The Dantzig Selector for Censored Linear Regression Models

Yi Li[1], Lee Dicker[2], and Sihai Dave Zhao[3]

[1] *University of Michigan,* [2] *Rutgers University,* [3] *University of Pennsylvania*

Supplementary Material

S1 Proof of Lemma 1

Decompose $\hat{S}(s_1, \boldsymbol{\beta}_1) - S(s_0) = \hat{S}(s_1, \boldsymbol{\beta}_1) - \hat{S}(s_0, \boldsymbol{\beta}_0) + \hat{S}(s_0, \boldsymbol{\beta}_0) - S(s_0)$. First study $\hat{S}(s_1, \boldsymbol{\beta}_1) - \hat{S}(s_0, \boldsymbol{\beta}_0)$. Using the arguments of Lai and Ying (1988), it follows that with probability 1,

$$\sup_{(s_1, \boldsymbol{\beta}_1) \in \mathcal{B}} |\hat{S}(s_1, \boldsymbol{\beta}_1) - \hat{S}(s_0, \boldsymbol{\beta}_0) - \xi(s_1, \boldsymbol{\beta}_1) + \xi(s_0, \boldsymbol{\beta}_0)| = o(n^{-1/2}),$$

where

$$\xi(s, \boldsymbol{\beta}) = \exp\left\{ - \sum_{i=1}^{n} \int_{-\infty}^{s_1} \frac{dE_x N_i(s, \boldsymbol{\beta})}{E_x \bar{Y}(s, \boldsymbol{\beta})} \right\} = \exp\left\{ - \frac{\sum_i G_i(s + \boldsymbol{\beta}'\mathbf{X}_i) f(s + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X}_i)}{\sum_i G_i(s + \boldsymbol{\beta}'\mathbf{X}_i) S(s + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X}_i)} ds \right\}$$

where E_x denote the expectation conditional on X, G_i is the survival function of C_i conditional on $\mathbf{X}_i$ and f is the density function of S. Note that $\xi(s, \boldsymbol{\beta}_0) = S(s)$.

Now denote by $\mathbf{d} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$ and by $\lambda(\cdot)$ the hazard function for S. Note that

$$\begin{aligned}
\xi(s, \boldsymbol{\beta}) &= \exp\left\{ \int_{-\infty}^{s} - \frac{\sum_i G_i(s + \boldsymbol{\beta}_0'\mathbf{X}_i + \mathbf{d}'\mathbf{X}_i) S(s + \mathbf{d}'\mathbf{X}_i)(\lambda(s + \mathbf{d}'\mathbf{X}_i) - \lambda(s))}{\sum_i G_i(s + \boldsymbol{\beta}_0'\mathbf{X}_i + \mathbf{d}'\mathbf{X}_i) S(s + \mathbf{d}'\mathbf{X}_i)} ds - \int_{-\infty}^{s} \lambda(s) ds \right\} \\
&= S(s) \exp\left\{ \int_{-\infty}^{s} - \frac{\sum_i G_i(s + \boldsymbol{\beta}_0'\mathbf{X}_i + \mathbf{d}'\mathbf{X}_i) S(s + \mathbf{d}'\mathbf{X}_i) \lambda^{(1)}(s + d_*'\mathbf{X}_i) \mathbf{d}'\mathbf{X}_i}{\sum_i G_i(s + \boldsymbol{\beta}_0'\mathbf{X}_i + \mathbf{d}'\mathbf{X}_i) S(s + \mathbf{d}'\mathbf{X}_i)} ds \right\} \\
&= S(s) \left\{ 1 - \mathbf{d}' \int_{-\infty}^{s} \frac{\sum_i G_i(s + \boldsymbol{\beta}_0'\mathbf{X}_i + \mathbf{d}'\mathbf{X}_i) S(s + \mathbf{d}'\mathbf{X}_i) \lambda^{(1)}(s + d_*'\mathbf{X}_i) \mathbf{X}_i}{\sum_i G_i(s + \boldsymbol{\beta}_0' X_i + \mathbf{d}'\mathbf{X}_i) S(s + \mathbf{d}'\mathbf{X}_i)} ds \right\} + o(||\mathbf{d}||),
\end{aligned}$$

where $\lambda^{(1)}(\cdot)$ denotes the first derivative. Hence,

$$\begin{aligned}
& \xi(s_1,\boldsymbol{\beta}_1)-\xi(s_0,\boldsymbol{\beta}_0)\\
= & \xi(s_1,\boldsymbol{\beta}_1)-\xi(s_1,\boldsymbol{\beta}_0)+\xi(s_1,\boldsymbol{\beta}_0)-\xi(s_0,\boldsymbol{\beta}_0)\\
= & -S(s_1)\mathbf{d}'\int_{-\infty}^{s}\frac{\sum_i G_i(s+\boldsymbol{\beta}_0'X_i+\mathbf{d}'\mathbf{X}_i)S(s+\mathbf{d}'\mathbf{X}_i)\lambda^{(1)}(s+\mathbf{d}_*'\mathbf{X}_i)\mathbf{X}_i}{\sum_i G_i(s+\boldsymbol{\beta}_0'\mathbf{X}_i+\mathbf{d}'\mathbf{X}_i)S(s+\mathbf{d}'\mathbf{X}_i)}ds\\
& +S(s_1)-S(s_0,\boldsymbol{\beta}_0)+o(||\mathbf{d}||)\\
= & -S(s_0)\mathbf{d}'\int_{-\infty}^{s}\frac{\sum_i G_i(s+\boldsymbol{\beta}_0'\mathbf{X}_i+\mathbf{d}'\mathbf{X}_i)S(s+\mathbf{d}'\mathbf{X}_i)\lambda^{(1)}(s+\mathbf{d}_*'\mathbf{X}_i)\mathbf{X}_i}{\sum_i G_i(s+\boldsymbol{\beta}_0'\mathbf{X}_i+\mathbf{d}'\mathbf{X}_i)S(s+\mathbf{d}'\mathbf{X}_i)}ds\\
& -f(s_0,\boldsymbol{\beta}_0)(s_1-s_0)+o(||\mathbf{d}||+|s_1-s_0|),
\end{aligned}$$

where $||\mathbf{d}_*||\leq||\mathbf{d}||$. Denote by

$$\Gamma^{(r)}(s,\boldsymbol{\beta}_0)=\text{plim}\frac{1}{n}\sum_i G_i(s+\boldsymbol{\beta}_0'\mathbf{X}_i)\mathbf{X}_i^{\otimes r} \tag{S1.1}$$

for $r=0,1,2$, where for a vector $\mathbf{a}$ $\mathbf{a}^{\otimes 0}=1,\mathbf{a}^{\otimes 1}=\mathbf{a}$ and $\mathbf{a}^{\otimes 2}=\mathbf{a}\mathbf{a}'$ and plim denote the probabilistic limit. The argument of Ying (1993, p.87) leads to

$$\int_{-\infty}^{s}\frac{\sum_i G_i(s+\boldsymbol{\beta}_0'\mathbf{X}_i+\mathbf{d}'\mathbf{X}_i)S(s+\mathbf{d}'\mathbf{X}_i)\lambda^{(1)}(s+\mathbf{d}_*'\mathbf{X}_i)\mathbf{X}_i}{\sum_i G_i(s+\boldsymbol{\beta}_0'\mathbf{X}_i+\mathbf{d}'\mathbf{X}_i)S(s+\mathbf{d}'\mathbf{X}_i)}ds=\int_{-\infty}^{s}\frac{\Gamma^{(1)}(s,\boldsymbol{\beta}_0)}{\Gamma^{(0)}(s,\boldsymbol{\beta}_0)}d\lambda(s)+o(||\mathbf{d}||)$$

Hence, $\hat{S}(s_1,\boldsymbol{\beta}_1)-\hat{S}(s_0,\boldsymbol{\beta}_0)=(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)'\left\{-\int_{-\infty}^{s}\frac{A_1(s,\boldsymbol{\beta}_0)}{A_2(s,\boldsymbol{\beta}_0)}d\lambda(s)\times S(s_0)\right\}-f(s_0,\boldsymbol{\beta}_0)(s_1-s_0)+o(n^{-1/2},||\boldsymbol{\beta}_1-\boldsymbol{\beta}_0||+|s_1-s_0|)$.

Finally, note that

$$\begin{aligned}
\hat{S}(s_0,\boldsymbol{\beta}_0)-S(s_0) &= S(s_0)\int_{-\infty}^{s_0}\frac{\sum_i dM_i(u,\boldsymbol{\beta}_0)}{\bar{Y}(u,\boldsymbol{\beta}_0)}+o_p(n^{-1/2})\\
&= n^{-1/2}S(s_0)Z(s_0)+o_p(n^{-1/2}),
\end{aligned}$$

where the last equality comes from the Martingale CLT, $M_i(u,\boldsymbol{\beta}_0)=N_i(u,\boldsymbol{\beta}_0)-\int_{-\infty}^{u}Y_i(u,\boldsymbol{\beta}_0)d\Lambda(u,\boldsymbol{\beta}_0)$ and $Z(s)$ is a version of $W(v(s))$, where $W(\cdot)$ is the Wiener process and

$$v(t)=\int_{-\infty}^{t}\lambda(s,\boldsymbol{\beta}_0)ds/\pi(s,\boldsymbol{\beta}_0). \tag{S1.2}$$

Here, $\pi(s,\boldsymbol{\beta}_0)=plim\frac{1}{n}\bar{Y}(s,\boldsymbol{\beta}_0)=S(s)\Gamma^{(0)}(s,\boldsymbol{\beta}_0)$.

Hence, the result follows by denoting $\mathcal{A}(s_0,\boldsymbol{\beta}_0)=-\int_{-\infty}^{s}\frac{\Gamma^{(1)}(s,\boldsymbol{\beta}_0)}{\Gamma^{(0)}(s,\boldsymbol{\beta}_0)}d\lambda(s)$. □

S2 Proof of Proposition 1

Denote by $\boldsymbol{\beta}_0$ the truth and

$$\tilde{Y}_i^0=E(Y_i|Y_i^*,\delta_i,\mathbf{X}_i,\boldsymbol{\beta}_0)=Y_i^*+(1-\delta_i)\frac{\int_{e_i(\boldsymbol{\beta}_0)}^{\infty}S(s)ds}{S\{e_i(\boldsymbol{\beta}_0)\})},$$

where S is the (true) survival function corresponding to the distribution function F, ie $S(.) = 1 - F(.)$. Then

$$\sum_{i=1}^{n} \mathbf{X}_i \left\{ \hat{Y}_i(\hat{\boldsymbol{\beta}}^{(0)}) - Y_i \right\} = \sum_{i=1}^{n} \mathbf{X}_i \left\{ \hat{Y}_i(\hat{\boldsymbol{\beta}}^{(0)}) - \tilde{Y}_i^0 \right\} + \sum_{i=1}^{n} \mathbf{X}_i(\tilde{Y}_i^0 - Y_i) \tag{S2.1}$$

The second term on the right hand side of (**??**) is $O_p(n^{1/2})$ by the CLT, we only need to consider the first term on the right hand side of (**??**), which is equal to

$$\sum_i \mathbf{X}_i(1-\delta_i) \left\{ \frac{\int_{e_i(\hat{\boldsymbol{\beta}}^{(0)})}^{\infty} \hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) ds}{\hat{S}(e_i(\hat{\boldsymbol{\beta}}^{(0)}), \hat{\boldsymbol{\beta}}^{(0)})} - \frac{\int_{e_i(\boldsymbol{\beta}_0)}^{\infty} S(s) ds}{S\{e_i(\boldsymbol{\beta}_0)\})} \right\}, \tag{S2.2}$$

where $\hat{S}(t, \boldsymbol{\beta})$ is the Nelson-Aalen estimator based on data $(Y_i^* - \mathbf{X}_i'\boldsymbol{\beta}, \delta_i), i = 1, \ldots, n$.

Equation (**??**) is asymptotically equal to

$$\sum_i \frac{\mathbf{X}_i(1-\delta_i)}{S\{e_i(\boldsymbol{\beta}_0)\}} \left\{ \int_{e_i(\hat{\boldsymbol{\beta}}^{(0)})}^{\infty} \hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) ds - \int_{e_i(\boldsymbol{\beta}_0)}^{\infty} S(s) ds \right\} \tag{S2.3}$$

$$- \sum_i \frac{\mathbf{X}_i(1-\delta_i) \int_{e_i(\boldsymbol{\beta}_0)}^{\infty} S(s) ds}{S^2\{e_i(\boldsymbol{\beta}_0)\}} \left\{ \hat{S}(e_i(\hat{\boldsymbol{\beta}}^{(0)}), \hat{\boldsymbol{\beta}}^{(0)}) - S\{e_i(\boldsymbol{\beta}_0)\} \right\} + O_p(1) \tag{S2.4}$$

Next consider (**??**). Note that

$$\begin{aligned}
& \int_{e_i(\hat{\boldsymbol{\beta}}^{(0)})}^{\infty} \hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) ds - \int_{e_i(\boldsymbol{\beta}_0)}^{\infty} S(s) ds \\
= & \int_{e_i(\boldsymbol{\beta}_0)}^{e_i(\hat{\boldsymbol{\beta}}^{(0)})} \hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) ds + \int_{e_i(\boldsymbol{\beta}_0)}^{\infty} \{\hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) - S(s)\} ds \\
= & \int_{e_i(\boldsymbol{\beta}_0)}^{e_i(\hat{\boldsymbol{\beta}}^{(0)})} S\{e_i(\boldsymbol{\beta}_0)\} ds + \int_{e_i(\boldsymbol{\beta}_0)}^{e_i(\hat{\boldsymbol{\beta}}^{(0)})} [\hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) - S\{e_i(\boldsymbol{\beta}_0)\}] ds \\
& + \int_{e_i(\boldsymbol{\beta}_0)}^{\infty} \{\hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) - S(s)\} ds.
\end{aligned}$$

It is obvious the first term in the above equation is equal to

$$\int_{e_i(\boldsymbol{\beta}_0)}^{e_i(\hat{\boldsymbol{\beta}}^{(0)})} S\{e_i(\boldsymbol{\beta}_0)\}) ds = S\{e_i(\boldsymbol{\beta}_0)\}\{e_i(\hat{\boldsymbol{\beta}}^{(0)}) - e_i(\boldsymbol{\beta}_0)\} = -S\{e_i(\boldsymbol{\beta}_0)\}\mathbf{X}_i'(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0).$$

For the second term, applying Lemma 1 and noting that $\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$ yields

$$\int_{e_i(\boldsymbol{\beta}_0)}^{e_i(\hat{\boldsymbol{\beta}}^{(0)})} [\hat{S}(s, \hat{\boldsymbol{\beta}}^{(0)}) - S\{e_i(\boldsymbol{\beta}_0)\}] ds = o_p(n^{-1/2}).$$

Finally, for the third term as

$$\sqrt{n}\{\hat{S}(s,\hat{\boldsymbol{\beta}}^{(0)}) - S(s)\} \to S(s)Z(t)$$

weakly, where $Z(t)$ is a version of $W(v(t))$, by the weak convergence of stochastic integrals (e.g Theorem 2.2 of Kurtz and Protter (1991)) and the Skorohod representation theorem, we have that

$$\int_{e_i(\boldsymbol{\beta}_0)}^{\infty} (\hat{S}(s,\hat{\boldsymbol{\beta}}^{(0)}) - S(s))ds = n^{-1/2}\int_{e_i(\boldsymbol{\beta}_0)}^{\infty} S(t)Z(t)dt + o_p(n^{-1/2}).$$

When applying Theorem 2.2 of Kurtz and Protter (1991), we need to verify that the variance of the integrand of the last integral (or the "change of the time" in the Gaussian process), which is $var\{S(t)Z(t)\} = S^2(t)v(t)$ is bounded at ∞. That is $\limsup_{t\to\infty} S^2(t)v(t) < \infty$. Indeed,

$$S^2(t)v(t) < \int_{-\infty}^{t} S^2(s)\frac{\lambda_0(s,\boldsymbol{\beta}_0)}{\pi(s,\boldsymbol{\beta}_0)}ds < \int_{-\infty}^{\infty}\frac{dF(s,\boldsymbol{\beta}_0)}{\Gamma^{(0)}(s,\boldsymbol{\beta}_0)} < \infty$$

by the regularity condition. Hence, (**??**) is equal (in distribution) to

$$\begin{aligned} & \sum_i \mathbf{X}_i(1-\delta_i)\Big\{(\tilde{A}_i(\boldsymbol{\beta}_0) - S\{e_i(\boldsymbol{\beta}_0)\}\mathbf{X}_i)'(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0) \\ & + n^{-1/2}\int_{e_i(\boldsymbol{\beta}_0)}^{\infty}\frac{S(t)}{S\{e_i(\boldsymbol{\beta}_0)\}}W(v(t))dt\Big\} + o_p(n^{1/2}) \\ = \; & O_p(n^{1/2}), \end{aligned}$$

where $\tilde{A}_i(\boldsymbol{\beta}_0) = \int_{e_i(\boldsymbol{\beta}_0)}^{\infty}\frac{S(s)}{S\{e_i(\boldsymbol{\beta}_0)\}}\mathcal{A}(\boldsymbol{\beta}_0,s)ds$ and the last equality stems from that $(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0) = O_p(n^{-1/2})$.

Finally consider (**??**). Using the Lemma, it is follows that

$$\begin{aligned} & \hat{S}(e_i(\hat{\boldsymbol{\beta}}^{(0)}),\hat{\boldsymbol{\beta}}^{(0)}) - S\{e_i(\boldsymbol{\beta}_0)\}) \\ \stackrel{d}{=} \; & S\{e_i(\boldsymbol{\beta}_0)\}\{\mathcal{A}(e_i(\boldsymbol{\beta}_0),\boldsymbol{\beta}_0) - \lambda(e_i(\boldsymbol{\beta}_0),\boldsymbol{\beta}_0)\mathbf{X}_i\}'(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0) + n^{-1/2}W(v(e_i(\boldsymbol{\beta}_0)))\} + o_p(n^{-1/2}), \end{aligned}$$

where $\stackrel{d}{=}$ is for equal in distribution.

Hence, (**??**) is equal, in distribution, to

$$\begin{aligned} & \sum_i \mathbf{X}_i(1-\delta_i)\frac{\int_{e_i(\boldsymbol{\beta}_0)}^{\infty} S(s)ds}{S\{e_i(\boldsymbol{\beta}_0)\}}\Big[\{\mathcal{A}(e_i(\boldsymbol{\beta}_0),\boldsymbol{\beta}_0) + \lambda(e_i(\boldsymbol{\beta}_0),\boldsymbol{\beta}_0)\mathbf{X}_i)\}(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0) \\ & + n^{-1/2}W(v(e_i(\boldsymbol{\beta}_0)))\Big] + o_p(n^{1/2}) \\ = \; & \sum_i \mathbf{X}_i(\tilde{Y}_i(\boldsymbol{\beta}_0) - Y_i^*)[\{\mathcal{A}(e_i(\boldsymbol{\beta}_0),\boldsymbol{\beta}_0) + \lambda(e_i(\boldsymbol{\beta}_0),\boldsymbol{\beta}_0)\mathbf{X}_i)\}(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_0) + O_p(n^{-1/2})] + o_p(n^{1/2}) \\ = \; & O_p(n^{1/2}). \end{aligned}$$

Combining (**??**) and (**??**) yields the result. □

S3 Proof of Lemma 2

Define the Lagrangian

$$L(\boldsymbol{\beta},\boldsymbol{\mu}) = ||\mathbf{W}\boldsymbol{\beta}||_1 + \boldsymbol{\mu}'\mathbf{Z}'(\hat{\mathbf{Y}} - \mathbf{Z}\mathbf{W}\boldsymbol{\beta}) - \lambda||\boldsymbol{\mu}||_1.$$

Then (11) and (12) imply

$$||\mathbf{W}\hat{\boldsymbol{\beta}}||_1 = L(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\mu}}) = \hat{\boldsymbol{\mu}}'\mathbf{Z}'\hat{\mathbf{Y}} - \lambda||\hat{\boldsymbol{\mu}}||_1.$$

Since

$$\begin{aligned}\inf_{\boldsymbol{\beta}} L(\boldsymbol{\beta},\boldsymbol{\mu}) &= \boldsymbol{\mu}'\mathbf{Z}'\hat{\mathbf{Y}} - \lambda||\boldsymbol{\mu}||_1 + \inf_{\boldsymbol{\beta}}\ (\mathrm{sgn}(\boldsymbol{\beta}) - \boldsymbol{\mu}\mathbf{Z}'\mathbf{Z})'\mathbf{W}\boldsymbol{\beta} \\ &= \begin{cases} \boldsymbol{\mu}'\mathbf{Z}'\hat{Y} - \lambda||\boldsymbol{\mu}||_1 & \text{if } ||\mathbf{Z}'\mathbf{Z}\boldsymbol{\mu}||_\infty \le 1 \\ -\infty & \text{otherwise,}\end{cases}\end{aligned}$$

and because (10) holds, we have

$$||\mathbf{W}\hat{\boldsymbol{\beta}}||_1 = \inf_{\boldsymbol{\beta}} L(\boldsymbol{\beta},\hat{\boldsymbol{\mu}}) \le \sup_{\boldsymbol{\mu}}\inf_{\boldsymbol{\beta}} L(\boldsymbol{\beta},\boldsymbol{\mu}) \le \sup_{\boldsymbol{\mu}} L(\tilde{\boldsymbol{\beta}},\boldsymbol{\mu})$$

for any $\tilde{\boldsymbol{\beta}}$. This, the inequality (9), and the fact that

$$\begin{aligned}\sup_{\boldsymbol{\mu}} L(\boldsymbol{\beta},\boldsymbol{\mu}) &= ||\mathbf{W}\boldsymbol{\beta}||_1 + \sup_{\boldsymbol{\mu}}\ \boldsymbol{\mu}'[\mathbf{Z}'(\hat{\mathbf{Y}} - \mathbf{Z}\mathbf{W}\boldsymbol{\beta}) - \lambda\mathrm{sgn}(\boldsymbol{\mu})] \\ &= \begin{cases} ||\mathbf{W}\boldsymbol{\beta}||_1 & \text{if } ||\mathbf{Z}'(\hat{\mathbf{Y}} - \mathbf{Z}\mathbf{W}\boldsymbol{\beta})||_\infty \le \lambda \\ \infty & \text{otherwise}\end{cases}\end{aligned}$$

imply that $||\mathbf{W}\hat{\boldsymbol{\beta}}||_1 \le ||\mathbf{W}\boldsymbol{\beta}||_1$ whenever $|\mathbf{Z}'(\hat{\mathbf{Y}} - \mathbf{Z}\mathbf{W}\boldsymbol{\beta})| \le \lambda$. This means that $\hat{\boldsymbol{\beta}}$ solves (ADSC). □

S4 Proof of Proposition 2

With $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ as in (13)-(16), we check that (9)-(12) hold with probability tending to 1. First note that

$$\mathbf{Z}_A'\mathbf{Z}\hat{\boldsymbol{\mu}} = \mathbf{Z}_A'\mathbf{Z}_A\hat{\boldsymbol{\mu}}_A = \mathrm{sgn}(\boldsymbol{\beta}_0)_A$$

and

$$\begin{aligned}\mathbf{Z}_{\bar{A}}'\mathbf{Z}\hat{\mu} &= \mathbf{Z}_{\bar{A}}\mathbf{Z}_A(\mathbf{Z}_A'\mathbf{Z}_A)^{-1}\mathrm{sgn}(\boldsymbol{\beta}_0)_A \\ &= (\mathbf{W}_{\bar{A},\bar{A}}^{-1})'\mathbf{X}_{\bar{A}}'\mathbf{P}_n\mathbf{X}_A(\mathbf{X}_A'\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{W}_{A,A}\mathrm{sgn}(\boldsymbol{\beta}_0)_A \\ &= o_P(1).\end{aligned}$$

This implies that (10) holds with probability tending to 1. To see that (9) holds with probability approaching 1, first observe that

$$\mathbf{Z}_A'(\hat{\mathbf{Y}} - \mathbf{Z}\mathbf{W}\hat{\boldsymbol{\beta}}) = \lambda\mathrm{sgn}(\hat{\boldsymbol{\mu}})_A. \tag{S4.1}$$

Furthermore,

$$\begin{aligned}\mathbf{Z}'_{\bar{A}}(\hat{\mathbf{Y}}-\mathbf{ZW}\hat{\boldsymbol{\beta}}) &= \mathbf{Z}'_{\bar{A}}\left[\mathbf{I}-\mathbf{Z}_A(\mathbf{Z}'_A\mathbf{Z}_A)^{-1}\mathbf{Z}'_A\right]\hat{\mathbf{Y}}+\lambda\mathbf{Z}'_{\bar{A}}\mathbf{Z}_A(\mathbf{Z}'_A\mathbf{Z}_A)^{-1}\mathrm{sgn}(\hat{\boldsymbol{\mu}})_A \\ &= \mathbf{W}^{-1}_{\bar{A},\bar{A}}\mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\hat{\mathbf{Y}} \qquad \text{(S4.2)} \\ &\quad +\lambda\mathbf{W}^{-1}_{\bar{A},\bar{A}}\mathbf{X}_{\bar{A}}\mathbf{P}_n\mathbf{X}_A(\mathbf{X}_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{W}_{A,A}\mathrm{sgn}(\hat{\boldsymbol{\mu}})_A.\end{aligned}$$

Proposition 1 implies that

$$\begin{aligned}\mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\hat{\mathbf{Y}} &= \mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\mathbf{Y} \\ &\quad +O_P(\sqrt{n}) \\ &= O_P(\sqrt{n}),\end{aligned}$$

where the second equality above holds because

$$\mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\mathbf{Y}=\mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\epsilon$$

and

$$\left|\mathbf{X}'_j\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\mathbf{P}_n\mathbf{X}_j\right|\le\mathbf{X}'_j\mathbf{X}_j,$$

which implies that

$$\frac{1}{\sqrt{n}}\mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\mathbf{Y}$$

has mean 0 and bounded variance. Since $\mathbf{W}^{-1}_{\bar{A},\bar{A}}=o_P(\lambda/\sqrt{n})$, it follows that

$$\mathbf{W}^{-1}_{\bar{A},\bar{A}}\mathbf{X}'_{\bar{A}}\mathbf{P}_n\left[\mathbf{I}-\mathbf{P}_n\mathbf{X}_A(\mathbf{X}'_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{P}_n\right]\hat{\mathbf{Y}}=o_P(\lambda).$$

Combining this with (**??**) and the fact that

$$\lambda\mathbf{W}^{-1}_{\bar{A},\bar{A}}\mathbf{X}_{\bar{A}}\mathbf{P}_n\mathbf{X}_A(\mathbf{X}_A\mathbf{P}_n\mathbf{X}_A)^{-1}\mathbf{W}_{A,A}\mathrm{sgn}(\hat{\boldsymbol{\mu}})_A=o_P(\lambda)$$

gives

$$\mathbf{Z}'_{\bar{A}}(\hat{\mathbf{Y}}-\mathbf{ZW}\hat{\boldsymbol{\beta}})=o_P(\lambda).$$

This fact, plus (**??**), implies that (9) holds with probability tending to 1. Since

$$\hat{\boldsymbol{\mu}}'\mathbf{Z}'\mathbf{ZW}\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\mu}}'_A\mathbf{Z}'_A\mathbf{Z}_A\mathbf{W}_{A,A}\hat{\boldsymbol{\beta}}_A=\mathrm{sgn}(\boldsymbol{\beta})'_A\mathbf{W}_{A,A}\hat{\boldsymbol{\beta}}_A$$

and $\mathrm{sgn}(\hat{\boldsymbol{\beta}})_A\xrightarrow{P}\mathrm{sgn}(\boldsymbol{\beta}_0)_A$, the probability that (refeq9) holds converges to 1. Lastly,

$$\hat{\boldsymbol{\mu}}'\mathbf{Z}'(\mathbf{P}_n\hat{\mathbf{Y}}-\mathbf{ZW}\hat{\boldsymbol{\beta}})=\hat{\boldsymbol{\mu}}'_A\mathbf{Z}'_A(\mathbf{P}_n\hat{\mathbf{Y}}-\mathbf{ZW}\hat{\boldsymbol{\beta}})=\lambda\hat{\boldsymbol{\mu}}'_A\mathrm{sgn}(\hat{\boldsymbol{\mu}})_A=\lambda||\hat{\boldsymbol{\mu}}||_1,$$

which implies that (12) holds. We conclude that (9)-(12) hold with probability tending to 1 and the proposition is proved. □

S5 Proof of Corollary 1

Let $\hat{\boldsymbol{\beta}}$ be any sequence of solutions to (ADSC), let $T = \{j;\ \hat{\beta}_j^1 \neq 0\}$ and let $E = \{j;\ |\mathbf{Z}_j'(\hat{\mathbf{Y}} - \mathbf{Z}\mathbf{W}\hat{\boldsymbol{\beta}})| = \lambda\}$. Proposition 2 implies that by slightly perturbing $\hat{\boldsymbol{\beta}}$ if necessary, we can assume that $E \subseteq A \subseteq T$. The conditions (9)-(12) in Lemma 2 imply that there exists $\mathbf{t} \in \{\pm 1\}^{|T|}$ such that

$$||\mathbf{W}^{-1}\mathbf{X}'\mathbf{P}_n\mathbf{X}_T(\mathbf{X}_T'\mathbf{P}_n\mathbf{X}_T)^{-1}\mathbf{W}_{T,T}\mathbf{t}||_\infty \leq 1. \tag{S5.1}$$

Since $w_j/w_k \xrightarrow{P} \infty$, whenever $j \in \bar{A}$ and $k \in A$, it follows that $T = A$, with probability tending to 1. Thus, $P(T = A) \to 1$ and $\hat{\boldsymbol{\beta}}$ is consistent for model selection. □

S6 Proof of Proposition 3

Since $\{T = A\} \subset \left\{\hat{\boldsymbol{\beta}}^{(0,T)} = \hat{\boldsymbol{\beta}}^{(0,A)}\right\}$, coupled with $P(T = A) \to 1$ implied by Proposition 2, it follows immediately that

$$P\left(\hat{\boldsymbol{\beta}}^{(0,T)} = \hat{\boldsymbol{\beta}}^{(0,A)}\right) \to 1.$$

Therefore, $\sqrt{n}(\hat{\boldsymbol{\beta}}_A^{(0,T)} - \boldsymbol{\beta}_{0,A}) = \sqrt{n}(\hat{\boldsymbol{\beta}}_A^{(0,A)} - \boldsymbol{\beta}_{0,A}) + o_P(1)$. Further, as Theorem 4 of Lai and Ying (1991) implies that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_A^{(0,A)} - \boldsymbol{\beta}_{0,A}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}),$$

the original claim is thus proved. □

References

Kurtz, T and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* **19**, 1035–1070.

Lai, T. and Ying, Z. (1991). Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis with Censored Data. *Annals of Statistics* **19**, 1370–1402.

Ying, Z. (1993). A Large Sample Study of Rank Estimation for Censored Regression Data. *Annals of Statistics* **21**, 76–99.