Supplementary Material for
"Testing for Marginal Linear Effects in Quantile Regression"
by Huixia Judy Wang, Ian W. McKeague and Min Qian

This supplement contains the proofs of Lemmas 1–4, Corollary 1 and Lemmas 5–6 in the main paper, and some additional simulation results.

For any vector $\mathbf{v} \in \mathbb{R}^p$, let v_k denote its kth element, and $\mathbf{v}_{(-k)}$ denotes its subset excluding the kth element, for any $k = 1, \ldots, p$. For notational simplicity, we write $\bar{k}_n\{\tau, \mathbf{b}_0(\tau)\}$ as $k_n(\tau)$, and omit the argument τ in various expressions such as $M_k(\tau, \cdot)$, $\pi_k(\tau, \cdot)$, $V_k(\tau, \cdot)$, $\mathbf{J}_k(\tau, \cdot)$ and etc. when necessary. We note that even though $\mathcal{T}$ is assumed to be a set consisting of L prespecified quantile levels in the main paper, the results in Theorem 1 and Lemmas 1-4 in fact hold uniformly over $\tau \in \mathcal{T}$ for any $\mathcal{T} \subset (0, 1)$.

S.1. Proofs of Lemmas 1–4

LEMMA **1** *Suppose that assumptions A1–A5 hold. For all τ's in $\mathcal{T}$ for which $\beta_0(\tau) \neq 0$ and $k_0(\tau)$ is unique, we have $\hat{k}_n(\tau) \stackrel{a.s.}{\to} k_0(\tau)$ and*

$$\frac{n^{1/2}\{\hat{\theta}_n(\tau) - \theta_n(\tau)\}}{\hat{\sigma}_n(\tau)} \stackrel{d}{\to} \frac{M_{p+k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\pi_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\} - M_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\mu_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}}{V_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\sigma_{k_0(\tau)}(\tau)}.$$

Proof: By the proof of the consistency part of Theorem 3 of Angrist et al. (2006), $(\hat{\alpha}_k, \hat{\theta}_k) - (\alpha_k, \theta_k) = o_{p^*}(1)$ uniformly in $\tau \in \mathcal{T}$. Under assumptions A1-A3, and the uniqueness condition of $k_0(\tau)$, by the Lipschitz property of ρ_τ, we have $E\{\rho_\tau(Y - \alpha_k - \theta_k X_k)\} - E\{\rho_\tau(\epsilon + \alpha_0 + \mathbf{X}^T\boldsymbol{\beta}_0 - \alpha_k - \theta_k X_k)\} \to 0$ uniformly over k and τ. Therefore, $k_n(\tau) = \text{argmin}_k \min_{\alpha,\theta} E\{\rho_\tau(Y - \alpha - \theta X_k)\} = \text{argmin}_k E[\rho_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}] \to k_0(\tau)$ uniformly in $\tau \in \mathcal{T}$. Note that the empirical process $(\tau, \boldsymbol{\beta}) \mapsto \mathbb{P}_n[\rho_\tau(Y - \mathbf{X}^T\boldsymbol{\beta})]$ is stochastically equicontinuous over $\mathcal{T} \times \mathcal{B}$, where $\mathcal{B}$ is any compact set. This together with the Lipschitz and convexity properties of ρ_τ and the strong law of large numbers leads to the following uniform convergence result:

$$\mathbb{P}_n[\rho_\tau\{Y - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau)X_k\}] - E[\rho_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\}] \stackrel{a.s.}{\to} 0. \qquad \text{(S.1)}$$

Thus we have $\hat{k}_n(\tau) - k_n(\tau) \stackrel{a.s.}{\to} 0$ and $\hat{k}_n(\tau) \stackrel{a.s.}{\to} k_0(\tau)$ uniformly in $\tau \in \mathcal{T}$, where $\hat{k}_n(\tau) = \text{argmin}_k \mathbb{P}_n[\rho_\tau\{Y - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau)X_k\}]$. By the proof of the asymptotic normality part of

Theorem 3 in Angrist et al. (2006), we have the following uniform asymptotic representation

$$\sqrt{n}\left\{\begin{pmatrix}\hat{\alpha}_k(\tau)\\ \hat{\theta}_k(\tau)\end{pmatrix}-\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}\right\}=\tilde{\mathbf{J}}_k^{-1}(\tau)\mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{Y-\alpha_k(\tau)-\theta_k(\tau)X_k\}\right]+o_{p^*}(1), \quad \text{(S.2)}$$

provided that $\tilde{\mathbf{J}}_k(\tau)=E\left[f_Y\{\alpha_k(\tau)+\theta_k(\tau)X_k|\mathbf{X}\}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^T\right]$ is positive definite for all $\tau\in\mathcal{T}$, where $f_Y(\cdot|\mathbf{X})$ is the conditional density of Y given $\mathbf{X}$, $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$ is the empirical process and $\psi_\tau(u)=\tau-I(u\le 0)$. It is easy to see that under the local model, $\tilde{\mathbf{J}}_k(\tau)=E\left[f_{\epsilon(\tau)}\{\alpha_k(\tau)+\theta_k(\tau)X_k-\alpha_0(\tau)-\mathbf{X}^T\boldsymbol{\beta}_n(\tau)|\mathbf{X}\}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^T\right]\to\mathbf{J}_k\{\tau,\boldsymbol{\beta}_0(\tau)\}$, which is positive definite by condition A5. In addition, by using similar arguments as in the Appendix A.1.4 of Angrist et al. (2006), we can show that $\hat{\sigma}_k(\tau)\to\sigma_k(\tau)$ uniformly in $\tau\in\mathcal{T}$. Then, since $P\{\hat{\theta}_n(\tau)=\hat{\theta}_{k_0(\tau)}(\tau)\}\to 1$, we have

$$\begin{aligned}\frac{n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}}{\hat{\sigma}_n(\tau)}&=\frac{n^{1/2}\{\hat{\theta}_{k_0(\tau)}(\tau)-\theta_{k_0(\tau)}(\tau)\}}{\sigma_{k_0(\tau)}(\tau)}+o_{p^*}(1)\\ &\xrightarrow{d}\frac{M_{p+k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\pi_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}-M_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\mu_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}}{V_{k_0(\tau)}\{\boldsymbol{\beta}_0(\tau)\}\sigma_{k_0(\tau)}(\tau)} \qquad \text{(S.3)}\end{aligned}$$

uniformly in $\tau\in\mathcal{T}$, where the last step comes from extracting the second element of the vector on the right side of (S.2) with $k=k_0(\tau)$.

LEMMA **2** *If assumptions A1–A5 hold, we have*

$$\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}=\begin{pmatrix}\alpha_0(\tau)\\ \beta_{n,k}(\tau)\end{pmatrix}+\{\mathbf{J}_k^{-1}(\tau,\mathbf{0})\}\mathbf{A}_k^T(\tau)\boldsymbol{\beta}_{n,(-k)}(\tau)+o(n^{-1/2}) \qquad \text{(S.4)}$$

uniformly over τ for which $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$, where $\mathbf{A}_k(\tau)=\left(E\{f_{\epsilon(\tau)}(0|\mathbf{X})\mathbf{X}_{(-k)}\},E\{f_{\epsilon(\tau)}(0|\mathbf{X})X_k\mathbf{X}_{(-k)}\}\right)$.

Proof: When $\boldsymbol{\beta}_0(\tau)=0$, we have $Q_\tau(Y|\mathbf{X})=\alpha_0(\tau)+n^{-1/2}\mathbf{X}^T\mathbf{b}_0(\tau)$. Under assumptions A1, A3 and A4, we can show that $Q_\tau(Y|\mathbf{X})\to\alpha_0(\tau)$ uniformly in $\tau\in\mathcal{T}$. In addition, recall that $(\alpha_k(\tau),\theta_k(\tau))$ is the population quantile coefficient vector obtained by regressing Y on X_k, so $Q_\tau(Y|X_k)=\alpha_k(\tau)+X_k\theta_k(\tau)$. Therefore, $Q_\tau(Y|X_k)\to\alpha_0(\tau)$, which implies that $(\alpha_k(\tau),\theta_k(\tau))\to(\alpha_0(\tau),0)$ uniformly in $\tau\in\mathcal{T}$.

We next establish the approximate representation of $(\alpha_k(\tau),\theta_k(\tau))$ under the local model with $\boldsymbol{\beta}_n(\tau)=n^{-1/2}\mathbf{b}_0(\tau)$. Under this model, $(\alpha_0(\tau),\boldsymbol{\beta}_n(\tau))=(\alpha_0(\tau),\beta_{n,k}(\tau),\boldsymbol{\beta}_{n,(-k)}^T(\tau))$

$= \underset{(\alpha,\beta_k,\boldsymbol{\beta}_{(-k)})}{\operatorname{argmin}} E\left\{\rho_\tau(Y - \alpha - X_k\beta - \mathbf{X}_{(-k)}^T\boldsymbol{\beta}_{(-k)})\right\}$ is the population quantile coefficient vector obtained by including both X_k and $\mathbf{X}_{(-k)}$ as predictors, while $(\alpha_k(\tau), \theta_k(\tau))$ are obtained by including only X_k in the regression. By applying the last formula in Section 2.3 of Angrist et al. (2006), we get

$$\begin{pmatrix} \alpha_k(\tau) \\ \theta_k(\tau) \end{pmatrix} = \begin{pmatrix} \alpha_0(\tau) \\ \beta_{n,k}(\tau) \end{pmatrix} + \left[E\left\{W_{k\tau}(\mathbf{X})\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^T\right\}\right]^{-1} E\left[W_{k\tau}(\mathbf{X})\tilde{\mathbf{X}}_k R_\tau(\mathbf{X})\right], \quad \text{(S.5)}$$

where $\tilde{\mathbf{X}}_k = (1, X_k)^T$, $R_\tau(\mathbf{X}) = Q_\tau(Y|\mathbf{X}) - (\alpha_0(\tau), \beta_{n,k})^T(\tau)\tilde{\mathbf{X}}_k = (\alpha_0(\tau) + \mathbf{X}^T\boldsymbol{\beta}_n(\tau)) - (\alpha_0(\tau) + \beta_{n,k}(\tau)X_k) = \mathbf{X}_{(-k)}^T\boldsymbol{\beta}_{n,(-k)}(\tau)$, $W_{k\tau}(\mathbf{X}) = 1/2\int_0^1 f_{\epsilon(\tau)}\left\{u\Delta_{k\tau}(\mathbf{X})|\mathbf{X}\right\}\, du$, $\Delta_{k\tau}(\mathbf{X}) = \{\alpha_k(\tau) + \theta_k(\tau)X_k\} - \{\alpha_0(\tau) + \mathbf{X}^T\boldsymbol{\beta}_n(\tau)\}$. The first expectation in the right side of (S.5) can be expressed as

$$1/2\mathbf{J}_k(\tau, \mathbf{0}) - E\left[\left\{1/2f_{\epsilon(\tau)}(0|\mathbf{X}) - W_{k\tau}(\mathbf{X})\right\}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^T\right], \quad \text{(S.6)}$$

and below we show that the second term above tends to the zero matrix uniformly in $\tau \in \mathcal{T}$. In a similar fashion, the second expectation in (S.5) can be expressed as

$$1/2\mathbf{A}_k^T(\tau)\boldsymbol{\beta}_{n,(-k)}(\tau) - E\left[\left\{1/2f_{\epsilon(\tau)}(0|\mathbf{X}) - W_{k\tau}(\mathbf{X})\right\}\tilde{\mathbf{X}}_k\mathbf{X}_{(-k)}^T\right]\boldsymbol{\beta}_{n,(-k)}(\tau), \quad \text{(S.7)}$$

the second term of which has order $o(1)\boldsymbol{\beta}_n(\tau) = o(n^{-1/2})$. The result then follows easily from (S.5) by retaining the leading terms in (S.6) and (S.7), and noting that the remainder term is of order $o(n^{-1/2})$.

It remains to show that the second term in (S.6) tends to the zero matrix. Recall that in the first paragraph of the proof we showed that $(\alpha_k(\tau), \theta_k(\tau)) \to (\alpha_0(\tau), 0)$ uniformly in $\tau \in \mathcal{T}$. It follows that $\Delta_{k(\tau)}(\mathbf{X}) \to 0$, so, by assumptions A1 and A4 (using the continuity of $f_{\epsilon(\tau)}(y|\mathbf{X})$ at $y = 0$ in this case) and the dominated convergence theorem, $W_{k\tau}(\mathbf{X}) \to 1/2f_{\epsilon(\tau)}(0|\mathbf{X})$ uniformly in $\tau \in \mathcal{T}$. Applying the dominated convergence theorem to the expectations in (S.6) and (S.7), using condition A3 and the second part of A4 to check that the integrands are dominated, completes the proof.

LEMMA 3 *If assumptions A1–A5 hold, we have*

$$\left(n^{1/2}\begin{pmatrix}\hat{\alpha}_k(\tau)-\alpha_0(\tau)\\ \hat{\theta}_k(\tau)-\beta_{n,k}(\tau)\end{pmatrix}\right)_{k=1}^{p} \xrightarrow{d} \left(\mathbf{J}_k^{-1}(\tau,\mathbf{0})\left\{\begin{pmatrix}M_k(\tau,\mathbf{0})\\ M_{p+k}(\tau,\mathbf{0})\end{pmatrix}+\mathbf{A}_k^T(\tau)\mathbf{b}_{0,(-k)(\tau)}\right\}\right)_{k=1}^{p}$$

uniformly in $\tau\in\mathcal{T}$ *for which* $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$.

Proof: From (S.2),

$$\tilde{\mathbf{J}}_k(\tau)n^{1/2}\left\{\begin{pmatrix}\hat{\alpha}_k(\tau)\\ \hat{\theta}_k(\tau)\end{pmatrix}-\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}\right\} = \mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)-\Delta_{k\tau}(\mathbf{X})\}\right]+o_{p^*}(1),$$

where $\tilde{\mathbf{J}}_k(\tau)=E\left[f_{\epsilon(\tau)}\{e_k(\tau)|\mathbf{X}\}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^T\right]$, and

$$\begin{aligned}\Delta_{k\tau}(\mathbf{X}) &= \{\alpha_k(\tau)+\theta_k(\tau)X_k\}-\{\alpha_0(\tau)+\mathbf{X}^T\boldsymbol{\beta}_n(\tau)\}\\ &= \tilde{\mathbf{X}}_k^T\left[\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{A}_k^T(\tau)\boldsymbol{\beta}_{n,(-k)}(\tau)+o(n^{-1/2})\right]-\mathbf{X}_{(-k)}^T\boldsymbol{\beta}_{n,(-k)}(\tau)\end{aligned}\qquad\text{(S.8)}$$

uniformly in $\tau\in\mathcal{T}$. By Lemma 2, when $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$, it is easy to show that $\tilde{\mathbf{J}}_k(\tau)=\mathbf{J}_k(\tau,\mathbf{0})+O(n^{-1/2})$ uniformly for all $k=1,\dots,p$ under assumptions A1, A3 and A4. Writing

$$\mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)-\Delta_{k\tau}(\mathbf{X})\}\right]=\mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}\right]+\mathbb{G}_n\left(\tilde{\mathbf{X}}_k[\psi_\tau\{\epsilon(\tau)-\Delta_{k\tau}(\mathbf{X})\}-\psi_\tau\{\epsilon(\tau)\}]\right),$$

we can show that the second term above converges to zero uniformly in $\tau\in\mathcal{T}$. This follows by applying Lemma 19.24 of van der Vaart (2000), and the fact that $(\tau,\boldsymbol{\beta})\mapsto\mathbb{G}_n[\psi_\tau(Y-\mathbf{X}^T\boldsymbol{\beta})\mathbf{X}]$ is stochastically equicontinuous over $\mathcal{T}\times\mathcal{B}$, where $\mathcal{B}$ is any compact set. To check the conditions of that lemma, first note that the class of functions $\mathcal{F}=\{y\mapsto\psi_\tau(y-\delta)-\psi_\tau(y)\colon\delta\in\mathbb{R}\}$ is P-Donsker (for any distribution P on the real line). Also note that the function $g\colon\delta\mapsto E[\psi_\tau\{\epsilon(\tau)-\delta\}-\psi_\tau\{\epsilon(\tau)\}]^2$ is continuous (by A4) and vanishes at $\delta=0$, and $\Delta_{k\tau}(\mathbf{X})\to 0$, so $g\{\Delta_{k\tau}(\mathbf{X})\}\to 0$, and therefore $g\{\Delta_{k\tau}(\mathbf{X})\}\to 0$ uniformly in $\tau\in\mathcal{T}$. This shows that $\mathbb{G}_n\left[\psi_\tau\{\epsilon(\tau)-\Delta_{k\tau}(\mathbf{X})\}-\psi_\tau\{\epsilon(\tau)\}\right]$ tends to zero in probability. A similar argument applies to $\mathbb{G}_n\left[X_k(\psi_\tau\{\epsilon(\tau)-\Delta_{k\tau}(\mathbf{X})\}-\psi_\tau\{\epsilon(\tau)\})\right]$, the second component of $\mathbb{G}_n\left[\tilde{\mathbf{X}}_k(\psi_\tau\{\epsilon(\tau)-\Delta_{k\tau}(\mathbf{X})\}-\psi_\tau\{\epsilon(\tau)\})\right]$.

Therefore, we have

$$n^{1/2}\left\{\begin{pmatrix}\hat{\alpha}_k(\tau)\\ \hat{\theta}_k(\tau)\end{pmatrix}-\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}\right\}=\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbb{G}_n[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}]+o_{p*}(1). \tag{S.9}$$

Combining Lemma 2 and (S.9), we get

$$\begin{aligned} & n^{1/2}\begin{pmatrix}\hat{\alpha}_k(\tau)-\alpha_0(\tau)\\ \hat{\theta}_k(\tau)-\beta_{n,k}(\tau)\end{pmatrix} \\ = \;& n^{1/2}\left[\begin{pmatrix}\hat{\alpha}_k(\tau)\\ \hat{\theta}_k(\tau)\end{pmatrix}-\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}\right]+n^{1/2}\left[\begin{pmatrix}\alpha_k(\tau)\\ \theta_k(\tau)\end{pmatrix}-\begin{pmatrix}\alpha_0(\tau)\\ \beta_{n,k}(\tau)\end{pmatrix}\right] \\ = \;& \mathbf{J}_k^{-1}(\tau,\mathbf{0})\left(\mathbb{G}_n[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}]+\mathbf{A}_k^T(\tau)\mathbf{b}_{0,(-k)}(\tau)\right)+o_{p^*}(1). \end{aligned} \tag{S.10}$$

Lemma 3 thus follows by the Slutsky's theorem and a multivariate central limit theorem.

LEMMA **4** *Under assumptions A1–A5, we have*

$$\begin{aligned}\frac{n^{1/2}\{\hat{\theta}_n(\tau)-\theta_n(\tau)\}}{\hat{\sigma}_n(\tau)} \stackrel{d}{\to} & \frac{M_{p+K(\tau)}(\tau,\mathbf{0})\pi(\tau,\mathbf{0})-M_{K(\tau)}(\tau,\mathbf{0})\mu_{K(\tau)}(\tau,\mathbf{0})}{V_{K(\tau)}(\tau,\mathbf{0})\sigma_{K(\tau)}(\tau)} \\ & +\left\{\frac{\mathbf{C}_{K(\tau)}(\tau)}{V_{K(\tau)}(\tau,\mathbf{0})}-\frac{\mathbf{C}_{\kappa_\tau\{\mathbf{b}_0(\tau)\}}(\tau)}{V_{\kappa_\tau\{\mathbf{b}_0(\tau)\}}(\tau,\mathbf{0})}\right\}^T\frac{\mathbf{b}_0(\tau)}{\sigma_{K(\tau)}(\tau)}\end{aligned}$$

uniformly over $\tau\in\mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau)=\mathbf{0}$, where

$$K(\tau)=\underset{k=1,\dots,p}{\operatorname{argmax}}\ \left\{\mathbf{M}_k(\tau)+\mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau)\right\}^T\mathbf{J}_k^{-1}(\tau,\mathbf{0})\left\{\mathbf{M}_k(\tau)+\mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau)\right\}$$

with $\mathbf{M}_k(\tau)=(M_k(\tau,\mathbf{0}),M_{k+p}(\tau,\mathbf{0}))^T$.

Proof: Note that

$$\hat{k}_n(\tau)=\underset{1\le k\le p}{\operatorname{argmin}}\,\mathbb{P}_n\left[\rho_\tau\{Y-\hat{\alpha}_k(\tau)-\hat{\theta}_k(\tau)X_k\}-\rho_\tau\{\epsilon(\tau)\}\right]. \tag{S.11}$$

Denote

$$\mathbf{Z}_k = \begin{pmatrix} 1 \\ X_k \\ \mathbf{X} \end{pmatrix}, \boldsymbol{\delta}_k(\tau) = n^{1/2} \begin{pmatrix} \alpha_k(\tau) - \alpha_0(\tau) \\ \theta_k(\tau) \\ -\boldsymbol{\beta}_n(\tau) \end{pmatrix}, \hat{\boldsymbol{\delta}}_k(\tau) = n^{1/2} \begin{pmatrix} \hat{\alpha}_k(\tau) - \alpha_0(\tau) \\ \hat{\theta}_k(\tau) \\ -\boldsymbol{\beta}_n(\tau) \end{pmatrix}.$$

Then $\mathbb{P}_n\left[\rho_\tau\{Y - \alpha_k(\tau) - \theta_k(\tau)X_k\} - \rho_\tau\{\epsilon(\tau)\}\right] = \mathbb{P}_n\left[\rho_\tau\{\epsilon(\tau) - n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)\} - \rho_\tau\{\epsilon(\tau)\}\right]$ is minimized at $\hat{\boldsymbol{\delta}}_k(\tau)$ for each $\tau \in \mathcal{T}$. By Knight's identity (1998)

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \{I(u \le s) - I(u \le 0)\}\, ds,$$

we get

$$\begin{aligned} & n\mathbb{P}_n\left[\rho_\tau\{\epsilon(\tau) - n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)\} - \rho_\tau\{\epsilon(\tau)\}\right] \\ = & -\mathbb{G}_n\left[\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)\psi_\tau\{\epsilon(\tau)\}\right] + n\mathbb{P}_n\left[U\{\delta_k(\tau), \mathbf{Z}_k\}\right], \end{aligned} \tag{S.12}$$

where

$$U(\delta_k, \mathbf{Z}_k) = \int_0^{n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k} \left[I\{\epsilon(\tau) \le s\} - I\{\epsilon(\tau) \le 0\}\right] ds.$$

Now write the second term of (S.12) as

$$n\mathbb{P}_n[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\}] = nE\left[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k)\right] + n\mathbb{P}_n\left[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\} - E\left(U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\}\right)\right].$$

By assumptions A3 and A4, we have that uniformly in $\tau \in \mathcal{T}$,

$$\begin{aligned} nE\left[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\}\right] &= nE\left[\int_0^{n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)} \left\{F_{\epsilon(\tau)}(s|\mathbf{X}) - F_{\epsilon(\tau)}(0|\mathbf{X})\right\} ds\right] \\ &= \frac{1}{2}\boldsymbol{\delta}_k^T(\tau)\mathbf{D}_k(\tau)\boldsymbol{\delta}_k(\tau) + o(1), \end{aligned} \tag{S.13}$$

where

$$\mathbf{D}_k(\tau) = E\left[f_{\epsilon(\tau)}(0|\mathbf{X})\,\mathbf{Z}_k\mathbf{Z}_k^T\right] = \begin{pmatrix} \mathbf{J}_k(\tau, \mathbf{0}) & \mathbf{B}_k^T(\tau) \\ \mathbf{B}_k(\tau) & \mathbf{C}(\tau) \end{pmatrix},$$

$\mathbf{B}_k(\tau) = E\left\{f_{\epsilon(\tau)}(0|\mathbf{X})\mathbf{X}\tilde{\mathbf{X}}_k^T\right\}$ and $\mathbf{C}(\tau) = E\left\{f_{\epsilon(\tau)}(0|\mathbf{X})\mathbf{X}\mathbf{X}^T\right\}$. Thus the bound

$$
\begin{aligned}
&\text{Var}\left\{n\mathbb{P}_n[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\}]\right. \\
\leq\ & nE\Big[\int_0^{n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)} [I\{\epsilon(\tau) \leq s\} - I\{\epsilon(\tau) \leq 0\}] - \left\{F_{\epsilon(\tau)}(s|\mathbf{X}) - F_{\epsilon(\tau)}(0|\mathbf{X})\right\} ds\Big]^2 \\
\leq\ & nE\Big[\Big|\int_0^{n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)} [I\{\epsilon(\tau) \leq s\} - I\{\epsilon(\tau) \leq 0\}] - \left\{F_{\epsilon(\tau)}(s|\mathbf{X}) - F_{\epsilon(\tau)}(0|\mathbf{X})\right\} ds\Big|\Big] \\
&\quad \times 2|n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)| \\
\leq\ & 4nE\left[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\}\right] |n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)|,
\end{aligned}
$$

since $U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\} \geq 0$. By (S.13) and assumption A3, $\text{Var}\left\{n\mathbb{P}_n[U\{\boldsymbol{\delta}_k(\tau), \mathbf{Z}_k\}]\right. \to 0$ uniformly in $\tau \in \mathcal{T}$ for any $\boldsymbol{\delta}_k(\tau) = O(1)$. Therefore, for any $\boldsymbol{\delta}_k(\tau) = O(1)$,

$$
\begin{aligned}
& n\mathbb{P}_n\left[\rho_\tau\{\epsilon(\tau) - n^{-1/2}\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)\} - \rho_\tau\{\epsilon(\tau)\}\right] \\
=\ & -\mathbb{G}_n\left[\mathbf{Z}_k^T\boldsymbol{\delta}_k(\tau)\psi_\tau\{\epsilon(\tau)\}\right] + \frac{1}{2}\boldsymbol{\delta}_k^T(\tau)\mathbf{D}_k(\tau)\boldsymbol{\delta}_k(\tau) + o_{p^*}(1).
\end{aligned}
$$

Under the local model with $\boldsymbol{\beta}_n(\tau) = n^{-1/2}\mathbf{b}_0(\tau)$, Lemma 2 suggests that $\hat{\boldsymbol{\delta}}_k(\tau) = O_p(1)$. Thus by the convexity lemma in Pollard (1991) and the stochastic equicontinuity of $(\tau, (\alpha, \boldsymbol{\beta})) \mapsto \mathbb{P}_n[\rho_\tau(Y - \alpha - \mathbf{X}^T\boldsymbol{\beta}) - \rho_\tau\{Y - \alpha_0(\tau) - \mathbf{X}^T\boldsymbol{\beta}_0(\tau)\}]$, we have

$$
\begin{aligned}
& n\mathbb{P}_n\left[\rho_\tau\{Y - \hat{\alpha}_k(\tau) - \hat{\theta}_k(\tau)X_k\} - \rho_\tau\{\epsilon(\tau)\}\right] \\
=\ & n\mathbb{P}_n\left[\rho_\tau\{\epsilon(\tau) - n^{-1/2}\mathbf{Z}_k^T\hat{\boldsymbol{\delta}}_k(\tau)\} - \rho_\tau\{\epsilon(\tau)\}\right] \\
=\ & -\mathbb{G}_n\left[\mathbf{Z}_k^T\hat{\boldsymbol{\delta}}_k(\tau)\psi_\tau\{\epsilon(\tau)\}\right] + 1/2\hat{\boldsymbol{\delta}}_k^T(\tau)\mathbf{D}_k(\tau)\hat{\boldsymbol{\delta}}_k(\tau) + o_{p^*}(1). \qquad \text{(S.14)}
\end{aligned}
$$

By (S.10), it is easy to show that

$$
\begin{aligned}
& n^{1/2}\begin{pmatrix}\hat{\alpha}_k(\tau) - \alpha_0(\tau) \\ \hat{\theta}_k(\tau)\end{pmatrix} = n^{1/2}\begin{pmatrix}\hat{\alpha}_k(\tau) - \alpha_0(\tau) \\ \hat{\theta}_k(\tau) - \boldsymbol{\beta}_{n,k}(\tau)\end{pmatrix} + \begin{pmatrix}0 \\ \mathbf{b}_{0,k}(\tau)\end{pmatrix} \\
=\ & \mathbf{J}_k^{-1}(\tau, \mathbf{0})\left[\mathbb{G}_n[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}] + \mathbf{A}_k(\tau)^T\mathbf{b}_{0,(-k)}(\tau) + \mathbf{J}_k(\tau, \mathbf{0})\begin{pmatrix}0 \\ b_{0,k}(\tau)\end{pmatrix}\right] + o_{p^*}(1) \\
=\ & \mathbf{J}_k^{-1}(\tau, \mathbf{0})\Big(\mathbb{G}_n[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}] + \mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau)\Big] + o_{p^*}(1).
\end{aligned}
$$

Thus

$$\hat{\boldsymbol{\delta}}_k(\tau) \quad = \quad n^{1/2}\begin{pmatrix}\hat{\alpha}_k(\tau)-\alpha_0(\tau)\\ \hat{\theta}_k(\tau)\\ -\boldsymbol{\beta}_n(\tau)\end{pmatrix} = \begin{pmatrix}\mathbf{J}_k^{-1}(\tau,\mathbf{0})\left(\mathbb{G}_n[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}]+\mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau)\right)\\ -\mathbf{b}_0(\tau)\end{pmatrix} + o_{p^*}(1).$$

Plugging $\hat{\boldsymbol{\delta}}_k(\tau)$ in (S.14), we get

$$\begin{aligned}
& n\mathbb{P}_n\left[\rho_\tau\{Y-\hat{\alpha}_k(\tau)-\hat{\theta}_k(\tau)X_k\}-\rho_\tau\{\epsilon(\tau)\}\right]\\
= \;& -\mathbb{G}_n\left[\tilde{\mathbf{X}}_k^T\psi_\tau\{\epsilon(\tau)\}\right]\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}\right]-\mathbb{G}_n\left[\psi_\tau\{\epsilon(\tau)\}\tilde{X}_k^T\right]\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{B}_k^T\mathbf{b}_0(\tau)\\
& +\mathbb{G}_n\left[\psi_\tau\{\epsilon(\tau)\}\mathbf{X}^T\right]\mathbf{b}_0(\tau)+\frac{1}{2}\mathbb{G}_n\left[\tilde{\mathbf{X}}_k^T\psi_\tau\{\epsilon(\tau)\}\right]\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}\right]\\
& -1/2\mathbf{b}_0^T(\tau)\left\{\mathbf{B}_k(\tau)\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{B}_k^T(\tau)-\mathbf{C}(\tau)\right\}\mathbf{b}_0(\tau)+o_p(1)\\
= \;& -1/2\mathbb{G}_n\left[\tilde{\mathbf{X}}_k^T\psi_\tau\{\epsilon(\tau)\}\right]\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbb{G}_n\left[\tilde{\mathbf{X}}_k\psi_\tau\{\epsilon(\tau)\}\right]\\
& -1/2\mathbf{b}_0^T(\tau)\left\{\mathbf{B}_k(\tau)\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{B}_k^T(\tau)-\mathbf{C}(\tau)\right\}\mathbf{b}_0(\tau)\\
& -\left(\mathbb{G}_n\left[\psi_\tau\{\epsilon(\tau)\}\tilde{\mathbf{X}}_k^T\right\}\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{B}_k^T(\tau)-\mathbb{G}_n\left\{\psi_\tau(\epsilon)\mathbf{X}^T\right]\right)\mathbf{b}_0(\tau)+o_p(1)\\
= \;& -1/2\mathbf{M}_k^T(\tau)\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{M}_k(\tau)-\{\mathbf{M}_k^T(\tau)\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{B}_k^T(\tau)-\mathbf{M}_{(2)}^T(\tau)\}\mathbf{b}_0(\tau)\\
& \qquad -1/2\mathbf{b}_0^T(\tau)\{\mathbf{B}_k\mathbf{J}_k^{-1}(\tau,\mathbf{0})\mathbf{B}_k^T(\tau)-\mathbf{C}(\tau)\}\mathbf{b}_0(\tau)+o_p(1), \qquad \text{(S.15)}
\end{aligned}$$

where $\mathbf{M}_k(\tau)=(M_k(\tau,\mathbf{0}),M_{p+k}(\tau,\mathbf{0}))^T$, and $\mathbf{M}_{(2)}(\tau)=(M_{p+1}(\tau,\mathbf{0}),\ldots,M_{2p}(\tau,\mathbf{0}))^T$. It is easy to see that the minimizer of the summation term on the right hand side of (S.15) with respect to k is equivalent to $K(\tau)$.

In addition, for all $j\neq k$, $\left(\mathbf{M}_j(\tau)+\mathbf{B}_j^T(\tau,\mathbf{0})\mathbf{b}_0(\tau)\right)^T\mathbf{J}_j^{-1}(\tau,\mathbf{0})\left(\mathbf{M}_j(\mathbf{0})+\mathbf{B}_j(\mathbf{0})^T\mathbf{b}_0(\tau)\right)\neq\left(\mathbf{M}_k(\tau)+\mathbf{B}_k(\tau)^T\mathbf{b}_0(\tau)\right)^T\mathbf{J}_k^{-1}(\tau,\mathbf{0})\left(\mathbf{M}_k(\tau)+\mathbf{B}_k(\tau)^T\mathbf{b}_0(\tau)\right)$ almost surely. Then it follows by Lemma 3 of McKeague and Qian (2015) that $K(\tau)$ is unique, a.s.. The lemma is thus proven by combining Lemma 3, (S.15), the uniform consistency of $\hat{\sigma}_k(\tau)$ and the continuity mapping theorem as used in the proof of Lemma 2 in McKeague and Qian (2015).

S.2. Proof of Corollary 1

For the homogenous case with constant error density, we have $f_{\epsilon(\tau)}(0|\mathbf{X})\equiv f_{\epsilon(\tau)}(0)$. Then

$$\mathbf{J}_k(\tau,\mathbf{0})=f_{\epsilon(\tau)}(0)\begin{pmatrix}1 & E(X_k)\\ E(X_k) & E(X_k^2)\end{pmatrix},\ \mathbf{J}_k^{-1}(\tau,\mathbf{0})=\frac{1}{f_{\epsilon(\tau)}(0)\mathrm{Var}(X_k)}\begin{pmatrix}E(X_k^2) & -E(X_k)\\ -E(X_k) & 1\end{pmatrix},$$

and $\mathbf{B}_k(\tau) = f_{\epsilon(\tau)}(0)E(\mathbf{X}\tilde{\mathbf{X}}_k^T)$. Therefore,

$$\begin{aligned}
& (\mathbf{M}_k(\tau) + \mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau))^T \mathbf{J}_k^{-1}(\tau, \mathbf{0})(\mathbf{M}_k(\tau) + \mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau)) \\
= \; & \frac{1}{f_{\epsilon(\tau)}(0)\mathrm{Var}(X_k)} \Big[\big\{ M_k^2(\tau)E(X_k^2) - 2M_k(\tau)M_{p+k}(\tau)E(X_k) + M_{p+k}^2(\tau) \big\} \\
& +2f_{\epsilon(\tau)}(0)\mathbf{b}_0^T(\tau)\big\{ M_k(\tau)\big(E(\mathbf{X})E(X_k^2) - E(X_k\mathbf{X})E(X_k)\big) + M_{p+k}(\tau)\mathrm{Cov}(X_k, \mathbf{X}) \big\} \\
& +f_{\epsilon(\tau)}^2(0)\mathbf{b}_0^T(\tau)\big\{ E(\mathbf{X})E(X_k^2) - E(X_k\mathbf{X})E(X_k) \big\} E(\mathbf{X}^T) \\
& \qquad +\mathrm{Cov}(X_k, \mathbf{X})E(X_k\mathbf{X}^T) \big\} \mathbf{b}_0(\tau) \Big],
\end{aligned}$$

where $M_j(\tau) = M_j(\tau, \mathbf{0})$, $j = 1, \ldots, 2p$. Note that $E(\mathbf{X})E(X_k^2) - E(X_k\mathbf{X})E(X_k) = E(\mathbf{X})\mathrm{Var}(X_k) - E(X_k)\mathrm{Cov}(X_k, \mathbf{X})$, and when $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, $M_1(\tau) = \cdots = M_p(\tau)$. Then

$$\begin{aligned}
& (\mathbf{M}_k(\tau) + \mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau))^T \mathbf{J}_k^{-1}(\tau, \mathbf{0})(\mathbf{M}_k(\tau) + \mathbf{B}_k^T(\tau)\mathbf{b}_0(\tau)) \\
= \; & \frac{1}{f_{\epsilon(\tau)}(0)\mathrm{Var}(X_k)} \Big[\{ M_{p+k}(\tau) - M_1(\tau)E(X_k) \}^2 + M_1^2(\tau)\mathrm{Var}(X_k) \\
& +2f_{\epsilon(\tau)}(0)\mathbf{b}_0^T(\tau)\mathrm{Cov}(X_k, \mathbf{X})\{ M_{p+k}(\tau) - M_1(\tau)E(X_k) \} \\
& +2f_{\epsilon(\tau)}(0)\mathbf{b}_0^T(\tau)M_1(\tau)E(\mathbf{X})\mathrm{Var}(X_k) \\
& +f_{\epsilon(\tau)}^2(0)\mathbf{b}_0^T(\tau)\{ E(\mathbf{X})E(\mathbf{X}^T)\mathrm{Var}(X_k) + \mathrm{Cov}(X_k, \mathbf{X})\mathrm{Cov}(X_k, \mathbf{X}^T) \}\mathbf{b}_0(\tau) \Big] \\
= \; & \frac{1}{f_{\epsilon(\tau)}(0)\mathrm{Var}(X_k)} \big\{ M_{p+k}(\tau) - M_1(\tau)E(X_k) + f_{\epsilon(\tau)}(0)\mathbf{b}_0^T(\tau)\mathrm{Cov}(X_k, \mathbf{X}) \big\}^2 \\
& +M_1^2(\tau)/f_{\epsilon(\tau)}(0) + 2\mathbf{b}_0^T(\tau)E(\mathbf{X})M_1(\tau) + f_{\epsilon(\tau)}(0)\mathbf{b}_0^T(\tau)E(\mathbf{X})E(\mathbf{X}^T)\mathbf{b}_0(\tau),
\end{aligned}$$

where $f_{\epsilon(\tau)}(0)$ and the last three terms do not depend on k. Thus

$$K(\tau) = \underset{k=1,\ldots,p}{\mathrm{argmax}} \big\{ M_{p+k}(\tau) - M_1 E(X_k) + f_{\epsilon(\tau)}(0)\mathbf{b}_0^T(\tau)\mathrm{Cov}(X_k, \mathbf{X}) \big\}^2 / \mathrm{Var}(X_k).$$

The limiting distribution of $n^{1/2}\{\hat{\theta}_n(\tau) - \theta_n\}/\hat{\sigma}_n(\tau)$ when $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$ can be simplified with some basic algebra. $\square$

S.3. Proof of Lemmas 5 and 6

LEMMA **5** *Suppose the assumptions in Theorem 1 hold. Then* $\hat{k}_n^*(\tau) \stackrel{P^M}{\to} k_0(\tau)$ *conditionally (on the data) a.s. and*

$$\frac{n^{1/2}(\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau))}{\hat{\sigma}_{n*}(\tau)} \stackrel{d}{\to} \frac{\{M_{p+k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\pi_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau)) - M_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\mu_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\}}{V_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\sigma_{k_0(\tau)}(\tau)}$$

for all $\tau \in \mathcal{T}$ *for which* $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$*, conditionally (on the data) in probability, where* $M_j\{\boldsymbol{\beta}_0(\tau)\} = M_j\{\tau, \boldsymbol{\beta}_0(\tau)\}$.

Proof: When $\boldsymbol{\beta}_0(\tau) \neq 0$, the local parameter $\mathbf{b}_0(\tau)$ in $\boldsymbol{\beta}_n(\tau)$ is negligible. For simplification, we prove the lemma under model (1). First note that by equation (13) in the main paper, $(\hat{\alpha}_k^*(\tau), \hat{\theta}_k^*(\tau))$ converges to $(\alpha_k(\tau), \theta_k(\tau))$ conditionally in probability. Thus, for $k = 1, \ldots, p$,

$$\begin{aligned}
&|\mathbb{P}_n^*\{\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k)\} - P\{\rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k)\}| \\
\leq &|(\mathbb{P}_n^* - \mathbb{P}_n)\{\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k)\}| + |(\mathbb{P}_n - P)\{\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k)\}| \\
&\qquad + |P[\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k) - \rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k)]| \\
\leq &\sup_{(\alpha,\theta)\in\Xi(\tau)} |(\mathbb{P}_n^* - \mathbb{P}_n)\{\rho_\tau(Y - \alpha - \theta X_k)\}| \\
&+ \sup_{(\alpha,\theta)\in\Xi(\tau)} |(\mathbb{P}_n - P)\{\rho_\tau(Y - \alpha - \theta X_k)\}| + o_{P_M}(1) \\
&+ |P[\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k) - \rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k)]| \\
= &|P[\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k) - \rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k)]| + o_{P_M}(1) \\
= &o_{P_M}(1)
\end{aligned}$$

conditionally in probability, where $\Xi(\tau) \in \mathbb{R}^2$ is a closed ball around $(\alpha_k(\tau), \theta_k(\tau))$, the second to last equality follows by the P-GC property of the function class $\{\rho_\tau(Y - \alpha - \theta X_k) : (\alpha, \theta) \in \Xi(\tau)\}$. The last equality follows by applying the bootstrap continuous mapping theorem to the function $g(\alpha, \theta) = P[\rho_\tau(Y - \alpha - \theta X_k)]$. Since the bootstrap estimate of

$k_0(\tau)$ satisfies $\hat{k}_n^*(\tau) = \underset{k}{\operatorname{argmin}}\, \mathbb{P}_n^*\{\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k)\}$, we have

$$\begin{aligned}
&P^M(\hat{k}_n^*(\tau) \neq k_0(\tau)) \\
=&P^M\left(\bigcup_{k:k\neq k_0(\tau)} \left\{\mathbb{P}_n^*\{\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k)\} < \mathbb{P}_n^*\{\rho_\tau(Y - \hat{\alpha}_{k_0(\tau)}^*(\tau) - \hat{\theta}_{k_0(\tau)}^*(\tau)X_{k_0})\}\right\}\right) \\
\leq& \sum_{k:k\neq k_0(\tau)} P^M\Big(\mathbb{P}_n^*\rho_\tau(Y - \hat{\alpha}_k^*(\tau) - \hat{\theta}_k^*(\tau)X_k) - P\rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k) \\
&\qquad + P\rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k) - P\rho_\tau(Y - \alpha_0(\tau) - \theta_0(\tau)X_{k_0}) \\
&\qquad < \mathbb{P}_n^*\rho_\tau(Y - \hat{\alpha}_{k_0}^*(\tau) - \hat{\theta}_{k_0}^*(\tau)X_{k_0}) - P\rho_\tau(Y - \alpha_0(\tau) - \theta_0(\tau)X_{k_0})\Big) \\
=& \sum_{k:k\neq k_0(\tau)} P^M\Big(P\{\rho_\tau(Y - \alpha_k(\tau) - \theta_k(\tau)X_k)\} - P\{\rho_\tau(Y - \alpha_0(\tau) - \theta_0(\tau)X_{k_0})\} < o_{P_M}(1)\Big)
\end{aligned}$$

which tends to zero in probability for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$, by the condition that $k_0(\tau)$ is unique. This together with the fact that $\hat{\sigma}_{k_0}(\tau) \xrightarrow{p} \sigma_{k_0(\tau)}(\tau)$ implies that $\hat{\sigma}_{n*}(\tau) \xrightarrow{P^M} \sigma_{k_0(\tau)}(\tau)$ conditionally in probability. Using equation (14) in the main paper and by bootstrap consistency of the sample mean, we have

$$\begin{aligned}
&\frac{\sqrt{n}(\hat{\theta}_n^*(\tau) - \hat{\theta}_n(\tau))}{\hat{\sigma}_{n*}(\tau)} \\
=&\frac{\sqrt{n}(\hat{\theta}_{\hat{k}_n^*(\tau)}^*(\tau) - \hat{\theta}_{k_0(\tau)}^*(\tau)) + \sqrt{n}(\hat{\theta}_{k_0(\tau)}^*(\tau) - \hat{\theta}_{k_0(\tau)}(\tau)) + \sqrt{n}(\hat{\theta}_{k_0(\tau)}(\tau) - \hat{\theta}_n(\tau))}{\hat{\sigma}_{n*}(\tau)} \\
=&\frac{\mathbf{J}_{k_0(\tau)}^{-1}(\tau, \boldsymbol{\beta}_0(\tau))\mathbb{G}_n^*[\psi_\tau(Y - \alpha_{k_0(\tau)}(\tau) - \theta_{k_0(\tau)}(\tau)X_{k_0(\tau)})\tilde{\mathbf{X}}_{k_0(\tau)}]}{\hat{\sigma}_{n*}(\tau)} + o_{P_M^*}(1) + o_p(1) \\
\xrightarrow{d}& \frac{\{M_{p+k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\pi_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau)) - M_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\mu_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\}}{V_{k_0(\tau)}(\boldsymbol{\beta}_0(\tau))\sigma_{k_0(\tau)}(\tau)}
\end{aligned}$$

for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) \neq \mathbf{0}$ conditionally in probability.

LEMMA **6** *Suppose all assumptions in Theorem 1 hold. Then $\mathbb{V}_n^*(\tau, \mathbf{b}_0(\tau))$ converges to the same limiting distribution as $\sqrt{n}(\hat{\theta}_n(\tau) - \theta_n(\tau))/\hat{\sigma}_n(\tau)$ for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$ conditionally (on the data) in probability.*

Proof: For any $\mathbf{b} \in \mathbb{R}^p$, define

$$\bar{k}_n(\tau, \mathbf{b}) = \underset{k=1,\ldots,p}{\operatorname{argmin}} \min_{\alpha,\theta} E[\rho_\tau\{\epsilon(\tau) + \alpha_0(\tau) + \mathbf{X}^T(\boldsymbol{\beta}_0(\tau) + n^{-1/2}\mathbf{b}) - \alpha - \theta X_k\}]. \quad \text{(S.16)}$$

Let $\mathbb{U}_n(\tau, \mathbf{b})$ and $L_n(\tau, \mathbf{b})$ be p-vectors, with the k-th components given by

$$\mathbb{U}_{n,k}(\tau, \mathbf{b}) = \left(\mathbb{G}_n\{\tilde{\mathbf{X}}_k\psi_\tau(\epsilon(\tau))\} + \widehat{\mathbf{B}}_k^T(\tau)\mathbf{b}\right)^T \widehat{\mathbf{J}}_k^{-1}(\tau) \left(\mathbb{G}_n\{\tilde{\mathbf{X}}_k\psi_\tau(\epsilon(\tau))\} + \widehat{\mathbf{B}}_k^T(\tau)\mathbf{b}\right)$$
$$\text{and } L_{n,k}(\tau, \mathbf{b}) = \min_{\alpha,\theta} E[\rho_\tau\{\epsilon(\tau) + \alpha_0(\tau) + \mathbf{X}^T(\boldsymbol{\beta}_0(\tau) + n^{-1/2}\mathbf{b}) - \alpha - \theta X_k\}],$$

respectively. Then $\bar{k}_n(\tau, \mathbf{b}) = \arg\min_k L_{n,k}(\tau, \mathbf{b})$. Let $\mathbb{W}_n(\tau, \mathbf{b})$ be a $p \times p$ matrix with the (j, k)-th element given by

$$\frac{(-\hat{\mu}_k(\tau), \hat{\pi}(\tau))\mathbb{G}_n\{\tilde{\mathbf{X}}_k\psi_\tau(\epsilon(\tau))\}}{\widehat{V}_k(\tau)\sigma_k(\tau)} + \left(\frac{\widehat{\mathbf{C}}_k(\tau)}{\widehat{V}_k(\tau)} - \frac{\widehat{\mathbf{C}}_j(\tau)}{\widehat{V}_j(\tau)}\right)^T \frac{\mathbf{b}}{\sigma_k(\tau)}.$$

Also let $\mathbb{D}_n(\tau, \mathbf{b})$ be a p-vector of zeros, apart from a 1 in the entry that maximizes $\mathbb{U}_n(\tau, \mathbf{b})$, and $D_n(\tau, \mathbf{b})$ be a p-vector of zeros, apart from a 1 in the $\bar{k}_n(\tau, \mathbf{b})$-th entry (i.e. the entry that minimizes $L_n(\tau, \mathbf{b})$). Then

$$\mathbb{V}_n(\tau, \mathbf{b}) = D_n(\tau, \mathbf{b})^T\mathbb{W}_n(\tau, \mathbf{b})\mathbb{D}_n(\tau, \mathbf{b}).$$

Similarly, define $\mathbb{U}(\tau, \mathbf{b})$, $\mathbb{W}(\tau, \mathbf{b})$, $\mathbb{D}(\tau, \mathbf{b})$ and $D(\tau, \mathbf{b})$ as processes of the same forms as $\mathbb{U}_n(\tau, \mathbf{b})$, $\mathbb{W}_n(\tau, \mathbf{b})$, $\mathbb{D}_n(\tau, \mathbf{b})$ and $D_n(\tau, \mathbf{b})$, except with $\mathbb{G}_n\{\tilde{\mathbf{X}}_k\psi_\tau(\epsilon(\tau))\}$ replaced by $\mathbf{M}_k(\tau)$, $\bar{k}_n(\tau, \mathbf{b})$ replaced by its limit $\kappa_\tau(\mathbf{b})$, and the sample variances/covariances replaced by their population versions. Lemma 4 implies that

$$\frac{n^{1/2}(\hat{\theta}_n(\tau) - \theta_n(\tau))}{\hat{\sigma}_n(\tau)} \xrightarrow{d} D(\tau, \mathbf{b}_0(\tau))^T\mathbb{W}(\tau, \mathbf{b}_0(\tau))\mathbb{D}(\tau, \mathbf{b}_0(\tau)) \quad \text{(S.17)}$$

for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$.

Let $\mathbb{D}_n^*(\tau, \mathbf{b})$ be a p-vector of zeros, apart from a 1 in the entry that maximizes $\mathbb{U}_n^*(\tau, \mathbf{b})$, and $D_n^*(\tau, \mathbf{b})$ be a p-vector of zeros, apart from a 1 in the entry that minimizes $\mathbb{L}_n(\tau, \mathbf{b})$, where $\mathbb{U}_n^*(\tau, \mathbf{b})$ and $\mathbb{L}_n(\tau, \mathbf{b})$ are defined at the beginning of the proof of Theorem 2. Let

$\mathbb{W}_n^*(\tau, \mathbf{b})$ be a $p \times p$ matrix with the (j,k)-th element given by

$$\frac{(-\hat{\mu}_k(\tau), \hat{\pi}(\tau))\mathbb{G}_n^*\{\tilde{\mathbf{X}}_k \psi_\tau(\hat{\epsilon}_n(\tau))\}}{\widehat{V}_k(\tau)\hat{\sigma}_k(\tau)} + \left(\frac{\widehat{\mathbf{C}}_k(\tau)}{\widehat{V}_k(\tau)} - \frac{\widehat{\mathbf{C}}_j(\tau)}{\widehat{V}_j(\tau)}\right)^T \frac{\mathbf{b}}{\hat{\sigma}_k(\tau)}.$$

Then $\mathbb{V}_n^*(\tau, \mathbf{b}_0(\tau)) = D_n^*(\tau, \mathbf{b}_0(\tau))^T \mathbb{W}_n^*(\tau, \mathbf{b}_0(\tau)) \mathbb{D}_n^*(\tau, \mathbf{b}_0(\tau))$.

Note that $\mathbb{G}_n^*\{\tilde{\mathbf{X}}_k \psi_\tau(\hat{\epsilon}_n(\tau))\}$ can be decomposed as

$$\mathbb{G}_n^*\{\tilde{\mathbf{X}}_k \psi_\tau(\hat{\epsilon}_n(\tau))\} = \mathbb{G}_n^*\{\tilde{\mathbf{X}}_k \psi_\tau(\epsilon(\tau))\} + \mathbb{G}_n^*\{\tilde{\mathbf{X}}_k[\psi_\tau(\hat{\epsilon}_n(\tau)) - \psi_\tau(\epsilon(\tau))]\}, \tag{S.18}$$

and $\hat{\epsilon}_n(\tau) = Y - \hat{\alpha}_n(\tau) - \hat{\theta}_n(\tau)X_{\hat{k}_n(\tau)} = \epsilon(\tau) + (\alpha_0(\tau) - \hat{\alpha}_n(\tau)) + \mathbf{X}^T\boldsymbol{\beta}_n(\tau) - \hat{\theta}_n(\tau)X_{\hat{k}_n}$. By bootstrap consistency of the sample mean,

$$\mathbb{G}_n^*\{\psi_\tau(\epsilon(\tau)), X_1\psi_\tau(\epsilon(\tau)), \ldots, X_p\psi_\tau(\epsilon(\tau))\}^T \xrightarrow{d} (M_1(\tau, \boldsymbol{\beta}_0), M_{p+1}(\tau, \boldsymbol{\beta}_0), \ldots, M_{2p}(\tau, \boldsymbol{\beta}_0))^T$$

conditionally in probability. To deal with the second term on the RHS of (S.18), define $\mathcal{F} = \{f(\mathbf{X}, \epsilon; \alpha, \boldsymbol{\beta}, \theta, k) = I(\epsilon < 0) - I\{\epsilon + (\alpha_0 - \alpha) + \mathbf{X}^T\boldsymbol{\beta} - \theta X_k < 0\} \colon (\alpha, \boldsymbol{\beta}, \theta, k) \in \mathbb{R}^{2+p} \times \{1, \ldots, p\}\}$. Then $\mathcal{F}$ is P-Donsker and $\mathbb{G}_n^*\{\tilde{\mathbf{X}}_k^T[\psi_\tau(\hat{\epsilon}_n(\tau)) - \psi_\tau(\epsilon(\tau))]\} = \mathbb{G}_n^*\big[\tilde{\mathbf{X}}_k^T f(\mathbf{X}, \epsilon(\tau); \hat{\alpha}_n(\tau), \boldsymbol{\beta}_n(\tau), \hat{\theta}_n(\tau), \hat{k}_n(\tau))\big]$. Note that $\hat{\alpha}_n(\tau) \xrightarrow{P} \alpha_0(\tau)$, $\boldsymbol{\beta}_n(\tau) = o(1)$, and $\hat{\theta}_n(\tau) = o_p(1)$. By Assumption A4, $\{g : (\alpha, \boldsymbol{\beta}, \theta, k) \to E[\tilde{\mathbf{X}}_k^T f(\mathbf{X}, \epsilon; \alpha, \boldsymbol{\beta}, \theta, k)]^2\}$ is continuous. Since $f(\mathbf{X}, \epsilon; \alpha_0, \mathbf{0}, 0, k) = 0$ for any $k \in \{1, \ldots, p\}$, we have

$$E[\tilde{\mathbf{X}}_k^T f(\mathbf{X}, \epsilon(\tau); \hat{\alpha}_n(\tau), \boldsymbol{\beta}_n(\tau), \hat{\theta}_n(\tau), \hat{k}_n(\tau))]^2 = o_p(1)$$

by the continuous mapping theorem. Using similar arguments as in the proof of van der Vaart (2000), we can show that $\mathbb{G}_n^*\{\tilde{\mathbf{X}}_k[\psi_\tau(\hat{\epsilon}_n(\tau)) - \psi_\tau(\epsilon(\tau))]\} = o_{P_M}(1)$ for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$, conditionally in probability.

By Slusky's Lemma and the continuous mapping Theorem,

$$(\mathbb{W}_n^*(\tau, \mathbf{b}_0(\tau)), \mathbb{U}_n^*(\tau, \mathbf{b}_0(\tau))) \xrightarrow{d} (\mathbb{W}(\tau, \mathbf{b}_0(\tau)), \mathbb{U}(\tau, \mathbf{b}_0(\tau)))$$

conditionally in probability. Using similar arguments to those at the end of the proof of Lemma 2 in McKeague and Qian (2015), along with the continuous mapping theorem, we have $\mathbb{D}_n^*(\tau, \mathbf{b}_0(\tau)) \xrightarrow{d} \mathbb{D}(\tau, \mathbf{b}_0(\tau))$, $D_n^*(\tau, \mathbf{b}_0(\tau)) \rightarrow D(\tau, \mathbf{b}_0(\tau))$ and

$$\mathbb{V}_n^*(\tau, \mathbf{b}_0(\tau)) = D_n^*(\tau, \mathbf{b}_0(\tau))^T \mathbb{W}_n^*(\tau, \mathbf{b}_0(\tau)) \mathbb{D}_n^*(\tau, \mathbf{b}_0(\tau)) \xrightarrow{d} D(\tau, \mathbf{b}_0(\tau))^T \mathbb{W}(\tau, \mathbf{b}_0(\tau)) \mathbb{D}(\tau, \mathbf{b}_0(\tau))$$

for all $\tau \in \mathcal{T}$ for which $\boldsymbol{\beta}_0(\tau) = \mathbf{0}$ conditionally in probability, This together with (S.17) implies the result.

S.4. Additional Simulation Results

Our proposed test statistic can be viewed as a maximum-type test statistic across p covariates. Similar to the discussion as in Chatterjee and Lahiri (2015) for mean regression, we may also consider an alternative statistic based on the sum of squared t-statistics, $\sum_{k=1}^{p} \hat{\theta}_k^2(\tau)/\hat{\sigma}_k^2(\tau)$. We conduct an additional simulation study to compare the performance of the bootstrap tests based on the maximum-type and the sum-type test statistics. The data are generated from the model: $Y = \mathbf{X}^T \boldsymbol{\beta} + \epsilon$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, $\epsilon \sim N(0, 1)$, and the covariate vector $\mathbf{X} = (X_1, \ldots, X_p)^T$ is from the multivariate normal distribution with mean zero, variance one and an exchangeable correlation of 0.5, truncated at -2 and 2. For power analysis, we set $\boldsymbol{\beta}$ in two different ways corresponding to sparse and dense signals respectively:

- (sparse): $\beta_1 = b$, and $\beta_j = 0$ for $j = 2, \ldots, p$;
- (dense): $\beta_j = b/p, j = 1, \ldots, p.$

Figure 1 plots the power curves of the bootstrap tests based on the maximum-type and sum-type test statistics for $p = 10$ and $p = 100$ in models with sparse and dense signals separately. We set the sample size as $n = 200$, and the nominal level as 0.05. The Type I errors from the bootstrap method for the sum-type statistic tend to be smaller than the nominal level. For fair comparison, we choose λ_n in the bootstrap procedure for the maximum-type statistic such that the resulting Type I errors are comparable to those of the sum-type test. The results in Figure 1 suggest that the maximum-type test is more powerful for detecting sparse signals, and the sum-type test has more power for detecting dense alternatives. This observation agrees with the findings in mean regression (Cai et al., 2014; Gregory et al., 2015; Chen and Qin, 2010; Fan et al., 2015).

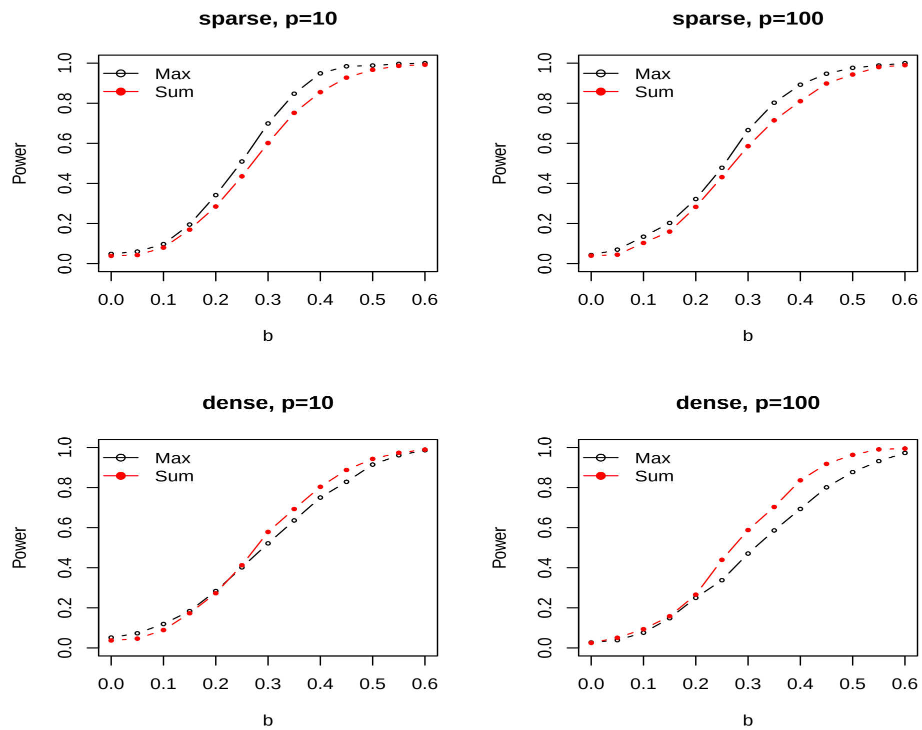

Figure 1: Power curves of the maximum-type and sum-type tests in models with sparse and dense signals.

References

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006) Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, **74**, 539–563.

Cai, T. T., Liu, W. and Xia, Y. (2014) Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society, Ser. B*, **76**, 349–372.

Chatterjee, A. and Lahiri, S. N. (2015) Comment on "an adaptive resampling test for detecting the presence of significant predictors" by I. McKeague and M. Qian. *Journal of American Statistical Association*, **110**, 1434–1438.

Chen, S. X. and Qin, Y. (2010) A two-sample test for high-dimensional data with applications to gene-set testing1. *The Annals of Statistics*, **38**, 808–835.

Fan, J., Liao, Y. and Yao, J. (2015) Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, **83**, 1497–1541.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V. and Lahiri, S. N. (2015) A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association*, **110**, 837–849.

McKeague, I. W. and Qian, M. (2015) An adaptive resampling test for detecting the presence of significant predictors. *Journal of American Statistical Association*, **110**, 1422–1433.

van der Vaart, A. W. (2000) *Asymptotic Statistics*. Cambridge: Cambridge University Press.