

Supplementary for “Semiparametric Efficient Estimation of Quantile Regression”

Zhanfeng Wang, Kani Chen, Yuanyuan Lin, Zhiliang Ying

School of Management, University of Science and Technology of China, Hefei, China

Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China

Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China

Department of Statistics, Columbia University, New York, NY 10027, United States

1. Proofs of main theorems

This section contains proofs of our main results.

Supplementary I: Proof of Theorem 1

For model (1.1), a semiparametric efficient score for $\beta(\tau)$ at any $\tau^* \in (0, 1)$ is derived by the least favorable submodel method (Begun *et al.*, 1983, Bickel *et al.*, 1993). Without loss of generality, we calculate the efficient score for the j -th component of $\beta(\tau)$ first. From van der Vaart (1998, section 25.4), we begin with the construction of a parametric submodel of

model (1.1) with parameter θ in a neighborhood of 0,

$$Q(\tau, \theta|X) = X^\top \boldsymbol{\beta}(\tau; \theta), \quad (\text{A.1})$$

where $\boldsymbol{\beta}(\tau; \theta) = \boldsymbol{\beta}_0(\tau) + \theta \mathbf{d}(\tau)$ and $\mathbf{d}(\tau)$ has a continuous derivative.

Let $\eta(t|X)$ be an inverse function in τ satisfying $X^\top \boldsymbol{\beta}(\tau; \theta) = t$. When $\theta = 0$, let $\eta(t|X) = \eta_0(t|X)$. From $t = X^\top \boldsymbol{\beta}(\tau; \theta)$, again, it follows that

$$\frac{\partial \tau}{\partial \theta} = - \frac{X^\top \mathbf{d}(\tau)}{X^\top \dot{\boldsymbol{\beta}}_0(\tau) + \theta X^\top \dot{\mathbf{d}}(\tau)}.$$

It follows from model (A.1) that the density function

$$f'(t; \theta|X)|_{t=X^\top \boldsymbol{\beta}(\tau; \theta)} = \frac{1}{X^\top \dot{\boldsymbol{\beta}}_0(\tau) + \theta X^\top \dot{\mathbf{d}}(\tau)}.$$

By Taylor expansion, we have

$$\begin{aligned} & f'(t; \theta|X)|_{t=X^\top \boldsymbol{\beta}(\tau; \theta)} \\ &= \frac{1}{X^\top \dot{\boldsymbol{\beta}}_0(\tau)} \Big|_{\tau=\eta_0(t|X)} + \theta g(t|X) \Big|_{t=X^\top \boldsymbol{\beta}_0(\tau)} + o(|\theta g(X^\top \boldsymbol{\beta}_0(\tau)|X)|), \end{aligned}$$

where

$$g(t|X) \Big|_{t=X^\top \boldsymbol{\beta}_0(\tau)} = \frac{X^\top \mathbf{d}(\tau) X^\top \ddot{\boldsymbol{\beta}}_0(\tau) - X^\top \dot{\mathbf{d}}(\tau) X^\top \dot{\boldsymbol{\beta}}_0(\tau)}{(X^\top \dot{\boldsymbol{\beta}}_0(\tau))^3}.$$

Observe that

$$f(X^\top \boldsymbol{\beta}_0(\tau)|X) = \frac{1}{X^\top \dot{\boldsymbol{\beta}}_0(\tau)}.$$

Then,

$$\begin{aligned} & f'(t; \theta|X)|_{t=X^\top \beta(\tau; \theta)} \\ &= f(t|X)|_{t=X^\top \beta_0(\tau)} + \theta g(t|X)|_{t=X^\top \beta_0(\tau)} + o(|\theta g(X^\top \beta_0(\tau)|X)|). \end{aligned}$$

Denoted by $G(t|X) = \int_{-\infty}^t g(u|X)du$. Then we have

$$G(X^\top \beta_0(\tau)|X) = -\frac{X^\top \mathbf{d}(\tau)}{X^\top \dot{\beta}_0(\tau)}. \quad (\text{A.2})$$

To guarantee $f'(t; \theta|X)$ being a density function of t for all θ , we need condition for $G(t|X)$,

$$\lim_{\tau \rightarrow 0} G(X^\top \beta_0(\tau)|X) = \lim_{\tau \rightarrow 1} G(X^\top \beta_0(\tau)|X) = 0,$$

that is,

$$\lim_{\tau \rightarrow 0} \frac{X^\top \mathbf{d}(\tau)}{X^\top \dot{\beta}_0(\tau)} = \lim_{\tau \rightarrow 1} \frac{X^\top \mathbf{d}(\tau)}{X^\top \dot{\beta}_0(\tau)} = 0. \quad (\text{A.3})$$

As $\mathbf{d}(\tau)$ is independent of X , equation (A.3) holds for all X . Then, $\mathbf{d}(\tau)$ can be obtained by solving a second-order differential equations. Without loss of generality, let

$$\mathbf{d}(0) = \mathbf{d}(1) = 0. \quad (\text{A.4})$$

Consider the parametric submodel (A.1) subject to constraint (A.4). It is well-known that $\text{Var}(\beta_j(\tau^*; \hat{\theta}))$ can be generally expressed as $d_j(\tau^*)^2 \text{Var}(\hat{\theta})$,

where $d_j(\tau^*)$ and $\beta_j(\tau^*; \hat{\theta})$ are the j -th components of $\mathbf{d}(\tau^*)$ and $\boldsymbol{\beta}(\tau^*; \hat{\theta})$ respectively, and $\hat{\theta}$ is an estimator of θ . When $d_j(\tau^*)$ equals to 1, $\text{Var}(\beta_j(\tau^*; \hat{\theta}))$ can be approximated by $\text{Var}(\hat{\theta})$. Thus, we set $d_j(\tau^*) = 1$.

It follows from (A.2) that

$$\begin{aligned} g(X^\top \boldsymbol{\beta}_0(\tau)|X)X^\top \dot{\boldsymbol{\beta}}_0(\tau) &= -\frac{\partial}{\partial \tau} \left\{ \frac{X^\top \mathbf{d}(\tau)}{X^\top \dot{\boldsymbol{\beta}}_0(\tau)} \right\} \\ &= -\frac{\partial \{f(X^\top \boldsymbol{\beta}_0(\tau)|X)X^\top \mathbf{d}(\tau)\}}{\partial \tau}. \end{aligned}$$

Then,

$$\left. \frac{\partial \log f'(t; \theta|X)}{\partial \theta} \right|_{\theta=0} = \frac{g(t|X)}{f(t|X)} = -\frac{\partial}{\partial \tau} \left\{ f(X^\top \boldsymbol{\beta}_0(\tau)|X)X^\top \mathbf{d}(\tau) \right\} \Big|_{\tau=\eta_0(t|X)} \equiv \xi.$$

The semiparametric efficient score is of $\boldsymbol{\beta}_j(\tau^*)$,

$$S_j(Y, X; \tau^*) = -\frac{\partial}{\partial \tau} \left\{ f(X^\top \boldsymbol{\beta}(\tau)|X)X^\top \mathbf{d}(\tau) \right\} \Big|_{\tau=\eta(Y|X)}, \quad (\text{A.5})$$

where $\mathbf{d}$ is a minimizer of

$$\mathcal{I}(\mathbf{d}) \equiv E \left(\left[\frac{\partial}{\partial \tau} \left\{ f(X^\top \boldsymbol{\beta}(\tau)|X)X^\top \mathbf{d}(\tau) \right\} \Big|_{\tau=\eta(Y|X)} \right]^2 \right)$$

over $\mathbf{d}(\cdot) = (d_1(\cdot), \dots, d_p(\cdot))^\top$ subject to $d_j(\tau^*) = 1$ and $\mathbf{d}(0) = \mathbf{d}(1) = 0$.

Since ξ is the score function of the parameter θ for the parametric submodel (A.1), the Fisher information is $\mathcal{I}(\mathbf{d}) = E(\xi^2)$, then the Cramér-Rao lower bound is $1/\mathcal{I}(\mathbf{d})$. As a result, it provides the semiparametric efficient score and the efficiency lower bound $1/\mathcal{I}(\mathbf{d})$ (Bickel et al., 1993).

Existence of $\mathbf{d}(\tau)$. One solution of $\mathbf{d}(\tau)$ can be obtained via minimizing the objection function $\mathcal{I}(\mathbf{d})$ subject to $d_j(\tau^*) = 1$ and $\mathbf{d}(0) = \mathbf{d}(1) = 0$. For notational convenience, we consider model (1.1) with $p = 2$, where $X = (1, \tilde{X})^\top$ and $\boldsymbol{\beta}_0(\tau) = (\beta_{10}(\tau), \beta_{20}(\tau))^\top$. Here, we assume that $\mathbf{d}(\tau) = (d_1(\tau), d_2(\tau))^\top$ has continuous derivatives up to the second order. Denoted by $B_X(\tau) = X^\top \mathbf{d}(\tau) / (X^\top \dot{\boldsymbol{\beta}}_0(\tau))$. Then, we have

$$\mathcal{I}(\mathbf{d}) = E\{(\dot{B}_X(\tau))^2|_{\tau=\eta_0(Y|X)}\} = E\left\{\int_0^1 (\dot{B}_X(\tau))^2 d\tau\right\}, \quad (\text{A.6})$$

where

$$\begin{aligned} \dot{B}_X(\tau) &= \frac{X^\top \dot{\mathbf{d}}(\tau) X^\top \dot{\boldsymbol{\beta}}_0(\tau) - X^\top \mathbf{d}(\tau) X^\top \ddot{\boldsymbol{\beta}}_0(\tau)}{(X^\top \dot{\boldsymbol{\beta}}_0(\tau))^2} \\ &= \frac{X^\top \dot{\mathbf{d}}(\tau) X^\top \dot{\boldsymbol{\beta}}_0(\tau) - X^\top \int_0^\tau \dot{\mathbf{d}}(u) du X^\top \ddot{\boldsymbol{\beta}}_0(\tau)}{(X^\top \dot{\boldsymbol{\beta}}_0(\tau))^2}. \end{aligned}$$

Let $\mathcal{I}'(\dot{\mathbf{d}}) = \mathcal{I}(\mathbf{d})$. Define a Hilbert space

$$\mathcal{H} = \{\dot{\mathbf{d}} \in L_2(0, 1) : \mathcal{I}'(\dot{\mathbf{d}}) < \infty, d_j(\tau^*) = 1, \mathbf{d}(0) = \mathbf{d}(1) = 0\},$$

with an inner product, for $\mathbf{v}, \mathbf{w} \in \mathcal{H}$,

$$\langle \mathbf{v}, \mathbf{w} \rangle = E\left\{\int_0^1 \frac{\partial}{\partial \tau} \left(\frac{X^\top \int_0^\tau \mathbf{v}(u) du}{X^\top \dot{\boldsymbol{\beta}}_0(\tau)} \right) \frac{\partial}{\partial \tau} \left(\frac{X^\top \int_0^\tau \mathbf{w}(u) du}{X^\top \dot{\boldsymbol{\beta}}_0(\tau)} \right) d\tau\right\},$$

where $L_2(0, 1)$ is square integratable space of functions in $(0, 1)$. The inner product induces a norm, for $\mathbf{v} \in \mathcal{H}$,

$$\|\mathbf{v}\| = (\mathcal{I}'(\mathbf{v}))^{1/2}.$$

It follows from Banach-Alaoglu Theorem that the closed ball with $0 < t < \infty$,

$$B_t = \{\mathbf{v} \in \mathcal{H} : \mathcal{I}'(\mathbf{v}) \leq t\}$$

is weak compact. Thence, for any convergence sequence $V = \{\mathbf{v}_l \in B_t, l = 1, 2, \dots\}$ with $\mathcal{I}'(\mathbf{v}_l)$ converging to a constant as $l \rightarrow \infty$, there exists a subsequence $\{\mathbf{w}_l \in V, l = 1, 2, \dots\}$ such that $\mathbf{w}_l$ converges in distribution to $\dot{\mathbf{d}}' \in B_t$. From the lower semicontinuity of $\mathcal{I}'(\dot{\mathbf{d}}')$, we have

$$\mathcal{I}'(\dot{\mathbf{d}}') \leq \liminf_{l \rightarrow \infty} \mathcal{I}'(\{\mathbf{w}_l\}),$$

which validates the existence of $\mathbf{d}(\tau) = \dot{\mathbf{d}}'(\tau)$.

A special case with $p = 1$. When there is only an intercept term or one-dimensional covariate $\tilde{X}$ without intercept in model (1.1), that is $X = 1$ or $\tilde{X}$, it shows

$$B_X(\tau) = \frac{d(\tau)}{\dot{\beta}_0(\tau)},$$

$$\mathcal{I}(d) = E\{(\dot{B}_X(\tau))^2|_{\tau=\eta_0(Y|X)}\} = E\{(\frac{\partial}{\partial \tau}\{\frac{d(\tau)}{\dot{\beta}_0(\tau)}\})^2|_{\tau=\eta_0(Y|X)}\}.$$

Considering the Fréchet derivative of $\mathcal{I}(d)$ with respect to d , we obtain

$$\ddot{B}_X(\tau) = 0 \quad \text{for all } \tau \neq \tau^*. \quad (\text{A.7})$$

Since $d(\tau^*) = 1$, it follows from (A.7) that

$$B_X(\tau) = \frac{d(\tau)}{\dot{\beta}_0(\tau)} = \begin{cases} a_1\tau + a_2, & \tau \leq \tau^*; \\ a_3\tau + a_4, & \tau > \tau^*. \end{cases} \quad (\text{A.8})$$

In view of $d(0) = d(1) = 0$, we have $a_2 = 0$ and $a_3 = -a_4$. In addition, due to $d(\tau^*) = 1$, it is not hard to check that $a_1\tau^* = 1/\dot{\beta}_0(\tau^*) = a_3(\tau^* - 1)$, which implies

$$a_1 = \frac{1}{\dot{\beta}_0(\tau^*)\tau^*}, \quad a_3 = \frac{1}{\dot{\beta}_0(\tau^*)(\tau^* - 1)}.$$

As a result,

$$\frac{d(\tau)}{\dot{\beta}_0(\tau)} = \begin{cases} \frac{\tau}{\dot{\beta}_0(\tau^*)\tau^*}, & \tau \leq \tau^*; \\ \frac{1-\tau}{\dot{\beta}_0(\tau^*)(1-\tau^*)}, & \tau > \tau^*, \end{cases}$$

and the lower bound is

$$\sigma_1^2(\tau^*) = \frac{1}{\mathcal{I}(d)} = \dot{\beta}_0(\tau^*)^2\tau^*(1-\tau^*).$$

Supplementary II

This section contains Lemmas 1, 2 and 3 and their proofs. More notations are needed. Let $\mathbf{U}$ be a $pL \times pL$ matrix satisfying

$$\mathbf{v}^\top \mathbf{U} \mathbf{v} \equiv \sum_{l=1}^{L+1} \frac{E\{f(X^\top \boldsymbol{\beta}_0(\tau_{l-1})|X)X^\top \mathbf{v}_{l-1} - f(X^\top \boldsymbol{\beta}_0(\tau_l)|X)X^\top \mathbf{v}_l\}^2}{\tau_l - \tau_{l-1}},$$

for any $\mathbf{v} = (\mathbf{v}_1^\top, \mathbf{v}_2^\top, \dots, \mathbf{v}_L^\top)^\top$, $\mathbf{v}_i \in R^p$, $i = 0, \dots, L+1$. Let $\mathbf{W}$ be a diagonal matrix with diagonal components the same as $1/\text{diag}(\mathbf{U}^{-1})$.

Denote $\mathbf{U}^* = \mathbf{U}^{-1}\mathbf{W}$. Let $\mathbf{D}_l^*$ be a $p \times p$ matrix, $l = 0, \dots, L+1$, satisfying $\mathbf{D}_0^* = \mathbf{D}_{L+1}^* = 0$, $(\mathbf{D}_1^*, \mathbf{D}_2^*, \dots, \mathbf{D}_L^*) = (\mathbf{u}_{k1}, \mathbf{u}_{k2}, \dots, \mathbf{u}_{kp})^\top$, where $\mathbf{u}_{kj}$ is the $\{(k-1)p+j\}$ -th column of matrix $\mathbf{U}^*$. For a quantile level τ^* of interest, let $\tau_k = \tau^*$ for some $1 \leq k \leq L$.

Lemma 1. *Under model (2.9), the semiparametric efficient score for $\boldsymbol{\beta}(\tau^*)$ is*

$$\begin{aligned} \mathbf{S}^*(Y, X; \tau^*) = & \sum_{l=1}^{L+1} \frac{f(X^\top \boldsymbol{\beta}(\tau_{l-1})|X)X^\top \mathbf{D}_{l-1}^* - f(X^\top \boldsymbol{\beta}(\tau_l)|X)X^\top \mathbf{D}_l^*}{\tau_l - \tau_{l-1}} \\ & \left[I\{X^\top \boldsymbol{\beta}(\tau_{l-1}) < Y < X^\top \boldsymbol{\beta}(\tau_l)\} - (\tau_l - \tau_{l-1}) \right]. \quad (\text{A.9}) \end{aligned}$$

Moreover, for the estimator of the j -th component of $\boldsymbol{\beta}(\tau^*)$, its variance has a lower bound

$$\sigma_j^{*2}(\tau^*) = \frac{1}{\mathbf{u}_{kj}^\top \mathbf{U} \mathbf{u}_{kj}}, \quad j = 1, 2, \dots, p. \quad (\text{A.10})$$

Proof of Lemma 1. For model (2.9), a semiparametric efficient score for $\boldsymbol{\beta}(\tau^*)$ is calculated by the least favorable submodel. Firstly, the efficient score for the j -th component of $\boldsymbol{\beta}(\tau^*)$ is derived. Similar to proof of Theorem 1, a parametric submodel of model (2.9) is constructed with parameter θ in a neighborhood of 0,

$$Q(\tau_l, \theta|X) = X^\top \boldsymbol{\beta}(\tau_l; \theta), \quad l = 1, \dots, L, \quad (\text{A.11})$$

where $\beta(\tau_l; \theta) = \beta_0(\tau_l) + \theta \mathbf{d}(\tau_l)$, $d_j(\tau^*) = 1$, and

$$G(X^\top \beta_0(\tau_l)|X) = -f(X^\top \beta_0(\tau_l)|X)X^\top \mathbf{d}(\tau_l), \quad l = 1, \dots, L. \quad (\text{A.12})$$

For the density function $f'(t|X)$, we can show

$$\left. \frac{\partial \log f'(t|X)}{\partial \theta} \right|_{\theta=0} = \frac{g(t|X)}{f(t|X)} \equiv \xi.$$

Letting $\mathbf{d}(\tau_0) = \mathbf{d}(\tau_{L+1}) = 0$ and by the Cauchy-Schwarz inequality, we have

$$E(\xi^2|X) \geq \sum_{l=1}^{L+1} \frac{\{f(X^\top \beta_0(\tau_{l-1})|X)X^\top \mathbf{d}(\tau_{l-1}) - f(X^\top \beta_0(\tau_l)|X)X^\top \mathbf{d}(\tau_l)\}^2}{\tau_l - \tau_{l-1}}.$$

Write

$$\begin{aligned} E(\xi^2) &= E(E(\xi^2|X)) \\ &\geq \sum_{l=1}^{L+1} \frac{E\{f(X^\top \beta_0(\tau_{l-1})|X)X^\top \mathbf{d}(\tau_{l-1}) - f(X^\top \beta_0(\tau_l)|X)X^\top \mathbf{d}(\tau_l)\}^2}{\tau_l - \tau_{l-1}} \\ &= \tilde{\mathbf{d}}^\top \mathbf{U} \tilde{\mathbf{d}}, \end{aligned}$$

where $\tilde{\mathbf{d}} = (\mathbf{d}(\tau_1)^\top, \mathbf{d}(\tau_2)^\top, \dots, \mathbf{d}(\tau_L)^\top)^\top$, and the equality holds if and only if

$$\frac{g(t|X)}{f^{1/2}(t|X)} = \sum_{l=1}^{L+1} \{a_l f^{1/2}(t|X) + b_l\} I\{X^\top \beta_0(\tau_{l-1}) < t < X^\top \beta_0(\tau_l)\}.$$

It follows from the constraint (A.12) that for $l = 1, \dots, L+1$,

$$b_l = 0, \quad a_l = \frac{\{f(X^\top \beta_0(\tau_{l-1})|X)\mathbf{d}(\tau_{l-1}) - f(X^\top \beta_0(\tau_l)|X)\mathbf{d}(\tau_l)\}^\top X}{\tau_l - \tau_{l-1}}.$$

Therefore, the semiparametric efficient score for $\beta_j(\tau^*)$ is

$$\begin{aligned} & \frac{g(t|X)}{f(t|X)} \Big|_{t=Y} \\ &= \sum_{l=1}^{L+1} \frac{f(X^\top \beta_0(\tau_{l-1})|X)X^\top \mathbf{d}(\tau_{l-1}) - f(X^\top \beta_0(\tau_l)|X)X^\top \mathbf{d}(\tau_l)}{\tau_l - \tau_{l-1}} \\ & \quad \times \left\{ I(X^\top \beta_0(\tau_{l-1}) < Y < X^\top \beta_0(\tau_l)) - (\tau_l - \tau_{l-1}) \right\}, \end{aligned}$$

where $\tilde{\mathbf{d}} = (\mathbf{d}(\tau_1)^\top, \mathbf{d}(\tau_2)^\top, \dots, \mathbf{d}(\tau_L)^\top)^\top$ such that $E(\xi^2)$ is minimized, which is equivalent to minimizing the quadratic function $\tilde{\mathbf{d}}^\top \mathbf{U} \tilde{\mathbf{d}}$ subject to $d_j(\tau^*) = 1$. By the Lagrange multiplier method, we have

$$L(\tilde{\mathbf{d}}, \lambda) = \tilde{\mathbf{d}}^\top \mathbf{U} \tilde{\mathbf{d}} + \lambda \{d_j(\tau^*) - 1\}.$$

Setting $\partial L(\tilde{\mathbf{d}}, \lambda) / \partial \tilde{\mathbf{d}} = 0$ and $d_j(\tau^*) = 1$ leads to

$$\tilde{\mathbf{d}} = \mathbf{u}_{kj},$$

where $\mathbf{u}_{kj}$ is the $\{(k-1)p + j\}$ -th column of matrix $\mathbf{U}^*$. As a result, for

$1 \leq j \leq p$, the semiparametric efficient score for $\beta_j(\tau^*)$ is

$$\begin{aligned} S_j^*(Y, X; \tau^*) &= \sum_{l=1}^{L+1} \frac{f(X^\top \beta(\tau_{l-1})|X)X^\top \mathbf{d}(\tau_{l-1}) - f(X^\top \beta(\tau_l)|X)X^\top \mathbf{d}(\tau_l)}{\tau_l - \tau_{l-1}} \\ & \quad \times [I\{X^\top \beta(\tau_{l-1}) < Y < X^\top \beta(\tau_l)\} - (\tau_l - \tau_{l-1})], \end{aligned}$$

with $\tilde{\mathbf{d}} = \mathbf{u}_{kj}$. Hence, the semiparametric efficient score for $\beta(\tau^*)$ is

$$\begin{aligned} \mathbf{S}^*(Y, X; \tau^*) &= \sum_{l=1}^{L+1} \frac{f(X^\top \beta(\tau_{l-1})|X)X^\top \mathbf{D}_{l-1}^* - f(X^\top \beta(\tau_l)|X)X^\top \mathbf{D}_l^*}{\tau_l - \tau_{l-1}} \\ & \quad \times [I\{X^\top \beta(\tau_{l-1}) < Y < X^\top \beta(\tau_l)\} - (\tau_l - \tau_{l-1})]. \end{aligned}$$

The proof of Lemma 1 is complete.

Lemma 2. *Under model (2.9), as $L \rightarrow \infty$ and the maximum of $\{\tau_l - \tau_{l-1}, l = 1, 2, \dots, L+1\}$ tends to 0, for $j = 1, \dots, p$,*

$$S_j^*(Y, X; \tau^*) \rightarrow S_j(Y, X; \tau^*) \quad \text{in probability,}$$

and

$$\sigma_j^{*2}(\tau^*) \rightarrow \sigma_j^2(\tau^*).$$

Proof of Lemma 2. The efficient score in (2.10) can be rewritten as

$$S_j^*(Y, X; \tau^*) = \sum_{l=1}^{L+1} \frac{f(X^\top \boldsymbol{\beta}(\tau_{l-1})|X)X^\top \mathbf{d}(\tau_{l-1}) - f(X^\top \boldsymbol{\beta}(\tau_l)|X)X^\top \mathbf{d}(\tau_l)}{\tau_l - \tau_{l-1}} I\{X^\top \boldsymbol{\beta}(\tau_{l-1}) < Y < X^\top \boldsymbol{\beta}(\tau_l)\}, \quad (\text{A.13})$$

where $S_j^*(Y, X; \tau^*)$ is the j -th component of $\mathbf{S}^*(Y, X; \tau^*)$, and

$\tilde{\mathbf{d}} = [\mathbf{d}(\tau_1)^\top, \dots, \mathbf{d}(\tau_L)^\top]^\top$ is a minimizer of the quadratic form

$$E[\{S_j^*(Y, X; \tau^*)\}^2] \equiv \tilde{\mathbf{d}}^\top \mathbf{U} \tilde{\mathbf{d}}$$

subject to $d_j(\tau^*) = 1$. Thus, Lemma 2 can be proved in a straightforward fashion by simple algebra.

Let $\hat{\boldsymbol{\beta}}^c(\tau)$ be the classical Koenker-Bassett regression quantile estimator of $\boldsymbol{\beta}_0(\tau)$ at any given quantile level τ and let h be the bandwidth for the estimation of $\dot{\boldsymbol{\beta}}_0(\tau)$, the derivative of $\boldsymbol{\beta}_0(\tau)$.

Hereafter, mathematical operators of vectors or matrices A and B , such as $A + B$ and A/B , stand for the operators of the corresponding elements or components of A and B .

Next, we intend to show that the resulting estimator $\hat{\beta}_j(\tau^*)$ achieves the semiparametric efficiency lower bound.

Lemma 3. *Under model (2.9), assume that assumptions $(A_1) - (A_3)$ hold.*

Then, for $j = 1, 2, \dots, p$,

$$\sqrt{n} \left\{ \hat{\beta}_j(\tau^*) - \beta_{0j}(\tau^*) \right\} \rightarrow N(0, \sigma_j^{*2}(\tau^*)) \quad (\text{A.14})$$

in distribution as $n \rightarrow \infty$, where $\beta_{0j}(\tau^)$ is the j -th component of $\beta_0(\tau^*)$.*

Moreover, the asymptotic variance of $\hat{\beta}_j(\tau^)$ achieves the semiparametric efficiency bound $\sigma_j^{*2}(\tau^*)$.*

Proof of Lemma 3. We prove this lemma in four steps.

Step 1. To prove

$$\sup_{-M_\epsilon < t < M_\epsilon} \left| \hat{f}(t|X) - f(t|X) \right| = O_p \left(\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right), \quad (\text{A.15})$$

where $\hat{f}(t|X) = 1/\{X^\top \hat{\beta}(\tau)\}$ with $t = X^\top \beta(\tau)$ for any fixed $\epsilon \leq \tau \leq 1 - \epsilon$,

$$\hat{\beta}(\tau) \equiv \frac{\hat{\beta}^c(\tau + h) - \hat{\beta}^c(\tau - h)}{2h},$$

and M_ϵ is certain constant large enough depending on ϵ and M . To this

end, first, standard approximation by the Taylor expansion gives

$$\frac{\beta_0(\tau + h) - \beta_0(\tau - h)}{2h} = \dot{\beta}_0(\tau) + O(h^2).$$

From Portnoy(2012, page 1733), condition (A2) and the boundedness of X in assumption (A1), we have

$$\sup_{\epsilon < \tau < 1-\epsilon} \left| \frac{1}{X^\top \hat{\beta}(\tau)} - \frac{1}{X^\top \dot{\beta}_0(\tau)} \right| = O_p \left(\frac{1}{\sqrt{nh^2}} + \frac{(\log n)^{3/2}}{nh} + h^2 \right),$$

which implies that (A.15) holds.

Step 2. Let $\hat{\mathbf{D}}_l^*$ be the estimator of $\mathbf{D}_l^*$ by replacing $\beta(\tau)$ and f with $\hat{\beta}^c(\tau)$ and $\hat{f}$ for $1 \leq l \leq L$, respectively. Similar to the proof of Step 1, it can be shown that

$$\sup_{1 \leq l \leq L} \left| \hat{\mathbf{D}}_l^* - \mathbf{D}_l^* \right| = O_p \left(\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right), \quad (\text{A.16})$$

and

$$\sup_{2 \leq l \leq L+1} \left| \frac{\mathbf{D}_l^*}{X^\top \dot{\beta}_0(\tau_l)} - \frac{\hat{\mathbf{D}}_{l-1}^*}{X^\top \hat{\beta}(\tau_l)} \right| = O_p \left(\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right). \quad (\text{A.17})$$

Step 3. To evaluate the estimated score function $\mathbf{S}^*(Y, X; \tau^*)$ in Lemma 1 by plugging in the estimates $\hat{f}(\cdot|X)$, $\hat{\mathbf{D}}_l^*$ and $\hat{\beta}^c(\tau_l)$, $l = 1, \dots, L$. To be

concise, we define

$$\hat{\mathbf{S}}^*(Y_i, X_i; \tau^*) \equiv \sum_{l=1}^{L+1} \frac{\hat{a}_{(l-1)i} - \hat{a}_{li}}{\tau_l - \tau_{l-1}} \left[I\{X_i^\top \hat{\boldsymbol{\beta}}^c(\tau_{l-1}) < Y_i < X_i^\top \hat{\boldsymbol{\beta}}^c(\tau_l)\} - (\tau_l - \tau_{l-1}) \right],$$

where $\hat{a}_{li} = \frac{X_i^\top \hat{\mathbf{D}}_l^*}{X_i^\top \hat{\boldsymbol{\beta}}(\tau_l)}$ and $a_{li} = \frac{X_i^\top \mathbf{D}_l^*}{X_i^\top \boldsymbol{\beta}_0(\tau_l)}$. By the monotonicity implied by the quantile regression model, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{S}}^*(Y_i, X_i; \tau^*) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{L+1} \frac{a_{(l-1)i} - a_{li}}{\tau_l - \tau_{l-1}} \times [I\{X_i^\top \boldsymbol{\beta}_0(\tau_{l-1}) < Y_i < X_i^\top \boldsymbol{\beta}_0(\tau_l)\} - (\tau_l - \tau_{l-1})] \\ &+ \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{L+1} \frac{(a_{(l-1)i} - a_{li})(\hat{\Delta}_{l-1}^i - \hat{\Delta}_l^i)}{\tau_l - \tau_{l-1}} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{L+1} \frac{(\hat{a}_{(l-1)i} - a_{(l-1)i}) - (\hat{a}_{li} - a_{li})}{\tau_l - \tau_{l-1}} \times (\hat{\Delta}_{l-1}^i - \hat{\Delta}_l^i) \\ &+ \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{L+1} \frac{(\hat{a}_{(l-1)i} - a_{(l-1)i}) - (\hat{a}_{li} - a_{li})}{\tau_l - \tau_{l-1}} \\ &\quad \times [I\{X_i^\top \boldsymbol{\beta}_0(\tau_{l-1}) < Y_i < X_i^\top \boldsymbol{\beta}_0(\tau_l)\} - (\tau_l - \tau_{l-1})] \\ &\equiv \frac{1}{n} \sum_{i=1}^n \mathbf{S}^0(Y_i, X_i; \tau^*) + \Pi_1 + \Pi_2 + \Pi_3, \end{aligned}$$

where $\hat{\Delta}_l^i = I\{X_i^\top \boldsymbol{\beta}_0(\tau_l) < Y_i < X_i^\top \hat{\boldsymbol{\beta}}^c(\tau_l)\}$. Clearly, $E(\hat{\Delta}_{l-1}^i) \neq 0$. Hence,

we can show

$$\sup_{1 \leq l \leq L} |\hat{\Delta}_l^i| = O_p \left(\frac{1}{\sqrt{nh^2}} + \frac{(\log n)^{3/2}}{nh} + h^2 \right),$$

which indicates that

$$\begin{aligned}\Pi_1 &= \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{L+1} \frac{(a_{(l-1)i} - a_{li})(\hat{\Delta}_{l-1}^i - \hat{\Delta}_l^i)}{\tau_l - \tau_{l-1}} \\ &= O_p \left(L \left[\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right] \right),\end{aligned}\quad (\text{A.18})$$

uniformly in i . Similarly, it can be shown that, uniformly in i ,

$$\Pi_2 = O_p \left(L \left[\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right]^2 \right), \quad (\text{A.19})$$

and

$$\Pi_3 = O_p \left(\frac{L}{\sqrt{n}} \left[\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right] \right). \quad (\text{A.20})$$

Combining (A.18), (A.19) and (A.20), we have

$$\begin{aligned}& \sup_{1 \leq k \leq L} \left| \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{S}}^*(Y_i, X_i; \tau^*) - \frac{1}{n} \sum_{i=1}^n \mathbf{S}^0(Y_i, X_i; \tau^*) \right| \\ &= O_p \left(L \left[\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right] \right).\end{aligned}\quad (\text{A.21})$$

For the variance estimation, since $\sigma_j^{*2}(\tau^*) = 1/\{\mathbf{u}_{kj}^\top \mathbf{U} \mathbf{u}_{kj}\}$, we estimate

it by the plug-in method. In view of (A.16), we have

$$\sup_{1 \leq l \leq L, 1 \leq j \leq p} |\hat{\sigma}_j^{*2}(\tau_l) - \sigma_j^{*2}(\tau_l)| = O_p \left(\frac{1}{\sqrt{nh^2}} + \frac{\{\log(n)\}^{3/2}}{nh} + h^2 \right).$$

Step 4. Recall the proposed one-step efficient estimation in Section 2,

which is, for $j = 1, 2, \dots, p$

$$\hat{\boldsymbol{\beta}}_j(\tau^*) = \hat{\boldsymbol{\beta}}_j^c(\tau^*) + \hat{\sigma}_j^{*2}(\tau^*) \frac{\sum_{i=1}^n \hat{\mathbf{S}}_j^*(Y_i, X_i; \tau^*)}{n}.$$

For ease of presentation, we define

$$\mathbf{S}^{*0}(Y_i, X_i; \tau^*, \boldsymbol{\beta}(\cdot)) \equiv \sum_{l=1}^{L+1} \frac{a_{(l-1)i} - a_{li}}{\tau_l - \tau_{l-1}} [I\{X_i^\top \boldsymbol{\beta}(\tau_{l-1}) < Y_i < X_i^\top \boldsymbol{\beta}(\tau_l)\} - (\tau_l - \tau_{l-1})].$$

According to the above definition, $\mathbf{S}^0(Y_i, X_i; \tau^*) = \mathbf{S}^{*0}(Y_i, X_i; \tau^*, \boldsymbol{\beta}_0(\tau^*))$.

Observe that, for $j = 1, 2, \dots, p$,

$$\begin{aligned} & \hat{\beta}_j(\tau^*) - \beta_{0,j}(\tau^*) \\ &= \{\hat{\beta}_j^c(\tau^*) - \beta_{0,j}(\tau^*)\} + (\hat{\sigma}_j^{*2}(\tau^*) - \sigma_j^{*2}(\tau^*)) \frac{1}{n} \sum_{i=1}^n \{\hat{S}_j^*(Y_i, X_i; \tau^*) - S_j^0(Y_i, X_i; \tau^*)\} \\ & \quad + \sigma_j^{*2}(\tau^*) \frac{1}{n} \sum_{i=1}^n \{\hat{S}_j^*(Y_i, X_i; \tau^*) - S_j^0(Y_i, X_i; \tau^*)\} \\ & \quad + \frac{1}{n} \sum_{i=1}^n S_j^0(Y_i, X_i; \tau^*) (\hat{\sigma}_j^{*2}(\tau^*) - \sigma_j^{*2}(\tau^*)) \\ & \quad + \sigma_j^{*2}(\tau^*) \frac{1}{n} \sum_{i=1}^n S_j^0(Y_i, X_i; \tau^*). \end{aligned}$$

Since S_j^{*0} is the efficient score, one can show that uniformly for $\|\boldsymbol{\beta}(\tau_l) - \boldsymbol{\beta}_0(\tau_l)\| \leq B_n$ and $1 \leq l \leq L$,

$$\begin{aligned} & \frac{|\sum_{i=1}^n \mathbf{S}^{*0}(Y_i, X_i; \tau^*, \boldsymbol{\beta}(\cdot)) - \sum_{i=1}^n \mathbf{S}^0(Y_i, X_i; \tau^*) + n A_n \{\boldsymbol{\beta}(\tau^*) - \boldsymbol{\beta}_0(\tau^*)\}|}{\sqrt{n} + n \|\boldsymbol{\beta}(\tau^*) - \boldsymbol{\beta}_0(\tau^*)\|^{1+\delta}} \\ &= o_p(1), \end{aligned} \tag{A.22}$$

where $A_n = (1/n) \sum_{i=1}^n \{\partial/(\partial \boldsymbol{\beta}(\tau^*)) S_j^{*0}(Y_i, X_i; \tau^*, \boldsymbol{\beta}(\cdot))\}$, $0 < \delta < 1$, and

$B_n \rightarrow 0$ as $n \rightarrow \infty$.

Second, by the central limit theorem,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n S_j^0(Y_i, X_i; \tau^*) \rightarrow N(0, \{\sigma_j^{*2}(\tau^*)\}^{-1}), \quad (\text{A.23})$$

in distribution as $n \rightarrow \infty$.

Third, by the law of large numbers, $-A_n \rightarrow \{\sigma_j^{*2}(\tau^*)\}^{-1}$ in probability as $n \rightarrow \infty$. Recall that $\hat{\sigma}_j^{*2}(\tau^*) - \sigma_j^{*2}(\tau^*) = o_p(1)$. Moreover, using (A.21) and (A.22), we have

$$\hat{\beta}^c(\tau^*) - \beta_0(\tau^*) + \sigma_j^{*2}(\tau^*) A_n \{\hat{\beta}^c(\tau^*) - \beta_0(\tau^*)\} = o_p(n^{-\frac{1}{2}}).$$

Finally,

$$\hat{\beta}_j(\tau^*) = \beta_{0,j}(\tau^*) + \sigma_j^{*2}(\tau^*) \frac{1}{n} \sum_{i=1}^n S_j^0(Y_i, X_i; \tau^*) + r_n,$$

where $r_n = O_p(L/(nh^2) + Lh^4)$ for n sufficiently large. In view of (A.23) and Assumption (A₃), we have shown

$$\sqrt{n}\{\hat{\beta}_j(\tau^*) - \beta_{0,j}(\tau^*)\} \rightarrow N(0, \sigma_j^{*2}(\tau^*))$$

in distribution as $n \rightarrow \infty$. The proof of Lemma 3 is complete.

Remark. Note that we need $r_n = o_p(1/\sqrt{n})$ to ensure the asymptotic normality, which requires

$$\sqrt{n}h^2 \rightarrow \infty, \text{ and } \sqrt{n}h^4 \rightarrow 0$$

as $n \rightarrow \infty$. It implies that we need to assume that $h = o(n^{-\delta})$ with $1/8 < \delta < 1/4$.

Supplementary III

Proof of Theorem 2. The first part of Theorem 2 has been shown in Lemma 2, and the proofs of the second part are analogous to those of Lemma 3, with L satisfying $\log(n)/c < L < c \log(n)$ for a positive constant $c > 1$. We omit the details.

Supplementary IV

Proof of Theorem 3.

According to the proofs of Lemma 3 and Theorem 2, we have a Bahadur representation of $\hat{\beta}_j(\tau)$ for $\tau \in (\epsilon, 1 - \epsilon)$,

$$\hat{\beta}_j(\tau) = \beta_{0,j}(\tau) + \sigma_j^2(\tau) \frac{1}{n} \sum_{i=1}^n S_j^0(Y_i, X_i; \tau) + r_n, \quad j = 1, \dots, p.$$

For any set $\mathbf{u} = (u_1, \dots, u_K)^\top$ with different values $u_k \in (\epsilon, 1 - \epsilon), k = 1, \dots, K$, let $\hat{\beta}_j(\mathbf{u}) = (\hat{\beta}_j(u_1), \dots, \hat{\beta}_j(u_K))^\top$ and $\beta_{0,j}(\mathbf{u}) = (\beta_{0,j}(u_1), \dots, \beta_{0,j}(u_K))^\top$.

Since the data $(Y_i, X_i), i = 1, \dots, n$ are independent and identically copies of (X, Y) from model (1), it can be easily shown that

$$\sqrt{n}\{\hat{\beta}_j(\mathbf{u}) - \beta_{0,j}(\mathbf{u})\} \rightarrow N(0, \tilde{\Sigma}) \quad (\text{A.24})$$

in distribution as $n \rightarrow \infty$, where $\tilde{\Sigma} = (g(u_l, u_k))_{K \times K}$ and $g(u, v) = \sigma_j^2(u)\sigma_j^2(v)$

$E\{S_j^0(Y, X; u)S_j^0(Y, X; v)\}$ for $u, v \in (\epsilon, 1 - \epsilon)$.

Next, we intend to show $\{S_j^*(Y, X; \tau) : \boldsymbol{\beta} \in \Theta, \tau \in (\epsilon, 1 - \epsilon)\}$ is a P -Donsker class, where Θ is the space of $\boldsymbol{\beta}$ satisfying Assumption (A_2) . For ease of presentation, we let one quantile grid point $\tau_k \equiv \tau$ vary across $(\epsilon, 1 - \epsilon)$, and the rest $L - 1$ quantile grid points, denoted by $\mathcal{A}$. For any quantile level τ , let $S_j^*(Y, X; \tau)$ be the semiparametric efficient score for $\beta_j(\tau)$ given in (2.10) by combining the information across the quantile levels in $\mathcal{A}$ and τ . It follows from (2.10) that

$$\begin{aligned}
\mathbf{S}^*(Y, X; \tau) &= (S_j^*(Y, X; \tau), \dots, S_p^*(Y, X; \tau))^\top \\
&= \sum_{l=1}^{L+1} \frac{f(X^\top \boldsymbol{\beta}(\tau_{l-1})|X)X^\top \mathbf{D}_{l-1}^* - f(X^\top \boldsymbol{\beta}(\tau_l)|X)X^\top \mathbf{D}_l^*}{\tau_l - \tau_{l-1}} \\
&\quad \times \left[I\{X^\top \boldsymbol{\beta}(\tau_{l-1}) < Y\} - (1 - \tau_{l-1}) \right] \\
&\quad + \sum_{l=1}^{L+1} \frac{f(X^\top \boldsymbol{\beta}(\tau_{l-1})|X)X^\top \mathbf{D}_{l-1}^* - f(X^\top \boldsymbol{\beta}(\tau_l)|X)X^\top \mathbf{D}_l^*}{\tau_l - \tau_{l-1}} \\
&\quad \times \left[(1 - \tau_l) - I\{X^\top \boldsymbol{\beta}(\tau_l) < Y\} \right] \\
&\equiv \Pi_1 + \Pi_2.
\end{aligned}$$

Under assumptions (A_1) and (A_2) , $f(X^\top \boldsymbol{\beta}(\tau_l)|X)$ and $X^\top \mathbf{D}_l^*$ are bounded for $l = 1, \dots, L+1$. Clearly, Π_1 and Π_2 are monotone in Y and bounded for all $\boldsymbol{\beta} \in \Theta$ and $\tau \in (\epsilon, 1 - \epsilon)$. Applying Lemma 4.1 in Kosorok (2006), we can show that $\{\Pi_1 : \boldsymbol{\beta} \in \Theta, \tau \in (\epsilon, 1 - \epsilon)\}$ and $\{\Pi_2 : \boldsymbol{\beta} \in \Theta, \tau \in (\epsilon, 1 - \epsilon)\}$ have bounded entropy (with bracketing) integral, which indicates that $\{\Pi_1 :$

$\beta \in \Theta, \tau \in (\epsilon, 1 - \epsilon)\}$ and $\{\Pi_2 : \beta \in \Theta, \tau \in (\epsilon, 1 - \epsilon)\}$ are P -Donsker classes. As a result, $\{S_j^*(Y, X; \tau) : \beta \in \Theta, \tau \in (\epsilon, 1 - \epsilon)\}$ is P -Donsker. According to Theorem 2.1 in Kosorok (2006), Lemma 3 and Theorem 2, the proof of Theorem 3 is complete.

Supplementary V

For the j th component of $\beta(\tau_k)$, $j = 1, \dots, p$, it follows from (A.13) that $\tilde{\mathbf{d}}$ can be solved by minimizing quadratic form $\tilde{\mathbf{d}}^\top \mathbf{U} \tilde{\mathbf{d}}$ over $\tilde{\mathbf{d}}$, subject to $d_j(\tau_k) = 1$. For $l \neq k$, letting the derivative of $\tilde{\mathbf{d}}^\top \mathbf{U} \tilde{\mathbf{d}}$ on $\mathbf{d}(\tau_l)$ be 0, we have

$$\begin{aligned} & (\tau_{l+1} - \tau_l) E\{f(X^\top \beta(\tau_{l-1})|X) f(X^\top \beta(\tau_l)|X) X X^\top\} \mathbf{d}(\tau_{l-1}) \\ & - (\tau_{l+1} - \tau_{l-1}) E\{f(X^\top \beta(\tau_l)|X) f_{Y|X}(X^\top \beta(\tau_l)) X X^\top\} \mathbf{d}(\tau_l) \\ & + (\tau_l - \tau_{l-1}) E\{f_{Y|X}(X^\top \beta(\tau_l)) f_{Y|X}(X^\top \beta(\tau_{l+1})) X X^\top\} \mathbf{d}(\tau_{l+1}) = 0. \end{aligned}$$

Based on model (1.1) with $p = 1$, we have

$$f(X\beta(\tau_l)|X)X = \frac{1}{\dot{\beta}(\tau_l)}, \quad l = 1, 2, \dots, L.$$

It follows that

$$\begin{aligned} 0 &= (\tau_{l+1} - \tau_l) \frac{\mathbf{d}(\tau_{l-1})}{\dot{\beta}(\tau_{l-1})} - (\tau_{l+1} - \tau_{l-1}) \frac{\mathbf{d}(\tau_l)}{\dot{\beta}(\tau_l)} + (\tau_l - \tau_{l-1}) \frac{\mathbf{d}(\tau_{l+1})}{\dot{\beta}(\tau_{l+1})} \\ &= (\tau_{l+1} - \tau_l) \left\{ \frac{\mathbf{d}(\tau_{l-1})}{\dot{\beta}(\tau_{l-1})} - \frac{\mathbf{d}(\tau_l)}{\dot{\beta}(\tau_l)} \right\} - (\tau_l - \tau_{l-1}) \left\{ \frac{\mathbf{d}(\tau_l)}{\dot{\beta}(\tau_l)} - \frac{\mathbf{d}(\tau_{l+1})}{\dot{\beta}(\tau_{l+1})} \right\}. \end{aligned}$$

Hence, the score (A.13) becomes

$$\begin{aligned}
& S_j^*(Y, X; \tau_k) \\
&= \sum_{l=1}^{L+1} \frac{f(X\beta(\tau_{l-1})|X)X\mathbf{d}(\tau_{l-1}) - f(X\beta(\tau_l)|X)X\mathbf{d}(\tau_l)}{\tau_l - \tau_{l-1}} I\{X\beta(\tau_{l-1}) < Y < X\beta(\tau_l)\} \\
&= \left\{ \frac{\mathbf{d}(\tau_k)/\dot{\beta}(\tau_k) - \mathbf{d}(\tau_{k+1})/\dot{\beta}(\tau_{k+1})}{\tau_{k+1} - \tau_k} - \frac{\mathbf{d}(\tau_{k-1})/\dot{\beta}(\tau_{k-1}) - \mathbf{d}(\tau_k)/\dot{\beta}(\tau_k)}{\tau_k - \tau_{k-1}} \right\} \\
& \quad \{\tau_k - I(Y < \beta(\tau_k)X)\}.
\end{aligned}$$

Supplementary VI

For illustration, we consider the estimate of $\beta_1(\tau_1)$ when $p \geq 2$ and $L = 2$.

If we only consider single quantile level τ_1 and ignore the model information at quantile level τ_2 , by (A.13) with $L = 1$, we have

$$E(S_1^*(Y, X; \tau_1)^2) = E \left[\frac{1}{\tau_1(1 - \tau_1)} \{f(X^\top \beta(\tau_1)|X)\}^2 \{X^\top \mathbf{d}(\tau_1)\}^2 \right] \equiv E(Q_1).$$

On the other hand, by incorporating the model information at τ_2 ($L = 2$ in (A.13)), we have

$$\begin{aligned}
& E(S_1^*(Y, X; \tau_1)^2) \\
&= E \left[\frac{1}{\tau_1} \{f(X^\top \beta(\tau_1)|X)\}^2 \{X^\top \mathbf{d}(\tau_1)\}^2 + \frac{1}{1 - \tau_2} \{f(X^\top \beta(\tau_2)|X)\}^2 \{X^\top \mathbf{d}(\tau_2)\}^2 \right. \\
& \quad \left. + \frac{1}{\tau_2 - \tau_1} \left\{ f(X^\top \beta(\tau_1)|X)X^\top \mathbf{d}(\tau_1) - f(X^\top \beta(\tau_2)|X)X^\top \mathbf{d}(\tau_2) \right\}^2 \right] \\
&\equiv E(Q_2).
\end{aligned}$$

Then,

$$\begin{aligned}
& Q_2 - Q_1 \\
&= \frac{1}{\tau_2 - \tau_1} \left[\frac{1 - \tau_2}{1 - \tau_1} \{f(X^\top \boldsymbol{\beta}(\tau_1)|X)\}^2 \{X^\top \mathbf{d}(\tau_1)\}^2 \right. \\
&\quad \left. + \frac{1 - \tau_1}{1 - \tau_2} \{f(X^\top \boldsymbol{\beta}(\tau_2)|X)\}^2 \{X^\top \mathbf{d}(\tau_2)\}^2 \right] \\
&\quad - \frac{2}{\tau_2 - \tau_1} f(X^\top \boldsymbol{\beta}(\tau_1)|X) \{X^\top \mathbf{d}(\tau_1)\} f(X^\top \boldsymbol{\beta}(\tau_2)|X) \{X^\top \mathbf{d}(\tau_2)\} \\
&\geq 0.
\end{aligned} \tag{A.25}$$

Notice that $E(Q_2)$ and $E(Q_1)$ are variances of the scores of $\boldsymbol{\beta}(\tau_1)$ with and without the model information at quantile τ_2 , respectively, and that $1/E(Q_2)$ and $1/E(Q_1)$ are the corresponding efficiency bounds of estimation of $\boldsymbol{\beta}(\tau_1)$. Since (A.25) implies $1/E(Q_2) \leq 1/E(Q_1)$, it can be concluded that combining information of a different quantile τ_2 leads to efficiency gain in estimating $\boldsymbol{\beta}(\tau_1)$. The proofs are complete.

Supplementary VII: more numerical results

We study the performance of the proposed method under different combinations of τ s: (0.5, 0.7), (0.5, 0.7, 0.9), (0.5, 0.6, 0.7, 0.9) and (0.5, 0.6, 0.7, 0.8, 0.9), corresponding to $L = 2, 3, 4$, and 5 , respectively. In Figure 1, we plot computation times and MSEs of regression coefficient estima-

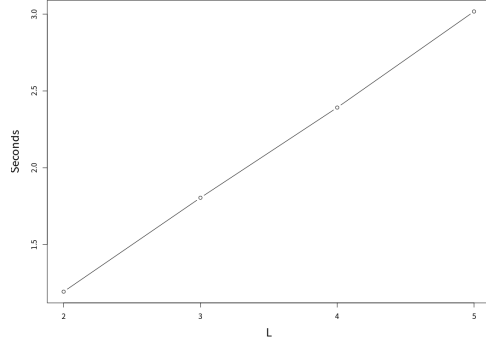
tion vs. different lengths of τ . As L increases, computation time becomes large. MSEs for the proposed method with $L = 4$ and 5 are much larger than those with $L = 2$ and 3. This is reasonable in our view, as density estimation involved in our proposed procedure becomes more challenging when the number of τ increases.

We also compare the numerical performance of the three methods with higher quantiles 0.8 and 0.9, which are shown in Table 1. Performance of the proposed method with more covariates in model (3.17) are shown in Table 2.

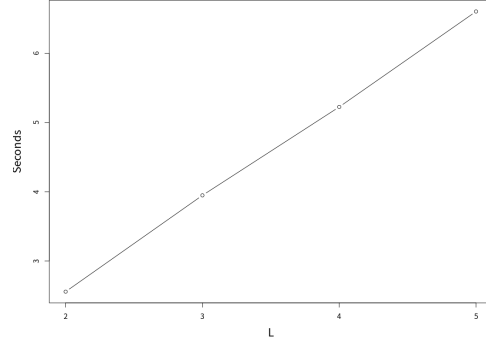
For the birth dataset, Figure 2 plots MSEs of median quantile estimation from different lengths of τ : (0.5, 0.7), (0.5, 0.7, 0.9), (0.5, 0.6, 0.7, 0.9) and (0.5, 0.6, 0.7, 0.8, 0.9). We observe that the combination with (0.5, 0.7, 0.9) has the smallest MSE.

References

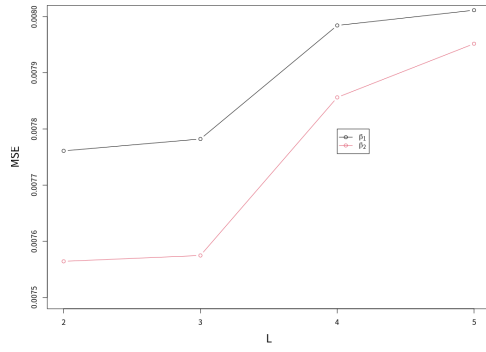
- [1] BEGUN, J. M., HALL, W. J., HUANG, W. M. AND WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432-452.
- [2] BICKEL, P. J., KLAASSEN, C. A., RITOV, Y. AND WELLNER, J. A.



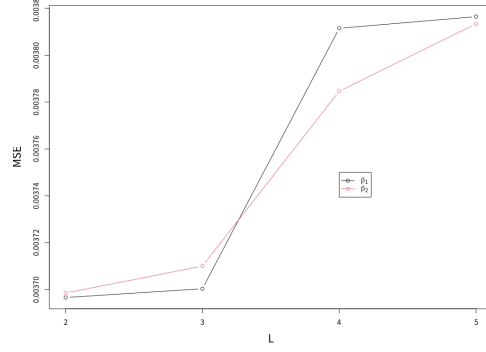
(a)



(b)



(c)

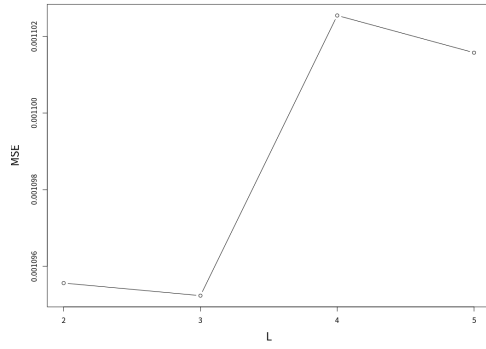


(d)

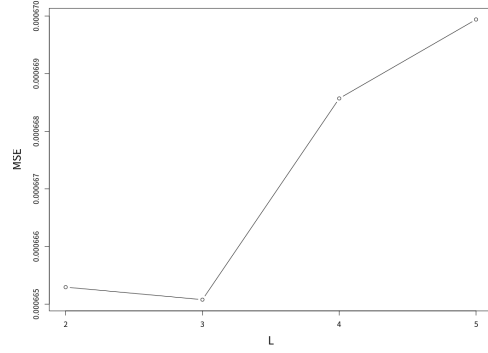
Figure 1: Computation times and MSE of regression coefficient estimation from different lengths of τ . (a): computation time with sample size 1000, (b): computation time with sample size 2000, (c): MSE with sample size 1000, (d): MSE with sample size 2000.

(1993). *Efficient and Adaptive Estimation for Semiparametric Models*.

New York: Springer-Verlag.



(a)



(b)

Figure 2: MSEs of median quantile estimation from different lengths of τ .

(a): male children, (b): female children.

- [3] Kosorok M. R. (2006). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer-Verlag.
- [4] PORTNOY, S. (2012). Nearly root- n approximation for regression quantile processes. *Ann. Statist* **40**, 1714-1736.
- [5] VAN DER VAART A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Table 1: Simulation results for five models with quantiles 0.8 and 0.9.

Model	n	$\tau = 0.8$		$\tau = 0.9$	
		$\beta_1(\tau)$	$\beta_2(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$
M1	True	2	1.8416	2	2.2816
	1000 TQE	2.007(0.0603)*	1.8332(0.1072)	2.0117(0.0725)	2.2734(0.1296)
	SEF	2.0098(0.0318)	1.8262(0.0642)	2.0158(0.0362)	2.2605(0.0784)
	EFF	2.0033(0.0293)	1.834(0.0618)	2.004(0.0372)	2.2747(0.0769)
	2000 TQE	2.0045(0.0412)	1.8361(0.0734)	2.009(0.0482)	2.2727(0.0877)
	SEF	2.0054(0.0192)	1.8335(0.0414)	2.0088(0.0231)	2.2698(0.0543)
	EFF	2.0014(0.0176)	1.8384(0.0411)	2.0029(0.0209)	2.2773(0.0515)
	M2	True	2.8416	3.2816	3.2816
	1000 TQE	2.8459(0.1358)	2.8335(0.1365)	3.2885(0.1622)	3.2751(0.1648)
M2	SEF	2.8438(0.0994)	2.8347(0.1022)	3.2818(0.1227)	3.2764(0.1237)
	EFF	2.8443(0.099)	2.8346(0.1009)	3.2843(0.1191)	3.2782(0.1206)
	2000 TQE	2.8454(0.0941)	2.8353(0.0936)	3.2861(0.1149)	3.275(0.1127)
	SEF	2.8412(0.0686)	2.8401(0.0673)	3.2836(0.0851)	3.2784(0.0862)
	EFF	2.8423(0.0678)	2.8394(0.0669)	3.2841(0.0821)	3.2802(0.0843)
	M3	True	2	3.1972	3.1972
	1000 TQE	2.013(0.1057)	2.3726(0.1876)	2.0246(0.141)	3.1834(0.2531)
	SEF	2.0192(0.0572)	2.3572(0.1135)	2.0354(0.072)	3.153(0.1533)
	EFF	2.0057(0.0548)	2.3733(0.1099)	2.0123(0.0668)	3.1796(0.1461)
M3	2000 TQE	2.0084(0.0721)	2.3771(0.1284)	2.0186(0.0938)	3.1807(0.1708)
	SEF	2.0108(0.0349)	2.3703(0.0732)	2.0201(0.0458)	3.1719(0.1061)
	EFF	2.0035(0.0321)	2.3794(0.0725)	2.0083(0.0407)	3.1861(0.1002)
	M4	True	2	4.0777	4.0777
	1000 TQE	2.026(0.1548)	2.3654(0.2754)	2.1044(0.4091)	4.0791(0.7658)
	SEF	2.0555(0.1174)	2.3(0.1921)	2.1874(0.2729)	3.9097(0.4643)
	EFF	2.0225(0.1386)	2.3431(0.2045)	2.1088(0.245)	3.9753(0.4437)
	2000 TQE	2.0168(0.104)	2.367(0.1876)	2.0754(0.2667)	4.0444(0.5028)
	SEF	2.0298(0.0675)	2.3352(0.1179)	2.1081(0.155)	3.9713(0.3199)
M4	EFF	2.0105(0.065)	2.3602(0.1124)	2.0615(0.1361)	4.0157(0.3069)
	M5	True	2.3863	3.1972	5.0777
	1000 TQE	2.3961(0.2664)	3.3753(0.3228)	3.2258(0.4602)	5.127(0.8171)
	SEF	2.3916(0.1973)	3.3608(0.238)	3.2108(0.3645)	5.1003(0.6241)
	EFF	2.3877(0.1954)	3.3741(0.2349)	3.2078(0.3441)	5.1077(0.5728)
	2000 TQE	2.3945(0.1843)	3.3708(0.2213)	3.2158(0.3218)	5.0801(0.5341)
	SEF	2.3844(0.134)	3.3746(0.1568)	3.2074(0.2518)	5.0805(0.4226)
	EFF	2.3867(0.1321)	3.377(0.1546)	3.2074(0.2408)	5.0878(0.3947)

* Standard deviations are in parentheses.

Table 2: Simulation results with more covariates.

τ	n		$\beta_1(\tau)$	$\beta_2(\tau)$	$\beta_3(\tau)$	$\beta_4(\tau)$
0.5		True	2	1	1	1
	1000	TQE	1.9921(0.3855)	0.998(0.2032)	1.0071(0.2558)	1.0072(0.2282)
		SEF	2.0014(0.2638)	0.9981(0.1634)	0.9986(0.1965)	1.0065(0.1825)
		EFF	1.9939(0.2678)	1.0007(0.1662)	0.9991(0.196)	1.0078(0.1799)
	2000	TQE	1.9973(0.2703)	1.0057(0.141)	0.9978(0.1838)	0.9963(0.1623)
		SEF	2.0054(0.1799)	1.0011(0.1169)	1.0007(0.1407)	0.9907(0.1235)
		EFF	2.0025(0.1817)	1.0021(0.1154)	1.0009(0.1419)	0.9911(0.1216)
0.7		True	2	1.5244	1.8473	1.7265
	1000	TQE	2.0172(0.4593)	1.5186(0.223)	1.847(0.292)	1.7305(0.2884)
		SEF	2.0465(0.3056)	1.5114(0.1896)	1.8314(0.2246)	1.7175(0.2268)
		EFF	1.9994(0.3198)	1.5193(0.1897)	1.8457(0.2298)	1.7378(0.2303)
	2000	TQE	1.9942(0.3091)	1.5306(0.1611)	1.8483(0.2088)	1.7321(0.2104)
		SEF	2.0196(0.2115)	1.5219(0.1333)	1.8448(0.1602)	1.7162(0.1547)
		EFF	2.0017(0.2174)	1.5252(0.1341)	1.8487(0.1621)	1.7257(0.1533)

* Standard deviations are in parentheses.

Table 3: Parameter estimate results for birth data with male child.

τ	model	Intercept			Mage			Fage			Nprevist		
		Est	Esd	P value	Est	Esd	P value	Est	Esd	P value	Est	Esd	P value
0.5	TQE	8.1311	0.0453	< 0.0001*	-0.0362	0.0139	0.0047	-0.0096	0.0044	0.0145	0.0357	0.0048	< 0.0001
	SEF	8.1137	0.0465	< 0.0001	-0.0313	0.0146	0.0162	-0.0092	0.0046	0.0225	0.0359	0.0052	< 0.0001
	EFF	8.1236	0.0497	< 0.0001	-0.0337	0.0149	0.012	-0.0104	0.0046	0.0118	0.0367	0.0049	< 0.0001
0.7	TQE	8.1217	0.0395	< 0.0001	-0.0154	0.0121	0.1026	-0.0039	0.004	0.1638	0.0357	0.0045	< 0.0001
	SEF	8.1197	0.0395	< 0.0001	-0.0146	0.0121	0.1141	-0.004	0.004	0.1573	0.0357	0.0045	< 0.0001
	EFF	8.1213	0.0452	< 0.0001	-0.015	0.014	0.1418	-0.0046	0.0043	0.1429	0.0365	0.0051	< 0.0001
0.9	TQE	8.1934	0.0549	< 0.0001	0.0106	0.0154	0.2459	-0.0091	0.0047	0.0269	0.0228	0.005	< 0.0001
	SEF	8.1788	0.0559	< 0.0001	0.015	0.0158	0.1707	-0.0091	0.0047	0.0272	0.023	0.0052	< 0.0001
	EFF	8.1847	0.0607	< 0.0001	0.014	0.0173	0.2094	-0.0099	0.0054	0.034	0.0232	0.006	< 0.0001

* means that p value is less than 0.0001.