Combining Regression Quantile Estimators

Kejia Shan and Yuhong Yang

Amylin Pharmaceuticals and University of Minnesota

Supplementary Material

This note contains the proofs of Theorems 1 and 2.

S1. Proof of Theorem 1:

The technical tools for model combination in the literature on mean regression (e.g., Catoni 2004; Yang 2004a) are not applicable due to the nature of the check loss. To overcome the difficulty, we introduce a surrogate loss $L_{\tau,a}(\xi) = L_\tau(\xi) + a\xi^2$ with $a > 0$, which serves as an intermediate quantity in our analysis. This surrogate loss satisfies Condition 7 in Yang (2004a). Let $h(z) = \exp\left(-\lambda L_\tau(z)\right)$ and $q^{n-n_0} = \frac{1}{M}\sum_{j=1}^{M}\prod_{i=n_0+1}^{n} h\left(y_i - \hat{q}_{\tau,j,i}(x_i)\right)$, where $\hat{q}_{\tau,j,i}(\cdot)$ is the estimator from the j^{th} candidate based on $\{(y_l, x_l)\}_{l=1}^{i}$.

For any j, we have $\log\left(1/q^{n-n_0}\right) \le \log(M) + \lambda\sum_{i=n_0+1}^{n} L_\tau\left(y_i - \hat{q}_{\tau,j,i}(x_i)\right)$. It can be verified that $q^{n-n_0} = \prod_{i=n_0+1}^{n}\sum_{j=1}^{M} W_{j,i} h\left(y_i - \hat{q}_{\tau,j,i}(x_i)\right)$. Therefore

$$\log\left(1/q^{n-n_0}\right) = -\sum_{i=n_0+1}^{n}\log\left(\sum_{j=1}^{M} W_{j,i} h\left(y_i - \hat{q}_{\tau,j,i}(x_i)\right)\right) = -\sum_{i=n_0+1}^{n}\log\left(E^J h\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right)\right), \tag{1}$$

where E^J is defined as the expectation with respect to J under discrete distribution $P(J = j) = W_{j,i}$ for fixed i.

By Lemma 3.6.1 of Catoni (2004, p. 85), we have

$$\log\left(E^J h\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right)\right) \le -\lambda E^J L_\tau\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right) + I, \tag{2}$$

where

$$\begin{aligned} I &= \frac{\lambda^2}{2} E^J\left[L_\tau\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right) - E^J L_\tau\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right)\right]^2 \\ &\times \exp\left(\bar{c}\lambda 2^2\left(|y_i - m_i| + \left(1 + \sup_{j\ge 1}|\hat{q}_{\tau,j,i}(x_i) - m_i|\right)\right)\right). \end{aligned}$$

As in the proof of Theorem 5 in Yang (2004a), with $\beta = 1$, we get

$$\begin{aligned} & E^J\left[L_\tau\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right) - E^J L_\tau\left(y_i - \hat{q}_{\tau,J,i}(x_i)\right)\right]^2 \\ \le\ & \bar{c}^2 2^{2\beta-1}\left(|y_i - m_i|^{2\beta} + \left(1 + \sup_{j\ge 1}|\hat{q}_{\tau,j,i}(x_i) - m_i|\right)^{2\beta}\right) \end{aligned}$$

$$\times E^J \left(\hat{q}_{\tau,J,i}(x_i) - E^J(\hat{q}_{\tau,J,i}(x_i))\right)^2.$$

Let $b_0 = y_i - \hat{q}_{\tau,\cdot,i}(x_i) = y_i - E^J(\hat{q}_{\tau,J,i}(x_i))$ and $b = y_i - \hat{q}_{\tau,j,i}(x_i)$. For the surrogate loss function $L_{\tau,a}$, we can show that $\forall a > 0$,

$$\begin{aligned}
& L_{\tau,a}(b) - (2ab_0 + \tau - 1_{b_0<0})(b - b_0) - L_{\tau,a}(b_0) \\
= \;& ab^2 + (\tau - 1_{b<0})b - ab_0^2 - (\tau - 1_{b_0<0})b_0 - 2abb_0 + 2ab_0^2 - \tau b + \tau b_0 + b1_{b_0<0} - b_0 1_{b_0<0} \\
= \;& a(b-b_0)^2 + \tau b - \tau b_0 - b1_{b<0} + b_0 1_{b_0<0} - \tau b + \tau b_0 + b1_{b_0<0} - b_0 1_{b_0<0} \\
= \;& a(b-b_0)^2 + b1_{b_0<0} - b1_{b<0} \\
= \;& a(b-b_0)^2 + \begin{cases} -b & \text{if } b_0 \geq 0 \text{ and } b < 0 \\ b & \text{if } b_0 < 0 \text{ and } b \geq 0 \\ 0 & \text{otherwise,} \end{cases} \\
\geq \;& a(b-b_0)^2.
\end{aligned}$$

Taking expectation under E^J and notice that $E^J(\hat{q}_{\tau,J,i}(x_i) - \hat{q}_{\tau,\cdot,i}(x_i)) = E^J(b - b_0) = 0$, we have

$$E^J L_{\tau,a}(y_i - \hat{q}_{\tau,J,i}(x_i)) - L_{\tau,a}(y_i - \hat{q}_{\tau,\cdot,i}(x_i)) \geq aE^J(\hat{q}_{\tau,J,i}(x_i) - \hat{q}_{\tau,\cdot,i}(x_i))^2. \tag{3}$$

With the notations in Condition 7 of Yang (2004a), it can be verified that when a is chosen such that $a \leq \min(\tau, 1-\tau)$, we have $\underline{c} = a$, $\beta = 1$ and $\overline{c} = \max(\tau, 1-\tau)$. Let E_i denote the expectation with respect to the random error ϵ_i given the previous observations and x_i.

Under the following two constraints on λ:

$$\frac{\lambda \underline{c}}{2} \geq \lambda^2 \overline{c}^2 2^{2\beta-2} e^{\overline{c}\lambda 2^\beta (A+1)^\beta} \left(1 + (A+1)^{2\beta}\right) H(\overline{c}\lambda 2^\beta) \tag{4}$$

$$\overline{c}\lambda 2^\beta \leq t_0, \tag{5}$$

we have

$$E_i(I) \leq \frac{\lambda}{2} E_i \left[E^J L_{\tau,a}(Y_i - \hat{q}_{\tau,J,i}(x_i)) - L_{\tau,a}(Y_i - \hat{q}_{\tau,\cdot,i}(x_i))\right]. \tag{6}$$

Let $B(\lambda) = e^{2\lambda \max(\tau,1-\tau)(A+1)} \left(1 + (A+1)^2\right) H(2\lambda \max(\tau, 1-\tau))$. It is easy to see that when a and λ are chosen such that $\lambda \leq \frac{t_0}{2\max(\tau,1-\tau)}$ and $a \geq 2\lambda (\max(\tau, 1-\tau))^2 B(\lambda)$, the constraints are met. Let $a_\lambda = 2\lambda (\max(\tau, 1-\tau))^2 B(t_0)$. Then under such a choice of (λ, a), we have

$$\begin{aligned}
& E_i \left[\log E^J \exp(-\lambda L_\tau(Y_i - \hat{q}_{\tau,J,i}(X_i)))\right] \\
\leq \;& -\lambda E_i \left[L_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i))\right] + \lambda E_i \left[L_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i)) - E^J L_\tau(Y_i - \hat{q}_{\tau,J,i}(X_i))\right] \\
& + \frac{\lambda}{2} E_i \left[E^J L_{\tau,a}(Y_i - \hat{q}_{\tau,J,i}(X_i)) - L_{\tau,a}(Y_i - \hat{q}_{\tau,\cdot,i}(X_i))\right] \\
= \;& -\lambda E_i \left[L_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i))\right] - \frac{\lambda}{2} E_i \left[E^J L_\tau(Y_i - \hat{q}_{\tau,J,i}(X_i)) - L_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i))\right] \\
& + \frac{\lambda a}{2} E_i \left[E^J (Y_i - \hat{q}_{\tau,J,i}(X_i))^2\right] - \frac{\lambda a}{2} E_i \left[(Y_i - \hat{q}_{\tau,\cdot,i}(X_i))^2\right] \\
\leq \;& -\lambda E_i \left[L_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i))\right] + \frac{\lambda a}{2}(C_2 + C_1^2).
\end{aligned}$$

The first inequality above holds because of (2) and (6), provided that constraints (4) and (5) hold; the equality above holds because $E^J \hat{q}_{\tau,J,i}(x) = \hat{q}_{\tau,\cdot,i}(x)$ by definition, $\forall x$; the second inequality holds because of Condition 3.

From all above, for $\lambda \leq \frac{t_0}{2\max(\tau,1-\tau)}$, with $a = a_\lambda$, since j is arbitrary, we get

$$\sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i)) \leq \inf_j \left\{ \frac{\log(M)}{\lambda} + \frac{a_\lambda(C_2 + C_1^2)(n-n_0)}{2} + \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,j,i}(X_i)) \right\}. \quad (7)$$

The constraints on λ imply $\lambda = O(a)$, this leads to an optimal choice of λ (and a) when $\frac{\log(M)}{\lambda} = \frac{a_\lambda(C_2+C_1^2)(n-n_0)}{2}$. This gives $\lambda_{opt} = \sqrt{\frac{\log(M)}{(\max(\tau,1-\tau))^2 B(t_0)(C_2+C_1^2)(n-n_0)}}$. It is clear that (4) and (5) are satisfied when $n - n_0$ is large enough. Under this optimal choice of $\lambda = \lambda_{opt}$, we have

$$\sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,\cdot,i}(X_i)) \leq \inf_j \left\{ \tilde{C}\sqrt{\log(M)} \times \sqrt{n-n_0} + \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,j,i}(X_i)) \right\}, \quad (8)$$

where $\tilde{C}$ is a constant that depends on τ, C_1, C_2. By convexity of $L_\tau(\cdot)$ in its argument, we also have

$$EL_\tau(Y - \hat{q}_{\tau,\cdot,\cdot}(X)) \leq \inf_j \left\{ \tilde{C}\sqrt{\frac{\log(M)}{n-n_0}} + \frac{1}{n-n_0} \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,j,i}(X_i) \right\}. \quad (9)$$

This completes the proof.

S2. Proof of Theorem 2:

We only provide a sketched proof of Theorem 2 here. Define $h_a(z) = \exp(-\lambda L_{\tau,a}(z))$ and $q^{n-n_0} = \frac{1}{M}\sum_{j=1}^{M}\prod_{i=n_0+1}^{n} h_a(y_i - \hat{q}_{\tau,j,i}(x_i))$, where $\hat{q}_{\tau,j,i}(\cdot)$ is the j^{th} candidate estimator based on $\{(y_l, x_l)\}_{l=1}^{i}$.

With L_τ replaced by $L_{\tau,a}$, we also similarly update I with I_a, $\hat{q}_{\tau,\cdot,i}$ with $\hat{q}^a_{\tau,\cdot,i}$ and $\hat{q}_\tau$ with $\hat{q}^a_\tau$. Following the proof of Theorem 1, when $a \leq \min(\tau, 1-\tau)$, with $\underline{c} = a$, $\beta = 1$ and $\overline{c} = \max(\tau, 1-\tau)$, we have

$$\log\left(E^J h_a(y_i - \hat{q}_{\tau,J,i}(x_i))\right) \leq -\lambda E^J L_{\tau,a}(y_i - \hat{q}_{\tau,J,i}(x_i)) + I_a, \quad (10)$$

where

$$\begin{aligned} I_a &= \frac{\lambda^2}{2} E^J \left[L_{\tau,a}(y_i - \hat{q}_{\tau,J,i}(x_i)) - E^J L_{\tau,a}(y_i - \hat{q}_{\tau,J,i}(x_i))\right]^2 \\ &\times \exp\left(\overline{c}\lambda 2^{2\beta}\left(|y_i - m_i|^\beta + \left(1 + \sup_{j\geq 1}|\hat{q}_{\tau,j,i}(x_i) - m_i|\right)^\beta\right)\right). \end{aligned}$$

Since $L_{\tau,a}$ satisfies Condition 7 in Yang (2004a), under the constraints on a and λ, we have

$$E_i(I_a) \leq \frac{\lambda}{2} E_i\left[E^J L_{\tau,a}(Y_i - \hat{q}_{\tau,J,i}(x_i)) - L_\tau(Y_i - \hat{q}^a_{\tau,\cdot,i}(x_i))\right]. \quad (11)$$

Then

$$\begin{aligned} & E_i\left[\log E^J \exp(-\lambda L_{\tau,a}(Y_i - \hat{q}_{\tau,J,i}(X_i)))\right] \\ \leq\quad & -\lambda E_i\left[L_{\tau,a}(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i))\right] + \lambda E_i\left[L_{\tau,a}(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i)) - E^J L_{\tau,a}(Y_i - \hat{q}_{\tau,J,i}(X_i))\right] \\ & + \frac{\lambda}{2} E_i\left[E^J L_{\tau,a}(Y_i - \hat{q}_{\tau,J,i}(X_i)) - L_{\tau,a}(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i))\right] \\ \leq\quad & -\lambda E_i\left[L_{\tau,a}(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i))\right]. \end{aligned}$$

Consequently, we get

$$\sum_{i=n_0+1}^{n} EL_{\tau,a}(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i)) \leq \inf_j \left\{ \frac{\log(M)}{\lambda} + \sum_{i=n_0+1}^{n} EL_{\tau,a}(Y_i - \hat{q}_{\tau,j,i}(X_i)) \right\}. \tag{12}$$

Since $L_{\tau,a}(\xi) = L_\tau(\xi) + a\xi^2$, the above inequality implies that

$$\begin{aligned} \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i)) \quad \leq \quad & \inf_j \left\{ \frac{\log(M)}{\lambda} + \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,j,i}(X_i)) \right. \\ & \left. + a \sum_{i=n_0+1}^{n} E(Y_i - \hat{q}_{\tau,j,i}(X_i))^2 - a \sum_{i=n_0+1}^{n} E(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i))^2 \right\}. \end{aligned}$$

Under Conditions 1 and 3, we conclude that

$$\begin{aligned} \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}^a_{\tau,\cdot,i}(X_i)) \quad \leq \quad & \inf_j \left\{ \frac{\log(M)}{\lambda} + \sum_{i=n_0+1}^{n} EL_\tau(Y_i - \hat{q}_{\tau,j,i}(X_i)) \right. \\ & \left. + 4a(n - n_0)(C_2 + C_1^2 + A^2) \right\}. \end{aligned}$$

The rest of the proof follows as before. This completes the proof.