Appendix to “Fast Nonparametric Quantile Regression with Arbitrary Smoothing Methods” published in the Journal of Computational and Graphical Statistics

Hee-Seok Oh, Thomas C. M. Lee and Douglas W. Nychka

A Proof of Proposition 1

Let $\phi_\tau(x) = \rho'_\tau(x)$ and $\psi_\tau(x) = \rho'_{\tau,c}(x)$. By the definition of quantile and pseudo quantile, it follows that

$$\psi_\tau(y-\alpha^*)f(y)dy = 0 \quad \text{and} \quad \phi_\tau(y-\alpha_0)f(y)dy = 0. \tag{12}$$

Next we calculate a bound for

$$\int_{\mathbb{R}} \big\{\psi_\tau(y-\alpha) - \phi_\tau(y-\alpha)\big\} f(y)dy.$$

With the substitution $u = y - \alpha$ we have

$$\begin{aligned}\int_{\mathbb{R}} \big\{\psi_\tau(u) - \phi_\tau(u)\big\} f(u+\alpha)du &= \int_{-c}^{c} \big\{\psi_\tau(u) - \phi_\tau(u)\big\} f(u+\alpha)du \\ &\le \int_{-c}^{c} \big\{\psi_\tau(u) - \phi_\tau(u)\big\} du \cdot \sup|f| \\ &= \Big\{\frac{c}{2}\tau + \frac{c}{2}(1-\tau)\Big\} \sup|f| = \frac{c}{2}\sup|f|. \end{aligned} \tag{13}$$

To derive the bound of interests, we consider

$$\begin{aligned} &\int_{\mathbb{R}} \big\{\psi_\tau(y-\alpha^*) - \phi_\tau(y-\alpha^*)\big\} f(y)dy \\ = &\int_{\mathbb{R}} \psi_\tau(y-\alpha^*) f(y)dy - \int_{\mathbb{R}} \phi_\tau(y-\alpha^*) f(y)dy. \end{aligned}$$

Given (12), it follows that

$$\begin{aligned}\int_{\mathbb{R}} \big\{\psi_\tau(y-\alpha^*) - \phi_\tau(y-\alpha^*)\big\} f(y)dy &= -\int_{\mathbb{R}} \phi_\tau(y-\alpha^*) f(y)dy \\ &= \int_{-\infty}^{\alpha^*} (1-\tau) f(y)dy - \int_{\alpha^*}^{\infty} \tau f(y)dy. \end{aligned}$$

Without loss of generality, we assume $\alpha_0 < \alpha^*$. Then, by the following fact which can be derived by (12)

$$\int_{\alpha_0}^{\infty} \tau f(y)dy = \int_{-\infty}^{\alpha_0} (1-\tau) f(y)dy, \tag{14}$$

we obtain

$$\begin{aligned}
\int_{-\infty}^{\alpha^*}(1-\tau)f(y)dy - \int_{\alpha^*}^{\infty} \tau f(y)dy &= \int_{-\infty}^{\alpha_0}(1-\tau)f(y)dy + \int_{\alpha_0}^{\alpha^*}(1-\tau)f(y)dy \\
&\quad -\Big\{\int_{\alpha_0}^{\infty} \tau f(y)dy - \int_{\alpha_0}^{\alpha^*} \tau f(y)dy\Big\} \\
&= \int_{\alpha_0}^{\alpha^*}(1-\tau)f(y)dy + \int_{\alpha_0}^{\alpha^*} \tau f(y)dy \\
&= \int_{\alpha_0}^{\alpha^*} f(y)dy = F(\alpha^*) - F(\alpha_0). \qquad (15)
\end{aligned}$$

Finally, by applying the general bound of (13) to (15), we establish the proposition. □

B Proof of Proposition 2

Suppose that α_0 is the optimal solution of (4).For a sufficiently small c, the increment resulted from subtracting the minimizer α_0 by δ_α can be expressed as

$$\begin{aligned}
&\sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0 + \delta_\alpha) - \sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0) \\
&\quad = \sum_{i=1}^{n} \Big[\tau((y_i - \alpha_0 + \delta_\alpha) - \frac{1}{2}c)\mathrm{I}(c \le y_i - \alpha_0 + \delta_\alpha) - \tau((y_i - \alpha_0) - \frac{1}{2}c)\mathrm{I}(c \le y_i - \alpha_0)\Big] \\
&\quad + \sum_{i=1}^{n} \Big[\frac{1}{2}\tau(y_i - \alpha_0 + \delta_\alpha)^2/c\mathrm{I}(0 \le y_i - \alpha_0 + \delta_\alpha < c) - \frac{1}{2}\tau(y_i - \alpha_0)^2/c\mathrm{I}(0 \le y_i - \alpha_0 < c)\Big] \\
&\quad + \sum_{i=1}^{n} \Big[\frac{1}{2}(1-\tau)(y_i - \alpha_0 + \delta_\alpha)^2/c\mathrm{I}(-c \le y_i - \alpha_0 + \delta_\alpha < 0) \\
&\qquad\qquad - \frac{1}{2}(1-\tau)(y_i - \alpha_0)^2/c\mathrm{I}(-c \le y_i - \alpha_0 < 0)\Big] \\
&\quad + \sum_{i=1}^{n} \Big[(\tau - 1)((y_i - \alpha_0 + \delta_\alpha) + \frac{1}{2}c)\mathrm{I}(y_i - \alpha_0 + \delta_\alpha < -c) \\
&\qquad\qquad - (\tau - 1)((y_i - \alpha_0) + \frac{1}{2}c)\mathrm{I}(y_i - \alpha_0 < -c)\Big].
\end{aligned}$$

Now we examine this one term at a time. The first term is

$$\sum_{i=1}^{n}\Big[\tau((y_i-\alpha_0+\delta_\alpha)-\frac{1}{2}c)\mathrm{I}(c\le y_i-\alpha_0+\delta_\alpha)-\tau((y_i-\alpha_0)-\frac{1}{2}c)\mathrm{I}(c\le y_i-\alpha_0)\Big]=\tau\delta_\alpha n_1, \quad (16)$$

where n_1 is the number of terms with $y_i \ge \alpha_0 + c$. Next, the second term becomes, for $0 \le y_i - \alpha_0 < c$,

$$\begin{aligned}
&\sum_{i=1}^{n}\Big[\frac{1}{2}\tau(y_i-\alpha_0+\delta_\alpha)^2/c\mathrm{I}(0\le y_i-\alpha_0+\delta_\alpha<c)-\frac{1}{2}\tau(y_i-\alpha_0)^2/c\mathrm{I}(0\le y_i-\alpha_0<c)\Big]\\
&=\frac{\tau}{2c}\sum_{i=1}^{n}\Big[(y_i-\alpha_0+\delta_\alpha)^2-(y_i-\alpha_0)^2\Big]\\
&\approx\delta_\alpha\frac{\tau}{c}\sum_{i=1}^{n}(y_i-\alpha)\mathrm{I}(0\le y_i-\alpha_0<c). \qquad (17)
\end{aligned}$$

Since $\sum_{i=1}^{n}(y_i-\alpha)\mathrm{I}(0\le y_i-\alpha_0<c)\le n_2c$, where n_2 is the number of terms with $0 \le y_i < \alpha_0 + c$, it follows that the second term of (17) is less than or equal to $n_2\delta_\alpha\tau$. That is,

$$\delta_\alpha\frac{\tau}{c}\sum_{i=1}^{n}(y_i-\alpha)\mathrm{I}(0\le y_i-\alpha_0<c)=n_2\delta_\alpha\tau-\kappa, \qquad (18)$$

where $\kappa(<\infty)$ represents the difference between (17) and $n_2\delta_\alpha\tau$. Similarly, the third term can be written as

$$\begin{aligned}
&\sum_{i=1}^{n}\Big[\frac{1}{2}(1-\tau)(y_i-\alpha_0+\delta_\alpha)^2/c\mathrm{I}(-c\le y_i-\alpha_0+\delta_\alpha<0\\
&\qquad-\frac{1}{2}(1-\tau)(y_i-\alpha_0)^2/c\mathrm{I}(-c\le y_i-\alpha_0<0)\Big]\\
&\approx(\tau-1)\delta_{\alpha_0}n_3+\kappa', \qquad (19)
\end{aligned}$$

where n_3 denotes the number of terms with $-c \le y_i - \alpha_0 < 0$ and κ' is a finite constant which is the difference between the third term and $(\tau-1)\delta_{\alpha_0}n_3$. Finally, we obtain the fourth term as

$$\begin{aligned}
&\sum_{i=1}^{n}\Big[(\tau-1)((y_i-\alpha_0+\delta_\alpha)+\frac{1}{2}c)\mathrm{I}(y_i-\alpha_0+\delta_\alpha<-c)\\
&\qquad-(\tau-1)((y_i-\alpha_0)+\frac{1}{2}c)\mathrm{I}(y_i-\alpha_0<-c)\Big]\\
&=n_4(\tau-1)\delta_\alpha, \qquad (20)
\end{aligned}$$

where n_4 denotes the number of terms with $y_i < \alpha_0 - c$. Hence, by combining (16)-(20), we obtain

$$\begin{aligned}&\sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0 + \delta_\alpha) - \sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0) \\ &\quad = \tau\delta_\alpha n_1 + \tau\delta_\alpha n_2 - \kappa + (\tau - 1)\delta_\alpha n_3 + \kappa' + (\tau - 1)\delta_\alpha n_4.\end{aligned}$$

Let $\zeta = \kappa' - \kappa$. Note that with a sufficiently small c, ζ is close to zero. In the median case of $\tau = 0.5$, ζ is exactly equal to zero. Define n_+ as the number of positive terms and n_- as the number of negative terms. Note that $n = n_+ + n_-$. Then

$$\begin{aligned}&\sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0 + \delta_\alpha) - \sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0) \\ &\quad = \tau\delta_\alpha(n_1 + n_2) + (\tau - 1)\delta_\alpha(n_3 + n_4) + \zeta \\ &\quad = \delta_\alpha[\tau n_+ + (\tau - 1)n_- + \Delta] \\ &\quad = \delta_\alpha[\tau n + \Delta - n_-], \qquad (21)\end{aligned}$$

where $\Delta = \zeta/\delta_\alpha$ which is negligible with a sufficiently small c. Hence, by requiring that (21) should be nonnegative, it results $\tau n + \Delta > n_-$.

Similarly, by letting $\zeta' = \kappa'' - \kappa'''$, we obtain the result

$$\begin{aligned}&\sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0 - \delta_\alpha) - \sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha_0) \\ &\quad = -\tau\delta_\alpha n_1 - \tau\delta_\alpha n_2 + \kappa'' + (1 - \tau)\delta_\alpha n_3 - \kappa''' + (1 - \tau)\delta_\alpha n_4 \\ &\quad = -\tau\delta_\alpha(n_1 + n_2) + (1 - \tau)\delta_\alpha(n_3 + n_4) + \zeta' \\ &\quad = \delta_\alpha[-\tau n_+ + (1 - \tau)n_- + \Delta'] \\ &\quad = \delta_\alpha[(1 - \tau)n + \Delta' - n_+], \qquad (22)\end{aligned}$$

where $\Delta' = \zeta'/\delta_\alpha$ which is also negligible with a sufficiently small c. Therefore, we obtain $(1 - \tau)n + \Delta' > n_+$. Combining with (21) and (22), we have

$$\tau - \frac{\Delta'}{n} < \frac{n_-}{n} < \tau + \frac{\Delta}{n}.$$

By the assumption $\Delta = O(n)$ and $\Delta' = O(n)$, we obtain

$$\frac{n_-}{n} \to \tau + \Delta'',$$

where Δ'' is negligible when c is chosen to be effectively zero relative to the magnitude of the data values. □