SUPPLEMENT TO "OPTIMALLY COMBINED ESTIMATION FOR TAIL QUANTILE REGRESSION"

Kehui Wang and Huixia Judy Wang

In this supplement, we provide the proofs for Theorems 1-3, Proposition 1 and the statements in Remark 2 in the main paper. We first introduce some notations. Denote

$$a_n = \frac{\sqrt{(1-\tau)n}}{F_0^{-1}(\tau) - F_0^{-1}(\widetilde{\tau}_m)} \text{ and } a_k = \frac{\sqrt{l_k(1-\tau)n}}{F_0^{-1}(\tau_k) - F_0^{-1}(\widetilde{\tau}_{mk})}, \tag{1}$$

where $m > 1$, $\widetilde{\tau}_m = 1 - m(1-\tau)$ and $\widetilde{\tau}_{mk} = 1 - m(1-\tau_k)$. For notational simplicity, we denote $F_i = F_Y(\cdot|\mathbf{x}_i)$ and $f_i = f_Y(\cdot|\mathbf{x}_i)$.

Proof of Theorem 1. At the τ_kth quantile, the local quantile estimator of the coefficients in model (2.1) is defined as

$$(\widehat{\alpha}_k, \widehat{\boldsymbol{\beta}}_k) = \underset{(\alpha,\boldsymbol{\beta})}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau_k}(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta}).$$

Denote $\widehat{t}_{n,k} = a_k(\widehat{\alpha}_k - \alpha_{0,k})$ and $\widehat{\mathbf{z}}_{n,k} = a_k(\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_0)$, $k = 1, \ldots, K$. By Theorem 5.1 of Chernozhukov (2005), we have

$$(\widehat{t}_{n,1}, \widehat{\mathbf{z}}_{n,1}, \ldots, \widehat{t}_{n,K}, \widehat{\mathbf{z}}_{n,K}^T)^T \xrightarrow{d} N\left(\mathbf{0},\ \left(\frac{m^{-\xi}-1}{\xi}\right)^{-2}\left\{\widetilde{\boldsymbol{\Gamma}} \otimes \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}^{-1}\right\}\right),$$

where $\widetilde{\boldsymbol{\Gamma}}$ is a $K \times K$ matrix with the (k,k')th element as $\min(l_k, l_{k'})/\sqrt{l_k l_{k'}}$. Therefore, we have

$$a_n(\widehat{\boldsymbol{\beta}}_{WQAE} - \boldsymbol{\beta}_0) = \sum_{k=1}^K \varpi_k \frac{a_n}{a_k}\widehat{\mathbf{z}}_{n,k} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\varpi}^T\Phi^{-1}(\xi)\boldsymbol{\Gamma}\Phi^{-1}(\xi)\boldsymbol{\varpi}\left(\frac{m^{-\xi}-1}{\xi}\right)^{-2}\mathbf{D}^{-1}\right).$$

Lemma 1. *For a sequence of quantiles $\tau_1, \ldots, \tau_K$ with $\tau_k \to 1$ and $(1-\tau_k)n \to \infty$,*

$$\frac{\min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'}}{1-\tau} \sim \min(l_k, l_{k'}),$$

where $\tau \to 1$, $(1-\tau)n \to \infty$, and $(1-\tau_k)/(1-\tau) \to l_k$ for $k = 1, \ldots, K$.

Proof. Let $\tau^* = 1-\tau$, $\tau_k^* = 1-\tau_k$. Therefore, $\tau_k^*/\tau^* \to l_k$, $k = 1, \ldots, K$, and $\tau^* \to 0$.

It's easy to show that $\min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'} = \min(\tau_k^*, \tau_{k'}^*) - \tau_k^*\tau_{k'}^*$. For any $k, k' = 1, \ldots, K$,

$$\begin{aligned}
\frac{\min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'}}{1-\tau} &= \frac{\min(\tau_k^*, \tau_{k'}^*) - \tau_k^*\tau_{k'}^*}{\tau^*} \\
&= \frac{\tau^*[\min(l_k, l_{k'}) + o(1) - \tau^*\{l_k + o(1)\}\{l_{k'} + o(1)\}]}{\tau^*} \\
&= \min\{l_k, l_{k'} + o(1)\} - \tau^*\{l_k + o(1)\}\{l_{k'} + o(1)\} \\
&\sim \min(l_k, l_{k'}).
\end{aligned}$$

Lemma 2. *Under conditions **A3**-**A5**, $a_k/a_n \to l_k^{\xi+1/2}$ for any $k = 1, \ldots, K$.*

Proof. Since $\partial F_0^{-1}(\tau)/\partial\tau = 1/f_0\{F_0^{-1}(\tau)\}$, **A4** means that for any $x > 0$

$$\frac{f_0\{F_0^{-1}(1-\tau^*)\}}{f_0\{F_0^{-1}(1-x\tau^*)\}} \sim x^{-\xi-1}, \text{ as } \tau^* \to 0. \tag{2}$$

For any $\delta > 0$, note that $dF_0^{-1}(1-s\delta)/ds = -\delta\left[f_0\{F_0^{-1}(1-s\delta)\}\right]^{-1}$. Therefore,

$$\int_1^m \frac{1}{f_0\{F_0^{-1}(1-s\delta)\}} ds = \frac{F_0^{-1}(1-\delta) - F_0^{-1}(1-m\delta)}{\delta}. \tag{3}$$

Combining (2) and (3) gives

$$\begin{aligned}
\frac{F_0^{-1}(\tau_k) - F_0^{-1}(\widetilde{\tau}_{mk})}{l_k(1-\tau)/f_0\{F_0^{-1}(\tau_k)\}} &\sim f_0\{F_0^{-1}(\tau_k)\}\left[\frac{F_0^{-1}\{1-(1-\tau_k)\} - F_0^{-1}\{1-m(1-\tau_k)\}}{1-\tau_k}\right] \\
&= f_0\{F_0^{-1}(\tau_k)\}\int_1^m \frac{1}{f_0[F_0^{-1}\{1-s(1-\tau_k)\}]} ds \\
&= \int_1^m \frac{f_0[F_0^{-1}\{1-(1-\tau_k)\}]}{f_0[F_0^{-1}\{1-s(1-\tau_k)\}]} ds \\
&\sim \int_1^m s^{-\xi-1} ds = \frac{m^{-\xi}-1}{-\xi} \quad (\ln m \text{ if } \xi = 0).
\end{aligned} \tag{4}$$

Therefore, applying (2) and (4), we have

$$\begin{aligned}
\frac{a_k}{a_n} &= \frac{\sqrt{nl_k(1-\tau)}}{\sqrt{n(1-\tau)}}\left\{\frac{F_0^{-1}(\tau) - F_0^{-1}(\widetilde{\tau}_m)}{F_0^{-1}(\tau_k) - F_0^{-1}(\widetilde{\tau}_{mk})}\right\} \\
&= \sqrt{l_k}\left[\frac{F_0^{-1}(\tau) - F_0^{-1}(\widetilde{\tau}_m)}{(1-\tau)/f_0\{F_0^{-1}(\tau)\}}\right]\left[\frac{(1-\tau)/f_0[F_0^{-1}(\tau)]}{l_k(1-\tau)/f_0\{F_0^{-1}(\tau_k)\}}\right]\left[\frac{l_k(1-\tau)/f_0\{F_0^{-1}(\tau_k)\}}{F_0^{-1}(\tau_k) - F_0^{-1}(\widetilde{\tau}_{mk})}\right] \\
&\sim \sqrt{l_k}\left(\frac{m^{-\xi}-1}{-\xi}\right)\left(\frac{1}{l_k} l_k^{\xi+1}\right)\left(\frac{-\xi}{m^{-\xi}-1}\right) = l_k^{\xi+\frac{1}{2}}.
\end{aligned}$$

Proof of Theorem 2. For notational simplicity, we write $\widehat{\boldsymbol{\theta}}_{WCRQ} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_K, \widehat{\boldsymbol{\beta}}^T)^T$ in this proof. Let $\widehat{u}_{n,k} = a_k(\widehat{\alpha}_k - \alpha_{0,k})$, $k = 1, \ldots, K$, and $\widehat{\mathbf{u}}_n = a_n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, where

$(\alpha_{0,1},\ldots,\alpha_{0,K},\boldsymbol{\beta}_0)$ are the true parameters. From (2.7), it is clear that $(\widehat{u}_{n,1},\ldots,\widehat{u}_{n,K},\widehat{\mathbf{u}}_n)$ is the minimizer of

$$L_n=\frac{a_n}{\sqrt{n(1-\tau)}}\sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\left\{\rho_{\tau_k}\left(y_i-\mathbf{x}_i^T\boldsymbol{\beta}_0-\alpha_{0,k}-\frac{u_k}{a_k}-\frac{\mathbf{x}_i^T\mathbf{u}}{a_n}\right)-\rho_{\tau_k}(y_i-\mathbf{x}_i^T\boldsymbol{\beta}_0-\alpha_{0,k})\right\}$$

with respect to $(u_1,\ldots,u_k,\mathbf{u})$. Using Knight's identity (Knight, 1998),

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

we can rewrite L_n as $L_n\doteq L_{n,1}+L_{n,2}$, where

$$L_{n,1}=\frac{a_n}{\sqrt{n(1-\tau)}}\sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\left(\frac{u_k}{a_k}+\frac{\mathbf{x}_i^T\mathbf{u}}{a_n}\right)\{I(\epsilon_{i,k}<0)-\tau_k\},$$

$$L_{n,2}=\frac{a_n}{\sqrt{n(1-\tau)}}\sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\int_0^{\frac{u_k}{a_k}+\frac{\mathbf{x}_i^T\mathbf{u}}{a_n}}\{I(\epsilon_{i,k}\leqslant s)-I(\epsilon_{i,k}\leqslant 0)\}ds,$$

and $\epsilon_{i,k}=y_i-\mathbf{x}_i^T\boldsymbol{\beta}_0-\alpha_{0,k}$. Denoting $\psi_{i,k}=I(\epsilon_{i,k}<0)-\tau_k$, we have

$$\begin{aligned}L_{n,1}&=\frac{a_n}{\sqrt{n(1-\tau)}}\sum_{k=1}^{K}\frac{\omega_k}{a_k}\sum_{i=1}^{n}\psi_{i,k}u_k+\frac{a_n}{\sqrt{n(1-\tau)}}\sum_{k=1}^{K}\frac{\omega_k}{a_n}\sum_{i=1}^{n}\psi_{i,k}\mathbf{x}_i^T\mathbf{u}\\&=\sum_{k=1}^{K}W_{n,k}u_k+\mathbf{W}_n^T\mathbf{u}\doteq\widetilde{\mathbf{W}}_n^T\widetilde{\mathbf{U}},\end{aligned}\tag{5}$$

where

$$W_{n,k}=\frac{\omega_k}{\sqrt{n(1-\tau)}}\frac{a_n}{a_k}\sum_{i=1}^{n}\psi_{i,k},\ \ \mathbf{W}_n=\frac{1}{\sqrt{n(1-\tau)}}\sum_{k=1}^{K}\sum_{i=1}^{n}\omega_k\psi_{i,k}\mathbf{x}_i,$$
$$\widetilde{\mathbf{W}}_n=(W_{n,1},\ldots,W_{n,K},\mathbf{W}_n^T)^T,\ \text{ and }\widetilde{\mathbf{U}}=(u_{n,1},\ldots,u_{n,K},\mathbf{u}_n^T)^T.$$

We next derive the limiting distribution of $\widetilde{\mathbf{W}}_n$. Denote

$$\mathbf{T}_i=\left(\frac{\omega_1}{\sqrt{(1-\tau)}}\frac{a_n}{a_1}\psi_{i,1},\ldots,\frac{\omega_K}{\sqrt{(1-\tau)}}\frac{a_n}{a_K}\psi_{i,K},\mathbf{S}_i^T\right)^T,$$

where $\mathbf{S}_i=(1-\tau)^{-1/2}\sum_{k=1}^K\omega_k\psi_{i,k}\mathbf{x}_i$. Note that $\mathbf{T}_i$ are i.i.d. with mean $\mathbf{0}$ and covariance matrix $\mathbf{V}_n$. For $k,k'=1,\ldots,K$, the (k,k')th element of $\mathbf{V}_n$ is,

$$\begin{aligned}\mathbf{V}_n(k,k')&=\operatorname{cov}\Big(\frac{\omega_k}{\sqrt{(1-\tau)}}\frac{a_n}{a_k}\psi_{i,k},\frac{\omega_{k'}}{\sqrt{(1-\tau)}}\frac{a_n}{a_{k'}}\psi_{i,k'}\Big)\\&=\frac{\omega_k\omega_{k'}}{(1-\tau)}\frac{a_n^2}{a_ka_{k'}}\{\min(\tau_k,\tau_{k'})-\tau_k\tau_{k'}\}\\&\to\ \omega_k\omega_{k'}l_k^{-\xi-\frac{1}{2}}l_{k'}^{-\xi-\frac{1}{2}}\min(l_k,l_{k'}),\end{aligned}\tag{6}$$

where Lemmas 1 and 2 are used to prove the last step. Under condition **A2**,

$$\begin{aligned} Var(\mathbf{S}_i) &= E\{Var(\mathbf{S}_i|\mathbf{x}_i)\} + Var\{E(\mathbf{S}_i|\mathbf{x}_i)\} \\ &= E\left\{\mathbf{x}_i\mathbf{x}_i^T \sum_{k=1}^{K}\sum_{k'=1}^{K} \omega_k\omega_{k'} \frac{\min(\tau_k,\tau_{k'}) - \tau_k\tau_{k'}}{1-\tau}\right\} + 0 \\ &= \mathbf{D}\sum_{k=1}^{K}\sum_{k'=1}^{K} \omega_k\omega_{k'} \frac{\min(\tau_k,\tau_{k'}) - \tau_k\tau_{k'}}{1-\tau} \to \mathbf{D}\boldsymbol{\omega}^T\boldsymbol{\Gamma}\boldsymbol{\omega}. \end{aligned}$$

In addition, for any $k = 1, \ldots, K$,

$$\begin{aligned} &\operatorname{cov}\left(\frac{\omega_k}{\sqrt{(1-\tau)}}\frac{a_n}{a_k}\psi_{i,k}, \mathbf{S}_i\right) = \frac{\omega_k}{(1-\tau)}\frac{a_n}{a_k} E\left\{\psi_{i,k}\left(\sum_{j=1}^{K}\omega_j\psi_{i,j}\mathbf{x}_i\right)\right\} \\ &= \frac{\omega_k}{(1-\tau)}\frac{a_n}{a_k} E\left\{E\left(\psi_{i,k}\sum_{j=1}^{K}\omega_j\psi_{i,j}\mathbf{x}_i \,\middle|\, \mathbf{x}_i\right)\right\} = \mathbf{0}, \end{aligned} \tag{7}$$

where the last step is due to the assumption that $E(\mathbf{X}) = \mathbf{0}$. Combining (6)-(7) gives the limit of $\mathbf{V}_n$

$$\mathbf{V}_n \to \mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & D\boldsymbol{\omega}^T\boldsymbol{\Gamma}\boldsymbol{\omega} \end{pmatrix}, \tag{8}$$

where $\mathbf{V}_1$ is a $K \times K$ matrix with the (k, k')th element $\omega_k\omega_{k'} l_k^{-\xi-\frac{1}{2}} l_{k'}^{-\xi-\frac{1}{2}} \min(l_k l_{k'})$, and $k, k' = 1, \ldots, K$. Applying the multivariate Central Limit Theorem and Slutsky theorem to $\mathbf{T}_i$, we can show that

$$\widetilde{\mathbf{W}}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{T}_i \xrightarrow{d} N(0, \mathbf{V}). \tag{9}$$

Now we consider the second part of the objective function L_n, $L_{n,2}$. By the definitions of a_n and a_k, we have

$$L_{n,2} = \sum_{k=1}^{K}\omega_k \frac{F_0^{-1}(\tau_k) - F_0^{-1}(\widetilde{\tau}_{mk})}{F_0^{-1}(\tau) - F_0^{-1}(\widetilde{\tau}_m)} G_n^k,$$

where

$$\begin{aligned} G_n^k &= \frac{a_k}{\sqrt{l_k(1-\tau)n}}\sum_{i=1}^{n}\int_0^{\frac{u_k}{a_k}+\frac{\mathbf{x}_i^T\mathbf{u}}{a_n}} \{I(\epsilon_{i,k} \leqslant s) - I(\epsilon_{i,k} \leqslant 0)\}ds \\ &= \sum_{i=1}^{n}\int_0^{u_k+\frac{a_k}{a_n}\mathbf{x}_i^T\mathbf{u}} \frac{I(\epsilon_{i,k} \leqslant \frac{s}{a_k}) - I(\epsilon_{i,k} \leqslant 0)}{\sqrt{l_k(1-\tau)n}} ds. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} E(G_n^k) &= nE\left[\int_0^{u_k+\frac{a_k}{a_n}\mathbf{x}_i^T\mathbf{u}} \frac{F_i\left\{F_i^{-1}(\tau_k)+s/a_k\right\}-F_i\{F_i^{-1}(\tau_k)\}}{\sqrt{l_k(1-\tau)n}}ds\right] \text{(iterated expectations)} \\ &\overset{(i)}{=} nE\left(\int_0^{u_k+\frac{a_k}{a_n}\mathbf{x}_i^T\mathbf{u}} \frac{f_i[F_i^{-1}(\tau_k)+o\{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})\}]s}{a_k\sqrt{l_k(1-\tau)n}}ds\right) \\ &\overset{(ii)}{\sim} nE\left[\int_0^{u_k+\frac{a_k}{a_n}\mathbf{x}_i^T\mathbf{u}} \frac{f_i\{F_i^{-1}(\tau_k)\}s}{a_k\sqrt{l_k(1-\tau)n}}ds\right] \\ &= nE\left[\frac{1}{2}(u_k+\frac{a_k}{a_n}\mathbf{x}_i^T\mathbf{u})^2\frac{f_i\{F_i^{-1}(\tau_k)\}}{a_k\sqrt{l_k(1-\tau)n}}\right] \\ &= E\left[\frac{1}{2}(u_k+\frac{a_k}{a_n}\mathbf{x}_i^T\mathbf{u})^2\frac{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})}{l_k(1-\tau)/f_i\{F_i^{-1}(\tau_k)\}}\right] \\ &\overset{(iii)}{\sim} E\left\{\frac{1}{2}(u_k+l_k^{\xi+\frac{1}{2}}\mathbf{x}_i^T\mathbf{u})^2K(\mathbf{x}_i)^{-\xi}\left(\frac{m^{-\xi}-1}{-\xi}\right)\right\}. \end{aligned} \tag{10}$$

By Taylor expansion and the fact that $(1-\tau)n \to \infty$,

$$s/a_k = s\{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})\}/\sqrt{l_k(1-\tau)n} = o\{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})\},$$

then equation (i) in (10) is proven. The equation (ii) holds because $f_i\{F_i^{-1}(\tau_k)+o\left(F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})\right)\} \sim f_i\{F_i^{-1}(\tau_k)\}$ as $\tau_k \to 1$, which is derived following the same arguments as in the proof of Lemma 9.6 in Chernozhukov (2005). The equation (iii) is proven as follows. By condition **A3**, as $\tau \to 1$,

$$f_i\left\{F_i^{-1}(\tau)\right\} = f_U\{\alpha(\tau)|\mathbf{x}_i\} \sim f_0\left\{F_0^{-1}(\tau)\right\}.$$

Therefore,

$$\frac{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})}{l_k(1-\tau)/f_i\{F_i^{-1}(\tau_k)\}} \sim \frac{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})}{l_k(1-\tau)/f_0\{F_0^{-1}(\tau_k)\}}. \tag{11}$$

Combining (11) and (4), we have

$$\frac{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})}{l_k(1-\tau)/f_i\{F_i^{-1}(\tau_k)\}} \sim \frac{m^{-\xi}-1}{-\xi}, \tag{12}$$

which together with Lemma 2 proves equation (iii). Furthermore, we can show that $Var(G_n^k) \to 0$ by following the same arguments as in the proof of Lemma 9.6 in Chernozhukov (2005). In Lemma 2 we showed that $\{F_0^{-1}(\tau)-F_0^{-1}(\widetilde{\tau}_m)\}/\{F_0^{-1}(\tau_k)-F_0^{-1}(\widetilde{\tau}_{mk})\} \sim l_k^{\xi}$. Therefore, we have

$$\begin{aligned} L_{n,2} &\xrightarrow{p} E\left\{\sum_{k=1}^K \omega_k l_k^{-\xi}\frac{1}{2}(u_k+l_k^{\xi+\frac{1}{2}}\mathbf{x}_i^T\mathbf{u})^2\left(\frac{m^{-\xi}-1}{-\xi}\right)\right\} \\ &= \left(\frac{m^{-\xi}-1}{-\xi}\right)\sum_{k=1}^K \omega_k\left(\frac{1}{2}u_k^2 l_k^{-\xi}+\frac{1}{2}l_k^{\xi+1}\mathbf{u}^T\mathbf{D}\mathbf{u}\right). \end{aligned} \tag{13}$$

Combining (5), (9) and (13), we get

$$L_n \xrightarrow{d} L_\infty \equiv \sum_{k=1} W_k u_k + \mathbf{W}^T\mathbf{u} + \left(\frac{m^{-\xi}-1}{-\xi}\right)\sum_{k=1}^{K}\omega_k\left(\frac{1}{2}u_k^2 l_k^{-\xi} + \frac{1}{2}l_k^{\xi+1}\mathbf{u}^T\mathbf{D}\mathbf{u}\right),$$

where $\widetilde{\mathbf{W}} = \left(W_1,\ldots,W_K,\mathbf{W}^T\right)^T$ is a random vector following the distribution $N(\mathbf{0},\mathbf{V})$ with $\mathbf{V}$ defined in (8). Since the objective function L_∞ is quadratic in $\widetilde{\mathbf{U}}$, the minimizer of L_∞ is

$$\begin{aligned} u_{k,\infty} &= \left\{\left(\frac{m^{-\xi}-1}{\xi}\right)\omega_k l_k^{-\xi}\right\}^{-1} W_k,\ \text{for } k=1,\ldots,K,\\ \mathbf{u}_\infty &= \left(\frac{m^{-\xi}-1}{-\xi}\right)^{-1}\{\boldsymbol{\phi}^T(\xi)\boldsymbol{\omega}\}^{-1}\mathbf{D}^{-1}\mathbf{W}, \end{aligned}$$

where $\boldsymbol{\phi}(\xi) = (l_1^{\xi+1},\ldots,l_K^{\xi+1})^T$. By the definition of $\mathbf{W}$, we have

$$\mathbf{u}_\infty \sim N\left(\mathbf{0},\ \frac{\boldsymbol{\omega}^T\boldsymbol{\Gamma}\boldsymbol{\omega}}{\{\boldsymbol{\phi}^T(\xi)\boldsymbol{\omega}\}^2}\left(\frac{m^{-\xi}-1}{-\xi}\right)^{-2}\mathbf{D}^{-1}\right).$$

Note that $\omega_k \geq 0,\ k=1,\ldots,K$, then the application of the convexity lemma in Pollard (1991) gives

$$a_n(\widehat{\boldsymbol{\beta}}_{WCRQ} - \boldsymbol{\beta}_0) = \widehat{\mathbf{u}}_n \xrightarrow{d} \mathbf{u}_\infty.$$

The proof of the statements in Remark 2 relies on the following Lemma 3.

Lemma 3. *Let $f(x) = (a^{\xi+1} - x^{\xi+1})/(a-x)$ for $a>0, x>0$ and $x \neq a$, then (i) when $\xi > 0$, $f(x)$ is an increasing function; (ii) when $-1/2 < \xi < 0$, $f(x)$ is a decreasing function.*

Proof. We first prove (i). Note that

$$f'(x) = \frac{a^{\xi+1} - (\xi+1)ax^{\xi} + \xi x^{\xi+1}}{(a-x)^2} \tag{14}$$

has the same sign as that of $(a/x)^{\xi+1} - (\xi+1)a/x + \xi$. Consider the function $s(t) = t^{\xi+1} - (\xi+1)t + \xi,\quad t>0$. For $\xi > 0$, $s''(t) = \xi(\xi+1)t^{\xi-1} > 0$, so $s(t)$ is a convex function that achieves its minimum at $t=1$. Since $s(1)=0$, $s(t)$ and $f'(x)$ are both nonnegative. Thus $f(x)$ is an increasing function for $\xi > 0$. To prove (ii), we can use the same technique to show that $s(t)$ is a concave function achieving its maximum at $t=1$, and thus $f'(x) \leq 0$ for all $x > 0$.

Proof of Remark 2. Recall that the matrix $\boldsymbol{\Gamma}$ is a $K \times K$ matrix with the (k,k')th element defined as $\min(l_k, l_{k'})$. Then it can be shown that $\boldsymbol{\Gamma}^{-1}$ is a band matrix with the following form

$$\boldsymbol{\Gamma}^{-1}=\begin{pmatrix}
\frac{1}{l_1-l_2} & -\frac{1}{l_1-l_2} & 0 & & 0 & 0\\
-\frac{1}{l_1-l_2} & \frac{1}{l_1-l_2}+\frac{1}{l_2-l_3} & -\frac{1}{l_2-l_3} & & 0 & 0\\
0 & -\frac{1}{l_2-l_3} & \frac{1}{l_2-l_3}+\frac{1}{l_3-l_4} & & 0 & 0\\
0 & 0 & -\frac{1}{l_3-l_4} & \ddots & -\frac{1}{l_{K-2}-l_{K-1}} & 0\\
0 & 0 & 0 & & \frac{1}{l_{K-2}-l_{K-1}}+\frac{1}{l_{K-1}-l_K} & -\frac{1}{l_{K-1}-l_K}\\
0 & 0 & 0 & & -\frac{1}{l_{K-1}-l_K} & \frac{l_{K-1}}{L_K(l_{K-1}-l_K)}
\end{pmatrix}.$$

Therefore, the optimal weights $\boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)/\mathbf{1}_K^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)=(\omega_k)_{k=1}^K$, where

$$\omega_1=c\left(\frac{l_1^{\xi+1}}{l_1-l_2}-\frac{l_2^{\xi+1}}{l_1-l_2}\right),\ \omega_K=c\left\{\frac{l_{K-1}}{l_{K-1}-l_K}(l_K^\xi-l_{K-1}^\xi)\right\},$$
$$\omega_k=c\left(\frac{l_k^{\xi+1}-l_{k+1}^{\xi+1}}{l_k-l_{k+1}}-\frac{l_{k-1}^{\xi+1}-l_k^{\xi+1}}{l_{k-1}-l_k}\right)\ \text{for } k=2,\ldots,K-1,$$

and $c=\mathbf{1}_K^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)$ is a positive constant. We consider the three different cases separately.

(i) Case 1 ($\xi>0$). Note that $l_1>l_2>\ldots>l_K$. Obviously $\omega_1>0$, $\omega_K<0$. For any $k=2,\ldots,K-1$, let $f(x)=\frac{l_k^{\xi+1}-x^{\xi+1}}{l_k-x}$, then $\omega_k=c\{f(l_{k+1})-f(l_k)\}<0$ by Lemma 3 (i).

(ii) Case 2 ($\xi=0$). It is easy to show that $\omega_1=1$ and $\omega_2=\ldots=\omega_K=0$.

(iii) Case 3 ($-1/2<\xi<0$). By Lemma 3 (ii) and the similar technique as used in the proof for case 1, we can show that $\omega_k>0$ for $k=1,\ldots,K$.

The proof of Proposition 1 relies on the following lemma.

Lemma 4 (Lemma 2 of Zhao and Xiao, 2013)**.** *Let* $\mathbf{S}$ *be a* $K\times K$ *symmetric positive-definite matrix and* $\boldsymbol{v}$ *be any non-zero* $K\times 1$ *column vector. Define* $\mathbf{M}=\boldsymbol{v}^T\mathbf{S}^{-1}\boldsymbol{v}\mathbf{S}-\boldsymbol{v}\boldsymbol{v}^T$*. Then (i) for any column vector* $\mathbf{z}$, $\mathbf{z}^T\mathbf{M}\mathbf{z}\geq 0$*; and (ii)* $\mathbf{z}^T\mathbf{M}\mathbf{z}=0$ *holds if and only if* $\mathbf{z}=c\mathbf{S}^{-1}\boldsymbol{v}$ *for some real constant* c*.*

Proof of Proposition 1. Since the matrix $\boldsymbol{\Gamma}$ is positive-definite and symmetric, and $\boldsymbol{\phi}(\xi)$ is non-zero, by Lemma 4 (i), we have

$$\boldsymbol{\omega}^T\boldsymbol{\phi}^T(\xi)\boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)\boldsymbol{\Gamma}\boldsymbol{\omega}-\boldsymbol{\omega}^T\boldsymbol{\phi}(\xi)\boldsymbol{\phi}^T(\xi)\boldsymbol{\omega}\geq 0,\ \text{ for any } \boldsymbol{\omega}\in\mathbb{R}^K. \tag{15}$$

Note that $\boldsymbol{\phi}^T(\xi)\boldsymbol{\Gamma}\boldsymbol{\phi}(\xi)$ is a scalar, (15) can be expressed as

$$\frac{\boldsymbol{\omega}^T\boldsymbol{\Gamma}\boldsymbol{\omega}}{\boldsymbol{\omega}^T\boldsymbol{\phi}(\xi)\boldsymbol{\phi}^T(\xi)\boldsymbol{\omega}}=\sigma^2_{WCRQ}(\boldsymbol{\omega})\geq\{\boldsymbol{\phi}^T(\xi)\boldsymbol{\Gamma}\boldsymbol{\phi}(\xi)\}^{-1},\ \text{ for any } \boldsymbol{\omega}\in\mathbb{R}^K.$$

Lemma 4 (ii) then implies that the equality in (15) holds if and only if $\boldsymbol{\omega}=c\boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)$ for some constant c.

Proof of Theorem 3. By the definitions, $\mathbf{B}(\boldsymbol{\theta})$ is the first derivative of $E\{\mathbf{A}(\boldsymbol{\theta})\}$. With the Taylor series expansion, we get

$$E\{\mathbf{A}(\widetilde{\boldsymbol{\theta}})\} = E\{\mathbf{A}(\boldsymbol{\theta}_0)\} + \mathbf{B}(\bar{\boldsymbol{\theta}})(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{16}$$

where $\bar{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}_0$ and $\widetilde{\boldsymbol{\theta}}$. Define

$$\begin{aligned} \boldsymbol{r}_n(\boldsymbol{\delta}) &= \mathbf{A}(\boldsymbol{\theta}+\boldsymbol{\delta}) - \mathbf{A}(\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n} \omega_k^{(o)} \mathbf{z}_{i,k} \left[I\{y_i - \mathbf{z}_{i,k}^T(\boldsymbol{\theta}+\boldsymbol{\delta}) < 0\} - I(y_i - \mathbf{z}_{i,k}^T\boldsymbol{\theta} < 0) \right]. \end{aligned}$$

Applying Lemma 4.1 of He and Shao (1996), we have the uniform approximation

$$\sup_{\boldsymbol{\delta}:||\boldsymbol{\delta}||\le C} ||\boldsymbol{r}_n(\boldsymbol{\delta}) - E\{\boldsymbol{r}_n(\boldsymbol{\delta})\}|| = O_p(\sqrt{n}\log n ||\boldsymbol{\delta}||^{1/2}), \text{ for some constant } C.$$

Since $\widetilde{\boldsymbol{\theta}}$ is an a_n-consistent estimator of $\boldsymbol{\theta}_0$,

$$||\{\mathbf{A}(\widetilde{\boldsymbol{\theta}}) - \mathbf{A}(\boldsymbol{\theta}_0)\} - [E\{\mathbf{A}(\widetilde{\boldsymbol{\theta}})\} - E\{\mathbf{A}(\boldsymbol{\theta}_0)\}]|| = O_p(\sqrt{n}\log n||\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_o||^{1/2}). \tag{17}$$

Combining (16) and (17) gives

$$\widehat{\boldsymbol{\theta}}_{OS} - \boldsymbol{\theta}_0 = -\mathbf{B}(\widetilde{\boldsymbol{\theta}})\mathbf{A}(\boldsymbol{\theta}_0) - \mathbf{R}_n,$$

where $\mathbf{R}_n = \mathbf{B}(\widetilde{\boldsymbol{\theta}})^{-1}\{\mathbf{B}(\widetilde{\boldsymbol{\theta}}) - \mathbf{B}(\bar{\boldsymbol{\theta}})\}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathbf{B}(\widetilde{\boldsymbol{\theta}})^{-1}O_p(\sqrt{n}\log n||\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_o||^{1/2})$. Following similar arguments as in the proof of Theorem 3 in Bradic, Fan and Wang (2011), we can show that $\mathbf{R}_n$ is $o_p(1/a_n)$. Consequently, to prove Theorem 3, we only need to consider $a_n\{\mathbf{B}(\widetilde{\boldsymbol{\theta}})\mathbf{A}(\boldsymbol{\theta}_0)\}$. Since $\widetilde{\boldsymbol{\theta}}$ is a consistent estimator, by Slusky theorem, it is sufficient to show the asymptotic normality of $a_n\{\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{A}(\boldsymbol{\theta}_0)\}$. By the regularly varying property in (12), Lemma 2 and the multivariate CLT, we can show that

$$a_n\{\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{A}(\boldsymbol{\theta}_0)\} \xrightarrow{d} N\left(\mathbf{0}, \mathbf{T}^{-1}\mathbf{J}\mathbf{T}^{-1}\right),$$

where

$$\mathbf{T} = \left(\frac{m^{-\xi}-1}{-\xi}\right)\begin{pmatrix} \omega_1^{(o)} l_1^{\xi+1} & & & \mathbf{0}^T \\ & \ddots & & \vdots \\ & & \omega_K^{(o)} l_K^{\xi+1} & \mathbf{0}^T \\ \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\omega}_{\text{opt}}^T \boldsymbol{\phi}(\xi)\mathbf{D} \end{pmatrix}, \quad \mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}\boldsymbol{\omega}_{\text{opt}}^T\boldsymbol{\Gamma}\boldsymbol{\omega}_{\text{opt}}, \end{pmatrix}$$

and $\mathbf{J}_1$ is a $K \times K$ matrix with the (k,k')th element $\omega_k\omega_{k'}\min(l_k, l_{k'})$. By some linear algebra, we can show that the lower right $p \times p$ block of $\mathbf{T}^{-1}\mathbf{J}\mathbf{T}^{-1}$ is

$$\frac{\boldsymbol{\omega}_{\text{opt}}^T\boldsymbol{\Gamma}\boldsymbol{\omega}_{\text{opt}}}{\{\boldsymbol{\phi}^T(\xi)\boldsymbol{\omega}_{\text{opt}}\}^2}\left(\frac{m^{-\xi}-1}{-\xi}\right)^{-2}\mathbf{D}^{-1}.$$

The proof is completed by plugging in $\boldsymbol{\omega}_{\text{opt}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)/\{\mathbf{1}_K^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\phi}(\xi)\}$.

References

Bradic, J., Fan, J. and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society B* **73**, 325–349.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics* **33**, 806–839.

He, X. and Shao, Q. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics* **24**, 2608–2630.

Knight, K. (1998). Limiting distributions for $L1$ regression estimators under general conditions. *The Annals of Statistics* **26**, 755–770.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186–199.