A DATA FUSION METHOD FOR QUANTILE TREATMENT EFFECTS

Yijiao Zhang and Zhongyi Zhu

Department of Statistics and Data Science, Fudan University

Supplementary Materials

Details for density estimation are provided in Section S1. Section S2 extends our proposed fused quantile treatment estimator (FQTE) to the missing at random (MAR) case. Theorems and lemmas of Section 4 are proved in Section S3. Section S4 provides additional simulation results.

S1 Density Estimation

To obtain estimators for the influence functions, we need to estimate $f_t(q_{t,p})$. We only need a consistent estimator regardless of the rate. Here we provide two possible ways. The first one is an IPW kernel estimator, which is formulated as

$$\hat{f}_1(y) = \frac{1}{nh}\sum_{i=1}^{n}\frac{T_i}{\hat{e}\left(X_i, S_i\right)}\kappa\left(\frac{Y_i - y}{h}\right), \quad \hat{f}_0(y) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1-T_i}{1-\hat{e}\left(X_i, S_i\right)}\kappa\left(\frac{Y_i - y}{h}\right), \tag{S1.1}$$

where $\kappa(u)$ is a kernel function such as $\kappa(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and $h \to 0$ is a bandwidth. We choose $h \asymp n^{-1/5}$. To relax the dependence of its consistency on the correct specification of the propensity score model, we can choose nonparametric estimating methods for obtaining $\hat{e}(X, S)$. For example, when the dimension of the covariates is low, we can use a logistic power series approximation for estimating $e(X, S)$, see Firpo (2007) for details. In cases where the dimension of covariates is relatively high, we may use flexible machine learning methods for estimating $e(X, S)$, such as random forests, boosting, neural nets, and various hybrids and ensembles of these methods, see, for example, Westreich et al. (2010) and references therein.

An alternative for density estimation is to follow the idea of constructing a doubly robust estimator. Specifically, given a pre-specified outcome model and propensity score model, we can estimate $f_t(y)$ by

$$\hat{f}_t(y) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{w}_{t,i}\left\{\frac{1}{h}\kappa\left(\frac{Y_i - y}{h}\right) - \frac{\mathrm{d}G_t\left(y \mid X_i, S_i; \hat{\theta}_t\right)}{\mathrm{d}y}\right\} + \frac{\mathrm{d}G_t\left(y \mid X_i, S_i; \hat{\theta}_t\right)}{\mathrm{d}y}\right], \tag{S1.2}$$

for $t = 0, 1$, where $\kappa(u)$ is a kernel function same as that in (S1.1) and $h \asymp n^{-1/5}$. Under regular conditions in kernel density estimation, the density estimator $\hat{f}_t(y)$ in (S1.2) is consistent for $f_t(y)$ for $t = 0, 1$, if either Assumption 3 or 4 holds, i.e, the density estimator also enjoys the property of double robustness.

S2 Extension to the MAR Case

Now we relax Assumption 5 to a weaker missing at random (MAR) assumption.

Assumption S2.1 (Missin At Random)**.** $R \perp\!\!\!\perp S | (Y, T, X)$.

Under Assumption S2.1, $\{Y_i, T_i, {X_i}^\top, i = 1, \ldots, N\}$ are no longer i.i.d. samples and the missing mechanism depends on the basic information (Y, T, X). However, we can also collect the two datasets by modeling the missing mechanism. Define $\pi(Y, T, X) = \mathrm{pr}(R = 1 | Y, T, X)$. Here we specify a parametric model $\pi(Y, T, X; \beta)$ (for example, a logistic regression model) for $\pi(Y, T, X)$.

Assumption S2.2 (Missing Probability)**.** *The parametric model* $\pi(Y, T, X; \beta)$ *is a correct specification for* $\pi(Y, T, X)$*; that is,* $\pi(Y, T, X) = \pi(Y, T, X; \beta^*)$ *with* β^* *the true model parameter.*

Similar to what we have done in Section 2.2, we can also obtain unbiased quantile and QTE estimators based on the validation sample. We still focus on the DR method and the same arguments can be done to obtain OR and IPW estimators. Let $\hat{\beta}$ be a $N^{1/2}$-consistent estimator for β^*. For simplicity, denote $\pi_(Y_i, T_i, X_i; \beta^*)$ as π_i and $\pi_i(Y_i, T_i, X_i; \hat{\beta})$ as $\hat{\pi}_i$. We propose the weighted quantile estimators $\hat{q}^{\mathcal{V}}_{t,p}$ by solving a weighted version of (2.2), i.e.

$$\sum_{i=1}^{n} \frac{1}{\hat{\pi}_i} \Psi_t(O_i; q, \hat{\eta}_t) \approx 0, \tag{S2.1}$$

where Ψ_t is defined in (2.1) and $\hat{\eta}_t = (\hat{\theta}_t, \hat{\alpha})$ is also obtained by solving corresponding weighted estimating equations,

$$\sum_{i=1}^{n} \frac{1}{\hat{\pi}_i} L_t(Y_i, T_i, X_i, S_i; \theta_t) = 0, \quad \sum_{i=1}^{n} \frac{1}{\hat{\pi}_i} h(T_i, X_i, S_i; \alpha) = 0.$$

The DR QTE estimator is then obtained as $\hat{\Delta}^{\mathcal{V}}_p = \hat{q}^{\mathcal{V}}_{1,p} - \hat{q}^{\mathcal{V}}_{0,p}$.

Let $\hat{q}^{\mathrm{Conf}}_{t,p}$ and $\hat{\eta}^{\mathrm{Conf}}_t$ same as that in Section 2.3, we can similarly modify (3.6) using a weigted version as

$$\hat{C}_t = \frac{1}{N} \sum_{i=1}^{n} \frac{1}{\hat{\pi}_i} \phi_t(U_i; \hat{q}^{\mathrm{Conf}}_{t,p}, \hat{\eta}^{\mathrm{Conf}}_t). \tag{S2.2}$$

The next theorem establishes the asymptotic normality of $\hat{q}^{\mathcal{V}}_{t,p}$ for $t = 0, 1$ in the MAR case.

Theorem S2.1. *Under Assumptions 3 or 4, S2.1, S2.2, Conditions S3.1 and S3.3 in Section S3.1, the DR quantile estimators $\hat{q}^{\mathcal{V}}_{1,p}$ and $\hat{q}^{\mathcal{V}}_{0,p}$ obtained by solving (S2.1) are asymptotically normal with*

$$\hat{q}^{\mathcal{V}}_{t,p} - q_{t,p} = \frac{1}{N} \sum_{i=1}^{N} \frac{R_i}{\pi_i} \psi_t(O_i; q_{t,p}, \eta^*_t) + o_p(\frac{1}{n^{1/2}}), \tag{S2.3}$$

where $\psi_t(O_i; q_{t,p}, \eta^*_t)$ is given in (2.4) in Section 2.3.

Lemma S2.1. *Under Assumptions S2.1, S2.2 and Conditions S3.2 and S3.3 in Section S3.1, we have*

$$\hat{C}_t = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{R_i}{\pi_i} - 1\right)\phi_t(U_i; q_{t,p}^{\mathrm{Conf},*}, \eta_t^{\mathrm{Conf},*}) + o_p(\frac{1}{n^{1/2}}). \tag{S2.4}$$

Based on the asymptotic linear forms, we can also establish the asymptotic normality with covariance in explicit forms. The following Proposition is an extension of Proposition 1.

Proposition S2.1. *Under the assumptions in Theorem S2.1 and Lemma S2.1, as $n \to \infty$, then*

$$n^{1/2}\begin{pmatrix} \hat{\Delta}_p^{\mathcal{V}} - \Delta_p \\ \hat{C} \end{pmatrix} \longrightarrow \mathcal{N}\left\{0, \begin{pmatrix} \sigma_{\mathcal{V}}^2 & \varrho^\top \\ \varrho & \Sigma_{\mathrm{ep}} \end{pmatrix}\right\},$$

in distribution, where $\sigma_{\mathcal{V}}^2 = \nu\,\mathrm{var}[\frac{R_i}{\pi_i}(\psi_{1,i} - \psi_{0,i})]$, $\varrho = \nu\,\mathrm{cov}[\frac{R_i}{\pi_i}(\psi_{1,i} - \psi_{0,i}), (\frac{R_i}{\pi_i} - 1)(\phi_{1,i}^\top, \phi_{0,i}^\top)^\top]$, $\Sigma_{\mathrm{ep}} = \begin{pmatrix} \Sigma_1 & \Sigma_{01}^\top \\ \Sigma_{01} & \Sigma_0 \end{pmatrix}$ *with* $\Sigma_{01} = \nu\,\mathrm{cov}[(\frac{R_i}{\pi_i} - 1)\phi_{0,i}, (\frac{R_i}{\pi_i} - 1)\phi_{1,i}]$ *and* $\Sigma_t = \nu\mathrm{var}[(\frac{R_i}{\pi_i} - 1)\phi_{t,i}]$, *for* $t = 0, 1$.

Based on Proposition S2.1, we can estimate $(\varrho, \Sigma_{\mathrm{ep}})$ by

$$\hat{\varrho} = \frac{\nu_n}{N}\sum_{i=1}^{N}\frac{R_i}{\hat{\pi}_i}\left(\frac{R_i}{\hat{\pi}_i} - 1\right)(\hat{\psi}_{1,i} - \hat{\psi}_{0,i})(\hat{\phi}_{1,i}^\top, \hat{\phi}_{0,i}^\top)^\top, \quad \hat{\Sigma}_t = \frac{\nu_n}{N}\sum_{i=1}^{N}\left(\frac{R_i}{\hat{\pi}_i} - 1\right)^2\hat{\phi}_{t,i}\hat{\phi}_{t,i}^\top$$

$$\hat{\Sigma}_{01} = \frac{\nu_n}{N}\sum_{i=1}^{N}\left(\frac{R_i}{\hat{\pi}_i} - 1\right)^2\hat{\phi}_{0,i}\hat{\phi}_{1,i}^\top, \quad \hat{\sigma}_{\mathcal{V}}^2 = \frac{\nu_n}{N}\sum_{i=1}^{N}\left(\frac{R_i}{\hat{\pi}_i}\right)^2\left(\hat{\psi}_{1,i} - \hat{\psi}_{0,i}\right)^2.$$

Then we can obtain our FQTE through (3.9).

S3 Proof for Main Theorems

S3.1 Regular Conditions

Condition S3.1. *(a) The parameter space $\mathcal{Q}_{t,p}$ of $q_{t,p}$ is compact for $t = 0, 1$. (b) The pth quantile for $F_t(\cdot)$ is unique for $t = 0, 1$. (c) The parameter spaces $\Theta_t \times \Omega$ of $\eta_t = (\theta_t, \alpha)$ are compact for $t = 0, 1$. (d) $\inf_{(X^\top, S^\top)^\top \in \mathbb{C}} \inf_\alpha e(X, S; \alpha) > \epsilon$, $\sup_{(X^\top, S^\top)^\top \in \mathbb{C}} \sup_\alpha e(X, S; \alpha) < 1 - \epsilon$ for some $\epsilon > 0$ where $\mathbb{C} = \mathbb{X} \times \mathbb{S}$. (e)(i) $e(X, S; \alpha)$ is differentiable at α^* with a bounded derivative; (ii) $G_t(q \mid X, S; \theta_t)$ is differentiable at θ_t^* uniformly in q in a neighbourhood of $q_{t,p}$, with bounded derivatives for $t = 0, 1$; (iii) $G_t(q \mid X, S; \theta_t)$ is differentiable at $q_{t,p}$ uniformly in θ_t in a neighbourhood of θ_t^*, with bounded derivatives for $t = 0, 1$.*

Condition S3.1(a) is commonly used for quantile estimation. Condition S3.1(b) is used for identification of $q_{t,p}^{\mathrm{Conf}}$. Condition S3.1(c)-6(e) are similar to conditions assumed in Han et al. (2019). It is easy to verify that Condition S3.1(e)(i) holds for logistic linear models and Condition S3.1(e)(ii),(iii) hold for normal linear models.

The following Condition is similar to Condition S3.1 made in the unconfounded case.

Condition S3.2. *(a) The parameter space $\mathcal{Q}_{t,p}^{\mathrm{Conf}}$ for $q_{t,p}^{\mathrm{Conf}}$ is compact for*

$t = 0, 1$. *(b) The parameter* $q_{t,p}^{\mathrm{Conf}}$ *is the unique solution to (2.5) for* $t = 0, 1$. *(c) The parameter spaces* $\Theta_t^{\mathrm{Conf}} \times \Omega^{\mathrm{Conf}}$ *of* $\eta_t^{\mathrm{Conf}} = (\theta_t^{\mathrm{Conf}}, \alpha^{\mathrm{Conf}})$ *are compact for* $t = 0, 1$. *(d)* $\inf_{X \in \mathbb{X}} \inf_{\alpha^{\mathrm{Conf}}} \tilde{e}(X; \alpha^{\mathrm{Conf}}) > \epsilon$, $\sup_{X \in \mathbb{X}} \sup_{\alpha^{\mathrm{Conf}}} \tilde{e}(X; \alpha^{\mathrm{Conf}}) < 1 - \epsilon$ *for some* $\epsilon > 0$. *(e)(i)* $\tilde{e}(X; \alpha^{\mathrm{Conf}})$ *is differentiable at* $\alpha^{\mathrm{Conf},*}$ *with a bounded derivative; (ii)* $\tilde{G}_t(q \mid X; \theta_t^{\mathrm{Conf},*})$ *is differentiable at* $\theta_t^{\mathrm{Conf},*}$ *uniformly in* q *in a neighbourhood of* $q_{t,p}^{\mathrm{Conf}}$ *with bounded derivatives for* $t = 0, 1$*; (iii)* $\tilde{G}_t(q \mid X; \theta_t^{\mathrm{Conf}})$ *is differentiable at* $q_{t,p}$ *uniformly in* θ_t^{Conf} *in a neighbourhood of* $\theta_t^{\mathrm{Conf},*}$ *with bounded derivatives for* $t = 0, 1$.

Condition S3.3. *(a)* $\inf_{(Y,T,X^\top) \in \mathbb{U}} \pi(Y, T, X) > \epsilon$*, for some* $\epsilon > 0$. *(b)* $\pi(Y, T, X; \beta)$ *is differentiable at* β^* *with a bounded derivative.*

S3.2 Proof of Theorem 1

We first establish the consistency of $\hat{q}_{t,p}^{\mathcal{V}}$. Note that it can be seen as a Z-estimator of the following Z-estimation equations:

$$\frac{1}{n} \sum_{i=1}^{n} \Psi_t(O_i; q, \hat{\eta}_t) = 0.$$

We first prove that the Z-estimating function $\Psi_t(O_i; q, \eta_t^*)$ in expection equals $F_t(q) - p$ for $t = 0, 1$.

By the law of iterated expectation, and if Assumption 3 holds, we have

$$\mathbb{E}\{\Psi_1(O_i;q,\eta_t^*)\} = \mathbb{E}\left[\mathbb{E}\left[\frac{T\{I(Y\le q)-G_1(q|X,S;\theta_1^*)\}}{e(X,S;\alpha^*)} + G_1(q|X,S;\theta_1^*)|X,S\right]\right] - p$$
$$= \mathbb{E}\left[\frac{\mathrm{pr}(T=1|X,S)}{e(X,S;\alpha^*)}\mathbb{E}\left[\{I(Y(1)\le q)-G_1(q|X,S;\theta_1^*)\}|T=1,X,S\right]\right.$$
$$\left. + G_1(q|X,S;\theta_1^*)\right] - p$$
$$= \mathbb{E}\left[\frac{\mathrm{pr}(T=1|X,S)}{e(X,S;\alpha^*)}\{G_1(q|X,S;\theta_1^*)-G_1(q|X,S;\theta_1^*)\} + G_1(q|X,S;\theta_1^*)\right] - p$$
$$= \mathrm{pr}\{Y(1)\le q\} - p,$$

and similarly we have $\mathbb{E}\{\Psi_0(O_i;q,\eta_t^*)\} = \mathrm{pr}\{Y(0)\le q\} - p$.

Note that if Assumption 4 holds, we also have

$$\mathbb{E}\{\Psi_1(O_i;q,\eta_t^*)\} = \mathbb{E}\left[\mathbb{E}\left[\frac{T\{I(Y\le q)-G_1(q|X,S;\theta_1^*)\}}{e(X,S;\alpha^*)} + G_1(q|X,S;\theta_1^*)|X,S\right]\right] - p$$
$$= \mathbb{E}\,[\ \frac{\mathrm{pr}(T=1|X,U)}{e(X,U;\alpha^*)}\mathbb{E}\{T[I(Y\le y)-G_1(q|X,S;\theta_1^*)]\,|T=1,X,S\}$$
$$+ G_1(q|X,S;\theta_1^*)\ ] - p$$
$$= \mathbb{E}\left[\mathbb{E}\{I\{Y(1)\le y\} - G_1(q|X,S;\theta_1^*)|X,S\} + G_1(q|X,S;\theta_1^*)\right] - p$$
$$= \mathrm{pr}\{Y(1)\le q\} - p,$$

and similarly $\mathbb{E}\{\Psi_0(O_i;q,\eta_t^*)\} = \mathrm{pr}\{Y(0)\le q\} - p$.

We then claim that

$$\sup_{q\in\mathcal{Q}_{t,p},\eta_t\in\Theta_t\times\Omega}\left|\frac{1}{n}\sum_{i=1}^n\Psi_t(O_i;q,\eta_t) - \mathbb{E}\{\Psi_t(O;q,\eta_t)\}\right| = o_p(1). \qquad \text{(S3.1)}$$

According to (S3.1), we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}\Psi_t(O_i;\hat{q}^{\mathcal{V}}_{t,p},\hat{\eta}_t)-\mathbb{E}\left\{\Psi_t(O;\hat{q}^{\mathcal{V}}_{t,p},\hat{\eta}_t)\right\}\right|=o_p(1). \tag{S3.2}$$

Since $\mathbb{E}\{\Psi_t(O;q,\eta_t)\}$ is continuous at $\eta_t=\eta_t^*$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}\Psi_t(O_i;\hat{q}^{\mathcal{V}}_{t,p},\hat{\eta}_t)-F_t(\hat{q}^{\mathcal{V}}_{t,p})+p\right|=o_p(1). \tag{S3.3}$$

Note that $\frac{1}{n}\sum_{i=1}^{n}\Psi_t(O_i;\hat{q}^{\mathcal{V}}_{t,p},\hat{\eta}_t)=o_p(1)$ and we have $\inf_{|q-q_{t,p}|\geq\epsilon}|F_t(q)-p|>0$ by Condition S3.1(b). Together with (S3.3), using similar techniques for proving Theorem 5.9 in Van der Vaart (2000), we have $\hat{q}^{\mathcal{V}}_{t,p}-q_{t,p}=o_p(1)$.

Now we turn to prove (S3.1) when $t=1$. The proof for $t=0$ is similar. By the definition of $\Psi_t(O_i;q_{t,p},\eta_t^*)$, it suffices to prove that $\mathcal{M}=\{\Psi_1(O;q,\eta_1);q\in\mathcal{Q}_{t,p},\eta_1=(\theta_1,\alpha)\in\Theta_1\times\Omega\}$ is a Glivenko-Cantelli class. Indeed, we have a stronger result that $\mathcal{M}$ is a Donsker class. To prove this, let us define $\mathcal{M}_1=\{\frac{1}{e(X,S;\alpha)},\alpha\in\Omega\}$, $\mathcal{M}_2=\{I(Y\leq q),q\in\mathcal{Q}_{t,p}\}$ and $\mathcal{M}_3=\{G_1(q|X,S;\theta_1),\theta_1\in\Theta_1\}$. It is well-known that $\mathcal{M}_2$ is Donsker. By Condition S3.1(c)) and S3.1(e)(i), we know the class $\{e(X,S;\alpha),\alpha\in\Omega\}$ is Donsker by Example 19.7 of Van der Vaart (2000). Together with Condition S3.1(d), we have that $\mathcal{M}_1$ is Donsker by Example 2.10.9 of Van der Vaart and Wellner (1996). Similarly, we have that $\mathcal{M}_3$ is Donsker under Conditions S3.1(a), S3.1(b), and S3.1(e)(ii),(iii). Since T is bounded, from the preservation of Donsker classes (corollary 9.32 of Kosorok (2008)), we know that $\mathcal{M}$ is Donsker. Hence (S3.1) holds and the part of consistency has been proved.

Now we establish the asymptotic normality of $\hat{q}_{t,p}^{\mathcal{V}}$. First, we show that $\Psi_t(O; q, \eta_t)$ is L_2 continuous at $(q_{t,p}, \eta_t^*)$. Indeed, we have

$$\int \{\Psi_1(O; 1, \eta_1) - \Psi_1(O; q_{1,p}, \eta_1^*)\}^2 \,\mathrm{dP} \leq \int \left\{\Psi_1(O; q, \eta_1) - \Psi_1^{\mathrm{DR}}(O; q_{1,p}, \eta_1)\right\}^2 \mathrm{dP}$$
$$+ \int \{\Psi_1(O; q_{1,p}, \eta_1) - \Psi_1(O; q_{1,p}, \eta_1^*)\}^2 \,\mathrm{dP} \leq C_1 \left|q - q_{1,p}\right|^2 + C_2 \left\|\eta_1 - \eta_1^*\right\|^2 \to 0,$$

as $(q, \eta_t) \to (q_{t,p}, \eta_t^*)$, where the second inequality uses Condition S3.1(d) and 6(e). Together with the fact that $\mathcal{M}$ is Donsker as we proved before, an application of Theorem 5.31 of Van der Vaart (2000) gives us

$$\begin{aligned} \hat{q}_{t,p}^{\mathcal{V}} - q_{t,p} = &- \frac{\mathbb{E}\{\Psi_t(O; q_{t,p}, \hat{\eta}_t)\} + \frac{1}{n}\sum_{i=1}^n \Psi_t(O_i; q_{t,p}, \eta_t^*)}{f_t(q_{t,p})} \\ &+ o_p\left(\frac{1}{n^{1/2}} + |\mathbb{E}\{\Psi_t(O; q_{t,p}, \hat{\eta}_t)\}|\right). \end{aligned} \tag{S3.4}$$

Under Conditions S3.1(d) and S3.1(e)(i), (ii), we know that $\Psi_t(O; q_{t,p}, \hat{\eta}_t)$ is differentiable at η_t^* with bounded derivatives, by delta method, we have that $n^{1/2}\mathbb{P}\Psi_t(O; q_{t,p}, \hat{\eta}_t) = O_p(1)$. For notational simplicity, define

$$\psi_{1,i}^{\mathrm{DR},*} = -\frac{\Psi_t(O_i; q_{t,p}, \eta_t^*)}{f_1(q_{1,p})} \tag{S3.5}$$

By a Taylor expansion of $\Psi_t(O; q_{t,p}, \hat{\eta}_t)$ at η_t^* in (S3.4), we have that

$$
\begin{aligned}
&\hat{q}_{t,p}^{\mathcal{V}} - q_{t,p} = \frac{1}{n}\sum_{i=1}^{n} \psi_{t,i}^{\mathrm{DR},*} + \mathbb{E}\left(\frac{\partial \psi_{t,i}^{\mathrm{DR},*}}{\partial {\theta_1}^{\mathrm{T}}}\right)(\hat{\theta}_1 - \theta_1^*) \\
&+ \mathbb{E}\left(\frac{\partial \psi_{t,i}^{\mathrm{DR},*}}{\partial \alpha^{\mathrm{T}}}\right)(\hat{\alpha} - \alpha^*) + o_p(\frac{1}{n^{1/2}}) \\
=&\frac{1}{n}\sum_{i=1}^{n} \psi_{t,i}^{\mathrm{DR},*} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\frac{\partial \psi_{t,i}^{\mathrm{DR},*}}{\partial {\theta_t}^{\mathrm{T}}}\right)(\mathbb{E}\dot{L}_{ti}^*)^{-1} L_{ti}^* \\
&- \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\frac{\partial \psi_{t,i}^{\mathrm{DR},*}}{\partial \alpha^{\mathrm{T}}}\right)(\mathbb{E}\dot{h_i}^*)^{-1} h_i^* + o_p(\frac{1}{n^{1/2}}).
\end{aligned}
$$

We have the following calculations:

$$
\mathbb{E}\left(\frac{\partial \psi_{1,i}^{\mathrm{DR},*}}{\partial \theta_1^{\mathrm{T}}}\right) = -\frac{1}{f_1(q_{a,p})}\mathbb{E}\left\{(1 - \frac{T_i}{e_i^*})\dot{G}_{1i,p}^*\right\}(\mathbb{E}\dot{L}_{1i}^*)^{-1} L_{1i}^*,
$$

$$
\mathbb{E}\left(\frac{\partial \psi_{0,i}^{\mathrm{DR},*}}{\partial \theta_0^{\mathrm{T}}}\right) = \frac{1}{f_0(q_{0,p})}\mathbb{E}\left\{(1 - \frac{1-T_i}{1-e_i^*})\dot{G}_{0i,p}^*\right\}(\mathbb{E}\dot{L}_{0i}^*)^{-1} L_{0i}^*,
$$

$$
\mathbb{E}\left(\frac{\partial \psi_{t,i}^{\mathrm{DR},*}}{\partial \alpha^{\mathrm{T}}}\right) = H_t \quad t = 0, 1.
$$

Note that under Assumption 4, $H_t = 0$ for $t = 0, 1$. Under Assumption 5,

$$
\mathbb{E}\left\{\frac{\partial h\left(T, X, S; \alpha^*\right)}{\partial \alpha^{\mathrm{T}}}\right\} = -\mathbb{E}\left\{h\left(T, X, S; \alpha^*\right)^{\otimes 2}\right\} = -\Sigma_\alpha
$$

Therefore, we can always replace $\mathbb{E}(\dot{h_i}^*)$ by $-\Sigma_\alpha$. Replacing the corresponding terms with the calculations, we obtain the influence functions for $\hat{q}_{t,p}^{\mathcal{V}}$ and hence for $\widehat{\Delta}_p^{\mathcal{V}}$.

S3.3 Difference-based QTE calibration estimator

Here we give a detailed discussion of the direct extension of the difference-based calibration method in (Yang and Ding, 2020) for QTE estimation. Consider the estimator $\hat{q}_{t,p}^{\mathcal{V},\mathrm{Conf}}$ obtained by solving (2.4). By similar arguments in the proof of Theorem 1, we can arrive at an analogous result of (S3.1), that is,

$$\sup_{q\in\mathcal{Q}_{t,p}^{\mathrm{Conf}},\eta_t^{\mathrm{Conf}}\in\Theta_t^{\mathrm{Conf}}\times\Omega^{\mathrm{Conf}}}\left|\frac{1}{n}\sum_{i=1}^{n}\phi_t(O_i;q,\eta_t)-\mathbb{E}\left\{\phi_t(U;q,\eta_t^{\mathrm{Conf}})\right\}\right|=o_p(1). \tag{S3.6}$$

However, the term $\mathbb{E}\{\phi_t(U;q,\eta_t^{\mathrm{Conf},*})\}$ can be complicated in this case. In fact, by the law of iterated expectation, we have

$$\begin{aligned}
&\mathbb{E}\{\phi_1(U;q,\eta_t^{\mathrm{Conf},*})\}\\
=&\mathbb{E}\left[\mathbb{E}\left[\frac{T\left[I\{Y\leq q\}-\tilde{G}_1(q|X;\theta_1^{\mathrm{Conf},*})\right]}{\tilde{e}(X;\alpha^{\mathrm{Conf},*})}+\tilde{G}_1(q|X;\theta_1^{\mathrm{Conf},*})|X,S\right]\right]-p\\
=&\mathbb{E}\left[\frac{\mathrm{pr}(T=1|X,S)}{\tilde{e}(X;\alpha^{\mathrm{Conf},*})}\mathbb{E}\Big[\left[I\{Y(1)\leq q\}-\tilde{G}_1(q|X;\theta_1^{\mathrm{Conf},*})\right]|T=1,X,S\Big]\right.\\
&\left.+\tilde{G}_1(q|X;\theta_1^{\mathrm{Conf},*})\right]-p\\
=&\mathbb{E}\left[\frac{\mathrm{pr}(T=1|X,S)}{\tilde{e}(X;\alpha^{\mathrm{Conf},*})}\left\{F_1(q\mid X,S)-\tilde{G}_1(q|X;\theta_1^{\mathrm{Conf},*})\right\}+\tilde{G}_1(q|X;\theta_1^{\mathrm{Conf},*})\right]-p.
\end{aligned}$$

Applying Theroem 5.31 in Van der Vaart (2000), we can also derive the influence function of $\hat{q}_{1,p}^{\mathcal{V},\mathrm{Conf}}$, which is a function of the derivative of $\mathbb{E}\{\phi_1(U_i;q,\eta_1^{\mathrm{Conf},*})\}$ with respect to q at $q_{1,p}^{\mathrm{Conf},*}$. Unlike that in the unconfounded case, due to the unmeasured confounders S, $\partial\mathbb{E}\{\phi_1(U_i; q, \eta_1^{\mathrm{Conf},*})\}/\partial q|_{q=q_{1,p}^{\mathrm{Conf},*}}$ involves two unknown terms related to S: the true propensity score $\mathrm{pr}(T = 1|X, S)$ and the true conditional distribution (density) $F_1(q|X, S)$ ($f_1(q|X, S)$). Similar conclusions hold for $\hat{q}_{0,p}^{\mathcal{V},\mathrm{Conf}}$, $\hat{q}_{1,p}^{\mathcal{O},\mathrm{Conf}}$ and $\hat{q}_{0,p}^{\mathcal{O},\mathrm{Conf}}$.

As pointed out by a reviewer, we may use the validation data with fully-observed confounders to estimate these terms. However, there are several disadvantages. Firstly, if we use parametric estimation methods, we need to correctly specify both the conditional distribution and propensity score, which will destroy the double robustness of the original DR estimators. Although nonparametric methods may serve as an alternative, they may be unstable when we are estimating the conditional densities. Last but not least, the sample size of the validation data is relatively small in our setting, which will further increase the instability of the estimates.

A special case is that the unconfounded assumption also holds given only observed confounders X, that is, S are auxiliary covariates. Under this stronger assumption and further assuming that either the conditional quantile or the propensity score model is correctly specified, we can also have $\mathbb{E}\{\phi_t(U; q, \eta_t^{\mathrm{Conf},*})\} = F_t(q) - p$, which leads to simpler expressions for $\hat{C}_{\mathrm{ep}}$. However, this assumption is less realistic and is contrary to the original intention of this article.

S3.4 Proof of Lemma 1

For simplicity, we denote $\phi_t(U_i;\hat{q}_{t,p}^{\text{Conf}},\hat{\eta}_t^{\text{Conf}})$ as $\phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}},\hat{\eta}_t^{\text{Conf}})$. Recall that $\phi_{t,i}$ is defined as $\phi_t(U_i;q_{t,p}^{\text{Conf},*},\eta_t^{\text{Conf},*})$. Then we have

$$\begin{aligned}\widehat{C}_t =& \frac{1}{n}\sum_{i=1}^n \phi_{t,i}(q_{t,p}^{\text{Conf},*},\eta_t^{\text{Conf},*}) + \underbrace{\frac{1}{n}\sum_{i=1}^n \phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}},\hat{\eta}_t^{\text{Conf}}) - \frac{1}{n}\sum_{i=1}^n \phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}},\eta_t^{\text{Conf},*})}_{\text{(S3.7.1)}} \\ &+ \underbrace{\frac{1}{n}\sum_{i=1}^n \phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}},\eta_t^{\text{Conf},*}) - \frac{1}{n}\sum_{i=1}^n \phi_{t,i}(q_{t,p}^{\text{Conf},*},\eta_t^{\text{Conf},*})}_{\text{(S3.7.2)}}. \end{aligned} \tag{S3.7}$$

Now, term-by-term,

$$\text{(S3.7.1)} = \frac{1}{n}\sum_{i=1}^n \left.\frac{\partial \phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}},\eta)}{\partial \eta}\right|_{\eta=\eta_t^{\text{Conf},*}} (\hat{\eta}_t^{\text{Conf}} - \eta_t^{\text{Conf},*}) + o_p(n^{-1/2}). \tag{S3.8}$$

Further define $\varphi_{t,i}(q) = \phi_{t,i}(q,\eta_t^{\text{Conf},*}) - \phi_{t,i}(q^{\text{Conf},*})$, $a_{t,i}(q) = \mathbb{E}\varphi_{t,i}(q)$, and $b_{t,i}(q) = \varphi_{t,i}(q) - \mathbb{E}\varphi_{t,i}(q)$. Then we have

$$\text{(S3.7.2)} = \frac{1}{n}\sum_{i=1}^n \varphi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}) = \frac{1}{n}\sum_{i=1}^n a_{t,i}(\hat{q}_{t,p}^{\text{Conf}}) + \frac{1}{n}\sum_{i=1}^n b_{t,i}(\hat{q}_{t,p}^{\text{Conf}}). \tag{S3.9}$$

For the first term on the right side of (S3.9), we have

$$\frac{1}{n}\sum_{i=1}^n a_{t,i}(\hat{q}_{t,p}^{\text{Conf}}) = \frac{1}{n}\sum_{i=1}^n \left[\left.\frac{\partial \mathbb{E}\{\phi_{t,i}(q,\eta_t^{\text{Conf},*})\}}{\partial q}\right|_{q=q_{t,p}^{\text{Conf},*}} \right] (\hat{q}_{t,p}^{\text{Conf}} - q_{t,p}^{\text{Conf},*}) + o_p(n^{-1/2}). \tag{S3.10}$$

Note that

$$\hat{q}_{t,p}^{\mathrm{Conf}} - q_{t,p}^{\mathrm{Conf},*} = -\left[\left. \frac{\partial \mathbb{E}\{\phi_{t,i}(q, \eta_t^{\mathrm{Conf},*})\}}{\partial q} \right|_{q=q_{t,p}^{\mathrm{Conf},*}} \right]^{-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} \phi_{t,i} \right.$$
$$\left. + \frac{1}{N} \sum_{i=1}^{N} \left. \frac{\partial \phi_{t,i}(q_{t,p}^{\mathrm{Conf},*}, \eta)}{\partial \eta} \right|_{\eta=\eta_t^{\mathrm{Conf},*}} (\hat{\eta}_t^{\mathrm{Conf}} - \eta_t^{\mathrm{Conf},*}) \right\} + o_p(n^{-1/2}). \tag{S3.11}$$

Plugging (S3.11) back into (S3.10) gives

$$\frac{1}{n} \sum_{i=1}^{n} a_{t,i}(\hat{q}_{t,p}^{\mathrm{Conf}}) = -\left\{ \frac{1}{N} \sum_{i=1}^{N} \phi_{t,i} + \frac{1}{N} \sum_{i=1}^{N} \left. \frac{\partial \phi_{t,i}(q_{t,p}^{\mathrm{Conf},*}, \eta)}{\partial \eta} \right|_{\eta=\eta_t^{\mathrm{Conf},*}} \right.$$
$$\left. (\hat{\eta}_t^{\mathrm{Conf}} - \eta_t^{\mathrm{Conf},*}) \right\} + o_p(n^{-1/2}). \tag{S3.12}$$

Now we can show that the second term on the right side of (S3.9) is $o_p(\frac{1}{n^{1/2}})$. Using similar arguments as we use in the proof of Theorem 1, we can show that

$$\left\{ \hat{\tilde{w}}_{t,i}\{I(Y_i \leqslant q) - \tilde{G}_t(q \mid X_i; \hat{\theta}_t^{\mathrm{Conf}})\} + \tilde{G}_t(q \mid X_i; \hat{\theta}_t^{\mathrm{Conf}}) - p, |q - q_{t,p}^{\mathrm{Conf},*}| < \epsilon \right\}$$

forms a Donsker class. Note that $\hat{\tilde{w}}_{t,i}\{I(Y_i \leqslant q) - \tilde{G}_t(q \mid X_i; \hat{\theta}_t^{\mathrm{Conf}})\} + \tilde{G}_t(q \mid X_i; \hat{\theta}_t^{\mathrm{Conf}}) - p$ is L_2 continuous at $q = q_{t,p}^{\mathrm{Conf},*}$. Then we have

$$\sup_{|q - q_{t,p}^{\mathrm{Conf},*}| < \epsilon} \left| \frac{1}{n} \sum_{i=1}^{n} b_{t,i}(q) \right| = o_p(n^{-1/2}).$$

Combining the analysis of terms (S3.7.1) and (S3.7.2), we have

$$\widehat{C}_t = \left\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \eta)}{\partial \eta}\bigg|_{\eta=\eta_t^{\text{Conf},*}} - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta)}{\partial \eta}\bigg|_{\eta=\eta_t^{\text{Conf},*}}}_{\text{(S3.13.1)}} \right\}$$
$$(\hat{\eta}_t^{\text{Conf}} - \eta_t^{\text{Conf},*}) + \frac{1}{n}\sum_{i=1}^{n}\phi_{t,i} - \frac{1}{N}\sum_{i=1}^{N}\phi_{t,i} + o_p(n^{-1/2}). \tag{S3.13}$$

It suffices to show that the term (S3.13.1) is $o_p(1)$. Note that

$$\begin{aligned}\text{(S3.13.1)} =& \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \eta)}{\partial \eta}\bigg|_{\eta=\eta_t^{\text{Conf},*}} - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta)}{\partial \eta}\bigg|_{\eta=\eta_t^{\text{Conf},*}} \\ &+ \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta)}{\partial \eta}\bigg|_{\eta=\eta_t^{\text{Conf},*}} - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta)}{\partial \eta}\bigg|_{\eta=\eta_t^{\text{Conf},*}}.\end{aligned}$$

The terms in the first line are $o_p(1)$ under Condition S3.2(e)(ii) and the terms in the second line are also $o_p(1)$ due to the weak law of large numbers. This completes the proof.

S3.5 Proof of Proposition 1

The asymptotic variance of $\hat{\Delta}_p^{\nu}$ is a direct application of Theorem 1. According to Lemma 1, we have

$$n^{1/2}\widehat{C}_t = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\phi_{t,i} - \frac{\nu^{1/2}}{N^{1/2}}\sum_{i=1}^{n}\phi_{t,i} + o_p(1).$$

Therefore, under Assumption 5, we can derive the asymptotic matrix for $\widehat{C}_t$ as

$$\Sigma_t = \mathrm{Var}(\phi_{t,i}) + \nu\,\mathrm{Var}(\phi_{t,i}) - 2\nu\,\mathrm{Var}(\phi_{t,i}) = (1-\nu)\,\mathrm{Var}(\phi_{t,i}),$$

for $t = 0, 1$. Similarly, we can write $\widehat{C}_t$ as

$$n^{1/2}\widehat{C}_t = \frac{1}{n^{1/2}}(1-\nu)\sum_{i=1}^{n}\phi_{t,i} - \frac{1}{n^{1/2}}\nu\sum_{i=n+1}^{N}\phi_{t,i} + o_p(1).$$

Hence, we have the asymptotic covariance as

$$\varrho = (1-\nu)\,\mathrm{cov}(\psi_{1,i} - \psi_{0,i}, (\phi_{1,i}^\top, \phi_{0,i}^\top)^\top)$$

for $t = 0, 1$. This completes the proof of Proposition 1.

S3.6 Proof of Theorem 2

Define $\tilde{\Delta}_p = \widehat{\Delta}_p^{\mathcal{V}} - \varrho^\top\Sigma_{\mathrm{ep}}^{-1}\widehat{C}$. By the delta method, we have

$$n^{1/2}(\tilde{\Delta}_p - \Delta_p) \xrightarrow{D} \mathcal{N}(0, \sigma^2).$$

Since $\hat{\varrho} = \varrho + o_p(1)$ and $\widehat{\Sigma}_{\mathrm{ep}} = \Sigma_{\mathrm{ep}} + o_p(1)$, the theorem follows after application of Slutsky's lemma.

S3.7 Proof of Theorem 3

The proof uses similar techniques that were used for establishing the consistency of variance estimators in Firpo (2007). We first establish the consistency of $\widehat{\Sigma}_{\mathrm{ep}}$. Denote $e(X_i, S_i)$ as e_i, $e(X_i, S_i; \hat{\alpha})$ as $\hat{\hat{e}}_i$ and $\tilde{e}(X_i; \hat{\alpha}^{\mathrm{Conf}})$ as

$\hat{\tilde{e}}_i$. Define

$$A_i = \frac{T_i}{e_i^*} I(Y_i \le q_{1,p}^*) - p, \quad B_i = G_1(q_{1,p}^* | X_i, S_i; \theta_1^*)(1 - \frac{T_i}{e_i^*}),$$

$$\tilde{A}_i = \frac{T_i}{\hat{e}_i} I(Y_i \le q_{1,p}^*) - p, \quad \tilde{B}_i = G_1(q_{1,p}^* | X_i, S_i; \theta_1^*)(1 - \frac{T_i}{\hat{e}_i}),$$

$$\widehat{A}_i = \frac{T_i}{\hat{e}_i} I(Y_i \le \hat{q}_{1,p}) - p, \quad \widehat{B}_i = G_1(\hat{q}_{1,p} | X_i, S_i; \theta_1^*)(1 - \frac{T_i}{\hat{e}_i})$$

$$D_i = \frac{T_i}{\tilde{e}_i^*} I(Y_i \le q_{1,p}^{\mathrm{Conf},*}) - p, \quad F_i = \tilde{G}_1(q_{1,p}^{\mathrm{Conf},*} | X_i; \theta_1^*)(1 - \frac{T_i}{\tilde{e}_i^*}),$$

$$\tilde{D}_i = \frac{T_i}{\hat{\tilde{e}}_i} I(Y_i \le q_{1,p}^{\mathrm{Conf},*}) - p, \quad \tilde{F}_i = \tilde{G}_1(q_{1,p}^{\mathrm{Conf},*} | X_i; \hat{\theta}_1^{\mathrm{Conf}})(1 - \frac{T_i}{\hat{\tilde{e}}_i}),$$

$$\widehat{D}_i = \frac{T_i}{\hat{\tilde{e}}_i} I(Y_i \le \hat{q}_{1,p}^{\mathrm{Conf}}) - p, \quad \widehat{F}_i = \tilde{G}_1(\hat{q}_{1,p}^{\mathrm{Conf}} | X_i; \hat{\theta}_1^{\mathrm{Conf}})(1 - \frac{T_i}{\hat{\tilde{e}}_i}).$$

Then we have $\hat{\phi}_{1,i} = \widehat{D}_i + \widehat{F}_i$. We now show that $\left| \frac{1}{N} \sum_{i=1}^N \hat{\phi}_{t,i}^2 - \frac{1}{N} \sum_{i=1}^N \phi_{t,i}^2 \right| = o_p(1)$. Note that

$$\left| \frac{1}{N} \sum_{i=1}^N \hat{\phi}_{t,i}^2 - \frac{1}{N} \sum_{i=1}^N \phi_{t,i}^2 \right| \le \underbrace{\left| \frac{1}{N} \sum_{i=1}^N (\widehat{D}_i^2 - \tilde{D}_i^2) \right|}_{\text{(S3.14.1)}} + \underbrace{\left| \frac{1}{N} \sum_{i=1}^N (\widehat{F}_i^2 - \tilde{F}_i^2) \right|}_{\text{(S3.14.2)}}$$
$$+ \underbrace{\left| \frac{2}{N} \sum_{i=1}^N (\widehat{D}_i \widehat{F}_i - \widehat{D}_i \tilde{F}_i) \right|}_{\text{(S3.14.3)}}. \tag{S3.14}$$

Now, term by term, we have

$$
\begin{aligned}
\text{(S3.14.1)} \leq & \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{\hat{\tilde{e}}_i^{\,2}} \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \right] \right| \\
& + \left| \frac{2}{N} \sum_{i=1}^{N} \frac{T_i}{\hat{\tilde{e}}_i} \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \right] \right| \\
\leq & \frac{\sup_{X \in \mathbb{X}} \tilde{e}_i}{\inf_{X \in \mathbb{X}} \hat{\tilde{e}}_i^{\,2}} \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{\tilde{e}_i} \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \right] \right| \\
& + \frac{\sup_{X \in \mathbb{X}} \tilde{e}_i}{\inf_{X \in \mathbb{X}} \hat{\tilde{e}}_i} \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{\tilde{e}_i} \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \right] \right| \\
\leq & C \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{\tilde{e}_i} \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \right] \right| \\
\leq & C \left| \hat{q}_{1,p}^{\text{Conf}} - q_{1,p}^{\text{Conf},*} \right| O_p(\mathbb{E} \left[(\frac{T_i}{\tilde{e}_i})^2 \right]^{1/2}) = o_p(1)
\end{aligned}
$$

where the third inequality follows from Condition S3.2(d) and C is a scalar constant. Similarly, for (S3.14.2), we have

$$
\begin{aligned}
\text{(S3.14.2)} = & \left| \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{T_i}{\hat{\tilde{e}}_i} \right)^2 \tilde{G}_1^2(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) - \tilde{G}_1^2(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right| \\
\leq & \frac{C}{N} \sum_{i=1}^{N} \left| \tilde{G}_1(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) + \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right| \left| \tilde{G}_1(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) \right. \\
& \left. - \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right| \leq C \left| \hat{q}_{1,p}^{\text{Conf}} - q_{1,p}^{\text{Conf},*} \right| = o_p(1).
\end{aligned}
$$

Finally, for (S3.14.3), we have

$$
\begin{aligned}
\text{(S3.14.3)} \leq & \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{\hat{\hat{e}}_i} \left(1 - \frac{T_i}{\hat{\hat{e}}_i}\right) \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) \tilde{G}_1(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) \right.\right. \\
& \left.\left. - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right] \right| \\
& + \left| \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{T_i}{\hat{\hat{e}}_i}\right) \left[\tilde{G}_1(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) - \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right] \right| \\
\leq & \frac{C_1}{N} \sum_{i=1}^{N} \left| \tilde{G}_1(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) - \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right| \\
& + \frac{C_2}{N} \sum_{i=1}^{N} \left| \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \left[I(Y_i \leq \hat{q}_{1,p}^{\text{Conf}}) - I(Y_i \leq q_{1,p}^{\text{Conf},*}) \right] \right| \\
\leq & \frac{C_1}{N} \sum_{i=1}^{N} \left| \tilde{G}_1(\hat{q}_{1,p}^{\text{Conf}} | X_i; \hat{\theta}_1^{\text{Conf}}) - \tilde{G}_1(q_{1,p}^{\text{Conf},*} | X_i; \hat{\theta}_1^{\text{Conf}}) \right| + C_2 \left| \hat{q}_{1,p}^{\text{Conf}} - q_{1,p}^{\text{Conf},*} \right| \\
= & o_p(1).
\end{aligned}
$$

Plugging the terms in to (S3.14), together with the fact that $\frac{1}{N}\sum_{i=1}^{N} \phi_{t,i}^2 = \Sigma_1 + o_p(1)$, we have derived the consistency for $\widehat{\Sigma}_1$. Since Σ_{01} and Σ_0 have a similar structure with Σ_1, we only prove the consistency of $\widehat{\Sigma}_1$ and similar procedure can be carried out for analysing $\widehat{\Sigma}_{01}$ and $\widehat{\Sigma}_0$.

Next we establish the consistency for $\hat{\sigma}_{\mathcal{V},1}^2$. The proof for $\hat{\sigma}_{\mathcal{V},0}^2$ is the same. Furthuer define $\Psi_{1,i} = A_i + B_i$, $\hat{\Psi}_{1,i} = \widehat{A}_i + \widehat{B}_i$ and $\tilde{\Psi}_{1,i} = \tilde{A}_i + \tilde{B}_i$. We denote $\psi_{t,i} - \psi_{1,i}^{\text{DR},*}$ as $\xi_{1,i}$, which is the additional term in the influence function due to nuisance parameters. Let $\hat{\xi}_{1,i}$ be its estimates by replacing $\mathbb{E}(\cdot)$ with the empirical measure and unknown parameters with their corresponding estimators. Note that $\hat{\sigma}_{\mathcal{V},1}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\Psi}_{1,i}}{\hat{f}_1(\hat{q}_{1,p})} + \hat{\xi}_{1,i}\right)^2$.

We first show that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\Psi}_{t,i}^2}{\hat{f}_1^2(\hat{q}_{1,p})}-\frac{1}{n}\sum_{i=1}^{n}\frac{\Psi_{1,i}^2}{f_1^2(q_{1,p})}\right|=o_p(1). \tag{S3.15}$$

Indeed, we have

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\Psi}_{1,i}^2}{\hat{f}_1^2(\hat{q}_{1,p})}-\frac{1}{n}\sum_{i=1}^{n}\frac{\Psi_{1,i}^2}{f_1^2(q_{1,p})}\right| \\
&\le \left|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\Psi}_{1,i}^2}{\hat{f}_1^2(\hat{q}_{1,p})}-\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\Psi}_{t,i}^2}{f_1^2(q_{1,p})}\right|+\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\Psi}_{1,i}^2-\Psi_{1,i}^2}{f_1^2(q_{1,p})}\right| \\
&\le \left|\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\Psi}_{1,i}^2-\Psi_{t,i}^2}{\hat{f}_1^2(\hat{q}_{1,p})}\right|+\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{\Psi}_{1,i}^2\left\{\frac{1}{\hat{f}_1^2(\hat{q}_{1,p})}-\frac{1}{f_1^2(q_{1,p})}\right\}\right|+\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\Psi}_{1,i}^2-\Psi_{t,i}^2}{f_1^2(q_{1,p})}\right|.
\end{aligned} \tag{S3.16}$$

The last term in the last line is obviously $o_p(1)$. Since the density estimator is consistent, we have $\left|\hat{f}_1(y)-f_1(y)\right|=o_p(1)$. The second term is bounded by

$$\begin{aligned}
&\frac{\left|\hat{f}_1(\hat{q}_{1,p})-f_1(q_{1,p})\right|\left|\hat{f}_1(\hat{q}_{1,p})+f_1(q_{1,p})\right|}{\hat{f}_1^2(\hat{q}_{1,p})f_1^2(q_{1,p})}\left|\frac{1}{n}\sum_{i=1}n\tilde{\Psi}_{1,i}^2\right| \\
&\le \frac{\left|\hat{f}_1(\hat{q}_{1,p})-f_1(\hat{q}_{1,p})\right|+|f_1(\hat{q}_{1,p})-f_1(q_{1,p})|}{\hat{f}_1^2(\hat{q}_{1,p})f_1^2(q_{1,p})}O_p(1)=o_p(1),
\end{aligned}$$

Now we turn to analyze the first term in the last line of (S3.16). Indeed, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\Psi}_{1,i}^{2}-\Psi_{t,i}^{2}}{\hat{f}_{1}^{2}(\hat{q}_{1,p})}\right| \leq O_p(1)\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{A}_i^2-\tilde{A}_i^2\right|}_{(S3.17.1)}+\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{B}_i^2-\tilde{B}_i^2\right|}_{(S3.17.2)}\right.$$
$$\left. + 2\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{A}_i\widehat{B}_i-\tilde{A}_i\tilde{B}_i\right|}_{(S3.17.3)}\right]. \tag{S3.17}$$

Using similar techniques as we used for analysing (S3.13.1)-(S3.13.3) together with Condition S3.1(d), we can show that $\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\Psi}_{1,i}^{2}-\Psi_{t,i}^{2}}{\hat{f}_{1}^{2}(\hat{q}_{1,p})}\right|$ is also $o_p(1)$. Plugging the terms in to (S3.16), we have proved (S3.15).

By similar arguments, we can also show that $\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{1,i}^{2}-\frac{1}{n}\sum_{i=1}^{n}\xi_{1,i}^{2}\right| = o_p(1)$ and $\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\Psi}_{1,i}}{\hat{f}_1(\hat{q}_{1,p})}\hat{\xi}_{1,i}-\frac{1}{n}\sum_{i=1}^{n}\frac{\Psi_{1,i}}{f_1(q_{1,p})}\xi_{1,i}\right| = o_p(1)$. Combining these results together, we have proved the consistency of $\sigma_{\mathcal{V},1}^2$. The proof for $\sigma_{\mathcal{V},0}^2$ is the same. This completes the proof of Theorem 3.

S3.8 Proof of Theorem S2.1

The proof is similar to that of Theorem 1. We first establish the consistency of $\hat{q}_{t,p}^{\mathcal{V}}$. Note that is can be seen as a Z-estimator of the following Z-estimating functions:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\hat{\pi}_i}\Psi_t(O_i;q,\hat{\eta}_t)=0.$$

Under Assumption 3 or 4 and S2.1, using similar arguments as in the proof of Theorem 1, it is easy to verify that the Z-estimating function $\frac{R_i}{\pi_i}\Psi_t(O_i; q, \hat{\eta}_t)$ in expectation equals $F_t(q) - p$ for $t = 0, 1$. Similar to (S3.1), we claim that

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}_{t,p},\eta_t\in\Theta_t\times\Omega}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\hat{\pi}_i}\Psi_t(O_i; q, \eta_t) - \mathbb{E}\left\{\frac{R_i}{\pi_i}\Psi_t(O; q, \eta_t)\right\}\right| \leq \\
&\sup_{q\in\mathcal{Q}_{t,p},\eta_t\in\Theta_t\times\Omega}\left|\frac{1}{N}\sum_{i=1}^{N}R_i(\frac{1}{\hat{\pi}_i} - \frac{1}{\pi_i})\Psi_t(O_i; q, \eta_t)\right| \\
&+ \sup_{q\in\mathcal{Q}_{t,p},\eta_t\in\Theta_t\times\Omega}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\Psi_t(O_i; q, \eta_t) - \mathbb{E}\left\{\frac{R_i}{\pi_i}\Psi_t(O; q, \eta_t)\right\}\right| = o_p(1).
\end{aligned} \tag{S3.18}$$

In fact, the first term on the right side of (S3.18)is bounded above by

$$\sup_{q\in\mathcal{Q}_{t,p},\eta_t\in\Theta_t\times\Omega}\left|\frac{1}{N}\sum_{i=1}^{N}R_i\frac{\Psi_t(O_i; q, \eta_t)}{\pi_i^2}\left.\frac{\partial\pi_i(\beta)}{\partial\beta^\top}\right|_{\beta=\beta^*}\right|\left|\hat{\beta} - \beta^*\right|,$$

which is $o_p(1)$ under Condition S3.3. According the proof of Theorem 1, it suffices to prove that $\tilde{\mathcal{M}} = \{\frac{R}{\pi(Y,T,X)}\Psi_1(O; q, \eta_1); q \in \mathcal{Q}_{t,p}, \eta_1 = (\theta_1, \alpha) \in \Theta_1 \times \Omega\}$ is a Donsker class. Since $\frac{R}{\pi(Y,T,X)}$ is bounded under Condition S3.3, $\tilde{\mathcal{M}}$ is Donsker from the preservation of Donsker classes (corollary 9.32 of Kosorok (2008)).

S3.9 Proof of Lemma S2.1

Using the same notations as that in the proof of Lemma 1, then we have

$$\begin{aligned}
\widehat{C}_t = & \underbrace{\frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \hat{\eta}_t^{\text{Conf}}) - \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \eta_t^{\text{Conf},*})}_{\text{(S3.19.1)}} \\
& + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \eta_t^{\text{Conf},*}) - \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta_t^{\text{Conf},*})}_{\text{(S3.19.2)}} \\
& + \underbrace{\frac{1}{N}\sum_{i=1}^{N}R_i(\frac{1}{\hat{\pi}_i} - \frac{1}{\pi_i})\phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \hat{\eta}_t^{\text{Conf}})}_{\text{(S3.19.3)}} + \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta_t^{\text{Conf},*})
\end{aligned} \tag{S3.19}$$

Now, term-by-term,

$$\text{(S3.19.1)} = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\left.\frac{\partial\phi_{t,i}(\hat{q}_{t,p}^{\text{Conf}}, \eta)}{\partial\eta}\right|_{\eta=\eta_t^{\text{Conf},*}}(\hat{\eta}_t^{\text{Conf}} - \eta_t^{\text{Conf},*}) + o_p(\frac{1}{n^{1/2}}). \tag{S3.20}$$

Using similar techniques as we use for analyzing (S3.13.2) in the proof of Lemma 1, we have

$$\text{(S3.19.2)} = -\frac{1}{N}\sum_{i=1}^{N}\left[\phi_{t,i} + \left.\frac{\partial\phi_{t,i}(q_{t,p}^{\text{Conf},*}, \eta)}{\partial\eta}\right|_{\eta=\eta_t^{\text{Conf},*}}(\hat{\eta}_t^{\text{Conf}} - \eta_t^{\text{Conf},*})\right] + o_p(\frac{1}{n^{1/2}}). \tag{S3.21}$$

Under Condition S3.3 and the consistency of $(\hat{q}_{t,p}^{\text{Conf}}, \hat{\eta}_t^{\text{Conf}}, \hat{\beta})$, using a similar argument in the analysis of (S3.18), we find that (S3.19.3) $= o_p(\frac{1}{n^{1/2}})$ by noting that $\frac{R_i}{\pi_i}\Psi_t(O; q, \eta_t)$ is equal to 0 in expectation.

Combining the analysis of terms (S3.19.1), (S3.19.2), and (S3.19.3), we have

$$\widehat{C}_t = \left\{ \underbrace{\frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\left.\frac{\partial \phi_{t,i}(\hat{q}_{t,p}^{\mathrm{Conf}},\eta)}{\partial \eta}\right|_{\eta=\eta_t^{\mathrm{Conf},*}} - \frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial \phi_{t,i}(q_{t,p}^{\mathrm{Conf},*},\eta)}{\partial \eta}\right|_{\eta=\eta_t^{\mathrm{Conf},*}}}_{\text{(S3.22.1)}} \right\}$$

$$(\hat{\eta}_t^{\mathrm{Conf}} - \eta_t^{\mathrm{Conf},*}) + \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i}(q_{t,p}^{\mathrm{Conf},*},\eta_t^{\mathrm{Conf},*}) - \frac{1}{N}\sum_{i=1}^{N}\phi_{t,i} + o_p(\frac{1}{n^{1/2}}). \tag{S3.22}$$

It is straightforward by noting that the first term in (S3.22.1) is asymptotically equivalent to the first term in (S3.13.1) under Assumption S2.1. Then (S3.22.1) is $o_p(1)$. Together with the results obtained above, we complete the proof.

S3.10 Proof of Proposition S2.1 and S2.2

According to Theorem S2.1 and Lemma S2.1, we have the following asymptotic linear representations,

$$n^{1/2}(\hat{q}_{t,p}^{\mathcal{V}} - q_{t,p}) = (\frac{\nu}{N})^{1/2}\sum_{i=1}^{N}\frac{R_i}{\pi_i}\phi_{t,i} + o_p(1),$$

$$n^{1/2}\widehat{C}_t = (\frac{\nu}{N})^{1/2}\sum_{i=1}^{N}\left(\frac{R_i}{\pi_i} - 1\right)\phi_{t,i} + o_p(1).$$

The results follow directly from the multivariate central limit theorem. To be detailed, for the asymptotic covariance, we have the following calculation,

$$\sigma^2_{\mathcal{V},t} = \mathrm{Var}\left\{ (\frac{\nu}{N})^{1/2} \sum_{i=1}^{N} \frac{R_i}{\pi_i} \phi_{t,i} \right\} = \nu\, \mathrm{var}(\frac{R_i}{\pi_i} \psi_{t,i}),$$

$$\begin{aligned}\rho_t &= \mathrm{Cov}\left\{ (\frac{\nu}{N})^{1/2} \sum_{i=1}^{N} \frac{R_i}{\pi_i} \phi_{t,i}, (\frac{\nu}{N})^{1/2} \sum_{i=1}^{N} \left(\frac{R_i}{\pi_i} - 1 \right) \phi_{t,i} \right\} \\ &= \nu\, \mathrm{Cov}[\frac{R_i}{\pi_i} \psi_{t,i}, (\frac{R_i}{\pi_i} - 1)\phi_{t,i}],\end{aligned}$$

$$\Sigma_t = \mathrm{Var}\left\{ (\frac{\nu}{N})^{1/2} \sum_{i=1}^{N} \left(\frac{R_i}{\pi_i} - 1 \right) \phi_{t,i} \right\} = \nu\, \mathrm{var}[(\frac{R_i}{\pi_i} - 1)\phi_{t,i}],$$

for $t = 0, 1$. This completes the proof of Proposition S2.1.

The proof of Proposition S2.2 is the same and thus omitted here.

S4 Additional Simulation Results

S4.1 Results for estimating $\Delta_{0.75}$ under Setting 3

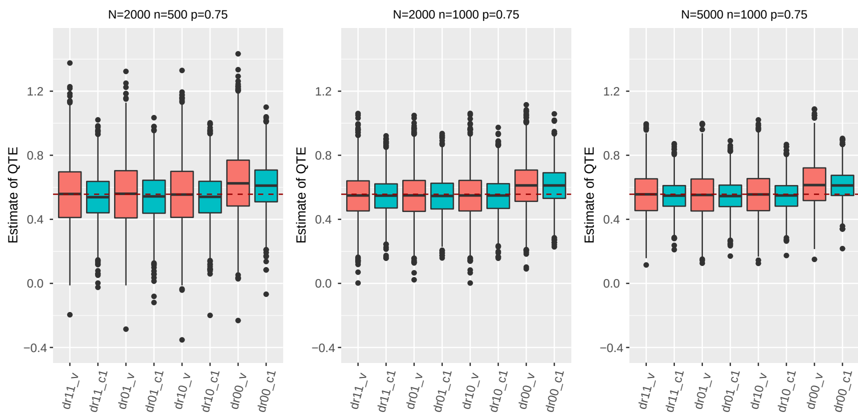

Figure S4.1: Point estimates of the DR estimators for $\Delta_{0.75}$ under Setting 3.

Table S4.1: Simulation Results for estimating $\Delta_{0.75}$ under Setting 3.

$\Delta_{0.75}$	$N = 2000, n = 500$				$N = 2000, n = 1000$				$N = 5000, n = 1000$			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
dr11_v	0.0000	0.0454	0.2131	0.9490	0.0076	0.0209	0.1487	0.9560	0.0001	0.0205	0.1490	0.9575
dr11_c1	0.0188	0.0236	0.1497	0.9425	0.0097	0.0140	0.1205	0.9515	0.0090	0.0093	0.1002	0.9615
dr01_v	0.0010	0.0465	0.2160	0.9465	0.0073	0.0217	0.1507	0.9510	0.0010	0.0211	0.1510	0.9590
dr01_c1	0.0178	0.0246	0.1528	0.9415	0.0094	0.0146	0.1226	0.9530	0.0102	0.0096	0.1024	0.9625
dr10_v	0.0005	0.0450	0.2135	0.9495	0.0074	0.0210	0.1489	0.9575	0.0001	0.0207	0.1492	0.9565
dr10_c1	0.0181	0.0235	0.1499	0.9445	0.0094	0.0141	0.1207	0.9480	0.0090	0.0093	0.1003	0.9620
dr00_v	0.0678	0.0508	0.2161	0.9380	0.0566	0.0251	0.1508	0.9375	0.0634	0.0252	0.1511	0.9375
dr00_c1	0.0505	0.0249	0.1465	0.9225	0.0550	0.0171	0.1199	0.9215	0.0550	0.0118	0.0971	0.9140

S4.2 Additional settings under MCAR assumption

We consider more settings under the MCAR assumption. In settings 1 and 2, we consider two simple cases with only two confounders, one commonly observed for the entire data and the other only observed for a subset of units. The two settings share the same distribution of the pre-treatment variables and the treatment assignment mechanism. The potential outcomes have the same noise levels but different signals in setting 1 and the same signals but different noise levels in setting 2. Setting 3 is what we described in the main text. In setting 4, 6 confounders are considered, 4 of which are fully observed and the rest 2 of which are observed only for a subset of units.

Setting 1:

$$W_{ki} \sim \text{Unif}(1-\sqrt{3}, 1+\sqrt{3}), k=1,2; X_i = W_{1i}, S_i = \sin((W_{1i} + W_{2i}/2));$$

$$\text{logit}\{P(T_i = 1 | X_i, S_i)\} = 0.25X_i - 0.25S_i;$$

$$Y_i(1) = 0.5X_i + 1.5S_i + \epsilon_i(1); Y_i(0) = 0.5X_i + 0.5S_i + \epsilon_i(0),$$

where $\epsilon_i(1) \sim \mathcal{N}(0,1)$, $\epsilon_i(0) \sim \mathcal{N}(0,1)$ and they are independent.

Setting 2:

$$W_{ki} \sim \text{Unif}(1-\sqrt{3}, 1+\sqrt{3}), k=1,2,3,4; X_i = W_{1i}, S_i = \sin((W_{1i} + W_{2i}/2));$$

$$\text{logit}\{P(T_i = 1 | X_i, S_i)\} = 0.25X_i - 0.25S_i;$$

$$Y_i(1) = X_i - S_i + \epsilon_i(1); Y_i(0) = X_i - S_i + \epsilon_i(0),$$

where $\epsilon_i(1) \sim \mathcal{N}(0,2^2)$, $\epsilon_i(0) \sim \mathcal{N}(0,1)$ and they are independent.

Setting 3:

$$W_{ki} \sim \text{Unif}(1-\sqrt{3}, 1+\sqrt{3}), k = 1,2,3; X_{1i} = W_{1i}, S_{1i} = \exp(W_{2i}/2),$$

$$S_{2i} = \log(W_{3i}+1), S_{3i} = \sin(3*(W_{1i}));$$

$$\text{logit}\{P(T_i = 1|X_i, S_i)\} = 0.25X_{1i} - 0.25S_{1i} + 0.25S_{2i} - 0.25S_{3i};$$

$$Y_i(1) = 0.5X_{1i} - 0.5S_{1i} + 0.5S_{2i} - 0.5S_{3i} + \epsilon_i(1);$$

$$Y_i(0) = 0.5X_{1i} - 0.5S_{1i} + 0.5S_{2i} - 0.5S_{3i} + \epsilon_i(0),$$

where $\epsilon_i(1) \sim \mathcal{N}(0, 2^2)$, $\epsilon_i(0) \sim \mathcal{N}(0, 1)$ and they are independent.

Setting 4:

$$W_{ki} \sim \text{Unif}(1-\sqrt{3}, 1+\sqrt{3}), k = 1, \cdots, 6; X_{1i} = \exp(W_{1i}/2), X_{2i} = \log(W_{2i}+1);$$

$$X_{3i} = \sin(W_{1i} + W_{3i}); X_{4i} = \exp(W_{4i}/3), S_{1i} = \cos(W_{5i} - W_{6i}), S_{2i} = \log(W_{6i}+1);$$

$$\text{logit}\{P(T_i = 1|X_i, S_i)\} = 0.25X_{1i} - 0.25X_{2i} + 0.25X_{3i} - 0.25X_{4i} + 0.25S_{1i} - 0.25S_{2i};$$

$$Y_i(1) = X_{1i} - X_{2i} + X_{3i} - X_{4i} + S_{1i} - S_{2i} + \epsilon_i(1);$$

$$Y_i(0) = X_{1i} - X_{2i} + X_{3i} - X_{4i} + S_{1i} - S_{2i} + \epsilon_i(0),$$

where $\epsilon_i(1) \sim \mathcal{N}(0, 2^2)$, $\epsilon_i(0) \sim \mathcal{N}(0, 1)$ and they are independent.

We specify a normal linear model both for the conditional distribution $F_t(y \mid X, S)$ and $F_t(y \mid X)$ and a log-linear model both for the propensity score $e(X, S)$ and $e(X)$. We use the random forest for the estimation of the propensity score when evaluating $f_t(q_{t,p})$. For the misspecified models, we take the transformation of the original covariates as $\tilde{X}_i = X_i^2$ and $\tilde{S}_i =$

$\exp(S_i)$, and then specified a normal linear model for $F_t(y|\tilde{X}, \tilde{S})$ and a log-linear model for $e(\tilde{X}, \tilde{S})$.

We first generated fully-observed data with a sample size $N = m + n$ from the underlying true model and then randomly select n units from the full data as a validation dataset. The covariates S of the rest m units are treated as missing. Since there is no explicit form for the true values of quantiles, we use a Monte Carlo of simulation of 10000 replications with a sample size of 1000000 to obtain a benchmark.

Table S4.2-S4.3 display the simulation results of FQTEs using a single calibrating quantile for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ in Setting 1,2,4.

As we can see from Table S4.2-S4.3, our fused QTE estimators enjoy a gain of efficiency compared to the initial estimators based on the validation dataset only. The coverage rates of our 95% confidence interval are all around 95% except "dr00", which suggests the consistency of our variance estimators. The coverage rates for "dr00" are lower than 95% since both the outcome and propensity score models are misspecified.

Table S4.2: Simulation results for estimating $\Delta_{0.5}$ in Setting 1,2 and 4.

$\Delta_{0.5}$	$N = 2000, n = 500$				$N = 2000, n = 1000$				$N = 5000, n = 1000$			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
						Setting 1						
dr11_v	0.0034	0.0217	0.1532	0.9595	0.0021	0.0112	0.1072	0.9450	0.0028	0.0106	0.1070	0.9605
dr11_c1	0.0048	0.0129	0.1171	0.9575	0.0047	0.0080	0.0907	0.9550	0.0072	0.0059	0.0792	0.9545
dr01_v	0.0032	0.0233	0.1565	0.9550	0.0027	0.0116	0.1096	0.9450	0.0029	0.0111	0.1094	0.9605
dr01_c1	0.0048	0.0143	0.1221	0.9580	0.0052	0.0084	0.0938	0.9545	0.0072	0.0065	0.0829	0.9530
dr10_v	0.0031	0.0215	0.1523	0.9605	0.0022	0.0111	0.1066	0.9450	0.0025	0.0105	0.1064	0.9610
dr10_c1	0.0050	0.0127	0.1165	0.9565	0.0048	0.0080	0.0902	0.9525	0.0069	0.0059	0.0788	0.9550
dr00_v	0.0414	0.0232	0.1560	0.9565	0.0358	0.0117	0.1092	0.9460	0.0356	0.0112	0.1090	0.9570
dr00_c1	0.0347	0.0134	0.1184	0.9560	0.0336	0.0081	0.0920	0.9520	0.0319	0.0060	0.0799	0.9515
						Setting 2						
dr11_v	0.0029	0.0407	0.2094	0.9535	0.0023	0.0194	0.1462	0.9610	0.0026	0.0190	0.1460	0.9655
dr11_c1	0.0122	0.0195	0.1431	0.9590	0.0064	0.0124	0.1165	0.9590	0.0038	0.0084	0.0947	0.9575
dr01_v	0.0013	0.0422	0.2127	0.9565	0.0016	0.0198	0.1485	0.9585	0.0035	0.0199	0.1483	0.9685
dr01_c1	0.0103	0.0210	0.1474	0.9545	0.0057	0.0130	0.1191	0.9560	0.0027	0.0092	0.0977	0.9525
dr10_v	0.0029	0.0409	0.2096	0.9535	0.0025	0.0195	0.1463	0.9580	0.0021	0.0190	0.1461	0.9650
dr10_c1	0.0119	0.0197	0.1434	0.9580	0.0065	0.0125	0.1167	0.9555	0.0041	0.0085	0.0949	0.9580
dr00_v	0.0066	0.0439	0.2128	0.9545	0.0071	0.0211	0.1485	0.9520	0.0115	0.0212	0.1483	0.9550
dr00_c1	0.0023	0.0213	0.1436	0.9460	0.0030	0.0137	0.1174	0.9495	0.0054	0.0096	0.0945	0.9330
						Setting 4						
dr11_v	0.0120	0.0663	0.2658	0.9540	0.0003	0.0317	0.1861	0.9550	0.0030	0.0321	0.1857	0.9565
dr11_c1	0.0105	0.0408	0.2077	0.9530	0.0018	0.0242	0.1583	0.9555	0.0030	0.0188	0.1395	0.9515
dr01_v	0.0126	0.0789	0.2924	0.9530	0.0051	0.0374	0.2044	0.9600	0.0050	0.0384	0.2040	0.9610
dr01_c1	0.0103	0.0520	0.2342	0.9515	0.0036	0.0295	0.1764	0.9530	0.0005	0.0238	0.1578	0.9495
dr10_v	0.0103	0.0651	0.2639	0.9525	0.0002	0.0313	0.1847	0.9560	0.0019	0.0320	0.1844	0.9585
dr10_c1	0.0083	0.0406	0.2066	0.9550	0.0016	0.0237	0.1574	0.9610	0.0039	0.0189	0.1390	0.9490
dr00_v	0.0164	0.0793	0.2941	0.9595	0.0317	0.0395	0.2057	0.9545	0.0297	0.0399	0.2053	0.9570
dr00_c1	0.0203	0.0499	0.2295	0.9510	0.0306	0.0297	0.1747	0.9515	0.0248	0.0231	0.1540	0.9495

Table S4.3: Simulation results for estimating $\Delta_{0.75}$ in Setting 1,2 and 4.

$\Delta_{0.75}$	$N = 2000, n = 500$				$N = 2000, n = 1000$				$N = 5000, n = 1000$			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
						Setting 1						
dr11_v	0.0096	0.0218	0.1573	0.9600	0.0069	0.0106	0.1097	0.9610	0.0038	0.0114	0.1099	0.9570
dr11_c1	0.0187	0.0108	0.1110	0.9640	0.0105	0.0072	0.0888	0.9675	0.0112	0.0051	0.0739	0.9555
dr01_v	0.0107	0.0224	0.1599	0.9625	0.0063	0.0111	0.1115	0.9620	0.0040	0.0118	0.1117	0.9545
dr01_c1	0.0200	0.0120	0.1165	0.9605	0.0100	0.0078	0.0918	0.9630	0.0114	0.0058	0.0780	0.9550
dr10_v	0.0097	0.0218	0.1576	0.9615	0.0072	0.0107	0.1099	0.9605	0.0031	0.0114	0.1101	0.9560
dr10_c1	0.0189	0.0109	0.1115	0.9625	0.0109	0.0072	0.0891	0.9635	0.0106	0.0052	0.0742	0.9575
dr00_v	0.0334	0.0225	0.1598	0.9630	0.0367	0.0110	0.1114	0.9625	0.0393	0.0118	0.1116	0.9550
dr00_c1	0.0256	0.0113	0.1135	0.9610	0.0335	0.0075	0.0904	0.9630	0.0325	0.0054	0.0756	0.9540
						Setting 2						
dr11_v	0.0018	0.0495	0.2268	0.9545	0.0019	0.0258	0.1594	0.9510	0.0037	0.0246	0.1596	0.9575
dr11_c1	0.0060	0.0264	0.1610	0.9455	0.0014	0.0170	0.1294	0.9480	0.0060	0.0116	0.1080	0.9475
dr01_v	0.0022	0.0520	0.2302	0.9505	0.0013	0.0266	0.1618	0.9510	0.0046	0.0254	0.1621	0.9600
dr01_c1	0.0065	0.0281	0.1654	0.9430	0.0008	0.0179	0.1322	0.9495	0.0069	0.0125	0.1112	0.9485
dr10_v	0.0016	0.0494	0.2274	0.9540	0.0023	0.0260	0.1598	0.9515	0.0040	0.0247	0.1601	0.9580
dr10_c1	0.0060	0.0264	0.1615	0.9460	0.0018	0.0171	0.1298	0.9455	0.0063	0.0117	0.1084	0.9440
dr00_v	0.0058	0.0522	0.2313	0.9535	0.0053	0.0278	0.1626	0.9520	0.0022	0.0267	0.1629	0.9555
dr00_c1	0.0016	0.0264	0.1609	0.9455	0.0058	0.0183	0.1304	0.9395	0.0000	0.0122	0.1072	0.9345
						Setting 4						
dr11_v	0.0081	0.0786	0.2847	0.9490	0.0106	0.0409	0.1999	0.9485	0.0075	0.0408	0.1996	0.9475
dr11_c1	0.0151	0.0541	0.2299	0.9420	0.0117	0.0315	0.1735	0.9480	0.0113	0.0257	0.1559	0.9450
dr01_v	0.0110	0.0987	0.3166	0.9450	0.0052	0.0498	0.2226	0.9505	0.0041	0.0503	0.2220	0.9485
dr01_c1	0.0174	0.0690	0.2587	0.9475	0.0057	0.0382	0.1945	0.9455	0.0073	0.0323	0.1757	0.9380
dr10_v	0.0078	0.0770	0.2825	0.9480	0.0105	0.0406	0.1985	0.9430	0.0077	0.0402	0.1982	0.9480
dr10_c1	0.0149	0.0525	0.2282	0.9495	0.0116	0.0314	0.1723	0.9450	0.0115	0.0253	0.1548	0.9450
dr00_v	0.0536	0.1041	0.3173	0.9465	0.0562	0.0541	0.2234	0.9465	0.0614	0.0542	0.2227	0.9430
dr00_c1	0.0494	0.0666	0.2488	0.9445	0.0571	0.0401	0.1903	0.9400	0.0601	0.0333	0.1677	0.9290

S4.3 Comparison to the difference-based calibration method

Here we compare the performance of our proposed FQTE with the difference-based calibration method in Yang et al. (2020). As we mentioned in Section 3.2 and Section S3.3, a direct extension of their method leads to $\hat{\Delta}_p^{\text{diff}} = \hat{\Delta}_p^{\mathcal{V}} - \hat{\Gamma}^\top \hat{V}^{-1}\hat{C}_{\text{ep}}$, where the covariance matrices Γ and V are rather complicated as discussed in S3.3. Additionally, we also consider the difference-based method using $\hat{q}_{1,p}^{\mathcal{V},\text{Conf}} - \hat{q}_{1,p}^{\mathcal{O},\text{Conf}}$ and $\hat{q}_{0,p}^{\mathcal{V},\text{Conf}} - \hat{q}_{0,p}^{\mathcal{O},\text{Conf}}$ as consistent estimators for zero. We call these two methods as "DIFF-c" and "DIFF-s" respectively. Here "s" and "c" stand for "separated" and "combined" respectively. Covariance matrices involved in the difference-based methods are estimated using bootstrap methods. We evaluate the BIAS, MSE, SE as well as the time consumed (in seconds) by both methods. Specifically, we considered Setting 1 and 3 mentioned in the Supplementary Materials Section S4.2 and varied the sample sizes (N, n) from $(1000, 200)$, $(2000, 500)$, $(2000, 1000)$ and $(5000, 1000)$. The simulation in each setting is based on 2000 replications and the number of bootstrap samples is 200. The simulations are conducted using R software on a Linux server with Intel(R) Xeon(R) CPU E5-2643 v4 @ 3.40 GHz. The CPU on the server has 24 cores in total. Table S4.4 and S4.5 display the comparison of the two methods for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ under Setting 1 and Setting 3 with various sample sizes when both the OR and PS models are correctly specified.

As we can see from Table S4.4 and Table S4.5, the performance of our method is equally matched to that of the difference-based approach in terms of MSE and SE. Our method has a slight advantage over the difference-based approaches in most cases. We also note that the BIASes of our method are slightly larger than that of the difference-based method, which may be caused by the variance estimation. However, since the variance term dominates the bias term, we still benefit from the larger decrease of variances. Moreover, our method enjoys a much shorter computing time compared to the difference-based approaches, especially when the sample size increases.

S4.4 FQTE under the MAR assumption

In this subsection, we present the empirical performance of our FQTE under the relaxed MAR assumption. We generate the entire data as before and consider the entire sample size $N \in \{1000, 2000, 5000\}$. We further set the the missing mechanism as $\text{logit}\{\pi(Y, T, X)\} = \text{pr}(R = 1|Y, T, X) = -0.7 - 0.1Y_i - 0.1X_{1i} - 0.3T_i$ and select each sample to the validation data with probability π_i. The size of the sample selected into the validation data

Table S4.4: Comparison between FQTE and the difference-based method for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ in Setting 1.

		FQTE					DIFF_c					DIFF_s				
Sample size	Method	BIAS	MSE	SE	CR	TIME	BIAS	MSE	SE	CR	TIME	BIAS	MSE	SE	CR	TIME
								$\Delta_{0.5}$								
$N = 1000$	dr11_v	0.0094	0.1013	0.3380	0.9550	0.0183	0.0094	0.1013	0.3380	0.9550	0.0183	0.0094	0.1013	0.3380	0.9550	0.0183
$n = 200$	**dr11_c1**	0.0154	0.0460	0.2308	0.9605	0.0379	0.0017	0.0426	0.2331	0.9660	22.7334	0.0004	0.0438	0.2302	0.9635	22.5769
$N = 2000$	dr11_v	0.0047	0.0369	0.2090	0.9655	0.0264	0.0047	0.0369	0.2090	0.9655	0.0264	0.0047	0.0369	0.2090	0.9655	0.0264
$n = 500$	**dr11_c1**	0.0045	0.0187	0.1428	0.9525	0.0564	0.0021	0.0187	0.1462	0.9580	27.2018	0.0001	0.0191	0.1447	0.9545	27.2691
$N = 2000$	dr11_v	0.0027	0.0199	0.1464	0.9515	0.0382	0.0027	0.0199	0.1464	0.9515	0.0382	0.0027	0.0199	0.1464	0.9515	0.0382
$n = 1000$	**dr11_c1**	0.0026	0.0130	0.1166	0.9555	0.0637	0.0007	0.0133	0.1193	0.9485	29.9578	0.0007	0.0135	0.1183	0.9460	30.0104
$N = 5000$	dr11_v	0.0040	0.0200	0.1462	0.9610	0.0522	0.0040	0.0200	0.1462	0.9610	0.0522	0.0040	0.0200	0.1462	0.9610	0.0522
$n = 1000$	**dr11_c1**	0.0058	0.0083	0.0949	0.9620	0.1301	0.0004	0.0084	0.0979	0.9625	48.5737	0.0017	0.0085	0.0969	0.9595	48.5852
								$\Delta_{0.75}$								
$N = 1000$	dr11_v	0.0061	0.1225	0.3614	0.9455	0.0200	0.0061	0.1225	0.3614	0.9455	0.0200	0.0061	0.1225	0.3614	0.9455	0.0200
$n = 200$	**dr11_c1**	0.0335	0.0679	0.2557	0.9400	0.0411	0.0043	0.0613	0.2790	0.9710	23.3149	0.0083	0.0641	0.2753	0.9625	23.1924
$N = 2000$	dr11_v	0.0003	0.0492	0.2270	0.9545	0.0264	0.0003	0.0492	0.2270	0.9545	0.0286	0.0003	0.0492	0.2270	0.9545	0.0286
$n = 500$	**dr11_c1**	0.0141	0.0260	0.1610	0.9520	0.0564	0.0005	0.0264	0.1749	0.9620	28.9842	0.0018	0.0270	0.1729	0.9580	29.0406
$N = 2000$	dr11_v	0.0002	0.0249	0.1597	0.9505	0.0440	0.0002	0.0249	0.1597	0.9505	0.0440	0.0002	0.0249	0.1597	0.9505	0.0440
$n = 1000$	**dr11_c1**	0.0071	0.0165	0.1295	0.9535	0.0776	0.0002	0.0173	0.1385	0.9565	36.7562	0.0009	0.0176	0.1374	0.9540	36.8414
$N = 5000$	dr11_v	0.0016	0.0262	0.1595	0.9465	0.0602	0.0016	0.0262	0.1595	0.9465	0.0602	0.0016	0.0262	0.1595	0.9465	0.0602
$n = 1000$	**dr11_c1**	0.0135	0.0123	0.1078	0.9450	0.1682	0.0022	0.0121	0.1165	0.9570	55.0355	0.0045	0.0123	0.1151	0.9525	54.9735

Table S4.5: Comparison between FQTE and the difference-based method for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ in Setting 3.

		FQTE					DIFF_c					DIFF_s				
Sample size	Method	BIAS	MSE	SE	CR	TIME	BIAS	MSE	SE	CR	TIME	BIAS	MSE	SE	CR	TIME
								$\Delta_{0.5}$								
$N = 1000$	dr11_v	0.0001	0.0962	0.3240	0.9560	0.0186	0.0001	0.0962	0.3240	0.9560	0.0186	0.0001	0.0962	0.3240	0.9560	0.0186
$n = 200$	**dr11_c1**	0.0270	0.0478	0.2265	0.9570	0.0350	0.0091	0.0494	0.2497	0.9665	22.2623	0.0123	0.0503	0.2458	0.9600	22.0893
$N = 2000$	dr11_v	0.0049	0.0375	0.2009	0.9580	0.0253	0.0049	0.0375	0.2009	0.9580	0.0253	0.0049	0.0375	0.2009	0.9580	0.0253
$n = 500$	**dr11_c1**	0.0055	0.0177	0.1411	0.9640	0.0499	0.0045	0.0193	0.1520	0.9665	26.2922	0.0021	0.0195	0.1499	0.9620	26.1966
$N = 2000$	dr11_v	0.0024	0.0185	0.1405	0.9630	0.0363	0.0024	0.0185	0.1405	0.9630	0.0363	0.0024	0.0185	0.1405	0.9630	0.0363
$n = 1000$	**dr11_c1**	0.0091	0.0122	0.1137	0.9575	0.0599	0.0037	0.0129	0.1185	0.9525	28.6843	0.0035	0.0129	0.1185	0.9510	28.5697
$N = 5000$	dr11_v	0.0012	0.0184	0.1407	0.9605	0.0455	0.0012	0.0184	0.1407	0.9605	0.0455	0.0012	0.0184	0.1407	0.9605	0.0455
$n = 1000$	**dr11_c1**	0.0066	0.0084	0.0943	0.9625	0.1066	0.0012	0.0087	0.1001	0.9660	45.5979	0.0027	0.0088	0.0984	0.9630	45.6621
								$\Delta_{0.75}$								
$N = 1000$	dr11_v	0.0036	0.1093	0.3410	0.9525	0.0209	0.0036	0.1093	0.3410	0.9525	0.0209	0.0036	0.1093	0.3410	0.9525	0.0209
$n = 200$	**dr11_c1**	0.0292	0.0562	0.2389	0.9480	0.0416	0.0005	0.0570	0.2658	0.9605	23.4753	0.0058	0.0594	0.2625	0.9515	23.3459
$N = 2000$	dr11_v	0.0040	0.0412	0.2121	0.9620	0.0272	0.0040	0.0412	0.2121	0.9620	0.0272	0.0040	0.0412	0.2121	0.9620	0.0272
$n = 500$	**dr11_c1**	0.0155	0.0201	0.1486	0.9590	0.0607	0.0008	0.0210	0.1637	0.9685	29.0447	0.0034	0.0214	0.1617	0.9645	29.0966
$N = 2000$	dr11_v	0.0024	0.0212	0.1488	0.9550	0.0417	0.0024	0.0212	0.1488	0.9550	0.0417	0.0024	0.0212	0.1488	0.9550	0.0417
$n = 1000$	**dr11_c1**	0.0071	0.0135	0.1206	0.9555	0.0792	0.0009	0.0145	0.1273	0.9535	37.3124	0.0008	0.0145	0.1263	0.9535	37.3713
$N = 5000$	dr11_v	0.0026	0.0197	0.1489	0.9640	0.0529	0.0026	0.0197	0.1489	0.9640	0.0529	0.0026	0.0197	0.1489	0.9640	0.0529
$n = 1000$	**dr11_c1**	0.0097	0.0090	0.0999	0.9620	0.1495	0.0008	0.0093	0.1071	0.9655	54.9016	0.0020	0.0093	0.1057	0.9625	54.8035

is around a quarter of the entire sample size. When estimating the QTE using the modified method, we first estimate π_i based on logistic regression to obtain $\hat{\pi}_i$. To illustrate the importance of our modified FQTE in Section S2 with data missing at random, we also compare it to the naive FQTE based on the MCAR assumption, which is violated here. Table S4.6 and Table S4.7 display the results for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ in Setting 3.

We can see that with data missing at random, the naive FQTEs based on the MCAR assumption lose their efficiency since in this case we can no longer treat all the entire data as i.i.d. In contrast, our modified FQTEs as proposed in Section S2 perform better than the naive FQTEs in terms of MSE, SE, as well as CR. This demonstrates the necessity of modification of our FQTE when the MCAR assumption no longer holds. Comparing the calibrated FQTEs to the corresponding uncalibrated estimators, we find considerable improvement in MSE and SE for both the modified and naive FQTEs. This shows the feasibility of our proposed calibration method. Besides, the coverage rates move closer to the nominal coverage rate of 95%, which confirms the asymptotic results established in Section S2.

Table S4.6: Results for estimating $\Delta_{0.5}$ in Setting 3 under the MAR assumption.

$\Delta_{0.5}$	Modified FQTE				Naive FQTE			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
				$N = 1000$				
dr11_v	0.0072	0.0716	0.3191	0.9750	0.2278	0.1291	0.2835	0.8650
dr11_c1	0.0215	0.0426	0.2230	0.9555	0.1103	0.0637	0.2278	0.9145
dr01_v	0.0077	0.0735	0.3218	0.9750	0.2270	0.1302	0.2868	0.8715
dr01_c1	0.0220	0.0451	0.2273	0.9530	0.1102	0.0658	0.2320	0.9200
dr10_v	0.0059	0.0737	0.3211	0.9745	0.2265	0.1288	0.2848	0.8680
dr10_c1	0.0201	0.0436	0.2252	0.9560	0.1099	0.0637	0.2293	0.9205
dr00_v	0.0569	0.0784	0.3248	0.9745	0.1646	0.1064	0.2875	0.9055
dr00_c1	0.0438	0.0440	0.2212	0.9500	0.0447	0.0533	0.2271	0.9360
				$N = 2000$				
dr11_v	0.0020	0.0355	0.2212	0.9760	0.2239	0.0863	0.1981	0.7975
dr11_c1	0.0112	0.0192	0.1529	0.9645	0.1029	0.0341	0.1586	0.9015
dr01_v	0.0009	0.0364	0.2232	0.9760	0.2246	0.0874	0.2004	0.8000
dr01_c1	0.0124	0.0202	0.1561	0.9675	0.1043	0.0357	0.1615	0.9060
dr10_v	0.0031	0.0358	0.2229	0.9775	0.2235	0.0863	0.1988	0.7965
dr10_c1	0.0101	0.0195	0.1543	0.9660	0.1029	0.0346	0.1595	0.9070
dr00_v	0.0661	0.0406	0.2256	0.9730	0.1614	0.0629	0.2009	0.8760
dr00_c1	0.0532	0.0213	0.1512	0.9535	0.0378	0.0248	0.1580	0.9505
				$N = 5000$				
dr11_v	0.0027	0.0130	0.1370	0.9805	0.2231	0.0643	0.1232	0.5480
dr11_c1	0.0053	0.0072	0.0943	0.9705	0.1010	0.0191	0.0983	0.8315
dr01_v	0.0027	0.0134	0.1383	0.9795	0.2227	0.0644	0.1246	0.5655
dr01_c1	0.0054	0.0076	0.0963	0.9635	0.1010	0.0196	0.1000	0.8390
dr10_v	0.0030	0.0133	0.1380	0.9780	0.2234	0.0644	0.1235	0.5465
dr10_c1	0.0052	0.0074	0.0951	0.9660	0.1016	0.0193	0.0987	0.8305
dr00_v	0.0674	0.0180	0.1397	0.9580	0.1596	0.0400	0.1249	0.7610
dr00_c1	0.0591	0.0105	0.0928	0.9190	0.0350	0.0099	0.0978	0.9480

Table S4.7: Results for estimating $\Delta_{0.75}$ in Setting 3 with data missing at random.

$\Delta_{0.5}$	Modified FQTE				Naive FQTE			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
				$N = 1000$				
dr11_v	0.0072	0.0716	0.3191	0.9750	0.2278	0.1291	0.2835	0.8650
dr11_c1	0.0215	0.0426	0.2230	0.9555	0.1103	0.0637	0.2278	0.9145
dr01_v	0.0077	0.0735	0.3218	0.9750	0.2270	0.1302	0.2868	0.8715
dr01_c1	0.0220	0.0451	0.2273	0.9530	0.1102	0.0658	0.2320	0.9200
dr10_v	0.0059	0.0737	0.3211	0.9745	0.2265	0.1288	0.2848	0.8680
dr10_c1	0.0201	0.0436	0.2252	0.9560	0.1099	0.0637	0.2293	0.9205
dr00_v	0.0569	0.0784	0.3248	0.9745	0.1646	0.1064	0.2875	0.9055
dr00_c1	0.0438	0.0440	0.2212	0.9500	0.0447	0.0533	0.2271	0.9360
				$N = 2000$				
dr11_v	0.0020	0.0355	0.2212	0.9760	0.2239	0.0863	0.1981	0.7975
dr11_c1	0.0112	0.0192	0.1529	0.9645	0.1029	0.0341	0.1586	0.9015
dr01_v	0.0009	0.0364	0.2232	0.9760	0.2246	0.0874	0.2004	0.8000
dr01_c1	0.0124	0.0202	0.1561	0.9675	0.1043	0.0357	0.1615	0.9060
dr10_v	0.0031	0.0358	0.2229	0.9775	0.2235	0.0863	0.1988	0.7965
dr10_c1	0.0101	0.0195	0.1543	0.9660	0.1029	0.0346	0.1595	0.9070
dr00_v	0.0661	0.0406	0.2256	0.9730	0.1614	0.0629	0.2009	0.8760
dr00_c1	0.0532	0.0213	0.1512	0.9535	0.0378	0.0248	0.1580	0.9505
				$N = 5000$				
dr11_v	0.0027	0.0130	0.1370	0.9805	0.2231	0.0643	0.1232	0.5480
dr11_c1	0.0053	0.0072	0.0943	0.9705	0.1010	0.0191	0.0983	0.8315
dr01_v	0.0027	0.0134	0.1383	0.9795	0.2227	0.0644	0.1246	0.5655
dr01_c1	0.0054	0.0076	0.0963	0.9635	0.1010	0.0196	0.1000	0.8390
dr10_v	0.0030	0.0133	0.1380	0.9780	0.2234	0.0644	0.1235	0.5465
dr10_c1	0.0052	0.0074	0.0951	0.9660	0.1016	0.0193	0.0987	0.8305
dr00_v	0.0674	0.0180	0.1397	0.9580	0.1596	0.0400	0.1249	0.7610
dr00_c1	0.0591	0.0105	0.0928	0.9190	0.0350	0.0099	0.0978	0.9480

S4.5 FQTE with more pseudo quantiles

We also consider calibrating the initial DR estimators using more pseudo quantiles. Here in Table S4.8 we display the results in Setting 3 for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ when both the OR and PS models are correctly specified. The results in other settings are similar and thus omitted here.

Table S4.8: Simulation Results for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ with more calibrating quantiles in Setting 3.

	$N = 2000, n = 500$				$N = 2000, n = 1000$				$N = 5000, n = 1000$			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
						$\Delta_{0.5}$						
dr11_v	0.0091	0.0358	0.2017	0.9565	0.0001	0.0185	0.1406	0.9610	0.0028	0.0188	0.1407	0.9535
dr11_c1	0.0131	0.0172	0.1418	0.9655	0.0021	0.0119	0.1139	0.9560	0.0063	0.0083	0.0943	0.9605
dr11_c2	0.0154	0.0171	0.1409	0.9630	0.0030	0.0120	0.1137	0.9570	0.0079	0.0083	0.0938	0.9585
dr11_c3	0.0092	0.0169	0.1390	0.9620	0.0011	0.0119	0.1132	0.9565	0.0049	0.0081	0.0926	0.9555
						$\Delta_{0.75}$						
dr11_v	0.0000	0.0454	0.2131	0.9490	0.0076	0.0209	0.1487	0.9560	0.0001	0.0205	0.1490	0.9575
dr11_c1	0.0188	0.0236	0.1497	0.9425	0.0097	0.0140	0.1205	0.9515	0.0090	0.0093	0.1002	0.9615
dr11_c2	0.0159	0.0233	0.1481	0.9350	0.0088	0.0139	0.1200	0.9525	0.0079	0.0092	0.0991	0.9575
dr11_c3	0.0133	0.0231	0.1477	0.9360	0.0079	0.0139	0.1199	0.9520	0.0065	0.0092	0.0988	0.9575

We can draw from Table S4.8 that a one-dimensional calibration quantile is adequate for efficiency improvement, more quantiles seem to contribute a little to the reduction of the standard error (SE).

S4.6 FQTE with a small validation sample size

We also test our fused estimators when the sample size of the validation dataset is small. Table S4.9 display the results of FQTEs for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ in Setting 3. The results in other settings are similar and thus omitted here.

From Table S4.9, we can see the role of the auxiliary dataset is of great importance for efficiency improvement when the sample size of the validation dataset is relatively small. Specifically, when $N = 2000$ and $n = 200$, the MSE of our FQTEs is reduced by more than half of the initial ones. This suggests we make good use of the auxiliary dataset when the sample size of the validation dataset is limited.

Table S4.9: Simulation Results for estimating $\Delta_{0.5}$ and $\Delta_{0.75}$ in Setting 3 with a small validation dataset.

	$N = 1000, n = 200$				$N = 1000, n = 500$				$N = 2000, n = 200$			
Method	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR	BIAS	MSE	SE	CR
						$\Delta_{0.5}$						
dr11_v	0.0098	0.0273	0.1738	0.9500	0.0009	0.0107	0.1072	0.9490	0.0036	0.0271	0.1729	0.9620
dr11_c1	0.0164	0.0141	0.1247	0.9580	0.0070	0.0075	0.0886	0.9555	0.0316	0.0123	0.1166	0.9590
dr01_v	0.0084	0.0278	0.1759	0.9515	0.0009	0.0108	0.1084	0.9530	0.0018	0.0280	0.1749	0.9590
dr01_c1	0.0176	0.0146	0.1275	0.9585	0.0070	0.0076	0.0900	0.9570	0.0293	0.0129	0.1195	0.9610
dr10_v	0.0104	0.0273	0.1744	0.9480	0.0020	0.0108	0.1076	0.9455	0.0022	0.0274	0.1735	0.9570
dr10_c1	0.0151	0.0142	0.1252	0.9580	0.0060	0.0075	0.0890	0.9550	0.0294	0.0124	0.1170	0.9610
dr00_v	0.0715	0.0293	0.1763	0.9440	0.0647	0.0116	0.1087	0.9425	0.0624	0.0287	0.1753	0.9565
dr00_c1	0.0479	0.0142	0.1227	0.9450	0.0571	0.0077	0.0882	0.9450	0.0383	0.0116	0.1138	0.9520
						$\Delta_{0.75}$						
dr11_v	0.0027	0.0343	0.2106	0.9675	0.0062	0.0137	0.1307	0.9640	0.0054	0.0338	0.2098	0.9590
dr11_c1	0.0363	0.0183	0.1550	0.9730	0.0088	0.0093	0.1086	0.9730	0.0407	0.0161	0.1466	0.9630
dr01_v	0.0032	0.0361	0.2141	0.9690	0.0056	0.0142	0.1327	0.9620	0.0041	0.0359	0.2133	0.9550
dr01_c1	0.0367	0.0196	0.1585	0.9690	0.0094	0.0097	0.1105	0.9685	0.0394	0.0177	0.1503	0.9620
dr10_v	0.0025	0.0338	0.2096	0.9640	0.0062	0.0135	0.1300	0.9625	0.0026	0.0335	0.2084	0.9575
dr10_c1	0.0360	0.0179	0.1534	0.9710	0.0086	0.0092	0.1079	0.9710	0.0371	0.0156	0.1449	0.9640
dr00_v	0.0592	0.0357	0.2083	0.9555	0.0692	0.0153	0.1293	0.9500	0.0609	0.0363	0.2072	0.9435
dr00_c1	0.0294	0.0177	0.1468	0.9495	0.0550	0.0103	0.1049	0.9470	0.0297	0.0159	0.1374	0.9375

References

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*(1), 259–276.

Han, P., L. Kong, J. Zhao, and X. Zhou (2019). A general framework for quantile estimation with incomplete data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*(2), 305–333.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Van der Vaart, A. W. and J. Wellner (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Westreich, D., J. Lessler, and M. J. Funk (2010). Propensity score estimation: neural networks, support vector machines, decision trees (cart), and meta-classifiers as alternatives to logistic regression. *Journal of clinical epidemiology 63*(8), 826–833.

Yang, S. and P. Ding (2020). Combining multiple observational data sources to estimate causal effects. *Journal of the American Statistical Association 115*(531), 1540–1554.

Yang, S., D. Zeng, and X. Wang (2020). Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *arXiv preprint arXiv:2007.12922.*