VARIABLE SELECTION IN PARTLY LINEAR REGRESSION MODEL WITH DIVERGING DIMENSIONS FOR RIGHT CENSORED DATA

Shuangge Ma and Pang Du

Yale University and Virginia Tech

Supplementary Material

We first describe the following results, which is Lemma 1 of Huang and Ma (2010).

Let Let $\tau = (\tau_1, \ldots, \tau_n)^T$ and $\xi_n = \max_{1 \le j \le p} |\xi_j|$. Suppose that conditions (A2) and (A3) hold. Then

$$\mathrm{E}(\xi_n) \le C_1 \sqrt{\log(p)} \big(\sqrt{2C_2 n \log(p)} + 4\log(2p) + C_2 n \big)^{1/2},$$

where $C_1, C_2 > 0$ are constants. In particular, when $\log(p)/n \to 0$,

$$\mathrm{E}(\xi_n) = O(1)\sqrt{n \log p}.$$

S1 Proof of Theorem 1

Examination of Theorem 1 of Zhang and Huang (2008) suggests that the normality assumption is not necessary. As a matter of fact, as long as the tail probability $\sim \exp(-x^2)$, Theorem 1 and its proof in Zhang and Huang (2008) holds. Part (a) of our Theorem 1 thus follows.

Under assumption (A1), $\min_{j \in A_1} |\beta_{0j}| > b_1 > 0$ for a constant b_1. Thus, if part (c) of Theorem 1 holds, then part (b) follows. Proof of part (c) proceeds as follows. The Lasso estimate satisfies

$$||\tilde{Y} - \tilde{X}\tilde{\boldsymbol{\beta}}||^2 + 2\lambda_n \sum_j |\tilde{\beta}_j| \le ||\tilde{Y} - \tilde{X}\boldsymbol{\beta}_0||^2 + 2\lambda_n \sum_j |\beta_{0j}|,$$

which leads to

$$||\tilde{Y} - \tilde{X}\tilde{\boldsymbol{\beta}}||^2 + 2\lambda_n \sum_{j \in A_1} |\tilde{\beta}_j| \le ||\tilde{Y} - \tilde{X}\boldsymbol{\beta}_0||^2 + 2\lambda_n \sum_{j \in A_1} |\beta_{0j}|.$$

Thus, we have

$$||\tilde{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||^2 - 2\tau^T \tilde{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \le 2\lambda_n \sum_{j \in A_1} |\tilde{\beta}_j - \beta_{0j}|.$$

We note that
$$\sum_{j\in A_1}|\tilde{\beta}_j-\beta_{0j}|\le\sqrt{|A_1|}||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||,$$
where $\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}=\{\tilde{\beta}_j : j\in A_1\cup\tilde{A}_1\}$ and $\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}=\{\beta_{0j} : j\in A_1\cup\tilde{A}_1\}$. Combining the above equations, we have

$$\begin{aligned}&||\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1})||^2-2\tau^T(\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}))\\&\le 2\lambda_n\sqrt{|A_1|}||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||.\end{aligned}$$

Define $\tau^*=\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{X}^T_{A_1\cup\tilde{A}_1}\tilde{X}_{A_1\cup\tilde{A}_1})^{-1}\tilde{X}^T_{A_1\cup\tilde{A}_1}\tau$. From the Cauchy-Schwarz inequality, we have

$$|2\tau^T(\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}))|\le 2||\tau^*||^2+\frac{1}{2}||\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1})||^2.$$

Combining the above equations,

$$||\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1})||^2\le 4||\tau^*||^2+4\lambda_n\sqrt{|A_1|}\times||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||.$$

Under assumption (A4),

$$||\tilde{X}_{A_1\cup\tilde{A}_1}(\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1})||^2\ge nc_*||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||^2.$$

Combining the above two equations, we have

$$nc_*||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||^2\le 4||\tau^*||^2+\frac{16\lambda_n^2|A_1|}{2nc_*}+\frac{1}{2}nc_*||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||^2.$$

It follows that

$$||\tilde{\boldsymbol{\beta}}_{A_1\cup\tilde{A}_1}-\boldsymbol{\beta}_{0A_1\cup\tilde{A}_1}||^2\le\frac{8||\tau^*||^2}{nc_*}+\frac{16\lambda_n^2|A_1|}{n^2c_*^2}. \tag{S1.1}$$

Under the SRC, we also have

$$||\tau^*||^2\le\frac{||\tilde{X}_{A_1\cup\tilde{A}_1}\tau||^2}{nc_*}\le\frac{max_{B:|B|\le p_1^*}||\tilde{X}_B\tau||^2}{nc_*}.$$

We also have
$$max_{B:|B|\le p_1^*}||\tilde{X}_B\tau||^2\le p_1^* max_j|\tilde{X}_j^T\tau|.$$
Applying the result described in the beginning of this section,

$$max_j|\tilde{X}_j^T\tau|=O(n\log(p)).$$

Thus,

$$||\tau^*||^2=O(\frac{p_1^*\log(p)}{c_*}). \tag{S1.2}$$

Part (c) follows from equations (S1.1) and (S1.2).

S2 Proof of Theorem 2

By the Karush-Kunh-Tucker condition, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$ is the adaptive Lasso estimate if

$$\begin{cases} \tilde{X}_j^T(\tilde{Y} - \tilde{X}\hat{\boldsymbol{\beta}}) = \lambda_n v_j sign(\hat{\beta}_j), & \hat{\beta}_j \neq 0 \\ |\tilde{X}_j^T(\tilde{Y} - \tilde{X}\hat{\boldsymbol{\beta}})| \leq \lambda_n v_j & \hat{\beta}_j = 0 \end{cases} \tag{S2.1}$$

and the vectors $\{\tilde{X}_j : j \in \hat{A}_1\}$ are linearly independent. Define $\tilde{s}_1 = (v_j sign(\beta_{0j}), j \in A_1)^T$, $\tilde{X}_{A_1} = (\tilde{X}_j, j \in A_1)$, and $\boldsymbol{\beta}_{0A_1} = (\beta_{0j}, j \in A_1)^T$. Define

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{A_1} &= (\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}(\tilde{X}_{A_1}^T \tilde{Y} - \lambda_n \tilde{s}_1) \\ &= \boldsymbol{\beta}_{0A_1} + \left(\tilde{X}_{A_1}^T \tilde{X}_{A_1}/n\right)^{-1} (\tilde{X}_{A_1}^T \tau - \lambda_n \tilde{s}_1)/n. \end{aligned} \tag{S2.2}$$

If $sign(\hat{\boldsymbol{\beta}}_{A_1}) = sign(\boldsymbol{\beta}_{0A_1})$, then (S2.1) holds for $\tilde{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{A_1}^T, 0^T)^T$. Since $\tilde{X}\tilde{\boldsymbol{\beta}} = \tilde{X}_{A_1}\hat{\boldsymbol{\beta}}_{A_1}^T$, we have

$$sign(\hat{\boldsymbol{\beta}}) = sign(\boldsymbol{\beta}_0) \quad \text{if} \quad \begin{cases} sign(\hat{\boldsymbol{\beta}}_{A_1}) = sign(\boldsymbol{\beta}_{0A_1}) \\ |\tilde{X}_j^T(\tilde{Y} - \tilde{X}_{A_1}\hat{\boldsymbol{\beta}}_{A_1})| \leq \lambda_n v_j, \forall j \notin A_1. \end{cases} \tag{S2.3}$$

Define $H_n = I - \tilde{X}_{A_1}(\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}\tilde{X}_{A_1}^T$. From the definition of $\hat{\boldsymbol{\beta}}_{A_1}$,

$$\tilde{Y} - \tilde{X}_{A_1}\hat{\boldsymbol{\beta}}_{A_1} = \tau - \tilde{X}_{A_1}(\hat{\boldsymbol{\beta}}_{A_1} - \boldsymbol{\beta}_{0A_1}) = H_n\tau + \tilde{X}_{A_1}(\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}\tilde{s}_1\lambda_n.$$

Thus, following (S2.3),

$$sign(\hat{\boldsymbol{\beta}}) = sign(\hat{\boldsymbol{\beta}}_0) \quad \text{if} \quad \begin{cases} sign(\beta_{0j})(\beta_{0j} - \hat{\beta}_j) \leq |\beta_{0j}|, & \forall j \in A_1 \\ |\tilde{X}_j^T\left(H_n\tau + \tilde{X}_{A_1}(\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}\tilde{s}_1\lambda_n\right| < \lambda_n v_j, & \forall j \notin A_1. \end{cases} \tag{S2.4}$$

Combining equations (S2.2) and (S2.4),

$$\begin{aligned} P\left\{sign(\hat{\boldsymbol{\beta}}) \neq sign(\boldsymbol{\beta}_0)\right\} \quad \leq \quad & P\left\{|e_j^T(\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}\tilde{X}_{A_1}^T\tau| \geq |\beta_{0j}|/2 \text{ for some } j \in A_1\right\} \\ & +P\left\{|e_j^T(\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}\tilde{s}_1|\lambda_n/n \geq |\beta_{0j}|/2 \text{ for some } j \in A_1\right\} \\ & +P\left\{|\tilde{X}_j^T H_n\tau| \geq \lambda_n v_j/2 \text{ for some } j \notin A_1\right\} \\ & +P\left\{|\tilde{X}_j^T \tilde{X}_{A_1}(\tilde{X}_{A_1}^T \tilde{X}_{A_1})^{-1}\tilde{s}_1| \geq v_j/2 \text{ for some } j \notin A_1\right\}, \end{aligned}$$

where e_j is the unit vector in the direction of the j-th coordinate. Following Huang et al. (2008), it can be proved that each of the above four probabilities converges to zero.