Supplementary Materials for "REGULARIZATION AFTER RETENTION IN ULTRAHIGH DIMENSIONAL LINEAR REGRESSION MODELS"

Haolei Weng, Yang Feng and Xingye Qiao

Columbia University, Columbia University and Binghamton University

This supplement contains proofs of all theoretical results and additional simulation results.

1 Proofs

Proof of Proposition 1. We refer to the general result of Theorem 3 in Wainwright (2009). By Condition 3, we have as $n \to \infty$,

$$\lambda_n^2 n/\{\rho(\Sigma_{S^c|S}) \log p_n\} > n^{\delta}/\rho(\Sigma_{S^c|S}) = 1/o(1) \to \infty, \tag{A1}$$

$$\begin{aligned} n/\{\rho(\Sigma_{S^c|S}) s_n \log(p_n - s_n)\} &> n/\{\rho(\Sigma_{S^c|S}) s_n \log p_n\} \\ &> n^{1-a_1-2a_2-\delta}/o(1) \to \infty. \end{aligned} \tag{A2}$$

With the same notations as in Wainwright (2009), (A1) implies $\phi_p \to \infty$, and (A2) shows that equation (34) in Wainwright (2009) holds. Then we need to show $\min_{j\in S} |\beta_j| > g(\lambda_n)$, for sufficient large n, where $g(\lambda_n)$ is defined by equation (33) in Wainwright (2009),

$$g(\lambda_n) \le c_3 \lambda_n s_n \Lambda^2_{\max}(\Sigma_{SS}^{-1/2}) + O\{(\log n/n)^{1/2}\} = O(n^{(\delta+a_1+2a_2-1)/2}) < \min_{j\in S} |\beta_j|,$$

due to $\Lambda^2_{\max}(\Sigma_{SS}^{-1/2}) = \Lambda_{\max}(\Sigma_{SS}^{-1}) = 1/\Lambda_{\min}(\Sigma_{SS})$ and Conditions 1, 3, 4. $\square$

Proof of Proposition 2. We refer to Theorem 4 in Fan and Song (2010) about uniform convergence of maximum marginal likelihood estimator in the generalized linear model. As in Fan and Song (2010), let $\boldsymbol{\beta}_j = (\beta_{j,0}, \beta_j)^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ be two-dimensional vectors and $\boldsymbol{X}_j = (1, X_j)^T$, where X_j is the jth predictor. Denote $\mathcal{B} = \{|\beta_{j,0}| \leq B, |\beta_j| \leq B\}$ and the true coefficient by $\boldsymbol{\beta}^\star$. We need to verify Conditions A', B', C' and D in Fan and Song (2010). Condition A' is trivial since $b(\theta) = \theta^2/2$. The expected marginal loglikelihood $E\{l(\boldsymbol{X}_j^T \boldsymbol{\beta}_j, Y)\}$ is a quadratic function of $\boldsymbol{\beta}_j$ with identity Hessian matrix. Thus, Condition C' is satisfied with $V = 1/2$. For Condition B',

$$|E\{b(\boldsymbol{X}_j^T \boldsymbol{\beta}) I(|X_j| > K_n)\}| = \beta_0^2/2P(|X_j| > K_n) + \beta_1^2/2E\{X_j^2 I(|X_j| > K_n)\} \tag{A3}$$

$$\leq \beta_0^2 e^{-K_n^2/2} + \beta_1^2(1 + K_n) e^{-K_n^2/2}. \tag{A4}$$

(A3) holds because X_j is symmetric. (A4) follows from the standard normal distribution of X_j. As mentioned in Section 5.2 in Fan and Song (2010), the optimal order of $K_n = n^{(1-2\kappa)/8}$. Taking B= $O(n^{(1-2\kappa)/8})$ and using Condition 5 , we get for any $\varepsilon > 0$,

$$\sup_{\boldsymbol{\beta} \in \mathcal{B},\ \|\boldsymbol{\beta} - \boldsymbol{\beta}_j^M\| \leq \varepsilon} |E\{b(\boldsymbol{X}_j^T \boldsymbol{\beta})(|X_j| > K_n)\}| \leq o(1/n).$$

With moment generating function of $|X_j|$, we know $\mathrm{pr}(|X_j| > t) \leq 2\exp(-t^2/2)$ and

$$E\{\exp\big(b(\boldsymbol{X}^T \boldsymbol{\beta}^\star + s_0) - b(\boldsymbol{X}^T \boldsymbol{\beta}^\star)\big)\} + E\{\exp\big(b(\boldsymbol{X}^T \boldsymbol{\beta}^\star - s_0) - b(\boldsymbol{X}^T \boldsymbol{\beta}^\star)\big)\} \tag{A5}$$

$$= 2\exp\Big(s_0^2\big(1 + (\boldsymbol{\beta}^\star)^T \Sigma \boldsymbol{\beta}^\star\big)/2\Big). \tag{A6}$$

Let positive constants $\alpha = 2, m_0 = 1/2, m_1 - s_1 = 2, s_0 = 1$. By Condition 6, there exists positive constant s_1 such that (A5) $\leq s_1$. Therefore, Condition D holds. $\square$

Proof of Theorem 1. Denote the design matrix by X, response vector by Y, and error vector by ε. Define $\bar{S}^c = \hat{R}^c \backslash S^c$. Let

$$\check{\beta} = \arg\min_{\beta} \left\{ (2n)^{-1} \|Y - X_{\hat{R}}\beta_{\hat{R}} - X_{\hat{R}^c}\beta_{\hat{R}^c}\|_2^2 + \lambda_n \|\beta_{\hat{R}^c}\|_1 \right\}, \tag{A7}$$

$$\bar{\beta} = \arg\min_{\beta_{S^c}=0} \left\{ (2n)^{-1} \|Y - X_S\beta_S\|_2^2 + \lambda_n \|\beta_{\bar{S}^c}\|_1 \right\}. \tag{A8}$$

Since $X_S^T X_S \sim W_{s_n}(\Sigma_{SS}, n)$, which is Wishart distribution, when the number of signals $s_n < n$, as in our scaling, X_S is of full rank with probability one. Therefore, (A8) is a strictly convex problem and $\bar{\beta}$ is unique with probability one.

By optimality conditions of convex problems (Bach et al., 2012), $\check{\beta}$ is a solution to (A7) if and only if

$$n^{-1} X^T (Y - X\check{\beta}) = \lambda_n \partial \|\check{\beta}_{\hat{R}^c}\|, \tag{A9}$$

where $\partial\|\check{\beta}_{\hat{R}^c}\|$ is the subgradient of $\|\beta_{\hat{R}^c}\|_1$ at $\beta = \check{\beta}$. Namely, the ith ($1 \le i \le p_n$) element of $\partial\|\check{\beta}_{\hat{R}^c}\|$ is

$$(\partial\|\check{\beta}_{\hat{R}^c}\|)_i = \begin{cases} 0 & \text{if } i \in \hat{R} \\ \operatorname{sign}(\check{\beta}_i) & \text{if } i \in \hat{R}^c \text{ and } \check{\beta}_i \neq 0 \\ t & \text{otherwise} \end{cases}$$

where t can be any real number with $|t| \le 1$. Similarly, $\bar{\beta}$ is the unique solution to (A8) if and only if

$$\bar{\beta}_{S^c} = 0, \quad n^{-1} X_S^T (Y - X_S\bar{\beta}_S) = \lambda_n \operatorname{sig}(\bar{\beta}_S), \tag{A10}$$

where $\operatorname{sig}(\bar{\beta}_S)$, a vector of length s_n, is the subgradient of $\|\beta_{\bar{S}^c}\|$ at $\beta_S = \bar{\beta}_S$. Then it is not hard to see that, the unique solution $\bar{\beta}$ is also a solution for (A7) if

$$\|n^{-1} X_{S^c}^T (Y - X_S\bar{\beta}_S)\|_\infty < \lambda_n, \tag{A11}$$

simply because (A10) and (A11) imply $\bar{\beta}$ satisfies (A9). Solving the equation in (A10) gives

$$\bar{\beta}_S = (X_S^T X_S)^{-1} \left[X_S^T Y - n\lambda_n \operatorname{sig}(\bar{\beta}_S) \right]. \tag{A12}$$

Using (A12) and $Y = X_S\beta_S + \varepsilon$, (A11) is equivalent to

$$\|X_{S^c}^T X_S(X_S^T X_S)^{-1}\mathrm{sig}(\bar{\beta}_S) + (n\lambda_n)^{-1}X_{S^c}^T\{I - X_S(X_S^T X_S)^{-1}X_S^T\}\varepsilon\|_\infty < 1. \tag{A13}$$

Based on the optimality conditions of convex problem, we have showed that if the optimization problem (A8)'s unique solution $\bar{\beta}$ satisfies (A13), then $\bar{\beta}$ is also a solution to (A7). On the other hand, it is easily seen that, for any solution $\check{\beta}$ to (A7), $\mathrm{sign}(\check{\beta}) = \mathrm{sign}(\beta)$ only if $\check{\beta}$ is also a solution to (A8). Therefore, If (A7) has a unique solution and $\bar{\beta}$ satisfies (A13), then $\bar{\beta}$ is that unique solution and $\mathrm{Supp}\{\bar{\beta}\} \subseteq S$. Furthermore, if the maximum gap $\|\bar{\beta}_S - \beta_S\|_\infty$ is upper bounded by the minimum absolute magnitude of β_S, we can achieve sign recovery. In summary, let

$$\begin{aligned}
W &= \{\check{\beta} \text{ is unique and } \mathrm{sign}(\check{\beta}) = \mathrm{sign}(\beta)\},\\
W_1 &= \{\text{(A7) has a unique solution and (A13) holds}\},\\
W_2 &= \{\min_{j\in S}|\beta_j| > \|\bar{\beta}_S - \beta_S\|_\infty\}.
\end{aligned}$$

Then, we have

$$\mathrm{pr}(W) \geq \mathrm{pr}(W_1 \cap W_2) \geq 1 - \mathrm{pr}(W_1^c) - \mathrm{pr}(W_2^c) = \mathrm{pr}(W_1) - \mathrm{pr}(W_2^c). \tag{A14}$$

In the following, we will show $P(W_1) \to 1$ and $P(W_2^c) \to 0$ in two steps separately. Since (A7) is similar to Lasso in random design, our proof mainly follows the proof of Theorem 3 in Wainwright (2009). The key difference is that the penalty term in (A7) is random due to the retention step of our method. To take care of that part, we need more notations. Let

$$\begin{aligned}
\mathcal{T} &= \{S_* : R \subseteq S_* \subseteq S\},\\
A &= \{R \subseteq \hat{R} \subseteq S\},\\
B &= \left\{\max_{1\leq j\leq p_n}|\hat{\beta}_j^M - \beta_j^M| \leq c_1 n^{-\kappa}\right\}.
\end{aligned}$$

Then $B \subseteq A$ and $\mathrm{pr}(B) = 1 - O(p_n \exp(-c_2 n^{(1-2\kappa)/4}))$, as we discussed in Section 3.2.

Step I. Let $F = X_{S^c}^T - \Sigma_{S^cS}\Sigma_{SS}^{-1}X_S^T$, and $F(j)$ be the jth row of F. By the property of conditional distribution of multivariate Gaussian, $F^1, \ldots, F^n$ are independently and identically distributed as $N(0, \Sigma_{S^c|S})$, and F is independent of X_S. After simple algebra calculation using $X_{S^c}^T = \Sigma_{S^cS}\Sigma_{SS}^{-1}X_S^T + F$, we get

$$
\begin{aligned}
&X_{S^c}^T X_S (X_S^T X_S)^{-1}\mathrm{sig}(\bar\beta_S) + (n\lambda_n)^{-1} X_{S^c}^T\{I - X_S(X_S^TX_S)^{-1}X_S^T\}\varepsilon \\
&= \Sigma_{S^cS}\Sigma_{SS}^{-1}\mathrm{sig}(\bar\beta_S) + FX_S(X_S^TX_S)^{-1}\mathrm{sig}(\bar\beta_S) \\
&\quad + (n\lambda_n)^{-1}F\{I - X_S(X_S^TX_S)^{-1}X_S^T\}\varepsilon. \qquad \text{(A15)}
\end{aligned}
$$

Let $K_1 = \Sigma_{S^cS}\Sigma_{SS}^{-1}\mathrm{sig}(\bar\beta_S)$ and $K_2 = FX_S(X_S^TX_S)^{-1}\mathrm{sig}(\bar\beta_S) + (n\lambda_n)^{-1}F\{I - X_S(X_S^TX_S)^{-1}X_S^T\}\varepsilon$. Then (A13) is equivalent to $\|K_1 + K_2\|_\infty < 1$. We analyze $\|K_1\|_\infty$ and $\|K_2\|_\infty$ on the high probability set A. Firstly, it is not hard to see,

$$
\begin{aligned}
\mathrm{pr}(\|K_1\|_\infty \le 1-\gamma) &= \mathrm{pr}(\{\|K_1\|_\infty \le 1-\gamma\} \cap A) + \mathrm{pr}(\{\|K_1\|_\infty \le 1-\gamma\} \cap A^c) \\
&\overset{(1)}{=} \mathrm{pr}(A) + \mathrm{pr}(\{\|K_1\|_\infty \le 1-\gamma\} \cap A^c),
\end{aligned}
$$

where (1) holds since when A holds, by Condition 10,

$$\|K_1\|_\infty \le \|\{\Sigma_{S^cS}(\Sigma_{SS})^{-1}\}_{S\cap R^c}\|_\infty \le 1-\gamma.$$

Under the scaling in Theorem 1, $\mathrm{pr}(A) \to 1, \quad \mathrm{pr}(\{\|K_1\|_\infty \le 1-\gamma\} \cap A^c) \le \mathrm{pr}(A^c) \to 0, \quad \text{as} \quad n \to \infty$. Hence,

$$\mathrm{pr}(\|K_1\|_\infty \le 1-\gamma) \to 1, \quad \text{as} \quad n \to \infty. \qquad \text{(A16)}$$

Similarly,

$$
\begin{aligned}
\mathrm{pr}(\|K_2\|_\infty > \frac{\gamma}{2}) &= \mathrm{pr}(\{\|K_2\|_\infty > \frac{\gamma}{2}\} \cap A) + \mathrm{pr}(\{\|K_2\|_\infty > \frac{\gamma}{2}\} \cap A^c) \\
&\leq \mathrm{pr}\left(\left\{\bigcup_{S_1 \in \mathcal{T}} \|K_2(S_1)\|_\infty > \frac{\gamma}{2}\right\} \cap A\right) + \mathrm{pr}(A^c) \\
&\leq \mathrm{pr}\left(\bigcup_{S_1 \in \mathcal{T}} \|K_2(S_1)\|_\infty > \frac{\gamma}{2}\right) + \mathrm{pr}(A^c), \qquad \text{(A17)}
\end{aligned}
$$

where $K_2(S_1)$ is the analogy of K_2 in (A15) by replacing $\hat{R}$ with S_1 in (A7) and (A8). Denote the corresponding solution to (A8) by $\bar{\beta}(S_1)$. Then,

$$
K_2(S_1) = FX_S(X_S^T X_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) + (n\lambda_n)^{-1}F\{I - X_S(X_S^T X_S)^{-1}X_S^T\}\varepsilon.
$$

By the definition of $\bar{\beta}(S_1)$, $\mathrm{sig}(\bar{\beta}_S(S_1))$ is a function of X_S and ε, so

$$
\begin{aligned}
&F(j)X_S(X_S^T X_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) \\
&+ (n\lambda_n)^{-1}F(j)\{I - X_S(X_S^T X_S)^{-1}X_S^T\}\varepsilon \mid (X_S, \varepsilon) \sim N(0, V_j), \qquad \text{(A18)}
\end{aligned}
$$

and

$$
\begin{aligned}
V_j &\leq (\Sigma_{S^c|S})_{jj}[\mathrm{sig}(\bar{\beta}_S(S_1))^T(X_S^T X_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) \\
&\quad +(n\lambda_n)^{-2}\varepsilon^T\{I - X_S(X_S^T X_S)^{-1}X_S^T\}\varepsilon] \\
&\leq \mathrm{sig}(\bar{\beta}_S(S_1))^T(X_S^T X_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) + (n\lambda_n)^{-2}\|\varepsilon\|_2^2,
\end{aligned}
$$

noticing that $\Sigma_{jj} = 1$ and $I - X_S(X_S^T X_S)^{-1}X_S^T$ is an idempotent and symmetric matrix. Let

$$
\begin{aligned}
H = \bigcup_{S_1 \in \mathcal{T}} \Big\{&\mathrm{sig}(\bar{\beta}_S(S_1))^T(X_S^T X_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) + (n\lambda_n)^{-2}\|\varepsilon\|_2^2 > \frac{s_n}{nC_{\min}}(8s_n^{1/2}n^{-1/2} + 1) \\
&+ (1 + s_n^{1/2}n^{-1/2})/(n\lambda_n^2)\Big\}.
\end{aligned}
$$

Then,

$$
\mathrm{pr}(\bigcup_{S_1 \in \mathcal{T}} \|K_2(S_1)\|_\infty > \gamma/2) \leq \mathrm{pr}(\bigcup_{S_1 \in \mathcal{T}} \|K_2(S_1)\|_\infty > \gamma/2 \mid H^c) + \mathrm{pr}(H). \qquad \text{(A19)}
$$

We first bound $\mathrm{pr}(H)$,

$$\begin{aligned}\mathrm{pr}(H) \quad &\leq \quad \mathrm{pr}\Big(\bigcup_{S_1\in\mathcal{T}} \mathrm{sig}(\bar{\beta}_S(S_1))^T(X_S^TX_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) > \frac{s_n}{nC_{\min}}(8s_n^{1/2}n^{-1/2}+1)\Big)\\ &\quad\quad +\mathrm{pr}\Big((n\lambda_n)^{-2}\|\varepsilon\|_2^2 > (1+s_n^{1/2}n^{-1/2})/(n\lambda_n^2)\Big).\end{aligned}$$

For any $S_1 \in \mathcal{T}$,

$$\begin{aligned}\mathrm{sig}(\bar{\beta}_S(S_1))^T(X_S^TX_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) \quad &\leq \quad s_n\|(X_S^TX_S)^{-1}\|_2\\ &\leq \quad s_n/n\Big(\|(X_S^TX_S/n)^{-1}-\Sigma_{SS}^{-1}\|_2+\|\Sigma_{SS}^{-1}\|_2\Big)\\ &\leq \quad s_n/n\Big(\|(X_S^TX_S/n)^{-1}-\Sigma_{SS}^{-1}\|_2+1/C_{\min}\Big).\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathrm{pr}\Big(\bigcup_{S_1\in\mathcal{T}} \mathrm{sig}(\bar{\beta}_S(S_1))^T(X_S^TX_S)^{-1}\mathrm{sig}(\bar{\beta}_S(S_1)) > \frac{s_n}{nC_{\min}}(8s_n^{1/2}n^{-1/2}+1)\Big)\\ &\leq \mathrm{pr}\Big(\|(X_S^TX_S/n)^{-1}-\Sigma_{SS}^{-1}\|_2 \geq \frac{8}{C_{\min}}s_n^{1/2}n^{-1/2}\Big) \leq 2\exp(-s_n/2), \qquad \text{(A20)}\end{aligned}$$

where we have used the concentration inequality of (58b) in Wainwright (2009). Since $\|\varepsilon\|_2^2 \sim \chi^2(n)$, using the inequality of (54a) in Wainwright (2009), we get

$$\begin{aligned}\mathrm{pr}\Big((n\lambda_n)^{-2}\|\varepsilon\|_2^2 > (1+s_n^{1/2}n^{-1/2})/(n\lambda_n^2)\Big) &\leq \mathrm{pr}\Big(\|\varepsilon\|_2^2 \geq (1+s_n^{1/2}n^{-1/2})n\Big)\\ &\leq \exp(-3/16 s_n), \qquad \text{(A21)}\end{aligned}$$

whenever $s_n/n < 1/2$. By the tail probability inequality of Gaussian distribution and (A18),

$$\begin{aligned}\mathrm{pr}(\bigcup_{S_1\in\mathcal{T}} \|K_2(S_1)\|_\infty \geq \gamma/2 \mid H^c) \quad &= \quad \frac{\mathrm{pr}((\bigcup_{S_1\in\mathcal{T}}\|K_2(S_1)\|_\infty \geq \gamma/2)\cap H^c)}{\mathrm{pr}(H^c)}\\ &= \quad \frac{E[\mathrm{pr}(\bigcup_{S_1\in\mathcal{T}}\|K_2(S_1)\|_\infty \geq \gamma/2 \mid X_S,\varepsilon)I(H^c)]}{\mathrm{pr}(H^c)}\\ &\leq \quad \frac{E[2^{s_n+1}(p_n-s_n)\exp(-\gamma^2/(8V))I(H^c)]}{\mathrm{pr}(H^c)}\\ &= \quad 2^{s_n+1}(p_n-s_n)\exp(-\gamma^2/(8V)),\end{aligned}$$

where $V = (1+s_n^{1/2}n^{-1/2})/(n\lambda_n^2) + s_n n^{-1} C_{\min}^{-1}(8s_n^{1/2}n^{-1/2}+1)$ and we used the cardinality of $\mathcal{T}$ is not larger than 2^{s_n}. Under the scaling of Theorem 1, it is easy to verify that

$$\log(p_n - s_n) + (s_n+1)\log 2 = o(\gamma^2/(8V)).$$

Hence, there exists $c_1 > 0$ so that

$$\mathrm{pr}(\bigcup_{S_1\in\mathcal{T}}\{\|K_2(S_1)\|_\infty \geq \gamma/2\} \mid H^c) \leq e^{-c_1 s_n}, \tag{A22}$$

for sufficiently large n. Putting (A19), (A20), (A21), and (A22) together, we proved that there exist positive constants c_2, c_3,

$$\mathrm{pr}\left(\bigcup_{S_1\in\mathcal{T}}\|K_2(S_1)\|_\infty > \frac{\gamma}{2}\right) \leq c_2 e^{-c_3 s_n}. \tag{A23}$$

(A17) and (A23) lead to

$$\mathrm{pr}(\|K_2\|_\infty > \frac{\gamma}{2}) \to 0, \quad \text{as} \quad n\to\infty. \tag{A24}$$

Then, (A16) and (A24) imply

$$\mathrm{pr}(\|K_1+K_2\|_\infty \leq 1-\frac{\gamma}{2}) \geq \mathrm{pr}(\|K_1\|_\infty \leq 1-\gamma) - \mathrm{pr}(\|K_2\|_\infty > \frac{\gamma}{2}) \to 1. \tag{A25}$$

So,

$$\begin{aligned}
\mathrm{pr}(W_1) \geq\ & \mathrm{pr}(A \cap \{\|K_1+K_2\|_\infty \leq 1-\frac{\gamma}{2}\} \text{ and (A7) has a unique solution}) \\
& + \mathrm{pr}(A^c \cap \{\|K_1+K_2\|_\infty \leq 1-\frac{\gamma}{2}\} \text{ and (A7) has a unique solution}) \\
& \overset{(2)}{=} \mathrm{pr}(A \cap \{\|K_1+K_2\|_\infty \leq 1-\frac{\gamma}{2}\}) \\
& + \mathrm{pr}(A^c \cap \{\|K_1+K_2\|_\infty \leq 1-\frac{\gamma}{2}\} \text{ and (A7) has a unique solution}) \\
& \to 1, \quad \text{as} \quad n\to\infty,
\end{aligned} \tag{A26}$$

where (2) is because when A and $\|K_1+K_2\|_\infty \leq 1-\frac{\gamma}{2}$ hold, (A7) always has a unique solution. If there exists another optimal solution to (A7), say β^*. Let $\bar{\beta}(\alpha) = \alpha\bar{\beta} + (1-\alpha)\beta^*, (0<\alpha<1)$.

Convexity of (A7) guarantees $\bar\beta(\alpha)$ is also a solution to (A7). By the optimality conditions and convexity, we have

$$\begin{aligned}\|n^{-1}X_{S^c}^T(Y-X\bar\beta(\alpha))\|_\infty &\le \alpha\|n^{-1}X_{S^c}^T(Y-X\bar\beta)\|_\infty+(1-\alpha)\|n^{-1}X_{S^c}^T(Y-X\beta^*)\|_\infty,\\ &< \alpha\lambda_n+(1-\alpha)\lambda_n=\lambda_n,\end{aligned}$$

where we have used $\|n^{-1}X_{S^c}^T(Y-X\bar\beta)\|_\infty<\lambda_n$ and $\|n^{-1}X_{S^c}^T(Y-X\beta^*)\|_\infty\le\lambda_n$. Therefore, $[\bar\beta(\alpha)]_{S^c}=0$. Then $\bar\beta(\alpha)$ is also a solution to (A8). The uniqueness of (A8) leads to $\bar\beta=\bar\beta(\alpha)$, implying $\bar\beta=\beta^*$. Hence the solution to (A7) is also unique.

Step II. Plugging $Y=X_S\beta_S+\varepsilon$ into (A12), we get,

$$\begin{aligned}\|\beta_S-\bar\beta_S\|_\infty &= \|\lambda_n(X_S^TX_S/n)^{-1}\mathrm{sig}(\bar\beta_S)-(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty\\ &\le \lambda_n\|(X_S^TX_S/n)^{-1}\|_\infty+\|(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty\\ &\le \lambda_n s_n^{1/2}\|(X_S^TX_S/n)^{-1}\|_2+\|(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty\\ &\le \lambda_n s_n^{1/2}(\|(X_S^TX_S/n)^{-1}-\Sigma_{SS}^{-1}\|_2+1/C_{\min})+\|(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty. \qquad \text{(A27)}\end{aligned}$$

Let $G=\Big\{\|(X_S^TX_S)^{-1}\|_2>9/(nC_{\min})\Big\}$, by the inequality (60) in Wainwright (2009), $\mathrm{pr}(G)\le 2\exp(-n/2)$. Since $(X_S^TX_S)^{-1}X_S^T\varepsilon\mid X_S\sim N(0,(X_S^TX_S)^{-1})$, similarly we condition on G to achieve,

$$\begin{aligned}\mathrm{pr}\Big(\|(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty>\frac{s_n^{1/2}}{n^{1/2}C_{\min}^{1/2}}\Big) &\le \mathrm{pr}\Big(\|(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty>\frac{s_n^{1/2}}{n^{1/2}C_{\min}^{1/2}}\mid G^c\Big)\\ &\quad+\mathrm{pr}(G)\\ &\le 2s_ne^{-s_n/18}+2e^{-n/2}\le 2e^{-c_3s_n}, \qquad \text{(A28)}\end{aligned}$$

for some positive c_3. (A20), (A27), and (A28) together imply that,

$$\|\bar\beta_S-\beta_S\|_\infty\le\lambda_n s_n^{1/2}\Big(\frac{8}{C_{\min}}s_n^{1/2}n^{-1/2}+1/C_{\min}\Big)+\frac{s_n^{1/2}}{n^{1/2}C_{\min}^{1/2}}$$

holds with probability larger than $1-2e^{-c_4 s_n}$ for a positive c_4. Under the scaling of Theorem 1 and Condition 10, it is easy to verify that

$$\min_{j\in S}|\beta_j| > \lambda_n s_n^{1/2}\Big(\frac{8}{C_{\min}}s_n^{1/2}n^{-1/2} + 1/C_{\min}\Big) + \frac{s_n^{1/2}}{n^{1/2}C_{\min}^{1/2}}, \tag{A29}$$

for sufficient large n. Thus,

$$\mathrm{pr}(W_2^c) = 1 - \mathrm{pr}(W_2) \le 1 - (1 - 2e^{-c_4 s_n}) \to 0, \quad \text{as} \;\; n \to \infty. \tag{A30}$$

Finally, (A14), (A26) and (A30) together show that,

$$\mathrm{pr}(\check{\beta} \text{ is unique and}, \mathrm{sign}(\check{\beta}) = \mathrm{sign}(\beta)) \to 1, \quad \text{as} \;\; n \to \infty. \quad \square$$

Proof of Theorem 2. Denote the design matrix by X, response vector by Y, and error vector by ε. Let $S = \{1 \le j \le p : \beta_j \ne 0\}, N = \{1 \le j \le p : \beta_j = 0\}$. Define the decomposition $S = \hat{S}_1 \cup \hat{S}_2, N = \hat{N}_1 \cup \hat{N}_2$, where $\hat{S}_2$ and $\hat{N}_2$ form the retention set.

Firstly, consider the second step,

$$\check{\beta} = \arg\min_{\beta}\left\{\frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda_n(\|\beta_{\hat{S}_1}\|_1 + \|\beta_{\hat{N}_1}\|_1)\right\}. \tag{A31}$$

We are going to show that with high probability,

$$\check{\beta}_{\hat{S}_1} \ne 0 \;\; \text{and} \;\; \check{\beta}_{\hat{N}_1} = 0. \tag{A32}$$

Define an oracle estimator of (A31),

$$\bar{\beta} = \arg\min_{\beta_{\hat{N}_1}=0}\left\{\frac{1}{2n}\|Y - X_{\hat{Q}}\beta_{\hat{Q}}\|_2^2 + \lambda_n\|\beta_{\hat{S}_1}\|_1\right\}. \tag{A33}$$

where $\hat{Q} = S \cup \hat{N}_2$. Similar as in Theorem 1, to show $\check{\beta}_{\hat{N}_1} = 0$, it is sufficient to prove,

$$\begin{aligned}&\|X_{\hat{Q}^c}^T X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1}\mathrm{sig}(\bar{\beta}_{\hat{Q}}) + (n\lambda_n)^{-1}X_{\hat{Q}^c}^T(I - X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1}X_{\hat{Q}}^T)(X_S\beta_S + \varepsilon)\|_\infty \\ &< 1,\end{aligned} \tag{A34}$$

and (A31) has a unique solution. Since $(I - X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1} X_{\hat{Q}}^T) X_{\hat{Q}} = 0$, (A34) can be simplified as

$$\|X_{\hat{Q}^c}^T X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1}\mathrm{sig}(\bar{\beta}_{\hat{Q}}) + (n\lambda_n)^{-1} X_{\hat{Q}^c}^T (I - X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1} X_{\hat{Q}}^T)\varepsilon\|_\infty < 1. \tag{A35}$$

Let

$$F = X_{\hat{Q}^c}^T - \Sigma_{\hat{Q}^c\hat{Q}}\Sigma_{\hat{Q}\hat{Q}}^{-1} X_{\hat{Q}}^T,$$

$$K_1 = \Sigma_{\hat{Q}^c\hat{Q}}\Sigma_{\hat{Q}\hat{Q}}^{-1}\mathrm{sig}(\bar{\beta}_{\hat{Q}}),$$

$$K_2 = F X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1}\mathrm{sig}(\bar{\beta}_{\hat{Q}}) + (n\lambda_n)^{-1} F\{I - X_{\hat{Q}}(X_{\hat{Q}}^T X_{\hat{Q}})^{-1} X_{\hat{Q}}^T\}\varepsilon.$$

Then, (A35) is equivalent to

$$\|K_1 + K_2\|_\infty < 1.$$

Different from the proof in Theorem 1, the subset $\hat{Q}$ is random now. To this end, introduce

$$\begin{aligned} A &= \{R \subset \hat{S}_2 \subset S, S \subset \hat{Q} \subset S \cup Z\}, \\ B &= \{S \subset \hat{Q} \subset S \cup Z\}, \\ C &= \{\hat{N}_2 \subset Z\}. \end{aligned}$$

From Proposition 2, it is not hard to show $P(A) \to 1$, under the scaling in Theorem 2. Note that $\mathrm{sig}(\bar{\beta}_{\hat{Q}})$ only has $\hat{S}_1$ non-zero entries, hence

$$\begin{aligned} \mathrm{pr}(\|K_1\|_\infty \le 1-\gamma) &\ge \mathrm{pr}(\{\|K_1\|_\infty \le 1-\gamma\} \cap A) \\ &\overset{(a)}{=} \mathrm{pr}(A), \end{aligned} \tag{A36}$$

where (a) holds because A and Condition 12 imply $\|K_1\|_\infty \le 1-\gamma$. To bound $\|K_2 + K_3\|_\infty$, let

$$H = \bigcup_{\substack{(Q,S_2)\\ S\subset Q\subset S\cup Z\\ R\subset S_2\subset S}} \Big\{\mathrm{sig}(\bar{\beta}_Q)^T (X_Q^T X_Q)^{-1}\mathrm{sig}(\bar{\beta}_Q) + (n\lambda_n)^{-2}\|\varepsilon\|_2^2 > \frac{s_n+z_n}{nC_{\min}}(8(s_n+z_n)^{1/2} n^{-1/2} + 1) + (1 + s_n^{1/2} n^{-1/2})/(n\lambda_n^2)\Big\}.$$

Note that $\bar{\beta}_Q$ is the analogy of $\bar{\beta}_{\hat{Q}}$ by replacing $\hat{Q}$ and $\hat{S_2}$ in (A33) with Q and S_2. Then,

$$\begin{aligned}\mathrm{pr}(\|K_2\|_\infty > \frac{\gamma}{2}) &\le \mathrm{pr}(\{\|K_2\|_\infty > \frac{\gamma}{2}\} \cap A) + \mathrm{pr}(A^c)\\ &\le \mathrm{pr}\Big(\Big\{ \bigcup_{\substack{(Q,S_2)\\ S\subset Q\subset S\cup Z\\ R\subset S_2\subset S}} \|K_2(Q,S_2)\|_\infty > \frac{\gamma}{2}\Big\} \cap A) + \mathrm{pr}(A^c)\\ &\le \mathrm{pr}(\bigcup_{\substack{(Q,S_2)\\ S\subset Q\subset S\cup Z\\ R\subset S_2\subset S}} \|K_2(Q,S_2)\|_\infty > \frac{\gamma}{2} \mid H^c) + \mathrm{pr}(H) + \mathrm{pr}(A^c). \qquad (A37)\end{aligned}$$

$\mathrm{pr}(H)$ can be bounded in the same way as in Theorem 1,

$$\begin{aligned}\mathrm{pr}(H) \le \mathrm{pr}(&\bigcup_{\substack{(Q,S_2)\\ S\subset Q\subset S\cup Z\\ R\subset S_2\subset S}} \Big\{\mathrm{sig}(\bar{\beta}_Q)^T (X_Q^T X_Q)^{-1} \mathrm{sig}(\bar{\beta}_Q) > \frac{s_n + z_n}{nC_{\min}} (8(s_n+z_n)^{1/2} n^{-1/2} + 1)\Big\})\\ &+ \mathrm{pr}((n\lambda_n)^{-2}\|\varepsilon\|_2^2 > (1 + s_n^{1/2} n^{-1/2})/(n\lambda_n^2))\\ &\le \mathrm{pr}(\bigcup_{S\subset Q\subset S\cup Z} \Big\{\|(X_Q^T X_Q/n)^{-1} - \Sigma_{QQ}^{-1}\|_2 \ge \frac{8}{C_{\min}} (s_n+z_n)^{1/2} n^{-1/2}\Big\}) + e^{-\frac{3}{16}s_n}\\ &\le \mathrm{pr}(\bigcup_{S\subset Q\subset S\cup Z} \Big\{\|(X_Q^T X_Q/n)^{-1} - \Sigma_{QQ}^{-1}\|_2 \ge \frac{8}{C_{\min}} (\mathrm{Card}(Q))^{1/2} n^{-1/2}\Big\}) + e^{-\frac{3}{16}s_n}\\ &\le 2^{z_n+1} \exp(-\frac{s_n}{2}) + \exp(-\frac{3}{16}s_n). \qquad (A38)\end{aligned}$$

We use similar arguments as in Theorem 1 for bounding the following,

$$\mathrm{pr}(\bigcup_{\substack{(Q,S_2)\\ S\subset Q\subset S\cup Z\\ R\subset S_2\subset S}} \|K_2(Q,S_2)\|_\infty > \frac{\gamma}{2} \mid H^c) \le 2^{s_n+1+z_n} (p_n - s_n) \exp(-\gamma^2/8V), \qquad (A39)$$

where $V = \frac{s_n+z_n}{nC_{\min}} (8(s_n+z_n)^{1/2} n^{-1/2} + 1) + (1 + s_n^{1/2} n^{-1/2})/(n\lambda_n^2)$.

Under the scaling in Theorem 2, (A36)(A37)(A38)(A39) show that (A34) holds with high probability. The uniqueness of (A31) can be proved by the same arguments as in Theorem 1.

We skip the proof here for simplicity. Next, we bound $\|\bar{\beta}_{\hat{Q}} - \beta_{\hat{Q}}\|_\infty$.

$$\begin{aligned}\|\bar{\beta}_{\hat{Q}} - \beta_{\hat{Q}}\|_\infty &= \|(X_{\hat{Q}}^T X_{\hat{Q}})^{-1}(X_{\hat{Q}}^T Y - n\lambda_n \mathrm{sig}(\bar{\beta}_{\hat{Q}})) - \beta_{\hat{Q}}\|_\infty \\ &\overset{(b)}{=} \|(X_{\hat{Q}}^T X_{\hat{Q}})^{-1} X_{\hat{Q}}^T \varepsilon - \lambda_n (X_{\hat{Q}}^T X_{\hat{Q}}/n)^{-1} \mathrm{sig}(\bar{\beta}_{\hat{Q}})\|_\infty \\ &\leq \|(X_{\hat{Q}}^T X_{\hat{Q}})^{-1} X_{\hat{Q}}^T \varepsilon\|_\infty + \|\lambda_n (X_{\hat{Q}}^T X_{\hat{Q}}/n)^{-1}\|_\infty,\end{aligned}$$

where (b) holds because $(X_{\hat{Q}}^T X_{\hat{Q}})^{-1} X_{\hat{Q}}^T X_S \beta_S - \beta_{\hat{Q}} = 0$. Let $U_n = \lambda_n (s_n + z_n)^{1/2} (\frac{8}{C_{\min}} (s_n + z_n)^{1/2} n^{-\frac{1}{2}} + \frac{1}{C_{\min}}) + \frac{(s_n+z_n)^{1/2}}{n^{1/2} C_{\min}^{1/2}}$. Then,

$$\begin{aligned}\mathrm{pr}(\|\bar{\beta}_{\hat{Q}} - \beta_{\hat{Q}}\|_\infty \geq U_n) &\leq \mathrm{pr}(\{\|\bar{\beta}_{\hat{Q}} - \beta_{\hat{Q}}\|_\infty \geq U_n\} \cap B) + \mathrm{pr}(B^c) \\ &\leq \mathrm{pr}(\bigcup_{S \subset Q \subset S \cup Z} \|(X_Q^T X_Q)^{-1} X_Q^T \varepsilon\|_\infty + \|\lambda_n (X_Q^T X_Q/n)^{-1}\|_\infty \geq U_n) \\ &\quad + \mathrm{pr}(B^c) \\ &\overset{(c)}{\leq} 2^{z_n} (2 s_n e^{-\frac{s_n}{18}} + 2e^{-n/2} + 2e^{-\frac{s_n}{2}}) + \mathrm{pr}(B^c),\end{aligned}$$

where (c) follows from the bounds (A27) and (A28) in the proof of Theorem 1. By Condition 10, it is not hard to verify $\min_{j \in S} |\beta_j| \gg U_n$. Thus,

$$\mathrm{pr}(\min_{j \in S} |\beta_j| > \|\bar{\beta}_{\hat{S}_1} - \beta_{\hat{S}_1}\|_\infty) \geq \mathrm{pr}(B) - 2^{z_n} (2 s_n e^{-\frac{s_n}{18}} + 2e^{-n/2} + 2e^{-\frac{s_n}{2}}), \tag{A40}$$

Since $P(B) \geq P(A) \to 1$ and $2^{z_n} (2 s_n e^{-\frac{s_n}{18}} + 2e^{-n/2} + 2e^{-\frac{s_n}{2}}) \to 0$ under the scaling in Theorem 2, (A40) implies that $\check{\beta}_{\hat{S}_1} \neq 0$ with high probability.

Let us now consider the third step. We have shown that, with high probability, the third step takes the form

$$\tilde{\beta} = \underset{\beta_{\hat{N}_1}=0}{\arg\min} \left\{ \frac{1}{2n} \|Y - X_{\hat{Q}} \beta_{\hat{Q}}\|_2^2 + \lambda_n^* (\|\beta_{\hat{S}_2}\|_1 + \|\beta_{\hat{N}_2}\|_1) \right\}. \tag{A41}$$

To prove $\mathrm{sign}(\tilde{\beta}) = \mathrm{sign}(\beta)$, it remains to show $\mathrm{sign}(\tilde{\beta}_S) = \mathrm{sign}(\beta_S)$ and $\tilde{\beta}_{\hat{N}_2} = 0$. We use similar arguments as in the second step. Define the oracle estimator of (A41),

$$\mathring{\beta} = \underset{\beta_{\hat{N}_1}=0,\ \beta_{\hat{N}_2}=0}{\arg\min} \left\{ \frac{1}{2n} \|Y - X_S \beta_S\|_2^2 + \lambda_n^* \|\beta_{\hat{S}_2}\|_1 \right\}.$$

Let

$$\tilde{F} = X_{\hat{N}_2}^T - \Sigma_{\hat{N}_2 S}\Sigma_{SS}^{-1}X_S^T,$$

$$\tilde{K}_1 = \Sigma_{\hat{N}_2 S}\Sigma_{SS}^{-1}\mathrm{sig}(\mathring{\beta}_S),$$

$$\tilde{K}_2 = \tilde{F}X_S(X_S^TX_S)^{-1}\mathrm{sig}(\mathring{\beta}_S) + (n\lambda_n^*)^{-1}\tilde{F}\{I - X_S(X_S^TX_S)^{-1}X_S^T\}\varepsilon.$$

Then,

$$\begin{aligned} P(\|\tilde{K}_1\|_\infty \leq 1-\alpha) &\geq P(\{\|\tilde{K}_1\|_\infty \leq 1-\alpha\} \cap C) \\ &\overset{(d)}{\geq} P(C), \end{aligned} \tag{A42}$$

where (d) holds because C and Condition 13 imply $\|\tilde{K}_1\|_\infty \leq 1-\alpha$. Let

$$\begin{aligned} \tilde{H} = \bigcup_{R\subset S_2 \subset S} & \Big\{\mathrm{sig}(\mathring{\beta}_S)^T(X_S^TX_S)^{-1}\mathrm{sig}(\mathring{\beta}_S) + (n\lambda_n^*)^{-2}\|\varepsilon\|_2^2 > \frac{s_n}{nC_{\min}}(8s_n^{1/2}n^{-1/2}+1) \\ & + (1+s_n^{1/2}n^{-1/2})/(n(\lambda_n^*)^2)\Big\}. \end{aligned}$$

Then,

$$\begin{aligned} P(\|\tilde{K}_2\|_\infty > \frac{\alpha}{2}) &\leq P(\{\|\tilde{K}_2\|_\infty > \frac{\alpha}{2}\} \cap A) + P(A^c) \\ &\leq P(\bigcup_{\substack{(N_2,S_2)\\ N_2\subset Z\\ R\subset S_2\subset S}} \Big\{\|\tilde{K}_2(N_2,S_2)\|_\infty > \frac{\alpha}{2}\Big\}) + P(A^c) \\ &\leq P(\bigcup_{\substack{(N_2,S_2)\\ N_2\subset Z\\ R\subset S_2\subset S}} \Big\{\|\tilde{K}_2(N_2,S_2)\|_\infty > \frac{\alpha}{2}\Big\} \mid \tilde{H}^c) + P(\tilde{H}) + P(A^c) \\ &\overset{(e)}{\leq} 2^{z_n+s_n+1}z_n e^{-\alpha^2/8\tilde{V}} + 2e^{-\frac{s_n}{2}} + e^{-\frac{3}{16}s_n} + P(A^c), \end{aligned} \tag{A43}$$

where (e) follows from (A20) and (A21) in the proof of Theorem 1 and $\tilde{V} = \frac{s_n}{nC_{\min}}(8s_n^{1/2}n^{-1/2} + 1) + (1+s_n^{1/2}n^{-1/2})/(n(\lambda_n^*)^2)$. Again, we skip the proof of uniqueness of (A41). Now we have shown that $\tilde{\beta}_{\hat{N}_2} = 0$ with high probability. The final step is to bound $\|\mathring{\beta}_S - \beta_S\|_\infty$.

$$\begin{aligned} \|\mathring{\beta}_S - \beta_S\|_\infty &= \|(X_S^TX_S)^{-1}(X_S^TY - n\lambda_n^*\mathrm{sig}(\mathring{\beta}_S)) - \beta_S\|_\infty \\ &\leq \|(X_S^TX_S)^{-1}X_S^T\varepsilon\|_\infty + \|\lambda_n^*(X_S^TX_S/n)^{-1}\|_\infty. \end{aligned}$$

Let $W_n = \lambda_n^* s_n^{1/2}(\frac{8}{C_{\min}} s_n^{1/2} n^{-\frac{1}{2}} + \frac{1}{C_{\min}}) + \frac{s_n^{1/2}}{n^{1/2} C_{\min}^{1/2}}$. By (A27) and (A28) in the proof of Theorem 1, we have $P(\|\mathring{\beta}_S - \beta_S\|_\infty \le W_n) \ge 1 - 2\exp(-c_2 s_n)$ for a positive c_2. Since $U_n \asymp W_n$,

$$P(\min_{j \in S} |\beta_j| > \|\mathring{\beta}_S - \beta_S\|_\infty) \ge 1 - 2\exp(-c_2 s_n), \tag{A44}$$

for sufficiently large n. Putting (A42)(A43)(A44) together, we have shown

$$P(\tilde{\beta} \text{ is unique and } \mathrm{sign}(\tilde{\beta}) = \mathrm{sign}(\beta)) \to 1, \quad \text{as } n \to \infty.$$

□

2 Additional simulation results

To make the simulation setting more challenging, we have investigated the following scenarios where the number of signals is increased to 20 while the rest of the settings are similar to Scenarios 1 and 2. We choose the same scaling between n and p_n: $p_n = \lfloor 100\exp(n^{0.2}) \rfloor$.

Scenario 3. The covariance matrix Σ is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & I \end{pmatrix}, \text{ where } \Sigma_{11} = (1-r)I + rJ \in \mathbb{R}^{(s_n+10)\times(s_n+10)},$$

in which I is the identity matrix and J is the matrix of all 1s.

(A). $r = 0.6, \sigma = 3.5, s_n = 20, \beta_S = (3, -2, 2, -2, 2, \cdots, -2, 2, -2)^T, \beta = (\beta_S^T, 0_{p-20}^T)^T$.

(B). $r = 0.6, \sigma = 1.2, s_n = 20, \beta_S = (1, 1, -1, 1, -1, \cdots, 1, -1, 1)^T, \beta = (\beta_S^T, 0_{p-20}^T)^T$.

Scenario 4. The covariance matrix Σ is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & I \end{pmatrix}, \text{ where } \Sigma_{11} = \begin{pmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega \end{pmatrix}, \Omega = \begin{pmatrix} 1 & r_0 & r_1 & r_3 \\ r_0 & 1 & r_2 & r_4 \\ r_1 & r_2 & 1 & 0 \\ r_3 & r_4 & 0 & 1 \end{pmatrix}$$

(C). $\Sigma_{11} \in \mathbb{R}^{40\times 40}$ is block-diagonal, $r_0 = 0.8, r_1 = -r_2 = r_3 = -r_4 = -0.1, \sigma = 2.5, s_n = 20, \beta_{11} = (2.5, -2, 0, 0, 2.5, -2, 0, 0, \cdots, 2.5, -2, 0, 0)^T, \beta = (\beta_{11}^T, 0_{p-40}^T)^T$.

(D). $\Sigma_{11} \in \mathbb{R}^{40\times 40}$ is block-diagonal, $r_0 = 0.75, r_1 = r_2 = r_3 = -r_4 = 0.2, \sigma = 2.5, s_n = 20, \beta_{11} = (2.5, -2, 0, 0, 2.5, -2, 0, 0, \cdots, 2.5, -2, 0, 0)^T, \beta = (\beta_{11}^T, 0_{p-40}^T)^T$.

From Tables 1 and 2, the results again demonstrate the superior performance of our proposed methods. In particular, the advantage of RAR+(MC+) is more significant compared to Scenarios 1 and 2. One possible reason is that since we have more signals, the retention step is able to keep more marginally important signals, leading to the easier discovery of additional signals compared with the other methods. Since the main message is similar to that in the two preceding scenarios, we skip the detailed comparisons.

Finally, in Tables 3-10, we present the relative estimation error as well as the model size over 200 simulation rounds for all the simulation examples.

References

Wainwright, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery. Information Theory, *IEEE Transactions on* **55**, 2183-2202.

Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with np-dimensionality. *The Annals of Statistics* **38**, 3567-3604.

Table 1: Sign recovery proportion over 200 simulation rounds.

Scenario 3 (A)	(350, 2520)	(450, 2976)	(550, 3420)	(650, 3856)	(750, 4288)
Lasso	0.000	0.000	0.000	0.000	0.000
SCAD	0.000	0.000	0.075	0.285	0.685
MC+	0.005	0.160	0.625	0.845	0.975
SIS-lasso	0.000	0.000	0.000	0.000	0.000
ISIS-lasso	0.000	0.000	0.000	0.000	0.000
Ada-lasso	0.000	0.000	0.000	0.000	0.000
SIS-MC+	0.000	0.000	0.000	0.000	0.000
ISIS-MC+	0.005	0.215	0.575	0.730	0.930
SC-lasso	0.000	0.000	0.000	0.000	0.000
SC-forward	0.000	0.000	0.000	0.000	0.000
SC-marginal	0.000	0.000	0.000	0.000	0.000
RAR_1	0.020	0.070	0.135	0.195	0.255
RAR_5	0.025	0.075	0.215	0.305	0.455
RAR_{30}	0.000	0.060	0.135	0.325	0.505
RAR(MC+)_{30}	0.010	0.350	0.765	0.830	0.895
RAR+_1	**0.055**	0.245	0.505	0.755	0.820
RAR+_5	0.030	0.135	0.335	0.570	0.720
RAR+_{30}	0.000	0.080	0.165	0.405	0.585
RAR+(MC+)_{30}	0.015	**0.365**	**0.805**	**0.915**	**0.990**

Scenario 3 (B)	(150, 1524)	(250, 2043)	(350, 2520)	(450, 2976)	(550, 3420)
Lasso	0.000	0.000	0.000	0.000	0.000
SCAD	0.000	0.015	0.425	0.890	0.980
MC+	0.000	0.295	0.940	0.995	**1.000**
SIS-lasso	0.000	0.000	0.000	0.000	0.000
ISIS-lasso	**0.015**	0.000	0.000	0.000	0.000
Ada-lasso	0.000	0.065	0.540	0.895	0.985
SIS-MC+	0.000	0.490	0.850	0.965	0.995
ISIS-MC+	0.000	0.375	0.895	0.995	**1.000**
SC-lasso	0.000	0.000	0.000	0.000	0.000
SC-forward	0.000	0.000	0.000	0.000	0.000
SC-marginal	0.000	0.000	0.045	0.180	0.505
RAR_1	0.000	0.000	0.000	0.000	0.000
RAR_5	0.000	0.000	0.000	0.000	0.000
RAR_{30}	0.000	0.000	0.000	0.000	0.000
RAR(MC+)_{30}	0.000	0.000	0.000	0.000	0.000
RAR+_1	0.005	0.215	0.605	0.895	0.955
RAR+_5	0.005	0.210	0.605	0.895	0.950
RAR+_{30}	0.005	0.215	0.605	0.895	0.955
RAR+(MC+)_{30}	0.000	**0.750**	**0.990**	**1.000**	**1.000**

Table 2: Sign recovery proportion over 200 simulation rounds.

Scenario 4 (C)	(300, 2285)	(400, 2750)	(500, 3199)	(600, 3639)	(700, 4073)
Lasso	0.000	0.000	0.000	0.000	0.000
SCAD	0.000	0.005	0.075	0.285	0.605
MC+	0.000	0.020	0.250	0.700	0.885
SIS-lasso	0.000	0.000	0.000	0.000	0.000
ISIS-lasso	0.000	0.000	0.000	0.000	0.000
Ada-lasso	0.000	0.000	0.000	0.000	0.000
SIS-MC+	0.000	0.000	0.000	0.000	0.000
ISIS-MC+	0.000	0.055	0.255	0.640	0.855
SC-lasso	0.000	0.000	0.000	0.000	0.000
SC-forward	0.000	0.000	0.000	0.000	0.000
SC-marginal	0.000	0.000	0.000	0.000	0.000
RAR_1	0.000	0.000	0.000	0.000	0.005
RAR_5	0.000	0.000	0.000	0.000	0.000
RAR_{30}	0.000	0.000	0.000	0.000	0.000
RAR(MC+)_{30}	**0.010**	0.505	0.820	0.860	0.830
RAR+_1	0.000	0.000	0.000	0.005	0.020
RAR+_5	0.000	0.000	0.000	0.000	0.000
RAR+_{30}	0.000	0.000	0.000	0.000	0.000
RAR+(MC+)_{30}	**0.010**	**0.530**	**0.915**	**1.000**	**1.000**

Scenario 4 (D)	(300, 2285)	(400, 2750)	(500, 3199)	(600, 3639)	(700, 4073)
Lasso	0.000	0.000	0.000	0.000	0.000
SCAD	0.005	0.025	0.090	0.310	0.605
MC+	0.005	0.070	0.265	0.635	0.900
SIS-lasso	0.000	0.000	0.000	0.000	0.000
ISIS-lasso	0.000	0.000	0.000	0.000	0.000
Ada-lasso	0.000	0.000	0.000	0.000	0.000
SIS-MC+	0.000	0.000	0.000	0.000	0.000
ISIS-MC+	**0.030**	0.170	0.435	0.650	0.845
SC-lasso	0.000	0.000	0.000	0.000	0.000
SC-forward	0.000	0.000	0.000	0.000	0.030
SC-marginal	0.000	0.000	0.000	0.000	0.000
RAR_1	0.000	0.000	0.000	0.000	0.000
RAR_5	0.000	0.000	0.000	0.000	0.000
RAR_{30}	0.000	0.000	0.000	0.000	0.000
RAR(MC+)_{30}	0.020	0.165	0.220	0.100	0.045
RAR+_1	0.000	0.000	0.000	0.000	0.015
RAR+_5	0.000	0.000	0.000	0.000	0.000
RAR+_{30}	0.000	0.000	0.000	0.000	0.000
RAR+(MC+)_{30}	**0.030**	**0.350**	**0.765**	**0.930**	**0.995**

Bach, F., Jenatton, R., Mairal, J. and Obozinski, G. (2012). Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.* **4(1)**, 1-106.

Department of Statistics, Columbia University, New York, NY 10027, U.S.A.

E-mail: hw2375@columbia.edu, yang.feng@columbia.edu

Department of Mathematical Sciences, Binghamton University, State University of New York, Binghamton, NY 13902, U.S.A.

E-mail: qiao@math.binghamton.edu

Table 3: Relative estimation error over 200 simulation rounds.

Scenario 1 (A)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	0.81 (0.12)	0.52 (0.13)	0.33 (0.10)	0.26 (0.07)	0.20 (0.05)
SCAD	0.56 (0.24)	0.12 (0.13)	0.03 (0.04)	0.01 (0.01)	0.01 (0.01)
MC+	0.54 (0.24)	0.10 (0.12)	0.02 (0.03)	0.01 (0.01)	0.01 (0.01)
SIS-lasso	0.85 (0.09)	0.75 (0.10)	0.71 (0.10)	0.67 (0.14)	0.66 (0.12)
ISIS-lasso	0.68 (0.17)	0.46 (0.11)	0.33 (0.09)	0.26 (0.07)	0.21 (0.05)
Ada-lasso	0.80 (0.11)	0.62 (0.13)	0.51 (0.16)	0.46 (0.17)	0.37 (0.16)
SIS-MC+	0.80 (0.12)	0.67 (0.14)	0.63 (0.12)	0.59 (0.17)	0.58 (0.13)
ISIS-MC+	0.52 (0.27)	0.11 (0.13)	0.03 (0.05)	0.01 (0.01)	0.01 (0.01)
SC-lasso	0.89 (0.18)	0.74 (0.16)	0.62 (0.14)	0.53 (0.19)	0.40 (0.23)
SC-forward	0.92 (0.16)	0.72 (0.22)	0.52 (0.19)	0.39 (0.18)	0.25 (0.19)
SC-marginal	0.90 (0.17)	0.75 (0.17)	0.66 (0.11)	0.64 (0.10)	0.63 (0.08)
RAR_1	0.67 (0.28)	0.28 (0.14)	0.15 (0.07)	0.11 (0.05)	0.09 (0.04)
RAR_5	0.70 (0.24)	0.28 (0.15)	0.13 (0.07)	0.10 (0.05)	0.07 (0.03)
RAR_{30}	0.74 (0.21)	0.31(0.16)	0.14 (0.08)	0.09 (0.05)	0.07 (0.03)
$RAR(MC+)_{30}$	0.59 (0.26)	0.16 (0.17)	0.02 (0.05)	0.01 (0.01)	0.01 (0.01)
$RAR+_1$	0.58 (0.27)	0.15 (0.15)	0.04 (0.05)	0.02 (0.02)	0.02 (0.02)
$RAR+_5$	0.66 (0.25)	0.17 (0.17)	0.04 (0.06)	0.02 (0.03)	0.01 (0.02)
$RAR+_{30}$	0.73 (0.23)	0.23 (0.19)	0.05 (0.08)	0.02 (0.03)	0.01 (0.02)
$RAR+(MC+)_{30}$	0.60 (0.26)	0.18 (0.17)	0.03 (0.05)	0.01 (0.01)	0.01 (0.01)

Scenario 1 (B)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	0.72 (0.09)	0.34 (0.10)	0.21 (0.05)	0.15 (0.03)	0.12 (0.03)
SCAD	0.43 (0.28)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
MC+	0.42 (0.28)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
SIS-lasso	0.74 (0.11)	0.64 (0.16)	0.54 (0.21)	0.46 (0.22)	0.39 (0.24)
ISIS-lasso	0.38 (0.15)	0.25 (0.07)	0.18 (0.04)	0.13 (0.03)	0.11 (0.03)
Ada-lasso	0.68 (0.13)	0.41 (0.18)	0.24 (0.17)	0.14 (0.11)	0.11 (0.11)
SIS-MC+	0.74 (0.14)	0.60 (0.23)	0.45 (0.26)	0.35 (0.25)	0.29 (0.25)
ISIS-MC+	0.23 (0.25)	0.02 (0.02)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
SC-lasso	0.84 (0.11)	0.72 (0.12)	0.57 (0.19)	0.28 (0.24)	0.09 (0.16)
SC-forward	0.86 (0.10)	0.76 (0.13)	0.53 (0.28)	0.18 (0.27)	0.05 (0.16)
SC-marginal	0.84 (0.14)	0.69 (0.15)	0.60 (0.16)	0.48 (0.17)	0.47 (0.15)
RAR_1	0.33 (0.16)	0.15 (0.06)	0.10 (0.03)	0.08 (0.03)	0.06 (0.02)
RAR_5	0.32 (0.19)	0.12 (0.05)	0.09 (0.03)	0.07 (0.02)	0.06 (0.02)
RAR_{30}	0.35 (0.21)	0.11 (0.04)	0.09 (0.03)	0.07 (0.02)	0.06 (0.02)
$RAR(MC+)_{30}$	0.27 (0.24)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)
$RAR+_1$	0.19 (0.17)	0.03 (0.03)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)
$RAR+_5$	0.20 (0.22)	0.03 (0.02)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)
$RAR+_{30}$	0.25 (0.24)	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
$RAR+(MC+)_{30}$	0.27 (0.24)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)

Table 4: Model size over 200 simulation rounds.

Scenario 1 (A)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	28.73 (20.96)	87.81 (26.23)	108.62 (21.10)	121.46 (25.03)	125.72 (22.76)
SCAD	27.38 (14.87)	45.41 (25.34)	25.31 (20.63)	12.51 (9.96)	7.22 (5.22)
MC+	9.57 (10.22)	14.77 (14.60)	7.03 (7.51)	4.64 (1.94)	4.56 (1.50)
SIS-lasso	11.56 (4.71)	19.15 (5.91)	23.52 (7.45)	27.87 (9.92)	32.48 (11.56)
ISIS-lasso	18.25 (5.07)	33.73 (3.28)	43.80 (3.75)	51.76 (4.57)	57.46 (5.17)
Ada-lasso	11.12 (13.18)	30.29 (29.73)	44.99 (36.32)	64.40 (44.09)	77.93 (41.74)
SIS-MC+	7.40 (3.60)	10.54 (4.38)	11.84 (5.04)	12.59 (7.82)	19.10 (14.65)
ISIS-MC+	10.71 (5.29)	13.82 (5.51)	9.28 (4.24)	7.19 (3.40)	5.36 (2.02)
SC-lasso	0.37 (0.54)	0.86 (0.50)	1.27 (0.68)	1.66 (0.95)	2.23 (1.12)
SC-forward	0.25 (0.46)	0.98 (0.72)	1.68 (0.72)	2.22 (0.84)	2.85 (1.01)
SC-marginal	0.30 (0.49)	0.83 (0.53)	1.16 (0.53)	1.28 (0.62)	1.39 (0.64)
RAR_1	30.14 (18.65)	50.58 (20.07)	51.85 (17.33)	55.92 (18.30)	60.02 (17.05)
RAR_5	27.83 (19.15)	50.64 (22.76)	49.67 (18.15)	52.20 (16.60)	53.75 (14.81)
RAR_{30}	28.09 (20.12)	55.59 (25.85)	52.51 (22.75)	51.70 (18.69)	51.48 (15.19)
RAR(MC+)_{30}	2.74 (2.04)	4.76 (2.43)	4.59 (1.38)	4.36 (0.79)	4.44 (0.87)
RAR+_1	10.80 (14.76)	7.49 (12.95)	5.35 (2.21)	5.33 (1.60)	5.66 (1.70)
RAR+_5	16.63 (20.47)	9.91 (20.74)	4.66 (2.02)	4.65 (1.62)	4.55 (0.86)
RAR+_{30}	21.85 (21.81)	16.03 (29.19)	4.88 (3.64)	4.29 (1.47)	4.27 (1.41)
RAR+(MC+)_{30}	2.60 (1.98)	4.41 (2.32)	4.39 (1.34)	4.05 (0.25)	4.01 (0.10)

Scenario 1 (B)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	40.46 (25.68)	110.42 (24.17)	122.44 (24.25)	135.25 (24.42)	142.64 (23.38)
SCAD	34.10 (16.44)	13.69 (9.31)	6.46 (2.36)	5.87 (3.06)	5.76 (2.86)
MC+	19.21 (15.90)	6.49 (3.17)	5.55 (2.60)	5.53 (2.50)	5.60 (2.54)
SIS-lasso	13.01 (4.42)	20.97 (8.35)	29.77 (12.60)	38.18 (13.56)	45.05 (14.77)
ISIS-lasso	20.74 (1.71)	35.29 (1.82)	45.94 (3.51)	54.19 (5.77)	62.11 (8.08)
Ada-lasso	21.79 (23.26)	57.25 (30.35)	59.42 (29.48)	61.77 (34.07)	59.09 (36.07)
SIS-MC+	8.57 (5.74)	15.34 (8.13)	17.82 (10.67)	16.11 (12.07)	13.94 (11.68)
ISIS-MC+	11.58 (4.22)	7.45 (2.31)	5.62 (0.97)	5.29 (0.74)	5.27 (0.64)
SC-lasso	0.83 (0.53)	1.44 (0.58)	2.21 (0.99)	3.68 (1.22)	4.57 (0.82)
SC-forward	0.74 (0.47)	1.24 (0.64)	2.37 (1.39)	4.14 (1.36)	4.79 (0.79)
SC-marginal	0.86 (0.69)	1.71 (0.92)	2.42 (1.00)	3.04 (0.87)	3.13 (0.81)
RAR_1	46.95 (19.42)	62.04 (20.03)	67.21 (20.51)	73.14 (21.45)	74.66 (20.69)
RAR_5	42.80 (17.76)	56.16 (16.84)	63.54 (18.78)	72.28 (17.98)	72.59 (18.31)
RAR_{30}	42.34 (17.36)	53.01 (15.53)	61.37 (18.18)	71.25 (17.21)	73.15 (15.75)
RAR(MC+)_{30}	5.14 (2.12)	7.53 (2.09)	9.44 (1.56)	10.23 (1.08)	10.34 (0.77)
RAR+_1	11.21 (10.08)	8.84 (4.78)	8.48 (1.69)	8.54 (1.64)	8.40 (1.73)
RAR+_5	10.48 (11.58)	7.28 (1.77)	7.73 (1.36)	7.86 (1.24)	7.79 (1.33)
RAR+_{30}	11.61 (14.60)	6.66 (1.80)	7.45 (1.31)	7.72 (1.21)	7.61 (1.21)
RAR+(MC+)_{30}	4.46 (1.69)	5.09 (0.52)	5.03 (0.17)	5.03 (0.16)	5.02 (0.14)

Table 5: Relative estimation error over 200 simulation rounds.

Scenario 2 (C)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	0.91 (0.09)	0.79 (0.14)	0.62 (0.15)	0.48 (0.14)	0.38 (0.10)
SCAD	0.70 (0.36)	0.16 (0.31)	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)
MC+	0.72 (0.33)	0.16 (0.31)	0.01 (0.02)	0.01 (0.01)	0.00 (0.00)
SIS-lasso	0.91 (0.10)	0.84 (0.10)	0.80 (0.08)	0.78 (0.06)	0.75 (0.10)
ISIS-lasso	0.79 (0.18)	0.64 (0.16)	0.56 (0.12)	0.46 (0.12)	0.38 (0.09)
Ada-lasso	0.89 (0.11)	0.79 (0.12)	0.72 (0.11)	0.67 (0.13)	0.61 (0.17)
SIS-MC+	0.88 (0.14)	0.78 (0.14)	0.72 (0.13)	0.69 (0.09)	0.66 (0.15)
ISIS-MC+	0.69 (0.33)	0.19 (0.29)	0.02 (0.03)	0.01 (0.01)	0.01 (0.01)
SC-lasso	0.96 (0.12)	0.83 (0.19)	0.73 (0.18)	0.67 (0.14)	0.64 (0.11)
SC-forward	0.98 (0.10)	0.84 (0.21)	0.71 (0.25)	0.54 (0.26)	0.44 (0.23)
SC-marginal	0.96 (0.13)	0.83 (0.19)	0.73 (0.18)	0.67 (0.14)	0.65 (0.11)
RAR_1	0.90 (0.25)	0.50 (0.30)	0.27 (0.17)	0.20 (0.10)	0.17 (0.07)
RAR_5	0.89 (0.20)	0.52 (0.32)	0.25 (0.18)	0.18 (0.09)	0.14 (0.06)
RAR_{30}	0.90 (0.15)	0.57 (0.32)	0.28 (0.22)	0.18 (0.11)	0.13 (0.06)
RAR(MC+)_{30}	0.78 (0.30)	0.27 (0.38)	0.04 (0.17)	0.01 (0.01)	0.01 (0.01)
RAR+_1	0.81 (0.25)	0.34 (0.36)	0.08 (0.17)	0.03 (0.05)	0.02 (0.04)
RAR+_5	0.86 (0.21)	0.39 (0.39)	0.08 (0.19)	0.02 (0.06)	0.01 (0.01)
RAR+_{30}	0.89 (0.17)	0.49 (0.40)	0.14 (0.27)	0.03 (0.11)	0.01 (0.05)
RAR+(MC+)_{30}	0.77 (0.29)	0.28 (0.37)	0.05 (0.16)	0.01 (0.01)	0.01 (0.01)

Scenario 2 (D)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	0.90 (0.10)	0.72 (0.17)	0.53 (0.16)	0.41 (0.13)	0.32 (0.10)
SCAD	0.61 (0.41)	0.14 (0.29)	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)
MC+	0.63 (0.39)	0.16 (0.30)	0.01 (0.02)	0.00 (0.01)	0.00 (0.00)
SIS-lasso	0.90 (0.10)	0.82 (0.12)	0.78 (0.12)	0.75 (0.13)	0.71 (0.17)
ISIS-lasso	0.77 (0.18)	0.60 (0.16)	0.49 (0.13)	0.40 (0.11)	0.33 (0.09)
Ada-lasso	0.87 (0.12)	0.76 (0.13)	0.66 (0.15)	0.60 (0.17)	0.53 (0.21)
SIS-MC+	0.87 (0.14)	0.76 (0.17)	0.70 (0.19)	0.67 (0.19)	0.63 (0.23)
ISIS-MC+	0.59 (0.37)	0.15 (0.28)	0.02 (0.03)	0.01 (0.01)	0.01 (0.01)
SC-lasso	0.96 (0.13)	0.81 (0.20)	0.74 (0.19)	0.69 (0.16)	0.68 (0.16)
SC-forward	0.97 (0.11)	0.82 (0.24)	0.66 (0.31)	0.56 (0.31)	0.49 (0.29)
SC-marginal	0.94 (0.15)	0.81 (0.20)	0.74 (0.19)	0.69 (0.15)	0.68 (0.12)
RAR_1	0.85 (0.31)	0.49 (0.28)	0.29 (0.15)	0.22 (0.10)	0.18 (0.07)
RAR_5	0.85 (0.26)	0.48 (0.31)	0.27 (0.15)	0.20 (0.09)	0.16 (0.06)
RAR_{30}	0.87 (0.20)	0.53 (0.33)	0.29 (0.21)	0.20 (0.09)	0.16 (0.06)
RAR(MC+)_{30}	0.71 (0.37)	0.30 (0.42)	0.02 (0.05)	0.01 (0.01)	0.01 (0.01)
RAR+_1	0.74 (0.31)	0.31 (0.33)	0.07 (0.12)	0.03 (0.04)	0.02 (0.02)
RAR+_5	0.80 (0.27)	0.32 (0.35)	0.05 (0.11)	0.02 (0.03)	0.01 (0.01)
RAR+_{30}	0.84 (0.22)	0.39 (0.38)	0.09 (0.18)	0.02 (0.04)	0.01 (0.01)
RAR+(MC+)_{30}	0.70 (0.36)	0.29 (0.40)	0.02 (0.05)	0.01 (0.01)	0.00 (0.00)

Table 6: Model size over 200 simulation rounds.

Scenario 2 (C)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	16.72 (16.71)	58.99 (48.10)	131.48 (55.65)	169.27 (41.09)	186.28 (37.59)
SCAD	20.20 (17.58)	34.18 (27.02)	17.29 (18.99)	8.30 (10.01)	6.81 (9.23)
MC+	8.45 (12.23)	13.79 (18.93)	5.80 (11.49)	3.11 (4.45)	3.50 (5.39)
SIS-lasso	10.01 (7.60)	21.30 (7.06)	27.29 (7.58)	31.15 (7.66)	34.57 (12.02)
ISIS-lasso	13.39 (9.37)	33.71 (9.01)	49.78 (6.00)	62.13 (4.25)	72.32 (4.80)
Ada-lasso	7.69 (8.58)	16.81 (21.86)	29.96 (46.00)	45.74 (61.53)	65.17 (80.59)
SIS-MC+	6.31 (5.65)	14.79 (5.97)	18.61 (5.96)	21.27 (6.75)	21.43 (9.09)
ISIS-MC+	9.59 (8.06)	14.55 (9.34)	7.17 (5.40)	4.07 (2.93)	2.59 (1.08)
SC-lasso	0.12 (0.34)	0.50 (0.57)	0.72 (0.49)	0.90 (0.43)	0.98 (0.36)
SC-forward	0.08 (0.29)	0.44 (0.55)	0.84 (0.74)	1.31 (0.78)	1.64 (0.67)
SC-marginal	0.14 (0.38)	0.48 (0.54)	0.73 (0.52)	0.87 (0.40)	0.97 (0.35)
RAR_1	19.90 (22.59)	64.49 (33.50)	81.28 (29.98)	88.62 (28.73)	98.68 (31.01)
RAR_5	19.02 (19.35)	61.52 (33.55)	77.38 (33.44)	83.52 (27.01)	88.77 (27.94)
RAR_{30}	17.16 (17.62)	61.05 (37.92)	80.22 (39.95)	81.14 (28.66)	83.48 (28.46)
RAR(MC+)_{30}	2.11 (2.48)	3.65 (3.16)	2.67 (1.45)	2.54 (1.11)	2.49 (0.97)
RAR+_1	9.64 (15.46)	17.97 (33.94)	9.14 (28.33)	3.87 (4.89)	4.36 (10.48)
RAR+_5	14.78 (17.17)	23.42 (37.29)	11.99 (35.87)	4.60 (17.24)	2.69 (0.82)
RAR+_{30}	15.76 (16.93)	32.93 (44.52)	26.13 (55.20)	8.42 (31.16)	3.85 (15.78)
RAR+(MC+)_{30}	2.01 (2.34)	3.48 (3.17)	2.37 (1.32)	2.17 (0.71)	2.10 (0.33)

Scenario 2 (D)	(100, 1232)	(200, 1791)	(300, 2285)	(400, 2750)	(500, 3199)
Lasso	17.81 (18.06)	75.94 (48.96)	135.83 (42.49)	159.51 (40.30)	172.10 (38.90)
SCAD	20.20 (17.67)	29.11 (28.29)	13.91 (18.14)	6.55 (7.65)	4.33 (5.23)
MC+	9.87 (14.68)	17.61 (30.66)	6.53 (11.59)	2.79 (2.35)	2.85 (3.08)
SIS-lasso	10.05 (7.04)	19.57 (7.19)	24.47 (9.09)	27.64 (11.37)	32.82 (15.18)
ISIS-lasso	14.90 (8.66)	34.42 (7.21)	49.34 (4.00)	59.79 (4.61)	68.28 (5.99)
Ada-lasso	9.94 (13.93)	25.73 (38.07)	46.73 (59.73)	58.56 (66.18)	76.77 (79.35)
SIS-MC+	6.58 (5.51)	12.82 (6.06)	14.77 (6.99)	15.53 (7.63)	14.37 (8.71)
ISIS-MC+	9.97 (7.27)	11.10 (9.83)	5.50 (6.69)	3.52 (2.77)	2.47 (1.16)
SC-lasso	0.16 (0.39)	0.62 (0.57)	0.95 (0.59)	1.15 (0.56)	1.36 (0.54)
SC-forward	0.13 (0.37)	0.57 (0.67)	1.13 (0.81)	1.51 (0.71)	1.72 (0.57)
SC-marginal	0.22 (0.46)	0.64 (0.60)	1.03 (0.63)	1.16 (0.56)	1.41 (0.52)
RAR_1	21.32 (21.68)	71.51 (41.23)	88.84 (27.77)	101.62 (31.03)	106.27 (30.42)
RAR_5	19.39 (19.22)	70.12 (40.15)	85.04 (28.81)	96.53 (27.18)	101.71 (27.97)
RAR_{30}	19.23 (19.60)	67.93 (40.90)	83.60 (37.14)	95.79 (29.60)	99.90 (27.50)
RAR(MC+)_{30}	2.37 (2.57)	3.42 (2.88)	3.37 (1.31)	3.34 (1.00)	3.29 (0.88)
RAR+_1	8.09 (12.03)	16.30 (28.40)	4.98 (4.84)	4.23 (4.85)	3.77 (1.98)
RAR+_5	12.91 (16.99)	20.65 (36.53)	4.78 (7.98)	3.33 (1.31)	3.02 (0.66)
RAR+_{30}	15.78 (18.26)	31.84 (47.07)	9.42 (27.34)	3.46 (4.80)	2.83 (0.47)
RAR+(MC+)_{30}	2.20 (2.57)	2.86 (2.74)	2.46 (1.06)	2.31 (0.74)	2.38 (0.49)

Table 7: Relative estimation error over 200 simulation rounds.

Scenario 3 (A)	(350, 2520)	(450, 2976)	(550, 3420)	(650, 3856)	(750, 4288)
Lasso	0.52 (0.10)	0.34 (0.07)	0.25 (0.05)	0.20 (0.03)	0.16 (0.02)
SCAD	0.04 (0.03)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
MC+	0.03 (0.02)	0.02 (0.01)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
SIS-lasso	0.92 (0.09)	0.89 (0.11)	0.89 (0.10)	0.86 (0.15)	0.85 (0.15)
ISIS-lasso	0.16 (0.10)	0.13 (0.05)	0.12 (0.03)	0.12 (0.02)	0.11 (0.03)
Ada-lasso	0.74 (0.17)	0.66 (0.20)	0.59 (0.21)	0.51 (0.21)	0.46 (0.20)
SIS-MC+	0.90 (0.11)	0.88 (0.12)	0.87 (0.12)	0.84 (0.16)	0.83 (0.17)
ISIS-MC+	0.04 (0.07)	0.02 (0.01)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
SC-lasso	0.98 (0.03)	0.96 (0.03)	0.94 (0.04)	0.92 (0.05)	0.89 (0.08)
SC-forward	0.97 (0.03)	0.95 (0.04)	0.93 (0.04)	0.90 (0.04)	0.88 (0.06)
SC-marginal	0.98 (0.03)	0.96 (0.03)	0.96 (0.03)	0.95 (0.03)	0.95 (0.03)
RAR_1	0.36 (0.17)	0.19 (0.10)	0.11 (0.05)	0.08 (0.03)	0.06 (0.02)
RAR_5	0.42 (0.15)	0.22 (0.11)	0.12 (0.06)	0.09 (0.04)	0.06 (0.03)
RAR_{30}	0.46 (0.13)	0.26 (0.10)	0.15 (0.07)	0.10 (0.05)	0.07 (0.03)
RAR(MC+)_{30}	0.39 (0.26)	0.04 (0.10)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
RAR+_1	0.29 (0.18)	0.12 (0.10)	0.05 (0.04)	0.02 (0.02)	0.02 (0.01)
RAR+_5	0.36 (0.17)	0.16 (0.11)	0.06 (0.05)	0.03 (0.03)	0.02 (0.01)
RAR+_{30}	0.42 (0.15)	0.19 (0.10)	0.08 (0.05)	0.04 (0.04)	0.02 (0.02)
RAR+(MC+)_{30}	0.41 (0.27)	0.05 (0.10)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)

Scenario 3 (B)	(150, 1524)	(250, 2043)	(350, 2520)	(450, 2976)	(550, 3420)
Lasso	0.87 (0.04)	0.62 (0.12)	0.29 (0.08)	0.17 (0.04)	0.12 (0.02)
SCAD	0.83 (0.07)	0.02 (0.05)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
MC+	0.83 (0.08)	0.02 (0.07)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
SIS-lasso	0.46 (0.24)	0.13 (0.15)	0.05 (0.06)	0.04 (0.02)	0.03 (0.02)
ISIS-lasso	0.08 (0.06)	0.06 (0.01)	0.05 (0.01)	0.04 (0.01)	0.04 (0.01)
Ada-lasso	0.34 (0.18)	0.08 (0.03)	0.05 (0.01)	0.03 (0.01)	0.02 (0.01)
SIS-MC+	0.46 (0.25)	0.10 (0.15)	0.03 (0.05)	0.01 (0.02)	0.01 (0.01)
ISIS-MC+	0.06 (0.06)	0.02 (0.01)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
SC-lasso	0.95 (0.03)	0.94 (0.04)	0.92 (0.04)	0.92 (0.04)	0.90 (0.08)
SC-forward	0.95 (0.02)	0.95 (0.01)	0.94 (0.02)	0.93 (0.04)	0.89 (0.06)
SC-marginal	0.95 (0.04)	0.81 (0.18)	0.52 (0.30)	0.28 (0.29)	0.11 (0.19)
RAR_1	0.28 (0.10)	0.10 (0.03)	0.07 (0.02)	0.06 (0.01)	0.04 (0.01)
RAR_5	0.28 (0.11)	0.10 (0.03)	0.07 (0.02)	0.06 (0.01)	0.04 (0.01)
RAR_{30}	0.29 (0.12)	0.10 (0.03)	0.07 (0.02)	0.06 (0.01)	0.04 (0.01)
RAR(MC+)_{30}	0.44 (0.15)	0.02 (0.02)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
RAR+_1	0.20 (0.11)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
RAR+_5	0.20 (0.12)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
RAR+_{30}	0.21 (0.12)	0.02 (0.01)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
RAR+(MC+)_{30}	0.45 (0.15)	0.02 (0.02)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)

Table 8: Model size over 200 simulation rounds.

Scenario 3 (A)	(350, 2520)	(450, 2976)	(550, 3420)	(650, 3856)	(750, 4288)
Lasso	273.33 (40.61)	315.99 (37.94)	341.34 (35.92)	367.96 (39.12)	381.97 (39.55)
SCAD	79.92 (31.09)	49.43 (23.25)	32.72 (14.22)	26.20 (7.67)	22.52 (3.35)
MC+	38.95 (16.02)	27.92 (9.80)	23.37 (3.85)	21.85 (2.31)	20.97 (1.22)
SIS-lasso	31.88 (14.65)	38.06 (18.36)	46.86 (24.08)	57.50 (29.46)	73.43 (33.73)
ISIS-lasso	58.95 (0.29)	72.81 (0.66)	86.44 (1.32)	97.69 (2.46)	107.52 (3.32)
Ada-lasso	166.19 (89.43)	201.85 (89.22)	263.53 (88.35)	300.00 (72.96)	310.65 (88.22)
SIS-MC+	15.59 (11.34)	18.58 (15.53)	24.62 (23.28)	29.34 (28.38)	37.61 (31.48)
ISIS-MC+	27.55 (4.50)	24.51 (3.40)	22.59 (2.31)	21.40 (1.53)	20.92 (1.17)
SC-lasso	0.58 (0.64)	0.89 (0.66)	1.27 (1.03)	1.77 (1.38)	2.83 (2.27)
SC-forward	0.77 (0.80)	1.28 (0.84)	1.65 (0.86)	2.27 (1.06)	2.91 (1.59)
SC-marginal	0.57 (0.62)	0.80 (0.73)	0.94 (0.71)	1.17 (0.93)	1.12 (0.90)
RAR_1	218.64 (60.94)	228.83 (69.80)	206.50 (57.03)	199.46 (51.52)	199.99 (48.71)
RAR_5	239.74 (58.02)	246.46 (66.92)	226.70 (65.84)	212.86 (62.83)	204.09 (56.31)
RAR_{30}	254.22 (52.28)	267.92 (64.09)	256.74 (67.66)	234.63 (73.18)	218.23 (67.62)
$RAR(MC+)_{30}$	20.40 (7.46)	22.18 (3.28)	21.03 (1.49)	20.82 (1.03)	20.70 (0.94)
$RAR+_1$	92.95 (84.72)	55.08 (48.71)	34.42 (23.43)	26.83 (11.77)	24.20 (6.91)
$RAR+_5$	131.48 (88.93)	80.40 (71.91)	43.02 (29.91)	31.13 (23.89)	24.28 (7.85)
$RAR+_{30}$	164.85 (95.99)	106.04 (83.80)	57.09 (46.21)	38.85 (40.98)	28.58 (31.31)
$RAR+(MC+)_{30}$	20.19 (8.28)	21.67 (3.57)	20.44 (1.21)	20.12 (0.37)	20.03 (0.16)

Scenario 3 (B)	(150, 1524)	(250, 2043)	(350, 2520)	(450, 2976)	(550, 3420)
Lasso	45.04 (40.01)	206.38 (38.38)	284.03 (30.34)	316.64 (40.11)	340.98 (40.12)
SCAD	49.11 (31.46)	46.71 (23.43)	23.59 (4.54)	21.32 (2.85)	20.85 (1.99)
MC+	36.14 (24.90)	28.02 (13.71)	21.31 (2.94)	20.79 (1.96)	20.54 (1.34)
SIS-lasso	27.02 (3.82)	43.83 (1.51)	57.53 (1.49)	70.73 (2.16)	83.61 (3.25)
ISIS-lasso	28.93 (0.31)	44.74 (0.58)	58.80 (0.53)	72.75 (0.57)	86.51 (0.84)
Ada-lasso	91.74 (19.04)	77.65 (16.54)	67.28 (12.78)	60.41 (8.76)	54.87 (6.83)
SIS-MC+	22.85 (5.63)	22.98 (5.28)	21.35 (2.50)	20.70 (1.34)	21.00 (1.71)
ISIS-MC+	26.34 (2.09)	23.38 (3.28)	21.22 (1.71)	20.88 (1.61)	20.89 (1.71)
SC-lasso	1.00 (0.66)	1.23 (0.71)	1.55 (0.79)	1.60 (0.84)	2.17 (1.61)
SC-forward	0.97 (0.23)	1.03 (0.22)	1.19 (0.46)	1.51 (0.82)	2.28 (1.19)
SC-marginal	1.28 (1.22)	4.14 (3.75)	10.02 (5.98)	14.73 (5.75)	18.03 (3.56)
RAR_1	102.92 (18.46)	133.64 (22.68)	155.96 (24.82)	176.36 (29.92)	161.19 (23.87)
RAR_5	102.24 (18.98)	133.68 (22.66)	155.96 (24.82)	176.36 (29.92)	161.14 (23.72)
RAR_{30}	102.50 (20.28)	133.82 (22.98)	155.90 (24.82)	176.36 (29.92)	161.32 (23.90)
$RAR(MC+)_{30}$	17.85 (3.66)	24.94 (1.73)	27.48 (0.84)	30.35 (1.03)	30.66 (1.19)
$RAR+_1$	35.79 (12.48)	26.00 (3.72)	25.73 (1.89)	26.57 (1.97)	27.32 (1.81)
$RAR+_5$	35.63 (12.75)	25.98 (3.72)	25.73 (1.89)	26.57 (1.97)	27.32 (1.79)
$RAR+_{30}$	35.54 (14.90)	25.96 (3.72)	25.73 (1.89)	26.57 (1.97)	27.31 (1.79)
$RAR+(MC+)_{30}$	16.41 (3.86)	20.45 (1.42)	20.02 (0.12)	20.00 (0.00)	20.01 (0.07)

Table 9: Relative estimation error over 200 simulation rounds.

Scenario 4 (C)	(300, 2285)	(400, 2750)	(500, 3199)	(600, 3639)	(700, 4073)
Lasso	0.88 (0.04)	0.81 (0.06)	0.68 (0.08)	0.50 (0.09)	0.36 (0.07)
SCAD	0.33 (0.34)	0.01 (0.02)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)
MC+	0.44 (0.36)	0.01 (0.04)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)
SIS-lasso	0.88 (0.04)	0.85 (0.04)	0.82 (0.03)	0.81 (0.03)	0.79 (0.03)
ISIS-lasso	0.48 (0.18)	0.26 (0.16)	0.14 (0.10)	0.10 (0.06)	0.08 (0.03)
Ada-lasso	0.87 (0.04)	0.81 (0.05)	0.77 (0.05)	0.72 (0.06)	0.69 (0.07)
SIS-MC+	0.85 (0.06)	0.80 (0.06)	0.76 (0.04)	0.73 (0.05)	0.70 (0.05)
ISIS-MC+	0.34 (0.22)	0.09 (0.12)	0.02 (0.05)	0.01 (0.01)	0.00 (0.00)
SC-lasso	0.98 (0.03)	0.96 (0.04)	0.94 (0.04)	0.91 (0.05)	0.88 (0.06)
SC-forward	0.98 (0.03)	0.97 (0.03)	0.95 (0.04)	0.92 (0.06)	0.87 (0.09)
SC-marginal	0.98 (0.03)	0.96 (0.03)	0.94 (0.04)	0.91 (0.05)	0.88 (0.06)
RAR_1	0.80 (0.10)	0.61 (0.14)	0.39 (0.14)	0.22 (0.08)	0.14 (0.05)
RAR_5	0.84 (0.07)	0.69 (0.12)	0.46 (0.14)	0.27 (0.09)	0.18 (0.06)
RAR_{30}	0.86 (0.05)	0.74 (0.10)	0.55 (0.14)	0.33 (0.10)	0.22 (0.06)
$RAR(MC+)_{30}$	0.74 (0.19)	0.09 (0.22)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
$RAR+_1$	0.77 (0.11)	0.56 (0.15)	0.32 (0.15)	0.13 (0.08)	0.05 (0.04)
$RAR+_5$	0.80 (0.09)	0.65 (0.13)	0.40 (0.16)	0.18 (0.10)	0.08 (0.05)
$RAR+_{30}$	0.84 (0.07)	0.71 (0.11)	0.50 (0.16)	0.24 (0.12)	0.11 (0.06)
$RAR+(MC+)_{30}$	0.74 (0.19)	0.09 (0.22)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)

Scenario 4 (D)	(300, 2285)	(400, 2750)	(500, 3199)	(600, 3639)	(700, 4073)
Lasso	0.86 (0.05)	0.74 (0.08)	0.54 (0.09)	0.37 (0.07)	0.28 (0.05)
SCAD	0.18 (0.27)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
MC+	0.29 (0.32)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
SIS-lasso	0.88 (0.04)	0.84 (0.04)	0.81 (0.04)	0.79 (0.04)	0.78 (0.04)
ISIS-lasso	0.43 (0.19)	0.22 (0.14)	0.10 (0.07)	0.07 (0.04)	0.07 (0.02)
Ada-lasso	0.85 (0.04)	0.79 (0.06)	0.72 (0.06)	0.66 (0.08)	0.61 (0.09)
SIS-MC+	0.86 (0.06)	0.80 (0.06)	0.76 (0.05)	0.73 (0.06)	0.71 (0.05)
ISIS-MC+	0.26 (0.21)	0.06 (0.09)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)
SC-lasso	0.98 (0.03)	0.96 (0.04)	0.94 (0.05)	0.91 (0.05)	0.88 (0.06)
SC-forward	0.99 (0.02)	0.97 (0.03)	0.95 (0.05)	0.90 (0.08)	0.80 (0.16)
SC-marginal	0.98 (0.03)	0.96 (0.04)	0.94 (0.05)	0.92 (0.05)	0.89 (0.06)
RAR_1	0.78 (0.11)	0.57 (0.15)	0.34 (0.12)	0.21 (0.08)	0.15 (0.05)
RAR_5	0.82 (0.09)	0.63 (0.13)	0.40 (0.13)	0.25 (0.08)	0.17 (0.06)
RAR_{30}	0.84 (0.07)	0.67 (0.13)	0.45 (0.13)	0.29 (0.09)	0.20 (0.06)
$RAR(MC+)_{30}$	0.56 (0.32)	0.09 (0.20)	0.01 (0.05)	0.01 (0.00)	0.00 (0.00)
$RAR+_1$	0.75 (0.12)	0.51 (0.16)	0.26 (0.13)	0.12 (0.08)	0.06 (0.04)
$RAR+_5$	0.79 (0.09)	0.57 (0.14)	0.32 (0.14)	0.15 (0.08)	0.07 (0.05)
$RAR+_{30}$	0.81 (0.08)	0.62 (0.14)	0.37 (0.14)	0.18 (0.09)	0.09 (0.05)
$RAR+(MC+)_{30}$	0.56 (0.31)	0.09 (0.20)	0.01 (0.04)	0.00 (0.00)	0.00 (0.00)

Table 10: Model size over 200 simulation rounds.

Scenario 4 (C)	(300, 2285)	(400, 2750)	(500, 3199)	(600, 3639)	(700, 4073)
Lasso	148.65 (53.33)	277.66 (81.33)	450.59 (65.89)	564.44 (51.95)	631.53 (55.97)
SCAD	111.59 (46.40)	45.05 (27.74)	29.10 (8.13)	24.03 (4.56)	23.19 (4.67)
MC+	85.45 (48.13)	35.48 (21.38)	23.66 (3.89)	22.26 (3.22)	22.29 (3.69)
SIS-lasso	49.21 (3.28)	62.41 (3.09)	74.89 (3.62)	84.70 (4.89)	93.89 (6.45)
ISIS-lasso	51.91 (1.27)	66.00 (0.00)	80.00 (0.00)	93.00 (0.00)	106.00 (0.00)
Ada-lasso	82.34 (42.62)	128.74 (81.82)	192.78 (114.44)	252.23 (137.11)	297.93 (149.76)
SIS-MC+	36.05 (5.56)	43.78 (5.30)	51.47 (5.14)	56.19 (5.88)	58.71 (6.79)
ISIS-MC+	36.04 (8.17)	27.24 (7.17)	23.08 (3.74)	21.77 (2.27)	21.51 (2.11)
SC-lasso	0.56 (0.73)	1.14 (1.05)	1.57 (1.21)	2.38 (1.45)	3.21 (1.71)
SC-forward	0.50 (0.66)	0.88 (0.84)	1.39 (1.21)	2.20 (1.58)	3.45 (2.47)
SC-marginal	0.56 (0.73)	1.08 (0.99)	1.57 (1.29)	2.39 (1.38)	3.15 (1.64)
RAR_1	180.54 (58.02)	323.93 (65.29)	432.16 (54.05)	464.14 (73.19)	471.97 (79.41)
RAR_5	166.57 (49.39)	308.80 (74.15)	439.25 (58.03)	492.87 (70.75)	511.66 (77.65)
RAR_{30}	157.47 (51.17)	291.75 (80.04)	443.62 (60.89)	514.98 (66.68)	545.32 (75.46)
RAR(MC+)_{30}	18.43 (7.09)	22.48 (5.05)	21.05 (1.30)	20.79 (1.09)	20.91 (1.38)
RAR+_1	70.92 (60.11)	132.89 (88.23)	178.72 (88.32)	129.23 (64.06)	77.42 (39.47)
RAR+_5	72.63 (64.51)	139.15 (106.46)	207.90 (101.73)	169.30 (77.60)	105.58 (51.73)
RAR+_{30}	105.98 (73.05)	155.62 (116.20)	239.47 (118.98)	209.94 (96.63)	143.95 (70.28)
RAR+(MC+)_{30}	17.55 (7.72)	22.01 (5.18)	20.29 (0.80)	20.08 (0.29)	20.10 (0.39)

Scenario 4 (D)	(300, 2285)	(400, 2750)	(500, 3199)	(600, 3639)	(700, 4073)
Lasso	171.26 (64.11)	335.36 (61.45)	461.26 (44.75)	539.62 (49.41)	604.03 (60.27)
SCAD	101.74 (44.90)	47.52 (27.61)	28.46 (7.60)	23.88 (4.67)	22.42 (3.72)
MC+	84.99 (44.22)	37.52 (20.12)	24.27 (4.73)	22.27 (3.89)	21.64 (2.77)
SIS-lasso	48.61 (3.23)	61.77 (3.34)	73.53 (4.30)	82.15 (6.47)	91.30 (7.94)
ISIS-lasso	52.00 (0.00)	66.00 (0.00)	80.00 (0.00)	93.00 (0.00)	106.00 (0.00)
Ada-lasso	98.95 (59.32)	172.17 (87.47)	240.71 (103.15)	299.51 (112.48)	359.44 (107.77)
SIS-MC+	35.42 (7.28)	43.16 (5.37)	48.31 (6.10)	51.56 (6.94)	51.92 (8.19)
ISIS-MC+	32.34 (7.99)	25.72 (5.28)	22.26 (2.67)	21.33 (2.00)	21.11 (1.67)
SC-lasso	0.68 (0.92)	1.31 (1.36)	2.15 (1.63)	3.35 (1.87)	4.64 (2.17)
SC-forward	0.45 (0.69)	0.93 (0.94)	1.73 (1.63)	3.67 (2.41)	6.82 (4.74)
SC-marginal	0.69 (0.93)	1.27 (1.22)	2.19 (1.54)	3.03 (1.66)	4.48 (2.07)
RAR_1	203.94 (58.08)	337.06 (55.88)	428.69 (55.35)	468.36 (70.64)	491.03 (86.23)
RAR_5	194.71 (61.69)	339.06 (58.72)	442.57 (54.91)	486.74 (67.00)	513.18 (90.01)
RAR_{30}	183.23 (64.52)	335.07 (58.49)	449.23 (48.38)	506.15 (62.94)	539.77 (78.42)
RAR(MC+)_{30}	22.11 (8.96)	23.93 (5.11)	22.38 (2.66)	23.17 (1.95)	23.77 (2.03)
RAR+_1	77.67 (65.12)	151.19 (79.52)	169.49 (84.70)	127.30 (70.25)	82.11 (45.90)
RAR+_5	82.79 (73.97)	164.93 (93.13)	199.56 (87.47)	151.95 (69.36)	104.80 (56.07)
RAR+_{30}	99.83 (80.28)	183.48 (106.48)	237.74 (98.37)	177.51 (75.64)	124.69 (58.79)
RAR+(MC+)_{30}	21.01 (9.51)	22.79 (5.11)	20.84 (3.16)	20.40 (0.75)	20.50 (0.95)