Dantzig-type penalization for multiple quantile regression with high dimensional covariates

Seyoung Park, Xuming He, and Shuheng Zhou

University of Michigan

Section S1 contains the technical proofs for the main results in the paper. Section S2 includes additional simulation results.

S1 Technical Details

S1.1 Notations

Let F_i be the conditional distribution of y_i given x_i for $i = 1, \ldots, n$, that is, $F_i(x) = \mathbb{P}\left[y_i \leq x \mid x_i\right]$ for all $x \in R$. Define the diagonal matrices

$$H_k = \text{diag}\left[f_1\left({x_1}^T\beta(\tau_k)\right), \cdots, f_n\left({x_n}^T\beta(\tau_k)\right)\right] \quad (k = 1, \ldots, K_n),$$

where $f_1, \cdots, f_n$ are defined in Condition 1 of the main paper. Then, for any vector $\delta \in \mathbb{R}^p$, define an intrinsic norm as in Belloni and Chernozhukov (2011),

$$\|\delta\|_{k,2} = \sqrt{\delta^T \frac{X^T H_k X}{n} \delta} \quad (k = 1, \ldots, K_n). \tag{S1.1}$$

For any positive constant c and the sets $T^{(k)}$ $(k = 1, \dots, K_n)$, defined in (2.2) in the main paper, let

$$A^{(k)}(c) = \left\{ \delta : \delta \neq 0, \delta \in \mathbb{R}^p, \ \|\delta_{\{T^{(k)}\}^c}\|_1 \leq c \|\delta_{T^{(k)}}\|_1 \right\}.$$

Define the function as follows: for $k = 1, \dots, K_n$,

$$\mathbb{Q}_n^{(k)}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - {x_i}^T \beta),$$

where the subdifferential of $\mathbb{Q}_n^{(k)}(\beta)$ at β is the following set of vectors (Wang et al. (2012)):

$$\partial \mathbb{Q}_n^{(k)}(\beta) = \left\{ \delta \in \mathbb{R}^p \mid \delta_j = -\frac{\tau}{n} \sum_i x_{ij} I(y_i > {x_i}^T \beta) + \frac{1-\tau}{n} \sum_i x_{ij} I(y_i < {x_i}^T \beta) - \frac{1}{n} \sum_i x_{ij} v_i \right\}.$$

Here x_{ij} is the jth component of x_i, and $v_i = 0$ if $y_i \neq {x_i}^T \beta$ and $v_i \in [\tau - 1, \tau]$ otherwise. For any $\mathcal{B} = [\beta^{(1)}, \cdots, \beta^{(K_n)}] \in \mathbb{R}^{p \times K_n}$, let

$$G(\mathcal{B}) = \sum_{k=1}^{K_n} \sum_{j=1}^{p} w_j^{(k)} |\beta_j^{(k)}| + \lambda \sum_{k=2}^{K_n} \frac{1}{|\tau_k - \tau_{k-1}|} \sum_{j=1}^{p} v_j^{(k)} |\beta_j^{(k)} - \beta_j^{(k-1)}|, \quad \text{(S1.2)}$$

which is the objective function of our optimization problem, as defined in (2.3). For any square matrix A, let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ be the maximum eigenvalue and the minimum eigenvalue of A, respectively.

S1.2 Preliminary Results

The following Lemma 1 controls the empirical error over all vectors in $A^{(k)}(c_0)$ for all $k = 1, \dots, K_n$ and is analogous to Lemma 5 of the Belloni and Chernozhukov (2011).

Lemma 1. *Let c_0 and $t_1, \cdots, t_{K_n}$ be positive numbers. Suppose Condition 1 and RE($2s_0, c_0$) hold. Let*

$$\tilde{\mathbb{Q}}^{(k)}(v) = E\left[\mathbb{Q}_n^{(k)}\{\beta(\tau_k)+v\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right] - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)+v\} + \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}$$

for any $v \in \mathbb{R}^p$. Then, we have

$$\mathbb{P}\left\{\sup_{v \in A^{(k)}(c_0), \|v\|_{k,2} \leq t_k} \left|\tilde{\mathbb{Q}}^{(k)}(v)\right| > C_1 \frac{1+c_0}{k(s_0,c_0)} t_k \sqrt{\frac{s_0 \log p}{n}} \ (k = 1, \ldots, K_n)\right\} \leq \frac{1}{n} \tag{S1.3}$$

for some absolute constant $C_1 > 0$.

S1.3 Proof of Theorem 1

We begin by providing several lemmas that will be used in the theoretical analysis.

Lemma 2. *Let c_0 be a positive number. Suppose* $\mathrm{RE}(2s_0, c_0)$ *holds. Then, we have for all $k = 1, \ldots, K_n$,*

$$\|\delta\|_1 \leq \sqrt{s_0} \frac{1+c_0}{\sqrt{\underline{f}} k(s_0,c_0)} \|\delta\|_{k,2}, \quad \|\delta\|_2 \leq \frac{1+c_0}{\sqrt{\underline{f}} k(2s_0,c_0)} \|\delta\|_{k,2}$$

for all $\delta \in A^{(k)}(c_0)$.

The following Lemma 3 is a fixed design version of (3.7) in Belloni and Chernozhukov (2011). Lemma 3 provides the lower bound of the difference of the expected values of quantile loss function over all vectors in the cone $A^{(k)}(c_0)$ for all $k = 1, \ldots, K_n$.

Lemma 3. *Let c_0 be a positive number. Suppose Condition 1 and RNI($2s_0, c_0$) hold. Then, we have for all $k = 1, \ldots, K_n$,*

$$E\left[\mathbb{Q}_n^{(k)}\{\beta(\tau_k)+\delta\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right] \geq \frac{3\underline{f}^{3/2}q(2s_0,c_0)}{8\bar{f}}\|\delta\|_{k,2} \wedge \frac{1}{4}\|\delta\|_{k,2}^2 \tag{S1.4}$$

for all $\delta \in A^{(k)}(c_0)$.

The following Lemma 4 shows that $\hat{\beta}^{(k)} - \beta(\tau_k)$ is included in the specific cone for all k.

Lemma 4. *Let η be any positive number. Let $[\hat{\beta}^{(1)}, \cdots, \hat{\beta}^{(K_n)}]$ be an optimum of (2.3) and (2.4) in the main paper. Suppose Condition 2 holds. Then, on event $\mathbb{E}_\eta$, defined in (3.3) in the main paper, we have*

$$\hat{\beta}^{(k)} - \beta(\tau_k) \in A^{(k)}\left(\frac{d_{\min}W_1 + 2\lambda(W_0 \vee W_1)}{d_{\min}W_2 - 2\lambda(W_0 \vee W_1)}\right) \quad (k = 1, \ldots, K_n),$$

where W_0, W_1, and W_2 are defined in Table 1 in the main paper.

Let c_0 be the parameter defined in Table 1 in the main paper. Let η_n be the sequence of numbers which satisfy the conditions in Theorem 1. Let $\delta_k = \hat{\beta}^{(k)} - \beta(\tau_k)$ $(k = 1, \ldots, K_n)$. Let E_2 be the event

$$\sup_{v \in A^{(k)}(c_0), \|v\|_{k,2} \leq \|\delta_k\|_{k,2}} \left|\tilde{\mathbb{Q}}^{(k)}(v)\right| \leq C_1 \frac{1+c_0}{k(s_0,c_0)}\|\delta_k\|_{k,2}\sqrt{\frac{s_0 \log p}{n}} \; (k = 1, \ldots, K_n),$$

where C_1 is the constant in Lemma 1. By Lemma 1, $P(E_2) \geq 1 - 1/n$.

Proof of (3.4) in Theorem 1. Throughout the proof, we assume $E_2 \cap \mathbb{E}_{\eta_n}$ holds. Lemma 4 implies that δ_k is in $A^{(k)}(c_0)$ for $k = 1, \ldots, K_n$. By Lemma

3, it holds that for $k = 1, \ldots, K_n$,

$$\begin{aligned}
& \frac{\|\delta_k\|_{k,2}^2}{4} \wedge \frac{3\underline{f}^{3/2} q(2s_0, c_0)}{8\bar{f}} \|\delta_k\|_{k,2} \\
\leq \; & \mathbb{E}\left[\mathbb{Q}_n^{(k)}\{\hat{\beta}^{(k)}\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right] \\
= \; & \mathbb{Q}_n^{(k)}\{\hat{\beta}^{(k)}\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + \left(\mathbb{E}\left[\mathbb{Q}_n^{(k)}\{\hat{\beta}^{(k)}\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right] - \mathbb{Q}_n^{(k)}\{\hat{\beta}^{(k)}\} + \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right) \\
\leq \; & \eta_n + \left(\mathbb{E}\left[\mathbb{Q}_n^{(k)}\{\hat{\beta}^{(k)}\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right] - \mathbb{Q}_n^{(k)}\{\hat{\beta}^{(k)}\} + \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\}\right) \\
\leq \; & \eta_n + C_1 \frac{1+c_0}{k(s_0, c_0)} \sqrt{\frac{s_0 \log p}{n}} \|\delta_k\|_{k,2}, \qquad \text{(S1.5)}
\end{aligned}$$

where C_1 is the absolute constant stated in Lemma 1.

Notice that (S1.5) implies that the first term in the left hand side must be less than the second term. Suppose otherwise, that is, $\|\delta_k\|_{k,2} \geq 3\underline{f}^{3/2} q(2s_0, c_0)/(2\bar{f})$. Then, we have

$$\frac{3\underline{f}^{3/2} q(2s_0, c_0)}{8\bar{f}} \|\delta_k\|_{k,2} \leq \eta_n + C_1 \frac{1+c_0}{k(s_0, c_0)} \sqrt{\frac{s_0 \log p}{n}} \|\delta_k\|_{k,2},$$

which contradicts the assumption $0 \leq \eta_n < 9\underline{f}^3 q^2(2s_0, c_0)/(32\bar{f}^2)$. Thus, we conclude

$$\frac{\|\delta_k\|_{k,2}^2}{4} \leq \eta_n + C_1 \frac{1+c_0}{k(s_0, c_0)} \sqrt{\frac{s_0 \log p}{n}} \|\delta_k\|_{k,2} \quad (k = 1, \ldots, K_n),$$

which yields

$$\|\delta_k\|_{k,2} \leq 4C_1 \frac{1+c_0}{k(s_0, c_0)} \sqrt{\frac{s_0 \log p}{n}} + 2\sqrt{\eta_n} \quad (k = 1, \ldots, K_n). \qquad \text{(S1.6)}$$

By Lemma 2 and (S1.6), we have

$$\|\delta_k\|_2 \leq 4C_1 \frac{(1+c_0)^2}{k(2s_0, c_0) k(s_0, c_0) \sqrt{\underline{f}}} \sqrt{\frac{s_0 \log p}{n}} + 2 \frac{1+c_0}{k(2s_0, c_0)\sqrt{\underline{f}}} \sqrt{\eta_n} \quad (k = 1, \ldots, K_n),$$

which implies

$$\begin{aligned}\|\hat{\beta}^{(k)}-\beta(\tau_k)\|_2 &\leq \frac{(1+c_0)^2}{k(2s_0,c_0)\sqrt{\underline{f}}}\left\{2+\frac{4C_1}{k(s_0,c_0)}\right\}\sqrt{\frac{s_0\log p}{n}+\eta_n}\\ &= \xi_1\sqrt{\frac{s_0\log p}{n}+\eta_n}, \qquad \text{(S1.7)}\end{aligned}$$

where $\xi_1=\frac{(1+c_0)^2}{k(2s_0,c_0)\sqrt{\underline{f}}}\left\{2+\frac{4C_1}{k(s_0,c_0)}\right\}$. This completes the proof. □

Proof of (3.5) in Theorem 1. Throughout the proof, we assume $E_2\cap\mathbb{E}_{\eta_n}$ holds. The main idea is to compare the objective functions of our optimization problem, as stated in (2.3), at $\hat{\mathcal{B}}$ and $\mathcal{B}^o$. Since $\mathcal{B}^o$ is feasible, $G(\hat{\mathcal{B}})$ must not be greater than $G(\mathcal{B}^o)$, where the function $G(\cdot)$ is defined in (S1.2). Hence, it holds that

$$\begin{aligned}0 &\leq G(\mathcal{B}^o)-G(\hat{\mathcal{B}})\\ &= \sum_{k=1}^{K_n}\sum_{j\in T^{(k)}} w_j^{(k)}|\beta_j(\tau_k)|+\sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j\in B^{(k)}} v_j^{(k)}|\beta_j(\tau_k)-\beta_j(\tau_{k-1})|\\ &- \sum_{k=1}^{K_n}\sum_{j\in T^{(k)}} w_j^{(k)}|\hat{\beta}_j^{(k)}|-\sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j\in B^{(k)}} v_j^{(k)}|\hat{\beta}_j^{(k)}-\hat{\beta}_j^{(k-1)}|-\sum_{k=1}^{K_n}\sum_{j\in\{T^{(k)}\}^c} w_j^{(k)}|\hat{\beta}_j^{(k)}|\\ &- \sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j\in\{B^{(k)}\}^c} v_j^{(k)}|\hat{\beta}_j^{(k)}-\hat{\beta}_j^{(k-1)}|.\end{aligned}$$

By triangle inequality, the above inequality implies

$$\begin{aligned}&\sum_{k=1}^{K_n}\sum_{j\in\{T^{(k)}\}^c} w_j^{(k)}|\hat{\beta}_j^{(k)}|+\sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j\in\{B^{(k)}\}^c} v_j^{(k)}|\hat{\beta}_j^{(k)}-\hat{\beta}_j^{(k-1)}|\\ \leq& \sum_{k=1}^{K_n}\sum_{j\in T^{(k)}} w_j^{(k)}|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)|+\sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j\in B^{(k)}} v_j^{(k)}|\hat{\beta}_j^{(k)}-\hat{\beta}_j^{(k-1)}-\beta_j(\tau_k)+\beta_j(\tau_{k-1})|\\ \leq& W_1\sum_{k=1}^{K_n}\|\{\hat{\beta}^{(k)}-\beta(\tau_k)\}_{T^{(k)}}\|_1+W_1\sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\|\{\hat{\beta}^{(k)}-\beta(\tau_k)\}_{B^{(k)}}\|_1\\ &+W_1\sum_{k=2}^{K_n}\frac{\lambda}{|\tau_k-\tau_{k-1}|}\|\{\hat{\beta}^{(k-1)}-\beta(\tau_{k-1})\}_{B^{(k)}}\|_1.\end{aligned}$$

Now, by the Cauchy-Schwarz inequality with $|T^{(k)}| \le s_0$ and $|B^{(k)}| \le 2s_0$, the above inequality implies

$$\begin{aligned} & \sum_{k=1}^{K_n} \sum_{j \in \{T^{(k)}\}^c} w_j^{(k)} |\hat{\beta}_j^{(k)}| + \sum_{k=2}^{K_n} \frac{\lambda}{|\tau_k - \tau_{k-1}|} \sum_{j \in \{B^{(k)}\}^c} v_j^{(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| \\ \le \;& W_1 \sqrt{K_n} \sqrt{s_0} \sqrt{\sum_{k=1}^{K_n} \|\{\hat{\beta}^{(k)} - \beta(\tau_k)\}\|_2^2} \\ + \;& 2W_1 \frac{\lambda}{\min_{k \ge 2} |\tau_k - \tau_{k-1}|} \sqrt{K_n} \sqrt{2s_0} \sqrt{\sum_{k=1}^{K_n} \|\{\hat{\beta}^{(k)} - \beta(\tau_k)\}\|_2^2} \\ \le \;& \xi_1 (W_1 + \sqrt{2} W_1) \sqrt{s_0} K_n \sqrt{\frac{s_0 \log p}{n} + \eta_n}, \end{aligned} \tag{S1.8}$$

where the last inequality follows from Condition 2. Now, by (S1.8) and the definition of W_2, (3.5) in the main paper holds. This completes the proof. □

S1.4 Proofs of Theorem 2

We begin by providing the following lemmas that will be used for the proof of Theorem 2. Lemma 5 is only used to show Lemma 6.

Lemma 5. *For an $n \times p$ design matrix $X = (x_1, \cdots, x_n)^T$, which is normalized to have column ℓ_2 norm $\sqrt{n}$, we have with probability at least $1 - 1/n$,*

$$\max_k \Big\| \sum_{i=1}^{n} x_i [\tau_k - I\{y_i \le {x_i}^T \beta(\tau_k)\}]/n \Big\|_\infty \le 3 \sqrt{\frac{\log p}{n}}. \tag{S1.9}$$

Recall the event E_1, defined in (5.1) in the matin paper: for all k,

$$\tilde{\lambda} \leq C_2\sqrt{\log p/n}, \quad \|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_2 \leq C_3\sqrt{s_0 \log p/n}, \quad \|\tilde{\beta}^{(k)}\|_0 \leq C_4 s_0. \qquad \text{(S1.10)}$$

The following lemma implies that we can find a proper η_n on event E_1.

Lemma 6. *Suppose the conditions of Theorem 2 hold. Then, we have* $P(\mathbb{E}_{\eta_n^*} \mid E_1) \geq 1 - 1/n$, *where*

$$\eta_n^* = \left(C_2 C_3\sqrt{C_4 + 1} + C_4 \max_k \Lambda_k\right) s_0 \log p/n.$$

Lemma 6 implies

$$\mathbb{P}(\mathbb{E}_{\eta_n^*} \cap E_1) = \mathbb{P}(E_1)\mathbb{P}(\mathbb{E}_{\eta_n^*} \mid E_1) \geq (1 - \mathbb{P}(E_1^c))\,(1 - 1/n) \geq 1 - \frac{1}{n} - \mathbb{P}(E_1^c).$$

Let $\delta_k = \hat{\beta}^{(k)} - \beta(\tau_k)$ $(k = 1, \ldots, K_n)$. On event E_3, we have

$$\sup_{v \in A^{(k)}(\psi_\lambda), \|v\|_{k,2} \leq \|\delta_k\|_{k,2}} \left|\tilde{\mathbb{Q}}^{(k)}(v)\right| \leq C_1 \frac{1 + \psi_\lambda}{k(s_0, \psi_\lambda)} \|\delta_k\|_{k,2} \sqrt{\frac{s_0 \log p}{n}} \; (k = 1, \ldots, K_n),$$

where $\psi_\lambda = (d_{\min} + 2\lambda)/(d_{\min} - 2\lambda)$, as defined in Table 1 in the main paper, and $\mathbb{P}(E_3) \geq 1 - 1/n$ by Lemma 1.

Proof of Theorem 2. Throughout the proof, we assume $\mathbb{E}_{\eta_n^*} \cap E_1 \cap E_3$, where $\mathbb{P}(\mathbb{E}_{\eta_n^*} \cap E_1 \cap E_3) \geq 1 - 2/n - \mathbb{P}(E_1^c)$. To utilize the results of Theorem 1, we will show that the conditions of Theorem 1 hold with $c_0 = \psi_\lambda$ in the current setting. Note that $W_0 \vee W_1 = W_2 = 1$ holds because the maximum absolute value of $P_{\zeta_n}(\cdot)$ is at most 1 and

$$P_{\zeta_n}\left(\tilde{\beta}_j^{(k)}\right) = 1 \; (j \in \{T^{(k)}\}^c), \quad P_{\zeta_n}\left(\tilde{\beta}_j^{(k)} - \tilde{\beta}_j^{(k-1)}\right) = 1 \; (j \in \{B^{(k)}\}^c).$$

These results follow from

$$\begin{aligned} |\tilde{\beta}_j^{(k)}| &\leq \|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_2 \leq C_3\sqrt{\frac{s_0 \log p}{n}} < \zeta_n \quad (j \in \{T^{(k)}\}^c), \\ |\tilde{\beta}_j^{(k)} - \tilde{\beta}_j^{(k-1)}| &\leq \|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_2 + \|\tilde{\beta}^{(k-1)} - \beta(\tau_{k-1})\|_2 \\ &\leq 2C_3\sqrt{\frac{s_0 \log p}{n}} \leq \zeta_n \quad (j \in \{B^{(k)}\}^c), \end{aligned}$$

where Condition 4 is used. Therefore, Condition 2 holds and

$$\frac{d_{\min} W_1 + 2\lambda(W_0 \vee W_1)}{d_{\min} W_2 - 2\lambda(W_0 \vee W_1)} \leq \psi_\lambda.$$

Since the conditions of Theorem 1 hold with $c_0 = \psi_\lambda$ and $\eta = \eta_n^*$, we can utilize the results of Theorem 1 with $\eta = \eta_n^*$ and $c_0 = \psi_\lambda$. Hence, we have

$$\begin{aligned} &\|\hat{\beta}^{(k)} - \beta(\tau_k)\|_2 \\ \leq\ & \frac{4d_{\min}^2}{(d_{\min} - 2\lambda)^2 k(2s_0, \psi_\lambda)\sqrt{\underline{f}}}\sqrt{\frac{s_0 \log p}{n} + \{C_2 C_3 \sqrt{C_4 + 1} + C_4 \max_k \Lambda_k\}\frac{s_0 \log p}{n}} \\ \leq\ & \xi_2 \sqrt{\frac{s_0 \log p}{n}}\ (k = 1, \ldots, K_n), \end{aligned} \tag{S1.11}$$

where

$$\xi_2 = \frac{4d_{\min}^2}{(d_{\min} - 2\lambda)^2 k(2s_0, \psi_\lambda)\sqrt{\underline{f}}}\sqrt{1 + C_2 C_3 \sqrt{C_4 + 1} + C_4 \max_k \Lambda_k}.$$

This completes the proof. □

S1.5 Proofs of Theorem 3

Let $C_5 = \{(3.7\alpha + C_3) \vee \xi_2\}$ and $C_6 = \{(3.7\alpha + 2C_3) \vee 2\xi_2\}/(K_n d_{\min})$, where $\alpha = \zeta_n(s_0 \log p/n)^{-0.5}$. First, we state the following lemma.

Lemma 7. *Suppose the conditions of Theorem 3 hold. Then, on event E_1, we have $W_1 = 0$.*

Proof of Theorem 3. Throughout the proof, we assume $\mathbb{E}_{\eta_n^*} \cap E_1 \cap E_3$. By Lemma 7,

$$G(\mathcal{B}^o) = \sum_{k=1}^{K_n} \sum_{j \in \{T^{(k)}\}^c} w_j^{(k)} |\beta_j(\tau_k)| + \sum_{k=2}^{K_n} \frac{\lambda}{|\tau_k - \tau_{k-1}|} \sum_{j \in \{B^{(k)}\}^c} v_j^{(k)} |\beta_j(\tau_k) - \beta_j(\tau_{k-1})| = 0,$$

where $G(\cdot)$ is the objective function of our optimization problem, as defined in (S1.2).

Now, notice that the proof of Theorem 1 and the result of Theorem 2 demonstrate that (3.5) in the main paper holds with $\eta = \eta_n^*$ and $c_0 = \psi_\lambda$. Then, the equation (3.5) and $W_1 = 0$ imply

$$\hat{\beta}^{(k)}_{\{T^{(k)}\}^c} = 0 \quad (k = 1, \ldots, K_n), \quad \{\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)}\}_{\{B^{(k)}\}^c} = 0 \quad (k = 2, \ldots, K_n). \tag{S1.12}$$

In addition, we have

$$\begin{aligned} \min_k \min_{j \in T^{(k)}} |\hat{\beta}_j^{(k)}| &\geq \min_k \min_{j \in T^{(k)}} |\beta_j(\tau_k)| - \max_k \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_2 \\ &> \xi_2 \sqrt{\frac{s_0 \log p}{n}} - \xi_2 \sqrt{\frac{s_0 \log p}{n}} = 0, \end{aligned} \tag{S1.13}$$

where the second inequality follows from the beta-min condition, as stated in Theorem 3. Similarly,

$$\begin{aligned} \min_{k \geq 2} \min_{j \in B^{(k)}} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| &\geq \min_{k \geq 2} \min_{j \in B^{(k)}} |\beta_j(\tau_k) - \beta_j(\tau_{k-1})| - 2 \max_k \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_2 \\ &> 2\xi_2 \sqrt{\frac{s_0 \log p}{n}} - 2\xi_2 \sqrt{\frac{s_0 \log p}{n}} = 0. \end{aligned} \tag{S1.14}$$

By (S1.12), (S1.13), and (S1.14), $\hat{\mathcal{B}}$ provides the exact model structure, which completes the proof. □

S1.6 Proofs of Theorem 4

Here we define the map T and new design matrix $z_i^{(k)}$, as stated in Section 6. First, we define a map $M : \{1, \ldots, p\} \times \{1, \ldots, K_n\} \to \mathbb{R}^{d_0}$ as follows:

1. If $\hat{\beta}_j^{(k)} = 0$, then $M(j,k) = 0$.

2. if $\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)}$, then $M(j,k) = M(j,k-1)$.

3. If $\hat{\beta}_j^{(k)} \neq 0$, $\hat{\beta}_{j'}^{(k')} = 0$ $(k' = 1, \ldots, K_n;\ j' = 1, \ldots, j-1)$, and $\hat{\beta}_j^{(k')} = 0$ $(k' = 1, \ldots, k-1)$, then $M(j,k) = 1$.

4. If $\hat{\beta}_j^{(k)} \neq 0$ and $\hat{\beta}_j^{(k)} \neq \hat{\beta}_j^{(k-1)}$, then

$$M(j,k) = 1 + \max(M_1, M_2),$$

   where

$$M_1 := \{M(j',k') :\ k' = 1, \ldots, K_n;\ j' = 1, \ldots, j-1\},$$

$$M_2 := \{M(j,k') :\ k' = 1, \ldots, k-1\}.$$

5. If $\hat{\beta}_j^{(1)} \neq 0$ for $j \geq 2$, then

$$M(j,1) = 1 + \max\{M(j',k') :\ k' = 1, \ldots, K_n;\ j' = 1, \ldots, j-1\}.$$

By utilizing the map M, we arrive at a new design matrix denoted by $z_i^{(k)} \in \mathbb{R}^{d_0}$ $(i = 1, \dots, n;\ k = 1, \dots, K_n)$. First, let

$$M(T^{(k)}, k) = \{M(j, k) : j \in T^{(k)}\} \quad (k = 1, \dots, K_n),$$

where the elements in $M(T^{(k)}, k)$ are in ascending order. Let

$$z^{(k)}{}_{i, M(T^{(k)}, k)} = x_{i, T^{(k)}}, \quad z_{i,j}^{(k)} = 0 \text{ for } j \in \{1, \cdots, d_0\} \setminus M(T^{(k)}, k).$$

Now, to define the map T, let

$$\mathrm{IM} = \{(j, k) : M(j, k) \neq 0,\ M(j, k) \neq M(j, k-1)\},$$

which is the location indices account for effective components. Then, for any $\mathcal{B} \in G$, $T(\mathcal{B}) \in \mathbb{R}^{d_0}$, where $T(\mathcal{B})_i = \mathcal{B}_{j,k}$ $(i = 1, \dots, d_0)$ for i satisfying $M(j, k) = i$ and $(j, k) \in \mathrm{IM}$. It is easily checked that for any $\mathcal{B} \in G$, $T(\mathcal{B})$ is a d_0–dimensional vector.

Remark 1. Illustrative example of the M and T.

Suppose that we consider the model, where $p = 5$ and $K_n = 3$ with the three quantile levels τ_1, τ_2, and τ_3. Assume that we obtain the following estimates from our Dantzig-type optimization problem:

$$\hat{B} = \begin{bmatrix} 0.9 & 0.9 & 0.0 \\ 1.1 & 1.5 & 1.5 \\ 0.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 1.0 \\ 0.0 & 0.2 & 0.2 \end{bmatrix} = \left[\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \hat{\beta}^{(3)}\right].$$

Then, M is the function such that

$$M(j,k) = \tilde{M}_{j,k} \quad \text{for} \quad j = 1, \cdots, 5, \quad \text{and} \quad k = 1, 2, 3,$$

where

$$\tilde{M} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 3 \\ 0 & 0 & 0 \\ 4 & 0 & 5 \\ 0 & 6 & 6 \end{bmatrix}$$

is the indices matrix follows from the model structure of $\hat{B}$.

Here $d_0 = 6$, and

$$T(\hat{B}) = [0.9, 1.1, 1.5, 0.5, 1.0, 0.2]^T.$$

Lemma 8. *Assume* $d_0 M_n^4 (\log n)^2 = o(n)$. *Let* $\Delta > 0$ *and* $\Theta = \{\theta \in R^{d_0} :$

$\|\theta\|_2 \leq \Delta\}$. *For any* $\theta \in \Theta$, *let*

$$I_2(\theta) = \frac{1}{\|\theta\|_2} \sum_k \sum_i \int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} I(\epsilon_i^{(k)} \leq x) - I(\epsilon_i^{(k)} \leq 0)dx,$$

where $\epsilon_i^{(k)} = y_i - x_i{}^T\beta(\tau_k) = y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o)$. *Then, with probability at least* $1 - n^{-9d_0 \log n} - 2\bar{f}K_n/n$,

$$\sup_{\theta\in\Theta} |I_2(\theta) - E[I_2(\theta)]| \leq 7n^{-1/4}K_n\Delta^{1/2}s_0^{3/4}(d_0)^{5/4}(\log n)^{3/2}.$$

Proof of Theorem 4. We will show that for any constant $\epsilon > 0$, there exists a sufficiently large constant $\Delta > 0$, satisfying

$$\mathbb{P}\Big[\inf_{\|\theta\|_2=\Delta,\ \theta\in R^{d_0}} L_n\Big(T(\mathcal{B}^o) + \sqrt{\frac{d_0}{n}}\theta\Big) > L_n(T(\mathcal{B}^o))\Big] \geq 1 - \epsilon, \qquad \text{(S1.15)}$$

where $L_n(\theta) = \sum_k \sum_i \rho_{\tau_k}[y_i - \{z_i^{(k)}\}^T\theta]$ for any $\theta \in R^{d_0}$. Since L_n is a strict convex function over $\theta \in R^{d_0}$, (S1.15) implies that the global minimum $T(\hat{\mathcal{B}})$ lies within the ball whose center is $T(\mathcal{B}^o)$ and the radius is $\Delta\sqrt{d_0/n}$, with probability at least $1 - \epsilon$, which proves the theorem. Let

$$G_n(\theta) = L_n\left(T(\mathcal{B}^o) + \sqrt{\frac{d_0}{n}}\theta\right) - L_n\left(T(\mathcal{B}^o)\right).$$

By the Knight's identity,

$$
\begin{aligned}
G_n(\theta) &= \sum_k \sum_i \rho_{\tau_k}\Big[y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o) - \sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta\Big] - \rho_{\tau_k}\Big[y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o)\Big] \\
&= \sqrt{\frac{d_0}{n}} \sum_k \sum_i \{z_i^{(k)}\}^T \theta \{I(\epsilon_i^{(k)} < 0) - \tau_k\} \\
&+ \sum_k \sum_i \int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} I(\epsilon_i^{(k)} \le x) - I(\epsilon_i^{(k)} \le 0) dx \\
&:= I_1(\theta) + I_2(\theta),
\end{aligned}
$$

where $\epsilon_i^{(k)}$ is defined in Lemma 8. First, consider $I_1(\theta)$. Let $v_i^{(k)} = I(\epsilon_i^{(k)} < 0) - \tau_k$ and $\Theta = \{\theta \in R^{d_0} : \|\theta\|_2 = \Delta\}$. Then, we have

$$
\begin{aligned}
\mathbb{E}\big[\sup_{\theta\in\Theta} I_1^2(\theta)\big] &= \frac{d_0}{n}\mathbb{E}\left[\sup_{\|\theta\|_2\in\Theta}\left\{(\sum_k\sum_i z_i^{(k)} v_i^{(k)})^T\theta\right\}^2\right] \\
&= \frac{d_0}{n}\mathbb{E}\big[\sup_{\|\theta\|_2\in\Theta} \theta^T Z Z^T \theta\big] \\
&\le \frac{d_0}{n}\Delta^2 \mathbb{E}[\lambda_{\max}(ZZ^T)], \qquad \text{(S1.16)}
\end{aligned}
$$

where $Z = \sum_k \sum_i \{z_i^{(k)} v_i^{(k)}\}$. We have noticed that ZZ^T is a zero matrix or a rank–one matrix, and that $Z^T Z$ is a eigenvalue of ZZ^T when ZZ^T is a rank–one matrix. Hence,

$$
\lambda_{\max}(ZZ^T) \le Z^T Z. \qquad \text{(S1.17)}
$$

Therefore (S1.16) and (S1.17) imply

$$\begin{aligned}\mathbb{E}\big[\sup_{\theta\in\Theta} I_1^2(\theta)\big] &\leq \frac{d_0}{n}\Delta^2\mathbb{E}[Z^TZ]\\ &= \frac{d_0}{n}\Delta^2\mathbb{E}[\sum_k\sum_{k'}\sum_i v_i^{(k)}v_i^{(k')}\{z_i^{(k)}\}^T z_i^{(k')}]\\ &= \frac{d_0}{n}\Delta^2\sum_k\sum_{k'}(\tau_k\wedge\tau_{k'}-\tau_k\tau_{k'})\sum_i\{z_i^{(k)}\}^T z_i^{(k')}\\ &\leq \Delta^2 K_n^2 d_0^2.\end{aligned}$$

Hence, by Markov inequality,

$$\mathbb{P}\left(\sup_{\|\theta\|_2\in\Theta}|I_1(\theta)|\geq\frac{\Delta K_n d_0}{\sqrt{\epsilon/2}}\right)\leq\frac{\epsilon}{2}.$$

Hence, with probability at least $1-\epsilon/2$, we have $\sup_{\|\theta\|_2\in\Theta}|I_1(\theta)|\leq\frac{\Delta K_n d_0}{\sqrt{\epsilon/2}}$.

Now, consider $I_2(\theta)$. Then, for any $\theta\in\Theta$,

$$\begin{aligned}\mathbb{E}\left(I_2(\theta)\right) &= \sum_k\sum_i\int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta}\mathbb{P}\left(y_i\leq {x_i}^T\beta(\tau_k)+x\right)-\mathbb{P}\left(y_i\leq {x_i}^T\beta(\tau_k)\right)dx\\ &= \sum_k\sum_i\int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} xf_i\left({x_i}^T\beta(\tau_k)\right)+\frac{x^2}{2}f_i'({x_i}^T\beta(\tau_k)+\tilde{x}_i^{(k)})dx\\ &\geq \sum_k\sum_i\frac{f_i\left({x_i}^T\beta(\tau_k)\right)}{2}\frac{d_0}{n}[\{z_i^{(k)}\}^T\theta]^2-\frac{\bar{f}}{6}\sum_k\sum_i\left(\frac{d_0}{n}\right)^{1.5}[\{z_i^{(k)}\}^T\theta]^3\\ &\geq \frac{d_0\underline{f}}{2}\sum_k\sum_i\theta^T\frac{1}{n}z_i^{(k)}\{z_i^{(k)}\}^T\theta-\frac{\Delta M_n\bar{f}d_0^{1.5}\sqrt{s_0}}{6\sqrt{n}}\sum_k\sum_i\theta^T\frac{1}{n}z_i^{(k)}\{z_i^{(k)}\}^T\theta,\end{aligned}$$

where $\tilde{x}_i^{(k)}\in(0,x)$ depends on i and k in the second line. The first and the second inequality follow from Condition 1 and the fact

$$|\{z_i^{(k)}\}^T\theta|\leq\|z_i^{(k)}\|_2\|\theta\|_2\leq M_n\sqrt{s_0}\Delta.$$

Since $d_0s_0=o(n)$ holds and the nonzero parts of $\sum_i z_i^{(k)}\{z_i^{(k)}\}^T/n$ is a

$s_k \times s_k$–dimensional sub–matrix of $\sum_i x_i x_i{}^T/n$, it holds that

$$\mathbb{E}\left(I_2(\theta)\right) \geq \frac{d_0 \underline{f}}{4} \sum_k \sum_i \theta^T \frac{1}{n} z_i^{(k)} \{z_i^{(k)}\}^T \theta \geq \frac{K_n d_0 \underline{f}}{4} k^2(s_0, 0)\Delta^2.$$

By Lemma 8 and Condition 6(a), we have

$$I_2(\theta) \geq \frac{K_n d_0 \underline{f}}{4} k^2(s_0, 0)\Delta^2 - \Delta^{3/2} o_p(K_n d_0),$$

where $o_p(1)$ is uniformly over $\theta \in \Theta$.

Hence, for any $\epsilon > 0$, with probability at least $1 - \epsilon/2$,

$$\inf_{\theta \in \Theta} G_n(\theta) \geq \frac{K_n d_0 \underline{f}}{4} k^2(s_0, 0)\Delta^2 - \Delta^{3/2} o_p(K_n d_0) - \frac{\Delta K_n d_0}{\sqrt{\epsilon/2}} > 0$$

with a sufficiently large Δ, which completes the proof. □

S1.7 Proofs of Theorem 5

Lemma 9. *Recall the matrices A_n and B_n defined in Theorem 5. We have*

$$\bar{f}^{-2}\phi^{-2}(s_0)k^2(s_0, 0)(\min_k \tau_k)(1 - \max_k \tau_k) \leq \lambda_{\min}(A_n{}^{-1} B_n A_n{}^{-1}),$$

$$\lambda_{\max}(A_n{}^{-1} B_n A_n{}^{-1}) \leq L_0^{-2}\phi(s_0)k^{-4}(s_0, 0).$$

Lemma 10. *Assume conditions of Theorem 5 hold. Then, for any sequence of $\alpha_n \in R^{d_0}$ with $\|\alpha_n\|_2 = 1$, the following asymptotic normality holds:*

$$n^{-1/2}\alpha_n^T (A_n{}^{-1} B_n A_n{}^{-1})^{-1/2} A_n{}^{-1} \sum_k \sum_i z_i^{(k)} \left(I(y_i - x_i{}^T \beta(\tau_k) < 0) - \tau_k \right) \to N(0, 1).$$

Proof of Theorem 5. Recall

$$T(\hat{\mathcal{B}}^{po}) = \arg\min_{\beta \in R^{d_0}} \sum_k \sum_i \rho_{\tau_k}(y_i - \{z_i^{(k)}\}^T \beta). \qquad \text{(S1.18)}$$

By $\theta = \sqrt{n/d_0}\,(\beta - T(\mathcal{B}^o))$, $T(\hat{\mathcal{B}}^{po}) = T(\mathcal{B}^o) + \sqrt{d_0/n}\hat{\theta}$, where

$$\hat{\theta} = \arg\min_{\theta \in R^{d_0}} \sum_k \sum_i \rho_{\tau_k}\Big[y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o) - \sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T \theta\Big]. \tag{S1.19}$$

Then, $\hat{\theta}$ can be written as $\hat{\theta} = G_n(\theta)$, where

$$\begin{aligned} G_n(\theta) = \arg\min_{\theta \in R^{d_0}} \quad & \sum_k \sum_i \rho_{\tau_k}\left(y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o) - \sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T \theta\right) \\ - \quad & \sum_k \sum_i \rho_{\tau_k}\left(y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o)\right). \end{aligned}$$

Consider θ over the set $\Theta_n = \{\theta \in R^{d_0} \mid \|\theta\|_2 \le C\}$ with some positive constant C independent of n. Decompose G_n into two terms:

$$G_n(\theta) = I_1(\theta) + I_2(\theta),$$

where

$$I_1(\theta) = \sqrt{\frac{d_0}{n}} \sum_k \sum_i \{z_i^{(k)}\}^T \theta \{I(\epsilon_i^{(k)} < 0) - \tau_k\},$$

$$I_2(\theta) = \sum_k \sum_i \int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} I(\epsilon_i^{(k)} \le x) - I(\epsilon_i^{(k)} \le 0)dx.$$

Consider the term $I_2(\theta)$. From the proof of Theorem 4,

$$\begin{aligned} & \Big|\mathbb{E}[I_2(\theta)] - \sum_k \sum_i \int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} f_i(x_i{}^T \beta(\tau_k)) x dx\Big| \\ \le & \ \Big|\sum_k \sum_i \int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} \frac{x^2}{2} f_i'(\tilde{x}_i^{(k)}) dx\Big| \\ \le & \ \frac{\bar{f}}{6} \sum_k \sum_i \big(\frac{d_0}{n}\big)^{1.5} |\{z_i^{(k)}\}^T \theta|^3 \\ \le & \ \frac{\bar{f}}{6} K_n \frac{d_0^{1.5}\sqrt{s_0} M_n}{\sqrt{n}} \|\theta\|_2^3 \phi(s_0) \\ = & \ o(\|\theta\|_2 K_n), \end{aligned}$$

where $\tilde{x}_i^{(k)} \in ({x_i}^T\beta(\tau_k), {x_i}^T\beta(\tau_k)+x)$ depends on i and k in the first inequality. The second inequality follows from Condition 1. The third inequality holds due to Sparse(s_0) and the fact that

$$|\{z_i^{(k)}\}^T\theta| \le \|z_i^{(k)}\|_2\|\theta\|_2 \le M_n\|\theta\|_2\sqrt{s_0}.$$

The last small o results follows from $M_n^2 d_0^3 s_0 = o(n)$. Moreover, Lemma 8 and the conditions of Theorem 5 imply

$$I_2(\theta) - \mathbb{E}[I_2(\theta)] = o_p(\|\theta\|_2 K_n),$$

where o_p is uniform over $\theta \in \Theta_n$. Hence, for all $\theta \in \Theta_n$,

$$I_2(\theta) = \sum_k \sum_i \frac{f_i({x_i}^T\beta(\tau_k))}{2}\frac{d_0}{n}[\{z_i^{(k)}\}^T\theta]^2 + o_p(\|\theta\|_2 K_n).$$

Thus, for all $\theta \in \Theta_n$, $G_n(\theta)$ can be written as

$$\begin{aligned} G_n(\theta) &= \sqrt{\frac{d_0}{n}}\sum_k\sum_i \{z_i^{(k)}\}^T\theta\left(I(\epsilon_i^{(k)} < 0) - \tau_k\right) \\ &+ \sum_k\sum_i \frac{f_i({x_i}^T\beta(\tau_k))}{2}\frac{d_0}{n}[\{z_i^{(k)}\}^T\theta]^2 + o_p(\|\theta\|_2 K_n). \end{aligned}$$

By matrix calculus,

$$\begin{aligned} \hat{\theta} &= \sqrt{\frac{n}{d_0}}\Big\{\sum_k\sum_i f_i({x_i}^T\beta(\tau_k))z_i^{(k)}\{z_i^{(k)}\}^T\Big\}^{-1}\sum_k\sum_i z_i^{(k)}\{I(\epsilon_i^{(k)} < 0) - \tau_k\} \\ &+ \left(\sum_k\sum_i \frac{f_i({x_i}^T\beta(\tau_k))}{2}\frac{d_0}{n}z_i^{(k)}\{z_i^{(k)}\}^T\right)^{-1} K_n o_p(1) \\ &= (nd_0)^{-0.5}{A_n}^{-1}\sum_k\sum_i z_i^{(k)}\{I(\epsilon_i^{(k)} < 0) - \tau_k\} + 2A_n^{-1}\frac{K_n}{d_0}o_p(1) \\ &= d_0^{-0.5}({A_n}^{-1}B_n{A_n}^{-1})^{\frac{1}{2}}\Big[n^{-0.5}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}{A_n}^{-1}\sum_k\sum_i z_i^{(k)}\{I(\epsilon_i^{(k)} < 0) - \tau_k\}\Big] \\ &+ \frac{1}{d_0}o_p(1), \end{aligned}$$

where $o_p(1)$ represents any d_0-dimensional vector whose ℓ_2 norm is $o_p(1)$.

For any $\alpha_n \in R^{d_0}$ with $\|\alpha_n\|_2 = 1$, Lemma 10 implies

$$\alpha_n^T\Big[n^{-0.5}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}{A_n}^{-1}\sum_k\sum_i z_i^{(k)}\{I(\epsilon_i^{(k)} < 0) - \tau_k\}\Big] \to N(0,1).$$

Hence, by Lemma 9,

$$\begin{aligned}\|\hat{\theta}\|_2 &\leq d_0^{-0.5}\lambda_{\max}\{({A_n}^{-1}B_n{A_n}^{-1})^{\frac{1}{2}}\}O_p\{\sqrt{d_0}\} + o_p(1)\\ &\leq L_0^{-1}\sqrt{\phi(s_0)}k^{-2}(s_0,0)O_p(1).\end{aligned}$$

Since C can be chosen to be much larger than $L_0^{-1}\sqrt{\phi(s_0)}k^{-2}(s_0,0)$, $\hat{\theta}$ is included in Θ_n. Hence, by Lemma 10 ,

$$\alpha_n^T\sqrt{n}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}\sqrt{\frac{d_0}{n}}\hat{\theta} \to N(0,1).$$

Thus,

$$\alpha_n^T\sqrt{n}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}\{T(\hat{\mathcal{B}}^{po}) - T(\mathcal{B}^o)\} \to N(0,1),$$

which completes the proof. □

S1.8 Proof of Lemmas

Proof of Lemma 1,2 and 3. The proofs essentially follow from the proofs of Lemmas 4 and 5 in Belloni and Chernozhukov (2011). □

Proof of Lemma 4. Suppose $\mathbb{E}_\eta$ holds. Then, $\beta(\tau_k) \in \mathbb{R}^{(k)}(r_k)$ $(k = 1, \ldots, K_n)$, where $\mathbb{R}^{(k)}(r_k)$ is defined in (3.1). This implies that

$$\mathcal{B}^{(k)} = [\hat{\beta}^{(1)}, \cdots, \hat{\beta}^{(k-1)}, \beta(\tau_k), \hat{\beta}^{(k+1)}, \cdots, \hat{\beta}^{(K)}]$$

is feasible for all k. We fix any k. Since $\hat{\mathcal{B}}$ is a global minimizer of (3.1), we have $G(\hat{\mathcal{B}}) \le G\left(\mathcal{B}^{(k)}\right)$, where $G(\cdot)$ is defined in (S1.2). This implies

$$\begin{aligned} &\sum_{j=1}^{p} w_j^{(k)}|\hat{\beta}_j^{(k)}| + \frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j=1}^{p} v_j^{(k)}|\hat{\beta}_j^{(k)}-\hat{\beta}_j^{(k-1)}| + \frac{\lambda}{|\tau_{k+1}-\tau_k|}\sum_{j=1}^{p} v_j^{(k+1)}|\hat{\beta}_j^{(k)}-\hat{\beta}_j^{(k+1)}| \\ \le\ &\sum_{j=1}^{p} w_j^{(k)}|\beta_j(\tau_k)| + \frac{\lambda}{|\tau_k-\tau_{k-1}|}\sum_{j=1}^{p} v_j^{(k)}|\beta_j(\tau_k)-\hat{\beta}_j^{(k-1)}| + \frac{\lambda}{|\tau_{k+1}-\tau_k|}\sum_{j=1}^{p} v_j^{(k+1)}|\beta_j(\tau_k)-\hat{\beta}_j^{(k+1)}|. \end{aligned}$$

By the triangle inequality and the definition of $d_{\min}$, it reduces to

$$\sum_{j\in\{T^{(k)}\}^c} w_j^{(k)}|\hat{\beta}_j^{(k)}| \le \sum_{j\in T^{(k)}} w_j^{(k)}(|\beta_j(\tau_k)|-|\hat{\beta}_j^{(k)}|) + \frac{\lambda}{d_{\min}}\sum_{j=1}^{p}(v_j^{(k)}+v_j^{(k+1)})|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)|.$$

Rearranging the terms yields

$$\sum_{j\in\{T^{(k)}\}^c}\left[w_j^{(k)} - \frac{\lambda}{d_{\min}}\{v_j^{(k)}+v_j^{(k+1)}\}\right]|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)| \le \sum_{j\in T^{(k)}}[w_j^{(k)}+\frac{\lambda}{d_{\min}}\{v_j^{(k)}+v_j^{(k+1)}\}]|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)|.$$

By the definition of W_2, W_1, and W,

$$\sum_{j\in\{T^{(k)}\}^c}\left(W_2 - \frac{2\lambda(W_0\vee W_1)}{d_{\min}}\right)|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)| \le \sum_{j\in T^{(k)}}\left(W_1 + \frac{2\lambda(W_0\vee W_1)}{d_{\min}}\right)|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)|.$$

Condition 2 implies $W_2-\frac{2\lambda}{d_{\min}}(W_0\vee W_1) > 0$, and we have for $k=1,\ldots,K_n$,

$$\sum_{j\in\{T^{(k)}\}^c}|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)| \le \frac{d_{\min}W_1+2\lambda(W_0\vee W_1)}{d_{\min}W_2-2\lambda(W_0\vee W_1)}\sum_{j\in T^{(k)}}|\hat{\beta}_j^{(k)}-\beta_j(\tau_k)|,$$

which completes the proof. □

Proof of Lemma 5. Lemma 1.5 in Ledoux and Talagrand (1991) implies that for any independent mean zero random variables $Z_1,\cdots,Z_n$ and positive constants $c_1,\cdots,c_n$, which satisfy $|Z_i|\le c_i$ $(i=1,\ldots,n)$, we have that for any $t>0$,

$$\mathbb{P}\left(|\sum_{i=1}^{n} Z_i| > t\right) \le 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\right). \qquad \text{(S1.20)}$$

Fix j, k, and any $t > 0$. Let $Z_i = x_{ij}[\tau_k - I\{y_i \leq {x_i}^T \beta(\tau_k)\}]/n$, where x_{ij} is the jth component of x_i. By (S1.20), it holds that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} x_{ij}[\tau_k - I\{y_i \leq {x_i}^T\beta(\tau_k)\}]/n\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} x_{ij}^2/n^2}\right) = 2\exp\left(-\frac{nt^2}{2}\right),$$

where we set $c_i = x_{ij}/n$. By the union bound,

$$\mathbb{P}\left(\max_k \max_j |\sum_{i=1}^{n} x_{ij}[\tau_k - I\{y_i \leq {x_i}^T\beta(\tau_k)\}]/n| \geq t\right) \leq 2K_n p \exp(-\frac{nt^2}{2}).$$

Letting $t = 3\sqrt{\log p/n}$ yields

$$\mathbb{P}\left(\max_k \max_j |\sum_{i=1}^{n} x_{ij}[\tau_k - I\{y_i \leq {x_i}^T\beta(\tau_k)\}]/n| \geq 3\sqrt{\log p/n}\right) \leq \frac{1}{n},$$

where we use $p > n \vee K_n$. This completes the proof. □

Proof of Lemma 6. Suppose E_1 holds. Then, we have for all $k = 1, \ldots, K_n$,

$$\begin{aligned}\|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_1 &\leq \sqrt{\|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_0}\|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_2 \\ &\leq \sqrt{(C_4+1)s_0}C_3\sqrt{\frac{s_0 \log p}{n}}. \qquad \text{(S1.21)}\end{aligned}$$

Note that (S1.21) uniformly holds for all k, with probability at least $1 - \gamma_n$.

In the Dantzig-type joint quantile regression setting, stated in Section 4, $r_k = \mathbb{Q}_n^{(k)}\{\tilde{\beta}^{(k)}\} + \Lambda_k \tilde{s} \log p/n$, where $\tilde{s} = \max_k \|\tilde{\beta}^{(k)}\|_0$. Hence, the event $\mathbb{E}_\eta$, defined in (3.3), is equivalent to

$$\mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} \leq \mathbb{Q}_n^{(k)}(\tilde{\beta}^{(k)}) + \Lambda_k \frac{\tilde{s}\log p}{n} \leq \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + \eta \quad (k = 1, \ldots, K_n). \qquad \text{(S1.22)}$$

To prove (S1.22), we use the fact that $\mathbb{Q}_n^{(k)}$ is a convex function and $-\sum_{i=1}^n x_i[\tau_k - I\{y_i \le x_i{}^T\beta(\tau_k)\}]/n$ is the subgradient of $\mathbb{Q}_n^{(k)}$ at $\beta(\tau_k)$. Hence, we have

$$\begin{aligned} & \mathbb{Q}_n^{(k)}\{\tilde{\beta}^{(k)}\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} \\ \ge\ & \left(-\frac{1}{n}\sum_{i=1}^n x_i[\tau_k - I\{y_i \le x_i{}^T\beta(\tau_k)\}]\right)^T \{\tilde{\beta}^{(k)} - \beta(\tau_k)\} \\ \ge\ & -\|\frac{1}{n}\sum_{i=1}^n (x_i[\tau_k - I\{y_i \le x_i{}^T\beta(\tau_k)\}])\|_\infty \|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_1. \quad \text{(S1.23)} \end{aligned}$$

Let E_4 be the event

$$E_4 = \left\{ \|\frac{1}{n}\sum_{i=1}^n (x_i[\tau_k - I\{y_i \le x_i{}^T\beta(\tau_k)\}])\|_\infty \le 3\sqrt{\frac{\log p}{n}} \right\}. \quad \text{(S1.24)}$$

By Lemma 5, $P(E_4) \ge 1 - 1/n$. Combining (S1.21), (S1.23), and (S1.24), we have on event E_4,

$$\begin{aligned} \mathbb{Q}_n^{(k)}\{\tilde{\beta}^{(k)}\} - \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} &\ge -3\sqrt{C_4+1}C_3\frac{s_0 \log p}{n}, \\ &\ge -\Lambda_k \frac{\tilde{s}\log p}{n}, \quad \text{(S1.25)} \end{aligned}$$

where the last inequality utilizes Condition 4. Hence, the first inequality of (S1.22) holds for all k.

Now, by using the fact that $\tilde{\beta}^{(k)}$s and $\tilde{\lambda}$ satisfy (S1.10) on event E_1, we can demonstrate that the second inequality of (S1.22) holds with $\eta = \eta_n^*$ as follows:

$$\begin{aligned}
\mathbb{Q}_n^{(k)}\{\tilde{\beta}^{(k)}\} + \Lambda_k \frac{\tilde{s}\log p}{n} &\leq \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + \tilde{\lambda}\{\|\beta(\tau_k)\|_1 - \|\tilde{\beta}^{(k)}\|_1\} + \Lambda_k \frac{\tilde{s}\log p}{n} \\
&\leq \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + \tilde{\lambda}\|\beta(\tau_k) - \tilde{\beta}^{(k)}\|_1 + \Lambda_k \frac{\tilde{s}\log p}{n} \\
&\leq \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + C_2C_3\sqrt{C_4+1}\frac{s_0\log p}{n} + \Lambda_k \frac{\tilde{s}\log p}{n} \\
&\leq \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + \{C_2C_3\sqrt{C_4+1} + C_4 \max_k \Lambda_k\}\frac{s_0\log p}{n} \\
&= \mathbb{Q}_n^{(k)}\{\beta(\tau_k)\} + \eta_n^*. \qquad \text{(S1.26)}
\end{aligned}$$

Here the first inequality follows from the definition of $\tilde{\beta}^{(k)}$. Combining (S1.25) and (S1.26) implies that (S1.22) holds with $\eta = \eta_n^*$, which completes the proof. □

Proof of Lemma 7. Suppose E_1 holds. Then, we have

$$\begin{aligned}
\min_k \min_{j\in T^{(k)}} |\tilde{\beta}_j^{(k)}| &\geq \min_k \min_{j\in T^{(k)}} |\beta_j(\tau_k)| - \max_k \|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_2 \\
&\geq (3.7\alpha + C_3)\sqrt{\frac{s_0\log p}{n}} - C_3\sqrt{\frac{s_0\log p}{n}} \\
&= 3.7\alpha\sqrt{\frac{s_0\log p}{n}} = 3.7\zeta_n, \qquad \text{(S1.27)}
\end{aligned}$$

where the second inequality follows from Condition 5. Similarly,

$$\begin{aligned}
\min_{k\geq 2} \min_{j\in B^{(k)}} |\tilde{\beta}_j^{(k)} - \tilde{\beta}_j^{(k-1)}| &\geq \min_{k\geq 2} \min_{j\in B^{(k)}} |\beta_j(\tau_k) - \beta_j(\tau_{k-1})| - 2\max_k \|\tilde{\beta}^{(k)} - \beta(\tau_k)\|_2 \\
&\geq (a\alpha + 2C_3)\sqrt{\frac{s_0\log p}{n}} - 2C_3\sqrt{\frac{s_0\log p}{n}} \\
&= a\alpha\sqrt{\frac{s_0\log p}{n}} \geq a\zeta_n. \qquad \text{(S1.28)}
\end{aligned}$$

By (S1.27) and (S1.28), we have $W_1 = 0$, which completes the proof. $\square$

Proof of Lemma 8. Fx $k \in \{1, \cdots, K_n\}$. Let

$$I_2^{(k)}(\theta) := \sum_i \frac{1}{\|\theta\|_2} \int_0^{\sqrt{\frac{d_0}{n}}\{z_i^{(k)}\}^T\theta} I(\epsilon_i^{(k)} \le x) - I(\epsilon_i^{(k)} \le 0) dx := \sum_i I_{2,i}^{(k)}(\theta).$$

Let $D := \{\theta \in \mathbb{R}^{d_0} \mid \|\theta\|_2 \le \frac{1}{nd_0 M_n \sqrt{n}}\}$. First, consider the case in which $\|\theta\|_2 \in D$. Then,

$$\left| \sqrt{\frac{d_0}{n}} \{z_i^{(k)}\}^T \theta \right| \le \sqrt{\frac{d_0 s_0}{n}} \frac{M_n}{nd_0 M_n \sqrt{n}} \le \frac{1}{n^2}.$$

Define the events $\mathbb{B}$ and $\mathbb{C}$ as follows:

$$\mathbb{B} = \left\{ |\epsilon_i^{(k)}| > \frac{1}{n^2}, \quad \text{for all } i. \right\}, \qquad \mathbb{C} = \left\{ \sup_{\theta:\ \theta \in D} I_2^{(k)}(\theta) = 0 \right\}.$$

Then, $\mathbb{P}(\mathbb{B}) \ge 1 - n\frac{2\bar{f}}{n^2} = 1 - \frac{2\bar{f}}{n}$, which implies $\mathbb{P}(\mathbb{C}) \ge 1 - \frac{2\bar{f}}{n}$. Moreover, it holds that

$$\sup_{\theta \in D} \left| \mathbb{E}[I_2^{(k)}(\theta)] \right| \le (1 - \mathbb{P}(\mathbb{C})) \frac{2\bar{f}}{n} M_n \sqrt{nd_0 s_0} \le \frac{2\bar{f} M_n \sqrt{d_0 s_0}}{\sqrt{n}}.$$

Hence, with probability at least $1 - 2\bar{f}/n$, we have

$$\sup_{\theta \in D} \left| I_2^{(k)}(\theta) - \mathbb{E}[I_2^{(k)}(\theta)] \right| \le \frac{2\bar{f} M_n \sqrt{d_0 s_0}}{\sqrt{n}}.$$

Now, consider the case in which $\|\theta\|_2 > 1/(nd_0 M_n \sqrt{n})$. We have for any $\lambda > 0$,

$$\mathbb{P}\left(|I_2^{(k)}(\theta) - \mathbb{E}[I_2^{(k)}(\theta)]| \ge t \right) \le \exp\left(-\lambda t - \lambda \mathbb{E}[I_2^{(k)}(\theta)] \right) \mathbb{E}\left[\exp(\lambda I_2^{(k)}(\theta)) \right]. \tag{S1.29}$$

We have

$$\begin{aligned}\mathbb{E}\left[\exp(\lambda I_2^{(k)}(\theta))\right] &= \prod_i \mathbb{E}\left[\exp(\lambda I_{2,i}^{(k)}(\theta))\right] \\ &= \prod_i \mathbb{E}\left[1+\lambda I_{2,i}^{(k)}(\theta)+\lambda^2(I_{2,i}^{(k)}(\theta))^2 O(1)\right] \\ &= \prod_i \left(1+\lambda\mathbb{E}[I_{2,i}^{(k)}(\theta)]+\lambda^2 O(\mathbb{E}[(I_{2,i}^{(k)}(\theta))^2])\right) \\ &\le \exp\left(\lambda\sum_i \mathbb{E}[I_{2,i}^{(k)}(\theta)]+\lambda^2\sum_i O(\mathbb{E}[(I_{2,i}^{(k)}(\theta))^2])\right), \quad \text{(S1.30)}\end{aligned}$$

where in the second equality $O(1)$ holds uniformly for all i and θ, provided that $\max_i |\lambda I_{2,i}^{(k)}(\theta)| \le \lambda M_n\sqrt{\frac{d_0 s_0}{n}} = o(1)$. Combining (S1.29) and (S1.30), we have

$$\begin{aligned}&\mathbb{P}\left(|I_2^{(k)}(\theta)-\mathbb{E}[I_2^{(k)}(\theta)]| \ge t\right) \\ \le\ & \exp\left(-\lambda t-\lambda\mathbb{E}[I_2^{(k)}(\theta)]+\lambda\sum_i \mathbb{E}[I_{2,i}^{(k)}(\theta)]+\lambda^2\sum_i O(\mathbb{E}[(I_{2,i}^{(k)}(\theta))^2])\right) \\ =\ & \exp\left(-\lambda t+\lambda^2\sum_i O(\mathbb{E}[(I_{2,i}^{(k)}(\theta))^2])\right) \\ =\ & \exp\left(-\lambda t+\lambda^2 O\left(\Delta\frac{s_0^{3/2}d_0^{3/2}}{\sqrt{n}}\right)\right). \quad \text{(S1.31)}\end{aligned}$$

Here we use the fact that

$$\sum_i \mathbb{E}[(I_{2,i}^{(k)}(\theta))^2] \le \frac{1}{\|\theta\|_2^2}\sqrt{\frac{d_0}{n}}\frac{\bar{f}d_0}{2n}\sum_i \|\{|z_i^{(k)}\}^T\theta\|_2^2 \max_i |\{z_i^{(k)}\}^T\theta| \le \frac{\Delta\bar{f}M_n^3}{2}\frac{s_0^{3/2}d_0^{3/2}}{\sqrt{n}}.$$

Since $\frac{tM_n}{\Delta s_0 d_0 \log n} = o(1)$, choosing $\lambda = \frac{t\sqrt{n}}{2\Delta s_0^{3/2}d_0^{3/2}\log n}$ in (S1.31) implies

$$\mathbb{P}(|I_2^{(k)}(\theta)-\mathbb{E}[I_2^{(k)}(\theta)]| \ge t) \le \exp\left(-\frac{t^2\sqrt{n}}{4\Delta s_0^{3/2}d_0^{3/2}\log n}\right).$$

Now, to apply the chaining argument, consider ϵ-size balls that cover

Θ. Let B be the set of centers of the balls. Then, we have

$$\mathbb{P}(\sup_{\theta\in B}|I_2^{(k)}(\theta)-\mathbb{E}[I_2^{(k)}(\theta)]|\geq t)\leq\exp\left(d_0\log\frac{2\Delta}{\epsilon}-\frac{t^2\sqrt{n}}{4\Delta s_0^{3/2}d_0^{3/2}\log n}\right).$$

Moreover, if $\theta_1,\theta_2\notin D$ and $|\theta_1-\theta_2|\leq\epsilon$, then

$$\begin{aligned}&\left|I_2^{(k)}(\theta_1)-\mathbb{E}[I_2^{(k)}(\theta_1)]-I_2^{(k)}(\theta_2)+\mathbb{E}[I_2^{(k)}(\theta_2)]\right|\\ \leq\ &\left|I_2^{(k)}(\theta_1)-I_2^{(k)}(\theta_2)\right|+\left|\mathbb{E}[I_2^{(k)}(\theta_1)]-\mathbb{E}[I_2^{(k)}(\theta_2)]\right|.\end{aligned}$$

Note that

$$\begin{aligned}&|I_2^{(k)}(\theta_1)-I_2^{(k)}(\theta_2)|\leq\frac{1}{\|\theta_1\|_2\|\theta_2\|_2}\left|\sum_i\|\theta_2\|_2\|\theta_1\|_2I_{2i}^{(k)}(\theta_1)-\sum_i\|\theta_1\|_2\|\theta_2\|_2I_{2i}^{(k)}(\theta_2)\right|\\ \leq\ & n^3d_0^2M^2\left(\|\theta_2\|_2\left|\sum_i\|\theta_1\|_2I_{2i}^{(k)}(\theta_1)-\sum_i\|\theta_2\|_2I_{2i}^{(k)}(\theta_2)\right|+|\|\theta_2\|_2-\|\theta_1\|_2|\left|\sum_i\|\theta_2\|_2I_{2i}^{(k)}(\theta_2)\right|\right)\\ \leq\ & n^3d_0^2M_n^2\left(\Delta n\sqrt{\frac{d_0}{n}}\sqrt{s_0}M_n\epsilon+\epsilon n\sqrt{\frac{d_0}{n}}\sqrt{s_0}M_n\Delta\right)\\ =\ & 2n^{3.5}d_0^{2.5}s_0^{0.5}M_n^3\epsilon\Delta.\end{aligned}$$

Similarly,

$$\begin{aligned}\left|\mathbb{E}[I_2^{(k)}(\theta_1)]-\mathbb{E}[I_2^{(k)}(\theta_2)]\right| &\leq n^3d_0^2M_n^2\left(\bar{f}\frac{d_0}{n}ns_0M_n^2\epsilon^2+\bar{f}\epsilon n\Delta^2\frac{d_0}{n}s_0M_n^2\right)\\ &\leq 2n^3d_0^3s_0M_n^2\epsilon\Delta^2.\end{aligned}$$

If we choose t such that $\epsilon n^{3.5}d_0^3M_n^3=o(t)$ and $\epsilon n^3d_0^4M_n^2=o(t)$ with ϵ being small enough, then we have

$$\mathbb{P}\left(\sup_{\theta\in\Theta\setminus D}|I_2^{(k)}(\theta)-\mathbb{E}[I_2^{(k)}(\theta)]|\geq 2t\right)\leq\exp\left(d_0\log\frac{2\Delta}{\epsilon}-\frac{t^2\sqrt{n}}{4\Delta s_0^{3/2}d_0^{3/2}\log n}\right).$$

Hence,

$$\mathbb{P}\left(\sup_{\theta\in\Theta\setminus D}|I_2(\theta)-\mathbb{E}[I_2(\theta)]|/\|\theta\|_2\geq 2tK_n\right)\leq\exp\left(\log K_n+d_0\log\frac{2\Delta}{\epsilon}-\frac{t^2\sqrt{n}}{4\Delta s_0^{3/2}d_0^{3/2}\log n}\right).$$

Letting $t = 3n^{-1/4}\Delta^{1/2}s_0^{3/4}d_0^{5/4}(\log n)^{3/2}$ and $\epsilon = n^{-9}$ with the growth condition $d_0 M_n^4 \log^2 n = o(n)$ yields

$$\mathbb{P}(\sup_{\theta\in\Theta\setminus D}|I_2(\theta)-\mathbb{E}[I_2(\theta)]|/\|\theta\|_2 \geq 6n^{-1/4}K_n\Delta^{1/2}s_0^{3/4}d_0^{5/4}(\log n)^{3/2}) \leq n^{-9d_0\log n}.$$

We have shown with probability at least $1-\frac{2\bar{f}K_n}{n}$ that

$$\sup_{\theta\in D}|I_2(\theta)-\mathbb{E}[I_2(\theta)]| \leq \frac{2\bar{f}M_n\sqrt{d_0 s_0}K_n}{\sqrt{n}}.$$

Therefore, we have with probability at least $1-n^{-9d_0\log n}-\frac{2\bar{f}K_n}{n}$,

$$\sup_{\theta\in\Theta}|I_2(\theta)-\mathbb{E}[I_2(\theta)]| \leq 7n^{-1/4}K_n\Delta^{1/2}s_0^{3/4}d_0^{5/4}(\log n)^{3/2}.$$

□

Proof of Lemma 9. First, we have

$$K_n\underline{f}k^2(s_0,0) \leq \lambda_{\min}(A_n) \leq \lambda_{\max}(A_n) \leq K_n\bar{f}\phi(s_0),$$

$$\big(\sum_{k,k'}\tau_k\wedge\tau_{k'}-\tau_k\tau_{k'}\big)k^2(s_0,0) \leq \lambda_{\min}(B_n) \leq \lambda_{\max}(B_n) \leq \big(\sum_{k,k'}\tau_k\wedge\tau_{k'}-\tau_k\tau_{k'}\big)\phi(s_0).$$

Hence, it holds that

$$\begin{aligned}
\lambda_{\min}({A_n}^{-1}B_n{A_n}^{-1}) &\geq \lambda^2_{\min}({A_n}^{-1})\lambda_{\min}(B_n) \\
&= \lambda^{-2}_{\max}(A_n)\lambda_{\min}(B_n) \\
&\geq \bar{f}^{-2}\phi^{-2}(s_0)k^2(s_0,0)\frac{\sum_{k,k'}\tau_k\wedge\tau_{k'}-\tau_k\tau_{k'}}{K_n^2} \\
&\geq \bar{f}^{-2}\phi^{-2}(s_0)k^2(s_0,0)(\min_k\tau_k)(1-\max_k\tau_k).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}\lambda_{\max}({A_n}^{-1}B_n{A_n}^{-1}) &\leq \lambda^2_{\max}({A_n}^{-1})\lambda_{\max}(B_n)\\ &= \lambda^{-2}_{\min}(A_n)\lambda_{\max}(B_n)\\ &\leq L_0^{-2}\phi(s_0)k^{-4}(s_0,0)\frac{\sum_{k,k'}\tau_k\wedge\tau_{k'}-\tau_k\tau_{k'}}{K_n^2}\\ &\leq L_0^{-2}\phi(s_0)k^{-4}(s_0,0),\end{aligned}$$

which completes the proof. □

Proof of Lemma 10. Recall $\epsilon_i^{(k)} = y_i - {x_i}^T\beta(\tau_k) = y_i - \{z_i^{(k)}\}^T T(\mathcal{B}^o)$. Now, define D_n as follows:

$$D_n = \alpha_n^T({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}{A_n}^{-1}n^{-0.5}\sum_i\sum_k z_i^{(k)}\{I(\epsilon_i^{(k)}<0)-\tau_k\} := \sum_i Z_{ni},$$

where $Z_{ni} = (n^{-0.5})\Big[\alpha_n^T({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}{A_n}^{-1}\sum_k z_i^{(k)}\{I(\epsilon_i^{(k)}<0)-\tau_k\}\Big]$. Then, $\mathbb{E}[Z_{ni}] = 0$ and

$$\begin{aligned}&\sum_i \mathrm{Var}(Z_{ni})\\ =\ &\sum_i \alpha_n^T({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}{A_n}^{-1}\sum_{k,k'}\frac{1}{n}z_i^{(k)}\{z_i^{(k)}\}^T\{\min(\tau_k,\tau_{k'})-\tau_k\tau_{k'}\}{A_n}^{-1}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}\alpha_n\\ =\ &\alpha_n^T({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}{A_n}^{-1}B_n{A_n}^{-1}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}\alpha_n\\ =\ &1.\end{aligned}$$

Consider an upper bound of Z_{ni} for all $i = 1,\dots,n$:

$$|Z_{ni}| \leq \left\|\sum_k z_i^{(k)}\{I(\epsilon_i^{(k)}<0)-\tau_k\}\right\|_2 \left\|{A_n}^{-1}({A_n}^{-1}B_n{A_n}^{-1})^{-\frac{1}{2}}\alpha_n\right\|_2/\sqrt{n}. \tag{S1.32}$$

Since $\sum_k z_i^{(k)}\{I(\epsilon_i^{(k)} < 0) - \tau_k\}$ is a d_0-dimensional vector and the absolute value of each components is upper bounded by $K_n M_n$,

$$\|\sum_k z_i^{(k)}\{I(\epsilon_i^{(k)} < 0) - \tau_k\}\|_2 \leq \sqrt{d_0} K_n M_n. \tag{S1.33}$$

Since $\|\alpha_n\|_2 = 1$, we have

$$\begin{aligned} \|({A_n}^{-1} B_n {A_n}^{-1})^{-\frac{1}{2}} \alpha_n\|_2 &\leq \lambda_{\max}\{({A_n}^{-1} B_n {A_n}^{-1})^{-\frac{1}{2}}\} \\ &= \{\lambda_{\min}({A_n}^{-1} B_n {A_n}^{-1})\}^{-0.5} \\ &\leq \bar{f}\phi(s_0) k^{-1}(s_0, 0)(\min_k \tau_k)^{-0.5}(1 - \max_k \tau_k)^{-0.5}, \end{aligned}$$

where the second inequality utilizes Lemma 9. Similarly, it holds that

$$\begin{aligned} &\|{A_n}^{-1}({A_n}^{-1} B_n {A_n}^{-1})^{-\frac{1}{2}} \alpha_n\|_2 \leq \lambda_{\max}({A_n}^{-1}) \|({A_n}^{-1} B_n {A_n}^{-1})^{-\frac{1}{2}} \alpha_n\|_2 \\ \leq\ & K_n^{-1} L_0^{-1} k^{-3}(s_0, 0) \bar{f}\phi(s_0)(\min_k \tau_k)^{-0.5}(1 - \max_k \tau_k)^{-0.5}. \end{aligned} \tag{S1.34}$$

Combing (S1.32), (S1.33), and (S1.34), it can be derived that

$$\max_i |Z_{ni}| \leq \sqrt{d_0/n} M_n L_0^{-1} k^{-3}(s_0, 0) \bar{f}\phi(s_0)(\min_k \tau_k)^{-0.5}(1 - \max_k \tau_k)^{-0.5}.$$

Hence, we have

$$\begin{aligned} \sum_i \mathbb{E}(|Z_{ni}|^3) &\leq \sum_i E(|Z_{ni}|^2) \sqrt{\frac{d_0}{n}} M_n L_0^{-1} k^{-3}(s_0, 0) \bar{f}\phi(s_0)(\min_k \tau_k)^{-0.5}(1 - \max_k \tau_k)^{-0.5} \\ &= \sqrt{d_0/n} M_n L_0^{-1} k^{-3}(s_0, 0) \bar{f}\phi(s_0)(\min_k \tau_k)^{-0.5}(1 - \max_k \tau_k)^{-0.5} \\ &\to 0. \end{aligned}$$

Thus, $\{Z_{ni}\}_{i=1}^{n}$ for all n are triangular array satisfying Lyapunov Condition. By central limit theorem for triangular arrays,

$$\sum_i Z_{ni} \to N(0,1),$$

which completes the proof. □

S2 Additional simulation results

This section includes the additional examples of the simulation study. See Section 7 of the main paper for details of the simulation settings. We considered the following additional examples to investigate the stability of selected models from four methods; Lasso, ALasso, FAL, and Dantzig, which are defined in Section 7. The performance measures are shown in Figure 1.

Example 1. Consider the model, which is same as Example 1 in the main paper except that ϵ_i's follow the standard Cauchy distribution.

Example 2. Consider the model, which is same as Example 1 in the main paper except that ϵ_i's follow the standard Laplace distribution.

Across all figures, the largest standard errors for the false positives, the false negatives, and the size of set differences are less than 0.9, 0.1, and 0.4, respectively. As shown in Figure 1, the results are consistent with those reported in the main paper.

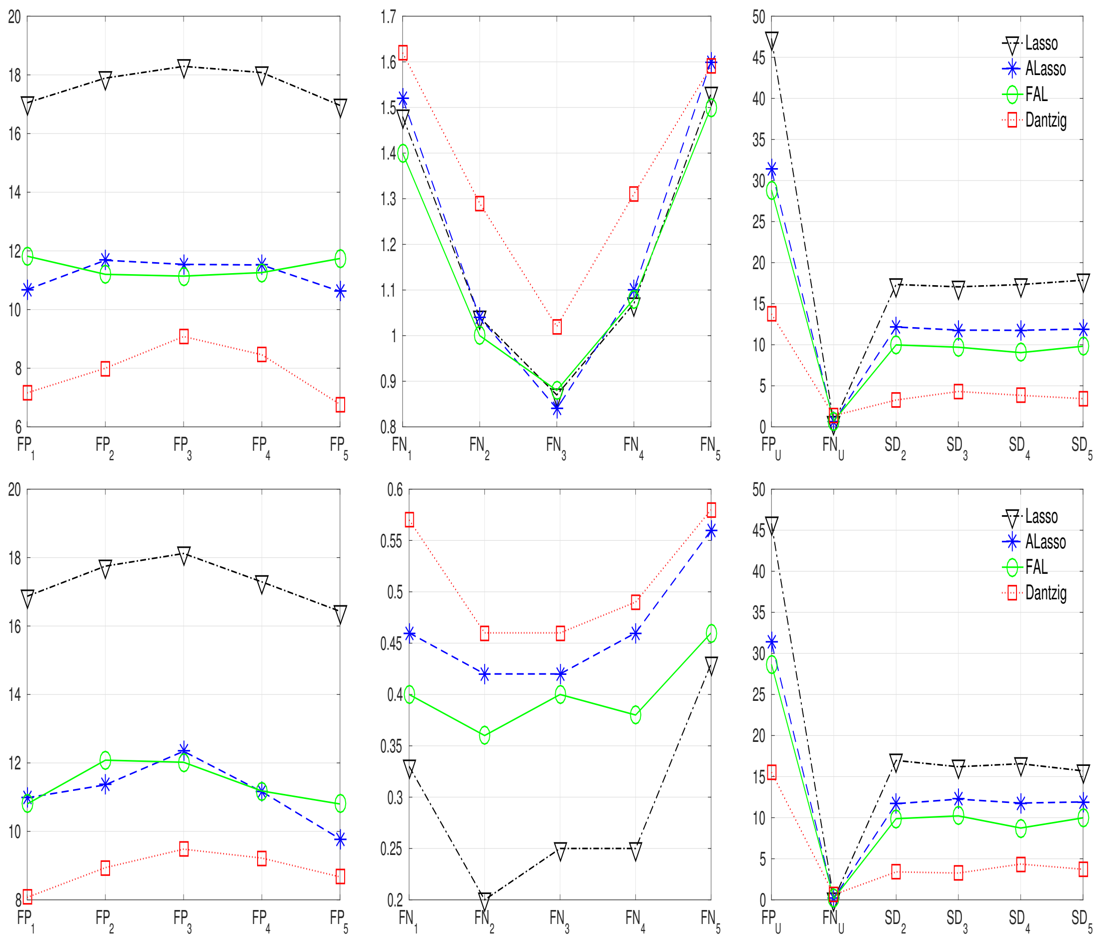

Figure 1: Results for Example 1 (top) and 2 (below): Each plot shows the false positives(left), the false negatives (middle), and the stability measures (right). Four competing procedures are evaluated: Lasso, ALasso, FAL, and Dantzig.

References

Belloni, A. and Chernozhukov, V. (2011). l_1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics* **39**, 82–130.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin.

Wang, L., Wu, Y., and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association* **107**, 214–222.