Reading Introduction

Estimating Counterfactual Treatment Effects to Assess External Validity

a. What is the question (of the paper)?

This paper deals with the *quantile counterfactual treatment effect* (*QCTE*) to evaluate potential treatment heterogeneity among individuals with some presumptions. Besides, it also discusses how to correctly choose a multiple bootstrap to build up the QCTE estimator and its uniform confidence interval.

b. Why should we care about it?

Most of the literature about program assessment attach importance to intrinsic validity instead of the extrinsic one. Since only little discussion on forecasting the treatment effects from one case to another, the paper tries to bridge the gap by testing the quantile treatment effect in the counterfactual case. To be more precise, it relaxes the *randomized experiment requirement* and extends the analysis to a unified method for QCTE. Additionally, compared with the previous approaches in literature review, it also generalizes the asymptotic analysis, employs the multiplier bootstrap, accounts for the treated environment, and puts forward a diverse and feasibly monotonizing means.

Real Word Example:

Statistics fails to explain casual inference, especially when evaluated with inaccurate substantial knowledge. In such cases, we need both priori and ex-post characteristics about the external treatment effects in various situations. For instance, we may care about a reported traumatic event, say, a severe earthquake, at the baseline assessment as a potential risk factor for incident depression during the follow-up period. Then, we may evaluate the effect of an index treatment (eg. getting comfort from a psychiatrist) in comparison of another treatment (eg. without comfort from others) on an outcome that can be binary or quantitative (eg. the measured level of incident depression psychologically). For every individual, the result can be observed only under one instead of both conditions whose result is counterfactual. Counterfactual effect is common since one can only be assigned to one environment at any fixed time. Thus, it is of importance to precisely estimate counterfactual treatment effects in order to evaluate external validity.

c. What is your (or the author's) answer?

(1) Job Corps is effective for individuals from 40^{th} to 85^{th} quantiles, but not for those below 40^{th} quantile.
(2) The strong economic performance may be a reason for the ineffectiveness of Job Corps at the low quantile while the Job Corps performs better than the counterfactual program.
(3) The skill hypothesis remains insignificant at the lower tail. Probably the low cognitive and non-cognitive skills result in the ineffective of Job Corps.

d. How did you (or the author) get there?

To begin with, the authors focuses on *the model framework*, *the parameters of interest*, and *the identification strategy* including *unconfoundedness* and *invariance of conditional distributions*. After that, they introduce the *estimation process* with regular conditions as well as the *essential asymptotic traits* for the robustness of the *multiplier bootstrap*, followed by the establishment of *uniform confidence bands* with implementation process. Then, the main case (i.e. *average counterfactual treatment effect*, or abbreviated as *ACTE*) and treated cases are discussed, after which they illustrate the *simulation* along with the *empirical research* to draw the main conclusions.

Notations

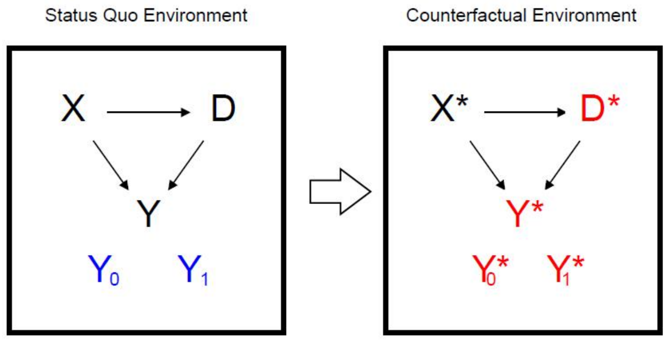

Figure 1: Model Framework

$Y_d = m_d(X, \varepsilon_d), \quad d = 0, 1$ (2.1)

Y_d : the corresponding potential outcomes for d, where d={0,1} is a binary treatment indicator, where the observed outcome is denoted as $Y = DY_1 + (1-D)Y_0$

$X = (X_1, \ldots, X_k)$: a k-dimension vector of pre-treatment covariates

ε_d : the unobserved error known as individual heterogeneity

m_d : the unknown function (i.e. a decision rule) determines Y_d given X and ε_d

$Y_d^* = m_d(X^*, \varepsilon_d), \quad d = 0, 1$ (2.2)

Y_d^* : the counterfactual potential outcome determined by the unknown decision rule

X^* : the observed counterfactual characteristics, where $X^* = \pi(X)$ for some unknown function π

ACTE: $\Delta^* = \mathbb{E}(Y_1^*) - \mathbb{E}(Y_0^*)$ (2.3)

QCTE: $\Delta^*(\tau) = \mathbb{Q}_{Y_1^*}(\tau) - \mathbb{Q}_{Y_0^*}(\tau)$ (2.4)

where $\mathbb{Q}_{Y_d^*}(\tau) = \inf\{y \in \mathcal{Y} : F_{Y_d^*}(y) \geq \tau\}$ with $\mathcal{Y}$ being the support of Y and $F_{Y_d^*}(y)$ the distribution function of Y_d^* . $\mathbb{Q}_{Y_d^*}(\tau)$ would be the ordinary inverse of $F_{Y_d^*}(y)$ if $F_{Y_d^*}(y)$ is continuous and strictly increasing from 0 to 1.

ACTT&QCTT: $\Delta_t^* = \mathbb{E}(Y_1^*|D^*=1) - \mathbb{E}(Y_0^*|D^*=1)$ and $\Delta_t^*(\tau) = \mathbb{Q}_{Y_1^*|D^*}(\tau|1) - \mathbb{Q}_{Y_0^*|D^*}(\tau|1)$ (2.5)

$D^* \in \{0,1\}$: the unknown counterfactual treatment assignment.

The propensity score for all $x \in \mathcal{X}$: $p(x) = \mathbb{P}(D=1|X=x)$

where $\mathcal{X}$ and $\mathcal{X}^*$ be the support of X and X* respectively.

Lemma 1: QCTE:

$$\Delta^*(\tau) = \inf_{y \in \mathcal{Y}} \left\{ \int_{\mathcal{X}} F_{Y|D,X}(y|1,x)\, dF_{X^*}(x) \geq \tau \right\} - \inf_{y \in \mathcal{Y}} \left\{ \int_{\mathcal{X}} F_{Y|D,X}(y|0,x)\, dF_{X^*}(x) \geq \tau \right\}$$

The distribution function estimator given by $\widetilde{F}_{Y_d^*}(y) = \frac{1}{n^*}\sum_{j=1}^{n^*} \widetilde{F}_{Y_d|X}(y|X_j^*)$ (3.1)

where a random sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ and a random sample $\{X_j^*\}_{j=1}^{n^*}$ where the sample size n and n*;

and $\widetilde{F}_{Y_d|X}(y|x)$ is the Nadaraya-Waston estimator

$$\widetilde{F}_{Y_d|X}(y|x) = \frac{\sum_{i=1}^n \mathbb{1}\{Y_i \leq y\}\, \mathbb{1}\{D_i = d\}\, \mathcal{K}_{x,h}(X_i - x)}{\sum_{i=1}^n \mathbb{1}\{D_i = d\}\, \mathcal{K}_{x,h}(X_i - x)}$$ (3.2)

$\mathbb{1}\{\cdot\}$: the indicator function

$\mathcal{K}_{x,h}(\cdot) = h^{-k}\mathcal{K}_x(\cdot/h)$: a higher-order boundary kernel whose shape adapts when x is near the boundary of $\mathcal{X}$ with $h = h_n$ the bandwidth factor.

Define functions $\phi_1, \phi_2 \text{ and } \phi$ s.t. for any function g with $\sup_{y\in\mathcal{Y}} g(y) > 0$,

$$\phi_1(g)(y) = \max\left\{0, \sup_{y'\le y} g(y')\right\}, \qquad \phi_2(g)(y) = \frac{g(y)}{\sup_{y'\in\mathcal{Y}} g(y')}, \qquad \phi = \phi_1 \circ \phi_2$$

Definition of the modified version of (3.1): $\widehat{F}_{Y_d^*}(y) = \phi(\widetilde{F}_{Y_d^*})(y)$ (3.3)

where $\widehat{F}_{Y_d^*}(y)$ is monotonically increasing and bounded between 0 and 1, yielding a proper distribution function estimator for $F_{Y_d^*}(y)$.

QECT estimator: $\widehat{\Delta}^*(\tau) = \widehat{\mathbb{Q}}_{Y_1^*}(\tau) - \widehat{\mathbb{Q}}_{Y_0^*}(\tau)$

where $\widehat{\mathbb{Q}}_{Y_d^*}(\tau) = \inf\{y \in \mathcal{Y} : \widehat{F}_{Y_d^*}(y) \ge \tau\}$ (3.4)

Lemma 2: $\sqrt{n}\left(\widehat{\boldsymbol{F}}(\cdot) - \boldsymbol{F}(\cdot)\right) \Rightarrow \mathcal{F}(\cdot)$

where $\mathcal{F}(y) = (\mathcal{F}_0(y_0), \mathcal{F}_1(y_1))^T$ is a 2-dimensional zero-mean Gaussian process with covariance function $\Psi^F(\boldsymbol{y}, \boldsymbol{y}') = \mathbb{E}[\varrho^F(\boldsymbol{y}, X)\varrho^F(\boldsymbol{y}', X)^T] + \mathbb{E}[\varphi^F(\boldsymbol{y}, X^*)\varphi^F(\boldsymbol{y}', X^*)^T]$, where $\varrho^F(\boldsymbol{y}, X) = (\varrho_0^F(y_0, X), \varrho_1^F(y_1, X))^T$ *and* $\varphi^F(\boldsymbol{y}, X^*) = (\varphi_0^F(y_0, X^*), \varphi_1^F(y_1, X^*))^T$ are given by

$$\varrho_d^F(y, X) = \frac{\mathbb{1}\{D = d\}\left[\mathbb{1}\{Y \le y\} - F_{Y_d|X}(y|X)\right]}{p(X)^d\left[1 - p(X)\right]^{1-d}} \frac{f_{X^*}(X)}{f_X(X)} \qquad (3.5)$$

$$\varphi_d^F(y, X^*) = \sqrt{\lambda}\left[F_{Y_d|X}(y|X^*) - F_{Y_d^*}(y)\right]$$

For d=0,1, and the convergence is in.

$\sqrt{n}(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y))$ is asymptotic linear in the following expression:

$$\begin{aligned}\sqrt{n}\left(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y)\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\mathbb{1}\{D_i = d\}\left[\mathbb{1}\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\right]}{p(X_i)^d\left[1 - p(X_i)\right]^{1-d}} \frac{f_{X^*}(X_i)}{f_X(X_i)} \\ &+ \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*} \sqrt{\lambda}\left[F_{Y_d|X}(y|X_j^*) - F_{Y_d^*}(y)\right] + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varrho_d^F(y, X_i) + \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*} \varphi_d^F(y, X_j^*) + o_p(1).\end{aligned}$$

Let x+x*, $\lambda = 1 \text{ and } Z = (Y, D, X, X^*)$, the sum of $\varrho_d^F(y, X) \text{ and } \varphi_d^F(y, X^*)$ would become:

$$\psi_d^F(y, Z) = \frac{\mathbb{1}\{D = d\}\left[\mathbb{1}\{Y \le y\} - F_{Y_d|X}(y|X)\right]}{p(X)^d\left[1 - p(X)\right]^{1-d}} + F_{Y_d|X}(y|X) - F_{Y_d}(y)$$

Thm1: $\sqrt{n}\left(\widehat{\Delta}^*(\cdot) - \Delta^*(\cdot)\right) \Rightarrow \mathcal{Q}(\cdot)$

where $\mathcal{Q}(\tau)$ is a Gaussian process with mean 0 and covariance function $\Psi(\tau) = \mathbb{E}[\varrho(\tau, X)\ \varrho(\tau, X)^T] + \mathbb{E}[\varphi(\tau, X^*)\varphi(\tau, X^*)^T]$, where

$$\varrho(\tau, X) = -\left[\frac{\varrho_1^F\left(\mathbb{Q}_{Y_1^*}(\tau), X\right)}{f_{Y_1^*}\left(\mathbb{Q}_{Y_1^*}(\tau)\right)} - \frac{\varrho_0^F\left(\mathbb{Q}_{Y_0^*}(\tau), X\right)}{f_{Y_0^*}\left(\mathbb{Q}_{Y_0^*}(\tau)\right)}\right] \qquad (3.6)$$

$$\varphi(\tau, X^*) = -\left[\frac{\varphi_1^F\left(\mathbb{Q}_{Y_1^*}(\tau), X^*\right)}{f_{Y_1^*}\left(\mathbb{Q}_{Y_1^*}(\tau)\right)} - \frac{\varphi_0^F\left(\mathbb{Q}_{Y_0^*}(\tau), X^*\right)}{f_{Y_0^*}\left(\mathbb{Q}_{Y_0^*}(\tau)\right)}\right]$$

where $\varrho_d^F(y, x)$ *and* $\varphi_d^F(y, x)$ are given in (3.5) and the convergence is in $\ell^\infty([0,1])$.

Test an effect of the counterfactual program for the media: $H_0 : \Delta^*(\tau) = 0 \quad \text{for } \tau = 0.5$

Ordinary t-stat: $$\widehat{\Psi}(\tau) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varrho}(\tau, X_i)\widehat{\varrho}(\tau, X_i)^T + \frac{1}{n^*}\sum_{j=1}^{n^*} \widehat{\varphi}(\tau, X_j^*)\widehat{\varphi}(\tau, X_j^*)^T$$

Functional hypothesis test for all individuals:

$$H_0 : \Delta^*(\tau) = 0 \quad \text{for } \tau \in [0,1] \quad \text{or} \quad H_0 : \Delta^*(\tau) \le 0 \quad \text{for } \tau \in [0,1]$$

With Assumption 4.1, estimate $\varrho(\tau, x) \text{ and } \varphi(\tau, x)$ by

$$\widehat{\varrho}(\tau, x) = -\left[\frac{\widehat{\varrho}_1^F(\widehat{\mathbb{Q}}_{Y_1^*}(\tau), x)}{\widehat{f}_{Y_1^*}(\widehat{\mathbb{Q}}_{Y_1^*}(\tau))} - \frac{\widehat{\varrho}_0^F(\widehat{\mathbb{Q}}_{Y_0^*}(\tau), x)}{\widehat{f}_{Y_0^*}(\widehat{\mathbb{Q}}_{Y_0^*}(\tau))}\right]$$

$$\widehat{\varphi}(\tau, x) = -\left[\frac{\widehat{\varphi}_1^F(\widehat{\mathbb{Q}}_{Y_1^*}(\tau), x)}{\widehat{f}_{Y_1^*}(\widehat{\mathbb{Q}}_{Y_1^*}(\tau))} - \frac{\widehat{\varphi}_0^F(\widehat{\mathbb{Q}}_{Y_0^*}(\tau), x)}{\widehat{f}_{Y_0^*}(\widehat{\mathbb{Q}}_{Y_0^*}(\tau))}\right] \qquad (4.1)$$

where $\widehat{\mathbb{Q}}_{Y_d^*}(\tau)$ is in (3.4) and

$$\widehat{\varrho}_d^F(y,x) = \frac{\mathbb{1}\{D_i = d\}\left[\mathbb{1}\{Y_i \le y\} - \widehat{F}_{Y_d|X}(y|x)\right]}{\widehat{p}(x)^d\,[1-\widehat{p}(x)]^{1-d}} \frac{\widehat{f}_{X^*}(x)}{\widehat{f}_X(x)} \quad (4.2)$$

$$\widehat{\varphi}_d^F(y,x) = \sqrt{\widehat{\lambda}}\left[\widehat{F}_{Y_d|X}(y|x) - \widehat{F}_{Y_d^*}(y)\right]$$

with $\widehat{\lambda} = n/n^*$.

The simulated process for $\mathcal{Q}(\tau)$:

$$\mathcal{Q}^u(\tau) = \begin{cases} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n} U_i\left[\widehat{\varrho}(\tau,X_i) + \widehat{\varphi}(\tau,X_i^*)\right] & \text{if } X^* = \pi(X) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n} U_i\widehat{\varrho}(\tau,X_i) + \dfrac{1}{\sqrt{n^*}}\displaystyle\sum_{j=1}^{n^*} U_j^*\widehat{\varphi}(\tau,X_j^*) & \text{if } X^* \perp\!\!\!\perp X. \end{cases} \quad (4.3)$$

Thm2: $\mathcal{Q}^u(\cdot) \overset{p}{\Rightarrow} \mathcal{Q}(\cdot)$

$$\widehat{F}_{Y_d|X}(y|x) = \phi_1(\widetilde{F}_{Y_d|X})(y|x) \quad (4.4)$$

The estimators for based on the kernel method:

$$\widetilde{p}(x) = \frac{\sum_{i=1}^{n} D_i\,\mathcal{K}_{x,h}(X_i - x)}{\sum_{i=1}^{n}\mathcal{K}_{x,h}(X_i - x)}, \quad \widetilde{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{x,h}(X_i - x), \quad \widetilde{f}_{X^*}(x) = \frac{1}{n^*}\sum_{j=1}^{n^*}\mathcal{K}_{x,h}(X_j^* - x) \quad (4.5)$$

$$\widehat{p}(x) = a_n\,\mathbb{1}\{\widetilde{p}(x) \le a_n\} + \widetilde{p}(x)\,\mathbb{1}\{a_n < \widetilde{p}(x) < 1 - a_n\} + (1-a_n)\,\mathbb{1}\{\widetilde{p}(x) \ge 1 - a_n\} \quad (4.6)$$

where $\{a_n \in (0, 1/2) : n \ge 1\}$ is a positive sequence converging to 0.

For any function g, $\chi(g)(y) = \max\{g(y), b_n\}$

where $\{b_n : n \ge 1\}$ is a non-increasing sequence of positive numbers that converges to 0.

The modified version of the density estimators in (4.5):

$$\widehat{f}_X(x) = \chi(\widetilde{f}_X)(x) \quad \text{and} \quad \widehat{f}_{X^*}(x) = \chi(\widetilde{f}_{X^*})(x) \quad (4.7)$$

The kernel estimator for $f_{Y_d^*}(y)$: $\widetilde{f}_{Y_d^*}(y) = \frac{1}{n^*}\sum_{j=1}^{n^*}\widetilde{f}_{Y_d|X}(y|X_j^*)$

where

$$\widetilde{f}_{Y_d|X}(y|x) = \frac{\sum_{i=1}^{n}\mathcal{W}_{y,\eta}(Y_i - y)\,\mathbb{1}\{D_i = d\}\,\mathcal{K}_{x,h}(X_i - x)}{\sum_{i=1}^{n}\mathbb{1}\{D_i = d\}\,\mathcal{K}_{x,h}(X_i - x)} \quad (4.8)$$

With $\mathcal{W}_{y,\eta}(\cdot) = \eta^{-1}\mathcal{W}(\cdot/\eta)$ a boundary kernel and $\eta = \eta_n$ the bandwidth in the y direction.

$$\widehat{f}_{Y_d^*}(y) = \chi(\widetilde{f}_{Y_d^*})(y) \quad (4.9)$$

For a nominal significant level α and for δ_ℓ and $\delta_u \in [0,1]$ with $\delta_\ell < \delta_u$, the critical values for the standardized one- and two-sided bands for that satisfy:

$$\widehat{C}_\alpha^{\text{1-sided}} = \inf_{y\in\mathcal{Y}}\left\{\mathbb{P}\left(\sup_{\tau\in[\delta_\ell,\delta_u]}\frac{\mathcal{Q}^u(\tau)}{\widehat{\sigma}(\tau)} \le y\right) \ge 1-\alpha\right\}$$

$$\widehat{C}_\alpha^{\text{2-sided}} = \inf_{y\in\mathcal{Y}}\left\{\mathbb{P}\left(\sup_{\tau\in[\delta_\ell,\delta_u]}\left|\frac{\mathcal{Q}^u(\tau)}{\widehat{\sigma}(\tau)}\right| \le y\right) \ge 1-\alpha\right\}$$

where $\widehat{\sigma}(\tau) = \widehat{\Psi}^{1/2}(\tau)$ and

$$\widehat{\Psi}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\varrho}(\tau,X_i)\widehat{\varrho}(\tau,X_i)^T + \frac{1}{n^*}\sum_{j=1}^{n^*}\widehat{\varphi}(\tau,X_j^*)\widehat{\varphi}(\tau,X_j^*)^T, \quad \tau\in[\delta_\ell,\delta_u]$$

The two standard one-sided uniform confidence bands for QCTE:

$$\left\{\left(-\infty,\quad \widehat{\Delta}^*(\tau) + \widehat{C}_\alpha^{\text{1-sided}}\frac{\widehat{\sigma}(\tau)}{\sqrt{n}}\right] : \tau\in[\delta_\ell,\delta_u]\right\} \quad (4.10)$$

$$\left\{\left[\widehat{\Delta}^*(\tau) - \widehat{C}_\alpha^{\text{1-sided}}\frac{\widehat{\sigma}(\tau)}{\sqrt{n}},\quad \infty\right) : \tau\in[\delta_\ell,\delta_u]\right\}$$

The standard two-sided uniform confidence bands for QCTE:

$$\left\{\left[\widehat{\Delta}^*(\tau) - \widehat{C}_\alpha^{\text{2-sided}}\frac{\widehat{\sigma}(\tau)}{\sqrt{n}},\quad \widehat{\Delta}^*(\tau) + \widehat{C}_\alpha^{\text{2-sided}}\frac{\widehat{\sigma}(\tau)}{\sqrt{n}}\right] : \tau\in[\delta_\ell,\delta_u]\right\} \quad (4.11)$$

ACTE under Assumption 5.1 and 5.2:

$$\Delta^* = \mathbb{E}(Y_1^*) - \mathbb{E}(Y_0^*) = \mathbb{E}_{X^*}\left[\mathbb{E}(Y|D=1,X) - \mathbb{E}(Y|D=0,X)\right]$$

ACTE estimator:

$$\widehat{\Delta}^* = \frac{1}{n^*}\sum_{j=1}^{n^*}\left[\widehat{\mathbb{E}}(Y_1|X=X_j^*) - \widehat{\mathbb{E}}(Y_0|X=X_j^*)\right]$$

where is the is the Nadaraya-Waston estimator:

Corollary 1: $\sqrt{n}(\widehat{\Delta}^* - \Delta^*) \Rightarrow \mathcal{N}(0, \mathbb{V}_{\Delta^*})$

where $\mathbb{V}_{\Delta^*} = \mathbb{E}\left[\varrho_{\Delta^*}(X)\right]^2 + \mathbb{E}\left[\varphi_{\Delta^*}(X^*)\right]^2$ with

$$\varrho_{\Delta^*}(X) = \left\{\frac{D\left[Y - \mathbb{E}(Y_1|X)\right]}{p(X)} - \frac{(1-D)\left[Y - \mathbb{E}(Y_0|X)\right]}{1-p(X)}\right\}\frac{f_{X^*}(X)}{f_X(X)}$$

$$\varphi_{\Delta^*}(X^*) = \sqrt{\lambda}\left[\mathbb{E}(Y_1|X^*) - \mathbb{E}(Y_0|X^*) - \Delta^*\right]$$

Semiparametric efficiency bound of the ATE estimator given in Hahn (1998):

$$\mathbb{E}\left\{\frac{\mathrm{Var}(Y_1|X)}{p(X)} + \frac{\mathrm{Var}(Y_0|X)}{1-p(X)} + \left[\mathbb{E}(Y_1 - Y_0|X) - \mathbb{E}(Y_1 - Y_0)\right]^2\right\}$$

Lemma 4: ACTT & QCTT:

$$\Delta_t^* = \int_{\mathcal{X}} \{\mathbb{E}(Y|X=x,D=1) - \mathbb{E}(Y|X=x,D=0)\}\frac{p(x)}{\mathbb{E}\left[p(X^*)\right]}\,dF_{X^*}(x)$$

$$\Delta_t^*(\tau) = \inf\left\{y \in \mathcal{Y} : \int_{\mathcal{X}} F_{Y|X,D}(y|x,1)\frac{p(x)}{\mathbb{E}\left[p(X^*)\right]}\,dF_{X^*}(x) \geq \tau\right\}$$
$$- \inf\left\{y \in \mathcal{Y} : \int_{\mathcal{X}} F_{Y|X,D}(y|x,0)\frac{p(x)}{\mathbb{E}\left[p(X^*)\right]}\,dF_{X^*}(x) \geq \tau\right\}$$

ACTT & QCTT estimators:

$$\widehat{\Delta}_t^* = \sum_{j=1}^{n^*}\widehat{p}(X_j^*)\left[\widehat{\mathbb{E}}(Y_1|X=X_j^*) - \widehat{\mathbb{E}}(Y_0|X=X_j^*)\right]\Big/\sum_{j=1}^{n^*}\widehat{p}(X_j^*)$$

$$\widehat{\Delta}_t^*(\tau) = \widehat{\mathbb{Q}}_{Y_1^*|D^*}(\tau|1) - \widehat{\mathbb{Q}}_{Y_0^*|D^*}(\tau|1)$$

where $\widehat{\mathbb{Q}}_{Y_d^*|D^*}(\tau|1) = \inf\{y \in \mathcal{Y} : \widehat{F}_{Y_d^*|D^*}(y|1) \geq \tau\}$ and

$$\widehat{F}_{Y_d^*|D^*}(y|1) = \sum_{j=1}^{n^*}\widehat{p}(X_j^*)\widehat{F}_{Y_d|X}(y|X_j^*)\Big/\sum_{j=1}^{n^*}\widehat{p}(X_j^*)$$

Corollary 2:

$$\sqrt{n}(\widehat{\Delta}_t^* - \Delta_t^*) \Rightarrow \mathcal{N}(0, \mathbb{V}_{\Delta_t^*}),$$

where $\mathbb{V}_{\Delta_t^*} = \mathbb{E}\left[\varrho_{\Delta_t^*}(X)\right]^2 + \mathbb{E}\left[\varphi_{\Delta_t^*}(X^*)\right]^2$ *with*

$$\varrho_{\Delta_t^*}(X) = \frac{p(X)}{\mathbb{E}\left[p(X^*)\right]}\left\{\frac{D\left[Y - \mathbb{E}(Y_1|X)\right]}{p(X)} - \frac{(1-D)\left[Y - \mathbb{E}(Y_0|X)\right]}{1-p(X)}\right\}\frac{f_{X^*}(X)}{f_X(X)},$$

$$\varphi_{\Delta_t^*}(X^*) = \sqrt{\lambda}\frac{p(X^*)}{\mathbb{E}\left[p(X^*)\right]}\left[\mathbb{E}(Y_1|X^*) - \mathbb{E}(Y_0|X^*) - \Delta^*\right].$$

Moreover,

$$\sqrt{n}\left(\widehat{\Delta}_t^*(\cdot) - \Delta_t^*(\cdot)\right) \Rightarrow \mathcal{Q}_t(\cdot),$$

where $\mathcal{Q}_t(\tau)$ *is a Gaussian process with mean zero and covariance function* $\Psi_t(\tau) = \mathbb{E}\left[\varrho_t(\tau, X)\,\varrho_t(\tau, X)^T\right] + \mathbb{E}\left[\varphi_t(\tau, X^*)\varphi_t(\tau, X^*)^T\right]$ *with*

$$\varrho_t(\tau, X) = -\left[\frac{\varrho_{1,t}^F\left(\mathbb{Q}_{Y_1^*|D^*}(\tau|1), X\right)}{f_{Y_1^*|D^*}\left(\mathbb{Q}_{Y_1^*|D^*}(\tau|1)|1\right)} - \frac{\varrho_{0,t}^F\left(\mathbb{Q}_{Y_0^*|D^*}(\tau|1), X\right)}{f_{Y_0^*|D^*}\left(\mathbb{Q}_{Y_0^*|D^*}(\tau|1)|1\right)}\right],$$

$$\varphi_t(\tau, X^*) = -\left[\frac{\varphi_{1,t}^F\left(\mathbb{Q}_{Y_1^*|D^*}(\tau|1), X^*\right)}{f_{Y_1^*|D^*}\left(\mathbb{Q}_{Y_1^*|D^*}(\tau|1)|1\right)} - \frac{\varphi_{0,t}^F\left(\mathbb{Q}_{Y_0^*|D^*}(\tau|1), X^*\right)}{f_{Y_0^*|D^*}\left(\mathbb{Q}_{Y_0^*|D^*}(\tau|1)|1\right)}\right],$$

where $\varrho^F_{d,t}(y,x)$ *and* $\varphi^F_{d,t}(y,x)$ *are given by*

$$\varrho^F_{d,t}(y,X) = \frac{p(X)}{\mathbb{E}\left[p(X^*)\right]} \frac{\mathbb{1}\{D=d\}\left[\mathbb{1}\{Y \le y\} - F_{Y_d|X}(y|X)\right]}{p(X)^d\left[1-p(X)\right]^{1-d}} \frac{f_{X^*}(X)}{f_X(X)}$$

$$\varphi^F_{d,t}(y,X^*) = \sqrt{\lambda}\frac{p(X^*)}{\mathbb{E}\left[p(X^*)\right]}\left[F_{Y_d|X}(y|X^*) - F_{Y_d^*}(y)\right],$$

and the convergence is in $\ell^\infty([0,1])$.

Simulated process to approximate $\mathbb{Q}_t(\cdot)$:

$$\mathcal{Q}^u_t(\tau) = \begin{cases} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n U_i\left[\widehat{\varrho}_t(\tau,X_i) + \widehat{\varphi}_t(\tau,X_i^*)\right] & \text{if } X^* = \pi(X) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n U_i\widehat{\varrho}_t(\tau,X_i) + \dfrac{1}{\sqrt{n^*}}\displaystyle\sum_{j=1}^{n^*} U_j^*\widehat{\varphi}_t(\tau,X_j^*) & \text{if } X^* \perp\!\!\!\perp X, \end{cases}$$

where $\widehat{\varrho}_t(\tau,x)$ and $\widehat{\varphi}_t(\tau,x)$ can be estimated given $\widehat{f}_{Y_d^*|D^*}(y|1) = \max\{\widetilde{f}_{Y_d^*|D^*}(y|1), b_n\}$ with

$$\widetilde{f}_{Y_d^*|D^*}(y|1) = \sum_{j=1}^{n^*}\widehat{p}(X_j^*)\widetilde{f}_{Y_d|X}(y|X_j^*) \Bigg/ \sum_{j=1}^{n^*}\widehat{p}(X_j^*)$$