Supplementary Material for High-dimensional Extreme Quantile Regression

Yiwei Tang*, Huixia Judy Wang**, Deyuan Li*

Fudan University, Rice University***

Supplementary Material

The supplementary material includes technical conditions, proofs of the theorems from the main paper, additional sensitivity analysis results, and descriptions of variables used in the auto insurance claims data analysis.

S1 Technical Conditions and Theoretical Proofs

We present technical conditions in Subsection S1.1, followed by the proofs of Theorems 1, 2, and 3 in Subsections S1.2, S1.3, and S1.4, respectively. Each subsection also includes relevant lemmas and remarks to further clarify the results.

We first introduce some notations. For simplicity, we denote $p_{n,k} = c_k(1-\tau_{0n})$, $k = 1, 2$ and $c_k > 0$, and $f_i(\cdot) := f_Y(\cdot \mid \mathbf{X}_i)$, $i = 1, \ldots, n$. For any $\boldsymbol{u} = (u_1, u_2, \ldots, u_{p+1})^T \in \mathbb{R}^{p+1}$, denote $\boldsymbol{u}_{[p]} = (u_2, \ldots, u_{p+1})^T \in \mathbb{R}^p$. For a random variable X, we use $\|X\|_{\psi_2}$ to denote its sub-Gaussian norm, i.e., $\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leq 2\}$. For a random vector $\mathbf{X} \in \mathbb{R}^d$, we use $\|\mathbf{X}\|_{\psi_2}$ to denote its sub-Gaussian norm, that is, $\|\mathbf{X}\|_{\psi_2} = \sup_{\mathbf{x}\in S^{d-1}} \|\langle \mathbf{X}, \mathbf{x}\rangle\|_{\psi_2}$, where $S^{d-1} = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\right\}$ is Euclidean sphere. Note that $F_Y(y \mid \mathbf{Z}) = F_Y(y \mid \mathbf{X})$ as $\mathbf{Z} = (1, \mathbf{X}^T)^T$. Let $\mathbb{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^\top, \mathbb{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^\top$, and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$.

For an empirical process $f \mapsto (\mathbb{E}_n - P) f = n^{-1} \sum_{i=1}^n (f(X_i) - Pf)$, we define the symmetrized process $f \mapsto \mathbb{E}_n^o f = n^{-1} \sum_{i=1}^n \varepsilon_i f(X_i)$, where $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. Rademacher random variables independent of $(X_1, \ldots, X_n)$. Let $Q_\tau(\boldsymbol{\beta}) = \mathbb{E}\left\{\rho_\tau\left(Y - \mathbf{Z}^T\boldsymbol{\beta}\right)\right\}$, $\widehat{Q}_\tau(\boldsymbol{\beta}) = \mathbb{E}_n\left\{\rho_\tau\left(Y_i - \mathbf{Z}_i^T\boldsymbol{\beta}\right)\right\}$. Define

$$M_n(\boldsymbol{u}, \tau) := \frac{1}{n}\left\{\sum_{i=1}^n \rho_\tau\left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}(\tau) - \mathbf{Z}_i^\top \boldsymbol{u}\right) - \rho_\tau\left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}(\tau)\right)\right\},$$

$$\mathbb{E}\{M_n(\boldsymbol{u}, \tau)\} := \frac{1}{n}\left[\sum_{i=1}^n \mathbb{E}\left\{\rho_\tau\left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}(\tau) - \mathbf{Z}_i^\top \boldsymbol{u}\right)\right\} - \mathbb{E}\left\{\rho_\tau\left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}(\tau)\right)\right\}\right].$$

S1.1 Conditions

Condition C1. For $i = 1, 2, \ldots, n$, $f_i(\cdot)$ is continuous differentiable, and there exist constants $\bar{f}, \bar{f}' > 0$, such that $|f_i(y)| < \bar{f}$, $|\partial f_i(y)/\partial y| < \bar{f}'$ for all $y \in \mathbb{R}$.

Condition C2. (1) Suppose covariance matrix $\Sigma := \mathbb{E}(\mathbf{X}_i\mathbf{X}_i^T)$ satisfies that $\Sigma_{ii} = 1, i = 1, 2, \ldots, p$ and $0 < C_{min} \leq \lambda_{min}(\Sigma) \leq \lambda_{max}(\Sigma) \leq C_{max} < \infty$

for some $C_{min}, C_{max} > 0$. (2) $\mathbf{X}$ is sub-Gaussian with $\|\mathbf{X}\|_{\psi_2} \leq \kappa$, where κ is some positive constant.

In order to state the next condition, for some $c_0 \geq 0$ and $\tau \in [\tau_{ln}, 1)$, define $A_\tau := \left\{\mathbf{u} \in \mathbb{R}^{p+1} : \left\|\mathbf{u}_{T_\tau^c}\right\|_1 \leq c_0 \left\|\mathbf{u}_{T_\tau}\right\|_1\right\}$, where $T_\tau = \text{support}(\boldsymbol{\beta}(\tau)) := \{j \in \{0, 1, \ldots, p\} : \beta_j(\tau) \neq 0\}$. A_τ is referred to as the restricted set. Define $\bar{T}_\tau(\mathbf{u}, m) \subset \{1, \ldots, p\} \backslash T_\tau$ as the support of the m largest in absolute value components of the vector $\mathbf{u}$ outside of T_τ. Obviously, $\bar{T}_\tau(\mathbf{u}, m)$ is the empty set if $m = 0$.

Condition C3. For some constants $m \geq 0$ and $c_0 \geq 9$, the matrix $\widetilde{\Sigma} = diag\,(1, \Sigma)$ satisfies

$$\kappa_m^2 := \inf_{1-\tau \in \mathcal{T}_n} \inf_{\mathbf{u} \in A_\tau, \mathbf{u} \neq 0} \frac{\mathbf{u}^T \widetilde{\Sigma} \mathbf{u}}{\left\|\mathbf{u}_{T_\tau \cup \bar{T}_\tau(\mathbf{u}, m)}\right\|^2} > 0$$

and $\log\left(L\kappa_0^2\right) \leq C_L \log\left(p \vee n\right)$ for some constant C_L. Moreover,

$$q := \inf_{1-\tau \in \mathcal{T}_n} \inf_{\mathbf{u} \in A_\tau, \mathbf{u} \neq 0} \frac{\mathbb{E}\left[\left|\mathbf{Z}_i' \mathbf{u}\right|^2\right]^{3/2}}{\mathbb{E}\left[\left|\mathbf{Z}_i' \mathbf{u}\right|^3\right]} > 0.$$

Condition C4. The density $f_Y(\cdot|\mathbf{X})$ is monotone in its upper tails for any given $\mathbf{X} \in \mathcal{X}$.

Condition C6. For each $j = 0, 1, \ldots, p$, $\beta_j(\cdot)$ is continuous and differentiable for all $\tau \in [\tau_{ln}, 1)$. Moreover, for any sequence $\tau_n \to 1$, we have

$\max_{1\le j\le p}\left\{|\beta_j'(\tau_n)|\right\} \lesssim (1-\tau_n)^{-\gamma-1}$, where $\beta_j'(\tau) = \partial\beta_j(\tau)/\partial\tau$.

Remark S.1 (Additional Explanation of Condition C5.)**.** *Condition C5 introduces an auxiliary parameter vector $\boldsymbol{\theta}_r$ to characterize the tail behavior of $Y \mid \mathbf{X}$ in high-dimensional settings. The linear component $\mathbf{X}^\top\boldsymbol{\theta}_r$ captures the central location, and the residual is defined as $U = Y - \mathbf{X}^\top\boldsymbol{\theta}_r$. The condition then describes the tail behavior of the conditional distribution $F_U(t_n \mid \mathbf{Z})$ and density $f_U(t_n \mid \mathbf{Z})$ as $t_n \to \infty$, assuming*

$$1 - F_U(t_n \mid \mathbf{Z}) \sim K(\mathbf{Z})\{1 - F_0(t_n)\}, \quad f_U(t_n \mid \mathbf{Z}) \sim K(\mathbf{Z})f_0(t_n),$$

along with second-order terms involving $K_1(\mathbf{Z})$ and $K_2(\mathbf{Z})$, which capture higher-order deviations in the distribution and density, respectively. These functions serve as covariate-dependent generalizations of the expansion coefficients used in fixed-dimensional models (e.g., Chernozhukov 2005, Wang, Li & He 2012).

Unlike these works, our framework allows the covariate dimension to grow with the sample size. Accordingly, $K(\cdot)$, $K_1(\cdot)$, and $K_2(\cdot)$ are permitted to diverge with n, and we impose additional high-dimensional conditions (e.g., $d_n\{1 - F_0(t_n)\} \to 0$ and $d_n f_0(t_n) \to 0$ as $n \to \infty$) to control their influence.

Importantly, our theoretical analysis only relies on the first-order expansion of f_U *(see Lemma 3 in the Supplementary Material), so we do not impose any specific condition on the rate of* $K_2(\cdot)$*. Although these functions may vary over* $\mathcal{Z}$*, we assume that* $K(\cdot)$ *and* $K_1(\cdot)$ *share common convergence rates* d_n *and* d_{1n}*, respectively. This assumption facilitates the theoretical analysis. For instance, in Example 1 with compact* $\mathcal{Z}$*, the rate of* $K(\mathbf{Z})$ *is determined by the sparsity of* $\tilde{\boldsymbol{\sigma}}$*. Even if* $K(\mathbf{Z})$ *lies within a range, e.g.,* $d_n \lesssim K(\mathbf{Z}) \lesssim D_n$*, our results remain valid with minor modifications.*

The second-order condition of U_0 *implies that* $U_0(tz)/U_0(t) \to z^\gamma$*, which can be interpreted as a first-order condition (see Condition C6 for the definition). The first-order condition is a necessary and sufficient condition for* F_0 *to belong to the maximum domain of attraction* $D(G_\gamma)$ *for heavy-tailed distributions (de Haan & Ferreira 2006). Most commonly used families of continuous distributions satisfies the second-order condition. Although both the extreme value index estimator and the extreme quantile estimator are based on the first-order condition, the second-order condition is essential for deriving their theoretical results.*

Note that Condition C5 is a sufficient but not necessary condition for establishing the rate at which $f_Y(Q_Y(\tau|\mathbf{X})|\mathbf{X}) \to 0$*. Nevertheless, it serves as a practical and verifiable condition under commonly used models, such* *as the location-scale-shift models (see Example 1 in Section 2.3).*

In summary, Condition C5 decomposes the tail quantile of $Y \mid \mathbf{X}$ *into a univariate baseline and covariate-dependent components through* $\boldsymbol{\theta}_r$, $K(\cdot)$, $K_1(\cdot)$, *and* $K_2(\cdot)$.

Remark S.2. *Conditions C1–C3 are standard in the literature on high-dimensional quantile regression. Condition C1 imposes mild smoothness assumptions on the conditional density of the response variable given covariates. Similar conditions appear in Belloni & Chernozhukov (2011) and Wang, Wu & Li (2012), but our assumption is even weaker since we do not require the conditional density to be bounded away from zero. This is more general than the commonly assumed Gaussian or sub-Gaussian conditions. Condition C2 posits a natural probabilistic model for the design matrix, assuming a well-behaved covariance structure and that* $\mathbf{X}$ *has sub-Gaussian rows. This setup is widely adopted in high-dimensional inference (e.g., Battey et al. (2018), Javanmard & Montanari (2014)). Condition C3 is a stronger version of the restricted eigenvalue condition and restricted nonlinear impact condition in Belloni & Chernozhukov (2011) because it necessitates a positive uniform lower bound for all vectors in* A_τ *when* $1-\tau$ *equals different values in* $\mathcal{T}_n$. *Various primitive conditions that imply bounds on* κ_m *can be found in the same reference. Actually, this condition is mild* *since we consider* $1 - \tau \in \mathcal{T}_n$ *is at the same rate. Condition C4 assumes regularity of the distribution function and its derivatives. It is a technical condition imposed to simplify the derivations, and is also assumed in Zhang (2018) (Assumption 4). Condition C6 is a regularity assumption on the tail behavior of the functions* $\beta_j(\tau)$*, requiring that they do not vary too rapidly as* $\tau \to 1$*. This is a mild condition, satisfied by both the location shift model and the location–scale shift model. If, however, there exists a coefficient function* $\beta_j(\tau)$ *that violates this condition — for example,* $\beta_j(\tau) = (1-\tau)^{-\gamma'}$*,with* $\gamma' > \gamma$ *— then the extreme value index of* $Y \mid \mathbf{X}$ *would necessarily be* γ'*, which contradicts the assumed tail behavior.The maximum domain of attraction condition covers a broad range of commonly encountered distributions. Specifically, many well-known distributions (e.g., Pareto, Student's t, Fréchet with* $\gamma > 0$*; normal, gamma, exponential with* $\gamma = 0$*; uniform with* $\gamma < 0$*) satisfy this condition.*

Remark S.3. *Here we provide some guidance on how to verify Conditions C1–C6 in Example 1 of the main paper. Condition C1 and C4 are regularity assumptions on the conditional distribution of* $Y \mid \mathbf{X}$*, and are typically easy to satisfy. For instance, one can let the error term* ϵ *follow a Student's t-distribution with degrees of freedom* v*. Condition C2 and C3 holds under mild assumptions on the distribution of* $\mathbf{X}$*. For example, C2 is sat-*

isfied if the components of $\mathbf{X}$ *are independently and uniformly distributed. Condition C5 can be verified using the techniques presented in the proofs of Lemmas 4 and 5, under the configuration specified in Example 1 of the main paper. Condition C6 is satisfied when the quantile function takes the form* $\beta_j(\tau) = \tilde{\sigma}_j F_0^{-1}(\tau)$, *where* $F_0^{-1}(\tau)$ *satisfies a first-order condition, and the effective sparsity level* s^* *depends on the number of nonzero entries in* $\tilde{\sigma}$.

S1.2 Proof of Theorem 1

In order to prove Theorem 1, we first introduce Lemmas 1-9. To obtain a uniform penalty level over the tail region, we define the random variable

$$\tilde{\Lambda} = n \sup_{1-\tau \in \mathcal{T}_n} \left\{ \max_{1 \le j \le p} \left| \mathbb{E}_n \left(\frac{X_{ij}\left(\tau - 1\left(u_i \le \tau\right)\right)}{\widehat{\sigma}_j \sqrt{\tau(1-\tau)}} \right) \right| \vee \left| \mathbb{E}_n \left(\frac{\tau - 1(u_i \le \tau)}{\sqrt{\tau\left(1-\tau\right)}} \right) \right| \right\},$$

where $u_1, \ldots, u_n$ are i.i.d. random variables from uniform$(0,1)$, independently distributed from the regressors, $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Lemma 1 provides an asymptotic upper bound for $\tilde{\Lambda}$, which is $O_p((n \log p)^{1/2} (1-\tau_{0n})^{-1/2})$, and this bound is used to determine the order of λ in the proof of Theorem 1. Lemma 2 shows that the error vector $\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)$ lies within a restricted set, uniformly for all $1-\tau \in \mathcal{T}_n$, with high probability if $\lambda \ge (c_0+3)/(c_0-3)\tilde{\Lambda}$. Lemma 4 and Lemma 5 build on the foundational result in Lemma 3 to derive second-order Taylor expansions for the conditional distribution and quantile functions at the tail. Lemma 6 offers a quadratic lower bound for $\mathbb{E}\left\{M_n\left(\boldsymbol{u},\tau\right)\right\}$ when n is large. Lemma 7 establishes a bound for the empirical error, $M_n\left(\boldsymbol{u},\tau\right)-\mathbb{E}\left\{M_n\left(\boldsymbol{u},\tau\right)\right\}$, uniformly across all vectors in the restricted set A_τ where $1-\tau\in\mathcal{T}_n$. Lemma 8 and Lemma 9 are auxiliary results that establish relationships between sub-exponential and sub-Gaussian squared random variables and provide a Bernstein-type inequality for exponential random variables, respectively.

Lemma 1. *Assume $1-\tau_{0n}\to 0$. There exists a universal constant $C_{\tilde{\Lambda}}>0$ such that,*

$$P\left(\tilde{\Lambda}\geq \tilde{K}C_{\tilde{\Lambda}}\sqrt{n\log p}/\sqrt{1-\tau_{0n}}\mid \mathbb{X}\right)\leq p^{-\tilde{K}^2+1},$$

for any $\tilde{K}\geq 1$.

Proof. We first consider $\mathbb{G}_n\left(Z_{ij}(\tau-1(u_i\leq\tau))/\widehat{\sigma}_j\right)$ for $1\leq j\leq p$ and $1-\tau\in\mathcal{T}_n$. Applying Lemma 1.5 in Ledoux & Talagrand (2013), we have

$$P\left(\left|\mathbb{G}_n\left\{Z_{ij}(\tau-1(u_i\leq\tau))/\widehat{\sigma}_j\right\}\right|\geq K\mid\mathbb{X}\right)\leq 2\exp\left(-K^2/2\right),$$

for any $K>0$.

Therefore, by Lemma 2.3.7 in van der Vaart & Wellner (1996), for

$K_2 \geq 2\sqrt{2}$, we have

$$P\left(\tilde{\Lambda} > K_2\sqrt{n} \mid \mathbb{X}\right) \leq 4p \max_{1\leq j\leq p} P\left(\sup_{1-\tau\in\mathcal{T}_n} |\mathbb{G}_n^o\left\{(\tau - 1(u_i \leq \tau))Z_{ij}/\widehat{\sigma}_j\right\}| > K_2/(4W_\tau) \mid \mathbb{X}\right),$$

where $\mathbb{G}_n^o(f) = \sqrt{n}\mathbb{E}_n^o(f)$ and $c_n = \max_{1-\tau\in\mathcal{T}_n}\left(\tau(1-\tau)\right)^{-1/2}$.

Next, we adapt the proof of Theorem 1 in Belloni & Chernozhukov (2011) by replacing $u \in \mathcal{U}$ with $1-\tau = c(1-\tau_{0n}), c \in \mathcal{T}$, where $\mathcal{T} = [c_1, c_2]$ is a compact set in $\mathbb{R}$. Specifically, for each $j = 1, 2, \ldots, p$, the function class

$$\mathcal{F}_j = \{Z_{ij}(\tau - 1(u_i \leq \tau))/\widehat{\sigma}_j : 1 - \tau = c(1-\tau_{0n}), c \in [c_1, c_2]\}$$

is a VC class. For $k \geq 1$, we obtain $P\left(\tilde{\Lambda} \geq kC_4 c_n\sqrt{n\log p} \mid \mathbb{X}\right) \leq p^{-k^2+1}$, where $C_4 > 0$ is a constant, and $c_n = \max_{1-\tau\in\mathcal{T}_n}\left(\tau(1-\tau)\right)^{-1/2} = \left(p_{n,1}(1-p_{n,1})\right)^{-1/2}$, with $p_{n,1} = c_1(1-\tau_{0n})$.

Since $1-\tau_{0n} \to 0$ as $n \to \infty$, for sufficiently large n and $k \geq 1$, we have

$$P\left(\tilde{\Lambda} \geq kC_{\tilde{\Lambda}}\sqrt{n\log p}/\sqrt{1-\tau_{0n}} \mid \mathbb{X}\right) \leq p^{-k^2+1},$$

where $C_{\tilde{\Lambda}} > 0$ is a constant. □

Lemma 2. *(i) Under Condition C2, for any $\epsilon \in (0,1)$, there exists some $N_\epsilon > 0$, such that for all $n > N_\epsilon$, we have that, for every $\mathbf{u} \in \mathbb{R}^{p+1}$,*

$$\frac{2}{3}\|\mathbf{u}\|_{1,n} \leq \|\mathbf{u}\|_1 \leq 2\|\mathbf{u}\|_{1,n}, \tag{S1.1}$$

with probability at least $1-\epsilon$.

(ii) Moreover, conditional on event $\{\lambda \geq (c_0+3)/(c_0-3)\tilde{\Lambda}\}$, *we have that, for* $n > N_\epsilon$, *uniformly in* $1-\tau \in \mathcal{T}_n$, $\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \in A_\tau$ *holds with probability at least* $1-\epsilon$.

Remark S.4. *We emphasize that part (ii) contains the main result, and its proof crucially depends on the conclusion established in part (i).*

Proof. (i) By Condition C2, $\mathbf{X}_i$ is sub-Gaussian, which implies that $\mathbf{X}_i^2$ is sub-exponential by Lemma 8 (see below). Moreover, by the concentration inequalities for sub-exponential random variables in Lemma 9 (see below), we conclude that, for $t \geq 0$ and $j = 1, 2, \ldots, p$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_{ij}^2 - 1\right| \geq t\right) \leq 2\exp\left(-cn\min\left(\frac{t^2}{\kappa^4}, \frac{t}{\kappa^2}\right)\right),$$

where c and κ are some positive constants. Therefore, for $t = 1/2$,

$$P\left(\max_{1\leq j\leq p}\left|\frac{1}{n}\sum_{i=1}^{n} X_{ij}^2 - 1\right| \geq \frac{1}{2}\right) \leq 2p\exp\left(-cn\min\left(\frac{1}{4\kappa^4}, \frac{1}{2\kappa^2}\right)\right) \to 0,$$

as $\log(p)/n \to 0$. Hence, for any $\epsilon \in (0,1)$, there exists some $N_\epsilon > 0$ such that for all $n > N_\epsilon$, $P\left(\max_{1\leq j\leq p}|\widehat{\sigma}_j - 1| \leq 1/2\right) > 1-\epsilon$.

Therefore, for all $n > N_\epsilon$, $1/2 \leq \widehat{\sigma}_j \leq 3/2$, which implies the statement (i).

(ii) The proof of this part is similar to that of Lemma 6(ii) in Belloni & Chernozhukov (2011). However, there are some differences since we include the intercept term but do not penalize it.

By the convexity of $\widehat{Q}_\tau(\cdot)$, we have $\mathbb{E}_n \left\{\mathbf{Z}_i a_i^*(\tau)\right\} \in \partial \widehat{Q}_\tau \left(\boldsymbol{\beta}(\tau)\right)$, where $a_i^*(\tau) = \tau - \mathbb{1}\left(Y_i \leq \mathbf{Z}^T \boldsymbol{\beta}(\tau)\right)$ are the rank scores. Therefore,

$$\widehat{Q}_\tau(\widehat{\boldsymbol{\beta}}(\tau)) \geq \widehat{Q}_\tau(\boldsymbol{\beta}(\tau)) + \mathbb{E}_n \left(\mathbf{Z}_i a_i^*(\tau)\right)^T \left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right).$$

Denote $\hat{D} = \operatorname{diag}\left[\widehat{\sigma}_0, \ldots, \widehat{\sigma}_p\right]$. Note that $\lambda\sqrt{\tau(1-\tau)}(c_0 - 3)/(c_0+3) \geq n\|\hat{D}^{-1}\mathbb{E}_n \left(\mathbf{Z}_i a_i^*(\tau)\right)\|_\infty$ holds with probability at least $1-\alpha$. By the definition of $\widehat{\boldsymbol{\beta}}(\tau)$, we have

$$\begin{aligned}
0 &\leq \widehat{Q}_\tau(\boldsymbol{\beta}(\tau)) - \widehat{Q}_\tau(\widehat{\boldsymbol{\beta}}(\tau)) + \frac{\lambda\sqrt{\tau(1-\tau)}}{n} \sum_{j=1}^p \widehat{\sigma}_j \left(|\beta_j(\tau)| - |\widehat{\beta}_j(\tau)|\right) \\
&\leq \left|\mathbb{E}_n \left(\mathbf{Z}_i a_i^*(\tau)\right)^T \left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right)\right| + \frac{\lambda\sqrt{\tau(1-\tau)}}{n} \sum_{j=1}^p \widehat{\sigma}_j \left(|\beta_j(\tau)| - |\widehat{\beta}_j(\tau)|\right) \\
&\leq \|\hat{D}^{-1}\mathbb{E}_n \left(\mathbf{Z}_i a_i^*(\tau)\right)\|_\infty \|\hat{D}\left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right)\|_1 + \frac{\lambda\sqrt{\tau(1-\tau)}}{n} \sum_{j=1}^p \widehat{\sigma}_j \left(|\beta_j(\tau)| - |\widehat{\beta}_j(\tau)|\right).
\end{aligned}$$

Thus, we have

$$\left(1-\frac{c_0-3}{c_0+3}\right)\sum_{j=0}^{p}\widehat{\sigma}_j|\beta_j(\tau)-\widehat{\beta}_j(\tau)|\leq\sum_{j=0}^{p}\widehat{\sigma}_j|\beta_j(\tau)-\widehat{\beta}_j(\tau)|+\sum_{j=1}^{p}\widehat{\sigma}_j\left(|\beta_j(\tau)|-|\widehat{\beta}_j(\tau)|\right).$$

Since $\sum_{j=1}^{p}\widehat{\sigma}_j\left(|\beta_j(\tau)-\widehat{\beta}_j(\tau)|+|\beta_j(\tau)|-|\widehat{\beta}_j(\tau)|\right)\leq 2\sum_{j\in T_\tau\backslash\{0\}}\widehat{\sigma}_j|\beta_j(\tau)-\widehat{\beta}_j(\tau)|$, we have

$$\left(1-\frac{c_0-3}{c_0+3}\right)\sum_{j=0}^{p}\widehat{\sigma}_j|\beta_j(\tau)-\widehat{\beta}_j(\tau)|\leq 2\sum_{j\in T_\tau}\widehat{\sigma}_j|\beta_j(\tau)-\widehat{\beta}_j(\tau)|,$$

which implies that $\|\widehat{\boldsymbol{\beta}}_{T_\tau^c}(\tau)\|_{1,n}\leq(c_0/3)\|\widehat{\boldsymbol{\beta}}_{T_\tau}(\tau)-\boldsymbol{\beta}_{T_\tau}(\tau)\|_{1,n}$.

Finally, by part (i) of this Lemma, the result follows. □

Before presenting Lemmas 4 and 5, we first introduce a foundational lemma on which they are built.

Lemma 3. *Assume Condition C5, we have*

$$1-F_Y\left(y\mid\mathbf{Z}\right)=\left(\frac{y}{c^*}\right)^{-\frac{1}{\gamma}}\left\{1+\frac{d^*\left(\mathbf{Z}\right)}{\rho^*}\left(\frac{y}{c^*}\right)^{\frac{\rho^*}{\gamma}}\left(1+o\left(1\right)\right)\right\},$$

$$U_Y\left(t\mid\mathbf{Z}\right)=c^*t^\gamma\left\{1+\frac{\gamma d^*\left(\mathbf{Z}\right)}{\rho^*}t^{\rho^*}\left(1+o\left(1\right)\right)\right\},$$

and $f_Y(y\mid\mathbf{Z})=c_0K(\mathbf{Z})y^{-1/\gamma-1}(1+o(1))$, *uniformly for* $\mathbf{Z}\in\mathcal{Z}$ *as* $t,y\to\infty$, *where* $c^*=cK\left(\mathbf{Z}\right)^\gamma$, $\rho^*=\max(\rho,-\delta)$,

$$
d^*(\mathbf{Z}) = \begin{cases} dK(\mathbf{Z})^{\rho}, & \text{if } \rho > -\delta, \\ \\ -(\delta/\rho)dK_1(\mathbf{Z})K(\mathbf{Z})^{-\delta}, & \text{if } \rho \leq -\delta, \end{cases}
$$

c, c_0 are some positive constants, and d is the same as in Condition C5 (iii). This means that $U_Y(t|\mathbf{Z})$ satisfies the second-order condition in Condition C5 (iii) indexed by (γ, ρ^*, d^*) where d^* plays the same role as d in A_1.

Proof. By Condition C5 (iii), we have $U_0(t) = ct^{\gamma}\{1 + A_1(t)/\rho(1 + o(1))\}$, $1-F_0(x) = (x/c)^{-1/\gamma}\left\{1 + (d/\rho)(x/c)^{\rho/\gamma}(1 + o(1))\right\}$ and $f_0(x) = c_0 x^{-1/\gamma-1}(1+o(1))$, as $t, x \to \infty$, where $A_1(t) = \gamma d t^{\rho}$ and $c, c_0 > 0$ are some constants.

By Condition C5 (ii), we have that, with notation $U = Y - \mathbf{Z}^T\boldsymbol{\theta}_r$,

$$
\begin{aligned}
1 - F_U(u \mid \mathbf{Z}) &= K(\mathbf{Z})\big(1 - F_0(u)\big)(1 + o(1)) \\
&= K(\mathbf{Z})\left(\frac{u}{c}\right)^{-1/\gamma}\left\{1 + \frac{d}{\rho}\left(\frac{u}{c}\right)^{\rho/\gamma}(1 + o(1))\right\} \\
&\quad + K(\mathbf{Z})K_1(\mathbf{Z})\left(\frac{u}{c}\right)^{-(1+\delta)/\gamma}\left\{1 + \frac{d}{\rho}\left(\frac{u}{c}\right)^{\rho/\gamma}(1 + o(1))\right\} \\
&= \left(\frac{u}{c^*}\right)^{-1/\gamma}\left\{1 + \frac{d^*(\mathbf{Z})}{\rho^*}\left(\frac{u}{c^*}\right)^{\rho^*/\gamma}(1 + o(1))\right\},
\end{aligned}
$$

and $f_U(u \mid \mathbf{Z}) = K(\mathbf{Z})f_0(u)(1+o(1)) = c_0 K(\mathbf{Z})u^{-1/\gamma-1}(1+o(1))$, uniformly for $\mathbf{Z} \in \mathcal{Z}$, where $c^* = cK(\mathbf{Z})^{\gamma}$, $\rho^* = \max(\rho, -\delta)$, and

$$
d^*(\mathbf{Z}) = \begin{cases} dK(\mathbf{Z})^{\rho}, & \text{if } \rho > -\delta, \\ \\ -(\delta/\rho)dK_1(\mathbf{Z})K(\mathbf{Z})^{-\delta}, & \text{if } \rho \leq -\delta. \end{cases}
$$

Thus, by straight forward calculation, we have $U_U(t \mid \mathbf{Z}) = c^* t^\gamma \{1 + \gamma d^*(\mathbf{Z}) t^{\rho^*}/\rho^*(1+o(1))\}$, see Lemma 2 in Wang, Li & He (2012) for example. Therefore, we have

$$U_Y(t \mid \mathbf{Z}) = \mathbf{Z}^T\boldsymbol{\theta}_r + U_U(t \mid \mathbf{Z}) = c^* t^\gamma \left\{1 + (\gamma d^*(\mathbf{Z})/\rho^*) t^{\rho^*}(1+o(1))\right\}$$

and $f_Y(y \mid \mathbf{Z}) = c_0 K(\mathbf{Z})(y + \mathbf{Z}^T\boldsymbol{\theta}_r)^{-1/\gamma-1}(1+o(1)) = c_0 K(\mathbf{Z}) y^{-1/\gamma-1}(1+o(1))$, uniformly for $\mathbf{Z} \in \mathcal{Z}$. Hence, uniformly for $\mathbf{Z} \in \mathcal{Z}$, we have

$$1 - F_Y(y \mid \mathbf{Z}) = (y/c^*)^{-1/\gamma}\left\{1 + (d^*(\mathbf{Z})/\rho^*)(y/c^*)^{\rho^*/\gamma}(1+o(1))\right\}.$$

□

Lemma 4. *Assume the condition of Lemma 3. If* $t_n \to \infty$, $d_n^{-\delta} d_{1n} t_n^{-\delta} \to 0$, *and* $\mathbf{Z}^T\boldsymbol{\theta}_r = o\left(\left(K(\mathbf{Z})^{\gamma+\rho} t_n^{\gamma+\rho}\right) \vee \left(K(\mathbf{Z})^{\gamma-\delta} K_1(\mathbf{Z}) t_n^{\gamma-\delta}\right)\right)$, *we have*

$$U_Y(t_n \mid \mathbf{Z}) = \mathbf{Z}^T\boldsymbol{\theta}_r + U_U(t_n \mid \mathbf{Z}) = c^* t_n^\gamma \left\{1 + \frac{\gamma d_1^*(\mathbf{Z})}{\rho_1^*} t_n^{\rho_1^*}(1+o(1))\right\},$$

where

$$\rho_1^* = \begin{cases} \rho, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} \to \infty, \\ -\delta, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} = O(1), \end{cases}$$

$$d_1^*(\mathbf{Z}) = \begin{cases} dK(\mathbf{Z})^\rho, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} \to \infty, \\ -\delta K_1(\mathbf{Z}) K(\mathbf{Z})^{-\delta}, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} = O(1). \end{cases}$$

Proof. Consider a sequence $u_n \to \infty$ such that $d_n u_n^{-1/\gamma} \to 0$ and $d_{1n} u_n^{-\delta/\gamma} \to 0$. Notice that $1-F_0(u_n) \sim u_n^{-1/\gamma}$ and $f_0(u_n) \sim u_n^{-\gamma-1}$, thus $d_n(1-F_0(u_n)) \to 0$ and $d_n f_0(u_n) \to 0$.

By Condition C5 (ii), we have

$$\begin{aligned} 1 - F_U(u_n \mid \mathbf{Z}) &= K(\mathbf{Z}) \left(\frac{u_n}{c}\right)^{-1/\gamma} \left\{1 + \frac{d}{\rho}\left(\frac{u_n}{c}\right)^{\rho/\gamma}(1+o(1))\right\} \\ &\quad + K(\mathbf{Z}) K_1(\mathbf{Z}) \left(\frac{u_n}{c}\right)^{-(1+\delta)/\gamma} \left\{1 + \frac{d}{\rho}\left(\frac{u_n}{c}\right)^{\rho/\gamma}(1+o(1))\right\}^{1+\delta} \\ &= \left(\frac{u_n}{c^*}\right)^{-1/\gamma} \left\{1 + \frac{d_2^*(\mathbf{Z})}{\rho_2^*}\left(\frac{u_n}{c^*}\right)^{\rho_2^*/\gamma}(1+o(1))\right\}, \end{aligned}$$

and $f_U(u_n \mid \mathbf{Z}) = c_0 K(\mathbf{Z}) u_n^{-1/\gamma-1}(1+o(1))$, uniformly for $\mathbf{Z} \in \mathcal{Z}$, where

$$\rho_2^* = \begin{cases} \rho, & \text{if } u_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} \to \infty, \\ -\delta, & \text{if } u_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} = O(1), \end{cases}$$

$$d_2^*(\mathbf{Z}) = \begin{cases} dK(\mathbf{Z})^{\rho}, & \text{if } u_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} \to \infty, \\ -\delta K_1(\mathbf{Z}) K(\mathbf{Z})^{-\delta}, & \text{if } u_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} = O(1). \end{cases}$$

Let $t_n = (1-F_U(u_n \mid \mathbf{Z}))^{-1}$. Obviously, $t_n \to \infty$ and $t_n^{-1} \sim K(\mathbf{Z}) u_n^{-1/\gamma}$. By $d_{1n} u_n^{-\delta/\gamma} \to 0$, we have $d_n^{-\delta} d_{1n} t_n^{-\delta} \to 0$ and thus $U_U(t_n \mid \mathbf{Z}) = c^* t_n^{\gamma}\{1 + \gamma d_1^*(\mathbf{Z}) t_n^{\rho_1^*}/\rho_1^*(1+o(1))\}$, where

$$\rho_1^* = \begin{cases} \rho, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} \to \infty, \\ -\delta, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} = O(1), \end{cases}$$

$$d_1^*(\mathbf{Z}) = \begin{cases} dK(\mathbf{Z})^{\rho}, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} \to \infty, \\ \\ -\delta K_1(\mathbf{Z}) K(\mathbf{Z})^{-\delta}, & \text{if } (d_n t_n)^{\rho+\delta} d_{1n}^{-1} = O(1). \end{cases}$$

Since $\mathbf{Z}^T\boldsymbol{\theta}_r = o\left(\left(K\left(\mathbf{Z}\right)^{\gamma+\rho} t_n^{\gamma+\rho}\right) \vee \left(K\left(\mathbf{Z}\right)^{\gamma-\delta} K_1\left(\mathbf{Z}\right) t_n^{\gamma-\delta}\right)\right)$, we have

$$U_Y\left(t_n \mid \mathbf{Z}\right) = \mathbf{Z}^T\boldsymbol{\theta}_r + U_U\left(t_n \mid \mathbf{Z}\right) = c^* t_n^{\gamma} \left\{1 + \frac{\gamma d_1^*\left(\mathbf{Z}\right)}{\rho_1^*} t_n^{\rho_1^*}\left(1 + o\left(1\right)\right)\right\}.$$

Thus, the statement is proved. □

Lemma 5. *Assume the condition of Lemma 3. If* $d_n y_n^{-\frac{1}{\gamma}} \to 0$, $d_{1n} y_n^{-\delta/\gamma} \to 0$, *and* $\mathbf{Z}^T\boldsymbol{\theta}_r = o\left(\left(y_n^{(\gamma+\rho)/\gamma}\right) \vee \left(K_1\left(\mathbf{Z}\right) y_n^{(\gamma-\delta)/\gamma}\right)\right)$, *we have*

$$1 - F_Y\left(y_n \mid \mathbf{Z}\right) = \left(\frac{y_n}{c^*}\right)^{-1/\gamma} \left\{1 + \frac{d_2^*\left(\mathbf{Z}\right)}{\rho_2^*}\left(\frac{y_n}{c^*}\right)^{\rho_2^*/\gamma}\left(1 + o\left(1\right)\right)\right\}$$

and $f_Y(y_n \mid \mathbf{Z}) = c_0 K(\mathbf{Z}) y_n^{-1/\gamma-1}(1 + o(1))$, *where*

$$\rho_2^* = \begin{cases} \rho, & \textit{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} \to \infty, \\ \\ -\delta, & \textit{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} = O(1), \end{cases}$$

$$d_2^*\left(\mathbf{Z}\right) = \begin{cases} dK(\mathbf{Z})^{\rho}, & \textit{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} \to \infty, \\ \\ -\delta K_1(\mathbf{Z}) K(\mathbf{Z})^{-\delta}, & \textit{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} = O(1). \end{cases}$$

Moreover, if we only need the first-order expansion, the conditions $\mathbf{Z}^T\boldsymbol{\theta}_r =$

$o\left(\left(K(\mathbf{Z})^{\gamma+\rho} t_n^{\gamma+\rho}\right) \vee \left(K(\mathbf{Z})^{\gamma-\delta} K_1(\mathbf{Z}) t_n^{\gamma-\delta}\right)\right)$ *and* $\mathbf{Z}^T\boldsymbol{\theta}_r =$ $o\left(\left(y_n^{(\gamma+\rho)/\gamma}\right) \vee \left(K_1(\mathbf{Z}) y_n^{(\gamma-\delta)/\gamma}\right)\right)$ *can be replaced by weaker conditions* $\mathbf{Z}^T\boldsymbol{\theta}_r =$ $o(K(\mathbf{Z})^\gamma t_n^\gamma)$ *and* $\mathbf{Z}^T\boldsymbol{\theta}_r = o(y_n)$, *respectively.*

Proof. Similar to proof of Lemma 4, if $\mathbf{Z}^T\boldsymbol{\theta}_r = o\left(\left(y_n^{(\gamma+\rho)/\gamma}\right) \vee \left(K_1(\mathbf{Z}) y_n^{(\gamma-\delta)/\gamma}\right)\right)$,

$$1 - F_Y(y_n \mid \mathbf{Z}) = \left(\frac{y_n}{c^*}\right)^{-1/\gamma} \left\{1 + \frac{d_2^*(\mathbf{X})}{\rho_2^*} \left(\frac{y_n}{c^*}\right)^{\rho_2^*/\gamma} (1 + o(1))\right\}$$

and $f_Y(y_n \mid \mathbf{Z}) = c_0 K(\mathbf{Z}) y_n^{-1/\gamma-1}(1 + o(1))$, uniformly for $\mathbf{Z} \in \mathcal{Z}$, where

$$\rho_2^* = \begin{cases} \rho, & \text{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} \to \infty, \\ -\delta, & \text{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} = O(1), \end{cases}$$

$$d_2^*(\mathbf{Z}) = \begin{cases} dK(\mathbf{Z})^\rho, & \text{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} \to \infty, \\ -\delta K_1(\mathbf{Z}) K(\mathbf{Z})^{-\delta}, & \text{if } y_n^{(\rho+\delta)/\gamma} d_{1n}^{-1} = O(1), \end{cases}$$

which completes the proof. □

Remark S.5. *Under Condition C5, by Lemma 3, we have* $\mathbf{Z}^T\boldsymbol{\beta}(\tau_{0n}) \sim c^*(1-\tau)^{-\gamma} \sim c d_n^\gamma (1-\tau)^{-\gamma}$. *Therefore, with the additional conditions* $t_n \to \infty$, $d_n t_n^{\gamma-1} \to \infty$, *and* $d_n^{-\delta} d_{1n} t_n^{-\delta} \to 0$, *we have* $f_Y(\mathbf{Z}^T\boldsymbol{\beta}(\tau_{0n}) \mid \mathbf{X}) \sim d_n^{-\gamma}(1 - \tau_{0n})^{\gamma+1}$.

Lemma 6. *Assume Condition C5,* $s_r d_n^{-\gamma}(1-\tau_{0n})^\gamma \to 0$ *and* $d_n^{-\delta} d_{1n} (1-\tau_{0n})^\delta \to 0$. *Then, for any* $C_S > (2c\gamma)^{-1} c_2^{\gamma+1} > 0$ *and* $C_T > 0$, *there exists some pos-*

itive integer N such that, for $n > N$,

$$\mathbb{E}\left\{M_n\left(\mathbf{u},\tau\right)\right\} \geq C_S d_n^{-\gamma}\left(1-\tau_{0n}\right)^{\gamma+1}\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\} - C_T\left(1-\tau_{0n}\right)\sqrt{\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\}},$$

uniformly for any $\mathbf{u}$ such that $\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\| \leq b_n$ and for $1-\tau \in \mathcal{T}_n$, where $b_n = o(d_n^{\gamma}\left(1-\tau_{0n}\right)^{-\gamma})$ and c is the same as in Lemma 3.

Proof. We first derive the lower and upper bounds for $1 - F_Y(\mathbf{Z}_i^T\boldsymbol{\beta}(\tau) \mid \mathbf{Z})$ and $1 - F_Y(\mathbf{Z}_i^T\boldsymbol{\beta}(\tau) + t \mid \mathbf{Z})$ for any $1-\tau \in \mathcal{T}_n$.

Given that $d_n^{-\delta} d_{1n}\left(1-\tau_{0n}\right)^{\delta} \to 0$ and $\mathbf{Z}^T\boldsymbol{\theta}_r = o(K(\mathbf{Z})^{\gamma}(1-\tau_{0n})^{-\gamma})$ and by Condition C5, we apply Lemmas 3 and 5 and have that, for any constants $\delta_1, \delta_2, \delta_3, \delta_4 \in (0,1)$, there exists a positive integer N_1 such that, for all $n > N_1$,

$$(1-\tau)(1-\delta_1) < 1 - F_Y(\mathbf{Z}_i^T\boldsymbol{\beta}(\tau) \mid \mathbf{Z}) < (1-\tau)(1+\delta_2),$$

$$(1-\tau)\left(1 + \frac{t}{cK(\mathbf{Z})^{\gamma}(1-\tau)^{-\gamma}}\right)^{-1/\gamma}(1-\delta_3) < 1 - F_Y(\mathbf{Z}_i^T\boldsymbol{\beta}(\tau) + t \mid \mathbf{Z}),$$

and

$$1 - F_Y(\mathbf{Z}_i^T\boldsymbol{\beta}(\tau) + t \mid \mathbf{Z}) < (1-\tau)\left(1 + \frac{t}{cK(\mathbf{Z})^{\gamma}(1-\tau)^{-\gamma}}\right)^{-1/\gamma}(1+\delta_4)$$

hold uniformly for all $1-\tau \in \mathcal{T}_n$ and for $t \in (0, b_n)$, where $c > 0$ is defined in Lemma 3.

Specifically, by Lemma 3, we have $\mathbf{Z}_i^T\boldsymbol{\beta}(\tau) + t = c^*(1-\tau)^{-\gamma}(1+o(1))$, uniformly for $t \in (0, b_n)$. Therefore, by Lemma 5, we have

$$\frac{1 - F_Y(\mathbf{Z}_i^T \boldsymbol{\beta}(\tau) + t \mid \mathbf{Z})}{(1-\tau)\left(1 + tK(\mathbf{Z})^{-\gamma}(1-\tau)^{\gamma}/c\right)^{-1/\gamma}} = 1 + \frac{d^*(\mathbf{Z})}{\rho^*}(1-\tau)^{-\rho^*}(1+tK(\mathbf{Z})^{-\gamma}(1-\tau)^{\gamma}/c)^{\rho^*/\gamma}(1+o(1)),$$

where $o(1)$ holds uniformly for $\mathbf{Z} \in \mathcal{Z}$. Since $d^*(\mathbf{Z})(1-\tau)^{-\rho^*}(1+tK(\mathbf{Z})^{-\gamma}(1-\tau)^{\gamma}/c)^{\rho^*/\gamma}/\rho^* \leq d^*(\mathbf{Z})c_2(1-\tau_{0n})/\rho^*$ uniformly for all $1-\tau \in \mathcal{T}_n$ and for $t \in (0, b_n)$, we have

$$\frac{1 - F_Y(\mathbf{Z}_i^T \boldsymbol{\beta}(\tau) + t \mid \mathbf{Z})}{(1-\tau)\left(1 + tK(\mathbf{Z})^{-\gamma}(1-\tau)^{\gamma}/c\right)^{-1/\gamma}} = 1 + o(1),$$

uniformly for all $1-\tau \in \mathcal{T}_n$, $t \in (0, b_n)$ and $\mathbf{Z} \in \mathcal{Z}$.

By the definition of $\mathbb{E}\{M_n(\mathbf{u}, \tau)\}$, we have

$$\begin{aligned}
&\mathbb{E}\{M_n(\mathbf{u}, \tau)\} \\
&= \mathbb{E}\left\{\rho_\tau(Y_i - \mathbf{Z}_i^T\boldsymbol{\beta}(\tau) - \mathbf{Z}_i^T\mathbf{u})\right\} - \mathbb{E}\left\{\rho_\tau(Y_i - \mathbf{Z}_i^T\boldsymbol{\beta}(\tau))\right\} \\
&= \mathbb{E}\left[\int_0^{\mathbf{Z}^T\mathbf{u}} \left\{F_Y(\mathbf{Z}^T\boldsymbol{\beta}(\tau) + t \mid \mathbf{Z}) - F_Y(\mathbf{Z}^T\boldsymbol{\beta}(\tau) \mid \mathbf{Z})\right\} dt\right] \\
&\geq \mathbb{E}\left[\int_0^{\mathbf{Z}^T\mathbf{u}} \left((1-\tau)(1-\delta_1) - (1-\tau)\left(1 + \frac{t}{cK(\mathbf{z})^{\gamma}(1-\tau)^{-\gamma}}\right)^{-1/\gamma}(1+\delta_4)\right) dt\right].
\end{aligned}$$

Since $\mathbf{Z}^T\mathbf{u} < b_n$ and $b_n = o(d_n^{\gamma}(1-\tau_{0n})^{-\gamma})$, we have, for any $t < \mathbf{Z}^T\mathbf{u}$ and $1-\tau \in \mathcal{T}_n$, $t\left\{cK(\mathbf{Z})^{\gamma}(1-\tau)^{-\gamma}\right\}^{-1} \to 0$.

Therefore, for any constant $\delta_5 \in (0,1)$, there exists a positive integer N_2 such that, for all $n > N_2$,

$$
\begin{aligned}
&\mathbb{E}\{M_n(\mathbf{u},\tau)\} \\
&\geq \mathbb{E}\left[\int_0^{\mathbf{Z}^T\mathbf{u}}\left\{(1-\tau)(1-\delta_1)-(1-\tau)\left(1-\frac{t}{c\gamma K(\mathbf{Z})^{\gamma}(1-\tau)^{-\gamma}}(1+\delta_5)\right)(1+\delta_4)\right\}dt\right] \\
&= \frac{(1+\delta_5)(1+\delta_4)}{2c\gamma(1-\tau)^{-\gamma-1}}\mathbf{u}^T\mathbb{E}(K(\mathbf{Z})^{-\gamma}\mathbf{Z}\mathbf{Z}^T)\mathbf{u}-(1-\tau)(\delta_4+\delta_5)\sqrt{\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\}} \\
&\geq \frac{(1+\delta_5)(1+\delta_4)}{2c\gamma d_n^{\gamma}(1-\tau)^{-\gamma-1}}\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\}-(1-\tau)(\delta_1+\delta_4)\sqrt{\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\}}.
\end{aligned}
$$

Thus, for any $\delta_1,\delta_4,\delta_5\in(0,1)$, there exists $N=\max(N_1,N_2)$ such that, for all $n>N$,

$$
\begin{aligned}
\mathbb{E}\{M_n(\mathbf{u},\tau)\} \geq{}& \frac{(1+\delta_5)(1+\delta_4)}{2cc_2^{-\gamma-1}\gamma d_n^{\gamma}(1-\tau_{0n})^{-\gamma-1}}\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\} \\
&-c_1(1-\tau_{0n})(\delta_1+\delta_4)\sqrt{\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\}},
\end{aligned}
$$

holds uniformly for any $\mathbf{u}$ such that $\|\tilde{\Sigma}^{1/2}\mathbf{u}\|\leq b_n$, and for $1-\tau\in\mathcal{T}_n$. Therefore, the lemma follows. □

Lemma 7. *For any $t>0$, let*

$$
\epsilon(t):=\sup_{1-\tau\in\mathcal{T}_n,\mathbf{u}\in A_\tau,\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\|\leq t}\left|M_n(\mathbf{u},\tau)-\mathbb{E}\{M_n(\mathbf{u},\tau)\}\right|.
$$

Under Conditions C1, C2, C3 and C4, there exists a universal constant $C_E>0$ such that for any $A>1$ and $\epsilon\in(0,1)$, there exists positive integer N_ϵ such that, for all $n>N_\epsilon$, with probability at least $1-3\epsilon-3p^{-A^2}$,

$$
\epsilon(t)\leq\frac{t}{\sqrt{n}}C_EA\frac{c_0+1}{\kappa_0}\{2+(1-c_1(1-\tau_{0n}))\}\sqrt{s}\sqrt{\log\left\{p\vee\left(\frac{1}{t}\frac{L\kappa_0}{c_0+1}\frac{1}{(1-\tau_{0n})\sqrt{s}}\beta^{(1)}\right)\right\}}, \tag{S1.2}
$$

where $\beta^{(1)} = \max_{1\le j\le p}\left\{|\beta_j^{'}\left(1-c_2\left(1-\tau_{0n}\right)\right)|\right\},\ \beta_j^{'}\left(\tau\right) = \partial\beta_j(\tau)/\partial\tau,\ L = c_2 - c_1 > 0.$

Proof. Let

$$\begin{aligned}\mathcal{A}\left(t\right) &= \sup_{1-\tau\in\mathcal{T}_n,\mathbf{u}\in A_\tau,\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\|\le t}\sqrt{n}\epsilon\left(t\right)\\ &= \sup_{1-\tau\in\mathcal{T}_n,\mathbf{u}\in A_\tau,\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\|\le t}\mathbb{G}_n\left\{\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)-\mathbf{Z}_i^\top\mathbf{u}\right)-\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)\right)\right\}.\end{aligned}$$

Notice that, for $\mathbf{u}$ satisfying $\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\|\le t$, it follows that

$$\mathrm{Var}\left\{\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)-\mathbf{Z}_i^\top\mathbf{u}\right)-\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)\right)\right\}\le\mathbb{E}\left\{(\mathbf{Z}_i^T\mathbf{u})^2\right\}\le t^2,$$

by the convexity of ρ_τ. Then by Lemma 2.3.7 in van der Vaart & Wellner (1996), we have that, for every $x > 0$,

$$\begin{aligned}P\left(\mathcal{A}\left(t\right)>x\right) &\le \frac{2}{1-\frac{4n}{x^2}\frac{t^2}{n}}P\left(\mathcal{A}^o\left(t\right)>\frac{1}{4}x\right)\\ &= \frac{2}{1-\frac{4t^2}{x^2}}\left\{P\left(\mathcal{A}^o\left(t\right)>\frac{1}{4}x\mid\Omega_1\right)+P\left(\Omega_1^C\right)\right\}.\end{aligned}$$

where

$$\mathcal{A}^o\left(t\right) = \sup_{1-\tau\in\mathcal{T}_n,\mathbf{u}\in A_\tau,\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\|\le t}\mathbb{G}_n^o\left\{\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)-\mathbf{Z}_i^\top\mathbf{u}\right)-\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)\right)\right\}$$

and Ω_1 is the event that $\frac{2}{3}\|\mathbf{u}\|_{1,n}\le\|\mathbf{u}\|_1\le 2\|\mathbf{u}\|_{1,n}$ holds for every $\mathbf{u}\in\mathbb{R}^{p+1}$.

By Lemma 2, for any $\epsilon \in (0,1)$, there exists positive integer N_ϵ such that, for all $n > N_\epsilon$, $P\left(\Omega_1^c\right) \leq \epsilon$. Notice that $\rho_\tau\left(y_i - \mathbf{Z}_i^T(\boldsymbol{\beta}(\tau)+\mathbf{u})\right) - \rho_\tau\left(y_i - \mathbf{Z}_i^T\boldsymbol{\beta}(\tau)\right) = \tau\mathbf{Z}_i^T\mathbf{u} + w_i\left(\mathbf{Z}_i^T\mathbf{u},\tau\right)$, where $w_i(b,\tau) = \left(y_i - \mathbf{Z}_i^T\boldsymbol{\beta}(\tau) - b\right)_- - \left(y_i - \mathbf{Z}_i^T\boldsymbol{\beta}(\tau)\right)_-$. Additionally, for $\mathbf{u} \in A_\tau$, $\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\| \leq t$, we have $\|\mathbf{u}\|_1 = \|\mathbf{u}_{\bar{T}_\tau}\|_1 + \|\mathbf{u}_{T_\tau}\|_1 \leq (c_0+1)\|\mathbf{u}_{T_\tau}\|_1 \leq (c_0+1)\sqrt{s}\|\mathbf{u}_{T_\tau}\|_2 \leq ((c_0+1)/\kappa_0)\sqrt{s}\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\| \leq at$, where $a = (c_0+1){\kappa_0}^{-1}\sqrt{s}$. Therefore, we have $\mathcal{A}^o(t) \leq \mathcal{B}^o(t) + \mathcal{C}^o(t)$, where

$$\mathcal{B}^o(t) = \sup_{1-\tau\in\mathcal{T}_n, \mathbf{u}\in A_\tau, \|\mathbf{u}\|_1\leq at} \left|\mathbb{G}_n^o\left(\tau\mathbf{Z}_i^T\mathbf{u}\right)\right|,$$

$$\mathcal{C}^o(t) = \sup_{1-\tau\in\mathcal{T}_n, \mathbf{u}\in A_\tau, \|\mathbf{u}\|_1\leq at} \left|\mathbb{G}_n^o\left\{w_i\left(\mathbf{Z}_i^T\mathbf{u},\tau\right)\right\}\right|.$$

For any $M > 0$,

$$\begin{aligned} P\left(\mathcal{B}^o(t) > M \mid \Omega_1\right) &\leq \min_{\lambda\geq 0}\left\{e^{-\lambda M}\mathbb{E}\left(e^{\lambda\mathcal{B}^o(t)} \mid \Omega_1\right)\right\} \\ &\leq \min_{\lambda\geq 0}\left\{e^{-\lambda M}2p\exp\left(2\lambda^2a^2t^2(1-c_1(1-\tau_{0n}))^2\right)\right\} \\ &= 2p\exp\left(-\frac{M^2}{8a^2t^2\left(1-c_1\left(1-\tau_{0n}\right)\right)^2}\right) \end{aligned}$$

and

$$\begin{aligned} P\left(\mathcal{C}^o(t) > M \mid \Omega_1\right) &\leq \min_{\lambda\geq 0}\left\{e^{-\lambda M}\mathbb{E}\left(e^{\lambda\mathcal{C}^o(t)} \mid \Omega_1\right)\right\} \\ &\leq \min_{\lambda\geq 0}\left\{e^{-\lambda M}2pL\delta^{-1}\exp\left(8\lambda^2a^2t^2\right)\right\} \\ &= \frac{2pL}{\delta}\exp\left(-\frac{M^2}{32a^2t^2}\right), \end{aligned}$$

where δ (see below) satisfies $s\delta\left(1-\tau_{0n}\right)\beta^{(1)} \leq at$. Hence,

$$
\begin{aligned}
& P\left\{\mathcal{A}^{o}\left(t\right) > (2+(1-c_1(1-\tau_{0n})))M \mid \Omega_1\right\} \\
\leq\ & P\left\{\mathcal{B}^{o}\left(t\right) > \left(1-c_1(1-\tau_{0n})\right)M \mid \Omega_1\right\} + P\left(\mathcal{C}^{o}\left(t\right) > 2M \mid \Omega_1\right) \\
\leq\ & 2p\left(1+L\delta^{-1}\right)\exp\left(-\frac{M^2}{8a^2t^2}\right).
\end{aligned}
$$

Recall that

$$
P\left(\mathcal{A}\left(t\right) > M\right) \leq \frac{2}{1-\frac{4t^2}{M^2}}\left\{P\left(\mathcal{A}^{o}\left(t\right) > \frac{1}{4}M \mid \Omega_1\right) + P\left(\Omega_1^{C}\right)\right\}.
$$

Let $M = (4M_1) \vee M_2$, where $M_2 = A\left\{2+\left(1-c_1\left(1-\tau_{0n}\right)\right)\right\}at\sqrt{\log\left\{2p^2\left(1+L\delta^{-1}\right)\right\}}$ for any $A>1$, and $M_1 = 2\sqrt{3}t$. Then, for all $n > N_\epsilon$,

$$
P\left(\mathcal{A}\left(t\right) > M\right) \leq 3\left\{P\left(\mathcal{A}^{o}\left(t\right) > \frac{1}{4}M \mid \Omega_1\right) + P\left(\Omega_1^{C}\right)\right\} \leq 3p^{-A^2} + 3\epsilon.
$$

Therefore, there exists a universal positive constant C_E such that

$$
\epsilon(t) \leq \frac{t}{\sqrt{n}}C_E A\frac{c_0+1}{\kappa_0}\left(2+\left(1-c_1\left(1-\tau_{0n}\right)\right)\right)\sqrt{s}\sqrt{\log\left\{p \vee \left(\frac{1}{t}\frac{L\kappa_0}{c_0+1}\frac{1}{\left(1-\tau_{0n}\right)\sqrt{s}}\beta^{(1)}\right)\right\}},
$$

and thus the statement follows.

For $\mathcal{B}^{o}(t)$, we have

$$
\begin{aligned}
\mathbb{E}\left[e^{\lambda \mathcal{B}^o(t)} \mid \Omega_1\right] &= \mathbb{E}\left(e^{\lambda \sup_{\tau,\mathbf{u}}\left|\mathbb{G}_n^o\left(\tau \mathbf{Z}_i^T \mathbf{u}\right)\right|} \mid \Omega_1\right) \\
&\leq \mathbb{E}\left(e^{\lambda \sup_{\tau}\left|2at\tau \mathbb{G}_n^o\left(\max_{1\leq j\leq p} Z_{ij}/\widehat{\sigma}_j\right)\right|} \mid \Omega_1\right) \\
&\leq \mathbb{E}\left(e^{\lambda\left|2at(1-c_1(1-\tau_{0n}))\mathbb{G}_n^o\left(\max_{1\leq j\leq p} Z_{ij}/\widehat{\sigma}_j\right)\right|} \mid \Omega_1\right) \\
&\leq 2p \max_{1\leq j\leq p} \mathbb{E}\left(e^{\lambda|2at(1-c_1(1-\tau_{0n}))\mathbb{G}_n^o(Z_{ij}/\widehat{\sigma}_j)|} \mid \Omega_1\right) \\
&\leq 2p \max_{1\leq j\leq p} \mathbb{E}\left(e^{2\lambda^2 a^2 t^2 (1-c_1(1-\tau_{0n}))^2} \mid \Omega_1\right) \\
&= 2p \exp\left(2\lambda^2 a^2 t^2 (1-c_1\left(1-\tau_{0n}\right))^2\right),
\end{aligned}
$$

where the first inequality is by Hölder's inequality, the third and forth inequalities are due to the following facts.

For any symmetric random variable $Z_j, j = 1, 2, \ldots, p$, we have

$$
\mathbb{E}\left(\max_{1\leq j\leq p} e^{|Z_j|}\right) \leq p \max_{1\leq j\leq p} \mathbb{E}\left(e^{|Z_j|}\right) \leq p \max_{1\leq j\leq p} \mathbb{E}\left(e^{Z_j} + e^{-Z_j}\right) \leq 2p \max_{1\leq j\leq p} \mathbb{E}\left(e^{Z_j}\right)
$$

and $\mathbb{E}\left\{\exp\left(2\lambda t \Gamma \mathbb{G}_n^o\left(X_{ij}\right)/\widehat{\sigma}_j\right) \mid \Omega_1, \mathbb{X}\right\} \leq \exp\left((2\lambda t\Gamma)^2/2\right)$ by the proof of Lemma 2.2.7 in van der Vaart & Wellner (1996).

For $\mathcal{C}^o(t)$, we first construct a $\delta-$net for interval $[c_1, c_2]$: $\mathcal{C}_K = \{\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_K\}$, where $\delta \leq ats^{-1}\left(1-\tau_{0n}\right)^{-1}(\beta^{(1)})^{-1}$ and $K\delta \leq L$. For each $1-\tau \in \mathcal{T}_n$, there exists a $\hat{c} \in \mathcal{C}_K$, such that $|(1-\tau)/(1-\tau_{0n}) - \hat{c}| \leq \delta$.

For any $i \in \{1, 2, \ldots, K\}$ and any c in $\hat{c}_i$'s δ neighborhood, i.e. $|c - \hat{c}_i| \leq$

δ, let $1-\tau = c\left(1-\tau_{0n}\right)$ and $1-\hat{\tau}_i = 1-\hat{c}_i\left(1-\tau_{0n}\right)$. We have

$$\|\boldsymbol{\beta}\left(\tau\right)-\boldsymbol{\beta}\left(\hat{\tau}_i\right)\|_1 \leq s\|\boldsymbol{\beta}\left(\tau\right)-\boldsymbol{\beta}\left(\hat{\tau}_i\right)\|_\infty \leq s\delta\left(1-\tau_{0n}\right)\beta^{(1)},$$

where $\beta^{(1)} = \max\left\{|\beta_j'\left(1-c_2\left(1-\tau_{0n}\right)\right)| : j=1,2,\ldots,p\right\}$ and the second inequality follows by monotonity of $f_Y\left(\cdot|\mathbf{X}\right)$ in Condition C4.

Since $s\delta\left(1-\tau_{0n}\right)\beta^{(1)} \leq at$, we have $\|\boldsymbol{\beta}\left(\tau\right)-\boldsymbol{\beta}\left(\hat{\tau}_i\right)\|_1 \leq at$ for all $1-\tau \in \mathcal{T}_n$. Therefore, we have

$$\begin{aligned}
\mathcal{C}^o(t) \leq & \sup_{|\frac{1-\tau}{1-\tau_{0n}}-\hat{c}_i|\leq\delta,\hat{c}_i\in\mathcal{C}_K} \sup_{\|\mathbf{u}\|_1\leq at} \left|\mathbb{G}_n^o\left\{w_i\left(\mathbf{Z}_i^T\left(\mathbf{u}+\boldsymbol{\beta}\left(\tau\right)-\boldsymbol{\beta}\left(\hat{\tau}_i\right)\right),\hat{\tau}_i\right)\right\}\right| \\
& + \sup_{|\frac{1-\tau}{1-\tau_{0n}}-\hat{c}_i|\leq\delta,\hat{c}_i\in\mathcal{C}_K} \left|\mathbb{G}_n^o\left\{w_i\left(\mathbf{Z}_i^T\boldsymbol{\beta}\left(\tau\right)-\mathbf{Z}_i^T\boldsymbol{\beta}\left(\hat{\tau}_i\right),\hat{\tau}_i\right)\right\}\right| \\
\leq & 2\sup_{\|\mathbf{u}\|_1\leq 2at,\hat{c}_i\in\mathcal{C}_K} \left|\mathbb{G}_n^o\left\{w_i\left(\mathbf{Z}_i^T\mathbf{u},\hat{\tau}_i\right)\right\}\right| =: \mathcal{D}^o\left(t\right).
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda\mathcal{D}^o(t)} \mid \Omega_1\right) \leq & K\max_{\hat{c}_i\in\mathcal{C}_K}\mathbb{E}\left\{\exp\left(2\lambda\sup_{\|\mathbf{u}\|_1\leq 2at}\left|\mathbb{G}_n^o\left\{w_i\left(\mathbf{Z}_i^T\mathbf{u},\hat{\tau}_i\right)\right\}\right|\right) \,\middle|\, \Omega_1\right\} \\
\leq & \frac{L}{\delta}\max_{\hat{c}_i\in\mathcal{C}_K}\mathbb{E}\left\{\exp\left(2\lambda\sup_{\|\mathbf{u}\|_1\leq 2at}\left|\mathbb{G}_n^o\left(\mathbf{Z}_i^T\mathbf{u}\right)\right|\right) \,\middle|\, \Omega_1\right\} \\
\leq & \frac{L}{\delta}\mathbb{E}\left[\exp\left\{4\lambda at\left|\mathbb{G}_n^o\left(\max_{1\leq j\leq p} Z_{ij}/\widehat{\sigma}_j\right)\right|\right\} \,\middle|\, \Omega_1\right] \\
\leq & \frac{2pL}{\delta}\max_{1\leq j\leq p}\mathbb{E}\left\{\exp\left(4\lambda at\mathbb{G}_n^o\left(Z_{ij}/\widehat{\sigma}_j\right)\right) \mid \Omega_1\right\} \\
\leq & \frac{2pL}{\delta}\exp\left(8\lambda^2a^2t^2\right),
\end{aligned}$$

it follows that $\mathbb{E}\left(e^{\lambda \mathcal{C}^o(t)} \mid \Omega_1\right) \leq (2pL/\delta) \exp\left(8\lambda^2 a^2 t^2\right)$. □

Lemma 8. *A random variable X is a sub-Gaussian if and only if X^2 is sub-exponential. Moreover, $\|X\|_{\psi_2}^2 \leq \|X^2\|_{\psi_1} \leq 2\|X\|_{\psi_2}^2$, where $\|X\|_{\psi_1}$ and $\|X\|_{\psi_2}$ are respectively sub-Gaussian and sub-exponential norm of X.*

Proof. The proof of this lemma can be found in Lemma 5.14 in Chapter 5 of Vershynin (2012). □

Lemma 9. *Let $X_1, \ldots, X_n$ be independent centered sub-exponential random variables, and $M = \max_{1\leq i\leq n} \|X_i\|_{\psi_1}$. Then for every $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and every $t \geq 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n a_i X_i \geq t\right) \leq \exp\left(-C_2 \min\left(\frac{t^2}{M^2\|a\|_2^2}, \frac{t}{M\|a\|_\infty}\right)\right).$$

Proof. The proof can be found in Proposition 5.16 in Chapter 5 of Vershynin (2012). □

Proof of Theorem 1.

Let C_F, C_S, C_T and $\tilde{C}$ be some positive constant such that $C_F > 1$, $C_S > (2c\gamma)^{-1}c_2^{\gamma+1}$, $C_T \leq 2\sqrt{A^2+1}C_{\tilde{\Lambda}}(c_0+1)/\kappa_0$ and

$$\tilde{C} > \frac{\left(2\sqrt{A^2+1}C_{\tilde{\Lambda}}\right) \vee \left(3C_E\sqrt{(1+\log C_F) \vee \left(2C_L + \frac{2}{\gamma+3}\right)}\right)}{C_S},$$

where $A > 1$ is any positive constant, and $C_{\tilde{\Lambda}}$, c and C_E are positive constants defined in Lemmas 1, 3 and 7, respectively. Denote $\lambda = \sqrt{A^2+1}C_{\tilde{\Lambda}} \sqrt{n\log p}/\sqrt{1-\tau_{0n}}$.

Consider the following three events:

(i) Ω_1: The event that (S1.1) holds and $\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \in A_\tau$ uniformly for $1-\tau \in \mathcal{T}_n$;

(ii) Ω_2: The event that (S1.2) in Lemma 7 holds;

(iii) Ω_3: The event that $\lambda \geq (c_0+3)/(c_0-3)\tilde{\Lambda}$ holds.

By Lemmas 1 and 2, there exists a positive integer N_ϵ such that, for all $n > N_\epsilon$, $P\left(\Omega_1\right) \geq P\left(\Omega_1 \cap \Omega_3\right) \geq 1 - p^{-A^2} - \epsilon$. By Lemma 7, for all $n > N_\epsilon$, $P\left(\Omega_2\right) \geq 1 - 3\epsilon - 3p^{-A^2}$. By Lemma 1, $P\left(\Omega_3\right) \geq 1 - p^{-A^2}$. Thus, for all $n > N_\epsilon$,

$$P\left(\cap_{i=1}^{3}\Omega_i\right) \geq 1 - 4\epsilon - 5p^{-A^2}. \tag{S1.3}$$

By Condition C6, there exists a positive integer N_F such that, for all $n > N_F$,

$$\beta^{(1)} \leq C_F(1-\tau_{0n})^{-\gamma-1}. \tag{S1.4}$$

By Lemma 6, there exists a positive integer N_{ST} such that, for all $n > N_{ST}$,

$$\mathbb{E}\left\{M_n\left(\mathbf{u},\tau\right)\right\} \geq C_S d_n^{-\gamma}\left(1-\tau_{0n}\right)^{\gamma+1}\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\} - C_T\left(1-\tau_{0n}\right)\sqrt{\mathbb{E}\{(\mathbf{Z}^T\mathbf{u})^2\}}, \tag{S1.5}$$

uniformly for $1-\tau \in \mathcal{T}_n$ and for any $\mathbf{u}$ such that $\left\|\tilde{\Sigma}^{1/2}\mathbf{u}\right\| \leq b_n$, where $b_n = o(d_n^{\gamma}(1-\tau_{0n})^{-\gamma})$.

Let $N = \max(N_\epsilon, N_F, N_{ST})$ and

$$t = 4\tilde{C}A\frac{c_0+1}{\kappa_0}d_n^{\gamma}\left(1-\tau_{0n}\right)^{-\gamma-1}\sqrt{\frac{s\log(p\vee n)}{n}}. \tag{S1.6}$$

We will show that, conditional on $\cap_{i=1}^3\Omega_i$, for all $n > N$, $\left\|\tilde{\Sigma}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\beta}}\left(\tau\right)-\boldsymbol{\beta}\left(\tau\right)\right)\right\| \leq t$, holds uniformly for $1-\tau \in \mathcal{T}_n$. We prove it by contradiction. Suppose that there exists $1-\tau \in \mathcal{T}_n$, such that $\left\|\tilde{\Sigma}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\beta}}\left(\tau\right)-\boldsymbol{\beta}\left(\tau\right)\right)\right\| > t$. By the definition of $\widehat{\boldsymbol{\beta}}\left(\tau\right)$, we have

$$\min_{\mathbf{u}\in A_\tau, \|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|\geq t}\Bigg[\sum_{i=1}^n\left\{\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)-\mathbf{Z}_i^\top\mathbf{u}\right)-\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0\left(\tau\right)\right)\right\}$$
$$+\lambda\sqrt{\tau(1-\tau)}\|\left(\mathbf{u}+\boldsymbol{\beta}\left(\tau\right)\right)_{[p]}\|_{1,n}-\lambda\sqrt{\tau(1-\tau)}\|\left(\boldsymbol{\beta}\left(\tau\right)\right)_{[p]}\|_{1,n}\Bigg] < 0.$$

By the convexity of the objective function, we have

$$\min_{\mathbf{u}\in A_\tau, \|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|=t}\left[\sum_{i=1}^{n}\left\{\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)-\mathbf{Z}_i^\top\mathbf{u}\right)-\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\right)\right\}\right.$$
$$\left.+\lambda\sqrt{\tau(1-\tau)}\|\left(\mathbf{u}+\boldsymbol{\beta}(\tau)\right)_{[p]}\|_{1,n}-\lambda\sqrt{\tau(1-\tau)}\|\left(\boldsymbol{\beta}(\tau)\right)_{[p]}\|_{1,n}\right]<0.$$

Since $\|\left(\mathbf{u}+\boldsymbol{\beta}(\tau)\right)_{[p]}\|_{1,n}-\|\left(\boldsymbol{\beta}(\tau)\right)_{[p]}\|_{1,n}\geq-\|\mathbf{u}\|_{1,n}\geq-2\|\mathbf{u}\|_1\geq-2\left(c_0+1\right)\|\mathbf{u}_{T_\tau}\|_1\geq -\frac{2\sqrt{s}(c_0+1)}{\kappa_0}\|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|_2$, for some $c_0>0$ and $\kappa_0>0$ and $\mathbf{u}\in\Omega_1$, we have

$$\min_{\mathbf{u}\in A_\tau, \|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|=t}\left[\sum_{i=1}^{n}\left\{\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)-\mathbf{Z}_i^\top\mathbf{u}\right)-\rho_\tau\left(Y_i-\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\right)\right\}\right.$$
$$\left.-\lambda\sqrt{\tau\left(1-\tau\right)}\frac{2\sqrt{s}\left(c_0+1\right)}{\kappa_0}\|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|_2\right]<0.$$

Note that since $t>(\kappa_0\sqrt{n})^{-1}$ and $1-\tau_{0n}>\sqrt{s\log(p)/n}$, it follows that,

$$\frac{1}{t}\frac{L\kappa_0}{c_0+1}\frac{1}{\left(1-\tau_{0n}\right)\sqrt{s}}\beta^{(1)}<L\kappa_0^2\frac{\sqrt{n}}{1-\tau_{0n}}\beta^{(1)}<C_FL\kappa_0^2n^{(\gamma+3)/2},$$

where the second inequality holds by (S1.4). Since

$$\log\left(\frac{1}{t}\frac{L\kappa_0}{c_0+1}\frac{1}{\left(1-\tau_{0n}\right)\sqrt{s}}\beta^{(1)}\right)<\log\left(C_F\right)+2\log\left(\sqrt{L}\kappa_0\right)+\frac{2}{\gamma+3}\log n,$$

it follows that

$$\log\left\{p\vee\left(\frac{1}{t}\frac{L\kappa_0}{c_0+1}\frac{1}{\left(1-\tau_{0n}\right)\sqrt{s}}\beta^{(1)}\right)\right\}<\left\{\left(1+\log C_F\right)\log p\right\}\vee\left\{\left(2C_L+\frac{2}{\gamma+3}\right)\log\left(p\vee n\right)\right\}$$
$$<\left\{\left(1+\log C_F\right)\vee\left(2C_L+\frac{2}{\gamma+3}\right)\right\}\log\left(p\vee n\right).$$

By Lemma 7, we have

$$\sqrt{n}\epsilon(t) \leq tC_E\sqrt{(1+\log C_F) \vee \left(2C_L + \frac{2}{\gamma+3}\right)}A\frac{c_0+1}{\kappa_0}\left(2+(1-p_{n,1})\right)\sqrt{s\log(p\vee n)},$$

and thus

$$\min_{\mathbf{u}\in A_\tau, \|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|=t}\left\{\mathbb{E}\left\{M_n(\mathbf{u},\tau)\right\} - \lambda\sqrt{\frac{\tau(1-\tau)}{n}}\frac{2\sqrt{s}(c_0+1)}{\kappa_0}\|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|_2\right.$$
$$\left. - tC_E\sqrt{(1+\log C_F) \vee \left(2C_L + \frac{2}{\gamma+3}\right)}A\frac{c_0+1}{\kappa_0}\left(2+(1-p_{n,1})\right)\sqrt{s\log(p\vee n)}\right\} < 0.$$

By the definition of λ, we have

$$\min_{\mathbf{u}\in A_\tau, \|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|=t}\left\{\mathbb{E}\left\{M_n(\mathbf{u},\tau)\right\}\right.$$
$$- \sqrt{A^2+1}\cdot C_{\tilde{\Lambda}}\frac{1}{\sqrt{1-\tau_{0n}}}\sqrt{n\log p}\sqrt{\frac{\tau(1-\tau)}{n}}\frac{2\sqrt{s}(c_0+1)}{\kappa_0}\|\tilde{\Sigma}^{\frac{1}{2}}\mathbf{u}\|_2$$
$$\left. - tC_E\sqrt{(1+\log C_F) \vee \left(2C_L + \frac{2}{\gamma+3}\right)}A\frac{c_0+1}{\kappa_0}\left(2+(1-p_{n,1})\right)\sqrt{s\log(p\vee n)}\right\} < 0.$$

By (S1.6) and the condition $1-\tau_{0n} > \sqrt{s\log p/n}$, we have $t < b_n$. Therefore, by (S1.5), we have

$$C_S d_n^{-\gamma}(1-\tau_{0n})^{\gamma+1}t^2 - (1-\tau_{0n})C_T t - \sqrt{A^2+1}2C_{\tilde{\Lambda}}\frac{(c_0+1)}{\kappa_0}\sqrt{s\log p}\, t$$
$$- tC_E\sqrt{(1+\log C_F) \vee \left(2C_L + \frac{2}{\gamma+3}\right)}A\frac{c_0+1}{\kappa_0}\left(2+(1-p_{n,1})\right)\sqrt{s\log(p\vee n)} < 0.$$

By the choice of $\tilde{C}$, and $1-\tau_{0n} < 1 < \sqrt{s \log p}$, we have,

$$0 = d_n^{-\gamma} (1-\tau_{0n})^{\gamma+1} t^2 - \tilde{C} A \frac{c_0+1}{\kappa_0} \sqrt{s \log (p \vee n)}\, t < 0,$$

which brings a contradiction.

Therefore, we have proven that, with t defined in (S1.6), conditioned on $\cap_{i=1}^3 \Omega_i$,

$$\forall 1-\tau \in \mathcal{T}_n, \qquad \|\tilde{\Sigma}^{\frac{1}{2}} \left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right) \| \le t.$$

Let $A = 2$, $C = 4\tilde{C}A(c_0+1)/\kappa_0$. Combining this with (S1.3) completes the proof. □

S1.3 Proof of Theorem 2

Recall that $k = n(1-\tau_{0n})$. The conditions in Theorem 1 on τ_{0n} can be expressed in terms of k: $k \to \infty$, $k/n \to 0$, $k/n > \sqrt{s \log p/n}$, $d_n^{\gamma}(k/n) \to 0$ and $s_r d_n^{-\gamma}(k/n)^{\gamma} \to 0$. Before proving Theorem 2, we present Corollary 1, which follows directly from Theorem 1.

Corollary 1. *Under the conditions of Theorem 1, we have*

$$\sup_{1-\tau \in \mathcal{T}_n} \left\| \widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| = O_P\left(d_n^{\gamma} \left(\frac{n}{k} \right)^{\gamma+1} \sqrt{\frac{s \log (p \vee n)}{n}} \right).$$

Proof. Note that $1-\tau_{0n} = k/n$ and that the conditions $k \to \infty$, $k/n \to$

0, $k/n > \sqrt{s\log p/n}$, $d_n^{\gamma}(k/n) \to 0$ and $s_r d_n^{-\gamma}(k/n)^{\gamma} \to 0$ imply that $1-\tau_{0n} > \sqrt{s\log p/n}$, $s_r d_n^{-\gamma}(1-\tau_{0n})^{\gamma} \to 0$ and $d_n^{\gamma}(1-\tau_{0n}) \to 0$, which are the required conditions in Theorem 1. Substituting $1-\tau_{0n} = k/n$ into Theorem 1 leads to the result. □

Proof of Theorem 2.

By the definition of $\hat{\gamma}$, we have

$$
\begin{aligned}
\hat{\gamma}-\gamma &= \frac{\sum_{j=1}^{J}\phi(l_j)\left\{\log\left(\frac{\widehat{Q}_Y(1-l_j(1-\tau_n)|\mathbf{x})}{\widehat{Q}_Y(1-l_1(1-\tau_n)|\mathbf{x})}\right)-\gamma\log(1/l_j)\right\}}{\sum_{j=1}^{J}\phi(l_j)\log(1/l_j)} \\
&= \frac{\sum_{j=1}^{J}\phi(l_j)\left\{\log\left(\frac{\widehat{Q}_Y(1-l_j(1-\tau_n)|\mathbf{x})}{Q_Y(1-l_j(1-\tau_n)|\mathbf{x})}\right)-\log\left(\frac{\widehat{Q}_Y(1-l_1(1-\tau_n)|\mathbf{x})}{Q_Y(1-l_1(1-\tau_n)|\mathbf{x})}\right)+\log\left(\frac{Q_Y(1-l_j(1-\tau_n)|\mathbf{x})}{Q_Y(1-l_1(1-\tau_n)|\mathbf{x})}l_j^{\gamma}\right)\right\}}{\sum_{j=1}^{J}\phi(l_j)\log(1/l_j)} \\
&=: A_{n1}-A_{n2}+A_{n3},
\end{aligned}
$$

where

$$
\begin{aligned}
A_{n1} &= \frac{\sum_{j=1}^{J}\phi(l_j)\log\left(\frac{\widehat{Q}_Y(1-l_j(1-\tau_n)|\mathbf{x})}{Q_Y(1-l_j(1-\tau_n)|\mathbf{x})}\right)}{\sum_{j=1}^{J}\phi(l_j)\log(1/l_j)}, \\
A_{n2} &= \frac{\sum_{j=1}^{J}\phi(l_j)\log\left(\frac{\widehat{Q}_Y(1-l_1(1-\tau_n)|\mathbf{x})}{Q_Y(1-l_1(1-\tau_n)|\mathbf{x})}\right)}{\sum_{j=1}^{J}\phi(l_j)\log(1/l_j)}, \\
A_{n3} &= \frac{\sum_{j=1}^{J}\phi(l_j)\log\left(\frac{Q_Y(1-l_j(1-\tau_n)|\mathbf{x})}{Q_Y(1-l_1(1-\tau_n)|\mathbf{x})}l_j^{\gamma}\right)}{\sum_{j=1}^{J}\phi(l_j)\log(1/l_j)}.
\end{aligned}
$$

We first claim that

$$
\frac{\widehat{Q}_Y(1-l_j(1-\tau_n)\mid\mathbf{x})}{Q_Y(1-l_j(1-\tau_n)\mid\mathbf{x})}-1 = O_P\left(\frac{ns}{k}\sqrt{\frac{p\log(p\vee n)}{n}}\right), \quad j=1,\ldots,J.
$$

By Lemma 3, $d_n^{-\gamma} n/k \to \infty$ and $s_r d_n^{-\gamma} (k/n)^{\gamma} \to 0$, we have $Q_Y(1-\ell_j(1-\tau_n) \mid \mathbf{x}) \sim d_n^{\gamma}(1-\tau_n)^{-\gamma}$.

Denote $\boldsymbol{\delta}_j = \widehat{\boldsymbol{\beta}}(1-\ell_j(1-\tau_n)) - \boldsymbol{\beta}(1-\ell_j(1-\tau_n))$. Since for $j = 1,\ldots,J$, $\|\boldsymbol{\delta}_j\| = O_P\left(d_n^{\gamma}(n/k)^{\gamma+1}\sqrt{s\log(p\vee n)/n}\right)$ holds by Corollary 1, it follows that, conditional on the event in Lemma 2 (ii), $\left|\mathbf{z}^T\boldsymbol{\delta}_j\right| \le C\|\boldsymbol{\delta}_j\|_1 \le C(c_0+1)\left\|\{\boldsymbol{\delta}_j\}_{T_\tau}\right\|_1 \le C(c_0+1)\sqrt{s}\left\|\{\boldsymbol{\delta}_j\}_{T_\tau}\right\|_2 \le C(c_0+1)\sqrt{s}\|\boldsymbol{\delta}_j\|_2$, where C is a constant such that $|\mathbf{x}_i| \le C, i = 1,2,\ldots,p$. Therefore, for $j = 1,2,\ldots,J$, we have

$$\left|\widehat{Q}_Y(1-\ell_j(1-\tau_n) \mid \mathbf{x}) - Q_Y(1-\ell_j(1-\tau_n) \mid \mathbf{x})\right| = \left|\mathbf{z}^T\boldsymbol{\delta}_j\right| = O_P\left(d_n^{\gamma}\left(\frac{n}{k}\right)^{\gamma+1} s\sqrt{\frac{\log(p\vee n)}{n}}\right),$$

and hence $A_{n1} = O_P\left((ns/k)\sqrt{\log(p\vee n)/n}\right)$ and $A_{n2} = O_P\left((ns/k)\sqrt{\log(p\vee n)/n}\right)$.

Now consider A_{n3}. By Lemma 3 and $d_n^{-\delta} d_{1n}(n/k)^{-\delta} \to 0$, $s_r = o\left(\left(d_n^{\gamma+\rho}(n/k)^{\gamma+\rho}\right) \vee \left(d_n^{\gamma-\delta} d_{1n}(n/k)^{\gamma-\delta}\right)\right)$, we have that, for each j,

$$\begin{aligned}
&\log\left(\frac{Q_Y(1-l_j(1-\tau_n) \mid \mathbf{x})}{Q_Y(1-l_1(1-\tau_n) \mid \mathbf{x})} l_j^{\gamma}\right)\\
&= \log\left[\frac{\left(\frac{c^*}{l_j(1-\tau_n)}\right)^{\gamma}\left\{1+\frac{\gamma d^*(\mathbf{z})}{\rho^*}\left(\frac{1}{l_j(1-\tau_n)}\right)^{\rho^*}(1+o(1))\right\}}{\left(\frac{c^*}{l_1(1-\tau_n)}\right)^{\gamma}\left\{1+\frac{\gamma d^*(\mathbf{z})}{\rho^*}\left(\frac{1}{l_1(1-\tau_n)}\right)^{\rho^*}(1+o(1))\right\}} l_j^{\gamma}\right]\\
&= \log\left[\left\{1+\frac{\gamma d^*(\mathbf{z})}{\rho^*}\left(\frac{1}{l_j(1-\tau_n)}\right)^{\rho^*}(1+o(1))\right\}\left\{1-\frac{\gamma d^*(\mathbf{z})}{\rho^*}\left(\frac{1}{l_1(1-\tau_n)}\right)^{\rho^*}(1+o(1))\right\}\right]\\
&= \log\left\{1+\frac{\gamma d^*(\mathbf{z})}{\rho^*}(1-\tau_n)^{-\rho^*}\left(l_j^{-\rho^*}-1\right)(1+o(1))\right\}\\
&= \frac{\gamma d^*(\mathbf{z})}{\rho^*}(1-\tau_n)^{-\rho^*}\left(l_j^{-\rho^*}-1\right)(1+o(1)),
\end{aligned}$$

where $c^* = cK(\mathbf{z})^\gamma$,

$$\rho^* = \begin{cases} \rho, & \text{if } d_n^{\rho+\delta}(1-\tau_n)^{-\rho-\delta}d_{1n}^{-1} \to \infty, \\ -\delta, & \text{if } d_n^{\rho+\delta}(1-\tau_n)^{-\rho-\delta}d_{1n}^{-1} = O(1), \end{cases}$$

$$d^*(\mathbf{z}) = \begin{cases} dK(\mathbf{z})^\rho, & \text{if } d_n^{\rho+\delta}(1-\tau_n)^{-\rho-\delta}d_{1n}^{-1} \to \infty, \\ -\delta K_1(\mathbf{z})K(\mathbf{z})^{-\delta}, & \text{if } d_n^{\rho+\delta}(1-\tau_n)^{-\rho-\delta}d_{1n}^{-1} = O(1). \end{cases}$$

Therefore, uniformly for $\mathbf{x} \in \mathcal{X}$, $A_{n3} = O_P\left(d_n^\rho (k/n)^{-\rho} \vee d_{1n}d_n^{-\delta}(k/n)^\delta\right)$. To summarize, the condition $d_n^\rho (k/n)^{-\rho} \vee d_{1n}d_n^{-\delta}(k/n)^\delta = o\left((ns/k)\sqrt{\log(p\vee n)/n}\right)$ implies that

$$\hat{\gamma} - \gamma = O_P\left(\frac{ns}{k}\sqrt{\frac{\log(p\vee n)}{n}}\right).$$

We demonstrate that for $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^n \mathbf{X}_i$,

$$\hat{\gamma}(\bar{\mathbf{x}}) - \gamma = O_P\left(\frac{n}{k}\sqrt{\frac{s\log(p\vee n)}{n}}\right).$$

Since $\bar{\mathbf{x}}$ is nonzero at only one position, for $j = 1, 2, \ldots, J$, we have

$$\left|\widehat{Q}_Y\left(1-\ell_j(1-\tau_n) \mid \bar{\mathbf{x}}\right) - Q_Y\left(1-\ell_j(1-\tau_n) \mid \bar{\mathbf{x}}\right)\right| = O_P\left(d_n^\gamma \left(\frac{n}{k}\right)^{\gamma+1}\sqrt{\frac{s\log(p\vee n)}{n}}\right),$$

by the Cauchy inequality. Using similar methods used before, we can verify the result. □

S1.4 Proof of Theorem 3

Notice that $\widehat{Q}_Y(\tau_n \mid \mathbf{x}) = \{(1-\tau_{0n})/(1-\tau_n)\}^{\hat{\gamma}} \, \mathbf{z}^T \widehat{\boldsymbol{\beta}}(\tau_{0n})$, where $1 - k/n = \tau_{0n}$ and $Q_Y(\tau_n \mid \mathbf{x}) = U_Y((1-\tau_n)^{-1} \mid \mathbf{x})$. By the regularly varying property of $U_Y(t \mid \mathbf{x})$, that is, $U_Y(tx \mid \mathbf{x})/U_Y(t \mid \mathbf{x}) \to x^{\gamma}$, as $t \to \infty$, we have

$$U_Y\left(\frac{1}{1-\tau_n} \,\middle|\, \mathbf{x}\right) = U_Y\left(\frac{n}{k} \,\middle|\, \mathbf{x}\right)\left(\left(\frac{k}{n(1-\tau_n)}\right)^{\gamma} + o(1)\right).$$

Therefore,

$$\begin{aligned}
\frac{\widehat{Q}_Y(\tau_n \mid \mathbf{x})}{Q_Y(\tau_n \mid \mathbf{x})} &= \left(\frac{k}{n(1-\tau_n)}\right)^{\hat{\gamma}-\gamma} \frac{\mathbf{z}^T\widehat{\boldsymbol{\beta}}(\tau_{0n})}{U_Y\left(\frac{n}{k} \mid \mathbf{x}\right)} \frac{1}{1+\left(\frac{n(1-\tau_n)}{k}\right)^{\gamma} o(1)} \\
&= \left\{1+\log\left(\frac{k}{n(1-\tau_n)}\right)(\hat{\gamma}-\gamma)(1+o(1))\right\} \times \left\{1+\frac{\mathbf{z}^T\widehat{\boldsymbol{\beta}}(\tau_{0n}) - U_Y\left(\frac{n}{k} \mid \mathbf{x}\right)}{U_Y\left(\frac{n}{k} \mid \mathbf{x}\right)}\right\} \\
&\quad \times \left\{1-\left(\frac{n(1-\tau_n)}{k}\right)^{\gamma} o(1)\right\} \\
&=: (1+R_{1n})(1+R_{2n})(1-R_{3n}).
\end{aligned}$$

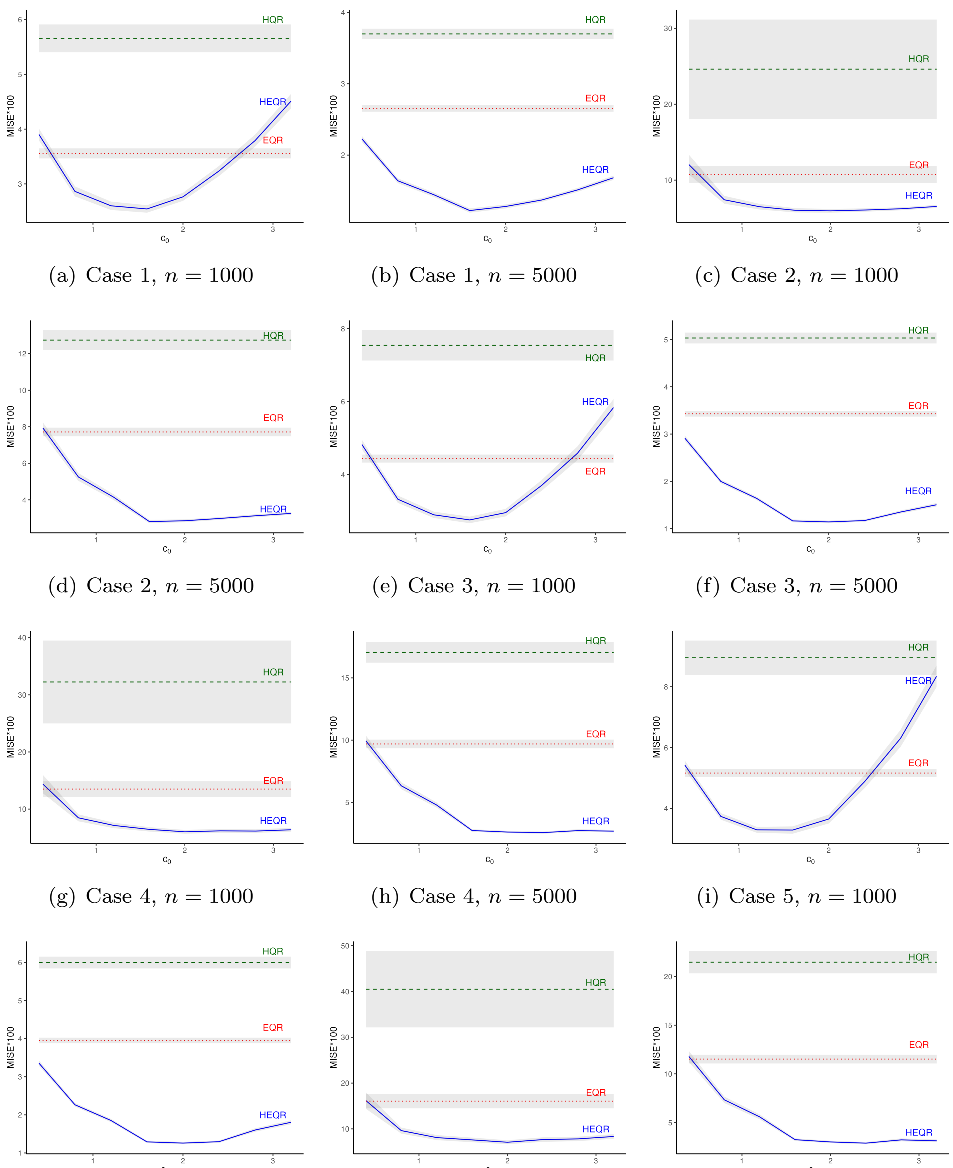
By Theorem 2 and Taylor expansion of the function $f(x) = \{k/n(1-\tau_n)\}^x$ at $x = 0$, we have $|R_{1n}| = O_P\left(\log[k/\{n(1-\tau_n)\}](ns/k)\sqrt{\log(p \vee n)/n}\right)$. Similar to the proof of Theorem 2, we have $|R_{2n}| = O_P\left((ns/k)^{\gamma+1}\sqrt{\log p/n}/(n/k)^{\gamma}\right) = O_P\left((ns/k)\sqrt{\log p/n}\right)$. For R_{3n}, we have $|R_{3n}| = o_P[\{n(1-\tau_n)/k\}^{\gamma}] = o_P(1)$, since $\gamma > 0$ and $n(1-\tau_n) = o(k)$.

To sum up, the leading term of $\widehat{Q}_Y(\tau_n \mid \mathbf{x})/Q_Y(\tau_n \mid \mathbf{x}) - 1$ is $|R_{1n}|$.

Combining with the condition $\log[k/\{n(1-\tau_n)\}](ns/k)\sqrt{\log(p\vee n)/n} \to 0$, we conclude the results. $\square$

S2 Sensitivity Analysis in the Simulation Study

In Section 3, we conduct a sensitivity analysis to evaluate the stability of the proposed method, denoted as HEQR, with respect to the choices of c_0, δ_1, and δ_2 in the rule of thumb $k = \lfloor c_0 n^{0.5+\delta_1}(\log p)^{0.5+\delta_2} \rfloor$. To illustrate the results, we present the following figures for all six cases. Figures S.1 present the Mean Integrated Squared Error (MISE) of the compared estimators for extreme conditional quantiles at $\tau_n = 0.995$ for Cases 1–6, plotted against $c_0 \in [0.4, 3]$, with (δ_1, δ_2) fixed at $(0.01, 0.05)$. The compared estimators include the extreme quantile regression (EQR) method without assuming a common slope in Wang, Li & He (2012), and the high-dimensional quantile regression (HQR) method from Belloni & Chernozhukov (2011). The pattern of MISE for $\tau_n = 0.999$ is similar and is therefore omitted. Figures S.2 present the MISE of HEQR as a function of $\delta_1 \in [0.005, 0.03]$, with $c_0 = 0.8$ and $\delta_2 = 0.05$, while Figures S.3 present the MISE of HEQR as a function of $\delta_2 \in [0.02, 0.08]$, with $c_0 = 0.8$ and $\delta_1 = 0.01$. The figures indicate that our proposed method remains stable for $c_0 \in [0.5, 2.3]$, $\delta_1 \in [0.005, 0.016]$, and $\delta_2 \in [0.01, 0.08]$, across various scenarios and sample sizes.



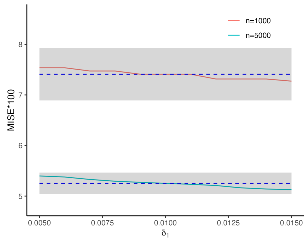
(a) Case 1, $n = 1000$ (b) Case 1, $n = 5000$ (c) Case 2, $n = 1000$

(d) Case 2, $n = 5000$ (e) Case 3, $n = 1000$ (f) Case 3, $n = 5000$

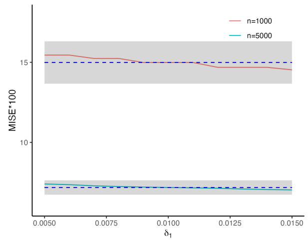
(g) Case 4, $n = 1000$ (h) Case 4, $n = 5000$ (i) Case 5, $n = 1000$

(j) Case 5, $n = 5000$ (k) Case 6, $n = 1000$ (l) Case 6, $n = 5000$

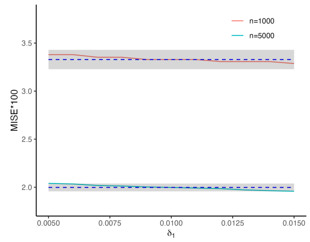
Figure S.1: The MISE of estimators for the extreme conditional quantile at $\tau_n = 0.995$ against c_0 in Cases 1-6 with $n = 1000$ and $n = 5000$.



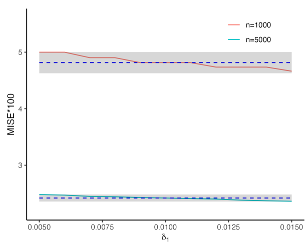
(a) Case 2, $\tau_n = 0.995$



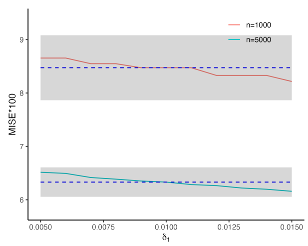
(b) Case 2, $\tau_n = 0.999$



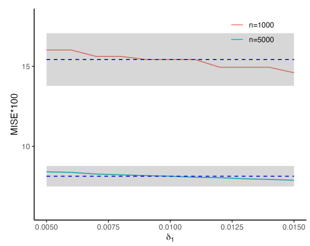
(c) Case 3, $\tau_n = 0.995$



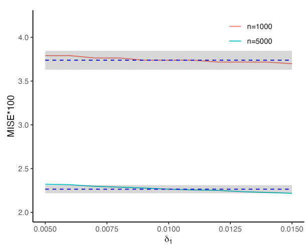
(d) Case 3, $\tau_n = 0.999$



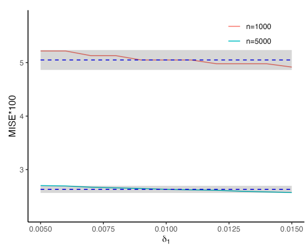
(e) Case 4, $\tau_n = 0.995$



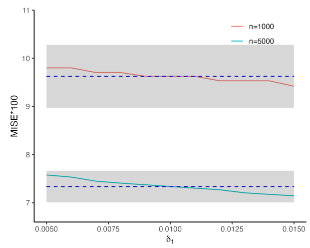
(f) Case 4, $\tau_n = 0.999$



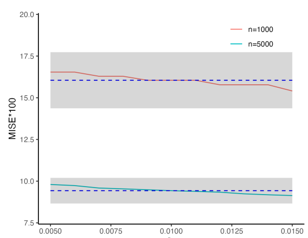
(g) Case 5, $\tau_n = 0.995$



(h) Case 5, $\tau_n = 0.999$

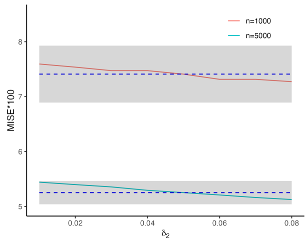


(i) Case 6, $\tau_n = 0.995$

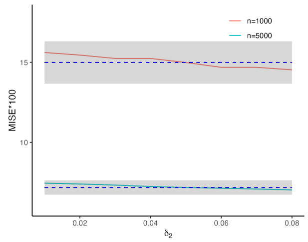


(j) Case 6, $\tau_n = 0.999$

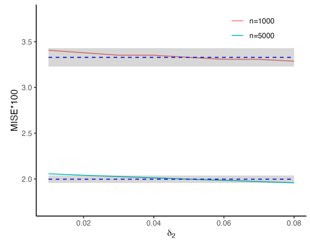
Figure S.2: The MISE of HEQR estimator for the extreme conditional quantile at $\tau_n = 0.995$ and 0.999 against δ_1 in Cases 2-6.



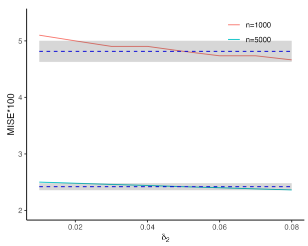
(a) Case 2, $\tau_n = 0.995$



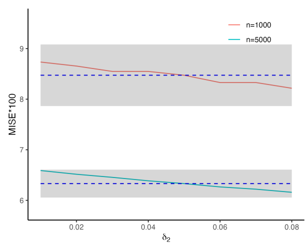
(b) Case 2, $\tau_n = 0.999$



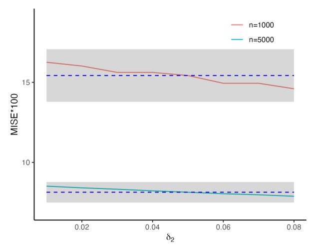
(c) Case 3, $\tau_n = 0.995$



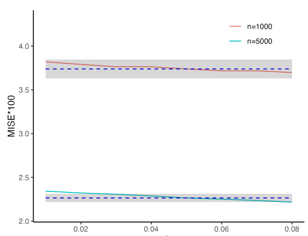
(d) Case 3, $\tau_n = 0.999$



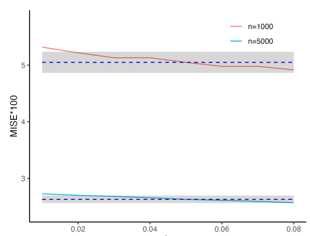
(e) Case 4, $\tau_n = 0.995$



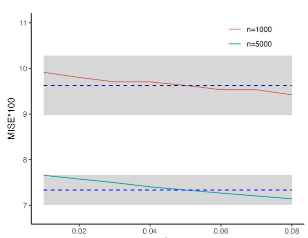
(f) Case 4, $\tau_n = 0.999$



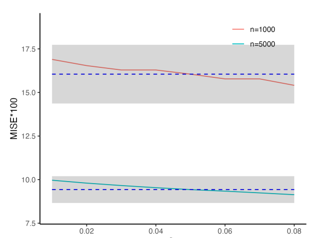
(g) Case 5, $\tau_n = 0.995$



(h) Case 5, $\tau_n = 0.999$



(i) Case 6, $\tau_n = 0.995$



(j) Case 6, $\tau_n = 0.999$

Figure S.3: The MISE of HEQR estimator for the extreme conditional quantile at $\tau_n = 0.995$ and 0.999 against δ_2 in Cases 2-6.

S3 Variables in the Auto Insurance Data

Tables S.1 and S.2 describe the continuous and binary variables in the auto insurance claims data analyzed in Section 4, respectively.

Table S.1: Description of continuous variables in the auto insurance data

Abbreviation	**Varaible**	**Description**	**Mean(SD)**
Age	Age		44.70(8.65)
HC	Home Children	count of children currently residing in a household	0.74(1.13)
YoJ	Years on Job		11.68(9.53)
TiF	Time In Force	the duration the insurance policy has been in effect	5.34(4.13)
Income	log(Income)		9.83(3.16)
TravelTime	log(Travel Time)		3.37(0.60)
VValue	log(Vehicle Value)		9.46(0.63)

Note: SD refers to standard deviation.

Bibliography

Battey, H., Fan, J., Liu, H., Lu, J. & Zhu, Z. (2018), 'Distributed testing and estimation under sparse high dimensional models', *Annals of Statistics* **46**(3), 1352–1382.

Belloni, A. & Chernozhukov, V. (2011), 'ℓ1-penalized quantile regression in high-dimensional sparse models', *Annals of Statistics* **39**(1), 82–130.

Chernozhukov, V. (2005), 'Extramal quantile regression', *Annals of Statistics* **33**(2), 806–839.

de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer Science & Business Media.

Javanmard, A. & Montanari, A. (2014), 'Confidence intervals and hypothesis testing for high-dimensional regression', *The Journal of Machine Learning Research* **15**(1), 2869–2909.

Ledoux, M. & Talagrand, M. (2013), *Probability in Banach Spaces: Isoperimetry and Processes*, Springer Science & Business Media.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.

Vershynin, R. (2012), *Compressed Sensing: Theory and Applications*, Cambridge University Press.

Wang, H. J., Li, D. & He, X. (2012), 'Estimation of high conditional quantiles for heavy-tailed distributions', *Journal of the American Statistical Association* **107**(500), 1453–1464.

Wang, L., Wu, Y. & Li, R. (2012), 'Quantile regression for analyzing heterogeneity in ultra-high dimension', *Journal of the American Statistical Association* **107**(497), 214–222.

Zhang, Y. (2018), 'Extremal quantile treatment effects', *Annals of Statistics* **46**(6B), 3707–3740.

Table S.2: Description of binary variables in the auto insurance data

Abbreviation	**Varaible**	**Description**
Hvalue1	Home Value 1	home value $\in (0, 150000]$
Hvalue2	Home Value 2	home value $\in (150000, 200000]$
Hvalue3	Home Value3	home value $\in (200000, \infty)$
Rural	Rural Population	living in rural areas
DC	Driving Children	there is at least one child in the vehicle while driving
SP	Single Parent	raising one or more children alone, without the support of a partner or spouse living in the household
MS	Marital Status	equals 1 if not married
Red	Red Car	
License	License Revoked	
Edu.L1	Bachelor's Degree	
Edu.L2	Higher Education	education level higher than Bachelor's degree
OldC1	Old Claims 1	total claims in last five years $\in (0, 5000)$
OldC2	Old Claims 2	total claims in last five years $\in (5000, \infty)$
FreqC1	Claims Freq 1	claim frequency in the last 5 years is 1
FreqC2	Claims Freq 2	claim frequency in the last 5 years is 2
FreqC3	Claims Freq 3	claim frequency in the last 5 years $\in (2, \infty)$
VPoint1	Vehicle Points 1	vehicle points $\in \{1, 2\}$ (vehicle points refer to a score or classification used by insurers to help determine the risk associated with insuring a specific vehicle.)
VPoint2	Vehicle Points 2	vehicle points $\in \{3, 4, \ldots\}$
VAge1	Car Age 1	car age $\in [5, 10)$
VAge1	Car Age 2	car age $\in [10, \infty)$
VType1	is commercial	
VType2	is van	
VType3	is sportcar	
VType4	is minivan	
VType5	is paneltruck	
VType6	is pickup	
Job1	is homemaker	
Job2	is student	
Job3	is Clerical	
Job4	is Manager	
Job5	is Professional	
Job6	is Doctor	