A Bootstrap Lasso + Partial Ridge Method to

Construct Confidence Intervals for Parameters in

High-dimensional Sparse Linear Models

Hanzhong Liu[1], Xin Xu[2], and Jingyi Jessica Li[3*]

[1] *Center for Statistical Science, Department of Industrial Engineering,*

Tsinghua University, Beijing, 100084, China

lhz2016@tsinghua.edu.cn

[2] *Department of Statistics,*

Yale University, New Haven, Connecticut, 06511, U.S.A.

xin.xu@yale.edu

[3] *Department of Statistics,*

University of California, Los Angeles, California, 90095, U.S.A.

[*] *To whom correspondence should be addressed: jli@stat.ucla.edu*

Supplementary Material

This supplementary material is organized as follows. Section S1 contains proofs of the results in the main

text. Section S2 provides examples that satisfy (or do not satisfy) Condition 8. Section S3 gives an example

that satisfies Condition 11. Section S4 and S5 provide simulation details and additional figures and tables of

simulation results, respectively. Section S6 provides the details of the real-data case study 2: neuroblastoma

gene expression data. Algorithms for rBLPR, pBLPR and cv(lasso+ols) are provided in Section S7.

S1 Proof of Theorems

S1.1 Proof of Theorem 1

We will follow the proof of the sign-consistency of the lasso in Zhao & Yu

(2006) with modifications when necessary. Before proving Theorem 1, we first

state the following Proposition 1 which is similar to the Proposition 1 in Zhao &

Yu (2006).

Proposition 1. *Assume Condition 5 holds with a constant $\eta > 0$, then*

$$\mathrm{pr}\left((\hat{\beta}_{\mathrm{lasso}})_S =_s \beta_S^0, \ (\hat{\beta}_{\mathrm{lasso}})_{S^c} = 0\right) \geq \mathrm{pr}(A_n \cap B_n) \qquad (\text{S1.1})$$

for

$$A_n = \left\{|C_{11}^{-1}W_S| < n^{\frac{1}{2}}\left(|\beta_S^0| - \lambda_1|C_{11}^{-1}\operatorname{sign}(\beta_S^0)| - |C_{11}^{-1}C_{12}\beta_{S^c}^0|\right)\right\},$$

$$B_n = \left\{|C_{21}C_{11}^{-1}W_S - W_{S^c}| \leq n^{\frac{1}{2}}\lambda_1\eta - |n^{\frac{1}{2}}(C_{21}C_{11}^{-1}C_{12} - C_{22})\beta_{S^c}^0|\right\},$$

where

$$W_S = n^{-\frac{1}{2}}X_S^{\mathrm{T}}\epsilon, \quad W_{S^c} = n^{-\frac{1}{2}}X_{S^c}^{\mathrm{T}}\epsilon.$$

Setting $\beta_{S^c}^0 = 0$, then Proposition 1 gives back to the same proposition in

Zhao & Yu (2006).

Proof. By Karush-Kuhn-Tucker condition for convex optimization, we obtain the following Lemma S1 without giving the proof.

Lemma S1. $\hat{\beta}_{\text{lasso}}$ *is the lasso estimator defined in* (2.2) *if and only if*

$$\frac{1}{2n}\frac{d\|Y-X\beta\|_2^2}{d\beta_j}\Big|_{\beta_j=(\hat{\beta}_{\text{lasso}})_j} = -\lambda_1 \operatorname{sign}\left((\hat{\beta}_{\text{lasso}})_j\right) \quad \text{for } j, \text{ such that } (\hat{\beta}_{\text{lasso}})_j \neq 0,$$

$$\frac{1}{2n}\left|\frac{d\|Y-X\beta\|_2^2}{d\beta_j}\Big|_{\beta_j=(\hat{\beta}_{\text{lasso}})_j}\right| \leq \lambda_1 \quad \text{for } j, \text{ such that } (\hat{\beta}_{\text{lasso}})_j = 0.$$

It is easy to obtain

$$\frac{1}{2n}\frac{d\|Y-X\beta\|_2^2}{d\beta} = -\frac{1}{n}X^{\mathrm{T}}(Y-X\beta) = C(\beta-\beta^0) - \frac{1}{n}X^{\mathrm{T}}\epsilon,$$

where $C = X^{\mathrm{T}}X/n$. Then by definition of the lasso (2.2) and Lemma S1, if there exist $\hat{\beta} = (\hat{\beta}_S^{\mathrm{T}}, 0_{S^c}^{\mathrm{T}})^{\mathrm{T}}$, such that the following holds:

$$n^{\frac{1}{2}}C_{11}(\hat{\beta}-\beta^0)_S - n^{\frac{1}{2}}C_{12}\beta_{S^c}^0 - X_S^{\mathrm{T}}\epsilon/n^{\frac{1}{2}} = -n^{\frac{1}{2}}\lambda_1 \operatorname{sign}(\beta_S^0), \qquad \text{(S1.2)}$$

$$-n^{\frac{1}{2}}\lambda_1\mathbf{1} \leq n^{\frac{1}{2}}C_{21}(\hat{\beta}-\beta^0)_S - X_{S^c}^{\mathrm{T}}\epsilon/n^{\frac{1}{2}} - n^{\frac{1}{2}}C_{22}\beta_{S^c}^0 \leq n^{\frac{1}{2}}\lambda_1\mathbf{1}, \qquad \text{(S1.3)}$$

$$|(\hat{\beta}-\beta^0)_S| < |\beta_S^0|, \qquad\qquad\qquad \text{(S1.4)}$$

then, $\hat{\beta}$ is the lasso solution, that is, $\hat{\beta} = \hat{\beta}_{\text{lasso}}$ and, hence, $(\hat{\beta}_{\text{lasso}})_{S^c} = \hat{\beta}_{S^c} = 0$ and $\operatorname{sign}((\hat{\beta}_{\text{lasso}})_S) = \operatorname{sign}(\hat{\beta}_S) = \operatorname{sign}(\beta_S^0)$. Let $W = X^{\mathrm{T}}\epsilon/n^{\frac{1}{2}}$, then, $W_S = X_S^{\mathrm{T}}\epsilon/n^{\frac{1}{2}}$ and $W_{S^c} = X_{S^c}^{\mathrm{T}}\epsilon/n^{\frac{1}{2}}$.

Substitute $(\hat{\beta} - \beta^0)_S$ and bound the absolute values, the existence of such $\hat{\beta}$ is implied by

$$|C_{11}^{-1}W_S| < n^{\frac{1}{2}} \left(|\beta_S^0| - \lambda_1 |C_{11}^{-1}\operatorname{sign}(\beta_S^0)| - |C_{11}^{-1}C_{12}\beta_{S^c}^0| \right), \qquad \text{(S1.5)}$$

$$|C_{21}C_{11}^{-1}W_S - W_{S^c}| \le n^{\frac{1}{2}}\lambda_1 \left(\mathbf{1} - |C_{21}C_{11}^{-1}\operatorname{sign}(\beta_S^0)| \right) - |n^{\frac{1}{2}} \left(C_{21}C_{11}^{-1}C_{12} - C_{22} \right) \beta_{S^c}^0|.$$

$$\text{(S1.6)}$$

$\{$(S1.5)$\}$ coincides with A_n and $\{$(S1.6)$\} \subset B_n$. This implies Proposition 1.

To prove Theorem 1, we can follow the proof of Theorem 4 in Zhao & Yu (2006), using our new Proposition 1.

First, by Proposition 1, we have

$$\operatorname{pr}\left((\hat{\beta}_{\text{lasso}})_S =_s \beta_S, \ (\hat{\beta}_{\text{lasso}})_{S^c} = 0 \right) \ge \operatorname{pr}(A_n \cap B_n).$$

On the other hand,

$$1 - \operatorname{pr}(A_n \cap B_n) \le \operatorname{pr}(A_n^c) + \operatorname{pr}(B_n^c)$$

$$\le \sum_{i=1}^{s} \operatorname{pr}\left(|z_i| \ge n^{\frac{1}{2}}(|\beta_i^0| - \lambda_1 b_i - h_i) \right) + \sum_{i=1}^{p-s} \operatorname{pr}\left(|\zeta_i| \ge n^{\frac{1}{2}}\lambda_1 \eta_i - m_i \right),$$

$$\text{(S1.7)}$$

where

$$z = (z_1, \ldots, z_s)^{\mathrm{T}} = C_{11}^{-1}W_S,$$

$$\zeta = (\zeta_1, \ldots, \zeta_{p-s})^{\mathrm{T}} = C_{21}C_{11}^{-1}W_S - W_{S^c},$$

$$b = (b_1, \ldots, b_s) = |C_{11}^{-1} \operatorname{sign}(\beta_S^0))|,$$

$$h = (h_1, \ldots, h_s) = |C_{11}^{-1} C_{12} \beta_{S^c}^0|,$$

$$m = (m_1, \ldots, m_{p-s}) = |n^{\frac{1}{2}} \left(C_{21} C_{11}^{-1} C_{12} - C_{22} \right) \beta_{S^c}^0|.$$

Due to Condition 1, ϵ_i are independent and identically distributed subGaussian random variables, with mean 0 and variance σ^2. Therefore, z_i's and ζ_i's are all subGaussian random variables, with mean 0. By simple algebra, we have

$$E(zz^{\mathrm{T}}) = \sigma^2 C_{11}^{-1}; \quad E(\zeta\zeta^{\mathrm{T}}) = \sigma^2 (C_{22} - C_{21} C_{11}^{-1} C_{12}).$$

Therefore,

$$E z_i^2 = \sigma^2 (C_{11}^{-1})_{ii} \leq \sigma^2 \Lambda_{\max}(C_{11}^{-1}) \leq \sigma^2 / \Lambda,$$

where the last inequality is due to Condition 3. Moreover,

$$E \zeta_i^2 = \sigma^2 (C_{22} - C_{21} C_{11}^{-1} C_{12})_{ii} \leq \sigma^2 (C_{22})_{ii} = \sigma^2,$$

where the last equality is because of Condition 2. Therefore, z_i's and ζ_i's are subGaussian random variables, with mean 0 and finite variance. Hence, there exits a constant $c > 0$, such that, for all $t > 0$,

$$\operatorname{pr}(|z_i| \geq t) \leq 2e^{-ct^2}; \quad \operatorname{pr}(|\zeta_i| \geq t) \leq 2e^{-ct^2}.$$

For $i = 1, \ldots, s$, using Cauchy-Schwarz inequality and Conditions 3, 4, and

7, we have

$$
\begin{aligned}
n^{\frac{1}{2}}\lambda_1|b_i| &\leq n^{\frac{1}{2}}\lambda_1\Lambda_{\max}(C_{11}^{-1})\|\operatorname{sign}(\beta_S^0)\|_2 \leq s^{\frac{1}{2}}n^{\frac{1}{2}}\lambda_1/\Lambda \\
&= O(n^{\frac{1}{2}}n^{\frac{c_1+c_4-1}{2}}) = o(n^{\frac{1}{2}}n^{\frac{c_3-1}{2}}),
\end{aligned}
$$

where the last inequality holds because $c_4 < c_3 - c_1$ (see Condition 7).

Condition 8 implies that $n^{\frac{1}{2}}h_i = O(1)$, for $i = 1, \ldots, s$. Combining with Condition 6, we have

$$
n^{\frac{1}{2}}\lambda_1|b_i| + n^{\frac{1}{2}}h_i = o(1)n^{\frac{1}{2}}|\beta_i^0|, \text{ for } i = 1, \cdots, s.
$$

Therefore,

$$
\begin{aligned}
\sum_{i=1}^{s}\operatorname{pr}\left(|z_i| \geq n^{\frac{1}{2}}(|\beta_i^0| - \lambda_1 b_i - h_i)\right) &\leq \sum_{i=1}^{s}\operatorname{pr}\left(|z_i| \geq (1+o(1))n^{\frac{1}{2}}|\beta_i^0|\right) \\
&\leq \sum_{i=1}^{s}\operatorname{pr}(|z_i| \geq n^{\frac{c_3}{2}}) \\
&= o(e^{-n^{c_2}}).
\end{aligned}
$$

(S1.8)

Due to Conditions 8 and 7, $m_i = o(n^{c_4/2})$, and $n^{\frac{1}{2}}\lambda_1 = O(n^{c_4/2})$. Then,

$$
\sum_{i=1}^{p-s}\operatorname{pr}\left(|\zeta_i| \geq n^{\frac{1}{2}}\lambda_1\eta_i - m_i\right) \leq \sum_{i=1}^{p-s}\operatorname{pr}\left(|\zeta_i| \geq O(n^{\frac{c_4}{2}})\right) = o(e^{-n^{c_2}}). \quad \text{(S1.9)}
$$

Theorem 1 follows immediately. $\qquad\square$

S1.2   Proof of Theorem 2

Proof. We have to check that the residual bootstrap version of Conditions $1 -$ 7 hold, with conditional probability, given ϵ, going to one. For the residual bootstrap sample, we have

$$Y^*_{\text{rboot}} = X\hat{\beta}_{\text{lasso+ols}} + \epsilon^*.$$

Conditions 2, 3 and 7 depend only on X and λ_1 which are the same for the original sample (X, Y) and bootstrap sample (X, Y^*_{rboot}), therefore, they hold obviously. We next show, one by one, the bootstrap version of Conditions 1, $4 -$ 8 hold, with probability going to one. We need the following Lemma.

Lemma S2. *Under Conditions $1 - 7$, and for the constant M in Condition 6, we have*

$$\text{pr}\left(||\hat{\beta}_{\text{lasso+ols}} - \beta^0||_\infty \leq 2Mn^{\frac{c_1-1}{2}}\right) \to 1. \tag{S1.10}$$

Lemma S2 bounds element-wise estimation error of the lasso+ols estimator, the proof of which can be founded in the following subsection S1.4.

Now, we can show that residual bootstrap version of Conditions 1, $4 - 8$ hold, with probability going to one. Under Conditions $1 - 7$ and using Theorem 1, the lasso $\hat{\beta}_{\text{lasso}}$ has sign-consistency, that is,

$$\text{pr}(\hat{S} = S) = 1 - o(e^{-n^{c_2}}) \to 1.$$

Replacing (β^0, ϵ, Y) with $(\hat{\beta}_{\text{lasso+ols}}, \epsilon^*, Y^*_{\text{rboot}})$.

In what follows, we always condition on $\{\hat{S} = S\}$. By Lemma S2, it is easy to show that

$$\mathrm{pr}\left((\hat{\beta}_{\mathrm{lasso+ols}})_S =_s \beta_S^0\right) \to 1,$$

which guarantees that bootstrap version of Conditions 4 – 8 hold, with probability going to one. Therefore, we only need to show the bootstrap version of Condition 1 holds, with probability going to 1, that is,

Condition S1. ϵ_i^* are conditionally independent and identically distributed sub-Gaussian random variables, with mean 0. That is, there exists constant $C^* > 0$ and $c^* > 0$, such that

$$\mathrm{pr}\left(|\epsilon_i^*| \geq t \mid \epsilon\right) \leq C^* e^{-c^* t^2}, \ \forall t \geq 0, \tag{S1.11}$$

holds in probability.

Lemma S3. *Conditions 1 – 10 imply Condition S1.*

The proof is similar to that in Liu & Yu (2013) with modifications accounting for cliff-weak-sparsity. Let $\mathbb{I}$ denote the indicator function. Note that $\mathrm{pr}(|\epsilon_i^*| \geq t \mid \epsilon) = \left(\sum_{i=1}^n \mathbb{I}_{|\hat{\epsilon}_i - \bar{\epsilon}| \geq t}\right)/n$, hence, S1.11 is equivalent to

$$\sup_{t \geq 0}\left\{\frac{1}{n}\sum_{i=1}^n e^{c^* t^2}\mathbb{I}_{|\hat{\epsilon}_i - \bar{\epsilon}| \geq t}\right\} \leq C^*. \tag{S1.12}$$

We know that

$$
\begin{aligned}
\hat{\epsilon}_i - \tilde{\epsilon} &= y_i - x_i^{\mathrm{T}}\hat{\beta}_{\mathrm{lasso+ols}} - (\overline{y} - \overline{x}^{\mathrm{T}}\hat{\beta}_{\mathrm{lasso+ols}}) \\
&= x_i^{\mathrm{T}}\beta^0 + \epsilon_i - x_i^{\mathrm{T}}\hat{\beta}_{\mathrm{lasso+ols}} - (\overline{x}^{\mathrm{T}}\beta^0 + \overline{\epsilon} - \overline{x}^{\mathrm{T}}\hat{\beta}_{\mathrm{lasso+ols}}) \\
&= x_i^{\mathrm{T}}(\beta^0 - \hat{\beta}_{\mathrm{lasso+ols}}) + \epsilon_i - \overline{\epsilon},
\end{aligned}
\tag{S1.13}
$$

where x_i^{T} is the ith row of X, $\overline{y} = \sum_{i=1}^{n} y_i/n$, $\overline{\epsilon} = \sum_{i=1}^{n} \epsilon_i/n$, and $\overline{x} = \sum_{i=1}^{n} x_i/n = 0$. It is easy to see that $\sup_{t\geq 0}\left\{(\sum_{i=1}^{n} e^{c^* t^2}\mathbb{I}_{|\hat{\epsilon}_i - \tilde{\epsilon}|\geq t})/n\right\}$ can be bounded by

$$
\frac{1}{n}\sum_{i=1}^{n}\left\{\sup_{t\geq 0}\left\{e^{c^* t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0 - \hat{\beta}_{\mathrm{lasso+ols}})|\geq t/3}\right\} + \sup_{t\geq 0}\left\{e^{c^* t^2}\mathbb{I}_{|\overline{\epsilon}|\geq t/3}\right\} + \sup_{t\geq 0}\left\{e^{c^* t^2}\mathbb{I}_{|\epsilon_i|\geq t/3}\right\}\right\}.
\tag{S1.14}
$$

We can bound the second and third terms exactly the same as those in Liu & Yu (2013), that is, there exist a constant $C_1^* > 0$, such that for $c^* = 1/(36\sigma^2)$,

$$
\mathrm{pr}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\geq 0}\left\{e^{c^* t^2}\mathbb{I}_{|\overline{\epsilon}|\geq t/3}\right\} \leq C_1^*\right) \to 1.
\tag{S1.15}
$$

$$
\mathrm{pr}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\geq 0}\left\{e^{c^* t^2}\mathbb{I}_{|\epsilon_i|\geq t/3}\right\} \leq C_1^*\right) \to 1.
\tag{S1.16}
$$

Since the proof is exactly the same, we omit it here. Next, we bound the first term, which is different from that in Liu & Yu (2013), because of the weaker cliff-weak-sparsity assumption. For the constant $D > 0$ appearing in Condition

10,

$$
\begin{aligned}
&\mathrm{pr}\left(\max_{1\leq i\leq n}|x_i^{\mathrm{T}}(\beta^0-\hat{\beta}_{\text{lasso+ols}})|\geq 2D\right)\\
=\ &\mathrm{pr}\left(\max_{1\leq i\leq n}|x_i^{\mathrm{T}}(\beta^0-\hat{\beta}_{\text{lasso+ols}})|\geq 2D,\hat{S}=S\right)\\
&+\mathrm{pr}\left(\max_{1\leq i\leq n}|x_i^{\mathrm{T}}(\beta-\hat{\beta}_{\text{lasso+ols}})|\geq 2D,\hat{S}\neq S\right)\\
\leq\ &\mathrm{pr}\left(\max_{1\leq i\leq n}|x_{i,S}^{\mathrm{T}}(\beta_S^0-(\hat{\beta}_{\text{lasso+ols}})_S)+x_{i,S^c}^{\mathrm{T}}\beta_{S^c}^0|\geq 2D\right)+\mathrm{pr}(\hat{S}\neq S)\\
\leq\ &\mathrm{pr}\left(\max_{1\leq i\leq n}|x_{i,S}^{\mathrm{T}}(\beta_S^0-(\hat{\beta}_{\text{lasso+ols}})_S)|\geq D\right)+\mathrm{pr}\left(\max_{1\leq i\leq n}|x_{i,S^c}^{\mathrm{T}}\beta_{S^c}^0|\geq D\right)\\
&+\mathrm{pr}(\hat{S}\neq S)\\
=\ &\mathrm{pr}\left(\max_{1\leq i\leq n}|x_{i,S}^{\mathrm{T}}(\beta_S^0-(\hat{\beta}_{\text{lasso+ols}})_S|\geq D\right)+\mathrm{pr}(\hat{S}\neq S), \qquad\qquad \text{(S1.17)}
\end{aligned}
$$

where the last equality holds because of Condition 10. Using Cauchy-Schwarz inequality and Lemma S2, we have

$$
\max_{1\leq i\leq n}|x_{i,S}^{\mathrm{T}}(\beta_S^0-(\hat{\beta}_{\text{lasso+ols}})_S)|\leq \max_{1\leq i\leq n}||x_{i,S}||_2||\beta_S^0-(\hat{\beta}_{\text{lasso+ols}})_S||_2.
$$

Conditional on $\{\hat{S}=S\}$, the lasso+ols estimator has the following form:

$$
(\hat{\beta}_{\text{lasso+ols}})_S=(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}Y=\beta_S^0+C_{11}^{-1}C_{12}\beta_{S^c}^0+(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon, \quad\text{(S1.18)}
$$

$$
(\hat{\beta}_{\text{lasso+ols}})_{S^c}=0.
$$

Therefore, together with Condition 8,

$$
\begin{aligned}
||\beta_S^0-(\hat{\beta}_{\text{lasso+ols}})_S||_2\ &\leq ||C_{11}^{-1}C_{12}\beta_{S^c}^0||_2+||(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon||_2\\
&=O\left((s/n)^{1/2}\right)+||(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon||_2. \quad\text{(S1.19)}
\end{aligned}
$$

Hence, by Condition 10,

$$\max_{1\le i\le n} |x_{i,S}^{\mathrm{T}}(\beta_S^0 - (\hat\beta_{\text{lasso+ols}})_S)|$$

$$\le\ o(n^{1/4}(s/n)^{1/2}) + \max_{1\le i\le n} ||x_{i,S}||_2||(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon||_2$$

$$=\ o(1) + \max_{1\le i\le n} ||x_{i,S}||_2||(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon||_2. \tag{S1.20}$$

It is easy to show that

$$\max_{1\le i\le n} ||x_{i,S}||_2||(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon||_2 = o_p(1), \tag{S1.21}$$

therefore,

$$\mathrm{pr}\left(\max_{1\le i\le n} ||x_{i,S}||_2||(X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}\epsilon||_2 \ge D\right) \to 0. \tag{S1.22}$$

Hence,

$$\mathrm{pr}\left(\max_{1\le i\le n} |x_i^{\mathrm{T}}(\beta^0 - \hat\beta_{\text{lasso+ols}})| \ge 2D\right) \to 0. \tag{S1.23}$$

Therefore,

$$\mathrm{pr}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\ge1}\left\{e^{c^*t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0-\hat\beta_{\text{lasso+ols}})|\ge t/3}\right\} \le e^{36D^2c^*}\right)$$

$$\ge\ \mathrm{pr}\left(\max_{1\le i\le n} |x_i^{\mathrm{T}}(\beta^0 - \hat\beta_{\text{lasso+ols}})| < 2D\right) \to 1. \tag{S1.24}$$

The above inequality holds, because it is easy to show that

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\sup_{t\ge1}\left\{e^{c^*t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0-\hat\beta_{\text{lasso+ols}})|\ge t/3}\right\} \le e^{36D^2c^*}\right\}$$

$$\supseteq\ \left\{\max_{1\le i\le n} |x_i^{\mathrm{T}}(\beta^0 - \hat\beta_{\text{lasso+ols}})| < 2D\right\}. \tag{S1.25}$$

It is clear that

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{0\leq t\leq 1}\left\{e^{c^*t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0-\hat{\beta}_{\text{lasso+ols}})|\geq t/3}\right\}\leq e^{c^*}.$$

Therefore, with probability going to 1, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sup_{t\geq 0}\left\{e^{c^*t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0-\hat{\beta}_{\text{lasso+ols}})|\geq t/3}\right\}\\
=\quad&\max\left\{\frac{1}{n}\sum_{i=1}^{n}\sup_{0\leq t\leq 1}\left\{e^{c^*t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0-\hat{\beta}_{\text{lasso+ols}})|\geq t/3}\right\},\right.\\
&\left.\frac{1}{n}\sum_{i=1}^{n}\sup_{t\geq 1}\left\{e^{c^*t^2}\mathbb{I}_{|x_i^{\mathrm{T}}(\beta^0-\hat{\beta}_{\text{lasso+ols}})|\geq t/3}\right\}\right\}\\
\leq\quad&\max\left\{e^{c^*},e^{36D^2c^*}\right\}.
\end{aligned}
\tag{S1.26}
$$

Let $C^*=2C_1^*+\max\left\{e^{c^*},e^{36D^2c^*}\right\}$, and combine (S1.26), (S1.15), and (S1.16),

$$\text{pr}\left(\sup_{t\geq 0}\left\{\frac{1}{n}\sum_{i=1}^{n}e^{c^*t^2}\mathbb{I}_{|\hat{\epsilon}_i-\tilde{\epsilon}|\geq t}\right\}\leq C^*\right)\to 1.$$

$\square$

S1.3   Proof of Theorem 3

Proof. First, by Theorem 1 and Theorem 2, both the lasso, $\hat{\beta}_{\text{lasso}}$, and the residual bootstrap lasso, $\hat{\beta}_{\text{rBlasso}}$, have model selection consistency. We can continue our argument by conditioning on $\{\hat{S}=S\}$ and $\{\hat{S}^*_{\text{rBlasso}}=S\}$.

Second, we next show that

$$n^{\frac{1}{2}}u^{\mathrm{T}}(\hat{\beta}_{\text{LPR}}-\beta^0)=n^{-\frac{1}{2}}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon+o_p(1);\tag{S1.27}$$

$$n^{\frac{1}{2}}u^{\mathrm{T}}(\hat{\beta}^*_{\mathrm{rBLPR}} - \hat{\beta}_{\mathrm{lasso+ols}}) = n^{-\frac{1}{2}}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon^* + o_p(1). \tag{S1.28}$$

By definition, $\hat{\beta}_{\mathrm{LPR}}$ is the solution of the following equation:

$$-\frac{1}{n}X^{\mathrm{T}}(Y - X\hat{\beta}_{\mathrm{LPR}}) + \lambda_2\left(\mathbf{0}^{\mathrm{T}}, \left(\hat{\beta}_{\mathrm{LPR},S^c}\right)^{\mathrm{T}}\right)^{\mathrm{T}} = 0.$$

Since $Y = X\beta^0 + \epsilon$, we have

$$\frac{1}{n}X^{\mathrm{T}}X(\hat{\beta}_{\mathrm{LPR}} - \beta^0) - \frac{1}{n}X^{\mathrm{T}}\epsilon + \lambda_2\left(\mathbf{0}^{\mathrm{T}}, \left(\hat{\beta}_{\mathrm{LPR},S^c}\right)^{\mathrm{T}}\right)^{\mathrm{T}} = 0.$$

Simple linear algebra gives

$$C_{\lambda_2}(\hat{\beta}_{\mathrm{LPR}} - \beta^0) = \frac{1}{n}X^{\mathrm{T}}\epsilon - \lambda_2\left(\mathbf{0}^{\mathrm{T}}, \left(\beta^0_{S^c}\right)^{\mathrm{T}}\right)^{\mathrm{T}}.$$

Therefore,

$$\begin{aligned} n^{\frac{1}{2}}u^{\mathrm{T}}(\hat{\beta}_{\mathrm{LPR}} - \beta^0) &= n^{-\frac{1}{2}}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon - \lambda_2 n^{\frac{1}{2}}u^{\mathrm{T}}C_{\lambda_2}^{-1}\left(\mathbf{0}^{\mathrm{T}}, \left(\beta^0_{S^c}\right)^{\mathrm{T}}\right)^{\mathrm{T}} \\ &= n^{-\frac{1}{2}}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon + o_p(1), \end{aligned} \tag{S1.29}$$

where the second equality is due to Condition 11 and $\lambda_2 \propto n^{-1}$ in Condition 7.

Similarly, note that $Y^*_{\mathrm{rboot}} = X\hat{\beta}_{\mathrm{lasso+ols}} + \epsilon^*$, and with probability going to 1, $\hat{\beta}_{\mathrm{lasso+ols},S^c} = 0$, we have

$$n^{\frac{1}{2}}u^{\mathrm{T}}(\hat{\beta}^*_{\mathrm{rBLPR}} - \hat{\beta}_{\mathrm{lasso+ols}}) = n^{-\frac{1}{2}}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon^* + o_p(1). \tag{S1.30}$$

Third, let $U = n^{-1/2}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon$, and $U^* = n^{-1/2}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon^*$. We can show that both U and $(U^* \mid \epsilon)$ converge in distribution to $N(0, \sigma_1^2)$, where $C =$

$X^{\mathrm{T}}X/n$ and

$$\sigma_1^2 = \lim_{n\to\infty} \left(u^{\mathrm{T}}C_{\lambda_2}^{-1}C(C_{\lambda_2}^{-1})^{\mathrm{T}}u\right)\sigma^2.$$

For simplicity, denote

$$\sigma_2^2 = \left(u^{\mathrm{T}}C_{\lambda_2}^{-1}C(C_{\lambda_2}^{-1})^{\mathrm{T}}u\right)\sigma^2.$$

We need to check Linderberg condition for the asymptotic normality. For deriving the asymptotic normality of $U = n^{-1/2}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon$, denote

$$v = n^{-1/2}X\left(C_{\lambda_2}^{-1}\right)^{\mathrm{T}}u = (v_1,\ldots,v_n)^{\mathrm{T}},$$

where $v_k = n^{-1/2}x_k^{\mathrm{T}}\left(C_{\lambda_2}^{-1}\right)^{\mathrm{T}}u$. It is easy to show that

$$\sum_{k=1}^{n} E(v_k\epsilon_k)^2 = \left(\sum_{k=1}^{n}v_k^2\right)\sigma^2 = \left(u^{\mathrm{T}}C_{\lambda_2}^{-1}C(C_{\lambda_2}^{-1})^{\mathrm{T}}u\right)\sigma^2 = \sigma_2^2. \qquad (S1.31)$$

The Linderberg condition holds if for any $\delta > 0$,

$$\frac{1}{\sigma_2^2}\sum_{k=1}^{n}v_k^2 E\left\{\epsilon_k^2 I_{|v_k\epsilon_k|>\delta\sigma_2}\right\} \to 0. \qquad (S1.32)$$

Since the errors ϵ_i are independent and identically distributed subGaussian random variables, it is easy to see that,

$$
\begin{aligned}
\frac{1}{\sigma_2^2}\sum_{k=1}^{n}v_k^2 E\left\{\epsilon_k^2 I_{|v_k\epsilon_k|>\delta\sigma_2}\right\} &\leq \frac{1}{\sigma^2}\max_{1\leq k\leq n} E\left\{\epsilon_k^2 I_{|v_k\epsilon_k|>\delta\sigma_2}\right\} \\
&\leq \frac{1}{\sigma^2}E\left\{\epsilon_1^2 I_{|\epsilon_1|>\frac{\delta}{\max_{1\leq k\leq n}|v_k|/\sigma_2}}\right\}, \\
&= o(1), \qquad\qquad\qquad (S1.33)
\end{aligned}
$$

where the last equality is because Condition 11. That is,

$$\max_{1 \leq k \leq n} |v_k|/\sigma_2 = n^{-\frac{1}{2}} \max_{1 \leq k \leq n} \left|u^{\mathrm{T}}C_{\lambda_2}^{-1}x_k\right| / \left\{\sigma^2 u^{\mathrm{T}}C_{\lambda_2}^{-1}C(C_{\lambda_2}^{-1})^{\mathrm{T}}u\right\}^{1/2} = o(1).$$

Finally, we prove the asymptotic normality of $U^* = n^{-1/2}u^{\mathrm{T}}C_{\lambda_2}^{-1}X^{\mathrm{T}}\epsilon^*$, given ϵ. By Lemma S3, ϵ_i^* are conditionally (given ϵ) independent and identically distributed subGaussian random variables, with mean 0 and variance $(\sigma^*)^2$. Similar arguments lead to the same asymptotic normality of U^*, given ϵ, as those for U, as long as $\sigma^* \to_p \sigma$. The reminder of the proof is to show that $\sigma^* \to_p \sigma$.

Note that

$$(\sigma^*)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\epsilon}_i - \tilde{\epsilon})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[x_i^{\mathrm{T}}(\beta^0 - \hat{\beta}_{\text{lasso+ols}}) + \epsilon_i - \bar{\epsilon}\right]^2.$$

By Strong Law of Large Number, we have

$$\frac{1}{n-1}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2 \to \sigma^2, \text{ almost surely.} \tag{S1.34}$$

Since

$$\frac{1}{n-1}\sum_{i=1}^{n}\left[x_i^{\mathrm{T}}(\beta^0 - \hat{\beta}_{\text{lasso+ols}})\right]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[x_{i,S}^{\mathrm{T}}(\beta_S^0 - (\hat{\beta}_{\text{lasso+ols}})_S) + x_{i,S^c}^{\mathrm{T}}\beta_{S^c}^0\right]^2$$

$$\leq 2\left\{\frac{1}{n-1}\sum_{i=1}^{n}\left[x_{i,S}^{\mathrm{T}}(\beta_S^0 - (\hat{\beta}_{\text{lasso+ols}})_S)\right]^2 + \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,S^c}^{\mathrm{T}}\beta_{S^c}^0)^2\right\}$$

$$\leq \frac{2n}{n-1}\left\{\max_{1 \leq i \leq n}\left[x_{i,S}^{\mathrm{T}}(\beta_S^0 - (\hat{\beta}_{\text{lasso+ols}})_S)\right]^2 + (\beta_{S^c}^0)^{\mathrm{T}}C_{22}(\beta_{S^c}^0)\right\}$$

$$= o_p(1), \tag{S1.35}$$

where the last equality holds because of (S1.20), (S1.21), and Condition 11.

Combining (S1.34) and (S1.35), we have

$$(\sigma^*)^2 \to_p \sigma^2.$$

$\square$

S1.4   Proof of Lemma S2

Proof. Under Conditions 1 – 7, and using Theorem 1, the lasso, $\hat{\beta}_{\text{lasso}}$, has model selection consistency, that is,

$$\text{pr}(\hat{S} = S) = 1 - o(e^{-n^{c_2}}) \to 1.$$

Conditional on $\{\hat{S} = S\}$, the lasso+ols estimator has the following form:

$$(\hat{\beta}_{\text{lasso+ols}})_S = (X_S^{\mathrm{T}} X_S)^{-1} X_S^{\mathrm{T}} Y = \beta_S^0 + C_{11}^{-1} C_{12} \beta_{S^c} + (X_S^{\mathrm{T}} X_S)^{-1} X_S^{\mathrm{T}} \epsilon;$$

$$(\hat{\beta}_{\text{lasso+ols}})_{S^c} = 0.$$

Therefore,

$$||\hat{\beta}_{\text{lasso+ols}} - \beta^0||_\infty \le ||C_{11}^{-1} C_{12} \beta_{S^c}^0||_\infty + ||(X_S^{\mathrm{T}} X_S)^{-1} X_S^{\mathrm{T}} \epsilon||_\infty + ||\beta_{S^c}^0||_\infty. \quad \text{(S1.36)}$$

By Condition 8, we have $||C_{11}^{-1} C_{12} \beta_{S^c}^0||_\infty = o(n^{(c_1-1)/2})$. Condition 6 gives $||\beta_{S^c}^0||_\infty \le M n^{-(1+c_1)/2}$. Since $(X_S^{\mathrm{T}} X_S)^{-1} X_S^{\mathrm{T}} \epsilon$ are subGaussian random variables, with covariance matrix $\sigma^2 C_{11}^{-1}/n$, it is not hard to show that

$$\text{pr}\left(||(X_S^{\mathrm{T}} X_S)^{-1} X_S^{\mathrm{T}} \epsilon||_\infty \le M n^{\frac{c_1-1}{2}} \right) \to 1.$$

Therefore,

$$\mathrm{pr}\left(||\hat{\beta}_{\text{lasso+ols}} - \beta^0||_\infty \leq 2Mn^{\frac{c_1-1}{2}}\right) \to 1.$$

$\square$

S2 Examples related to Condition 8

We provide three examples of design matrices, which satisfy or do not satisfy Condition 8.

Example 1. Orthogonal design. X is orthogonal such that $X^{\mathrm{T}}X/n$ is an identity matrix. In this case, C_{11} and C_{22} are identity matrices, and C_{12} is a zero matrix. As $\|\beta^0_{S^c}\|_\infty = o\left(n^{-1/2}\right)$, we have

$$\left\|C_{11}^{-1}C_{12}\beta^0_{S^c}\right\|_\infty = 0,$$

$$\left\|n^{\frac{1}{2}}(C_{21}C_{11}^{-1}C_{12} - C_{22})\beta^0_{S^c}\right\|_\infty = \left\|n^{\frac{1}{2}}\beta^0_{S^c}\right\|_\infty = O(1).$$

Thus, this example satisfies Condition 8.

Example 2. Exponential decay. In this example, X has the following

pattern:

$$\frac{1}{n}X^{\mathrm{T}}X = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & 1 & \rho & \cdots & \rho^{p-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \cdots & 1 \end{pmatrix}.$$

In this case,

$$C_{11} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{s-1} \\ \rho & 1 & \rho & \cdots & \rho^{s-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{s-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{s-1} & \rho^{s-2} & \rho^{s-3} & \cdots & 1 \end{pmatrix}.$$

Using mathematical induction, we can prove that

$$(1 - \rho^2)C_{11}^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1 + \rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix},$$

$$
C_{12} = \begin{pmatrix} \rho^s & \rho^{s+1} & \cdots & \rho^{p-1} \\ \rho^{s-1} & \rho^s & \cdots & \rho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho^2 & \cdots & \rho^{p-s} \end{pmatrix}.
$$

Then,

$$
C_{11}^{-1}C_{12} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \rho - \rho^3 & \rho^2 - \rho^4 & \cdots & \rho^{p-s} - \rho^{p-s+2} \end{pmatrix}.
$$

As $\|\beta_{S^c}^0\|_\infty = o\left(n^{-1/2}\right)$, we have

$$
\left\|n^{\frac{1}{2}}C_{11}^{-1}C_{12}\beta_{S^c}^0\right\|_\infty \le n^{\frac{1}{2}}(\rho + \rho^2 - \rho^{p-s+1} - \rho^{p-s+2})\left\|\beta_{S^c}^0\right\|_\infty = O(1),
$$

$$
\left\|n^{\frac{1}{2}}(C_{21}C_{11}^{-1}C_{12} - C_{22})\beta_{S^c}^0\right\|_\infty \le \left\|n^{\frac{1}{2}}C_{21}C_{11}^{-1}C_{12}\beta_{S^c}^0\right\|_\infty + \left\|n^{\frac{1}{2}}C_{22}\beta_{S^c}^0\right\|_\infty,
$$

$$
\left\|n^{\frac{1}{2}}C_{21}C_{11}^{-1}C_{12}\beta_{S^c}^0\right\|_\infty \le n^{\frac{1}{2}}(\rho^2 + \rho^3 - \rho^{p-s+2} - \rho^{p-s+3})\left\|\beta_{S^c}^0\right\|_\infty = O(1),
$$

$$
\left\|n^{\frac{1}{2}}C_{22}\beta_{S^c}^0\right\|_\infty < \frac{n^{\frac{1}{2}}}{1 - \rho}\left\|\beta_{S^c}^0\right\|_\infty = O(1).
$$

Thus, this example satisfies Condition 8.

Example 3. Equal correlation. The design matrix X satisfies

$$\frac{1}{n}X^{\mathrm{T}}X = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}.$$

In this case,

$$C_{11} = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}, \quad C_{12} = \begin{pmatrix} \rho & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & \rho \end{pmatrix}.$$

It is easy to show that $\mathbf{1} = (1, 1, \cdots, 1)^{\mathrm{T}}$ is an eigenvector of C_{11}, hence, it is

also an eigenvector of C_{11}^{-1}. Let Sum_{S^c} be the sum of elements in $\beta_{S^c}^0$. Then,

$$\left\| C_{11}^{-1}C_{12}\beta_{S^c}^0 \right\| = \frac{\rho}{1 + \rho(s-1)} \left| Sum_{S^c} \right|.$$

As we do not assume a bound for Sum_{S^c}, this example does not always satisfy

Condition 8.

S3   Examples related to Condition 11

When the correlation between covariates satisfies $\mathrm{cor}(X_i, X_j) = \rho^{|i-j|}$, with

$\rho < 1/5$, Condition 11 holds.

In this case, $p \leq n$, and

$$(\beta^0_{S^c})^T C_{22}(\beta^0_{S^c}) = \sum_{\substack{s<i\leq p \\ s<j\leq p}} (C_{22})_{ij}\beta^0_i\beta^0_j = \sum_{\substack{s<i\leq p \\ s<j\leq p}} \rho^{|i-j|}\beta^0_i\beta^0_j = o(1).$$

Lemma S4. *For any $p \times 1$ vector u, v, and $p \times p$ symmetric matrix A, we have*

$\mu_p(A) \leq \frac{u^T A v}{u^T v} \leq \mu_1(A)$, *where $\mu_p(A)$ and $\mu_1(A)$ are the smallest and largest*

eigenvalues of A, respectively.

From the above lemma,

$$\mu_p(C^{-1}_{\lambda_2}) \leq \frac{u^T C^{-1}_{\lambda_2} x_k}{u^T x_k} \leq \mu_1(C^{-1}_{\lambda_2}).$$

Assume that $\rho < \frac{1}{5}$, by Gershgorin circle theorem, there exists a $\delta > 0$, such that

$$3 > \mu_1(C_{\lambda_2}) \geq \mu_2(C_{\lambda_2}) \geq \ldots \mu_p(C_{\lambda_2}) > \delta > 0.$$

Then we have

$$\tfrac{1}{\delta} > \mu_1(C^{-1}_{\lambda_2}) \geq \mu_2(C^{-1}_{\lambda_2}) \geq \ldots \mu_p(C^{-1}_{\lambda_2}) > \tfrac{1}{3}.$$

Thus,

$$\left|\frac{u^T C^{-1}_{\lambda_2} x_k}{u^T x_k}\right| \leq \tfrac{1}{\delta}, \quad \max_{1\leq k\leq n} |u^T C^{-1}_{\lambda_2} x_k| \leq \tfrac{1}{\delta} \max_{1\leq k\leq n} |u^T x_k|.$$

Therefore, Condition 11 is guaranteed by assuming

$$\max_{1\leq k\leq n} |u^T x_k| = o(\sqrt{n}), \quad u^T_{S^c}\beta^0_{S^c} = o(\sqrt{n}).$$

S4   Simulation Details

This section is organized as follows. Subsection S4.1 introduces the simulation setups. Subsection S4.2 studies the impact of the partial ridge tuning parameter λ_2 on the coverage probabilities and the mean interval lengths of the confidence intervals constructed by the rBLPR and pBLPR methods. In Subsection S4.3, we compare the performance of the rBLPR and pBLPR methods with that of the bootstrap lasso+ols method. Subsection S4.4 presents the comparison results of rBLPR, pBLPR, LDPE, JM, and BLDPE. We investigate the robustness of the rBLPR and pBLPR methods by varying signal-to-noise ratios in Subsection S4.5. In Subsection S4.6, we present the comparison results of different methods under a misspecified model.

We use R package "glmnet" to compute the lasso solution path and select the tuning parameter, λ_1, by 5-fold Cross Validation cv(lasso+ols); see Algorithm S3 for details. The number of replications in the bootstrap is 1000, that is, $B = 1000$.

S4.1   Simulation setups

We consider two generative models for data simulation.

(1) Linear regression model. The simulated data are drawn from the linear

model:

$$y_i = x_i^{\mathrm{T}}\beta^0 + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n. \tag{S4.1}$$

We fix $n = 200$ and $p = 500$. We generate the design matrix X in three scenarios, using the R package "mvtnorm". In Scenarios 1 and 2, we choose σ such that the Signal-to-Noise-Ratio equals ten, that is, $\mathrm{SNR} = ||X\beta^0||_2^2/(n\sigma^2) = 10$. We also examine other values of n, p and σ, but they are not reported here because the conclusions are similar.

Scenario 1 (Normal): Predictor vectors x_i, for $i = 1, \ldots, n$, are generated independently from a multivariate normal distribution $N(\mathbf{0}, \Sigma)$ with covariance matrix Σ. We consider three types of Σ, following the setup in (Dezeure et al., 2014).

$$\text{Toeplitz}: \quad \Sigma_{ij} = \rho^{|i-j|}, \text{ with } \rho = 0.5, 0.9,$$

$$\text{Exponential decay}: \quad (\Sigma^{-1})_{ij} = \rho^{|i-j|}, \text{ with } \rho = 0.5, 0.9,$$

$$\text{Equal correlation}: \quad \Sigma_{ij} = \rho, \text{ with } \rho = 0.5, 0.9.$$

Scenario 2 (t_2): Predictor vectors x_i, for $i = 1, \ldots, n$, are generated independently from a multivariate t distribution, with two degrees of freedom, $t_2(\mathbf{0}, \Sigma)$, where Σ is a Toeplitz-type matrix: $\Sigma_{ij} = \rho^{|i-j|}$, with $\rho = 0.5, 0.9$.

Scenario 3 (fMRI data): A 200×500 design matrix X is generated by random sampling, without replacement, from the real 1750×2000 design matrix

in the fMRI data (see Section 5 for more details on this data). Every column of X is normalized to have zero mean and unit variance, and we choose σ, such that SNR $= 1, 5$ or 10.

We also consider two cases to generate β^0.

Case 1 (hard sparsity): β^0 has 10 nonzero elements whose indices are randomly sampled, without replacement, from $\{1, \ldots, p\}$, and whose values are generated from $U[1/3, 1]$, a uniform distribution on the interval $[1/3, 1]$. The remaining 490 elements are set to be zero.

Case 2 (weak sparsity): The setup is similar to that in (Zhang & Zhang, 2014). β^0 has 10 large elements whose indices are randomly sampled, without replacement, from $\{1, \ldots, p\}$, and whose values are generated from a normal distribution, $N(1, 0.001)$. The remaining 490 elements decay at a rate of $1/(j + 3)^2$, that is, $\beta_j^0 = 1/(j + 3)^2$.

The values of x_i and β^0 are generated once and then kept fixed. The average absolute correlations among the covariates with large coefficients are 0.08, 0.06, and 0.47 for the normal design with a Toeplitz type covariance matrix, normal design with an Exponential decay type covariance matrix, and t_2 design with a Toeplitz type covariance matrix, respectively. After $X = (x_1^{\mathrm{T}}, \ldots, x_n^{\mathrm{T}})^{\mathrm{T}}$ and β^0 are generated, we simulate $Y = (y_1, \ldots, y_n)^{\mathrm{T}}$ from the linear model (S4.1) by generating independent error terms for 1000 replications. Then we construct

confidence intervals for each regression coefficient, and compute their coverage probabilities and mean interval lengths.

(2) Misspecified linear model. The simulation is based on a real data set: fMRI (see Section 5 for more details). Let X and Y^f (distinguished from the simulated response Y below) denote the design matrix (with $n = 1750$ observations and $p = 2000$ predictors) and the actual response (of the ninth voxel) in the fMRI data set. The original design matrix in the fMRI data set has 10921 predictors, but we first removed the predictors with variances no more than $1e^{-4}$ and selected $p = 2000$ predictors that have the largest absolute correlations with the response. We compute the lasso+ols estimator $\beta^f_{\text{lasso+ols}}$ (selecting the tuning parameter λ_1 by 5-fold cross validation on lasso+ols):

$$\beta^f_{\text{lasso}} = \arg\min_{\beta} \left\{ \frac{1}{2n} ||Y^f - X\beta||_2^2 + \lambda_1 ||\beta||_1 \right\},$$

$$\beta^f_{\text{lasso+ols}} = \arg\min_{\beta:\beta_j=0,\ j\notin S} \frac{1}{2n} ||Y^f - X\beta||_2^2,$$

where $S = \{j : (\beta^f_{\text{lasso}})_j \neq 0\}$ is the set of relevant predictors. We re-ordered the predictors by sorting the values of $\beta^f_{\text{lasso+ols}}$ in a decreasing order, such that the first four predictors corresponds to the largest 4 nonzero elements of $\beta^f_{\text{lasso+ols}}$. Then we generate the simulated response $Y = (y_1, \ldots, y_n)^\mathsf{T}$ from the following model:

$$y_i = E(y_i|x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \tag{S4.2}$$

$$E(y_i|x_i) = x_i^{\mathrm{T}}\beta_{\text{lasso+ols}}^f + \sum_{j=1}^{4}\alpha_j x_{ij}^2 + \sum_{1 \le j < k \le 4}\alpha_{jk}x_{ij}x_{ik},$$

where α_j, for $j = 1,..,4$, and α_{jk}, for $1 \le j \ne k \le 4$, are independently generated from a uniform distribution, $U(0, 0.1)$. The values of α_j and α_{jk} are generated once and then kept fixed. We set σ such that $\text{SNR} = \sum_{i=1}^{n}E(y_i|x_i)^2/(n\sigma^2) = 0.5,\ 1$ or 5. Since the quadratic and interaction terms are not included in the design matrix $X = (x_1^{\mathrm{T}}, \ldots, x_n^{\mathrm{T}})^{\mathrm{T}}$, a linear model between Y and X, $y_i = x_i^{\mathrm{T}}\beta^0 + \epsilon_i$, is misspecified. In this misspecified linear model, the parameter vector β^0 we are interested in is the projection coefficient of $E(Y \mid X)$ onto the subspace spanned by the relevant predictors:

$$\beta_S^0 = (X_S^{\mathrm{T}}X_S)^{-1}X_S^{\mathrm{T}}E(Y \mid X); \quad \beta_{S^c}^0 = \mathbf{0}.$$

Again, in order to compute the coverage probabilities and mean confidence interval lengths, we generate Y by simulating independent error terms ϵ_i's in equation (S4.2) for 1000 times. The confidence level is set to 95%.

S4.2 Selection of the partial ridge tuning parameter λ_2

We first study the effects of the partial ridge tuning parameter λ_2 on the performance of the bootstrap LPR methods (rBLPR and pBLPR). Figure S1 compares the coverage probabilities and mean confidence interval lengths produced by different values of λ_2, based on the following simulation setup: the predictors

are generated from a Normal distribution as in Scenario 1, with a Toeplitz type covariance matrix corresponding to $\rho = 0.5$, and β^0 is hard sparse. We also compare the results for other simulation setups, but the conclusions are essential the same and are not reported here. In order to give a better view, in the following figures without further emphasizing, we sort the elements of β^0 in a decreasing order (in absolute value) and only plot the results for the largest 25 elements of β^0. We can see that both the coverage probabilities and mean confidence interval lengths are very stable with respect to a large range of λ_2 values. Our simulation experiments show that fixing λ_2 at $1/n$ works quite well for a wide range of noise levels. For the sake of simplicity, we take $\lambda_2 = 1/n$ in this study, but acknowledging that further research is needed to find a more systematic approach for selecting λ_2.

S4.3 Comparison of bootstrap lasso+ols and bootstrap LPR methods

We now compare the performance of the rBLPR and pBLPR methods with that of the bootstrap lasso+ols method. Figure S2 shows the comparison results in terms of coverage probabilities and mean confidence interval lengths for the Normal distributed design matrix in Scenario 1 with a Toeplitz type covariance matrix corresponding to $\rho = 0.5$ or 0.9, and for β^0 with hard or weak sparsity. For other design matrices, the conclusions are similar. We see that the rBLPR

and pBLPR have similar performance, while the latter performs slightly better, therefore, we only present the results for pBLPR in the following contents. In the hard sparsity cases, all the methods work very well. In the weak sparsity cases, however, the bootstrap lasso+ols method gives very poor coverage probabilities (less than 50% for 95% confidence intervals) for the small, but nonzero elements of β^0. This is because these elements are seldom selected by the lasso and, therefore, a large proportion of their bootstrap lasso+ols estimates are zero, producing noncoverage confidence intervals, such as $[0, 0]$. The pBLPR method dramatically improve the performance of the bootstrap lasso+ols method. It produces promising coverage probabilities, at the price of slightly increasing the confidence interval lengths. However, for medium-size components of β^0, pBLPR has problems covering true values even when design matrices are generated from a normal distribution (The coverage probability for one particular such component is only 63%). This is because the lasso cannot identify these medium-size components with high probability.

S4.4   Comparison of bootstrap LPR and de-sparsified methods

Figures S3, S4, and S5 show the comparison results of pBLPR, LDPE, JM, and BLDPE, under a Normal design matrix with a Toeplitz type covariance matrix, with an Equi.corr type covariance matrix, and a t_2 distributed design matrix with

a Toeplitz type covariance matrix, respectively. From Figure S3, we see that the pBLPR gives promising results. Overall, it has good performance for large and small components of β^0, and in some cases it outperforms LDPE and JM, by producing confidence intervals with, on average, 50% shorter lengths (see the comparison results in Tables S1, S2, S3, and S4, which show the mean coverage probabilities and the mean lengths of the confidence intervals for large coefficients and small including zero coefficients, respectively). When the predictors have high correlations (see the results for $\rho = 0.9$), pBLPR gives confidence intervals with higher coverage probabilities for large coefficients, and for small and zero coefficients, it gives shorter confidence interval lengths with good coverage probabilities. Following the evaluation scheme in (Dezeure et al., 2014), we also show more details of the comparison results in Figures S8 – S11, which display the $1,000$ confidence intervals and their empirical coverage of the true coefficients (blue line), for five methods: pBLPR, rBLPR, LDPE, JM, and BLDPE. The black and red colors in these figures are used to indicate whether confidence intervals cover the truth or not. The first 10 coefficients are the 10 largest (in absolute values) nonzero coefficients. For each method, the 15 zero (Figures S8 and S9) or small, but nonzero (Figures S10 and S11) coefficients shown are those with the worst empirical coverage probabilities. The numbers above confidence intervals are the empirical coverage probabilities in percentages. These figures

clearly show the advantegeous performance of the pBLPR in constructing confidence intervals for a broad range of coefficients.

Under a Normal design matrix with an Equi.corr type covariance matrix (see Figure S4), the JM does not work well when $\rho = 0.9$, because it dramatically overestimates the noise variance. Our method also has unsatisfactory performance in terms of coverage probabilities for large coefficients, because the lasso cannot correctly select the large predictors due to the strong collinearity among the predictors. Under a t_2 design matrix, Figure S5 shows that no methods perform well, leaving large space for improvement. For other covariance structures, the comparison results are shown in Figures S6 and S7.

The bootstrap version LDPE method (BLDPE) does improve the performance of LDPE. It has the best coverage probabilities among the considered methods, but its confidence interval lengths are close to or slightly shorter than the better one of LDPE and JM and, hence, larger than the pBLPR method.

The selection frequency of each coefficient in the 1000 simulation runs is shown in Figure S16 and S17. Although some important coefficients are missed by the lasso, their empirical coverage probabilities are still good. This maybe because the bootstrap runs help to correct the selection and the LPR estimator is no longer sparse due to the partial ridge penalty.

The comparison results for rBLPR can be found in Figures S18 to S22.

In addition, we also compare the bias, standard deviation (SD) and root-mean square error (RMSE) of the de-sparsified estimators and the LPR estimator, in order to see to what extent these methods reduce the lasso bias. Figure S12 shows the results. We found that, compared with LDPE and JM, the LPR estimator has smaller biases (99% and 72% smaller, on average, than that of LDPE and JM, respectively) for almost all coefficients, but the LPR estimator has larger SDs (30% and 62% larger, on average, than that of LDPE and JM, respectively) for large coefficients. Overall, LPR has 60% smaller RMSE than LDPE, but 42% larger RMSE than JM. Another interesting finding is that although de-sparsified estimators can dramatically decrease the biases of the lasso by more than 40% for large β_j^*'s, they can increase the biases more than twice for small, or zero β_j^*'s.

S4.5 Robustness to signal-to-noise ratios

Figure S13 shows the comparison results under varying signal-to-noise ratios (SNRs). We can see that the coverage performance of the de-sparsified methods is more robust to SNR changes. On the other hand, the pBLPR method works well when SNR is high (say, larger than 5), but it may have low coverage probabilities for nonzero coefficients when SNR is low. This is reasonable because the lasso cannot identify nonzero coefficients with high probability when SNR is

low. The pBLPR method depends more on the model selection performance of the lasso. However, it has much shorter (more than 20%, on average) confidence interval lengths for zero coefficients even when SNR is low.

S4.6 Comparison of different methods under the misspecified model

Figure S14 compares the performance of pBLPR, LDPE, JM, and BLDPE under the misspecified linear model. The pBLPR performs slightly worse than the other three methods in terms of coverage probabilities, but it produces more than 50%, on average, shorter confidence intervals.

S5 Figures and Tables

List of Figures

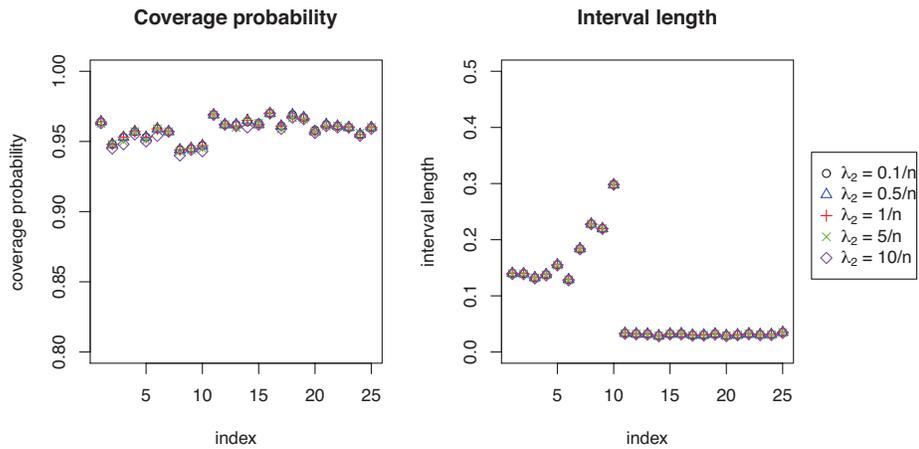**Coverage probability**   **Interval length**

Figure S1: The effects of λ_2 on coverage probabilities and mean confidence interval lengths. The predictors are generated from a Normal distribution in Scenario 1 with a Toeplitz type covariance matrix, and $\rho = 0.5$. The coefficient vector β^0 is hard sparse.
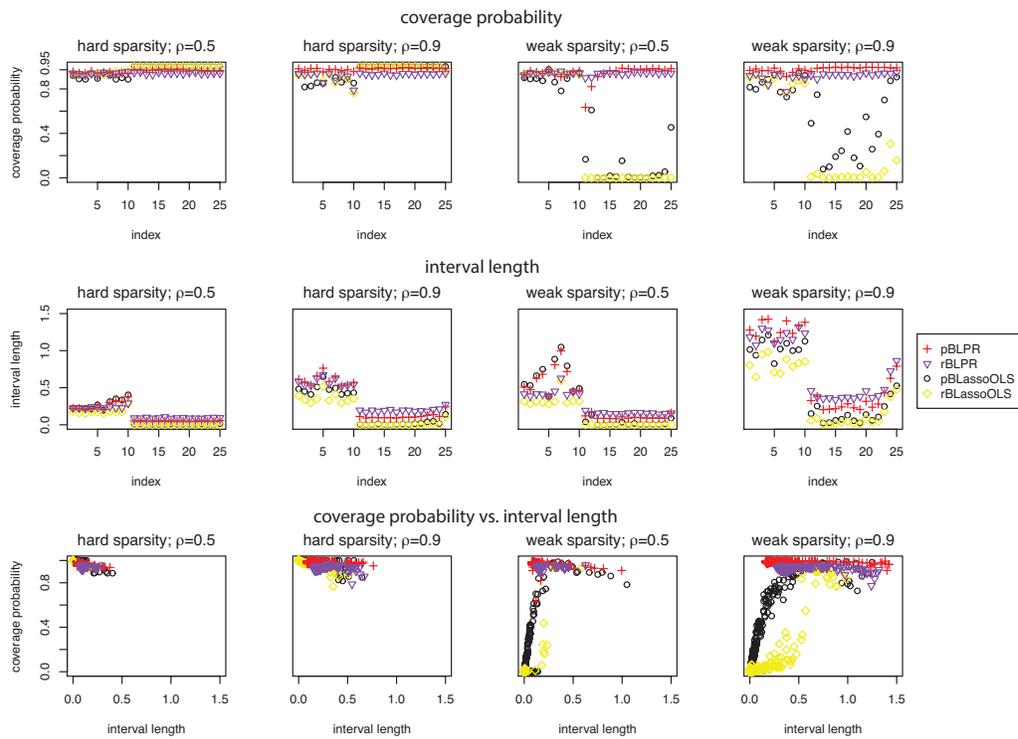
Figure S2: Comparison of coverage probabilities (the first row) and mean confidence interval lengths (the second row) produced by four methods: rBLPR, pBLPR, residual bootstrap lasso+OLS (denoted by rBLassoOLS) and paired bootstrap lasso+OLS (denoted by pBLassoOLS). The third row shows the coverage probabilities v.s. mean interval lengths. The design matrix is generated from a Normal distribution with a Toeplitz type covariance matrix.
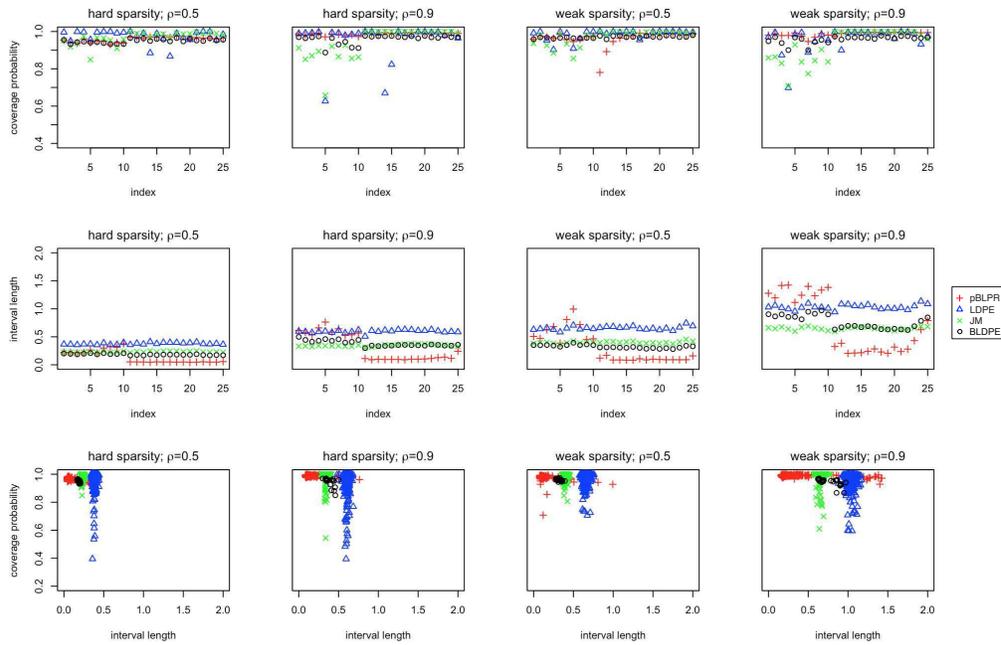
Figure S3: Comparison of coverage probabilities (the first row) and mean interval lengths (the second row) produced by pBLPR, LDPE, JM, and BLDPE. The third row shows the coverage probabilities v.s. mean interval lengths. The design matrix is generated from a Normal distribution with a Toeplitz type covariance matrix.
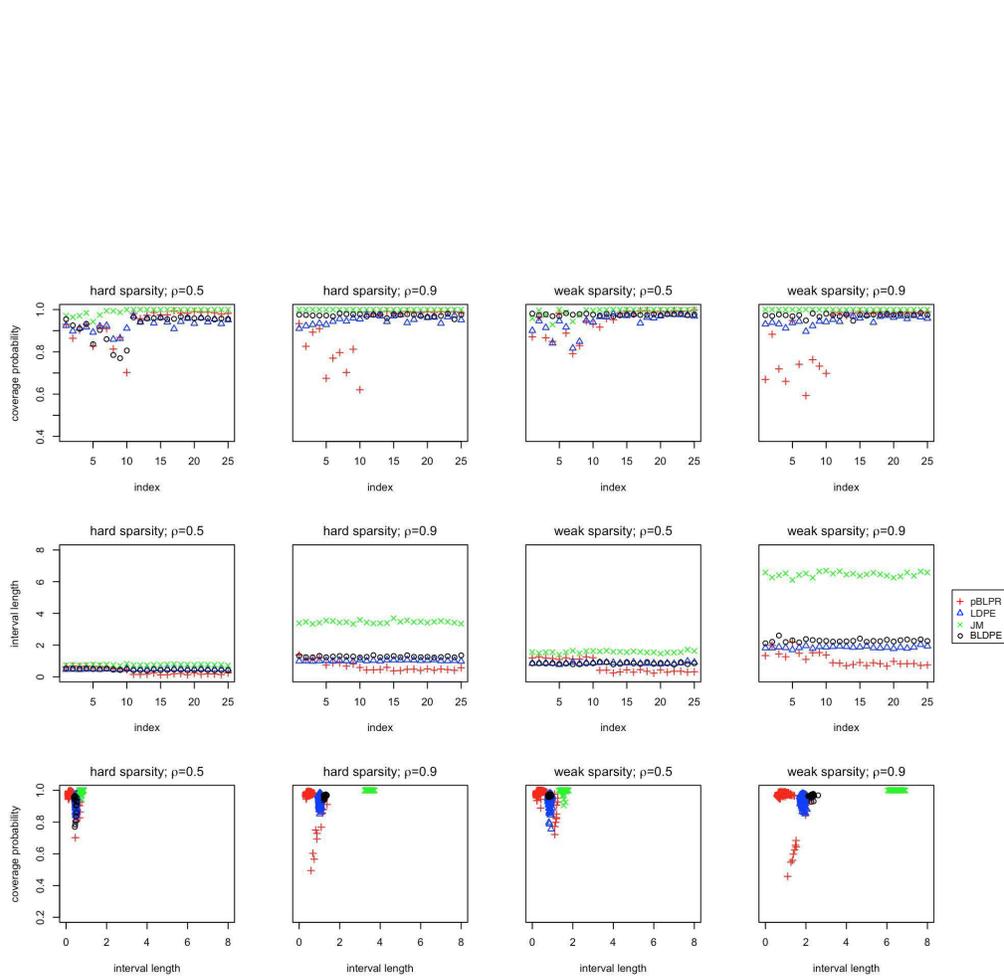
Figure S4: See caption of Figure S3 with the only difference being that the covariance matrix is an Equi.corr type.
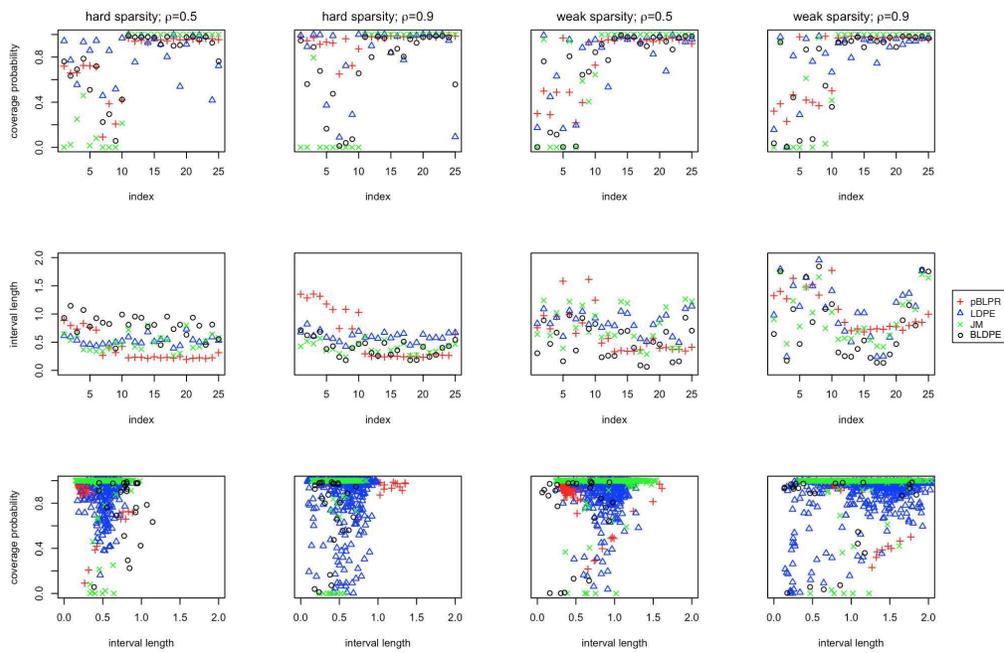
Figure S5: See caption of Figure S3 with the only difference being the type of design matrix. In this plot, the design matrix is generated from t_2 distribution with a Toeplitz type covariance matrix.
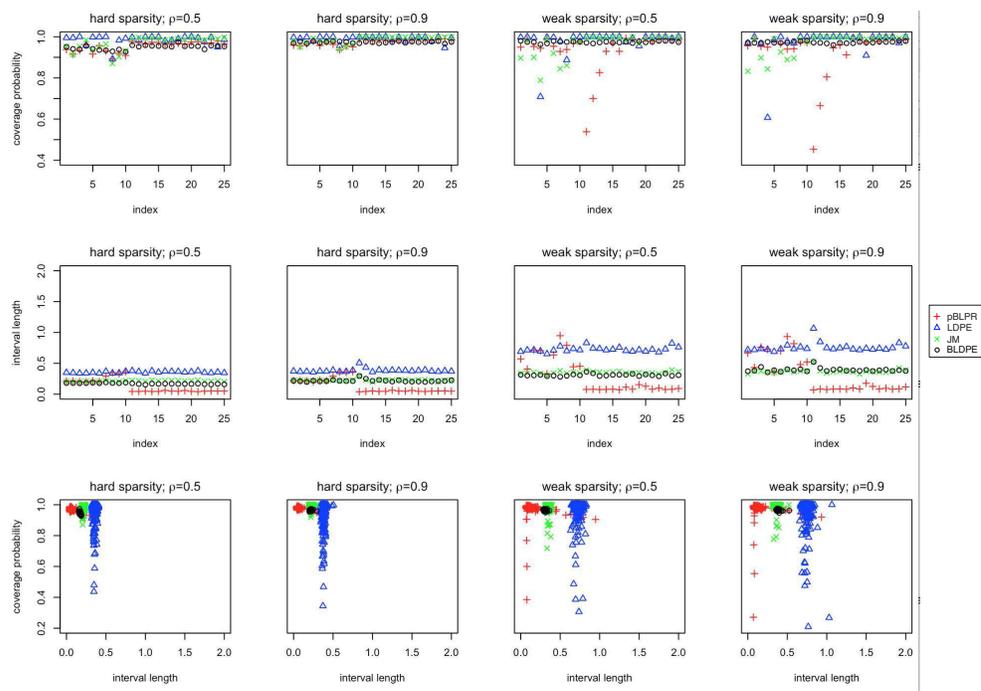
Figure S6: See caption of Figure S3 with the only difference being that the covariance matrix is Exp.decay type.
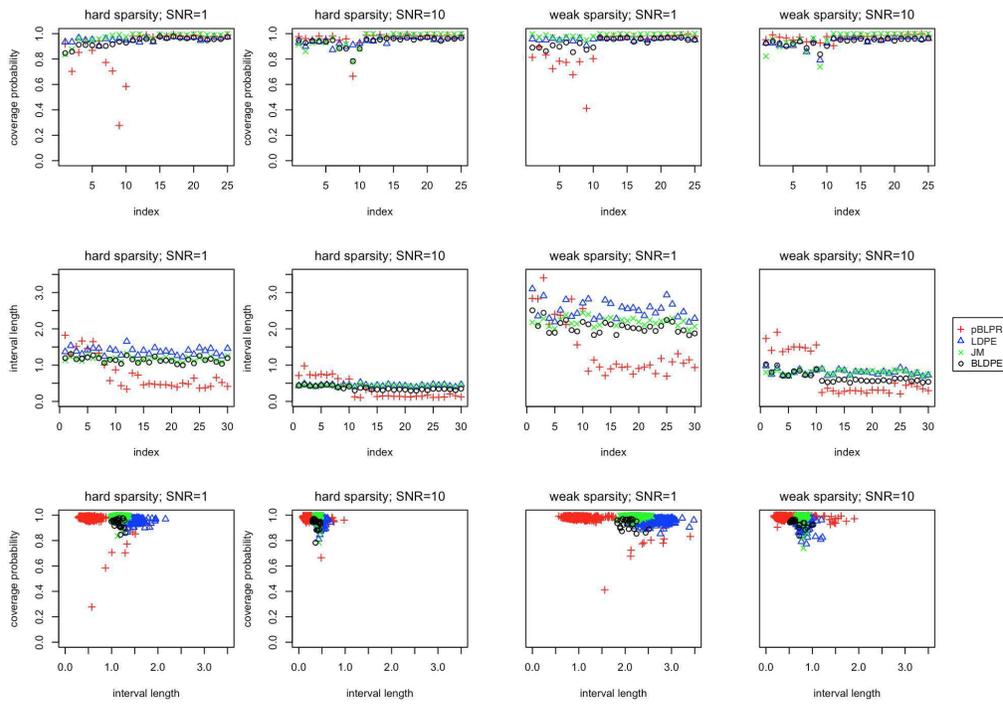
Figure S7: See caption of Figure S3 with the only difference being that the design matrix is generated from the fMRI data.
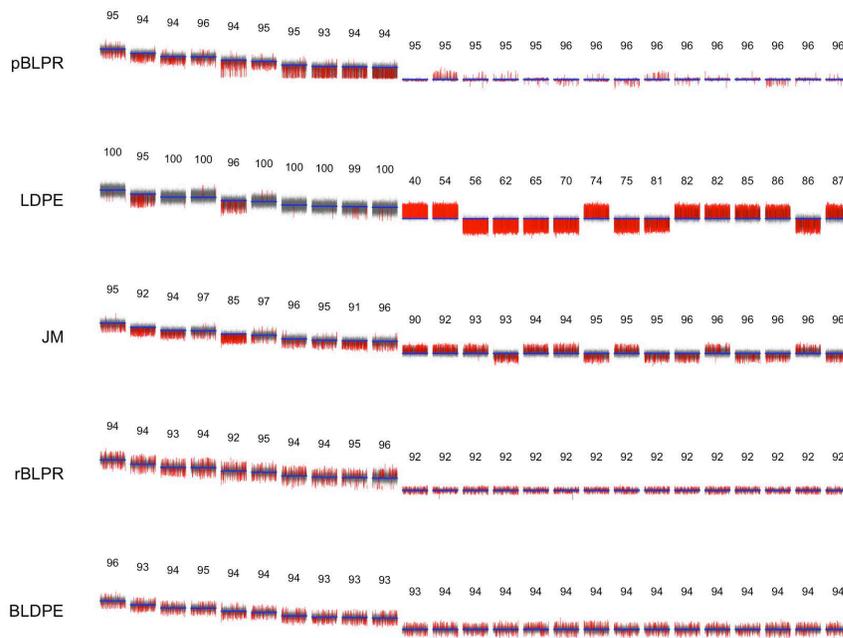
Figure S8: $1,000$ confidence intervals and their empirical coverage of the true coefficients (blue line). Black confidence intervals cover the truth, whereas red confidence intervals do not. The first 10 coefficients are the largest 10 (non-zero). The remaining 15 coefficients shown are those with the worst coverage for that method. The numbers above the intervals are the empirical coverage probabilities in percentages. This plot is for hard sparsity and a Normal design matrix with a Toeplitz type covariance matrix, and $\rho = 0.5$.

Figure S9: See caption of Figure S8 with the only difference being $\rho = 0.9$.
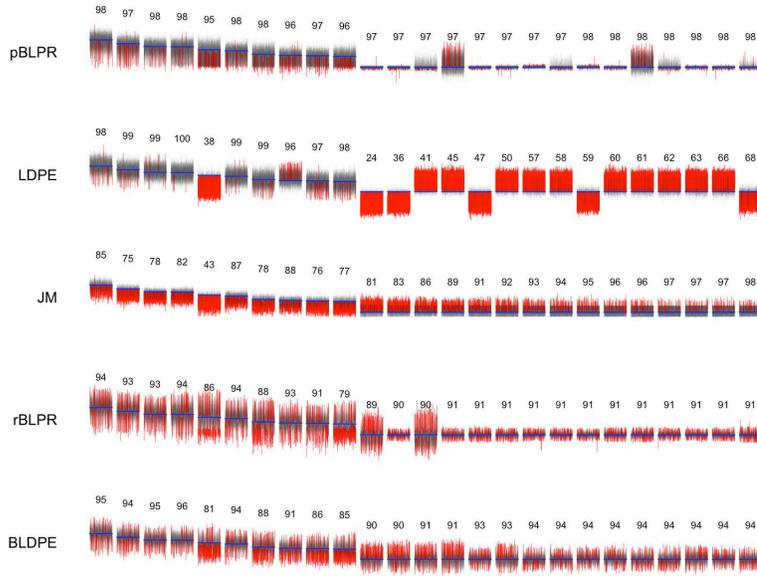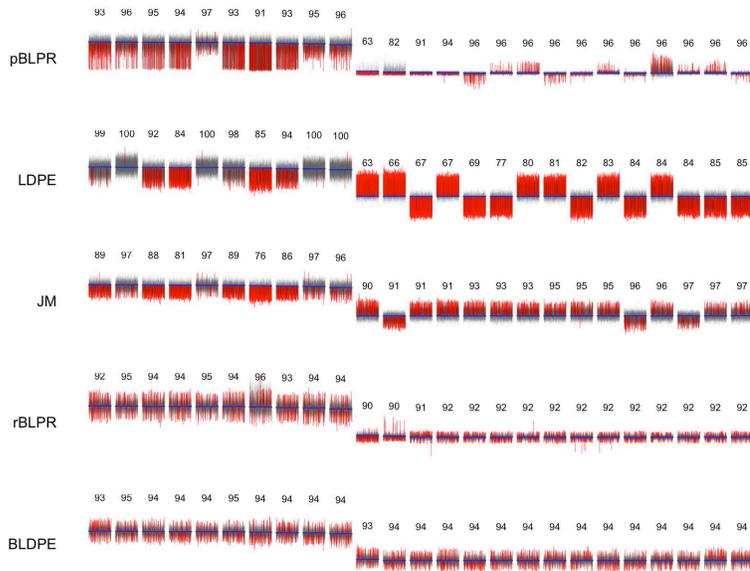

Figure S10: See caption of Figure S8 with the only difference being weak sparsity.
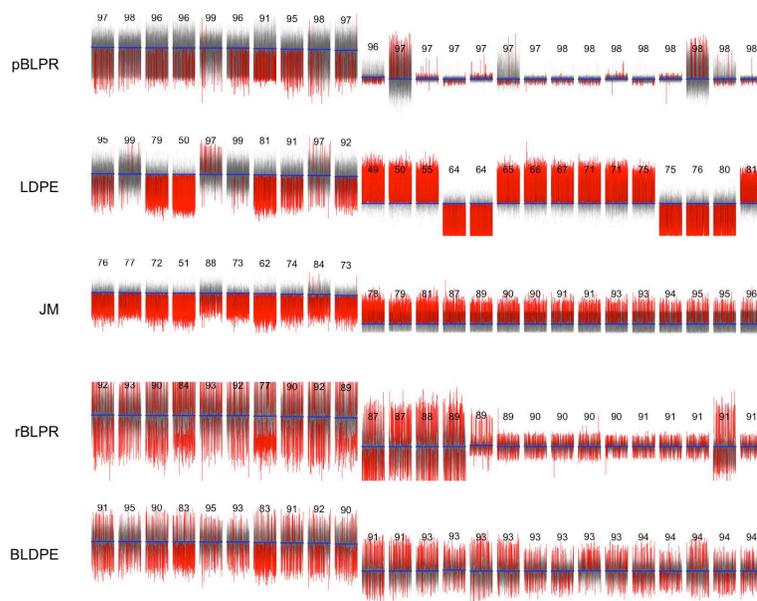
Figure S11: See caption of Figure S8 with the only differences being weak sparsity and $\rho = 0.9$.
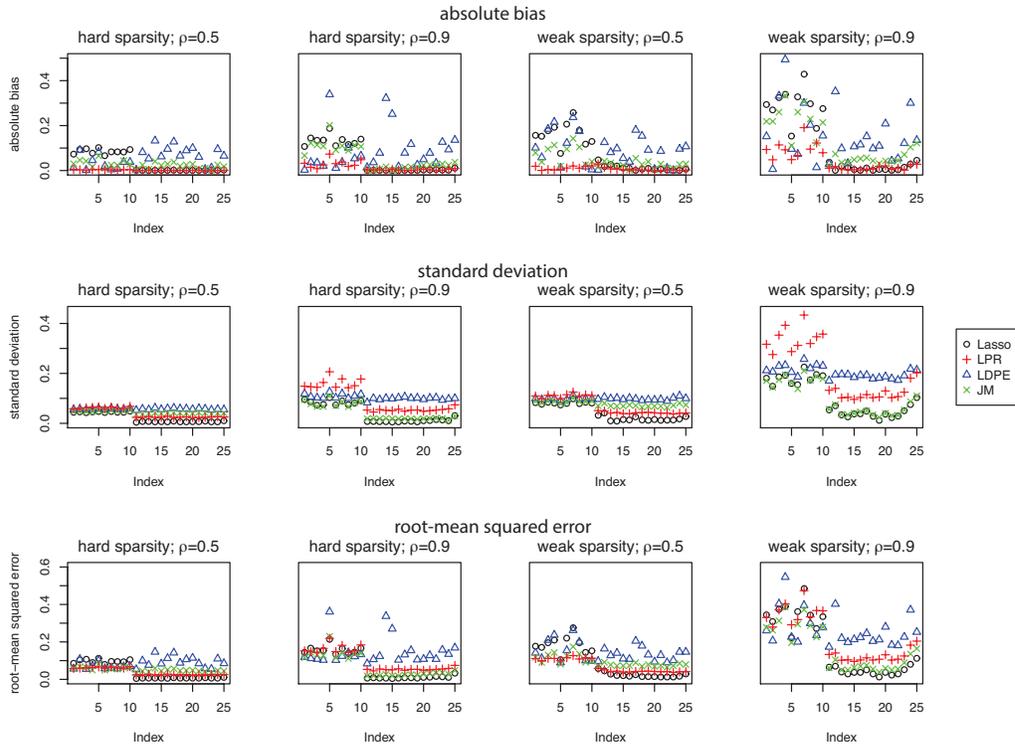
Figure S12: Comparison of bias, standard deviation and root-mean squared error. The design matrix is generated from the Normal distribution with a Toeplitz type covariance matrix.

Figure S13: Comparison of coverage probabilities (first row) and mean interval lengths (second row) produced by pBLPR, LDPE, JM, and BLDPE, when SNR changes. The third row shows the coverage probabilities v.s. mean interval lengths. The design matrix is generated from a Normal distribution with a Toeplitz type covariance matrix, and $\rho = 0.5$.
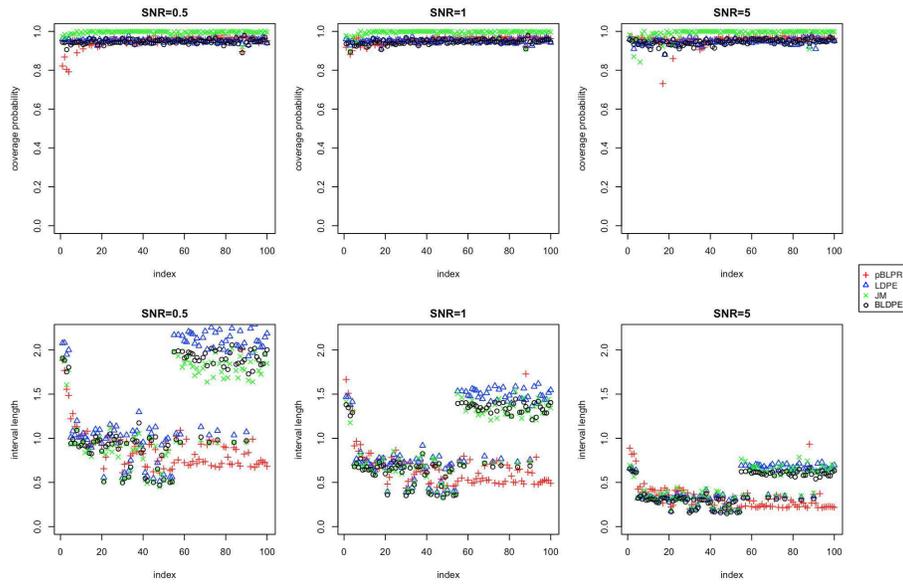
Figure S14: Comparison of coverage probabilities and mean interval lengths produced by pBLPR, LDPE, JM, and BLDPE. The results is based on data simulated from the misspecified linear model (S4.2).
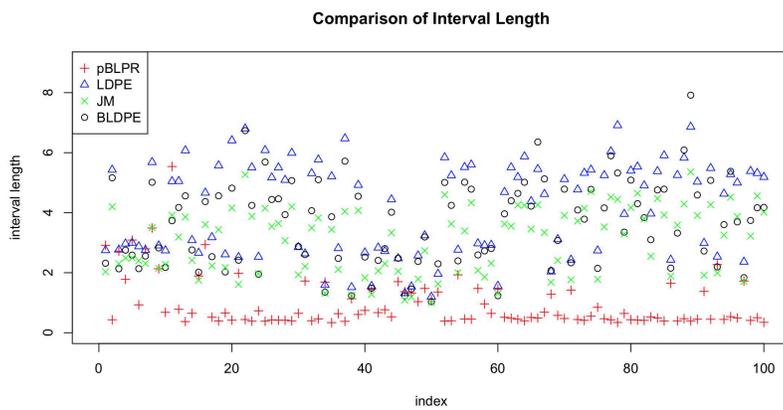
Figure S15: Comparison of interval lengths produced by pBLPR, LDPE, JM, and BLDPE. The plot is generated using the ninth voxel as the response in the fMRI data.
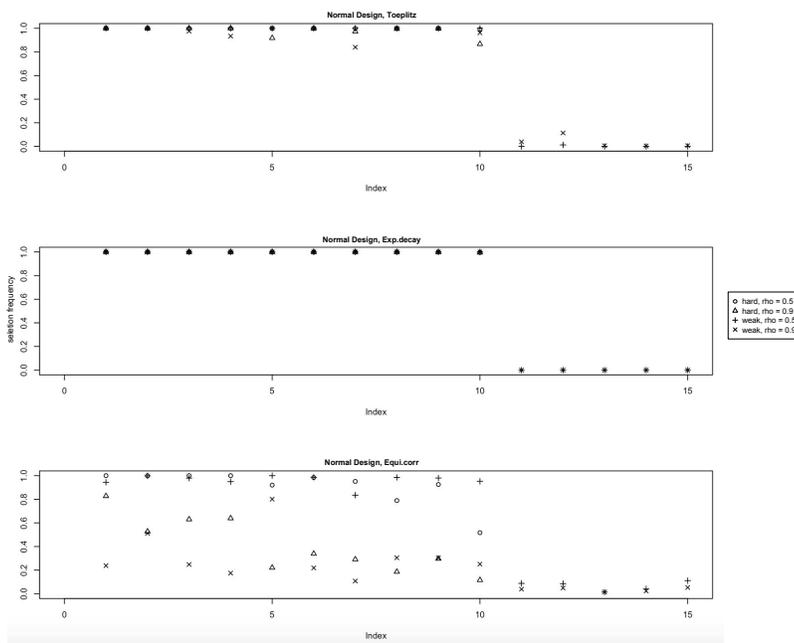
Figure S16: The selection frequency of each coefficient in 1000 simulation runs by the lasso (the 10 nonzero coefficients in hard sparsity case and the first 15 largest coefficients in absolute values in weak sparsity case).
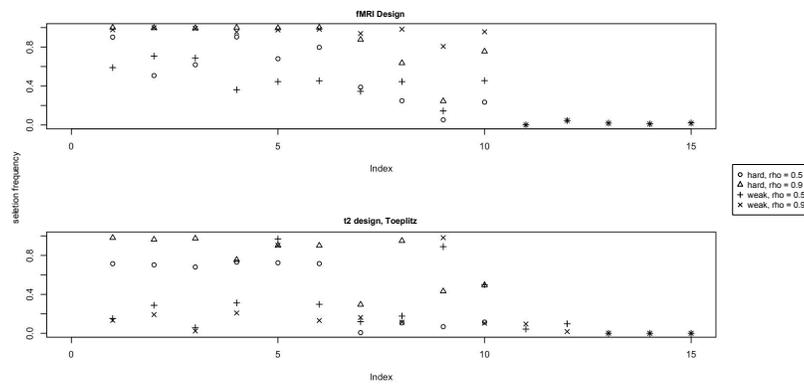
Figure S17: The selection frequency of each coefficient in 1000 simulation runs by the lasso (the 10 nonzero coefficients in hard sparsity case and the first 15 largest coefficients in absolute values in weak sparsity case).
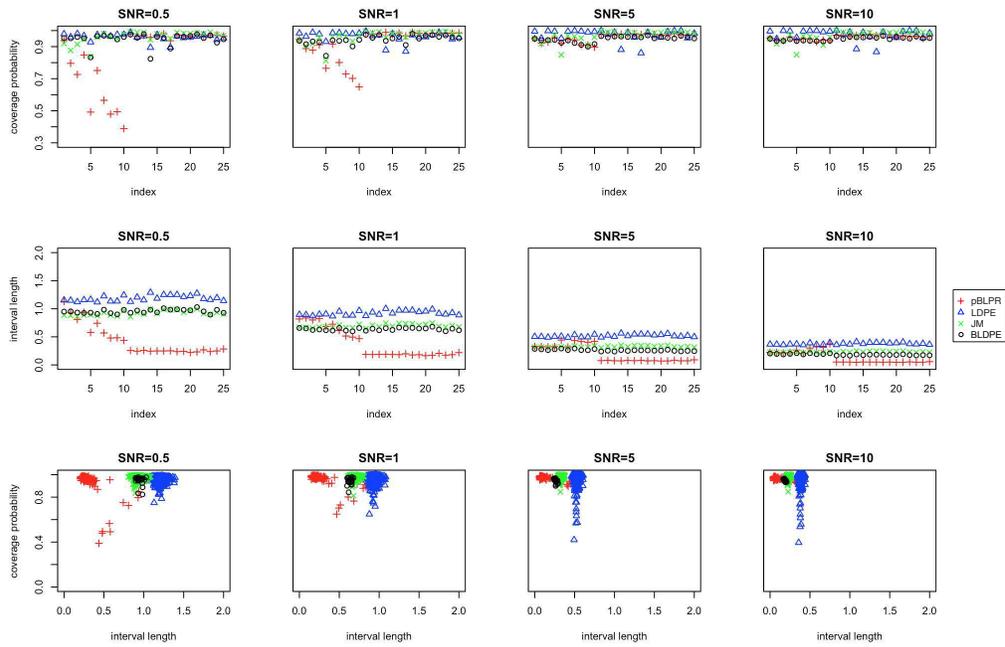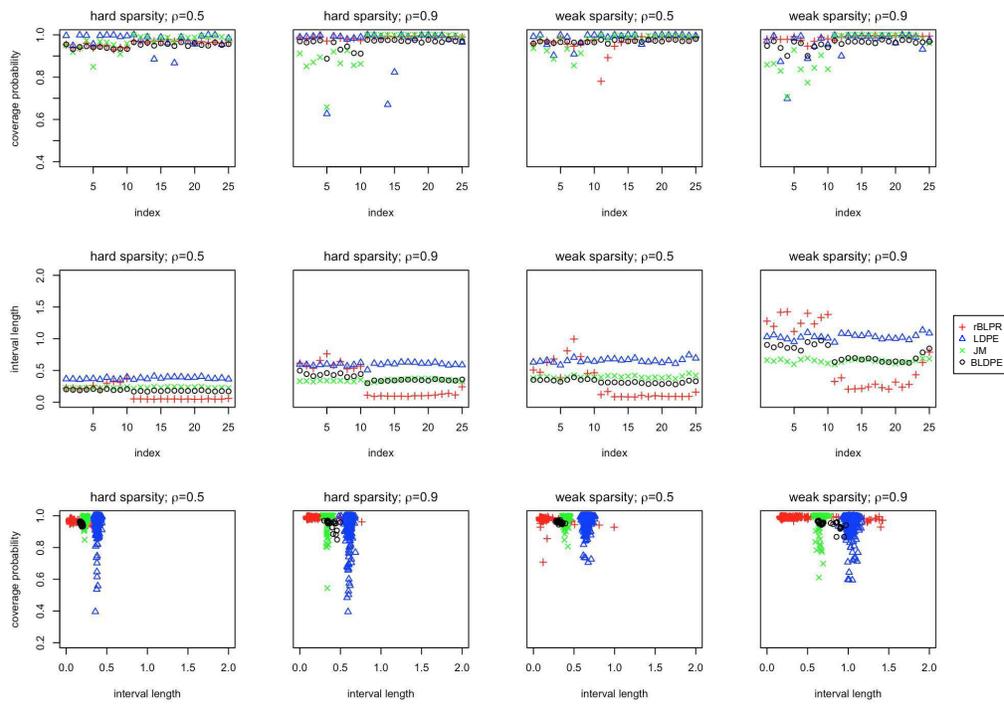
Figure S18: Comparison of coverage probabilities (the first row) and mean interval lengths (the second row) produced by rBLPR, LDPE, JM, and BLDPE. The third row shows the coverage probabilities v.s. mean interval lengths. The design matrix is generated from a Normal distribution with a Toeplitz type covariance matrix.
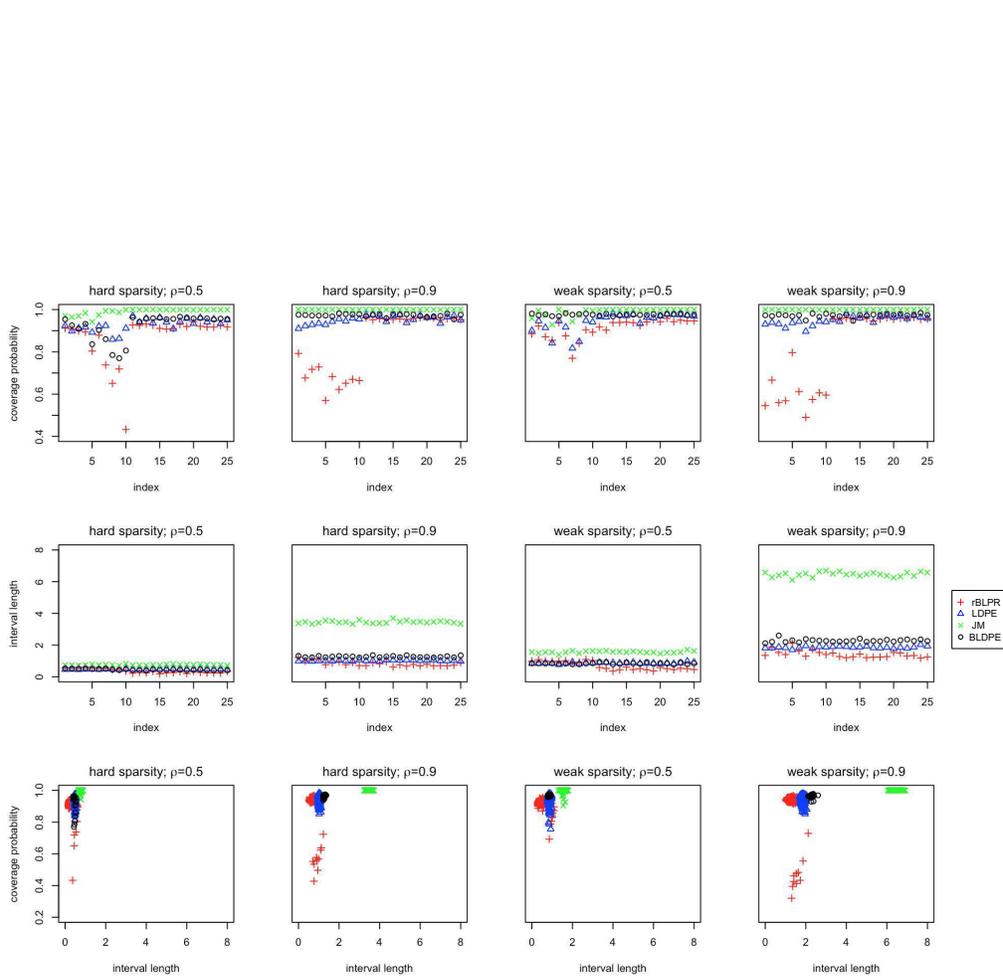
Figure S19: See caption of Figure S18 with the only difference being that the covariance matrix is an Equi.corr type.
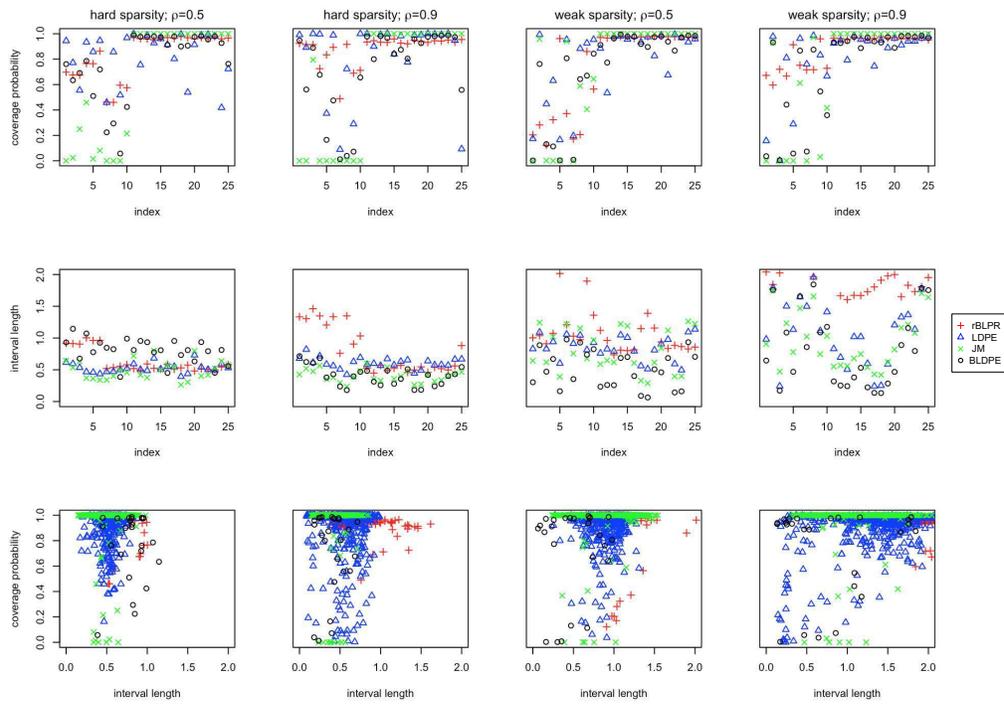
Figure S20: See caption of Figure S18 with the only difference being the type of design matrix. In this plot, the design matrix is generated from t_2 distribution with a Toeplitz type covariance matrix.
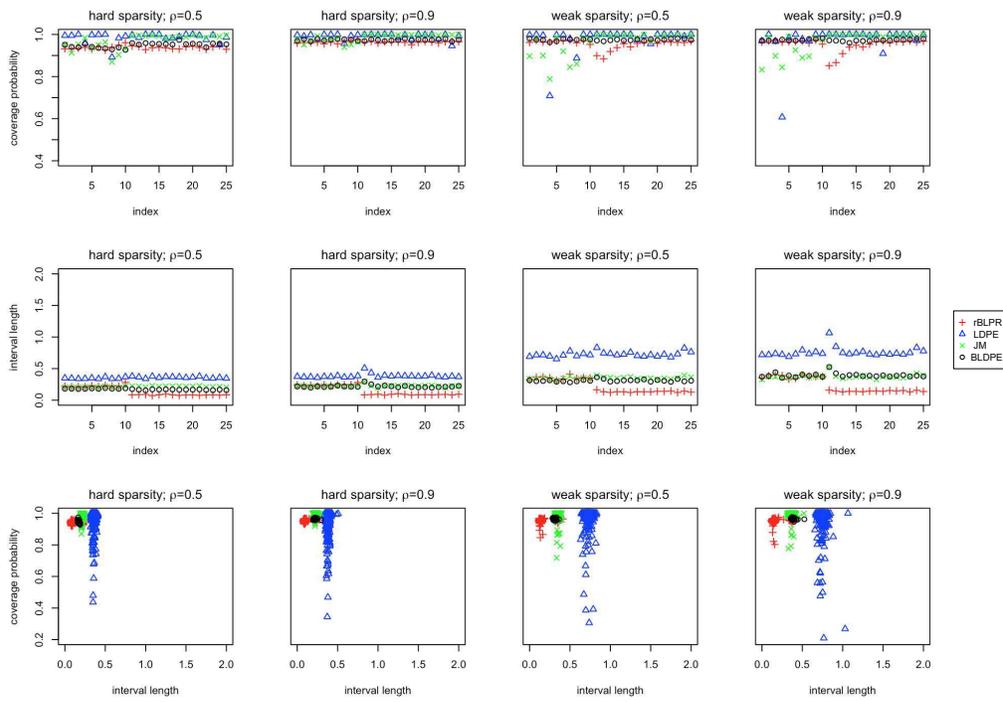
Figure S21: See caption of Figure S18 with the only difference being that the covariance matrix is Exp.decay type.
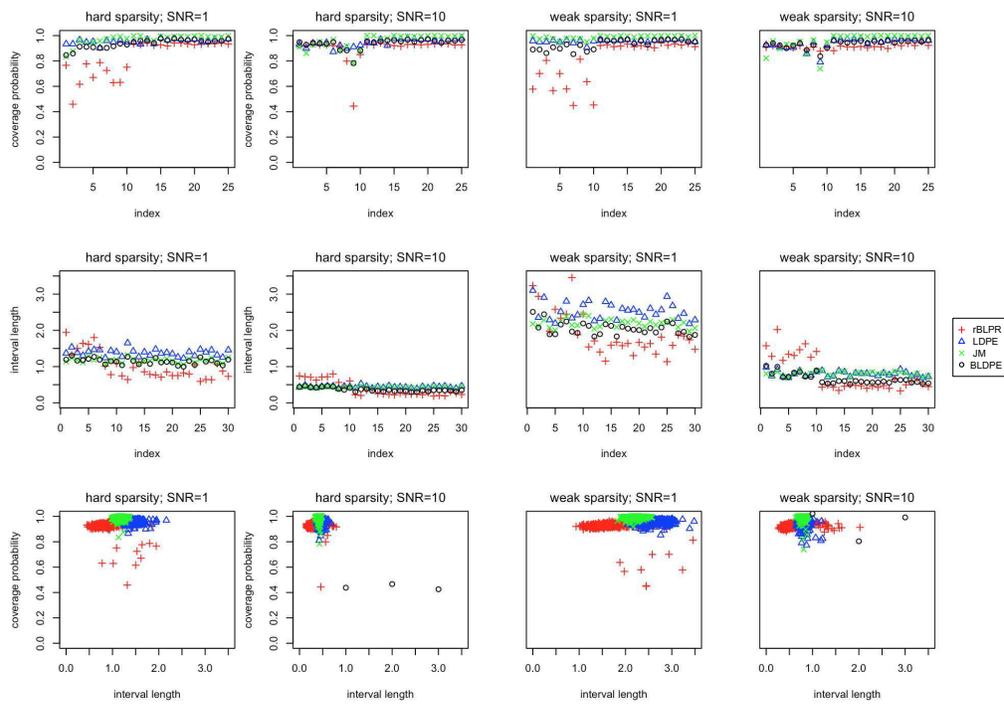
Figure S22: See caption of Figure S18 with the only difference being that the design matrix is generated from the fMRI data.

S6 Details of real-data case study 2: neuroblastoma gene expression data

S6.1 Overview

In this section, we apply our pBLPR and rBLPR methods, as well as three de-sparsified lasso methods, LDPE, JM, and BLDPE, to a data set containing $43,827$ gene expression measurements from the Illumina RNA sequencing of 498 neuroblastoma samples (Gene Expression Omnibus accession number GSE62564, with the file name `GSE62564_SEQC_NB_RNA-Seq_log2RPM.txt.gz`) generated by the Sequencing Quality Control (SEQC) consortium (Wang et al., 2014; Su et al., 2014; Munro et al., 2014; Su et al., 2014). Each neuroblastoma sample was labeled as high-risk (HR) or non-HR, indicating whether the sample belonged to a HR patient based on clinical evidence. There were 176 HR samples and 322 non-HR samples. We encode the sample labels as a binary vector $Z \in R^{498}$, with $Z_i = 1$ if the ith sample is HR, and $Z_i = 0$ otherwise. For the jth gene, we calculate the Pearson correlation between its gene expression vector $X_j \in R^{498}$ and Z, and we check the ten genes with the highest correlations. Among these ten genes, we find a gene *CAMTA1*, which has been reported as a gene related to medulloblastoma (Wu et al., 2012), a type of cancer closely related to neuroblastoma, and included in the Candidate Cancer Gene

Database (CCGD) (Abbott et al., 2015). We use the gene expression vector of *CAMTA1* as the response vector Y, and we consider the gene expression matrix of the other $43,826$ genes as the design matrix of dimensions $498 \times 43,826$. Our goal is to find genes that have significant effects on predicting the expression of *CAMTA1* in a multiple linear model. Given our lack of knowledge on the complex regulatory relationships between genes, the linear model is almost certainly a misspecified model. However, this case study would serve as a reasonable real-data example to demonstrate the ability of our pBLPR and rBLPR methods and three de-sparsified lasso methods (LDPE, JM and BLDPE) to identify significant predictors in a misspecified linear model.

S6.2   Results

We apply five methods (pBLPR, rBLPR, LDPE, JM, and BLDPE) to the linear model with the gene expression levels of *CAMTA1* as the response Y and the other 2000 genes which have the largest correlations with *CAMTA1* as features. We define the "significant genes" found by each method as those features whose 95% confidence intervals of their coefficients do not contain zero. Based on this definition, the numbers of significant genes found by the five methods are summarized in Table S5.

The results show that LDPE and its bootstrap version (BLDPE) find the

most significant genes; pBLPR and rBLPR find 91 and 26 significant genes, respectively; JM finds only one significant genes. We investigate the biological functions of those significant genes by performing a Gene Ontology (GO) analysis using a bioinformatics online tool GOrilla (Eden et al., 2009). Specifically, between one of our methods (pBLPR or rBLPR) and one of the de-sparsified lasso methods (LDPE, JM, or BLDPE), we check the significant genes found by one method but not the other, and we obtain the functions (i.e., Biological Process GO terms) enriched in those genes by GOrilla. An interesting observation is that the functions related to natural and regulated cell deaths (e.g., apoptosis and autophagy), which are key processes used to prevent cancer, are only enriched in the significant genes found by pBLPR or rBLPR, but not in those found by any of the de-sparisied lasso methods. On the other hand, only general functions, such as basic processes in cells, are enriched in the significant genes found by a de-sparsified lasso method, but not by our methods. Table S6 provides a summary of the numbers of the enriched GO terms and the specific terms related to apoptosis or autophagy. The detailed GO analysis results are provided in the Supplementary File.

The results of this case study suggest that pBLPR and rBLPR find significant features that are more reasonable and interpretable, based on domain knowledge, implying that pBLPR and rBLPR are robust to model misspecifi-

cation.

Table S1: Mean coverage probabilities over large β_j^0's (first 10 largest in absolute value).

				Normal design, Toeplitz					
β^0	ρ	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
hard	.5	.94	.94	.94	.91	.85	.30	.99	.94
hard	.9	.90	.97	.90	.87	.83	.31	.92	.77
weak	.5	.94	.94	.94	.90	.83	.27	.95	.90
weak	.9	.89	.96	.87	.84	.82	.32	.88	.73
				Normal design, Exponential decay					
hard	.5	.94	.93	.94	.89	.84	.26	.98	.94
hard	.9	.94	.93	.94	.88	.83	.25	.99	.95
weak	.5	.94	.93	.94	.88	.80	.20	.93	.86
weak	.9	.94	.94	.94	.88	.80	.19	.91	.87
				Normal design, Equal correlation					
hard	.5	.78	.87	.71	.60	.65	.40	.90	.98
hard	.9	.46	.66	.20	.40	.19	.33	.90	1.00
weak	.5	.79	.82	.67	.48	.59	.33	.84	.96
weak	.9	.34	.57	.15	.34	.15	.28	.88	1.00
				t_2 design, Toeplitz					
hard	.5	.65	.53	.39	.45	.23	.03	.78	.10
hard	.9	.80	.89	.64	.77	.46	.10	.73	.08
weak	.5	.41	.53	.33	.47	.16	.09	.63	.35
weak	.9	.74	.51	.20	.46	.13	.20	.64	.30
				fMRI design					
β^0	SNR	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
hard	1	.68	.76	.36	.61	.26	.37	.95	.94
hard	5	.78	.90	.73	.75	.63	.45	.92	.91
hard	10	.86	.93	.84	.79	.73	.46	.92	.90
weak	1	.63	.75	.30	.59	.22	.36	.95	.97
weak	5	.83	.93	.73	.69	.60	.43	.91	.90
weak	10	.91	.96	.88	.75	.79	.44	.90	.90

Table S2: Mean confidence interval lengths over large β_j^0's (first 10 largest in absolute value).

β^0	ρ	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
				Normal design, Toeplitz					
hard	.5	.24	.27	.18	.27	.15	.20	.37	.23
hard	.9	.57	.60	.36	.47	.30	.35	.59	.34
weak	.5	.44	.61	.33	.65	.28	.39	.64	.39
weak	.9	1.20	1.30	.81	1.04	.65	.76	1.02	.65
				Normal design, Exponential decay					
hard	.5	.23	.26	.17	.27	.15	.20	.35	.21
hard	.9	.24	.27	.18	.28	.15	.20	.37	.22
weak	.5	.35	.60	.26	.67	.24	.34	.71	.34
weak	.9	.37	.64	.26	.71	.24	.34	.73	.37
				Normal design, Equal correlation					
hard	.5	.53	.61	.38	.51	.34	.44	.48	.75
hard	.9	.94	.94	.33	.49	.32	.44	1.01	3.45
weak	.5	.99	1.18	.72	.94	.62	.81	.86	1.55
weak	.9	1.59	1.52	.49	.74	.5	.68	1.84	6.44
				t_2 design, Toeplitz					
hard	.5	.79	.61	.24	.42	.14	.27	.51	.46
hard	.9	1.20	1.11	.45	.64	.23	.39	.62	.41
weak	.5	1.25	1.03	.40	.76	.16	.47	.9	.87
weak	.9	2.63	1.89	.54	.96	.22	.51	1.33	1.11

β^0	SNR	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
				fMRI design					
hard	1	1.42	1.32	.57	.69	.38	.48	1.40	1.18
hard	5	.87	.89	.46	.63	.38	.48	.63	.60
hard	10	.66	.71	.37	.53	.32	.42	.44	.43
weak	1	2.79	2.50	.86	1.17	.61	.82	2.56	2.20
weak	5	1.89	1.89	.89	1.15	.72	.91	1.15	1.12
weak	10	1.45	1.53	.73	1.09	.63	.83	.81	.80

Table S3: Mean coverage probabilities over small β_j^0's (except for the first 10 largest in absolute value).

β^0	ρ	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
				Normal design, Toeplitz					
hard	.5	.94	.97	1.00	1.00	.96	1.00	.98	.99
hard	.9	.93	.99	1.00	1.00	.97	1.00	.96	1.00
weak	.5	.94	.98	.01	.06	.36	.33	.98	.99
weak	.9	.93	.99	.03	.15	.20	.35	.96	1.00
				Normal design, Exponential decay					
hard	.5	.94	.97	1.00	1.00	.96	1.00	.98	.99
hard	.9	.94	.97	1.00	1.00	.96	1.00	.96	1.00
weak	.5	.94	.98	.00	.06	.42	.32	.98	.99
weak	.9	.94	.98	.01	.05	.42	.31	.97	1.00
				Normal design, Equal correlation					
hard	.5	.92	.98	.98	1.00	.98	1.00	.95	1.00
hard	.9	.93	.98	.98	1.00	.98	1.00	.94	1.00
weak	.5	.91	.99	.16	.37	.07	.46	.95	1.00
weak	.9	.93	.97	.07	.25	.04	.35	.94	1.00
				t_2 design, Toeplitz					
hard	.5	.97	.95	.99	1.00	.99	1.00	.93	1.00
hard	.9	.94	.98	.99	1.00	.98	1.00	.91	1.00
weak	.5	.97	.95	.05	.06	.07	.10	.92	1.00
weak	.9	.96	.97	.04	.09	.05	.13	.9	1.00
				fMRI design					
β^0	SNR	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
hard	1	.93	.98	.99	1.00	.97	1.00	.96	.99
hard	5	.93	.98	.99	1.00	.97	1.00	.96	.99
hard	10	.93	.98	.99	1.00	.98	1.00	.96	.99
weak	1	.93	.98	.05	.22	.05	.35	.96	.99
weak	5	.92	.99	.08	.27	.07	.42	.96	1.00
weak	10	.92	.99	.07	.26	.07	.44	.96	1.00

Table S4:   Mean confidence interval lengths over small β_j^0's (except for the first 10 largest in absolute value).

β^0	ρ	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
				Normal design, Toeplitz					
hard	.5	.09	.05	.00	.01	.01	.02	.38	.23
hard	.9	.20	.12	.02	.03	.03	.03	.61	.35
weak	.5	.16	.10	.00	.02	.03	.04	.66	.40
weak	.9	.39	.24	.04	.06	.05	.06	1.04	.66
				Normal design, Exponential decay					
hard	.5	.09	.05	.00	.01	.01	.02	.36	.22
hard	.9	.09	.05	.00	.01	.01	.02	.38	.23
weak	.5	.14	.09	.00	.02	.02	.04	.72	.35
weak	.9	.14	.09	.00	.02	.02	.04	.74	.37
				Normal design, Equal correlation					
hard	.5	.27	.17	.07	.05	.06	.05	.49	.76
hard	.9	.75	.46	.13	.10	.12	.09	1.03	3.45
weak	.5	.49	.34	.15	.12	.12	.10	.87	1.56
weak	.9	1.36	.83	.22	.18	.23	.17	1.86	6.44
				t_2 design, Toeplitz					
hard	.5	.53	.24	.02	.03	.01	.02	.54	.51
hard	.9	.56	.31	.04	.05	.02	.03	.60	.43
weak	.5	.88	.39	.03	.05	.02	.03	.90	.90
weak	.9	1.86	.83	.06	.08	.02	.04	1.33	1.12

β^0	SNR	rBLPR	pBLPR	rBlassoOLS	pBlassoOLS	rBlasso	pBlasso	LDPE	JM
				fMRI design					
hard	1	.83	.5	.08	.09	.08	.08	1.40	1.17
hard	5	.37	.23	.04	.05	.05	.05	.63	.58
hard	10	.26	.16	.03	.03	.04	.04	.44	.42
weak	1	1.63	1.01	.19	.18	.15	.15	2.54	2.21
weak	5	.75	.50	.12	.12	.10	.10	1.13	1.13
weak	10	.52	.36	.07	.09	.08	.08	.80	.81

Method	pBLPR	rBLPR	LDPE	JM	BLDPE
# significant genes	91	26	501	1	135

Table S5: Numbers of significant genes found by the 95% confidence intervals constructed by five methods.

$A\backslash B$	LDPE	JM	BLDPE	$B\backslash A$	LDPE	JM	BLDPE
pBLPR	4/19	6/15	4/11	pBLPR	0/67	0/0	0/18
rBLPR	2/14	5/18	2/11	rBLPR	0/78	0/0	0/18

Table S6: The numbers of Biological Process GO terms enriched in the significant genes found by method A, but not by method B. The numerators are the numbers of GO terms related to apoptosis or autophagy, and the denominators are the total numbers of GO terms enriched in the significant genes. For example, $4/19$ in the left table indicates that there are 19 GO terms enriched in the significant genes found by pBLPR, but not by LDPE, and among these 19 terms, 4 terms are related to apoptosis or autophagy.

S7   Algorithms

Algorithm S1   Residual Bootstrap LPR (rBLPR) procedure for confidence interval construction

---

Input: Data (X, Y); Confidence level $1 - \alpha$; Number of replications B.

Output: Confidence interval $[l_j, u_j]$ of β_j^0, for $j = 1, \ldots, p$.

1: Compute the Lasso+OLS estimator $\hat{\beta}_{\text{Lasso+OLS}}$, given data (X, Y);

2: Compute residual vector $\hat{\epsilon} = (\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n)^T = Y - X\hat{\beta}_{\text{Lasso+OLS}}$;

3: Re-sample from the empirical distribution of the centered residual $\{\hat{\epsilon}_i - \bar{\hat{\epsilon}}, i = 1, \ldots, n\}$, where $\bar{\hat{\epsilon}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_i$, to form $\epsilon^* = (\epsilon_1^*, \ldots, \epsilon_n^*)^T$;

4: Generate residual Bootstrap response $Y_{\text{rboot}}^* = X\hat{\beta}_{\text{Lasso+OLS}} + \epsilon^*$;

5: Compute the residual Bootstrap LPR, $\hat{\beta}_{\text{rBLPR}}^*$, based on (X, Y_{rboot}^*) as in equations (2.6) and (2.7);

6: Repeat steps 3-5 for B times, and obtain $\hat{\beta}_{\text{rBLPR}}^{*(1)}, \ldots, \hat{\beta}_{\text{rBLPR}}^{*(B)}$;

7: For each $j = 1, \ldots, p$, compute the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\left\{ (\hat{\beta}_{\text{rBLPR}}^{*(b)})_j \right\}_{b=1}^{B}$, and denote them as a_j and b_j, respectively; let $l_j = (\hat{\beta}_{\text{LPR}})_j + (\hat{\beta}_{\text{Lasso+OLS}})_j - b_j$ and $u_j = (\hat{\beta}_{\text{LPR}})_j + (\hat{\beta}_{\text{Lasso+OLS}})_j - a_j$;

8: **return** $1 - \alpha$ confidence interval $[l_j, u_j]$, for $j = 1, \ldots, p$.

Algorithm S2   Paired Bootstrap LPR (pBLPR) procedure for confidence

interval construction

Input: Data (X, Y); Confidence level $1 - \alpha$; Number of replications B.

Output: Confidence interval $[l_j, u_j]$ of β_j^0, for $j = 1, \ldots, p$.

1: Generate a Bootstrap sample $(X^*_{\text{pboot}}, Y^*_{\text{pboot}}) = \{(x_i^*, y_i^*), i = 1, \ldots, n\}$ from

    the empirical distribution of $\{(x_i, y_i), i = 1, \ldots, n\}$;

2: Based on $(X^*_{\text{pboot}}, Y^*_{\text{pboot}})$, compute the paired Bootstrap Lasso estimator,

    $\hat{\beta}^*_{\text{pBLasso}}$, as in equation (2.8) and its selected predictor set, $\hat{S}^*_{\text{pBLasso}}$; and then

    compute the paired Bootstrap LPR estimator, $\hat{\beta}^*_{\text{pBLPR}}$, as in equation (2.9);

3: Repeat steps 1-2 for B times and obtain $\hat{\beta}^{*(1)}_{\text{pBLPR}}, \ldots, \hat{\beta}^{*(B)}_{\text{pBLPR}}$;

4: For each $j = 1, \ldots, p$, compute the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\left\{ (\hat{\beta}^{*(b)}_{\text{pBLPR}})_j \right\}_{b=1}^{B}$,

    and denote them as l_j and u_j, respectively;

5: **return** $1 - \alpha$ confidence interval $[l_j, u_j]$, for $j = 1, \ldots, p$.

Algorithm S3 K-fold cross validation based on lasso+ols: cv(lasso+ols)

Input: Design matrix X, response Y, a sequence of tuning parameter values

$\lambda_1, \ldots, \lambda_J$, and number of folds K.

Output: The optimal tuning parameter selected by cv(lasso+ols): $\lambda_{optimal}$.

1: Randomly divide the data $z = (X, Y)$ into K roughly equal folds $\{z_k, k = 1, \ldots, K\}$;

2: For each $k = 1, \ldots, K$, denote $\hat{S}^{(k)}(\lambda_0) = \emptyset$ and $\hat{\beta}^{(k)}_{\text{lasso+ols}}(\lambda_0) = 0$.

- Fit the model with parameters λ_j, $j = 1, \ldots, J$ to the other $K-1$ folds,

  $z_{-k} = z \setminus z_k$, of the data, giving the lasso solution path $\hat{\beta}^{(k)}(\lambda_j)$, $j = 1, \ldots, J$, and compute the sets of selected covariates on the path

  $$\hat{S}^{(k)}(\lambda_j) = \left\{ l : \hat{\beta}^{(k)}_l(\lambda_j) \neq 0 \right\}, \text{ for } j = 1, \ldots, J;$$

- For each $j = 1, \ldots, J$, compute the lasso+ols estimator:

  $$\hat{\beta}^{(k)}_{\text{lasso+ols}}(\lambda_j) = \begin{cases} \arg\min_{\beta: \, \beta_j = 0, \, j \notin \hat{S}^{(k)}(\lambda_j)} \left\{ \frac{1}{2|z_{-k}|} \sum_{i \in z_{-k}} (y_i - x_i^T \beta)^2 \right\}, & \text{if } \hat{S}^{(k)}(\lambda_j) \neq \hat{S}^{(k)}(\lambda_{j-1}), \\ \hat{\beta}^{(k)}_{\text{lasso+ols}}(\lambda_{j-1}), & \text{otherwise}; \end{cases}$$

- Compute the prediction error $PE^{(k)}$ on the kth fold of the data:

  $$PE^{(k)}(\lambda_j) = \frac{1}{|z_k|} \sum_{i \in z_k} \left(y_i - x_i^T \hat{\beta}^{(k)}_{\text{lasso+ols}}(\lambda_j) \right)^2;$$

3: Compute cross validated error $CVE(\lambda_j)$, $j = 1, \ldots, J$:

$$CVE(\lambda_j) = \frac{1}{K} \sum_{k=1}^{K} PE^{(k)}(\lambda_j);$$

4: Compute the optimal λ selected by cv(lasso+ols):

$$\lambda_{optimal} = \underset{\lambda_j:\ j=1,\dots,J}{\arg\min}\ CVE(\lambda_j);$$

5: **return** $\lambda_{optimal}$.

Bibliography

Abbott, K., Nyre, E., Abrahante, J. *et al.* (2015). The Candidate Cancer Gene Database: a database of cancer driver genes from forward genetic screens in mice. *Nucleic Acids Res.* **D1**, D844–D848.

Dezeure, R., Bühlmann, P., Meier, L. and Meinshausen, N. (2014). High-dimensional Inference: Confidence intervals, p-values and R-Software hdi. *Stat. Sci.* **30**, 533–558.

Eden, E., Navon, R., Steinfeld, I. *et al.* (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* **10**, 48.

Liu, H. and Yu, B. (2013). Asymptotic properties of lasso+mLS and lasso+Ridge in sparse high-dimensional linear regression. *Electron. J. Stat.* **7**, 3124–3169.

Munro, S. A., Lund, S. P., Pine, P. S. *et al.* (2014). Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. *Nat. Commun.* **5**, 5125.

Su, Z., Fang, H., Hong, H. *et al.* (2014). An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. *Genome Biol.* **15**, 523.

Su, Z., Łabaj, P. P., Li, S. *et al.* (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914.

Wang, C., Gong, B., Bushel, P. R. *et al.* (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* **32**, 926–932.

Wu, X., Northcott, P. A., Dubuc, A. *et al.* (2012). Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature.* **482**, 529–533.

Zhang C-H., and Zhang S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Statist. Soc.* B **76**, 217–242.

Zhao, P. and Yu, B. (2006). On Model Selection Consistency of lasso. *J. Mach. Learn. Res.* **7**, 2541–2563.