Supplementary Material for "Ensemble and calibration multiply robust estimation for quantile treatment effect"

Xiaohong He[1] and Lei Wang[1]
[1]School of Statistics and Data Science & LPMC, Nankai University, Tianjin 300071, China

S1 Proofs of Theorems 1-3

Proof of Theorem 1. Assume that true value of q^t is q_0^t, $t=0,1$. (C3) guarantees the identifiability of q_0^t, which means that q_0^t is the unique solution to $E\{\psi_\tau(Y_t-q)\}=0$. According to Theorem 5.9 in van der Vaart (1998), we need to check the uniform convergence in order to prove the consistency of multiply robust estimator. For treatment group, we need to check

$$\sup_{|q-q_0^1|<\epsilon}\left|\sum_{i\in S_1}\hat{w}_i\psi_\tau(Y_i-q)-E\{\psi_\tau(Y_1-q)\}\right|=o_p(1). \tag{S1}$$

For control group, we need to check

$$\sup_{|q-q_0^0|<\epsilon}\left|\sum_{i\in S_0}\hat{w}_i\psi_\tau(Y_i-q)-E\{\psi_\tau(Y_0-q)\}\right|=o_p(1). \tag{S2}$$

(a) Correctly specified model for $\pi(x)$ is contained in $\mathcal{W}$.

Without loss of generality, let $\pi_1(X,\beta_1)$ be the correct model and $\beta_{1,0}$ be the true value of β_1, so that $\pi_1(X,\beta_{1,0})=\pi(X)$. Define

$$\hat\lambda_{11}=\hat\theta_1(\hat\beta_1)\hat\rho_{11}-1,\ \hat\lambda_{1j}=\hat\theta_1(\hat\beta_1)\hat\rho_{1j},\ \hat\lambda_{01}=(1-\hat\theta_1(\hat\beta_1))\hat\rho_{01}-1,\ \hat\lambda_{0j}=(1-\hat\theta_1(\hat\beta_1))\hat\rho_{0j},$$

for $j=2,\dots,J$, then $\hat\lambda_1^T=(\hat\lambda_{11},\dots,\hat\lambda_{1J})$ and $\hat\lambda_0^T=(\hat\lambda_{01},\dots,\hat\lambda_{0J})$ satisfy

$$\frac{1}{n_1}\sum_{i\in S_1}\frac{\hat g_i(\hat\beta,\hat q^1,\hat\gamma^1)}{1+\hat\rho_1^T\hat g_i(\hat\beta,\hat q^1,\hat\gamma^1)}=\frac{\hat\theta_1(\hat\beta_1)}{n_1}\sum_{i\in S_1}\frac{\hat g_i(\hat\beta,\hat q^1,\hat\gamma^1)/\pi_{1i}(\hat\beta_1)}{1+\hat\lambda_1^T\hat g_i(\hat\beta,\hat q^1,\hat\gamma^1)/\pi_{1i}(\hat\beta_1)}=0,$$
$$\frac{1}{n_0}\sum_{i\in S_0}\frac{\hat g_i(\hat\beta,\hat q^0,\hat\gamma^0)}{1+\hat\rho_0^T\hat g_i(\hat\beta,\hat q^0,\hat\gamma^0))}=\frac{1-\hat\theta_1(\hat\beta_1)}{n_0}\sum_{i\in S_0}\frac{\hat g_i(\hat\beta,\hat q^0,\hat\gamma^0)/(1-\pi_{1i}(\hat\beta_1))}{1+\hat\lambda_0^T\hat g_i(\hat\beta,\hat q^0,\hat\gamma^0)/(1-\pi_{1i}(\hat\beta_1))}=0.$$

[1]Corresponding author: lwangstat@nankai.edu.cn

Therefore, weights $\hat{w}_i$'s can be expressed as

$$\hat{w}_i = \frac{1}{n_1}\frac{\hat{\theta}_1(\hat{\beta}_1)/\pi_{1i}(\hat{\beta}_1)}{1+\hat{\lambda}_1^T\hat{g}_i(\hat{\beta},\hat{q}^1,\hat{\gamma}^1)/\pi_{1i}(\hat{\beta}_1)}, \quad \text{for } i\in S_1,$$
$$\hat{w}_i = \frac{1}{n_0}\frac{(1-\hat{\theta}_1(\hat{\beta}_1))/(1-\pi_{1i}(\hat{\beta}_1))}{1+\hat{\lambda}_0^T\hat{g}_i(\hat{\beta},\hat{q}^0,\hat{\gamma}^0)/(1-\pi_{1i}(\hat{\beta}_1))}, \quad \text{for } i\in S_0.$$

Denote β_{m*}, γ_{k*}^t, q_{k*}^t, θ_{m*} and η_{k*}^t as the probability limits of $\hat{\beta}_m$, $\hat{\gamma}_k^t$, $\hat{q}_k^t$, $\hat{\theta}_m(\hat{\beta}_m)$ and $\hat{\eta}_k^t(\hat{q}_k^t,\hat{\gamma}_k^t)$ as $n\to\infty$, respectively. Notice that q_{k*}^t and η_{k*}^t do not depend on L. It is clear that $\beta_{1*}=\beta_{1,0}$, $\theta_{m*}=E\{\pi_m(\beta_{m*})\}$ and $\eta_{k*}^t = E[\psi_\tau\{Y^l(\gamma_{k*}^t)-q_{k*}^t\}]$. Write $\beta_*^T=\{\beta_{1*}^T,\ldots,\beta_{M*}^T\}$, $(\gamma_*^t)^T=\{(\gamma_{1*}^t)^T,\ldots,(\gamma_{K*}^t)^T\}$, $(q_*^t)^T=\{q_{1*}^t,\ldots,q_{K*}^t\}$ and

$$\begin{aligned} g(\beta_*,q_*^t,\gamma_*^t)^T =& \Big((\pi_1(\beta_{1*})-\theta_{1*})(-1)^{t+1},\ldots,(\pi_M(\beta_{M*})-\theta_{M*})(-1)^{t+1},\\ & \frac{1}{L}\sum_{l=1}^L\psi_\tau\{Y^l(\gamma_{1*}^t)-q_{1*}^t\}-\eta_{1*}^t,\ldots,\frac{1}{L}\sum_{l=1}^L\psi_\tau\{Y^l(\gamma_{K*}^t)-q_{K*}^t\}-\eta_{K*}^t\Big). \end{aligned}$$

Since that

$$E\Big\{\frac{T}{\pi(X)}g(\beta_*,q_*^1,\gamma_*^1)\Big\}=0, \quad E\Big\{\frac{1-T}{1-\pi(X)}g(\beta_*,q_*^0,\gamma_*^0)\Big\}=0,$$

and $\beta_{1*}=\beta_{1,0}$, 0's are the solutions of λ_1 and λ_0 in

$$E\Big\{\frac{Tg(\beta_*,q_*^1,\gamma_*^1)/\pi_1(\beta_{1*})}{1+\lambda_1^Tg(\beta_*,q_*^1,\gamma_*^1)/\pi_1(\beta_{1*})}\Big\}=0,$$
$$E\Big\{\frac{(1-T)g(\beta_*,q_*^0,\gamma_*^0)/(1-\pi_1(\beta_{1*}))}{1+\lambda_0^Tg(\beta_*,q_*^0,\gamma_*^0)/(1-\pi_1(\beta_{1*}))}\Big\}=0.$$

Thus, it can be obtained that $\hat{\lambda}_1=o_p(1)$ and $\hat{\lambda}_0=o_p(1)$ by the theory of empirical likelihood. For treatment group, uniform convergence (S1) satisfies

$$\sup_{|q-q_0^1|<\epsilon}\Big|\sum_{i\in S_1}\hat{w}_i\psi_\tau(Y_i-q)-E\{\psi_\tau(Y_1-q)\}\Big|\le A_1+B_1+C_1,$$

with

$$A_1=\sup_{|q-q_0^1|<\epsilon}\Big|\sum_{i\in S_1}\hat{w}_i\psi_\tau(Y_i-q)-\frac{1}{n}\sum_{i=1}^n\frac{T_i}{\pi_{1i}(\hat{\beta}_1)}\psi_\tau(Y_i-q)\Big|,$$
$$B_1=\sup_{|q-q_0^1|<\epsilon}\Big|\frac{1}{n}\sum_{i=1}^n\frac{T_i}{\pi_{1i}(\hat{\beta}_1)}\psi_\tau(Y_i-q)-\frac{1}{n}\sum_{i=1}^n\frac{T_i}{\pi_{1i}(\beta_{1,0})}\psi_\tau(Y_i-q)\Big|,$$
$$C_1=\sup_{|q-q_0^1|<\epsilon}\Big|\frac{1}{n}\sum_{i=1}^n\frac{T_i}{\pi_{1i}(\beta_{1,0})}\psi_\tau(Y_i-q)-E\{\psi_\tau(Y_1-q)\}\Big|.$$

Since $\hat{\theta}_1(\hat{\beta}_1) = n_1/n + o_p(1)$, it is clear that for $i \in S_1$, $\hat{w}_i = 1/n\pi_{1i}(\hat{\beta}_1) + o_p(1)$ and

$$A_1 = \sup_{|q-q_0^1|<\epsilon} \Big| \frac{1}{n} \sum_{i \in S_1} \frac{1}{\pi_{1i}(\hat{\beta}_1)} \psi_\tau(Y_i - q) - \frac{1}{n} \sum_{i=1}^n \frac{T_i}{\pi_{1i}(\hat{\beta}_1)} \psi_\tau(Y_i - q) \Big| + o_p(1) = o_p(1).$$

By Taylor expansion for B_1, it can be shown that

$$B_1 \le \sup_{|q-q_0^1|<\epsilon} \Big| \frac{1}{n} \sum_{i=1}^n \frac{T_i \psi_\tau(Y_i - q)}{\pi_{1i}(\beta_{1,0})^2} \frac{\partial \pi_{1i}(\beta)}{\partial \beta^T} \Big|_{\beta=\beta_{1,0}} |\hat{\beta}_1 - \beta_{1,0}| = o_p(1).$$

Han et al. (2019) proved that

$$\Big\{ \frac{T_i}{\pi_{1i}(\beta_{1,0})} \psi_\tau(Y_i - q) : |q - q_0^1| < \epsilon \Big\},$$

forms a Donsker class and $T_i \psi_\tau(Y_i - q_0^1)/\pi_{1i}(\beta_{1,0})$ is L_2 continuous at q_0^1. Combining with the fact that

$$\frac{1}{n} \sum_{i=1}^n \frac{T_i}{\pi_{1i}(\beta_{1,0})} \psi_\tau(Y_i - q_0^1) = O_p(n^{-1/2}),$$

and $E\{\psi_\tau(Y_1 - q_0^1)\} = 0$, it can be concluded that C_1 is $O_p(n^{-1/2})$. Hence, uniform convergence (S1) can be checked. For control group, uniform convergence (S2) satisfies

$$\sup_{|q-q_0^0|<\epsilon} \Big| \sum_{i \in S_0} \hat{w}_i \psi_\tau(Y_i - q) - E\{\psi_\tau(Y_0 - q)\} \Big| \le A_0 + B_0 + C_0,$$

with

$$A_0 = \sup_{|q-q_0^0|<\epsilon} \Big| \sum_{i \in S_0} \hat{w}_i \psi_\tau(Y_i - q) - \frac{1}{n} \sum_{i=1}^n \frac{1 - T_i}{1 - \pi_{1i}(\hat{\beta}_1)} \psi_\tau(Y_i - q) \Big|,$$

$$B_0 = \sup_{|q-q_0^0|<\epsilon} \Big| \frac{1}{n} \sum_{i=1}^n \frac{1 - T_i}{1 - \pi_{1i}(\hat{\beta}_1)} \psi_\tau(Y_i - q) - \frac{1}{n} \sum_{i=1}^n \frac{1 - T_i}{1 - \pi_{1i}(\beta_{1,0})} \psi_\tau(Y_i - q) \Big|,$$

$$C_0 = \sup_{|q-q_0^0|<\epsilon} \Big| \frac{1}{n} \sum_{i=1}^n \frac{1 - T_i}{1 - \pi_{1i}(\beta_{1,0})} \psi_\tau(Y_i - q) - E\{\psi_\tau(Y_0 - q)\} \Big|.$$

Since $1 - \hat{\theta}_1(\hat{\beta}_1) = n_0/n + o_p(1)$, it is clear that for $i \in S_0$, $\hat{w}_i = 1/n(1 - \pi_{1i}(\hat{\beta}_1)) + o_p(1)$ and

$$A_0 = \sup_{|q-q_0^0|<\epsilon} \Big| \frac{1}{n} \sum_{i \in S_0} \frac{1}{1 - \pi_{1i}(\hat{\beta}_1)} \psi_\tau(Y_i - q) - \frac{1}{n} \sum_{i=1}^n \frac{1 - T_i}{1 - \pi_{1i}(\hat{\beta}_1)} \psi_\tau(Y_i - q) \Big| + o_p(1) = o_p(1).$$

By Taylor expansion for B_0, we have

$$B_0 \le \sup_{|q-q_0^0|<\epsilon} \Big| \frac{1}{n} \sum_{i=1}^n \frac{(1 - T_i)\psi_\tau(Y_i - q)}{(1 - \pi_{1i}(\beta_{1,0}))^2} \frac{\partial \pi_{1i}(\beta)}{\partial \beta^T} \Big|_{\beta=\beta_{1,0}} \Big| |\hat{\beta}_1 - \beta_{1,0}| = o_p(1).$$

Similarly, it can be seen that

$$\left\{\frac{1-T_i}{1-\pi_{1i}(\beta_{1,0})}\psi_\tau(Y_i-q) : |q-q_0^0|<\epsilon\right\},$$

forms a Donsker class and $(1-T_i)\psi_\tau(Y_i-q_0^0)/(1-\pi_{1i}(\beta_{1,0}))$ is L_2 continuous at q_0^0. Combining with the facts that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1-T_i}{1-\pi_{1i}(\beta_{1,0})}\psi_\tau(Y_i-q_0^0)=O_p(n^{-1/2}),$$

and $E\{\psi_\tau(Y_0-q_0^0)\}=0$, it can be concluded that C_0 is $O_p(n^{-1/2})$ and (S2) can be checked. Since the MR estimator $\hat{q}_{mr}$ is the difference between $\hat{q}_{mr}^1$ and $\hat{q}_{mr}^0$, the consistency obviously holds.

(b) Correctly specified model for $f(Y|X)$ is contained in $\mathcal{F}$.

Without loss of generality, let $f_1(X,\gamma_1)$ be the correct model and $\gamma_{1,0}$ be the true value of γ_1, so that $f_1^t(X,\gamma_1^t)=f(Y_t|X)$ and $\gamma_{1*}^t=\gamma_{1,0}^t$. Similarly, it is clear that $g(\hat{\beta},\hat{q}^t,\hat{\gamma}^t)\to g(\beta_*,q_*^t,\gamma_*^t)$. Denote ρ_{t*} as the probability limit of $\hat{\rho}_t$, and one of the constraints g is

$$\sum_{i\in S_t}\hat{w}_i\Big[\frac{1}{L}\sum_{l=1}^{L}\psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t)-\hat{q}_1^t\}\Big]=\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{L}\sum_{l=1}^{L}\psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t)-\hat{q}_1^t\}\Big].$$

For treatment ($T=1$) and control ($T=0$) groups, define $T^{(1)}=T$ and $T^{(0)}=1-T$. By triangle inequality,

$$\begin{aligned}&\sup_{|q-q_0^t|<\epsilon}\Big|\sum_{i\in S_t}\hat{w}_i\psi_\tau(Y_i-q)-E\{\psi_\tau(Y_t-q)\}\Big|\\ \le&\sup_{|q-q_0^t|<\epsilon}\Big|\sum_{i\in S_t}\hat{w}_i\big[\psi_\tau(Y_i-q)-\frac{1}{L}\sum_{l=1}^{L}\psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t)-\hat{q}_1^t\}\big]\Big|+\sup_{|q-q_0^t|<\epsilon}|\hat{\eta}_1^t-E\{\psi_\tau(Y_t-q)\}|.\end{aligned}$$

It can be seen that left-hand side in (S1) or (S2) can be expressed as

$$\sup_{|q-q_0^t|<\epsilon}\Big|\sum_{i\in S_t}\hat{w}_i\psi_\tau(Y_i-q)-E\{\psi_\tau(Y_t-q)\}\Big|\le M_{t1}+M_{t2}+M_{t3}+M_{t4}+M_{t5}+M_{t6}+M_{t7},$$

with

$$M_{t1} = \sup_{|q-q_0^t|<\epsilon} \Big| \sum_{i\in S_t} \hat{w}_i \big[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t) - \hat{q}_1^t\}\big]$$

$$- \frac{1}{n_t}\frac{1}{1+\rho_{t*}^T g(\beta_*, q_*^t, \gamma_*^t)} \sum_{i=1}^{n} T_i^{(t)} \big[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t) - \hat{q}_1^t\}\big] \Big|,$$

$$M_{t2} = \Big|\frac{1}{n_t}\frac{1}{1+\rho_{t*}^T g(\beta_*, q_*^t, \gamma_*^t)} \sum_{i=1}^{n} T_i^{(t)} \frac{1}{L}\sum_{l=1}^{L} [\psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t) - \hat{q}_1^t\} - \psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\}]\Big|,$$

$$M_{t3} = \sup_{|q-q_0^t|<\epsilon} \Big|\frac{1}{n_t}\frac{1}{1+\rho_{t*}^T g(\beta_*, q_*^t, \gamma_*^t)} \sum_{i=1}^{n} T_i^{(t)} \Big[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\}\Big]$$

$$- \frac{n}{n_t}\frac{1}{1+\rho_{t*}^T g(\beta_*, q_*^t, \gamma_*^t)} E[T^{(t)}\{\psi_\tau(Y_t - q) - \psi_\tau(Y_t - q_0^t)\}]\Big|,$$

$$M_{t4} = \sup_{|q-q_0^t|<\epsilon} \Big|\frac{n}{n_t}\frac{1}{1+\rho_{t*}^T g(\beta_*, q_*^t, \gamma_*^t)} E[T^{(t)}\{\psi_\tau(Y_t - q) - \psi_\tau(Y_t - q_0^t)\}]\Big|,$$

$$M_{t5} = \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t) - \hat{q}_1^t\} - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\}\Big|,$$

$$M_{t6} = \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\} - E\{\psi_\tau(Y_t - q_0^t)\}\Big|,$$

$$M_{t7} = \sup_{|q-q_0^t|<\epsilon} |E\{\psi_\tau(Y_t - q_0^t)\} - E\{\psi_\tau(Y_t - q)\}|.$$

Obviously, we have

$$M_{t1} = \sup_{|q-q_0^t|<\epsilon} \Big|\frac{1}{n_t}\sum_{i\in S_t}\Big[\frac{1}{1+\hat{\rho}_t^T g_i(\hat{\beta}, \hat{q}^t, \hat{\gamma}^t)} - \frac{1}{1+\rho_{t*}^T g_i(\beta_*, q_*^t, \gamma_*^t)}\Big]$$

$$\big[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t) - \hat{q}_1^t\}\big]\Big|.$$

Using Taylor expansion based on derivatives and subderivatives, it can be verified that M_{t1} is $o_p(1)$. Similarly, it can be proved that M_{12} and M_{15} are both $o_p(1)$ by using Taylor expansion about $\hat{q}^t$ and $\hat{\gamma}^t$ on $\psi_\tau\{\tilde{Y}_i^l(\hat{\gamma}_1^t) - \hat{q}_1^t\} - \psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\}$. Under (C3), the unique solution to $E\{\psi_\tau(Y_t - q)\} = 0$ is q_0^t, so

$$\sup_{|q-q_0^t|<\epsilon} \big|E\{\psi_\tau(Y_t - q_0^t)\} - E\{\psi_\tau(Y_t - q)\}\big| = o_p(1),$$

then M_{14} and M_{17} are both $o_p(1)$. Besides, it can be shown that M_{16} is $o_p(1)$ by the weak law of large numbers and M_{13} is $O_p(n^{-1/2})$ by noting that

$$\Big\{T_i^{(t)}\big[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^{L} \psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\}\big] : |q - q_0^t| < \epsilon\Big\}$$

forms a Donsker class and $T_i^{(t)}\big[\psi_\tau(Y_i - q_0^t) - \frac{1}{L}\sum_{l=1}^{L}\psi_\tau\{\tilde{Y}_i^l(\gamma_{1,0}^t) - q_0^t\}\big]$ is L_2 continuous at q_0^t. Then uniform convergence (S1) and (S2) can be proved and consistency of $\hat{q}_{mr}^t$ holds. This completes the proof.

Proof of Theorem 2. Express the difference $\hat{q}_{mr} - q_0$ to be proved as two parts, $\hat{q}_{mr} - q_0 = (\hat{q}_{mr}^1 - q_0^1) - (\hat{q}_{mr}^0 - q_0^0)$, and consider them separately. If correct model for $\pi(x)$ is contained in $\mathcal{W}$, note that for treatment group,

$$
\begin{aligned}
0 =& \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\hat{\beta},\hat{q}^1,\hat{\gamma}^1)/\pi_{1i}(\hat{\beta}_1)}{1+\lambda_1^T g_i(\hat{\beta},\hat{q}^1,\hat{\gamma}^1)/\pi_{1i}(\hat{\beta}_1)} - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\hat{\beta},\hat{q}^1,\hat{\gamma}^1)}{\pi_{1i}(\hat{\beta}_1)} && \text{(S3)}\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\hat{\beta},\hat{q}^1,\hat{\gamma}^1)}{\pi_{1i}(\hat{\beta}_1)} - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)}{\pi_{1i}(\beta_{1*})} && \text{(S4)}\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)}{\pi_{1i}(\beta_{1*})} - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,q_*^1,\hat{\gamma}^1)}{\pi_{1i}(\beta_{1*})} && \text{(S5)}\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,q_*^1,\hat{\gamma}^1)}{\pi_{1i}(\beta_{1*})} - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,q_*^1,\gamma_*^1)}{\pi_{1i}(\beta_{1*})} && \text{(S6)}\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,q_*^1,\gamma_*^1)}{\pi_{1i}(\beta_{1*})}.
\end{aligned}
$$

Taylor expansion can be applied to (S3) and (S4), then

$$(S3) = -\frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\hat{\beta},\hat{q}^1,\hat{\gamma}^1)^{\otimes 2}}{\pi_{1i}(\hat{\beta}_1)^2}\hat{\lambda}_1 + o_p(n^{-1/2}).$$

Denote $R_i(\beta,q^1,\gamma^1) = g_i(\beta,q^1,\gamma^1)/\pi_{1i}(\beta_1)$. Notice that only the first row of $\partial g_i(0,\beta_*,\hat{q}^1,\hat{\gamma}^1)/\partial\beta_1$ has non-zero values and only the mth row of $\partial g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)/\partial\beta_m$ has non-zero values, $m = 2,\ldots,M$, therefore,

$$\frac{\partial R_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)}{\partial\beta_1} = \frac{\pi_{1i}(\beta_{1*})\partial g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)/\partial\beta_1 - g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)(\partial\pi_{1i}(\beta_{1*})/\partial\beta_1)^T}{\pi_{1i}(\beta_{1*})^2},$$

and

$$\frac{\partial R_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)}{\partial\beta_m} = \frac{1}{\pi_{1i}(\beta_{1*})}\frac{\partial g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)}{\partial\beta_m}.$$

Then

$$(S4) = -\frac{1}{n}\sum_{i=1}^{n}\frac{T_i g_i(\beta_*,\hat{q}^1,\hat{\gamma}^1)(\partial\pi_{1i}(\beta_{1*})/\partial\beta_1)^T}{\pi_{1i}(\beta_{1*})^2}(\hat{\beta}_1 - \beta_{1*}) + o_p(n^{-1/2}).$$

Similarly, it can be verified that $\{T_i R_i(\beta_*,q^1,\hat{\gamma}^1) : \|q^1 - q_*^1\| < \epsilon\}$ forms a Donsker class and $T_i R_i(\beta_*,q_*^1,\hat{\gamma}^1)$ is L_2 continuous at q_*^1. Then we have,

$$(S5) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial E\{T_i R_i(\beta_*,q_*^1,\hat{\gamma}^1)\}}{\partial q^1}(\hat{q}^1 - q_*^1) + o_p(n^{-1/2}).$$

It is also clear that $\{T_iR_i(\beta_*, q_*^1, \gamma^1) : \|\gamma^1 - \gamma_*^1\| < \epsilon\}$ forms a Donsker class and $T_iR_i(\beta_*, q_*^1, \gamma_*^1)$ is L_2 continuous at γ_*^1. Therefore,

$$(S6) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial E\{T_iR_i(\beta_*, q_*^1, \gamma_*^1)\}}{\partial \gamma^1}(\hat{\gamma}^1 - \gamma_*^1) + o_p(n^{-1/2}).$$

It is straightforward to see that both $E\{T_iR_i(\beta_*, q_*^1, \hat{\gamma}^1)\}$ and $E\{T_iR_i(\beta_*, q_*^1, \gamma_*^1)\}$ are 0. Therefore, both (S5) and (S6) are zeros. Hence, define

$$M_1 = E\Big\{\frac{g(\beta_*, q_*^1, \gamma_*^1)(\partial \pi_1(\beta_{1*})/\partial \beta_1)^T}{\pi_1(\beta_{1*})}\Big\}.$$

It can be concluded from (C7) that

$$\sqrt{n}\hat{\lambda}_1 = (G_1)^{-1}\Big\{n^{-1/2}\sum_{i=1}^{n}\frac{T_i - \pi_{1i}(\beta_{1*})}{\pi_{1i}(\beta_{1*})}g_i(\beta_*, q_*^1, \gamma_*^1) - n^{-1/2}\sum_{i=1}^{n} M_1 E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}\Big\} + o_p(1).$$

Next, note that

$$\begin{aligned}
0 =& \frac{1}{n}\sum_{i=1}^{n}\frac{T_i/\pi_{1i}(\hat{\beta}_1)}{1 + \hat{\lambda}_1^T g_i(\hat{\beta}, \hat{q}^1, \hat{\gamma}^1)/\pi_{1i}(\hat{\beta}_1)}\psi_\tau(Y_i - \hat{q}_{mr}^1) \\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{T_i/\pi_{1i}(\hat{\beta}_1)}{1 + \hat{\lambda}_1^T g_i(\hat{\beta}, \hat{q}^1, \hat{\gamma}^1)/\pi_{1i}(\hat{\beta}_1)}\psi_\tau(Y_i - \hat{q}_{mr}^1) - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\pi_{1i}(\hat{\beta}_1)}\psi_\tau(Y_i - \hat{q}_{mr}^1) \quad \text{(S7)} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\pi_{1i}(\hat{\beta}_1)}\psi_\tau(Y_i - \hat{q}_{mr}^1) - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - \hat{q}_{mr}^1) \quad \text{(S8)} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - \hat{q}_{mr}^1) - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - q_0^1) \quad \text{(S9)} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - q_0^1).
\end{aligned}$$

It can be shown that

$$(S7) = -\Big\{\frac{1}{n}\sum_{i=1}^{n} T_i\frac{\psi_\tau(Y_i - \hat{q}_{mr}^1)}{\pi_{1i}(\hat{\beta}_1)^2}g_i(\hat{\beta}, \hat{q}^1, \hat{\gamma}^1)^T\Big\}\hat{\lambda}_1 + o_p(n^{-1/2}),$$

$$(S8) = -\Big\{\frac{1}{n}\sum_{i=1}^{n}\frac{T_i\psi_\tau(Y_i - \hat{q}_{mr}^1)}{\pi_{1i}(\beta_{1*})^2}\Big(\frac{\partial \pi_{1i}(\beta_{1*})}{\partial \beta_1}\Big)^T\Big\}(\hat{\beta}_1 - \beta_{1*}) + o_p(n^{-1/2}),$$

and

$$(S9) = -f_1(q_0^1)(\hat{q}_{mr}^1 - q_0^1) + o_p(n^{-1/2}).$$

Let

$$B_1 = E\Big\{\frac{\psi_\tau(Y_1 - q_0^1)}{\pi_1(\beta_{1,0})}\Big(\frac{\partial \pi_1(\beta_{1,0})}{\partial \beta_1}\Big)^T\Big\},$$

it can be verified that

$$f_1(q_0^1)\sqrt{n}(\hat{q}_{mr}^1 - q_0^1) = -A_1\sqrt{n}\hat{\lambda}_1 - B_1\sqrt{n}(\hat{\beta}_1 - \beta_{1*}) + n^{-1/2}\sum_{i=1}^n \frac{T_i\psi_\tau(Y_i - q_0^1)}{\pi_{1i}(\beta_{1,0})^2} + o_p(1)$$

$$= n^{-1/2}\sum_{i=1}^n [Q_{1i}(\beta_{1,0}) - \{B_1 - A_1(G_1)^{-1}M_1\}E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}] + o_p(1).$$

According to the generalized information equality in Newey (1990), we have

$$B_1 - A_1(G_1)^{-1}M_1 = -E\Big\{\frac{\partial Q_1(\beta_{1,0})}{\partial \beta_1}\Big\} = E(Q_1\Phi_1^T),$$

so

$$f_1(q_0^1)\sqrt{n}(\hat{q}_{mr}^1 - q_0^1) = n^{-1/2}\sum_{i=1}^n [Q_{1i}(\beta_{1,0}) - E(Q_1\Phi_1^T)E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}] + o_p(1). \quad \text{(S10)}$$

For control group, note that

$$0 = \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\hat{\beta},\hat{q}^0,\hat{\gamma}^0)/(1-\pi_{1i}(\hat{\beta}_1))}{1+\lambda_0^T g_i(\hat{\beta},\hat{q}^0,\hat{\gamma}^0)/(1-\pi_{1i}(\hat{\beta}_1))} - \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\hat{\beta},\hat{q}^0,\hat{\gamma}^0)}{1-\pi_{1i}(\hat{\beta}_1)} \quad \text{(S11)}$$

$$+ \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\hat{\beta},\hat{q}^0,\hat{\gamma}^0)}{1-\pi_{1i}(\hat{\beta}_1)} - \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,\hat{q}^0,\hat{\gamma}^0)}{1-\pi_{1i}(\beta_{1*})} \quad \text{(S12)}$$

$$+ \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,\hat{q}^0,\hat{\gamma}^0)}{1-\pi_{1i}(\beta_{1*})} - \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,q_*^0,\hat{\gamma}^0)}{1-\pi_{1i}(\beta_{1*})} \quad \text{(S13)}$$

$$+ \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,q_*^0,\hat{\gamma}^0)}{1-\pi_{1i}(\beta_{1*})} - \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,q_*^0,\gamma_*^0)}{1-\pi_{1i}(\beta_{1*})} \quad \text{(S14)}$$

$$+ \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,q_*^0,\gamma_*^0)}{1-\pi_{1i}(\beta_{1*})}.$$

Denote $R_i(\beta, q^0, \gamma^0)$ as $g_i(\beta, q^0, \gamma^0)/(1-\pi_{1i}(\beta_1))$. Similarly, we have

$$(S11) = -\frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\hat{\beta},\hat{q}^0,\hat{\gamma}^0)^{\otimes 2}}{(1-\pi_{1i}(\hat{\beta}_1))^2}\hat{\lambda}_0 + o_p(n^{-1/2}),$$

and

$$(S12) = \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)g_i(\beta_*,\hat{q}^0,\hat{\gamma}^0)(\partial\pi_{1i}(\beta_{1*})/\partial\beta_1)^T}{(1-\pi_{1i}(\beta_{1*}))^2}(\hat{\beta}_1 - \beta_{1*}) + o_p(n^{-1/2}).$$

It can be shown that $\{(1-T_i)R_i(\beta_*, q^0, \hat{\gamma}^0) : \|q^0 - q_*^0\| < \epsilon\}$ forms a Donsker class and $(1-T_i)R_i(\beta_*, q_*^0, \hat{\gamma}^0)$ is L_2 continuous at q_*^0. Therefore,

$$(S13) = \frac{1}{n}\sum_{i=1}^n \frac{\partial E\{(1-T_i)R_i(\beta_*, q_*^0, \hat{\gamma}^0)\}}{\partial q^0}(\hat{q}^0 - q_*^0) + o_p(n^{-1/2}).$$

Similarly, we have that $\{(1-T_i)R_i(\beta_*, q_*^0, \gamma^0) : \|\gamma^0 - \gamma_*^0\| < \epsilon\}$ forms a Donsker class and $(1-T_i)R_i(\beta_*, q_*^0, \gamma_*^0)$ is L_2 continuous at γ_*^0. Then

$$(S14) = \frac{1}{n}\sum_{i=1}^n \frac{\partial E\{(1-T_i)R_i(\beta_*, q_*^0, \gamma_*^0)\}}{\partial \gamma^0}(\hat\gamma^0 - \gamma_*^0) + o_p(n^{-1/2}).$$

It is straightforward to see that both $E\{(1-T_i)R_i(\beta_*, q_*^0, \hat\gamma^0)\}$ and $E\{(1-T_i)R_i(\beta_*, q_*^0, \gamma_*^0)\}$ are zeros. Therefore, both (S13) and (S14) are zeros. Hence, by defining

$$M_0 = E\Big\{\frac{g(\beta_*, q_*^0, \gamma_*^0)(\partial\pi_1(\beta_{1*})/\partial\beta_1)^T}{1-\pi_1(\beta_{1*})}\Big\},$$

and (C7), it can be proved that

$$\sqrt{n}\hat\lambda_0 = (G_0)^{-1}\Big\{n^{-1/2}\sum_{i=1}^n \frac{\pi_{1i}(\beta_{1*}) - T_i}{1-\pi_{1i}(\beta_{1*})} g_i(\beta_*, q_*^0, \gamma_*^0) + n^{-1/2}\sum_{i=1}^n M_0 E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}\Big\} + o_p(1).$$

Next, note that

$$\begin{aligned}
0 =& \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)/(1-\pi_{1i}(\hat\beta_1))}{1+\hat\lambda_0^T g_i(\hat\beta, \hat q^0, \hat\gamma^0)/(1-\pi_{1i}(\hat\beta_1))}\psi_\tau(Y_i - \hat q_{mr}^0) \\
=& \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)\psi_\tau(Y_i - \hat q_{mr}^0)/(1-\pi_{1i}(\hat\beta_1))}{1+\hat\lambda_0^T g_i(\hat\beta, \hat q^0, \hat\gamma^0)/(1-\pi_{1i}(\hat\beta_1))} - \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)\psi_\tau(Y_i - \hat q_{mr}^0)}{1-\pi_{1i}(\hat\beta_1)} && \text{(S15)} \\
&+ \frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\pi_{1i}(\hat\beta_1)}\psi_\tau(Y_i - \hat q_{mr}^0) - \frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - \hat q_{mr}^0) && \text{(S16)} \\
&+ \frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - \hat q_{mr}^0) - \frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - q_0^0) && \text{(S17)} \\
&+ \frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\pi_{1i}(\beta_{1*})}\psi_\tau(Y_i - q_0^0).
\end{aligned}$$

It can be shown that

$$(S15) = -\Big\{\frac{1}{n}\sum_{i=1}^n (1-T_i)\frac{\psi_\tau(Y_i - \hat q_{mr}^0)}{(1-\pi_{1i}(\hat\beta_1))^2} g_i(\hat\beta, \hat q^0, \hat\gamma^0)^T\Big\}\hat\lambda_0 + o_p(n^{-1/2}),$$

$$(S16) = \Big\{\frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)\psi_\tau(Y_i - \hat q_{mr}^0)}{(1-\pi_{1i}(\beta_{1*}))^2}\Big(\frac{\partial\pi_{1i}(\beta_{1*})}{\partial\beta_1}\Big)^T\Big\}(\hat\beta_1 - \beta_{1*}) + o_p(n^{-1/2}),$$

and

$$(S17) = -f_0(q_0^0)(\hat q_{mr}^0 - q_0^0) + o_p(n^{-1/2}).$$

Let

$$B_0 = E\Big\{\frac{\psi_\tau(Y_0 - q_0^0)}{1-\pi_1(\beta_{1,0})}\Big(\frac{\partial\pi_1(\beta_{1,0})}{\partial\beta_1}\Big)^T\Big\},$$

we have

$$f_0(q_0^0)\sqrt{n}(\hat{q}_{mr}^0 - q_0^0) = -A_0\sqrt{n}\hat{\lambda}_0 + B_0\sqrt{n}(\hat{\beta}_1 - \beta_{1*}) + n^{-1/2}\sum_{i=1}^{n}\frac{(1-T_i)\psi_\tau(Y_i - q_0^0)}{1-\pi_{1i}(\beta_{1,0})} + o_p(1)$$
$$= n^{-1/2}\sum_{i=1}^{n}[Q_{0i}(\beta_{1,0}) + \{B_0 - A_0(G_0)^{-1}M_0\}E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}] + o_p(1).$$

From the generalized information equality in Newey (1990), it can be shown that

$$B_0 - A_0(G_0)^{-1}M_0 = E\Big\{\frac{\partial Q_0(\beta_{1,0})}{\partial \beta_1}\Big\} = -E(Q_0\Phi_1^T),$$

and

$$f_0(q_0^0)\sqrt{n}(\hat{q}_{mr}^0 - q_0^0) = n^{-1/2}\sum_{i=1}^{n}[Q_{0i}(\beta_{1,0}) - E(Q_0\Phi_1^T)E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}] + o_p(1). \quad \text{(S18)}$$

Finally, it can be concluded using (S10) and (S18) that

$$\sqrt{n}(\hat{q}_{mr} - q_0) = \sqrt{n}\{(\hat{q}_{mr}^1 - q_0^1) - (\hat{q}_{mr}^0 - q_0^0)\}$$
$$= f_1(q_0^1)^{-1}\big\{n^{-1/2}\sum_{i=1}^{n}[Q_{1i}(\beta_{1,0}) - E(Q_1\Phi_1^T)E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}]\big\}$$
$$- f_0(q_0^0)^{-1}\big\{n^{-1/2}\sum_{i=1}^{n}[Q_{0i}(\beta_{1,0}) - E(Q_0\Phi_1^T)E(\Phi_1^{\otimes 2})^{-1}\Phi_{1i}]\big\} + o_p(1).$$

This completes the proof.

Proof of Theorem 3. The consistency of least square estimator guarantees that if $\mathcal{W}$ contains a correct model, $\hat{\alpha}_\pi$ resembles a unit vector with the position of correct model taking the value of 1 and all the rest 0. If $\mathcal{F}$ contains the correct model, $\hat{\alpha}_l^t$ has the same result. Then this becomes a special case of theorem 1 with constrains only have two conditions, $g_i(X_i, \hat{\beta}, \hat{q}^1, \hat{\gamma}^1) = (\tilde{\pi}_i - \tilde{\theta}, \tilde{m}_i^1 - \tilde{\eta}^1)^T$ for $i \in S_1$ and $g_i(X_i, \hat{\beta}, \hat{q}^0, \hat{\gamma}^0) = (\tilde{\theta} - \tilde{\pi}_i, \tilde{m}_i^0 - \tilde{\eta}^0)^T$ for $i \in S_0$. So consistency of ensemble estimator can be established here.

References

[1] Newey, W.K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics,* 5, 99–135.

[2] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

S2 Non-normal outcome

In specific, we consider the outcome models for Y_t as

$$\begin{aligned} Y_t =& \gamma_{11}^t + \gamma_{12}^t W_1 + \gamma_{13}^t W_2 + \gamma_{14}^t W_3 + \gamma_{15}^t W_4 + \gamma_{16}^t W_5 + \gamma_{17}^t W_6 + \epsilon_t, \\ & \text{with transformation } W_k = X_k^2,\ k = 1,2,3,4,5,6, \end{aligned}$$

where X_i is generated from a 8-dimensional normal distribution with mean zero and identity covariance matrix, and ϵ_t's are independently from $\chi^2(1) - 1$ for $t = 0, 1$. We generate T from a Bernoulli distribution with probability $\pi(X)$ and consider two choices for $\pi(X)$:

$$\begin{aligned} (1)\ \pi(X) =& 1/\{1 + \exp(\beta_{11} + \beta_{12}X_1 + \beta_{13}X_2 + \beta_{14}X_4 + \beta_{15}X_5 + \beta_{16}X_7 + \beta_{17}X_8)\}, \\ (2)\ \pi(X) =& 1/\{1 + \exp(\beta_{21} + \beta_{22}W_1 + \beta_{23}W_2 + \beta_{24}W_4 + \beta_{25}W_5 + \beta_{26}W_7 + \beta_{27}W_8)\}, \\ & \text{with transformation } W_k = exp(X_k/2),\ k = 1,2,4,5,7,8. \end{aligned}$$

The true parameters are set as

$$\begin{aligned} &\gamma_{1,0}^1 = (1, 1.5, 1.5, 1.5, 3.5, 3.5, 3.5)^T,\ \gamma_{1,0}^0 = (-1.5, 1.5, 1.5, 1.5, 3.5, 3.5, 3.5)^T, \\ &\beta_{1,0} = (-0.7, 0.7, 0.3, 0.7, 0.3, 0.7, 0.3)^T,\ \beta_{2,0} = (-0.7, 0.7, 0.3, 0.7, 0.3, 0.3, -0.7)^T. \end{aligned}$$

To implement our proposed methods, we postulate two functions for $f(Y_t|X)$ as follows:

$$(A)\quad f_1^t(X,\gamma_1^t) = \frac{1}{2^{1/2}\Gamma(1/2)} e^{-\frac{Y-(\gamma_{10}^t+\gamma_{11}^t X_1^2+\ldots+\gamma_{18}^t X_8^2-1)}{2}} \sqrt{Y - (\gamma_{10}^t + \gamma_{11}^t X_1^2 + \ldots + \gamma_{18}^t X_8^2 - 1)},$$

when $Y - (\gamma_{10}^t + \gamma_{11}^t X_1^2 + \ldots + \gamma_{18}^t X_8^2 - 1) \geq 0$, and 0 otherwise.

$$(B)\quad f_2^t(X,\gamma_2^t) = \frac{1}{\sqrt{2\pi}} exp\Big\{ -\frac{1}{2}\{Y - (\gamma_{20}^t + \gamma_{21}^t exp(X_1/2) + \ldots + \gamma_{28}^t exp(X_8/2))\}^2 \Big\},$$

and two models for $\pi(X)$ are the same as before. The results based on 1,000 replications are given in Table S1 and the proposed estimators have the similar performance with the normal outcome variables.

S3 Computation time

According to your suggestion, we calculated the average computation time of the estimators $\hat{q}_{mr(1111)}$ and $\hat{q}_{es(1111)}$ for 50th percentile QTE based on $n = 500$ with 100

Table S1: Bias (×100), standard deviation (SD) and mean square error (MSE) of the QTE estimators with non-normal outcomes when f_1 is true and $n = 500$.

Method	π_1 is true			π_2 is true		
	Bias	SD	MSE	Bias	SD	MSE
			$\tau = 0.25$			
$\hat{q}_{naive}$	-26.2392	0.7684	0.6594	-44.4387	0.8385	0.9006
$\hat{q}_{ipw(1000)}$	3.7349	0.9155	0.8396	-73.8911	0.9227	1.3973
$\hat{q}_{ipw(0100)}$	101.2199	0.8530	1.7521	-10.2209	0.9139	0.8457
$\hat{q}_{mi(0010)}$	1.1427	0.1781	0.0318	0.3398	0.1878	0.0353
$\hat{q}_{mi(0001)}$	357.0578	1.1531	14.0787	-14.5027	1.2022	1.4663
$\hat{q}_{aipw(1010)}$	-0.8238	0.6055	0.3667	-22.7071	0.5957	0.4064
$\hat{q}_{aipw(1001)}$	24.1214	1.1308	1.3368	-92.2679	0.9744	1.8007
$\hat{q}_{aipw(0110)}$	22.8054	0.5883	0.3981	-1.5518	0.5638	0.3181
$\hat{q}_{aipw(0101)}$	128.2054	1.0241	2.6924	-23.8312	0.9935	1.0438
$\hat{q}_{mr(1110)}$	0.0135	0.4646	0.2159	-1.1094	0.4501	0.2027
$\hat{q}_{mr(1101)}$	9.7674	0.8918	0.8048	-11.4204	0.8661	0.7631
$\hat{q}_{mr(0111)}$	10.8311	0.4548	0.2185	-0.6411	0.4527	0.2049
$\hat{q}_{mr(1011)}$	0.0092	0.4754	0.2260	-7.0105	0.4386	0.1973
$\hat{q}_{mr(1111)}$	0.3021	0.4745	0.2252	-0.9825	0.4513	0.2038
$\hat{q}_{es(1110)}$	0.5484	0.4396	0.1933	-1.1412	0.4365	0.1906
$\hat{q}_{es(1101)}$	12.3452	0.9119	0.8469	-18.8812	0.8698	0.7921
$\hat{q}_{es(0111)}$	6.7407	0.4330	0.1920	0.1216	0.4442	0.1973
$\hat{q}_{es(1011)}$	0.5560	0.4413	0.1947	-2.9825	0.4316	0.1872
$\hat{q}_{es(1111)}$	0.5623	0.4446	0.1977	0.2567	0.4348	0.1890
			$\tau = 0.5$			
$\hat{q}_{naive}$	-31.7115	0.9806	1.0621	-64.8359	1.0693	1.5638
$\hat{q}_{ipw(1000)}$	16.5392	1.2693	1.6385	-111.8816	1.1661	2.6116
$\hat{q}_{ipw(0100)}$	171.8833	1.1686	4.3199	-15.2507	1.2087	1.4843
$\hat{q}_{mi(0010)}$	-0.1447	0.1620	0.0262	0.2422	0.1753	0.0307
$\hat{q}_{mi(0001)}$	237.3227	0.9780	6.5887	-32.1504	1.0145	1.1326
$\hat{q}_{aipw(1010)}$	-4.0017	0.6464	0.4194	-1.5704	0.6690	0.4478
$\hat{q}_{aipw(1001)}$	36.3969	1.3585	1.9779	-124.2742	1.3601	3.3942
$\hat{q}_{aipw(0110)}$	-0.2743	0.6185	0.3826	5.3990	0.6625	0.4418
$\hat{q}_{aipw(0101)}$	216.1140	1.3963	6.6203	-17.0298	1.4205	2.0467
$\hat{q}_{mr(1110)}$	0.7907	0.5135	0.2637	3.5194	0.5047	0.2559
$\hat{q}_{mr(1101)}$	27.2148	1.2344	1.5978	-19.2658	1.1038	1.2554
$\hat{q}_{mr(0111)}$	7.1908	0.4981	0.2532	3.7916	0.5025	0.2539
$\hat{q}_{mr(1011)}$	0.3458	0.5159	0.2662	-0.4105	0.4947	0.2448
$\hat{q}_{mr(1111)}$	1.0543	0.5185	0.2690	3.6672	0.5107	0.2622
$\hat{q}_{es(1110)}$	0.0573	0.4910	0.2411	2.3996	0.4894	0.2401
$\hat{q}_{es(1101)}$	30.5048	1.2677	1.7001	-32.5951	1.1434	1.4136
$\hat{q}_{es(0111)}$	3.4852	0.4728	0.2248	3.5262	0.4833	0.2348
$\hat{q}_{es(1011)}$	-0.1860	0.4974	0.2474	-0.5224	0.4795	0.2300
$\hat{q}_{es(1111)}$	0.2747	0.4903	0.2404	3.6215	0.4879	0.2394

Table S1: Continued.

Method	π_1 is true			π_2 is true		
	Bias	SD	MSE	Bias	SD	MSE
			$\tau = 0.75$			
$\hat{q}_{naive}$	-49.4363	1.4203	2.2617	-96.0108	1.4848	3.1264
$\hat{q}_{ipw(1000)}$	17.4341	1.8824	3.5738	-164.2372	1.6591	5.4499
$\hat{q}_{ipw(0100)}$	249.6236	1.8287	9.5754	-26.2986	1.8061	3.3312
$\hat{q}_{mi(0010)}$	0.5811	0.1896	0.0360	1.2686	0.2141	0.0460
$\hat{q}_{mi(0001)}$	82.8655	1.0315	1.7506	-71.5473	1.0786	1.6754
$\hat{q}_{aipw(1010)}$	-4.7348	0.9467	0.8985	20.8097	0.8697	0.7996
$\hat{q}_{aipw(1001)}$	31.8906	1.8089	3.3737	-143.7272	1.6337	4.7349
$\hat{q}_{aipw(0110)}$	-27.8102	0.8619	0.8201	9.5660	0.9041	0.8266
$\hat{q}_{aipw(0101)}$	248.3869	1.8786	9.6988	-35.4038	1.7823	3.3021
$\hat{q}_{mr(1110)}$	0.6859	0.7476	0.5589	2.9267	0.6900	0.4770
$\hat{q}_{mr(1101)}$	25.2273	1.7812	3.2361	-31.6701	1.6877	2.9487
$\hat{q}_{mr(0111)}$	3.1589	0.7094	0.5043	3.1041	0.6853	0.4705
$\hat{q}_{mr(1011)}$	1.0178	0.7414	0.5498	3.0364	0.6781	0.4607
$\hat{q}_{mr(1111)}$	1.6141	0.7479	0.5596	2.7471	0.6882	0.4744
$\hat{q}_{es(1110)}$	1.2045	0.6877	0.4731	3.9707	0.6893	0.4768
$\hat{q}_{es(1101)}$	30.5546	1.8009	3.3367	-45.5932	1.7060	3.1184
$\hat{q}_{es(0111)}$	2.0497	0.6763	0.4578	3.5178	0.6650	0.4435
$\hat{q}_{es(1011)}$	-0.6385	0.7031	0.4945	1.7815	0.6581	0.4334
$\hat{q}_{es(1111)}$	0.0857	0.6900	0.4761	4.3042	0.6637	0.4423

Table S2: Average computation time(s) of $\hat{q}_{mr(1111)}$ and $\hat{q}_{es(1111)}$.

Method		$\hat{q}_{mr(1111)}$	$\hat{q}_{es(1111)}$
π_1 is true	f_1 is true	1.8211	1.3089
	f_2 is true	1.8246	1.4244
π_2 is true	f_1 is true	1.8505	1.3442
	f_2 is true	1.7298	1.3553

replications. The results are shown in Table S2. As expected, it can be seen that the ensemble estimator effectively reduces computational burden with less time than the ordinary multiply robust estimator.