \max_{1\le j\le p_n} P(|II_{nj2}|>\delta/n)\\
&\quad +\max_{1\le j\le p_n} P(|II^{(1)}_{nj1}|\ge \delta/n)+\max_{1\le j\le p_n} P(|II^{(2)}_{nj1}|\ge \delta/n)\\
&\le 2\exp(-\frac{\delta^2}{2(n+2\delta/3)})+2\exp(-\frac{\delta^2}{8(c_1M_1n^{1-\varsigma_1}+\delta/3)})\\
&\quad +2\exp(-\frac{\delta^2}{8(c_2M_2n^{1-\varsigma_2}+\delta/3)})\\
&\le 6\exp(-\frac{\delta^2}{c_3n+c_4\delta}) \qquad \text{(S1.8)}
\end{aligned}
$$

for some positive constants $c_3\ge\max(2,8c_1M_1,8c_2M_2)$ and $c_4\ge 8/3$, provided that $\delta/(2n)\ge\max(c_1M_1n^{-\varsigma_1},c_2M_2n^{-\varsigma_2})$.

Using the basic inequality that $||x|-|y||\le|x-y|$ and the union bound of probability, it follows that

$$
\begin{aligned}
P(\max_{1\le j\le p_n}|\widehat{u}_j-u_j|\ge Cn^{-\kappa}) &\le p_n\max_{1\le j\le p_n}P(|\widehat{u}_j-u_j|\ge Cn^{-\kappa})\\
&\le p_n\max_{1\le j\le p_n}P(|II_{nj1}+II_{nj2}|\ge\sqrt{\tau(1-\tau)\iota(1-\iota)}Cn^{-\kappa}).
\end{aligned}
$$

Hence, this in conjunction with the result (S1.8) proves the desired result by letting $\delta=\sqrt{\tau(1-\tau)\iota(1-\iota)}Cn^{1-\kappa}/3$.

(ii) Prove the second assertion. By the choice of ν_n and using condition (C2), we have

$$
\begin{aligned}
&P(\mathcal{M}_*\subset\widehat{\mathcal{M}_a})\ge P\Big(\min_{j\in\mathcal{M}_*}\widehat{u}_j>\nu_n\Big)\ge P\Big(\min_{j\in\mathcal{M}_*}(\widehat{u}_j-u_j)>\nu_n-\min_{j\in\mathcal{M}_*}u_j\Big)\\
&\ge P\Big(\min_{j\in\mathcal{M}_*}u_j-\max_{j\in\mathcal{M}_*}\big|\widehat{u}_j-u_j\big|>\nu_n\Big)\ge 1-P\Big(\max_{j\in\mathcal{M}_*}\big|\widehat{u}_j-u_j\big|\ge\nu_n\Big)\\
&\ge 1-6s_n\exp(-\tilde{c}_1n^{1-2\kappa}).
\end{aligned}
$$

COX SCREENING

Thus, this completes the proof. □

Proof of Proposition 2.4 Observe that the expected false discovery rate can be rewritten as

$$\mathrm{E}\Big\{\frac{|\widehat{\mathcal{M}}_{a,\delta}\cap(\mathcal{M}_a^*)^c|}{|(\mathcal{M}_a^*)^c|}\Big\}=\frac{1}{p_n-|\mathcal{M}_a^*|}\sum_{j\in(\mathcal{M}_a^*)^c}P\big(\sqrt{n}[\widehat{\Omega}_1(\tau,\iota;\tau,\iota)]^{-1/2}|\widehat{\varrho}_{Y,X_j}(\tau,\iota)|\geq\delta\big).$$

By Theorem 2.1 and the Berry-Esséen result, it follows that there exists some constant $c_a>0$ such that

$$\sup_z\big|P\big(\sqrt{n}[\widehat{\Omega}_1(\tau,\iota;\tau,\iota)]^{-1/2}\widehat{\varrho}_{Y,X_j}(\tau,\iota)\leq z\big)-\Phi(z)\big|\leq c_a n^{-1/2}.$$

Combining the above results, we obtain

$$\mathrm{E}\Big\{\frac{|\widehat{\mathcal{M}}_{a,\delta}\cap(\mathcal{M}_a^*)^c|}{|(\mathcal{M}_a^*)^c|}\Big\}\leq\frac{1}{p_n-|\mathcal{M}_a^*|}\sum_{j\in(\mathcal{M}_a^*)^c}(2(1-\Phi(\delta))+c_a/\sqrt{n}).$$

Plugging $\delta=\Phi^{-1}(1-\bar{d}_n/(2p_n))$ into the above inequality yields the result. □

Proof of Theorem 3.1 We first observe that $\mathrm{E}\{\psi_\tau(Y-\mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z}\}=0$ and $\mathrm{E}\{\psi_\tau(X-\mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\}=0$, where $\mathbf{Z}^T\boldsymbol{\alpha}^0=\mathbf{Z}^T\boldsymbol{\alpha}^0$ and $\mathbf{Z}^T\boldsymbol{\theta}^0=\mathbf{Z}^T\boldsymbol{\theta}^0$. Using the arguments in Li et al. (2015) and Koenker (2005), we can obtain

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)=\big\{\mathrm{E}[f_{Y_i|\mathbf{Z}_i}(\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\mathbf{Z}_i\mathbf{Z}_i^T]\big\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\mathbf{Z}_i+o_p(1),\tag{S1.9}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)=\big\{\mathrm{E}[f_{X_i|\mathbf{Z}_i}(\mathbf{Z}_i^T\boldsymbol{\theta}^0)\mathbf{Z}_i\mathbf{Z}_i^T]\big\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\iota(X_i-\mathbf{Z}_i^T\boldsymbol{\theta}^0)\mathbf{Z}_i+o_p(1).\tag{S1.10}$$

Let

$$U_n(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}})=\frac{1}{n}\sum_{i=1}^n[I(Y_i\leq\mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}},X_i\leq\mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})-\mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0,\mathbf{Z}^T\boldsymbol{\theta}^0)\}].$$

Then, it follows that

$$\sqrt{n}\{\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota)-\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)\}=\frac{\sqrt{n}U_n(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}})}{\sqrt{\tau(1-\tau)\iota(1-\iota)}},$$

where

$$\begin{aligned}\sqrt{n}U_n(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[I(Y_i\le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}},X_i\le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})-\mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\widehat{\boldsymbol{\alpha}},\mathbf{Z}^T\widehat{\boldsymbol{\theta}})\}\big]\\ &\quad+\big[\mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\widehat{\boldsymbol{\alpha}},\mathbf{Z}^T\widehat{\boldsymbol{\theta}})\}-\mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0,\mathbf{Z}^T\boldsymbol{\theta}^0)\}\big]\\ &\hat{=} K_{n1}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}})+K_{n2}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}})\ \text{(say)}.\end{aligned}\tag{S1.11}$$

Employing the result of empirical processes (Corollary 2.3.12, van der Vaart and Wellner (1996), p115), we can obtain

$$\begin{aligned}K_{n1}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[I(Y_i\le \mathbf{Z}_i^T\boldsymbol{\alpha}^0,X_i\le \mathbf{Z}_i^T\boldsymbol{\theta}^0)\\ &\qquad -\mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0,\mathbf{Z}^T\boldsymbol{\theta}^0)\}\big]+o_p(1).\end{aligned}\tag{S1.12}$$

By Taylor expansion and using (S1.9) and (S1.10), we have

$$\begin{aligned}K_{n2}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}}) &= \mathrm{E}\big\{F_{X|\mathbf{Z},Y=\mathbf{Z}^T\boldsymbol{\alpha}^0}(\mathbf{Z}^T\boldsymbol{\theta}^0)f_{Y|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z}\big\}^T(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)\\ &\quad+\mathrm{E}\big\{F_{Y|\mathbf{Z},X=\mathbf{Z}^T\boldsymbol{\theta}^0}(\mathbf{Z}^T\boldsymbol{\alpha}^0)f_{X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\big\}^T(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\Delta_{12}^T\Delta_{11}^{-1}\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\mathbf{Z}_i\\ &\qquad+\Delta_{21}^T\Delta_{22}^{-1}\psi_\iota(X_i-\mathbf{Z}_i^T\boldsymbol{\theta}^0)\mathbf{Z}_i\Big\}+o_p(1).\end{aligned}\tag{S1.13}$$

Finally, combining (S1.11)-(S1.13) and recalling the definitions of $\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota)$ and $\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)$, we have

$$
\begin{aligned}
&\sqrt{\tau(1-\tau)\iota(1-\iota)}\sqrt{n}\{\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota)-\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)\} \\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\big[I(Y_i\le \mathbf{Z}_i^T\boldsymbol{\alpha}^0, X_i\le \mathbf{Z}_i^T\boldsymbol{\theta}^0)-\mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0,\mathbf{Z}^T\boldsymbol{\theta}^0)\}\big] \\
&\qquad +\Delta_{12}^T\Delta_{11}^{-1}\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\mathbf{Z}_i+\Delta_{21}^T\Delta_{22}^{-1}\psi_\iota(X_i-\mathbf{Z}_i^T\boldsymbol{\theta}^0)\mathbf{Z}_i\Big\}+o_p(1), \quad \text{(S1.14)}
\end{aligned}
$$

which implies that $\sqrt{n}\{\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota)-\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)\}\stackrel{a}{=}\frac{1}{\sqrt{n}}\sum_{i=1}^n[\zeta(X_i,Y_i,\mathbf{Z}_i;\tau,\iota)-\mathrm{E}\zeta(X,Y,\mathbf{Z};\tau,\iota)]$. It can be checked that $\{\zeta(X_i,Y_i,\mathbf{Z}_i;\tau,\iota):(\tau,\iota)\in[a,b]^2\}$ is Donsker. Therefore, by Donsker Theorem, we can conclude the result. □

Proof of Theorem 3.2 This result can be proved using the same arguments in the proof of Theorem 3.1. We omit the details. □

To prove Theorem 3.3, we need the following two lemmas which can be verified using the same arguments in Ma et al. (2017) and the proofs are omitted.

Lemma S1.2. *Under conditions (D1)(i) and (D2), for any given constant $c_1^*>0$ there exist some constant c_2^* such that*

$$P(\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\|>c_1^*n^{-\kappa})\le 3\exp(-c_2^*q_n^{-1}n^{1-2\kappa}).$$

Lemma S1.3. *Under conditions (D1)(ii) and (D2), for every $1\le j\le p_n$ and for any given constant $c_3^*>0$ there exist some constant c_4^* such that*

$$P(\|\widehat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j^0\|>c_3^*n^{-\kappa})\le 3\exp(-c_4^*q_n^{-1}n^{1-2\kappa}).$$

Proof of Theorem 3.3 To prove the first part of the result, we need to only prove the following fact. Under conditions (D1) and (D2), for every $1\le j\le p_n$ and for any given constant $c_5^* > 0$, there exist some positive constants c_6^* such that

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\big[\psi_\tau(Y_i-\mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})\psi_\iota(X_{ij}-\mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_j)-\mathrm{E}\{\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\psi_\iota(X_{ij}-\mathbf{Z}_i^T\boldsymbol{\theta}_j^0)\}\big]\Big|\geq c_5^* n^{-\kappa}\Big)$$
$$\leq 12\exp(-c_6^* q_n^{-1} n^{1-2\kappa}). \tag{S1.15}$$

To this end, we first make a decompose as

$$\frac{1}{n}\sum_{i=1}^{n}\big[\psi_\tau(Y_i-\mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})\psi_\iota(X_{ij}-\mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_j)-\mathrm{E}\{\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\psi_\iota(X_{ij}-\mathbf{Z}_i^T\boldsymbol{\theta}_j^0)\}\big]$$
$$\hat{=} N_{nj1}+N_{nj2}+N_{nj3},$$

where

$$\begin{aligned}
N_{nj1} &= \frac{1}{n}\sum_{i=1}^{n}\big[\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\psi_\iota(X_{ij}-\mathbf{Z}_i^T\boldsymbol{\theta}_j^0)-\mathrm{E}\{\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\psi_\iota(X_{ij}-\mathbf{Z}_i^T\boldsymbol{\theta}_j^0)\}\big],\\
N_{nj2} &= \frac{1}{n}\sum_{i=1}^{n}\big[\psi_\tau(Y_i-\mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})-\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\big]\psi_\iota(X_{ij}-\mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_j),\\
N_{nj3} &= \frac{1}{n}\sum_{i=1}^{n}\big[\psi_\iota(X_{ij}-\mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_j)-\psi_\iota(X_{ij}-\mathbf{Z}_i^T\boldsymbol{\theta}_j^0)\big]\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0).
\end{aligned}$$

Next, we are going to find the probability bounds for N_{nj1}, N_{nj2} and N_{nj3}. For N_{nj1}, let $\Delta_{nij}=\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\psi_\iota(X_{ij}-\mathbf{Z}_i^T\boldsymbol{\theta}_j^0)$. Since $|\Delta_{nij}|\leq 1$ and thence $\mathrm{var}(\Delta_{nij})\leq 1$, we can obtain that for any $\delta>0$,

$$P(|N_{nj1}|\geq \delta/n)\leq 2\exp(-\frac{\delta^2}{2(n+\delta/3)}), \tag{S1.16}$$

by applying Bernstein's inequality (Lemma 2.2.9, van der Vaart and Wellner (1996)). For N_{nj2}, observe that $|N_{nj2}|\leq \frac{1}{n}\sum_{i=1}^{n}\big|\psi_\tau(Y_i-\mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})-\psi_\tau(Y_i-\mathbf{Z}_i^T\boldsymbol{\alpha}^0)\big|$. It follows that for any

$\delta > 0$,

$$\begin{aligned} P(|N_{nj2}| \geq \delta/n) &\leq P\Big\{\frac{1}{n}\sum_{i=1}^{n}\big|\psi_\tau(Y_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}) - \psi_\tau(Y_i - \mathbf{Z}_i^T\boldsymbol{\alpha}^0)\big| \geq \delta/n, \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\| \leq c_1^* n^{-\kappa}\Big\} \\ &\quad + P(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\| > c_1^* n^{-\kappa}). \qquad \text{(S1.17)} \end{aligned}$$

Let $\Pi_{ni} = \sup_{\|\mathbf{u}\|\leq 1}\big|\psi_\tau(Y_i - \mathbf{Z}_i^T(\boldsymbol{\alpha}^0 + c_1^* n^{-\kappa}\mathbf{u})) - \psi_\tau(Y_i - \mathbf{Z}_i^T\boldsymbol{\alpha}^0)\big|$. Under condition (D1)(i) and (D2)(ii), there exist a finite positive constant c_5 and a $\mathbf{u}^*$ with $\|\mathbf{u}^*\| \leq 1$ such that $\mathrm{E}(\Pi_{ni}) \leq \mathrm{E}\{|\int_{\mathbf{Z}_i^T\boldsymbol{\alpha}^0}^{\mathbf{Z}_i^T\boldsymbol{\alpha}^0 + c_1^* n^{-\kappa}\mathbf{Z}_i^T\mathbf{u}^*} f_{Y|\mathbf{Z}=\mathbf{z}}(y)\mathrm{d}y|\} \leq c_1^* c_5 n^{-\kappa}\mathrm{E}\{|\mathbf{Z}_i^T\mathbf{u}^*|\} \leq c_1^* c_5 c_{\max}^{1/2} n^{-\kappa} = c_6 n^{-\kappa}$, where $c_6 = c_1^* c_5 c_{\max}^{1/2}$ the last inequality uses Cauchy-Schwarz inequality. Analogously, we can obtain $\mathrm{E}(\Pi_{ni}^2) \leq c_7 n^{-\kappa}$ for some positive constant c_7. Notice that $|\Pi_{ni}| \leq 1$. Hence, applying Bernstein's inequality (Lemma 2.2.9, van der Vaart and Wellner (1996)) yields that for any $\epsilon > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\Pi_{ni} - \mathrm{E}(\Pi_{ni})\Big| \geq \epsilon\Big) \leq 2\exp\Big(-\frac{n^2\epsilon^2}{2(c_7 n^{1-\kappa} + n\epsilon/3)}\Big).$$

Letting $\delta \geq 2c_6 n^{1-\kappa}$, that is $\delta/n - c_6 n^{-\kappa} \geq \delta/(2n)$, and using the above inequality, we can obtain

$$\begin{aligned} P\Big(\frac{1}{n}\sum_{i=1}^{n}\Pi_{ni} \geq \delta/n\Big) &\leq P\Big(\frac{1}{n}\sum_{i=1}^{n}\Pi_{ni} - \mathrm{E}(\Pi_{ni}) \geq \delta/n - c_6 n^{-\kappa}\Big) \\ &\leq P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\Pi_{ni} - \mathrm{E}(\Pi_{ni})\Big| \geq \delta/(2n)\Big) \\ &\leq 2\exp\Big(-\frac{\delta^2}{8(c_7 n^{1-\kappa} + \delta/6)}\Big). \qquad \text{(S1.18)} \end{aligned}$$

Because the first probability on the right hand side of (S1.17) is bounded by (S1.18). So, by Lemma S1.2 and the results (S1.17) and (S1.18), we have

$$P(|N_{nj2}| \geq \delta/n) \;\;\leq\;\; 2\exp(-\frac{\delta^2}{8(c_7 n^{1-\kappa}+\delta/6)}) + 3\exp(-c_2^* q_n^{-1} n^{1-2\kappa}) \qquad \text{(S1.19)}$$

provided that $\delta \geq 2c_6 n^{1-\kappa}$. Similarly, using Lemma S1.3, and under conditions (D1)(ii) and D2(ii), we have

$$P(|N_{nj3}| \geq \delta/n) \;\;\leq\;\; 2\exp(-\frac{\delta^2}{8(c_8 n^{1-\kappa}+\delta/6)}) + 3\exp(-c_4^* q_n^{-1} n^{1-2\kappa}) \qquad \text{(S1.20)}$$

for some positive constant c_8, provided that $\delta \geq 2c_9 n^{1-\kappa}$ for some constant $c_9 > 0$. Let $c_{10} = \max(c_7, c_8)$ and $c_{11} = \max(c_2^*, c_4^*)$. Combining the results (S1.16), (S1.19) and (S1.20) gives

$$\begin{aligned}&\max_{1\leq j\leq p_n} P(|N_{nj1}+N_{nj2}+N_{nj3}| \geq 3\delta/n)\\ &\leq 2\exp(-\frac{\delta^2}{2(n+\delta/3)}) + 4\exp(-\frac{\delta^2}{8(c_{10}n^{1-\kappa}+\delta/6)}) + 6\exp(-c_{11}q_n^{-1}n^{1-2\kappa}),\end{aligned} \qquad \text{(S1.21)}$$

provided that $\delta \geq 2\max(c_6, c_9)n^{1-\kappa}$. The result in (S1.15) follows by letting $\delta = 3\max(c_6, c_9)n^{1-\kappa}$ in (S1.21). In addition, applying (S1.15) and using similar arguments in the proof of Theorem 2.3, we can conclude the first part of Theorem 3.3.

For the second part of the result, we can verify it by following the proof of Theorem 2.3. □

Proof of Proposition 3.4 The proof directly follows the proof of Proposition 2.4. The details are omitted. □

Appendix B: A Compared Screening Procedure T-SIS

We first describe the the testing-based screening procedure (T-SIS) mentioned in the simulation section in the main manuscript, which is suggested by one referee.

- First, consider p_n marginal hypothesis testing problems, where for each $j(1 \le j \le p_n)$, we test the null hypothesis $H_0 : \varrho_{Y,X_j}(\tau, \iota) = 0$ versus $H_1 : \varrho_{Y,X_j}(\tau, \iota) \neq 0$.
- Second, propose a self-studentised testing statistic for each hypothesis problem as $T_j \hat{=} \sqrt{n}\widehat{\Omega}_1^{-1/2}\widehat{\varrho}_{Y,X_j}(\tau, \iota)$, which is asymptotically distributed as $N(0, 1)$ under H_0, according to Theorem 2.1. Given the sample data, we can calculate the self-studentised testing statistic as t_j and its associated p-value as $2[1 - \Phi(|t_j|)]$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$.
- Third, rank these p-values in an increasing order and select the first d_n variables as an active set.

Appendix C: Additional Numerical Results

(1) Inference Performance for CC and CPC

We consider two simulation examples with fixed dimension p in this subsection, and illustrate the practical performance of estimated CC and CPC, respectively. We consider sample size $n = 200$ and 400 and set the number of repetitions to be $N = 5000$ in Examples 1 and 2.

Example S1. We generate the response from two models

- (a1) $Y = \exp(2X_1) + \exp((2 + c_0)X_2)$, where (X_1, X_2) is from a standard bivariate normal distribution with $corr(X_1, X_2) = \rho$ and

- (a2) $Y = 2X_{01} + (2 + c_0)X_{02} + \varepsilon$,

where in model (a2), covariates X_1 and X_2 are both generated from a mixture distribution of a normal distribution with probability 0.9 and a cauchy distribution with probability 0.1, that is, $X_1 = 0.9X_{01} + 0.1\epsilon_1$ and $X_2 = 0.9X_{02} + 0.1\epsilon_2$ with (X_{01}, X_{02}) following a standard bivariate normal distribution with $corr(X_{01}, X_{02}) = \rho$, $\epsilon_1 \sim \frac{1}{5}Cauchy(0, 1)$ and $\epsilon_2 \sim \frac{1}{5}Cauchy(0, 1)$ are independent and the model error ε is generated from $N(0, 1)$. Our interest of this example is to test $H_0 : \varrho_{Y,X_1}(\tau, \iota) = \varrho_{Y,X_2}(\tau, \iota)$ at the significance level $\alpha_0 = 0.05$ under various values of c_0. We consider different ρ's and set $c_0 = 0, 1, 2, 4$, where $c_0 = 0$ implies that H_0 holds true, whereas H_0 should be rejected with large probability for other values of c_0. We report the empirical size and power for each setup over 5000 runs in Table 1. Observing Table 1, we can see that our proposed CC testing procedure based on Theorem 2 performs satisfactorily across different quantile levels (τ, ι) since the values of empirical size are close to nominal level $\alpha_0 = 0.05$ and enlarging the sample size from 200 to 400 generally tends to improve the performance. Also we can see that when c_0 runs away from 0, the empirical power increases to 1, and when the correlation between X_1 and X_2 is low, the performance will be better.

Example S2. In this example, we generate conditional variables $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)^T$ from the 4-dimensional multivariate normal distribution $N(\mathbf{0}_4, \Sigma)$ with $\Sigma = (\rho^{|j-k|})_{1\le j,k\le 4}$.

We generate Y from the model

- $Y = 2X_1 + (2 + c_0)X_2 + \mathbf{Z}^T\mathbf{b} + \varepsilon$,

where $\mathbf{b} = (3, 4, 3, 4)^T$, $X_1 = \mathbf{Z}^T\mathbf{b} + \epsilon_1$ and $X_2 = \mathbf{Z}^T\mathbf{b} + \epsilon_2$, where $\epsilon_1 \sim \frac{1}{3}t(3)$ and $\epsilon_2 \sim \frac{1}{3}t(3)$ are independent. The other setups are the same as Example 1. In this example, our interest is to test $H_0 : \varrho_{Y,X_1|\mathbf{Z}}(\tau, \iota) = \varrho_{Y,X_2|\mathbf{Z}}(\tau, \iota)$ at the significance level $\alpha_0 = 0.05$. Here, we consider the performance of CPC with $(\tau, \iota) = (0.5, 0.5)$, and the corresponding empirical size and power are reported in Table 2. A similar conclusion can be drawn as in Example S1. These numerical results empirically demonstrate that our theoretical result in Theorem 5 is valid.

(2) Empirical performance of CC and CPC estimates

In this subsection, we conduct some additional simulations, part of which can be viewed as a response to the questions raised by the reviewers, namely, how well the plug-in variance estimates of CC and CPC perform? To answer this question, we conduct two simulation examples to evaluate the performance of the proposed plug-in estimates of variance of CC estimator and CPC estimator, respectively.

Example S3 Assume that X_1^* and X_2^* are independent and standard normally distributed. We simulate covariates X_1 and X_2 from two mixture distributions: $X_1 = 0.9X_1^* + 0.1\epsilon_1$ and $X_2 = 0.9X_2^* + 0.1\epsilon_2$, respectively, where ϵ_1 and ϵ_2 are independent and have the same

Table 1: Empirical size and power for testing $H_0 : \varrho_{Y,X_1}(\tau,\iota) = \varrho_{Y,X_2}(\tau,\iota)$ using the CC across $(\tau,\iota) = (0.25, 0.25)$, $(0.5, 0.5)$ and $(0.75, 0.75)$ for Example S1.

			Model (a1)				Model (a2)			
ρ	Method(τ,ι)	n	$c_0 = 0$	1	2	4	$c_0 = 0$	1	2	4
0	$\mathrm{CC}_{(0.25,0.25)}$	200	0.056	0.07	0.07	0.08	0.051	0.48	0.90	1.00
		400	0.051	0.07	0.09	0.08	0.055	0.78	1.00	1.00
	$\mathrm{CC}_{(0.5,0.5)}$	200	0.059	0.33	0.69	0.94	0.057	0.56	0.95	1.00
		400	0.054	0.57	0.93	1.00	0.048	0.85	1.00	1.00
	$\mathrm{CC}_{(0.75,0.75)}$	200	0.057	0.77	0.99	1.00	0.053	0.49	0.90	1.00
		400	0.055	0.96	1.00	1.00	0.050	0.78	1.00	1.00
0.5	$\mathrm{CC}_{(0.25,0.25)}$	200	0.055	0.06	0.08	0.14	0.048	0.27	0.62	0.94
		400	0.056	0.06	0.08	0.18	0.052	0.45	0.90	1.00
	$\mathrm{CC}_{(0.5,0.5)}$	200	0.057	0.26	0.56	0.85	0.050	0.29	0.68	0.96
		400	0.053	0.44	0.83	0.98	0.049	0.51	0.93	1.00
	$\mathrm{CC}_{(0.75,0.75)}$	200	0.054	0.71	0.99	1.00	0.052	0.26	0.63	0.93
		400	0.052	0.95	1.00	1.00	0.043	0.47	0.89	1.00
0.9	$\mathrm{CC}_{(0.25,0.25)}$	200	0.055	0.08	0.17	0.47	0.048	0.10	0.19	0.43
		400	0.052	0.09	0.29	0.75	0.052	0.14	0.35	0.73
	$\mathrm{CC}_{(0.5,0.5)}$	200	0.061	0.17	0.36	0.63	0.048	0.09	0.20	0.45
		400	0.050	0.27	0.59	0.87	0.048	0.15	0.36	0.75
	$\mathrm{CC}_{(0.75,0.75)}$	200	0.045	0.52	0.91	0.99	0.053	0.09	0.20	0.43
		400	0.047	0.85	1.00	1.00	0.049	0.14	0.36	0.72

Table 2: Empirical size and power for testing $H_0 : \varrho_{Y,X_1|\mathbf{Z}}(\tau,\iota) = \varrho_{Y,X_2|\mathbf{Z}}(\tau,\iota)$ using CPC with $(\tau,\iota) = (0.5, 0.5)$ in Example S2.

		$\varepsilon \sim N(0,1)$				$\varepsilon \sim \frac{1}{3}Cauchy(0,1)$			
ρ	n	$c_0 = 0$	1	2	4	$c_0 = 0$	1	2	4
0	200	0.074	0.40	0.84	1.00	0.084	0.45	0.87	1.00
	400	0.067	0.65	0.98	1.00	0.073	0.72	0.99	1.00
0.5	200	0.073	0.40	0.83	1.00	0.081	0.46	0.86	1.00
	400	0.066	0.65	0.98	1.00	0.068	0.72	0.99	1.00
0.9	200	0.077	0.40	0.84	1.00	0.080	0.46	0.86	1.00
	400	0.069	0.65	0.98	1.00	0.070	0.72	0.99	1.00

distribution as $\frac{1}{5}Cauchy(0,1)$. We simulate the response variable Y from the model:

$$Y = 2\exp(c_0)X_1^* + 2X_2^* + \varepsilon,$$

where the random error ε is distributed as one of the following two distributions: (Normal error) $N(0,1)$ and (Cauchy error) $\frac{1}{3}Cauchy(0,1)$. In this example, our interest is to consider the finite-sample performance of CC estimator $\widehat{\rho}_{\tau,\iota}(Y, X_1)$ by computing the mean of CC estimator, empirical standard error and the mean of estimated standard deviation over 5000 runs. The corresponding results are reported in Tables 3 and 4.

From Tables 3 and 4, we can first observe that the values of estimated standard deviation are very close to empirical standard errors, which indicates that our proposed plug-in estimator of asymptotic variance of CC estimator works reasonably well as expected. Furthermore, when the sample size increases, the SD and SE gradually decrease to zero and, meanwhile, the estimation of CC keeps stable. For large CC (corresponding to $c_0 = 2$), the SD and SE tend to be smaller than those for small CC (corresponding to $c_0 = 0$). The

CC SCREENING

Table 3: Empirical performance of CC estimator for Example S3, where Ave., SE and SD represent the average of CC estimator, empirical standard error and the mean of estimated standard deviation, respectively, over 5000 runs.

	(τ, ι)		Normal error			Cauchy error		
			Ave.	SE	SD	Ave.	SE	SD
$c_0 = 0$	(0.25, 0.25)	$n = 30$	0.417	0.200	0.189	0.417	0.170	0.168
		$n = 50$	0.410	0.154	0.148	0.415	0.130	0.130
		$n = 80$	0.399	0.116	0.119	0.419	0.103	0.103
		$n = 100$	0.399	0.105	0.106	0.416	0.091	0.092
		$n = 200$	0.405	0.073	0.074	0.418	0.064	0.065
		$n = 400$	0.404	0.052	0.052	0.418	0.045	0.046
	(0.5, 0.5)	$n = 30$	0.417	0.170	0.168	0.402	0.171	0.170
		$n = 50$	0.415	0.130	0.130	0.401	0.133	0.131
		$n = 80$	0.419	0.103	0.103	0.402	0.102	0.103
		$n = 100$	0.416	0.091	0.092	0.406	0.093	0.092
		$n = 200$	0.418	0.064	0.065	0.406	0.065	0.065
		$n = 400$	0.418	0.045	0.046	0.405	0.045	0.046
	(0.75, 0.75)	$n = 30$	0.416	0.195	0.197	0.393	0.198	0.197
		$n = 50$	0.412	0.151	0.151	0.390	0.152	0.152
		$n = 80$	0.400	0.118	0.119	0.382	0.118	0.119
		$n = 100$	0.402	0.101	0.106	0.380	0.106	0.106
		$n = 200$	0.402	0.073	0.074	0.384	0.074	0.075
		$n = 400$	0.404	0.052	0.052	0.385	0.053	0.053

differences of SD and SE between normal random error and cauchy random error are minor.

Example S4: Suppose that we generate the response variable from

$$Y = 3\exp(c_0)X + 1.5\mathbf{Z}^T\mathbf{b} + \varepsilon,$$

where $X = \mathbf{Z}^T\mathbf{b} + \epsilon$ with $\epsilon \sim \frac{1}{5}t(4)$, $\mathbf{b} = (2, 2)^T$, and $\mathbf{Z} = (Z_1, Z_2)^T$, in which Z_1 and Z_2 are independent and standard normally distributed. We consider random error ε to

Table 4: Empirical performance of CC estimator for Example S3, where Ave., SE and SD represent the average of CC estimator, empirical standard error and the mean of estimated standard deviation, respectively, over 5000 runs.

			Normal error			Cauchy error		
	(τ, ι)		Ave.	SE	SD	Ave.	SE	SD
$c_0 = 2$	(0.25, 0.25)	$n = 30$	0.826	0.142	0.137	0.800	0.148	0.141
		$n = 50$	0.820	0.103	0.105	0.800	0.110	0.108
		$n = 80$	0.810	0.080	0.090	0.785	0.085	0.093
		$n = 100$	0.811	0.070	0.080	0.786	0.075	0.083
		$n = 200$	0.816	0.049	0.054	0.792	0.051	0.056
		$n = 400$	0.819	0.035	0.037	0.795	0.037	0.039
	(0.5, 0.5)	$n = 30$	0.793	0.116	0.118	0.775	0.119	0.122
		$n = 50$	0.803	0.089	0.089	0.781	0.091	0.093
		$n = 80$	0.808	0.066	0.069	0.789	0.070	0.072
		$n = 100$	0.807	0.061	0.062	0.789	0.062	0.064
		$n = 200$	0.810	0.041	0.043	0.791	0.043	0.044
		$n = 400$	0.812	0.029	0.030	0.793	0.031	0.031
	(0.75, 0.75)	$n = 30$	0.828	0.145	0.162	0.804	0.148	0.166
		$n = 50$	0.825	0.103	0.119	0.796	0.111	0.123
		$n = 80$	0.809	0.079	0.091	0.784	0.084	0.094
		$n = 100$	0.811	0.069	0.080	0.786	0.076	0.083
		$n = 200$	0.817	0.049	0.054	0.792	0.053	0.056
		$n = 400$	0.818	0.034	0.037	0.795	0.038	0.039

be distributed from two cases: (Normal error) $N(0,1)$ and (Cauchy error) $\frac{1}{3}Cauchy(0,1)$. In such an example, we aim to investigate the empirical performance of CPC estimator $\widehat{\rho}_{\tau,\iota}(Y, X_1|\mathbf{Z})$ similarly to that in Example S3. The simulation results are given in Tables 5 and 6. From these two tables, a similar conclusion can be made.

Table 5: Empirical performance of CPC estimator for Example S4, where Ave., SE and SD represent the average of CPC estimator, empirical standard error and the mean of estimated standard deviation, respectively, over 5000 runs.

			Normal error			Cauchy error		
	(τ, ι)		Ave.	SE	SD	Ave.	SE	SD
$c_0 = 0$	(0.25, 0.25)	$n = 30$	0.370	0.203	0.177	0.441	0.211	0.181
		$n = 50$	0.377	0.155	0.143	0.455	0.155	0.143
		$n = 80$	0.379	0.123	0.117	0.465	0.120	0.116
		$n = 100$	0.380	0.109	0.105	0.473	0.107	0.104
		$n = 200$	0.379	0.075	0.075	0.477	0.075	0.074
		$n = 400$	0.380	0.053	0.053	0.483	0.053	0.052
	(0.5, 0.5)	$n = 30$	0.365	0.176	0.152	0.476	0.166	0.148
		$n = 50$	0.377	0.134	0.124	0.499	0.126	0.118
		$n = 80$	0.380	0.104	0.101	0.513	0.099	0.095
		$n = 100$	0.383	0.093	0.091	0.520	0.084	0.085
		$n = 200$	0.386	0.065	0.065	0.530	0.061	0.060
		$n = 400$	0.388	0.046	0.046	0.535	0.042	0.043
	(0.75, 0.75)	$n = 30$	0.326	0.196	0.190	0.380	0.197	0.194
		$n = 50$	0.351	0.151	0.134	0.422	0.149	0.138
		$n = 80$	0.365	0.119	0.108	0.445	0.118	0.110
		$n = 100$	0.366	0.107	0.099	0.456	0.105	0.100
		$n = 200$	0.374	0.076	0.073	0.471	0.074	0.072
		$n = 400$	0.379	0.052	0.053	0.478	0.052	0.051

(3) Performance of CC-SIS compared with some popular screening methods

The following example is used to evaluate the finite-sample performance of CC-SIS.

Example S5. Let $\mathbf{X}^* = (X_1^*, \ldots, X_{p_n}^*)^T$ be a latent random vector having the p_n-dimensional normal distribution $N(\mathbf{0}_{p_n}, \Sigma)$ with $\Sigma = (\rho^{|k-l|})_{1\le k,l\le p_n}$, where we set the correlation $\rho = 0.4$ and 0.8. We write $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_{p_n})^T$ with each component ϵ_j be-

Table 6: Empirical performance of CPC estimator for Example S4, where Ave., SE and SD represent the average of CPC estimator, empirical standard error and the mean of estimated standard deviation, respectively, over 5000 runs.

			Normal error			Cauchy error		
	(τ, ι)		Ave.	SE	SD	Ave.	SE	SD
$c_0 = 2$	(0.25, 0.25)	$n = 30$	0.822	0.147	0.149	0.779	0.163	0.156
		$n = 50$	0.842	0.099	0.107	0.803	0.115	0.111
		$n = 80$	0.854	0.073	0.080	0.819	0.083	0.086
		$n = 100$	0.857	0.064	0.070	0.826	0.073	0.076
		$n = 200$	0.865	0.042	0.049	0.839	0.048	0.052
		$n = 400$	0.869	0.030	0.034	0.847	0.033	0.036
	(0.5, 0.5)	$n = 30$	0.801	0.106	0.122	0.775	0.118	0.125
		$n = 50$	0.831	0.077	0.083	0.808	0.085	0.087
		$n = 80$	0.844	0.059	0.062	0.830	0.065	0.065
		$n = 100$	0.849	0.052	0.055	0.837	0.056	0.057
		$n = 200$	0.860	0.036	0.037	0.850	0.038	0.039
		$n = 400$	0.864	0.025	0.026	0.860	0.026	0.027
	(0.75, 0.75)	$n = 30$	0.738	0.131	0.188	0.697	0.153	0.191
		$n = 50$	0.794	0.093	0.123	0.754	0.109	0.126
		$n = 80$	0.823	0.072	0.087	0.793	0.080	0.091
		$n = 100$	0.834	0.062	0.075	0.802	0.070	0.078
		$n = 200$	0.854	0.042	0.050	0.827	0.048	0.053
		$n = 400$	0.863	0.029	0.034	0.842	0.032	0.036

ing independent of other components and having the standard Cauchy distribution, i.e., $\epsilon_j \sim Cauchy(0,1)$. We generate covariates $\mathbf{X}$ from a mixture distribution: $\mathbf{X} = 0.8\mathbf{X}^* + 0.2\boldsymbol{\epsilon}$, and simulate the response data from the following three models:

- (b1) $Y = 3X_1^* + 3X_2^* + 2X_3^* + 2X_4^* + 2X_5^* + \varepsilon$,

- (b2) $Y = 5X_1^* I(X_1^* < 0) + 5X_2^* I(X_2^* > 0) + 5\sin(X_{10}^*) + \varepsilon$,

- (b3) $Y = \exp\{3\beta_1 \sin(X_1^*) + 2\beta_2 \exp(X_2^*) + 1.5\beta_3 I(X_3^* > 0) + 2\log(|X_4^*|)\} + \varepsilon$,

where ε is simulated from two scenarios: $\varepsilon \sim N(0,1)$ and $\varepsilon \sim Cauchy(0,1)$ and, in model (b3), we set $\beta_j = c_j(-1)^U(4\log n/\sqrt{n} + Z_0)$ for $j = 1, 2$ and 3, where $U \sim Bernoulli(0.4)$, $Z_0 \sim N(0, 0.5^2)$ and $(c_1, c_2, c_3) = (1, 0.5, 1)$. The resulting screening results in terms of MMS and $\mathcal{P}$ are presented in Tables 7 and 8.

Eyeballing Tables 7 and 8, we can make some key observations. Under models (b1) and (b2), our CC-SIS outperforms SIS, SIRS, DC-SIS and QC-SIS. In this case, both response and covariates are heavy-tailed and thus traditional linear correlation screening methods all fail to work. Our methods are also comparable to the nonparametric Kendall's τ which achieves high accuracy but is slower due to the numerical integration in its implementation. (b3) is a difficult case and very hard to screen accurately. Under this case, our CC-SIS still performs much better than other methods.

Example S6. We generate the response from the following two models

- (d1) $Y = \beta X_1 + \beta X_2 + \beta X_3 - 3\beta\sqrt{\rho}X_4 + \varepsilon$,
- (d2) $Y = \beta X_1^* + \beta X_2^* + \beta X_3^* - 3\beta\sqrt{\rho}X_4^* + \varepsilon$,

where $\beta = 4$ for both models and ε is simulated as $N(0,1)$ or $\frac{1}{3}Cauchy(0,1)$. In model (d1), we consider the observed covariates $\mathbf{X} \sim N(\mathbf{0}_{p_n}, \Sigma)$ with $\Sigma = (\sigma_{ij})_{1\le i,j\le p_n}$, implying that covariates $\mathbf{X}$ have no outlier and the response Y is fully dependent on observed $\mathbf{X}$ up to a random noise. Under this setting, it is desired to expect that the QPC-SIS performs better than our CPC-SIS since $\mathbf{X}$ is normal. In model (d2), we generate covariates $\mathbf{X}$ from a mixture distribution $0.9\mathbf{X}^* + 0.1\boldsymbol{\epsilon}$, where $\mathbf{X}^* \sim N(\mathbf{0}_{p_n}, \Sigma)$ with $\Sigma = (\sigma_{ij})_{1\leq i,j\leq p_n}$, and each element of $\boldsymbol{\epsilon}$ is independent and distributed as $\frac{1}{5}Cauchy(0,1)$. In this model, the covariates are contaminated with outliers, while the heterogeneity of response Y stems merely from the random error ε. In addition, in model (d1), we let $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho, j \neq i$ except that $\sigma_{4j} = \sigma_{j4} = \sqrt{\rho}$. In model (d2), we set $\sigma_{ij} = 0$ if $i > 4$ or $j > 4$ and the rest are the same as model (d1). For both models, at the population level, the covariate X_4 is marginally uncorrelated with Y. We consider two cases of $\rho = 0.95$ and 0.5 for simulation comparison.

Tables 9 and 10 report the screening results regarding the rank R_j and MMS. From the tables, we can see that all the marginal screening approaches fail to pick out the covariate X_4 with very large values of the rank R_4. QPC-SIS and our CPC-SIS work much better than those marginal methods. Particularly, under model (d1), our CPC-SIS has a very competitive performance to QPC-SIS. Under model (d2), when the covariates are highly correlated ($\rho = 0.95$), our proposal CPC-SIS$_{(0.5,0.5)}$ has the best performance.

(4) A further comparison of CC-SIS to other methods

Secondly, we further compare our proposed CC-SIS with two variable screening procedures: mCC-SIS and T-SIS, which are mentioned in the main manuscript by two additional simulation examples.

Example S7 Let $\mathbf{X}^* = (X_1^*, \ldots, X_{p_n}^*)^T$ be a latent random vector having the p_n-dimensional normal distribution $N(\mathbf{0}_{p_n}, \Sigma)$ with $\Sigma = (\rho^{|k-l|})_{1\leq k,l\leq p_n}$, where we set the correlation coefficient $\rho = 0.5$ for whole simulation. We write $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_{p_n})^T$ with each component ϵ_j

being independent of other components and having the standard Cauchy distribution, i.e., $\epsilon_j \sim Cauchy(0,1)$. We generate covariates $\mathbf{X}$ from a mixture distribution: $\mathbf{X} = 0.8\mathbf{X}^* + 0.2\boldsymbol{\epsilon}$, and simulate the response data from the following model:

- $Y = 3X_1^* + 3X_2^* + 2X_3^* + 2X_4^* + 2X_5^* + \varepsilon$,

where ε is simulated from two scenarios: $\varepsilon \sim N(0,1)$ and $\varepsilon \sim Cauchy(0,1)$. The simulation results are reported in Table 11 for $p_n = 1000$. From Table 11, we can observe that when sample size is very small ($n = 30$), our proposed CC-SIS performs best at the median level of (τ, ι), while the mCC-SIS behaves best since it has the smallest model size and RSD when sample size is large.

Example S8. We generate the response data from the following model: $Y = 5X_1^* I(X_1^* < 0) + 5X_2^* I(X_2^* > 0) + 5\sin(X_{10}^*) + \varepsilon$, where ε is simulated from two scenarios: $\varepsilon \sim N(0,1)$ and $\varepsilon \sim Cauchy(0,1)$ and, the covariates $\mathbf{X}$ are simulated in the same manner as in Example S7 above. The resulting screening results in terms of MMS and R_j are presented in Table 12. From Table 12, we can observe that our proposed CC-SIS performs best at the median level of (τ, ι) for small sample size and the mCC-SIS dominates other methods for large sample size.

(5) Overlaps for breast cancer data

Table 13 gives the overlaps of selected genes probes using several approaches in the breast cancer data analysis.

Bibliography

Koenker R. (2005). *Quantile Regression*, New York: Cambridge University Press.

Li G., Li Y. and Tsai C.L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association*, **110**, 246-261.

Ma S., Li R. and Tsai C. (2017). Variable screening via quantile partial correlation. *Journal of the American Statistical Association*, **112**, 650-663.

Serfling R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

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

Xiaochao Xia (Corresponding author), College of Science, Huazhong Agricultural University, Wuhan, China.

E-mail: (xia_xiao_chao@126.com)

Jialiang Li, Department of Statistics and Applied Probability, National University of Singapore; Duke-NUS Graduate Medical School; Singapore Eye Research Institute.

E-mail: (stalj@nus.edu.sg)

Table 7: Simulation results for Example S5, where MMS stands for the median of the minimum model size, and robust standard deviations (RSD) are given in parentheses, $\mathcal{P}$ is the proportion of screened sets that cover all active predictors with screening parameter $d_n = \lfloor n/\log n \rfloor$.

	$\rho = 0.4$				$\rho = 0.8$			
	$\varepsilon \sim N(0,1)$		$\varepsilon \sim Cauchy(0,1)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim Cauchy(0,1)$	
Method(τ,ι)	MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$
Model (b1)								
SIS	655 (396)	0.03	790 (227)	0.00	522 (431)	0.12	608 (389)	0.03
SIRS	615 (176)	0.00	669 (158)	0.00	513 (149)	0.00	547 (138)	0.00
DC-SIS	634 (441)	0.09	725 (350)	0.04	460 (527)	0.20	615 (489)	0.13
Kendall-SIS	5 (0)	0.99	5 (1)	0.96	5 (0)	1.00	5 (0)	1.00
CC-SIS$_{(0.25,0.25)}$	9 (18)	0.80	20 (41)	0.70	5 (0)	1.00	5 (0)	1.00
CC-SIS$_{(0.5,0.5)}$	6 (7)	0.93	9 (14)	0.84	5 (0)	1.00	5 (0)	1.00
CC-SIS$_{(0.75,0.75)}$	9 (15)	0.85	14 (40)	0.70	5 (0)	1.00	5 (0)	1.00
QC-SIS$_{(0.25)}$	710 (315)	0.01	782 (271)	0.00	649 (444)	0.06	678 (406)	0.03
QC-SIS$_{(0.5)}$	716 (334)	0.01	790 (250)	0.02	665 (438)	0.07	693 (398)	0.06
QC-SIS$_{(0.75)}$	711 (297)	0.01	751 (314)	0.01	599 (379)	0.04	670 (408)	0.02
CQC-SIS	573 (156)	0.00	614 (149)	0.00	442 (118)	0.00	474 (104)	0.00
Model (b2)								
SIS	482 (408)	0.13	638 (376)	0.01	351 (414)	0.15	686 (331)	0.03
SIRS	551 (161)	0.00	589 (195)	0.00	523 (173)	0.00	543 (158)	0.00
DC-SIS	284 (585)	0.30	607 (493)	0.12	230 (569)	0.35	506 (514)	0.17
Kendall-SIS	3 (0)	1.00	3 (0)	1.00	4 (1)	1.00	4 (1)	1.00
CC-SIS$_{(0.25,0.25)}$	6 (20)	0.78	13 (45)	0.68	5 (2)	0.99	7 (4)	0.94
CC-SIS$_{(0.5,0.5)}$	3 (1)	0.98	4 (3)	0.95	4 (2)	1.00	5 (1)	1.00
CC-SIS$_{(0.75,0.75)}$	7 (15)	0.85	14 (43)	0.66	6 (3)	0.99	7 (3)	0.96
QC-SIS$_{(0.25)}$	615 (356)	0.06	588 (391)	0.05	411 (418)	0.07	598 (411)	0.05
QC-SIS$_{(0.5)}$	584 (460)	0.07	640 (373)	0.05	539 (474)	0.13	592 (425)	0.10
QC-SIS$_{(0.75)}$	588 (414)	0.03	579 (353)	0.02	486 (374)	0.09	604 (401)	0.06
CQC-SIS	512 (153)	0.00	562 (155)	0.00	441 (122)	0.00	481 (150)	0.00

Table 8: (Continued) Simulation results for Example S5, where MMS stands for the median of the minimum model size, and robust standard deviations (RSD) are given in parentheses, $\mathcal{P}$ is the proportion of screened sets that cover all active predictors with screening parameter $d_n = \lfloor n/\log n \rfloor$.

	$\rho = 0.4$				$\rho = 0.8$			
	$\varepsilon \sim N(0,1)$		$\varepsilon \sim Cauchy(0,1)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim Cauchy(0,1)$	
Method(τ, ι)	MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$
				Model (b3)				
SIS	737 (296)	0.00	768 (204)	0.00	524 (399)	0.01	619 (368)	0.01
SIRS	767 (215)	0.00	745 (182)	0.00	664 (240)	0.00	703 (203)	0.00
DC-SIS	790 (245)	0.00	789 (213)	0.01	722 (276)	0.02	719 (309)	0.01
Kendall-SIS	504 (467)	0.12	475 (442)	0.10	5 (192)	0.65	16 (377)	0.60
CC-SIS$_{(0.25,0.25)}$	475 (234)	0.02	679 (317)	0.02	252 (365)	0.21	279 (455)	0.14
CC-SIS$_{(0.5,0.5)}$	483 (357)	0.09	478 (337)	0.07	18 (341)	0.56	53 (357)	0.47
CC-SIS$_{(0.75,0.75)}$	426 (419)	0.17	281 (460)	0.19	12 (309)	0.59	113 (349)	0.47
QC-SIS$_{(0.25)}$	791 (199)	0.00	795 (191)	0.00	794 (222)	0.01	781 (225)	0.01
QC-SIS$_{(0.5)}$	785 (210)	0.00	810 (209)	0.01	750 (266)	0.02	796 (219)	0.04
QC-SIS$_{(0.75)}$	762 (257)	0.00	793 (240)	0.00	698 (408)	0.01	714 (311)	0.02
CQC-SIS	743 (216)	0.00	749 (177)	0.00	623 (210)	0.00	628 (260)	0.00

Table 9: Simulation results for Example S6 (d1), where R_j indicates the median of the rank of the relevant predictors and MMS stands for the median of the minimum model size and its robust standard deviations (RSD) are given in parenthesis.

		$\varepsilon \sim N(0,1)$					$\varepsilon \sim \frac{1}{3}Cauchy(0,1)$				
ρ	Method(τ, ι)	R_1	R_2	R_3	R_4	MMS (RSD)	R_1	R_2	R_3	R_4	MMS (RSD)
						Model (d1)					
0.5	SIS	2	2	2	450	450 (235)	3	3	3	411	476 (270)
	SIRS	3	3	3	469	501 (243)	3	3	4	432	466 (258)
	DC-SIS	2	2	2	497	497 (280)	2	2	2	464	464 (220)
	Kendall-SIS	2	2	2	454	454 (228)	2	2	2	467	468 (280)
	CC-SIS$_{(0.25,0.25)}$	2	2	2	367	372 (331)	2	3	2	382	390 (311)
	CC-SIS$_{(0.5,0.5)}$	2	2	2	415	425 (272)	2	2	2	396	408 (275)
	CC-SIS$_{(0.75,0.75)}$	2	2	2	371	387 (324)	2	3	2	378	388 (311)
	QC-SIS$_{(0.25)}$	2	2	2	478	493 (270)	2	2	2	473	477 (245)
	QC-SIS$_{(0.5)}$	2	2	2	437	437 (236)	2	2	2	461	463 (331)
	QC-SIS$_{(0.75)}$	2	2	2	437	438 (227)	2	2	2	408	421 (247)
	QPC-SIS$_{(0.25)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	QPC-SIS$_{(0.5)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	QPC-SIS$_{(0.75)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	CPC-SIS$_{(0.25,0.25)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	CPC-SIS$_{(0.5,0.5)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	CPC-SIS$_{(0.75,0.75)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
0.95	SIS	3	3	3	470	498 (306)	214	269	262	466	735 (353)
	SIRS	28	52	45	474	541 (331)	24	26	43	473	584 (376)
	DC-SIS	3	3	3	506	556 (220)	10	12	15	495	610 (240)
	Kendall-SIS	3	3	3	482	529 (296)	3	3	3	486	557 (325)
	CC-SIS$_{(0.25,0.25)}$	91	95	104	303	481 (276)	88	99	88	318	485 (292)
	CC-SIS$_{(0.5,0.5)}$	56	49	67	365	499 (383)	53	68	47	354	610 (377)
	CC-SIS$_{(0.75,0.75)}$	114	92	94	297	470 (301)	125	63	138	318	492 (289)
	QC-SIS$_{(0.25)}$	3	3	3	470	512 (387)	3	3	3	467	522 (375)
	QC-SIS$_{(0.5)}$	3	3	3	465	499 (346)	3	3	3	467	506 (382)
	QC-SIS$_{(0.75)}$	3	3	3	468	520 (379)	3	3	3	469	508 (329)
	QPC-SIS$_{(0.25)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	QPC-SIS$_{(0.5)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	QPC-SIS$_{(0.75)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	CPC-SIS$_{(0.25,0.25)}$	2	2	2	4	4 (1)	2	2	2	5	5 (1)
	CPC-SIS$_{(0.5,0.5)}$	2	2	2	4	4 (0)	2	2	2	4	4 (0)
	CPC-SIS$_{(0.75,0.75)}$	2	2	2	4	4 (1)	2	2	2	4	4 (1)

Table 10: Simulation results for Example S6 (d2), where R_j indicates the median of the rank of the relevant predictors, and MMS stands for the median of the minimum model size, and robust standard deviations (RSD) are given in parentheses.

		$\varepsilon \sim N(0,1)$					$\varepsilon \sim \frac{1}{3}Cauchy(0,1)$				
ρ	Method(τ,ι)	R_1	R_2	R_3	R_4	MMS (RSD)	R_1	R_2	R_3	R_4	MMS (RSD)
					Model (d2)						
0.5	SIS	2	2	2	530	566 (359)	10	12	14	535	667 (351)
	SIRS	99	93	100	453	485 (304)	107	110	112	478	507 (356)
	DC-SIS	2	2	2	558	629 (393)	2	2	2	554	638 (322)
	Kendall-SIS	2	2	2	514	514 (362)	2	2	2	578	578 (381)
	CC-SIS$_{(0.25,0.25)}$	3	3	3	437	442 (362)	4	4	4	453	456 (333)
	CC-SIS$_{(0.5,0.5)}$	3	3	3	339	348 (358)	2	3	3	459	463 (343)
	CC-SIS$_{(0.75,0.75)}$	3	3	3	440	446 (315)	3	4	5	449	451 (333)
	QC-SIS$_{(0.25)}$	3	3	3	499	594 (352)	3	3	2	531	618 (363)
	QC-SIS$_{(0.5)}$	2	2	2	514	583 (377)	2	3	2	501	533 (322)
	QC-SIS$_{(0.75)}$	3	3	3	532	588 (345)	3	3	3	460	537 (337)
	QPC-SIS$_{(0.25)}$	3	3	3	2	4 (1)	3	3	3	2	4 (1)
	QPC-SIS$_{(0.5)}$	2	3	3	2	4 (0)	3	3	3	2	4 (0)
	QPC-SIS$_{(0.75)}$	3	3	3	2	4 (1)	3	3	3	2	4 (3)
	CPC-SIS$_{(0.25,0.25)}$	2	2	2	4	4 (2)	2	2	2	4	5 (3)
	CPC-SIS$_{(0.5,0.5)}$	3	3	3	1	4 (0)	3	3	3	2	4 (0)
	CPC-SIS$_{(0.75,0.75)}$	2	2	2	4	4 (1)	2	2	3	4	5 (3)
0.95	SIS	182	169	200	557	712 (285)	518	434	435	496	770 (282)
	SIRS	355	388	375	507	729 (290)	450	401	403	507	720 (305)
	DC-SIS	238	191	236	540	762 (335)	347	329	365	507	825 (262)
	Kendall-SIS	192	154	181	534	675 (274)	224	222	196	482	734 (309)
	CC-SIS$_{(0.25,0.25)}$	266	281	268	443	695 (197)	254	260	255	438	688 (205)
	CC-SIS$_{(0.5,0.5)}$	292	224	303	473	672 (305)	318	305	213	462	666 (300)
	CC-SIS$_{(0.75,0.75)}$	262	265	265	441	694 (199)	268	423	284	431	693 (195)
	QC-SIS$_{(0.25)}$	313	295	338	595	791 (259)	393	394	357	480	755 (280)
	QC-SIS$_{(0.5)}$	274	304	308	529	795 (283)	291	282	290	480	753 (258)
	QC-SIS$_{(0.75)}$	259	341	279	529	746 (276)	362	374	376	459	806 (268)
	QPC-SIS$_{(0.25)}$	112	150	158	4	530 (465)	163	186	162	5	583 (475)
	QPC-SIS$_{(0.5)}$	106	147	103	3	518 (524)	148	102	96	3	514 (535)
	QPC-SIS$_{(0.75)}$	148	164	214	3	574 (437)	146	196	126	6	591 (483)
	CPC-SIS$_{(0.25,0.25)}$	14	11	11	3	270 (478)	13	16	34	6	311 (591)
	CPC-SIS$_{(0.5,0.5)}$	5	5	6	1	110 (404)	4	4	4	1	24 (371)
	CPC-SIS$_{(0.75,0.75)}$	11	13	15	3	225 (453)	13	39	14	5	299 (537)

Table 11: Simulation results for Example S7, where R_j indicates the median of the rank of the relevant predictors, and MMS stands for the median of the minimum model size, and robust standard deviations (RSD) are given in parentheses.

		$\varepsilon \sim N(0,1)$						$\varepsilon \sim \frac{1}{3}Cauchy(0,1)$					
	Method(τ,ι)	R_1	R_2	R_3	R_4	R_5	MMS (RSD)	R_1	R_2	R_3	R_4	R_5	MMS (RSD)
$n=30$	T-SIS$_{(0.25,0.25)}$	103	68	86	106	175	577 (207)	125	95	89	118	423	577 (254)
	T-SIS$_{(0.5,0.5)}$	55	24	34	69	219	442 (336)	90	36	48	121	231	464 (287)
	T-SIS$_{(0.75,0.75)}$	183	131	155	187	306	577 (152)	232	160	206	239	237	595 (242)
	CC-SIS$_{(0.25,0.25)}$	70	5	67	74	101	404 (185)	78	70	71	78	380	408 (193)
	CC-SIS$_{(0.5,0.5)}$	25	5	22	29	134	162 (240)	28	22	24	34	133	164 (245)
	CC-SIS$_{(0.75,0.75)}$	69	49	63	70	157	393 (178)	76	68	74	79	84	401 (193)
	mCC-SIS	44	12	30	52	127	245 (319)	107	21	57	92	161	343 (344)
$n=50$	T-SIS$_{(0.25,0.25)}$	90	16	19	89	131	355 (313)	90	18	78	102	195	416 (332)
	T-SIS$_{(0.5,0.5)}$	11	5	10	26	76	196 (223)	20	7	16	41	71	245 (290)
	T-SIS$_{(0.75,0.75)}$	74	57	67	92	109	432 (356)	100	68	74	114	134	468 (278)
	CC-SIS$_{(0.25,0.25)}$	20	3	5	21	32	152 (270)	21	3	18	25	125	160 (259)
	CC-SIS$_{(0.5,0.5)}$	4	3	6	11	28	93 (173)	8	4	6	20	29	103 (143)
	CC-SIS$_{(0.75,0.75)}$	5	3	4	22	28	151 (319)	22	4	5	26	31	165 (255)
	mCC-SIS	14	4	6	22	57	99 (107)	25	4	19	40	66	164 (209)
$n=100$	T-SIS$_{(0.25,0.25)}$	8	3	6	17	78	98 (98)	22	4	12	20	51	123 (181)
	T-SIS$_{(0.5,0.5)}$	4	2	2	4	13	27 (55)	4	2	3	5	27	48 (79)
	T-SIS$_{(0.75,0.75)}$	11	3	5	17	80	99 (132)	20	5	9	19	88	130 (203)
	CC-SIS$_{(0.25,0.25)}$	3	2	3	4	29	39 (81)	4	2	3	5	21	101 (190)
	CC-SIS$_{(0.5,0.5)}$	3	2	2	4	9	16 (26)	3	2	3	4	14	31 (46)
	CC-SIS$_{(0.75,0.75)}$	3	1	3	4	31	40 (86)	4	2	3	5	34	108 (191)
	mCC-SIS	3	1	3	4	12	16 (24)	4	2	3	5	18	26 (66)

Table 12: Simulation results for Example 6, where R_j indicates the median of the rank of the relevant predictors, and MMS stands for the median of the minimum model size, and robust standard deviations (RSD) are given in parentheses.

		$\varepsilon \sim N(0,1)$				$\varepsilon \sim \frac{1}{3}Cauchy(0,1)$			
	Method(τ,ι)	R_1	R_2	R_3	MMS (RSD)	R_1	R_2	R_3	MMS (RSD)
$n=30$	T-SIS$_{(0.25,0.25)}$	82	138	112	479 (361)	108	390	149	521 (241)
	T-SIS$_{(0.5,0.5)}$	80	93	56	303 (297)	107	89	108	406 (389)
	T-SIS$_{(0.75,0.75)}$	259	127	216	566 (292)	260	197	267	587 (337)
	CC-SIS$_{(0.25,0.25)}$	65	84	76	387 (246)	74	355	89	405 (196)
	CC-SIS$_{(0.5,0.5)}$	26	29	25	147 (303)	31	28	32	154 (253)
	CC-SIS$_{(0.75,0.75)}$	86	51	77	396 (391)	162	70	81	399 (203)
	mCC-SIS	33	37	96	151 (148)	91	92	152	294 (326)
$n=50$	T-SIS$_{(0.25,0.25)}$	20	125	105	305 (284)	28	207	111	418 (318)
	T-SIS$_{(0.5,0.5)}$	24	23	17	105 (166)	48	50	37	208 (259)
	T-SIS$_{(0.75,0.75)}$	111	56	114	425 (407)	358	78	123	459 (264)
	CC-SIS$_{(0.25,0.25)}$	3	31	25	144 (311)	4	126	28	161 (258)
	CC-SIS$_{(0.5,0.5)}$	14	8	7	56 (167)	23	24	21	96 (175)
	CC-SIS$_{(0.75,0.75)}$	25	2	27	144 (327)	138	7	28	166 (257)
	mCC-SIS	12	4	30	54 (66)	28	23	45	116 (170)
$n=100$	T-SIS$_{(0.25,0.25)}$	3	32	18	96 (145)	6	83	29	130 (223)
	T-SIS$_{(0.5,0.5)}$	4	3	2	14 (26)	5	7	3	33 (64)
	T-SIS$_{(0.75,0.75)}$	37	3	17	94 (97)	101	6	75	150 (208)
	CC-SIS$_{(0.25,0.25)}$	1	11	4	35 (88)	1	32	11	109 (196)
	CC-SIS$_{(0.5,0.5)}$	3	3	2	8 (17)	3	4	3	16 (44)
	CC-SIS$_{(0.75,0.75)}$	16	1	4	37 (84)	37	1	25	116 (193)
	mCC-SIS	2	2	3	5 (7)	3	2	7	16 (31)

Table 13: The overlaps of selected genes probes using various approaches for the breast cancer data, where the screening threshold parameter is set as $d_n = \lfloor n/\log n \rfloor = 21$ for each method. The CPC-SISa$_1$ means the CPC-SIS in Case 1, the CPC-SISa$_2$ indicates the CPC-SIS in Case 2, and the CPC-SISa$_3$ stands for the CPC-SIS in Case 3.

					QC-SIS(τ)			QPC-SIS(τ)			CC-SIS(τ, ι)			CPC-SISa$_1$	CPC-SISa$_2$	CPC-SISa$_3$
	SIS	SIRS	DC-SIS	Kendall-SIS	0.25	0.5	0.75	0.25	0.5	0.75	(0.25,0.25)	(0.5,0.5)	(0.75,0.75)	(0.5,0.5)	(0.5,0.5)	(0.5,0.5)
SIS	21	8	7	4	1	7	5	1	1	0	2	1	1	0	1	0
SIRS	8	21	13	9	1	15	2	1	0	0	3	5	0	0	1	0
DC-SIS	7	13	21	11	0	8	3	0	0	0	1	6	0	0	1	0
Kendall	4	9	11	21	0	3	4	0	0	0	0	3	0	0	0	0
QC-SIS$_{(0.25)}$	1	1	0	0	21	1	0	0	0	0	1	0	0	0	0	0
QC-SIS$_{(0.5)}$	7	15	8	3	1	21	0	1	0	0	3	7	0	0	1	0
QC-SIS$_{(0.75)}$	5	2	3	4	0	0	21	0	1	0	0	0	5	0	0	0
QPC-SIS$_{(0.25)}$	1	1	0	0	0	1	0	21	0	0	0	0	0	0	0	0
QPC-SIS$_{(0.5)}$	1	0	0	0	0	0	1	0	21	0	0	0	0	0	0	0
QPC-SIS$_{(0.75)}$	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0
CC-SIS$_{(0.25,0.25)}$	2	3	1	0	1	3	0	0	0	0	21	1	0	0	0	0
CC-SIS$_{(0.5,0.5)}$	1	5	6	3	0	7	0	0	0	0	1	21	0	0	1	0
CC-SIS$_{(0.75,0.75)}$	1	0	0	0	0	0	5	0	0	0	0	0	21	0	1	0
CPC-SISa$_{1(0.5,0.5)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0
CPC-SISa$_{2(0.5,0.5)}$	1	1	1	0	0	1	0	0	0	0	0	1	1	0	21	0
CPC-SISa$_{3(0.5,0.5)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21