Regularized Quantile Regression and Robust Feature Screening for Single Index Models

Wei Zhong, Liping Zhu, Runze Li and Hengjian Cui

Xiamen University, Shanghai University of Finance and Economics
Pennsylvania State University and Capital Normal University

Supplementary Material

S1 Double Robust SIS based on Distance Correlation

In this section, we propose to quantify the importance of X_k through its distance correlation between the respective marginal distribution functions of X_k and Y. That is,

$$\rho_k = \mathrm{dcorr}\{F_k(X_k), F(Y)\}, \tag{S1.1}$$

where $F_k(x) = E\{\mathbf{1}(X_k \le x)\}$, $F(y) = E\{\mathbf{1}(Y \le y)\}$ and $\mathbf{1}(\cdot)$ denotes an indicator function.

Let $\{(\mathbf{x}_i, Y_i), i = 1, \cdots, n\}$ be a random sample from the population $(\mathbf{x}, Y)$. We first estimate the distance covariance between $F_k(X_k)$ and $F(Y)$ through the moment estimation method,

$$\widehat{\mathrm{dcov}}^2\{F_k(X_k), F(Y)\} = \widehat{S}_{k,1} + \widehat{S}_{k,2} - 2\widehat{S}_{k,3}, \tag{S1.2}$$

where

$$\begin{aligned}
\widehat{S}_{k,1} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|F_{nk}(X_{ik}) - F_{nk}(X_{jk})|\,|F_n(Y_i) - F_n(Y_j)|\,,\\
\widehat{S}_{k,2} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|F_{nk}(X_{ik}) - F_{nk}(X_{jk})|\,\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|F_n(Y_i) - F_n(Y_j)|\,, \text{ and}\\
\widehat{S}_{k,3} &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}|F_{nk}(X_{ik}) - F_{nk}(X_{lk})|\,|F_n(Y_j) - F_n(Y_l)|\,.
\end{aligned}$$

are corresponding estimators of $S_{k,1}$, $S_{k,2}$ and $S_{k,3}$, $F_{nk}(x)$ is the empirical distribution function of X_k and $F_n(y)$ is the empirical distribution function of Y, i.e. $F_{nk}(x) = n^{-1}\sum_{i=1}^{n}\mathbf{1}(X_{ik} \le x)$ and $F_n(y) = n^{-1}\sum_{i=1}^{n}\mathbf{1}(Y_i \le y)$. Then we can estimate ρ_k with

$$\widehat{\rho}_k = \widehat{\mathrm{dcorr}}\{F_k(X_k), F(Y)\} = \frac{\widehat{\mathrm{dcov}}(F_k(X_k), F(Y))}{\sqrt{\widehat{\mathrm{dcov}}(F_k(X_k), F_k(X_k))\widehat{\mathrm{dcov}}(F(Y), F(Y))}}. \tag{S1.3}$$

We retain the covariates with the $\widehat{\rho}_k$ values larger than a user-specified threshold. Denote

$$\widehat{\mathcal{A}} = \left\{k : \widehat{\rho}_k \ge cn^{-\kappa},\ \text{for } 1 \le k \le p\right\}$$

for some pre-specified thresholds $c > 0$ and $0 \leq \kappa < 1/2$. The constants c and κ control the signal strength. We refer to this approach as double robust independence screening procedure based on distance correlation (DC-2RoSIS).

Without imposing any moment conditions on the covariates $\mathbf{x}$ or the response variable Y, the following theorem demonstrates the sure screening property of the proposed DC-2RoSIS.

Theorem 1. (Sure Screening Property) There exists a positive constant c_1 such that

$$\Pr\left(\max_{1\leq k\leq p}|\widehat{\rho}_k - \rho_k| \geq cn^{-\kappa}\right) \leq O\Big(p\exp\left\{-c_1 n^{1-2\kappa}\right\}\Big). \tag{S1.4}$$

If the marginal utility defined in (S1.1) satisfies $\min_{k\in\mathcal{A}}\rho_k \geq 2cn^{-\kappa}$, for some constants $c > 0$ and $0 \leq \kappa < 1/2$, it follows that

$$\Pr\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}\right) \geq 1 - O\Big(s_n \exp\left\{-c_1 n^{1-2\kappa}\right\}\Big), \tag{S1.5}$$

where s_n is the cardinality of $\mathcal{A}$. Thus, $\Pr\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}\right) \to 1$ as $n \to \infty$.

Proof: For notational clarity, we use c_1 to denote a generic positive constant. First we assume $F(y)$ and $F_k(x)$ are known. Then, we define that

$$\widehat{\mathrm{dcov}}^{*2}\{F_k(X_k), F(Y)\} = \widehat{S}^*_{k1} + \widehat{S}^*_{k2} - 2\widehat{S}^*_{k3},$$

where

$$\begin{aligned}
\widehat{S}^*_{k,1} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |F_k(X_{ik}) - F_k(X_{jk})|\,|F(Y_i) - F(Y_j)|\,,\\
\widehat{S}^*_{k,2} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |F_k(X_{ik}) - F_k(X_{jk})|\,\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |F(Y_i) - F(Y_j)|\,, \text{ and}\\
\widehat{S}^*_{k,3} &= \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n |F_k(X_{ik}) - F_k(X_{lk})|\,|F(Y_j) - F(Y_l)|\,.
\end{aligned}$$

Similarly, we define $\widehat{\rho}^*_k = \widehat{\mathrm{dcorr}}^{*2}\{F_k(X_k), F(Y)\}$. Note that two distribution functions $F(y)$ and $F_k(x)$ are bounded, we follow the proof of Theorem 1 of Li, Zhong, and Zhu (2012) to easily obtain that there exists a positive constant $c_1 > 0$ such that

$$\Pr\left(\max_{1\leq k\leq p}|\widehat{\rho}^*_k - \rho_k| \geq cn^{-\kappa}\right) \leq O\left(p\exp\left\{-c_1 n^{1-2\kappa}\right\}\right). \tag{S1.6}$$

To prove Theorem 1, it thus suffices to show the difference between $\widehat{\rho}^*_k$ and $\widehat{\rho}_k$ defined in (S1.3) is ignorable enough when n is large. Next, we recall that $\widehat{S}^*_{k1} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |F_k(X_{ik}) -$

$F_k(X_{jk})||F(Y_i) - F(Y_j)|$ and $\widehat{S}_{k1} = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n |F_{nk}(X_{ik}) - F_{nk}(X_{jk})||F_n(Y_i) - F_n(Y_j)|$.

$$\begin{aligned}
& \Pr\left(\max_{1\le k\le p} |\widehat{S}^*_{k1} - \widehat{S}_{k1}| \ge \varepsilon\right) \\
= \quad & \Pr\left(\max_{1\le k\le p} n^{-2}\sum_{i=1}^n\sum_{j=1}^n \Big||F_k(X_{ik}) - F_k(X_{jk})| - |F_{nk}(X_{ik}) - F_{nk}(X_{jk})|\Big||F(Y_i) - F(Y_j)| \ge \frac{\varepsilon}{2}\right) \\
+ \quad & \Pr\left(\max_{1\le k\le p} n^{-2}\sum_{i=1}^n\sum_{j=1}^n \Big||F(Y_i) - F(Y_j)| - |F_n(Y_i) - F_n(Y_j)|\Big||F_{nk}(X_{ik}) - F_{nk}(X_{jk})| \ge \frac{\varepsilon}{2}\right) \\
=: \quad & T_1 + T_2.
\end{aligned}$$

Using the argument that $\big||x| - |y|\big| \le |x - y| \le |x| + |y|$, we obtain that

$$\begin{aligned}
& \big|\,|F_n(Y_i) - F_n(Y_j)| - |F(Y_i) - F(Y_j)|\,\big| \\
\le \quad & |F_n(Y_i) - F(Y_i)| + |F_n(Y_j) - F(Y_j)| \le 2\max_{1\le i\le n} |F_n(Y_i) - F(Y_i)|\,.
\end{aligned}$$

Also because $|F_{nk}(X_{ik}) - F_{nk}(X_{jk})| \le 1$, we have

$$\begin{aligned}
T_2 \quad \le \quad & \Pr\left(\max_{1\le k\le p} 4\max_{1\le i\le n} |F_n(Y_i) - F(Y_i)| \ge \varepsilon\right) \\
\le \quad & \Pr\left(\max_{y\in R} |F_n(y) - F(y)| \ge \varepsilon/4\right) \le 2\exp(-n\varepsilon^2/8),
\end{aligned} \tag{S1.7}$$

where the last inequality follows by Dvoretzky-Kiefer-Wolfowitz inequality.

Similarly, we can prove that

$$\begin{aligned}
T_1 \quad \le \quad & \Pr\left(\max_{1\le k\le p} 4\max_{1\le i\le n} |F_{nk}(X_{ik}) - F_k(X_{ik})| \ge \varepsilon\right) \\
\le \quad & p\Pr\left(\max_{x\in R} |F_{nk}(x) - F_k(x)| \ge \varepsilon/4\right) \le 2p\exp(-n\varepsilon^2/8).
\end{aligned} \tag{S1.8}$$

Then, we have

$$\Pr\left(\max_{1\le k\le p} |\widehat{S}^*_{k1} - \widehat{S}_{k1}| \ge \varepsilon\right) \le 2(p+1)\exp(-n\varepsilon^2/8).$$

We can follow the above argument to obtain the similar upper bounds for the probabilities $\Pr\left(\max_{1\le k\le p} |\widehat{S}^*_{km} - \widehat{S}_{km}| \ge \varepsilon\right)$ with $m = 2, 3$. Thus, it is not difficult to show that

$$\Pr\left(\max_{1\le k\le p} |\widehat{\rho}^*_k - \widehat{\rho}_k| \ge cn^{-\kappa}\right) \le O\left(p\exp\left\{-c_1 n^{1-2\kappa}\right\}\right). \tag{S1.9}$$

Therefore, (S1.6) and (S1.9) together complete the proof of the first part of Theorem 1. The second part follows directly by the proof of Theorem 1 of Li, Zhong, and Zhu (2012). □

Example S.1. Next we investigate the numerical performance of the DC-2RoSIS and compare it with other existing sure independence screening procedures. We consider two regression models in this example.

Model (I): $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_7 X_7 + \varepsilon$;

Model (II): $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_7 X_7^2 + \varepsilon$;

where $\beta_1 = 3, \beta_2 = 1.5$ and $\beta_7 = 2$, which imply that only X_1, X_2 and X_7 are truly important. The random error ε is independently generated from either standard normal or standard Cauchy distribution. We consider two cases for generating the covariates. In case (1), the covariate vector is drawn from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 0.5^{|i-j|}$; In case (2), all covariates are drawn independently from standard Cauchy distribution. We fix $p = 1,000$ and $n = 200$. The simulation results are summarized in Table 1.

It can be seen from Table 1 that both DC-RoSIS and DC-2RoSIS have very similar performance and both are superior to other independence screening methods when extreme values are merely present in the response realizations. When both the covariates and the response are extremely heavily tailed, for example, in case (2), DC-2RoSIS is slightly better than DC-RoSIS.

Table 1: Performance comparison among different independence screening methods for two regression models in Example S.1.

		$\varepsilon \sim \mathcal{N}(0,1)$					$\varepsilon \sim$ Cauchy Distribution				
Covariates $\mathbf{x}$	Method	$\mathcal{S}$	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_a$	Size	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_a$
Model (I). Linear Model											
Case (1) Normal	SIS	3.0(0.0)	1.00	1.00	1.00	1.00	225.0(517.7)	0.72	0.60	0.47	0.39
	DC-SIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	0.99	0.99	0.93	0.93
	SIRS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
	RRCS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	1.00	1.00
	DC-RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	1.00	1.00
	DC-2RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	1.00	1.00
Case (2) Cauchy	SIS	31.5(45.9)	0.96	0.82	0.91	0.75	39.5(48.3)	0.94	0.82	0.90	0.71
	DC-SIS	12.0(17.2)	0.99	0.94	0.98	0.94	16.5(19.6)	1.00	0.94	0.95	0.92
	SIRS	3.0(9.3)	0.99	0.98	1.00	0.97	4.0(8.6)	1.00	0.99	0.98	0.97
	RRCS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
	DC-RoSIS	3.0(0.0)	0.98	0.96	1.00	0.94	3.0(0.0)	0.99	0.99	0.96	0.94
	DC-2RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
Model (II). Additive Model											
Case (1) Normal	SIS	249.0(348.3)	1.00	1.00	0.23	0.23	558.0(375.9)	0.73	0.67	0.06	0.03
	DC-SIS	4.0(0.9)	1.00	1.00	1.00	1.00	7.0(10.5)	0.99	0.99	0.83	0.83
	SIRS	328.0(396.3)	1.00	1.00	0.12	0.12	554.0(392.4)	1.00	1.00	0.10	0.10
	RRCS	304.0(333.9)	1.00	1.00	0.19	0.19	546.0(343.1)	1.00	1.00	0.07	0.07
	DC-RoSIS	4.0(1.7)	1.00	1.00	1.00	1.00	6.0(3.9)	1.00	1.00	0.99	0.99
	DC-2RoSIS	6.0(4.7)	1.00	1.00	1.00	1.00	17.5(16.0)	1.00	1.00	0.92	0.93
Case (2) Cauchy	SIS	533.0(470.2)	0.26	0.13	0.94	0.04	400.5(455.2)	0.28	0.17	0.97	0.04
	DC-SIS	201.5(180.4)	0.41	0.23	1.00	0.12	219.5(184.1)	0.43	0.26	1.00	0.18
	SIRS	142.0(422.6)	0.98	0.97	0.39	0.37	218.5(529.5)	0.98	0.93	0.32	0.27
	RRCS	318.0(355.2)	1.00	1.00	0.14	0.14	384.0(414.5)	1.00	0.94	0.16	0.16
	DC-RoSIS	3.0(0.0)	0.99	0.94	0.98	0.91	3.0(10.3)	0.96	0.98	0.98	0.92
	DC-2RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(1.5)	1.00	0.98	1.00	0.98

S2 Simulation Example 1 with More Active Covariates

We further consider the simulation Example 1 in the paper. The simulation settings are the same as Example 1 except that 6 active covariates are considered here. The true coefficient is set as

$$\boldsymbol{\beta} = (3, 1.5, 0, 0, 0, 0, 2, 0, 0, 0, 3, 1.5, 0, 0, 0, 0, 2, 0, \ldots, 0)^{\mathrm{T}},$$

where X_1, X_2, X_7, X_{11}, X_{12} and X_{17} are truly important. We repeat each experiment 500 times and summarize the simulation results in Tables 2 and 3 using the same criteria as in Example 1.

The simulation results deliver similar messages as in Table 4.1 in the paper. Even when more active covariates are considered, DC-RoSIS still improves DC-SIS, especially when the response variable is heavily-tailed or has some extreme values. DC-RoSIS outperforms other robust existing independence screening procedures when the active covariates exhibit symmetric patterns with the response. This example further demonstrates that the proposed DC-RoSIS provides an alternative procedure for ultrahigh dimensional screening literature.

S3 Simulation Example 1 with Weak Signal-to-Noise Ratios

We further consider the simulation Example 1 in the paper. The simulation settings are the same as Example 1 except that weaker signal-to-noise ratios are considered. To be specific, we change the true coefficients in two ways: $\boldsymbol{\beta} = (3, 1.5, 0, 0, 0, 0, 2, 0, \ldots, 0)^{\mathrm{T}}/2$ and $\boldsymbol{\beta} = (3, 1.5, 0, 0, 0, 0, 2, 0, \ldots, 0)^{\mathrm{T}}/4$. We repeat each experiment 500 times and summarize the simulation results using the same criteria in Tables 4 and 5 as in Example 1.

The simulation results deliver similar messages as in Table 4.1 in the paper. As the signal-to-noise ratios become weaker, the performances of all independence screening methods become worse. However, the DC-RoSIS can still improve the original DC-SIS especially when the response variable is heavily-tailed or has some extreme values. Both DC-RoSIS and DC-SIS outperform other robust independence screenings when the active covariates exhibits symmetric patterns with the response.

References

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics* **27**, 642–669.

Li, R., Zhong, W., and Zhu, L.(2012). Feature screening via distance correlation learning, *Journal of the American Statistical Association* **107**, 1129–1139.

Table 2: Performance comparison among different independence screening methods for four regression models with 6 active covariates and normal random errors.

	Method	$\mathcal{S}$	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_{s11}$	$\mathcal{P}_{s12}$	$\mathcal{P}_{s17}$	$\mathcal{P}_a$
Model (1) ($\lambda = 1$)	SIS	8.0(1.5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	DC-SIS	8.0(2.2)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SIRS	8.0(2.2)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	RRCS	8.0(2.2)	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	DC-RoSIS	8.0(2.2)	1.00	1.00	1.00	1.00	1.00	1.00	0.99
Model (1) ($\lambda = 0.25$)	SIS	19.0(32.8)	1.00	1.00	0.91	1.00	1.00	0.88	0.81
	DC-SIS	10.0(8.9)	1.00	1.00	0.98	1.00	1.00	0.96	0.94
	SIRS	8.0(1.5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	RRCS	8.0(2.2)	1.00	1.00	1.00	1.00	1.00	0.99	0.99
	DC-RoSIS	8.0(2.2)	1.00	1.00	1.00	1.00	1.00	0.99	0.99
Model (2)	SIS	340.0(333.2)	0.83	0.69	0.47	0.86	0.71	0.39	0.09
	DC-SIS	125.0(177.9)	0.98	0.88	0.66	0.99	0.93	0.60	0.36
	SIRS	18.0(27.6)	1.00	1.00	0.94	1.00	1.00	0.88	0.82
	RRCS	16.0(23.1)	1.00	1.00	0.94	1.00	1.00	0.90	0.84
	DC-RoSIS	19.0(30.6)	1.00	1.00	0.95	1.00	1.00	0.87	0.82
Model (3)	SIS	883.0(127.2)	0.09	0.07	0.08	0.08	0.07	0.06	0.00
	DC-SIS	78.5(97.1)	1.00	0.97	0.80	1.00	0.98	0.69	0.49
	SIRS	72.0(63.6)	1.00	0.98	0.80	1.00	1.00	0.68	0.53
	RRCS	878.5(149.6)	0.08	0.09	0.09	0.09	0.09	0.05	0.00
	DC-RoSIS	83.0(83.5)	1.00	0.97	0.76	1.00	0.99	0.67	0.50
Model (4)	SIS	592.5(332.3)	1.00	1.00	0.19	1.00	1.00	0.15	0.03
	DC-SIS	17.0(17.9)	1.00	1.00	0.97	1.00	1.00	0.94	0.92
	SIRS	646.0(312.3)	1.00	1.00	0.14	1.00	1.00	0.11	0.01
	RRCS	643.5(319.9)	1.00	1.00	0.15	1.00	1.00	0.10	0.02
	DC-RoSIS	23.0(22.4)	1.00	1.00	0.95	1.00	1.00	0.94	0.90

Table 3: Performance comparison among different independence screening methods for four regression models with 6 active covariates and Cauchy random errors.

	Method	$\mathcal{S}$	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_{s11}$	$\mathcal{P}_{s12}$	$\mathcal{P}_{s17}$	$\mathcal{P}_a$
Model (1) ($\lambda = 1$)	SIS	458.5(529.4)	0.69	0.58	0.46	0.67	0.60	0.42	0.23
	DC-SIS	11.0(22.0)	0.99	0.97	0.91	0.99	0.97	0.88	0.84
	SIRS	9.0(5.2)	1.00	1.00	0.99	1.00	1.00	0.98	0.97
	RRCS	9.0(6.7)	1.00	1.00	0.99	1.00	1.00	0.97	0.97
	DC-RoSIS	9.0(6.7)	1.00	1.00	0.99	1.00	1.00	0.97	0.96
Model (1) ($\lambda = 0.25$)	SIS	899.5(115.5)	0.08	0.07	0.08	0.08	0.12	0.06	0.00
	DC-SIS	842.0(178.7)	0.16	0.13	0.13	0.17	0.19	0.10	0.02
	SIRS	8.0(3.9)	1.00	1.00	1.00	1.00	1.00	0.97	0.97
	RRCS	9.0(5.2)	1.00	1.00	0.99	1.00	1.00	0.97	0.96
	DC-RoSIS	9.0(5.4)	1.00	1.00	0.99	1.00	1.00	0.96	0.95
Model (2)	SIS	483.5(417.9)	0.73	0.57	0.39	0.77	0.60	0.35	0.07
	DC-SIS	195.0(232.5)	0.95	0.82	0.54	0.96	0.85	0.49	0.23
	SIRS	21.0(38.8)	1.00	0.99	0.94	1.00	1.00	0.85	0.79
	RRCS	19.5(35.1)	1.00	0.99	0.95	1.00	1.00	0.87	0.82
	DC-RoSIS	23.0(38.3)	1.00	1.00	0.94	1.00	1.00	0.84	0.78
Model (3)	SIS	899.0(116.6)	0.09	0.08	0.07	0.08	0.07	0.06	0.00
	DC-SIS	337.0(290.7)	0.76	0.56	0.37	0.80	0.60	0.29	0.08
	SIRS	70.0(59.9)	1.00	0.99	0.80	1.00	1.00	0.66	0.53
	RRCS	893.0(121.6)	0.07	0.09	0.08	0.08	0.09	0.07	0.00
	DC-RoSIS	86.5(74.4)	1.00	0.98	0.77	1.00	0.99	0.65	0.48
Model (4)	SIS	753.5(282.1)	0.70	0.64	0.11	0.70	0.62	0.10	0.01
	DC-SIS	62.0(124.5)	0.98	0.96	0.71	0.98	0.96	0.69	0.55
	SIRS	679.5(313.6)	1.00	1.00	0.14	1.00	1.00	0.08	0.00
	RRCS	681.5(300.2)	1.00	1.00	0.12	1.00	1.00	0.09	0.01
	DC-RoSIS	44.0(58.4)	1.00	1.00	0.83	1.00	1.00	0.81	0.67

Table 4: Performance comparison among different independence screening methods for four regression models with two different random errors when $\boldsymbol{\beta} = (3, 1.5, 0, 0, 0, 0, 2, 0, \ldots, 0)^{\mathrm{T}}/2$.

		$\varepsilon \sim \mathcal{N}(0,1)$					$\varepsilon \sim$ Cauchy Distribution				
	Method	$\mathcal{S}$	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_a$	Size	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_a$
Model (1) ($\lambda = 1$)	SIS	3.0(0.0)	1.00	1.00	1.00	1.00	488.5(456.5)	0.50	0.44	0.25	0.17
	DC-SIS	3.0(0.0)	1.00	1.00	1.00	1.00	5.0(18.8)	0.97	0.95	0.84	0.83
	SIRS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	0.99	0.99
	RRCS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	0.99	0.99
	DC-RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	0.99	0.99
Model (1) ($\lambda = 0.25$)	SIS	3.0(0.7)	1.00	1.00	1.00	1.00	789.0(202.4)	0.07	0.08	0.07	0.00
	DC-SIS	3.0(0.0)	1.00	1.00	1.00	1.00	764.0(236.0)	0.09	0.09	0.07	0.01
	SIRS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	1.00	1.00
	RRCS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	0.99	0.99
	DC-RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	0.99	0.99
Model (2)	SIS	3.0(0.0)	1.00	1.00	1.00	1.00	6.0(99.6)	0.89	0.83	0.73	0.69
	DC-SIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	1.00	1.00	0.98	0.98
	SIRS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
	RRCS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
	DC-RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
Model (3)	SIS	796.0(216.6)	0.08	0.07	0.08	0.00	807.0(220.7)	0.09	0.08	0.08	0.00
	DC-SIS	5.0(5.2)	1.00	1.00	0.97	0.97	74.5(146.5)	0.91	0.82	0.55	0.50
	SIRS	9.0(9.7)	1.00	1.00	0.98	0.98	13.0(13.4)	1.00	1.00	0.97	0.97
	RRCS	780.0(219.0)	0.10	0.09	0.09	0.00	780.0(215.5)	0.09	0.08	0.06	0.00
	DC-RoSIS	10.0(13.4)	1.00	1.00	0.96	0.96	14.0(15.8)	1.00	1.00	0.94	0.94
Model (4)	SIS	307.5(404.5)	1.00	1.00	0.24	0.24	633.0(358.9)	0.54	0.43	0.06	0.03
	DC-SIS	4.0(1.5)	1.00	1.00	0.99	0.99	36.5(99.6)	0.95	0.92	0.66	0.65
	SIRS	432.5(425.4)	1.00	1.00	0.13	0.13	451.0(385.3)	1.00	1.00	0.12	0.12
	RRCS	442.5(402.2)	1.00	1.00	0.13	0.14	430.5(399.6)	1.00	1.00	0.10	0.10
	DC-RoSIS	4.0(1.5)	1.00	1.00	0.99	0.99	13.0(16.6)	1.00	1.00	0.94	0.94

Table 5: Performance comparison among different independence screening methods for four regression models with two different random errors when $\boldsymbol{\beta} = (3, 1.5, 0, 0, 0, 0, 2, 0, \ldots, 0)^{\mathrm{T}}/4$.

		$\varepsilon \sim \mathcal{N}(0,1)$					$\varepsilon \sim$ Cauchy Distribution				
	Method	$\mathcal{S}$	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_a$	Size	$\mathcal{P}_{s1}$	$\mathcal{P}_{s2}$	$\mathcal{P}_{s7}$	$\mathcal{P}_a$
Model (1) ($\lambda = 1$)	SIS	3.0(0.0)	1.00	1.00	1.00	1.00	718.5(298.7)	0.25	0.18	0.11	0.02
	DC-SIS	3.0(0.7)	1.00	1.00	1.00	1.00	84.0(177.6)	0.91	0.82	0.52	0.49
	SIRS	3.0(0.7)	1.00	1.00	1.00	1.00	6.0(11.9)	1.00	1.00	0.92	0.92
	RRCS	3.0(0.7)	1.00	1.00	1.00	1.00	9.0(19.1)	1.00	1.00	0.90	0.89
	DC-RoSIS	3.0(0.7)	1.00	1.00	1.00	1.00	7.0(13.4)	1.00	1.00	0.92	0.92
Model (1) ($\lambda = 0.25$)	SIS	4.0(2.7)	1.00	0.99	0.99	0.98	787.0(232.3)	0.08	0.11	0.05	0.00
	DC-SIS	3.0(0.7)	1.00	1.00	1.00	1.00	766.5(221.3)	0.09	0.11	0.06	0.00
	SIRS	3.0(0.0)	1.00	1.00	1.00	1.00	6.0(14.2)	1.00	1.00	0.89	0.88
	RRCS	3.0(0.7)	1.00	1.00	1.00	1.00	9.0(20.9)	1.00	1.00	0.85	0.85
	DC-RoSIS	3.0(0.7)	1.00	1.00	1.00	1.00	7.0(17.9)	1.00	1.00	0.87	0.87
Model (2)	SIS	3.0(0.0)	1.00	1.00	1.00	1.00	45.0(404.1)	0.83	0.74	0.58	0.53
	DC-SIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.7)	0.99	0.98	0.96	0.96
	SIRS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
	RRCS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
	DC-RoSIS	3.0(0.0)	1.00	1.00	1.00	1.00	3.0(0.0)	1.00	1.00	1.00	1.00
Model (3)	SIS	769.0(243.6)	0.12	0.09	0.07	0.01	797.0(201.5)	0.07	0.07	0.08	0.00
	DC-SIS	5.0(9.2)	1.00	1.00	0.96	0.96	133.5(200.7)	0.87	0.74	0.45	0.38
	SIRS	13.0(16.6)	1.00	1.00	0.93	0.93	26.5(32.1)	1.00	1.00	0.82	0.82
	RRCS	773.5(241.2)	0.10	0.08	0.09	0.00	772.5(223.5)	0.11	0.07	0.07	0.00
	DC-RoSIS	15.0(20.3)	1.00	1.00	0.89	0.90	31.0(38.8)	1.00	1.00	0.78	0.78
Model (4)	SIS	352.5(392.5)	1.00	1.00	0.20	0.20	710.0(282.1)	0.22	0.17	0.09	0.01
	DC-SIS	5.0(4.5)	1.00	1.00	0.99	1.00	187.0(237.3)	0.92	0.82	0.26	0.23
	SIRS	458.0(399.6)	1.00	1.00	0.12	0.12	494.5(400.4)	1.00	1.00	0.07	0.07
	RRCS	433.0(386.7)	1.00	1.00	0.10	0.10	509.0(384.1)	1.00	0.99	0.06	0.06
	DC-RoSIS	7.0(7.4)	1.00	1.00	0.98	0.98	50.0(77.8)	1.00	1.00	0.60	0.60