Supplementary Materials for

Subgroup Analysis in Censored Linear Regression

Xiaodong Yan

School of Economics, Shandong University, Jinan, China

Guosheng Yin

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong

Xingqiu Zhao

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

In the supplementary materials, we provide the proofs for Proposition 1, and Theorems 1–3.

Proof of Proposition 1

By the definition of $Q(\eta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu})$, we have

$$Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\nu}^{(k+1)} \mid \boldsymbol{\theta}^{(k)}) - Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) = \frac{1}{\varphi}\|\boldsymbol{\nu}^{(k+1)} - \boldsymbol{\nu}^{(k)}\|^2. \tag{A.1}$$

Since $\boldsymbol{\alpha}^{(k+1)}$ is the minimizer of $Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}, \boldsymbol{\nu}^{(k)})$, we have

$$Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) - Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) \leq 0. \tag{A.2}$$

Moreover, $\boldsymbol{\beta} \mapsto Q(\eta^{(k+1)}, \boldsymbol{\beta}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)})$ and $\eta \mapsto Q(\eta, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)})$ are both convex, because the Hessian matrix $(\widetilde{\mathbb{X}}\mathcal{Q}_Z\boldsymbol{X} + \varphi\Omega^\top\Omega)$ and $\widetilde{\mathbb{Z}}\boldsymbol{Z}$ are both positive definite. Thus there exist constants $c_1 > 0$ and $c_2 > 0$ such that the following inequalities hold:

$$Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) - Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) \leq -\frac{c_1}{2}\|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\|^2 \tag{A.3}$$

and

$$Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) - Q(\eta^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) \leq -\frac{c_2}{2}\|\eta^{(k+1)} - \eta^{(k)}\|^2. \tag{A.4}$$

Summing (A.1)-(A.4), we have

$$\begin{aligned} & Q(\eta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\nu}^{(k+1)} \mid \boldsymbol{\theta}^{(k)}) - Q(\eta^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}) \\ & \leq \frac{1}{\varphi}\|\boldsymbol{\nu}^{(k+1)} - \boldsymbol{\nu}^{(k)}\|^2 - \frac{c_1}{2}\|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\|^2 - \frac{c_2}{2}\|\eta^{(k+1)} - \eta^{(k)}\|^2. \end{aligned} \tag{A.5}$$

Since $\{\boldsymbol{\alpha}^{(k)}\}_{k=1}^{\infty}$ is bounded, by the ADMM iterative procedure, $\boldsymbol{\beta}^{(k)}$ and $\eta^{(k)}$ are also both bounded. Thus $Q(\eta^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)})$ and $\{\eta^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)}\}_{k=1}^{\infty}$ are bounded. For convenience, we note

$$\mathcal{A}^{(k)} = Q(\eta^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)} \mid \boldsymbol{\theta}^{(k)}), \mathcal{B}^{(k)} = \frac{c_1}{2}\|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\|^2 + \frac{c_2}{2}\|\eta^{(k+1)} - \eta^{(k)}\|^2, \mathcal{C} = \frac{1}{\varphi}\|\nu^{(k+1)} - \boldsymbol{\nu}^{(k)}\|^2.$$

Since $\mathcal{A}^{(k)}$ is bounded, then there exists a subsequence $\{\mathcal{A}^{(k_j)}\}$, such that

$$\lim_{k_j \to \infty} \mathcal{A}^{(k_j)} = \text{liminf}_{k \to \infty} \mathcal{A}^{(k)}.$$

By Lemma A.5 and $\lim_{k \to \infty} \mathcal{C}^{(k)} \to 0$, we have

$$\begin{aligned} \text{liminf}_{k_j \to \infty} \mathcal{A}^{(k_j)} & \leq \text{liminf}_{k_j \to \infty} (\mathcal{A}^{(k_j)} - \mathcal{A}^{(k_j+1)} + \mathcal{C}^{(k_j)}) \\ & = \text{liminf}_{k \to \infty} \mathcal{A}^{(k)} - \text{liminf}_{k_j \to \infty} \mathcal{A}^{(k_j+1)} \leq 0. \end{aligned}$$

As $\mathcal{B}^{(k_j)} \geq 0$, thus $\text{liminf}_{k_j \to \infty} \mathcal{B}^{(k_j)} = 0$, which means

$$\text{liminf}_{k_j \to \infty} \{c_1\|\boldsymbol{\beta}^{(k_j+1)} - \boldsymbol{\beta}^{(k_j)} + c_2\|\eta^{(k_j+1)} - \eta^{(k_j)}\|\} = 0,$$

together with the last step of ADMM iteration and $\|\boldsymbol{\nu}^{(k+1)} - \boldsymbol{\nu}^{(k)}\| \to 0$, we have

$$\text{liminf}_{k_j \to \infty} \|\boldsymbol{\alpha}^{(k_j+1)} - \boldsymbol{\alpha}^{(k_j)}\| = 0.$$

Therefore, the sequence $\{\eta^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\nu}^{(k)}\}_{k=1}^{\infty}$ has a subsequence $\{\eta^{(k_j)}, \boldsymbol{\beta}^{(k_j)}, \boldsymbol{\alpha}^{(k_j)}, \boldsymbol{\nu}^{(k_j)}\}_{k_j=1}^{\infty}$ which converges to a point $\{\eta^*, \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\nu}^*\}$, and we have

$$\beta_i^* - \beta_j^* - \alpha_{ij}^* = 0, \forall 1 \leq i < j \leq n.$$

Define

$$W_F(t) = t - \frac{\int_t^\infty s dF(s)}{1 - F(t)}, \qquad W_F(t, h) = h(t) - \frac{\int_t^\infty h(s) dF(s)}{1 - F(t)},$$

$$\mathcal{M}(s,t \mid F) = I(t \le s) - \frac{\int_{-\infty}^{s} I(t \ge u)dF(u)}{1 - F(s-)},$$

and

$$\mathcal{S}(s,t \mid F) = tI(t \le s) + \frac{\int_{s}^{\infty} udF(u)}{1 - F(s)} I(t > s).$$

The Buckley–James type least squares estimating function for the oracle estimator $\widehat{\boldsymbol{\phi}}^{or}$ is equivalent to

$$\Psi_n(\boldsymbol{\phi}) = n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi}) \ge u)(U_i - \widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u))W_{\widetilde{F}_{\boldsymbol{\phi}}}(u)d\mathcal{M}(u,\epsilon_i(\boldsymbol{\phi}) \mid \widetilde{F}_{\boldsymbol{\phi}})$$

(Proposition 3.2 of Ritov (1990)). Define

$$\widetilde{\Psi}_n(\boldsymbol{\phi}) = n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi}) \ge u)(U_i - \mathbb{D}_{\boldsymbol{\phi},i}(u))W_{F_{\boldsymbol{\phi}}}(u)d\mathcal{M}(u,\epsilon_i(\boldsymbol{\phi}) \mid F_{\boldsymbol{\phi}}). \tag{A.6}$$

Lemma 1. *For a given small constant* ε*,*

(i) $\sup\{\left|W_{\widetilde{F}_{\boldsymbol{\phi}}}(t) - W_{F_{\boldsymbol{\phi}}}(t)\right| : \|\boldsymbol{\phi}\| \le \kappa, \sum_{i=1}^{n} I(\upsilon_i(\boldsymbol{\phi}) \ge s) \ge \frac{cn^{1-\varsigma}}{2}, t \le s \le b_0\} = O(n^{-1/2+4\varsigma+\varepsilon})$ *a.s., then* $\sup\{\left|W_{\widetilde{F}_{\boldsymbol{\theta}}}(t) - W_{F_{\boldsymbol{\theta}}}(t)\right| : \sup_i \|\theta_i\| \le \kappa, \sum_{i=1}^{n} I(\upsilon_i(\theta_i) \ge s) \ge \frac{cn^{1-\varsigma}}{2}, t \le s \le b_0\} = O(n^{-1/2+4\varsigma+\varepsilon})$ *a.s.*

(ii) $\sup\{\|n^{-1}\sum_{i=1}^{n}[\delta_i U_i - \delta_i \mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))]\| : \|\boldsymbol{\phi}\| \le \kappa\} = O(n^{-1/2+\varepsilon})$ *a.s.*

(iii) $\sup\{\|\widehat{D}_{\boldsymbol{\phi}}^{(j)}(u) - D_{\boldsymbol{\phi}}^{(j)}(u)\| : u \le b_0, \|\boldsymbol{\phi}\| \le \kappa, j = 1,2\} = O(n^{-1/2+\varepsilon})$ *a.s.*

Proof of Lemma 1

By Lemma 2 of Lai and Ying (1991), we have

$$\begin{aligned}&\sup\Big\{\Big|\frac{\int_t^{b_0} sd\widetilde{F}_{\boldsymbol{\phi}}(s)}{1 - \widetilde{F}_{\boldsymbol{\phi}}(t)} - \frac{\int_t^{b_0} sdF_{\boldsymbol{\phi}}(s)}{1 - F_{\boldsymbol{\phi}}(t)}\Big| : \|\boldsymbol{\phi}\| \le \kappa, t \le s \le b_0, \sum_{i=1}^{n} I(\upsilon_i(\boldsymbol{\phi}) \ge s) \ge \frac{cn^{1-\varsigma}}{2}\Big\}\\ &= O(n^{-1/2+4\varsigma+\varepsilon}) \quad a.s.\end{aligned}$$

for every $0 \le \varsigma < 1$ and $\varepsilon > 0$, and thus Lemma 1 (i) holds. We obtain Lemma 1 (ii) using

$$U_i - \mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi})) = [(Z_i - D_{\boldsymbol{\phi}}^{(1)}(\epsilon_i(\boldsymbol{\phi})))^\top, (X_i - D_{\boldsymbol{\phi}}^{(2)}(\epsilon_i(\boldsymbol{\phi})))^\top \pi_{i1}, \cdots, (X_i - D_{\boldsymbol{\phi}}^{(2)}(\epsilon_i(\boldsymbol{\phi})))^\top \pi_{iR}]^\top$$

with

$$D_{\boldsymbol{\phi}}^{(1)}(u) = E[Z_i \mid Y_i^* - U_i^\top \boldsymbol{\phi} \ge u] = E[Z_i \mid Y_i^* - U_i^\top \boldsymbol{\phi} \ge u, \delta_i = 1]$$

and

$$D^{(2)}_{\boldsymbol{\phi}}(u) = E[X_i \mid Y_i^* - U_i^\top \boldsymbol{\phi} \geq u] = E[X_i \mid Y_i^* - U_i^\top \boldsymbol{\phi} \geq u, \delta_i = 1]$$

(Lai and Ying, 1991) and

$$E(\delta_i Z_i) = E[\delta_i D^{(1)}_{\boldsymbol{\phi}}(\epsilon_i(\boldsymbol{\phi}))], \quad E[\delta_i X_i] = E[\delta_i D^{(2)}_{\boldsymbol{\phi}}(\epsilon_i(\boldsymbol{\phi}))].$$

We conclude Lemma 1 (iii) from the definitions of $D^{(1)}_{\boldsymbol{\phi}}(u)$ and $D^{(2)}_{\boldsymbol{\phi}}(u)$.

Lemma 2. $\sup_{\|\boldsymbol{\phi}\|\leq\kappa} \|\Psi_n(\boldsymbol{\phi}) - \widetilde{\Psi}_n(\boldsymbol{\phi})\| = O(n^{-1/2+3\varsigma+2\varepsilon})$ *a.s.*

Proof of Lemma 2

Note that $\Psi_n(\boldsymbol{\phi}) - \widetilde{\Psi}_n(\boldsymbol{\phi}) = J_{1n}(\boldsymbol{\phi}) + J_{2n}(\boldsymbol{\phi}) + J_{3n}(\boldsymbol{\phi})$, where

$$J_{1n}(\boldsymbol{\phi}) = n^{-1/2}\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}) \geq u)(U_i - \widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u))\{W_{\widetilde{F}_{\boldsymbol{\phi}}}(u) - W_{F_{\boldsymbol{\phi}}}(u)\} d\mathcal{M}(u, \epsilon_i(\boldsymbol{\phi}) \mid \widetilde{F}_{\boldsymbol{\phi}}),$$
$$J_{2n}(\boldsymbol{\phi}) = n^{-1/2}\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}) \geq u)(U_i - \widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)) W_{F_{\boldsymbol{\phi}}}(u) d\{\mathcal{M}(u, \epsilon_i(\boldsymbol{\phi}) \mid \widetilde{F}_{\boldsymbol{\phi}}) - \mathcal{M}(u, \epsilon_i(\boldsymbol{\phi}) \mid F_{\boldsymbol{\phi}})\},$$
$$J_{3n}(\boldsymbol{\phi}) = n^{-1/2}\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}) \geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u) - \widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\} W_{F_{\boldsymbol{\phi}}}(u) d\mathcal{M}(u, \epsilon_i(\boldsymbol{\phi}) \mid F_{\boldsymbol{\phi}}).$$

For J_{1n}, we consider the process

$$L_n(\boldsymbol{\phi}) = J_{1n}(\boldsymbol{\phi}) - Q_{1n}(\boldsymbol{\phi}),$$

where $Q_{1n}(\boldsymbol{\phi}) = n^{-1/2}\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}) \geq u)\{W_{\widetilde{F}_{\boldsymbol{\phi}}}(u) - W_{F_{\boldsymbol{\phi}}}(u)\}\{\mathbb{D}_{\boldsymbol{\phi},i}(u) - \widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\} dN_i(\boldsymbol{\phi}, u)$, and $N_i(\boldsymbol{\phi}, u) = I(\epsilon_i(\boldsymbol{\phi}) \leq u)$. By Lemma 1 (i) and (ii), we have

$$\begin{aligned}
\big\|L_n(\boldsymbol{\phi})\big\|\| &= \big|\|n^{-1/2}\sum_{i=1}^n \{W_{\widetilde{F}_{\boldsymbol{\phi}}}(\epsilon_i(\boldsymbol{\phi})) - W_{F_{\boldsymbol{\phi}}}(\epsilon_i(\boldsymbol{\phi}))\} \\
&\quad \times \big[U_i - \{\mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi})) - \widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))\} - \sum_{j=1}^n U_j I(\upsilon_j(\boldsymbol{\phi}) \geq \epsilon_i(\boldsymbol{\phi})) \frac{d\widetilde{F}_{\boldsymbol{\phi}}(\epsilon_i(\boldsymbol{\phi}))}{1 - \widetilde{F}_{\boldsymbol{\phi}}(\epsilon_i(\boldsymbol{\phi}))}\big]\delta_i\big|\| \\
&= \big|\|n^{-1/2}\sum_{i=1}^n \{W_{\widetilde{F}_{\boldsymbol{\phi}}}(\epsilon_i(\boldsymbol{\phi})) - W_{F_{\boldsymbol{\phi}}}(\epsilon_i(\boldsymbol{\phi}))\}\big[U_i - \mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))\big]\delta_i\big|\| \\
&\leq \sup_{\|\boldsymbol{\phi}\|\leq\kappa, t\leq b_0} \big|W_{\widetilde{F}_{\boldsymbol{\phi}}}(t) - W_{F_{\boldsymbol{\phi}}}(t)\big|\|\big|n^{-1/2}\sum_{i=1}^n \{\delta_i U_i - \delta_i \mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))\}\big|\| \\
&= O(n^{-1/2+3\varsigma+2\varepsilon}) \ a.s.
\end{aligned}$$

On the other hand, using Lemma 1 (i) and (iii), we have

$$\begin{aligned}
\|Q_{1n}(\boldsymbol{\phi})\| &\leq \|n^{-1/2}\sum_{i=1}^{n}\int\{W_{\widetilde{F}_{\boldsymbol{\phi}}}(u)-W_{F_{\boldsymbol{\phi}}}(u)\}\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}dN_i(\boldsymbol{\phi},u)\| \\
&= \|n^{-1/2}\sum_{i=1}^{n}\delta_i\{W_{\widetilde{F}_{\boldsymbol{\phi}}}(\epsilon_i(\boldsymbol{\phi}))-W_{F_{\boldsymbol{\phi}}}(\epsilon_i(\boldsymbol{\phi}))\}\{\mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))\}\| \\
&\leq \sup_{\|\boldsymbol{\phi}\|\leq\kappa,t\leq b_0}\left|W_{\widetilde{F}_{\boldsymbol{\phi}}}(t)-W_{F_{\boldsymbol{\phi}}}(t)\right|\|n^{-1/2}\sum_{i=1}^{n}(\mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi})))\| \\
&= O(n^{-1/2+3\varsigma+2\varepsilon})\ a.s.
\end{aligned}$$

Therefore, $\|J_{1n}(\boldsymbol{\phi})\|\leq\|L_n(\boldsymbol{\phi})\|+\|Q_{1n}(\boldsymbol{\phi})\|=O(n^{-1/2+3\varsigma+2\varepsilon})\ a.s.$

For J_{2n}, by $W_{F_{\boldsymbol{\phi}}}(u)\leq 2b_0$ and Lemma 1 (i) and (iii),

$$\begin{aligned}
\|J_{2n}(\boldsymbol{\phi})\| &\leq n^{-1/2}2b_0\sup\|\sum_{i=1}^{n}\delta_i[U_i-\widehat{\mathbb{D}}_{\boldsymbol{\phi}}(\epsilon_i(\boldsymbol{\phi}))]\|\sup|\frac{\int_{-\infty}^{b_0}ud\widetilde{F}_{\boldsymbol{\phi}}(u)}{1-\widetilde{F}_{\boldsymbol{\phi}}(u)}-\frac{\int_{-\infty}^{b_0}udF_{\boldsymbol{\phi}}(u)}{1-F_{\boldsymbol{\phi}}(u)}| \\
&= O(n^{-1/2+3\varsigma+2\varepsilon})\ a.s.
\end{aligned}$$

Since

$$\sum_{i=1}^{n}\{\mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))\}\epsilon_i(\boldsymbol{\phi})\delta_i=\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi})\geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}u\frac{d\widetilde{F}_{\boldsymbol{\phi}}(u)}{1-\widetilde{F}_{\boldsymbol{\phi}}(u)},$$

then J_{3n} can be written as

$$\begin{aligned}
J_{3n}(\boldsymbol{\phi}) &= n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi})\geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}d\mathcal{S}(u,\epsilon_i(\boldsymbol{\phi})\mid F_{\boldsymbol{\phi}}) \\
&= n^{-1/2}\sum_{i=1}^{n}\{\mathbb{D}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(\epsilon_i(\boldsymbol{\phi}))\}\epsilon_i(\boldsymbol{\phi})\delta_i \\
&\quad -n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi})\geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}u\frac{dF_{\boldsymbol{\phi}}(u)}{1-F_{\boldsymbol{\phi}}(u)} \\
&= n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi})\geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}u\frac{d\widetilde{F}_{\boldsymbol{\phi}}(u)}{1-\widetilde{F}_{\boldsymbol{\phi}}(u)} \\
&\quad -n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi})\geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}u\frac{dF_{\boldsymbol{\phi}}(u)}{1-F_{\boldsymbol{\phi}}(u)} \\
&= n^{-1/2}\sum_{i=1}^{n}\int I(\zeta_i(\boldsymbol{\phi})\geq u)\{\mathbb{D}_{\boldsymbol{\phi},i}(u)-\widehat{\mathbb{D}}_{\boldsymbol{\phi},i}(u)\}u\{\frac{d\widetilde{F}_{\boldsymbol{\phi}}(u)}{1-\widetilde{F}_{\boldsymbol{\phi}}(u)}-\frac{dF_{\boldsymbol{\phi}}(u)}{1-F_{\boldsymbol{\phi}}(u)}\} \\
&\leq n^{1/2}\sup_{u\leq b_0}\{\|\widehat{D}_{\boldsymbol{\phi}}^{(j)}(u)-D_{\boldsymbol{\phi}}^{(j)}(u)\|,j=1,2\}\sup|\frac{\int_{-\infty}^{b_0}ud\widetilde{F}_{\boldsymbol{\phi}}(u)}{1-\widetilde{F}_{\boldsymbol{\phi}}(u)}-\frac{\int_{-\infty}^{b_0}udF_{\boldsymbol{\phi}}(u)}{1-F_{\boldsymbol{\phi}}(u)}|=O(n^{-1/2+3\varsigma+2\varepsilon})\ a.s.
\end{aligned}$$

Hence, we complete the proof of Lemma 2.

Lemma 3. $n^{1/2}\widetilde{\Psi}_n(\boldsymbol{\phi}) = n^{1/2}\widetilde{\Psi}_n(\boldsymbol{\phi}_0) + V_n(\boldsymbol{\phi}-\boldsymbol{\phi}_0) + o\{\max(n^{1/2}, \mathbb{E}_{\max}(\boldsymbol{U}^\top\boldsymbol{U})\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|)\}$ *a.s. for* $\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\| \le n^{-\gamma}$.

Proof of Lemma 3

Set

$$\begin{aligned}
\widetilde{\Psi}_{n1}(a,\boldsymbol{\phi}) &= n^{-1/2}\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}) \ge u)(U_i - \mathbb{D}_{\boldsymbol{\phi},i}(u))W_{F_{\boldsymbol{\phi}}}(u)dF(u+aU_i),\\
\widetilde{\Psi}_{n2}(a,\boldsymbol{\phi}) &= n^{-1/2}\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}) \ge u)(U_i - \mathbb{D}_{\boldsymbol{\phi},i}(u))W_{F_{\boldsymbol{\phi}}}(u)\frac{\int_u^\infty dF(s+aU_i)}{1-F_{\boldsymbol{\phi}}(u-)}dF_{\boldsymbol{\phi}}(u).
\end{aligned}$$

Under the condition $\sup_i \|U_i\| \le c_2 + c_3$, we have

$$\widetilde{\Psi}_n(\boldsymbol{\phi}) = \widetilde{\Psi}_{n1}(\boldsymbol{\phi}-\boldsymbol{\phi}_0, \boldsymbol{\phi}) - \widetilde{\Psi}_{n2}(\boldsymbol{\phi}-\boldsymbol{\phi}_0, \boldsymbol{\phi}) + o(1)$$

for $\boldsymbol{\phi}-\boldsymbol{\phi}_0 \le n^{-\gamma}$. Taking Taylor's expansion for $F_{\boldsymbol{\phi}}(u+aU_i)$ and $F_{\boldsymbol{\phi}}(s+aU_i)$, as $\boldsymbol{\phi} \to \boldsymbol{\phi}_0$,

$$\begin{aligned}
&\widetilde{\Psi}_{n1}(\boldsymbol{\phi}-\boldsymbol{\phi}_0, \boldsymbol{\phi}) - \widetilde{\Psi}_{n1}(0,\boldsymbol{\phi})\\
=\ & n^{-1/2}\Big\{\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}_0) \ge u)U_i(U_i - \mathbb{D}_{\boldsymbol{\phi}_0,i}(u))^\top W_F(u)df(u)\Big\}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)\\
&+o(n^{-1/2}\mathbb{E}_{\max}(\boldsymbol{U}^\top\boldsymbol{U})\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|)\\
=\ & n^{-1/2}\Big\{\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}_0) \ge u)U_i(U_i - \mathbb{D}_{\boldsymbol{\phi}_0,i}(u))^\top W_F(u)\frac{f'(u)}{f(u)}dF(u)\Big\}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)\\
&+o(n^{-1/2}\mathbb{E}_{\max}(\boldsymbol{U}^\top\boldsymbol{U})\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|),
\end{aligned}$$

and

$$\begin{aligned}
&\widetilde{\Psi}_{n2}(\boldsymbol{\phi}-\boldsymbol{\phi}_0, \boldsymbol{\phi}) - \widetilde{\Psi}_{n2}(0,\boldsymbol{\phi})\\
=\ & n^{-1/2}\Big\{\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}_0) \ge u)U_i(U_i - \mathbb{D}_{\boldsymbol{\phi}_0,i}(u))^\top W_F(u)\frac{\int_u^\infty df(s)}{1-F(u-)}dF(u)\Big\}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)\\
&+o(n^{-1/2}\mathbb{E}_{\max}(\boldsymbol{U}^\top\boldsymbol{U})\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|)\\
=\ & n^{-1/2}\Big\{\sum_{i=1}^n \int I(\zeta_i(\boldsymbol{\phi}_0) \ge u)U_i(U_i - \mathbb{D}_{\boldsymbol{\phi}_0,i}(u))^\top W_F(u)\frac{\int_u^\infty \frac{f'(s)}{f(s)}dF(s)}{1-F(u-)}dF(u)\Big\}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)\\
&+o(n^{-1/2}\mathbb{E}_{\max}(\boldsymbol{U}^\top\boldsymbol{U})\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|).
\end{aligned}$$

Therefore,

$$\begin{aligned}
n^{1/2}\widetilde{\Psi}_n(\boldsymbol{\phi}) - n^{1/2}\widetilde{\Psi}_n(\boldsymbol{\phi}_0) &= n^{1/2}\{\widetilde{\Psi}_{n1}(\boldsymbol{\phi}-\boldsymbol{\phi}_0,\boldsymbol{\phi}) - \widetilde{\Psi}_{n2}(\boldsymbol{\phi}-\boldsymbol{\phi}_0,\boldsymbol{\phi})\}\\
&\quad -n^{1/2}\{\widetilde{\Psi}_{n1}(0,\boldsymbol{\phi}_0) - \widetilde{\Psi}_{n2}(0,\boldsymbol{\phi}_0)\} + o(n^{1/2})\\
&= V_n(\boldsymbol{\phi}-\boldsymbol{\phi}_0) + o(\max\{n^{1/2}, \mathbb{E}_{\max}(\boldsymbol{U}^\top\boldsymbol{U})\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|\}),
\end{aligned}$$

where

$$V_n = \sum_{i=1}^{n} \int I(\zeta_i(\boldsymbol{\phi}_0) \geq u) U_i (U_i - \mathbb{D}_{\boldsymbol{\phi}_0,i}(u))^\top W_F(u) W_F(u, f'/f) dF(u).$$

Proof of Theorem 1 (i)

Lemma 2 is equivalent to

$$\sup_{\|\boldsymbol{\phi}\| \leq \kappa} \|\Psi_n(\boldsymbol{\phi}) - \widetilde{\Psi}_n(\boldsymbol{\phi})\| = o(n^{-1/2+4\varsigma}) \ a.s. \tag{A.7}$$

under the condition $\lim_{n\to\infty} n^{1/2-4\varsigma}\Big\{ \inf_{\boldsymbol{\phi} \leq \kappa, \|\boldsymbol{\phi}-\boldsymbol{\phi}_0\| \geq n^{-\gamma}} \|\widetilde{\Psi}(\boldsymbol{\phi})\| \Big\} = \infty$ and (A.7),

$$P\{\Psi_n(\boldsymbol{\phi}) \text{ have a zero-crossing on } \|\boldsymbol{\phi} - \boldsymbol{\phi}_0\| \geq n^{-\gamma} \text{ and } \|\boldsymbol{\phi}\| \leq \kappa \text{ for large } n\} = 0.$$

Since $\Psi_n(\widehat{\boldsymbol{\phi}}^{or}) = 0$, then by Lemma 3 and conditions $\mathbb{E}_{\max}(\boldsymbol{U}^\top \boldsymbol{U}) \leq n$ and $4\varsigma + \gamma > 1$ with $\frac{1}{8} \leq \varsigma < 1$, we have

$$\sup_{\|\widehat{\boldsymbol{\phi}}^{or} - \boldsymbol{\phi}_0\| \leq n^{-\gamma}} \|\widetilde{\Psi}_n(\boldsymbol{\phi}_0) + n^{-1/2} V_n(\widehat{\boldsymbol{\phi}}^{or} - \boldsymbol{\phi}_0)\| = o(n^{-1/2+4\varsigma}) \ a.s. \tag{A.8}$$

Since $E\{\widetilde{\Psi}_n(\boldsymbol{\phi}_0)\} = 0$, we have $\|n^{1/2}\widetilde{\Psi}_n(\boldsymbol{\phi}_0)\| = O(n^{1/2+\varepsilon})$ $a.s.$ Therefore, under $\|V_n^{-1}\| \leq \frac{1}{c_4} |\mathcal{G}_{\min}|^{-1}$,

$$\|\widehat{\boldsymbol{\phi}}^{or} - \boldsymbol{\phi}_0)\| = o(\max\{n^{1/2}/\mathcal{G}_{\min}, n^{4\varsigma}/\mathcal{G}_{\min}\}) \ a.s.,$$

and

$$\|\widehat{\boldsymbol{\rho}}^{or} - \boldsymbol{\rho}_0\| = \|\widehat{\eta}^{or} - \eta_0\| = o(\max\{n^{1/2}/\mathcal{G}_{\min}, n^{4\varsigma}/\mathcal{G}_{\min}\}) \ a.s.$$

Moreover,

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}}^{or} - \boldsymbol{\beta}_0\|^2 &= \textstyle\sum_{l=1}^{L} \sum_{i\in\mathcal{G}_l} (\widehat{\rho}_l^{or} - \rho_{0l})^2 \leq \mathcal{G}_{\max} \sum_{l=1}^{L} (\widehat{\rho}_l^{or} - \rho_{0l})^2 \\ &= o(\max\{n\mathcal{G}_{\max}/\mathcal{G}_{\min}^2, n^{8\varsigma}\mathcal{G}_{\max}/\mathcal{G}_{\min}^2\}) \ a.s. \end{aligned}$$

and

$$\sup_i \|\widehat{\beta}_i^{or} - \beta_{0i}\| = \sup_l \|\widehat{\rho}_l^{or} - \rho_{0l}\| \leq \|\widehat{\boldsymbol{\rho}}^{or} - \boldsymbol{\rho}_0\| = o(\max\{n^{1/2}/\mathcal{G}_{\min}, n^{4\varsigma}/\mathcal{G}_{\min}\}) \ a.s.$$

Proof of Theorem 1 (ii)

It follows from Theorem 1 (i) and equation (A.8) that

$$\begin{aligned}(\widehat{\phi}^{or} - \phi_0) &= -n^{1/2} V_n^{-1} \widetilde{\Psi}_n(\phi_0) + o(n^{4\varsigma}/\mathcal{G}_{\min}) \\ &= \sum_{i=1}^{n} V_n^{-1} B_i(\phi_0) + o(n^{4\varsigma}/\mathcal{G}_{\min}),\end{aligned} \tag{A.9}$$

where $B_i(\phi_0) = \int I(\zeta_i(\phi_0) \geq u)(U_i - \mathbb{D}_{\phi_0}(u)) W_F(u) d\mathcal{M}(u, \epsilon_i(\phi_0) \mid F)$. Next we verify the Lindeberg–Feller condition. Note that

$$\begin{aligned}E\|V_n^{-1} B_i(\phi_0)\|^4 &= E\{B_i(\phi_0)^\top V_n^{-1} V_n^{-1} B_{ni}(\phi_0)\}^2 \\ &\leq \|V_n^{-1}\|^4 E\{B_i(\phi_0)^\top B_i(\phi_0)\}^2 = O(1/\mathcal{G}_{\min}^4), \\ P(\|V_n^{-1} B_i(\phi_0)\| > \varepsilon) &\leq \|V_n^{-1}\|^2 E\|B_i(\phi_0)\|^2/\varepsilon^2 = O(1/(\mathcal{G}_{\min}^2 \varepsilon)).\end{aligned}$$

Therefore, under the condition $v_n \to 0$, we have

$$\begin{aligned}& \sum_{i=1}^{n} E\|V_n^{-1} B_i(\phi_0)\|^2 I(\|V_n^{-1} B_i(\phi_0)\| > \varepsilon) \\ =\ & nE\|V_n^{-1} B_1(\phi_0)\|^2 I(\|V_n^{-1} B_1(\phi_0)\| > \varepsilon) \\ \leq\ & n\{E\|V_n^{-1} B_i(\phi_0)\|^4\}^{1/2} \{P(\|V_n^{-1} B_i(\phi_0)\| > \varepsilon)\}^{1/2} \\ =\ & O(n/\mathcal{G}_{\min}^3) \to 0.\end{aligned}$$

By noting that $\sum_{i=1}^n \mathrm{var}\{V_n^{-1} B_i(\phi_0)\} = E(V_n^{-1} \Sigma_n V_n^{-1})$, where

$$\Sigma_n = \sum_{i=1}^{n} \int I(\zeta_i(\phi_0) \geq u)(U_i - \mathbb{D}_{\phi_0}(u))(U_i - \mathbb{D}_{\phi_0}(u))^\top W_F^2(u) dF(u),$$

and applying the Lindeberg–Feller central limit theorem (van der Vaart 1998), we have

$$G_n \mathcal{V}_n^{-1/2} (\widehat{\phi}^{or} - \phi_0) \to \mathcal{N}(0, 1).$$

Proof of Theorem 2

Define

$$\begin{aligned}
\ell(\eta,\boldsymbol{\beta}) &= \frac{1}{2}\|\widetilde{\boldsymbol{Y}}(\boldsymbol{\theta},\widetilde{F}_{\boldsymbol{\theta}})-\boldsymbol{Z}\eta-\boldsymbol{X}\boldsymbol{\beta}\|^2-\frac{n}{2}\{\bar{Y}(\boldsymbol{\theta},\widetilde{F}_{\boldsymbol{\theta}})-\bar{Z}^\top\eta-\bar{\boldsymbol{X}}^\top\boldsymbol{\beta}\}^2,\\
P_\lambda(\boldsymbol{\beta}) &= \sum_{1\le i<j\le n}\lambda\varrho_\lambda(\|\beta_i-\beta_j\|),\\
\ell^{\mathcal{G}}(\eta,\boldsymbol{\rho}) &= \frac{1}{2}\|\widetilde{\boldsymbol{Y}}(\boldsymbol{\phi},\widetilde{F}_{\boldsymbol{\phi}})-\boldsymbol{Z}\eta-\boldsymbol{X}\Pi\boldsymbol{\rho}\|^2-\frac{n}{2}\{\bar{Y}(\boldsymbol{\phi},\widetilde{F}_{\boldsymbol{\phi}})-\bar{U}^\top\boldsymbol{\phi}\}^2,\\
P_\lambda^{\mathcal{G}}(\boldsymbol{\rho}) &= \sum_{1\le r<r'\le R}\lambda|\mathcal{G}_r||\mathcal{G}_{r'}|\varrho_\lambda(\|\rho_r-\rho_{r'}\|),
\end{aligned}$$

and let $\ell_P(\eta,\boldsymbol{\beta})=\ell(\eta,\boldsymbol{\beta})+P_\lambda(\boldsymbol{\beta})$, and $\ell_P^{\mathcal{G}}(\eta,\boldsymbol{\rho})=\ell^{\mathcal{G}}(\eta,\boldsymbol{\rho})+P_\lambda^{\mathcal{G}}(\boldsymbol{\rho})$. Let $H: M_{\mathcal{G}}\to\mathcal{R}^{Rp}$ be the mapping that $H(\boldsymbol{\beta})$ is the $Rp\times 1$ vector consisting of R vectors with dimension p and its rth vector component equals the common value of β_i for $i\in\mathcal{G}_r$. Let $H^*:\mathcal{R}^{np}\to\mathcal{R}^{Rp}$ be the mapping that $H^*(\boldsymbol{\beta})=\{|\mathcal{G}_r|^{-1}\sum_{i\in\mathcal{G}_r}\beta_i^\top, r=1,\ldots,R\}^\top$.

Consider the neighborhood of $(\eta_0,\boldsymbol{\beta}_0)$:

$$\Theta=\{\eta\in\mathcal{R}^q,\boldsymbol{\beta}\in\mathcal{R}^{np}:\|\eta-\eta_0\|\le cv_n,\ \sup_i\|\beta_i-\beta_{0i}\|\le cv_n\},$$

where $v_n=\max\{n^{1/2}/\mathcal{G}_{\min}, n^{4\varsigma}/\mathcal{G}_{\min}\}$. We show that $(\widehat{\eta}^{or\top},\widehat{\boldsymbol{\beta}}^{or\top})^\top$ is a strictly local minimizer of the proposed penalized objective function almost surely through the following two steps:

(i) In event A_1, where $A_1=\{\|\widehat{\eta}^{or}-\eta_0\|\le cv_n,\ \sup_i\|\widehat{\beta}_i^{or}-\beta_{0i}\|\le cv_n\}$, $\ell_P(\eta,\boldsymbol{\beta}^*)>\ell_P(\widehat{\eta}^{or},\widehat{\boldsymbol{\beta}}^{or})$ for any $(\eta^\top,\boldsymbol{\beta}^{*\top})^\top\in\Theta$ and $(\eta^\top,\boldsymbol{\beta}^{*\top})^\top\neq(\widehat{\eta}^{or\top},\widehat{\boldsymbol{\beta}}^{or\top})^\top$, where $\boldsymbol{\beta}^*=H^{-1}(H^*(\boldsymbol{\beta}))$.

(ii) There is an event A_2 such that $P(A_2^C)\le\frac{2}{n}$ and in $A_1\cap A_2$, there is a neighborhood Θ_n of $(\widehat{\eta}^{or\top},\widehat{\boldsymbol{\beta}}^{or\top})^\top$, and for $(\eta^\top,\boldsymbol{\beta}^\top)^\top\in\Theta_n\cap\Theta$, $\ell_P(\eta,\boldsymbol{\beta})>\ell_P(\eta,\boldsymbol{\beta}^*)$.

It is easy to show (i) following Ma and Huang (2016). To show the result in (ii), we consider $\Theta_n=\{\beta_i:\sup_i\|\beta_i-\widehat{\beta}_i^{or}\|\le s_n\}$ for a positive sequence s_n. For $(\eta^\top,\boldsymbol{\beta}^\top)^\top\in\Theta_n\cap\Theta$, by Taylor's expansion, we have

$$\ell_P(\eta,\boldsymbol{\beta})-\ell_P(\eta,\boldsymbol{\beta}^*)=\mathcal{H}_1+\mathcal{H}_2,$$

where

$$\mathcal{H}_1=\mathbb{S}(\widetilde{\boldsymbol{\theta}},\widetilde{F}_{\widetilde{\boldsymbol{\theta}}})^\top\mathbb{X}(\boldsymbol{\beta}-\boldsymbol{\beta}^*),\quad\text{and}\quad\mathcal{H}_2=\sum_{i=1}^n\frac{\partial P_\lambda(\widetilde{\boldsymbol{\beta}})}{\partial\beta_i^\top}(\beta_i-\beta_i^*).$$

Here, $\mathbb{S}(\widetilde{\boldsymbol{\theta}},\widetilde{F}_{\widetilde{\boldsymbol{\theta}}})$ is an n-vector with the ith component equal to $\mathcal{S}(\zeta_i(\widetilde{\theta}_i),\epsilon_i(\widetilde{\theta}_i)\mid\widetilde{F}_{\widetilde{\boldsymbol{\theta}}})$, $\widetilde{\boldsymbol{\beta}}=a\boldsymbol{\beta}+(1-a)\boldsymbol{\beta}^*$, $\widetilde{\boldsymbol{\theta}}=a\boldsymbol{\theta}+(1-a)\boldsymbol{\theta}^*$, and $\boldsymbol{\theta}^*=(\eta^\top,\boldsymbol{\beta}^{*\top})^\top$.

Note that

$$\mathcal{H}_2 \geq \sum_{r=1}^{R} \sum_{i,j\in\mathcal{G}_r, i<j} \lambda \varrho_\lambda'(4s_n)\|\beta_i - \beta_j\|.$$

Setting $\boldsymbol{Q} = (Q_1^\top, \dots, Q_n^\top)^\top = \{\mathbb{S}(\widetilde{\boldsymbol{\theta}}, \widetilde{F}_{\widetilde{\boldsymbol{\theta}}})^\top \widetilde{\mathbb{X}}\}^\top$, we have

$$\begin{aligned}
Q_i &= X_i\{\mathcal{S}(\zeta_i(\widetilde{\theta}_i), \epsilon_i(\widetilde{\theta}_i) \mid \widetilde{F}_{\widetilde{\boldsymbol{\theta}}}) - \frac{1}{n}\sum_{j=1}^{n} \mathcal{S}(\zeta_j(\widetilde{\theta}_j), \epsilon_j(\widetilde{\theta}_j) \mid \widetilde{F}_{\widetilde{\boldsymbol{\theta}}})\},\\
\mathcal{H}_1 &= -\sum_{l=1}^{L} \sum_{i,j\in\mathcal{G}_l, i<j} \frac{(Q_j - Q_i)^\top(\beta_j - \beta_i)}{|\mathcal{G}_l|},\\
\sup_i \|Q_i\| &\leq \mathcal{P}_1 + \mathcal{P}_2 + \mathcal{P}_3,
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{P}_1 &= \sup_i \|X_i\| \sup_i \big\{|\mathcal{S}(\zeta_i(\widetilde{\theta}_i), \epsilon_i(\widetilde{\theta}_i) \mid F_{\widetilde{\boldsymbol{\theta}}}) - E\epsilon_i(\widetilde{\theta}_i)|\big\},\\
\mathcal{P}_2 &= \sup_i \|X_i\| \big\{|\frac{1}{n}\sum_{j=1}^{n} \mathcal{S}(\zeta_j(\widetilde{\theta}_j), \epsilon_j(\widetilde{\theta}_j) \mid F_{\widetilde{\boldsymbol{\theta}}}) - E\epsilon_j(\widetilde{\theta}_j)|\big\},\\
\mathcal{P}_3 &= 2\sup_i \|X_i\| \big\{\sup_t \big|W_{\widetilde{F}_{\widetilde{\boldsymbol{\theta}}}}(t) - W_{F_{\widetilde{\boldsymbol{\theta}}}}(t)\big|\big\}.
\end{aligned}$$

For $\mathcal{P}_1$, since

$$\begin{aligned}
& P\left(\sup_i \big|\mathcal{S}(\zeta_i(\widetilde{\theta}_i), \epsilon_i(\widetilde{\theta}_i) \mid F_{\widetilde{\boldsymbol{\theta}}}) - E\epsilon_i(\widetilde{\theta}_i)\big| > \sqrt{2\log(n)/c_1}\right)\\
\leq & \sum_{i=1}^{n} P\left(\big|\mathcal{S}(\zeta_i(\widetilde{\theta}_i), \epsilon_i(\widetilde{\theta}_i) \mid F_{\widetilde{\boldsymbol{\theta}}}) - E\epsilon_i(\widetilde{\theta}_i)\big| > \sqrt{2\log(n)/c_1}\right)\\
\leq & \frac{2}{n},
\end{aligned}$$

we conclude that there is an event A_2 such that $P(A_2^C) \leq \frac{2}{n}$, and under the event A_2 and conditions (C3) (i),

$$\mathcal{P}_1 \leq c_2(\sqrt{2\log(n)/c_1}), \quad \mathcal{P}_2 \leq \mathcal{P}_1.$$

By Lemma 1 (i),

$$\mathcal{P}_3 \leq 2c_2(cn^{-1/2+4\varsigma}).$$

Thus, we have

$$\begin{aligned}
\Big|\frac{(Q_j - Q_i)^\top(\beta_j - \beta_i)}{|\mathcal{G}_l|}\Big| &\leq 2\mathcal{G}_{\min}^{-1} \sup_i \|Q_i\|\|\beta_j - \beta_i\|\\
&\leq 4c_2\mathcal{G}_{\min}^{-1}[\sqrt{2\log(n)/c_1} + cn^{-1/2+4\varsigma}]\|\beta_j - \beta_i\|, \qquad \text{(A.10)}
\end{aligned}$$

and

$$\ell_P(\boldsymbol{\beta}) - \ell_P(\boldsymbol{\beta}^*) \geq \sum_{r=1}^{R} \sum_{i,j\in\mathcal{G}_r, i<j} \{\lambda\varrho'_\lambda(4s_n) - 4c_2\mathcal{G}_{\min}^{-1}[\sqrt{2\log(n)/c_1} + cn^{-1/2+4\varsigma}]\}\|\beta_i - \beta_j\|.$$

Let $s_n \to 0$, and then $\lambda\varrho'_\lambda(4s_n) \to c\lambda$. Since $\lambda \gg \max(\sqrt{\log(n)}/\mathcal{G}_{\min}, n^{-1/2+4\varsigma}/\mathcal{G}_{\min})$, we have $\ell_P(\boldsymbol{\beta}) - \ell_P(\boldsymbol{\beta}^*) \geq 0$ for a sufficiently large n, which completes the proof of Theorem 2.

Proof of Theorem 3

Following the similar arguments used in the proof of Theorem 1, we can conclude the results of Theorem 3 (i) and (ii) by letting $\boldsymbol{X}\Pi = \boldsymbol{x}$ and $\mathcal{G}_{\min} = \mathcal{G}_{\max} = n$. Here we give a simplified proof similar to that of Theorem 2.

Define $\mathbb{M} = \{\boldsymbol{\beta} \in \mathcal{R}^{np} : \beta_1 = \cdots = \beta_n\}$. Note that $\beta_i = \rho$ for all i. Let $\mathbb{H} : \mathbb{M} \to \mathcal{R}^p$ be the mapping that $\mathbb{H}(\boldsymbol{\beta})$ is the p-vector equal to ρ. Let $\mathbb{H}^* : \mathcal{R}^{np} \to \mathcal{R}^p$ be the mapping that $\mathbb{H}(\boldsymbol{\beta}) = \{n^{-1}\sum_{i=1}^n \beta_i\}$. Clearly, when $\boldsymbol{\beta} \in \mathbb{H}$, $\mathbb{H}(\boldsymbol{\beta}) = \mathbb{H}^*(\boldsymbol{\beta})$. Define the neighborhood of $\boldsymbol{\beta}_0$:

$$\Theta' = \{\boldsymbol{\beta} \in \mathcal{R}^{np} : \sup_i \|\beta_i - \beta_{0i}\| \leq cv'_n\},$$

where $v'_n = \max(n^{-1/2}, n^{4\varsigma-1})$. We show that $\widehat{\boldsymbol{\beta}}^{or}$ is a strictly local minimizer of the proposed penalized objective function with probability approaching 1 through the following two steps.

(i) In the event A'_1, where $A'_1 = \{\sup_i \|\widehat{\beta}_i^{or} - \beta_{0i}\| \leq cv'_n\}$, $\ell_P(\boldsymbol{\beta}^*) > \ell_P(\widehat{\boldsymbol{\beta}}^{or})$ for any $\boldsymbol{\beta}^* \in \Theta$ and $\boldsymbol{\beta}^* \neq \widehat{\boldsymbol{\beta}}^{or}$, where $\boldsymbol{\beta}^* = \mathbb{H}^{-1}(\mathbb{H}^*(\boldsymbol{\beta}))$.

(ii) There is an event A'_2 such that $P(A_2^{'C}) \leq \frac{2}{n}$ and in $A'_1 \cap A'_2$, there is a neighborhood Θ'_n of $\widehat{\boldsymbol{\beta}}^{or}$, and for any $\boldsymbol{\beta} \in \Theta'_n \cap \Theta'$, we have $\ell_P(\boldsymbol{\beta}) > \ell_P(\boldsymbol{\beta}^*)$.

Using the idea of Ma and Huang (2016), we can obtain (i). Next we show (ii). For a positive sequence s_n, $\Theta'_n = \{\beta_i : \sup_i \|\beta_i - \widehat{\beta}_i^{or}\| \leq s_n\}$. For $(\eta^\top, \boldsymbol{\beta}^\top)^\top \in \Theta'_n \cap \Theta'$, by Taylor's expansion, we have

$$\ell_P(\boldsymbol{\beta}) - \ell_P(\boldsymbol{\beta}^*) = \mathcal{H}'_1 + \mathcal{H}'_2,$$

where

$$\begin{aligned}
\mathcal{H}'_1 &= \mathbb{S}(\widetilde{\boldsymbol{\theta}}, \widetilde{F}_{\widetilde{\boldsymbol{\theta}}})^\top \widetilde{\mathbb{X}}(\boldsymbol{\beta} - \boldsymbol{\beta}^*), \\
\mathcal{H}'_2 &= \sum_{i=1}^{n} \frac{\partial P_\lambda(\widetilde{\boldsymbol{\beta}})}{\partial \beta_i^\top}(\beta_i - \beta_i^*),
\end{aligned}$$

with $\widetilde{\boldsymbol{\beta}} = a\boldsymbol{\beta} + (1-a)\boldsymbol{\beta}^*$, $\widetilde{\boldsymbol{\theta}} = a\boldsymbol{\theta} + (1-a)\boldsymbol{\theta}^*$, and $\boldsymbol{\theta}^* = (\eta^\top, \boldsymbol{\beta}^{*\top})^\top$.

Note that

$$\begin{aligned} \mathcal{H}_2' &\geq \sum_{i<j} \lambda \varrho_\lambda'(4s_n)\|\beta_i - \beta_j\|, \\ \mathcal{H}_1' &= -n^{-1}\sum_{i<j}(Q_j - Q_i)^\top(\beta_j - \beta_i). \end{aligned}$$

Following the similar proof of (A.10), under event A_2' such that $P(A_2'^C) \leq \frac{2}{n}$, we have

$$\begin{aligned} n^{-1}\Big|(Q_j - Q_i)^\top(\beta_j - \beta_i)\Big| &\leq n^{-1}2\sup_i \|Q_i\|\|\beta_j - \beta_i\| \\ &\leq 4c_2 n^{-1}[\sqrt{2\log(n)/c_1} + cn^{-1/2+4\varsigma}]\|\beta_j - \beta_i\|. \end{aligned}$$

Then,

$$\ell_P(\boldsymbol{\beta}) - \ell_P(\boldsymbol{\beta}^*) \geq \sum_{i<j}\{\lambda\varrho_\lambda'(4s_n) - 4c_2 n^{-1}[\sqrt{2\log(n)/c_1} + cn^{-1/2+4\varsigma}]\}\|\beta_i - \beta_j\|.$$

Let $s_n \to 0$, and then $\lambda\varrho_\lambda'(4s_n) \to c\lambda$. Since $\lambda \gg \max(\sqrt{\log(n)}/n, n^{-3/2+4\varsigma})$, we have $\ell_P(\boldsymbol{\beta}) \geq \ell_P(\boldsymbol{\beta}^*)$ for a sufficiently large n, and thus this completes the proof of Theorem 3.

References

Lai, T. L. and Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 31–546.

Ma, S. and Huang, J. (2016). Estimating subgroup-specific treatment effects via concave fusion. *arXiv preprint arXiv:1607.03717.*

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303–328.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization1. *J. Optim. Theory Appl.* **109**, 475–494.

van der vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, New York.