Supplemental Appendix for "Multivalued Treatments and Decomposition Analysis: An application to the WIA Program"

Wallice Ao[*] Sebastian Calonico[†] Ying-Ying Lee[‡]

August 23, 2019

This supplement contains proofs of all results and other technical details.

Contents

[*]Institute for Defense Analyses. Email: wao@ida.org.

[†]Columbia University, Department of Health Policy and Management. Email: nsc2136@columbia.edu

[‡]University of California Irvine, Department of Economics. Email: yingying.lee@uci.edu

A Functional Delta Method applied to Quantile Effects

We illustrate Corollary 1 by letting Γ be the τ-quantile operator on $\theta_0(y) \equiv F_{Y\langle t|c\rangle}(y)$, i.e., Γ is a generalized inverse $\theta_0^{-1} : (0,1) \to \mathcal{Y}$ given by $\theta_0^{-1}(\tau) = \inf\{y : \theta_0(y) \geq \tau\}$. For the quantile treatment effects for the treated, $\theta_0^{-1}(\tau)$ is the τ-quantile function of $Y(t)$ for the treated s, denoted by $Q_\tau = Q_\tau(Y(t)|T = c) = F^{-1}_{Y(t)|T}(\tau|c)$. Hadamard-differentiability requires $F_{Y\langle t|c\rangle}(y)$ to be continuously differentiable at the τ-quantile, with the derivative being strictly positive and bounded over a compact neighborhood. Additional assumptions might be needed for different policy functionals. For instance, Bhattacharya (2007) gives regularity conditions for Hadamard-differentiability of Lorenz and Gini functionals.

Corollary A.1 (Quantile treatment effect for the treated) *Assume the conditions in Corollary 1. Then uniformly in* $\tau \in [a,b] \subset (0,1)$, $\sqrt{n}\Big(\hat{Q}_\tau - Q_\tau\Big) = n^{-1/2}\sum_{i=1}^n \psi^Q_{t|c}(Z_i;\tau) + o_p(1)$, *where the influence function is*

$$\psi^Q_{t|c}(Z_i;\tau) \equiv \frac{-D_{ti}}{p_c f_{Y(t)|T}(Q_\tau|c)}\,\frac{P_c(X_i)}{P_t(X_i)}\Big(\mathbf{1}\{Y_i \leq Q_\tau\} - F_{Y|TX}(Q_\tau|t,X_i)\Big) + \frac{-D_{ci}}{p_c f_{Y(t)|T}(Q_\tau|c)}\Big(F_{Y|TX}(Q_\tau|t,X_i) - \tau\Big).$$

The quantile process $\sqrt{n}\big(\hat{Q}_\cdot - Q_\cdot\big) = n^{-1/2}\sum_{i=1}^n \psi^Q_{t|c}(Z_i;\cdot) + o_p(1) \Longrightarrow -\mathbb{G}_t\big(Q_\cdot\big)\Big/\theta_0'(Q_\cdot) \equiv \mathbb{G}^Q_t(\cdot)$, *where* $\mathbb{G}^Q_t$ *is a Gaussian process indexed by* τ *in the metric space* $l^\infty([a,b])$ *with mean zero and covariance kernel* $Cov\big(\mathbb{G}^Q_t(\tau_1), \mathbb{G}^Q_t(\tau_2)\big) = \mathbb{E}\big[\psi^Q_{t|c}(Z;\tau_1)\psi^Q_{t|c}(Z;\tau_2)\big]$, *for any* $\tau_1 < \tau_2 \in [a,b]$.

To carry out point-wise inference, the asymptotic variance can be estimated by $n^{-1}\sum_{i=1}^n \hat{\psi}^Q_{t|c}(Z_i;\tau)^2$. Alternatively, Section 3.2 describes a simulation approach to conduct uniform inference.

B Proof of Main Theorems

Notation Let $(Z_1, Z_1, ..., Z_n)$ be an *i.i.d.* sequence of random variables taking values in a probability space $(\mathcal{Z}, \mathcal{B})$ with distribution $\mathcal{P}$. For some measurable function $\phi : \mathcal{Z} \to \mathbb{R}$, define $\mathbb{E}\phi = \int \phi d\mathcal{P}$ and $G_n\phi = \sqrt{n}(n^{-1}\sum_i \phi(Z_i) - \mathbb{E}\phi)$ for the empirical process at ϕ. Denote the true parameters and functions with the superscript $*$, i.e., $e^*(X) \equiv \mathbb{E}[Y|T = t, X]$ and $p^*_t(x) = \mathbb{P}(T = t|X = x)$. Let the true parameters $\gamma_c \equiv \gamma^*$ and $\gamma_{t|c} = \gamma^*_t$ by suppressing the subscript of c for simplicity. Let $\|\cdot\|_\infty$ denote the sup-norm in all arguments for functions. Let $\bar{O}_p(a_n)$ and $\bar{o}_p(a_n)$

be $O_p(a_n)$ and $o_p(a_n)$ uniformly in $y \in \mathcal{Y}$. Denote

$$\hat{F}^{IPW}_{Y\langle t|c\rangle}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \hat{\varphi}_1(Z_i; y, t, c), \text{ where } \varphi_1(Z; y, t, c) \equiv \frac{D_t}{P_t(X)}\mathbf{1}\{Y \leq y\}\frac{P_c(X)}{p_c};$$

$$\hat{F}^{EIF}_{Y\langle t|c\rangle}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \hat{\varphi}_2(Z_i; y, t, c), \text{ where}$$

$$\varphi_2(Z; y, t, c) \equiv \varphi_1(Z; y, t, c) + F_{Y|TX}(y|t, X)\left(\frac{D_c}{P_c(X)} - \frac{D_t}{P_t(X)}\right)\frac{P_c(X)}{p_c}.$$

Assumption B.1 *(i) (a) The class of functions* $\{\theta \mapsto m(\cdot;\theta) : \theta \in \Theta\}$ *is Glivenko-Cantelli; (b)* $\mathbb{E}\big[\sup_{\theta\in\Theta} |m(Y(t);\theta)|\big] < \infty$*; (c)* $\{\theta \mapsto e^*_t(\cdot;\theta) : \theta \in \Theta\}$ *is Glivenko-Cantelli.*

(ii) **(IPW)** *For some* $\delta > 0$*: (a)* $\{\theta \mapsto m(\cdot;\theta) : |\theta - \gamma^*_t| < \delta\}$ *is a Donsker class; (b) there exist constants* $C > 0$ *and* $r \in (0, 1)$ *such that* $\mathbb{E}\big[\sup_{|\theta-\tilde{\theta}|<\delta} |m(Y(t);\theta) - m(Y(t);\tilde{\theta})|^2\big] \leq C\delta^{2r}$ *for all* $\tilde{\theta} \in \Theta$*; (c)* $\mathbb{E}\big[\sup_{|\theta-\gamma^*_t|} |m(Y(t);\theta)|^2\big] < \infty$*;*

(ii) **(EIF)** *For some* $\delta > 0$*, and for all* $x \in \mathcal{X}$ *and all* θ *such that* $|\theta - \gamma^*_t| < \delta$*: (a)* $e^*_t(x,\theta)$ *is continuously differentiable with derivative given by* $\partial_\theta e^*_t(x;\theta) \equiv (\partial/\partial\theta)e^*_t(x;\theta)$ *with* $\mathbb{E}[\sup_{|\theta-\gamma^*_t|<\delta} |\partial_\theta e^*_t(X;\theta)|] < \infty$*; and (b) there exist* $\epsilon > 0$ *and a measurable function* $b(x)$*, with* $\mathbb{E}[|b(x)|] < \infty$*, such that* $|\partial_\theta e_t(x;\theta) - \partial_\theta e^*_t(x;\theta)| \leq b(x)\|e_t - e^*_t\|^\epsilon_\infty$ *for all function* $e_t(\theta) \in \mathcal{E}$ *such that* $\|e_t - e^*_t\|_\infty < \delta$*, where* $\mathcal{E}$ *is a subspace of smooth functions on* $\mathcal{X}$*, endowed with the supremum norm.*

Assumption B.2 *(a)* $p^*_t(\cdot)$ *and* $e^*_t(\cdot, \gamma^*_t)$ *are s times differentiable with* $s/d_x > 5\eta/2 + 1/2$*, where* $\eta = 1$ *or* $1/2$ *depending on whether power series or splines are used as basis functions, respectively; (b) X is continuously distributed with density bounded and bounded away from zero on its compact support* $\mathcal{X}$*; and (c) for some* $\delta > 0$*,* $var[m(Y(t);\theta)|X = x]$ *is uniformly bounded for all* $x \in \mathcal{X}$ *and all* θ *such that* $|\theta - \gamma^*_t| < \delta$*.*

The following assumptions guarantee the existence of the efficiency bounds.

Assumption B.3 *For all* $t \in \mathcal{T}$*: (a)* $\mathbb{E}\big[m(Y\langle t|c\rangle;\theta)^2\big] < \infty$ *and* $\mathbb{E}\big[m(Y\langle t|c\rangle;\theta)\big]$ *is differentiable in* $\theta \in \Theta$ *at* $\gamma_{t|c}$*; (b) Define the gradient matrix*

$$\Gamma_{*|c} \equiv \begin{bmatrix} \Gamma_{0|c} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Gamma_{1|c} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Gamma_{J|c} \end{bmatrix}, \textit{where } \Gamma_{t|c} \equiv \frac{\partial}{\partial\theta^\top}\mathbb{E}\big[m(Y\langle t|c\rangle;\theta)\big]\Big|_{\theta=\gamma_{t|c}}$$

and $\mathbf{0}$ *is a* $d_m \times d_\theta$ *matrix of zeros. The rank of* $\Gamma_{*|c}$ *is* $(J+1)d_\theta$.

Assumption B.4 *For all* $t \in \mathcal{T}$*: (a)* $\mathbb{E}\big[m(Y\langle t|X_1c\rangle;\theta)^2\big] < \infty$ *and* $\mathbb{E}\big[m(Y\langle t|X_1c\rangle;\theta)\big]$ *is differentiable in* $\theta \in \Theta$ *at* $\lambda_{t|c}$*; and (b) Define the gradient matrix*

$$\Gamma_{*|X_1c} \equiv \begin{bmatrix} \Gamma_{0|X_1c} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Gamma_{1|X_1c} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Gamma_{J|X_1c} \end{bmatrix}, \text{ where } \ \Gamma_{t|X_1c} \equiv \frac{\partial}{\partial \theta^\top}\mathbb{E}\big[m(Y\langle t|X_1c\rangle;\theta)\big]\Big|_{\theta=\lambda_{t|c}}$$

and $\mathbf{0}$ *is a* $d_m \times d_\theta$ *matrix of zeros. The rank of* $\Gamma_{*|X_1c}$ *is* $(J+1)d_\theta$.

Assumption B.5 *The nonparametric estimators for* p^* *and* e^* *are described in Section 5.4 in Cattaneo (2010) with* $K = n^v$, $4s/d_x - 6\eta > 1/v > 4\eta + 2$, $\eta = 1$ *or* $\eta = 1/2$ *depending on whether power series or splines are used as basis functions.*

Assumption B.6 *For any* $t, c \in \mathcal{T}$ *and* $y, y_1, y_2 \in \mathcal{Y}$,

(a) $P_t(X) \in \mathcal{C}^\alpha_M(\mathcal{X})$ *and* $F_{Y|TX}(y|t,X) \in \mathcal{C}^\alpha_M(\mathcal{X})$ *for* $\alpha > d_x/2$. $\sup_{x\in\mathcal{X}} \big|F_{Y|TX}(y_1|t,x) - F_{Y|TX}(y_2|t,x)\big| < C|y_1 - y_2|^{1/2}$ *for some positive constant* C.[1]

(b) $\sup_{x\in\mathcal{X}} \|\partial^q_x \hat{P}_t(x) - \partial^q_x P_t(x)\| = o_p(1)$ *and* $\sup_{x\in\mathcal{X}} \big\|\partial^q_x \hat{F}_{Y|TX}(y|t,x) - \partial^q_x F_{Y|TX}(y|t,x)\big\| = o_p(1)$ *for all* $q < d_x/2$.

(c) **(EIF)** $\int \big(\hat{F}_{Y|TX}(y|t,x) - F_{Y|TX}(y|t,x)\big)\big(\hat{P}_c(x) - P_c(x)\big)\ f_X(x)dx = \bar{o}_p(n^{-1/2})$.

(IPW) $\int F_{Y|TX}(y|t,x)\frac{P_c(x)}{p_c}\Big(\frac{\hat{P}_c(x)}{P_c(x)} - \frac{\hat{P}_t(x)}{P_t(x)}\Big)\ f_X(x)dx = n^{-1}\sum_{i=1}^n F_{Y|TX}(y|t,X_i)\frac{P_c(X_i)}{p_c}\Big(\frac{D_{ci}}{P_c(X_i)} - \frac{D_{ti}}{P_t(X_i)}\Big) + \bar{o}_p(n^{-1/2})$.[2]

---

[1] By van der Vaart and Wellner (1996) (P. 154), $\mathcal{C}^\alpha_M(\mathcal{X})$ is defined on a bounded set $\mathcal{X}$ in $\mathbb{R}^{d_x}$ as follows: For any vector $q = (q_1, ..., q_d)$ of q_d integers, let D^q denote the differential operator $D^q = \partial^{q\cdot}/\partial x_1^{q_1} ... \partial x_d^{q_d}$. Denote $q. = \sum_{l=1}^d q_l$ and $\underline{\alpha}$ to be the greatest integer strictly smaller than α. Let $\|g\|_\alpha = \max_{q.\le\underline{\alpha}} \sup_x |D^q g(x)| + \max_{q.\le\underline{\alpha}} \sup_{x\neq x'} |D^q g(x) - D^q g(x')|/\|x - x'\|^{\alpha-\underline{\alpha}}$ where $\max_{q.\le\underline{\alpha}}$ denotes the maximum over $(q_1, ..., q_d)$ such that $q. \le \underline{\alpha}$ and the suprema are taken over the interior of $\mathcal{X}$. Then $\mathcal{C}^\alpha_M(\mathcal{X})$ is the set of all continuous functions $g : \mathcal{X} \subset \mathbb{R}^{d_x} \mapsto \mathbb{R}$ with $\|g\|_\alpha \le M$.

[2] This condition is analogous to the condition (b) in Lemma B.1 that the nonparametrically estimated propensity score captures the correction term in the efficient influence function.

B.1 Proof of Theorem 1

The proof of the asymptotic theorem for $\hat{\gamma}^{IPW}$ and $\hat{\gamma}^{EIF}$ follows the proofs Cattaneo (2010): Lemma B.1 below combines Theorems 2 to 5 and Lemma B.2 modifies Theorem 8. The result follows. $\square$

Lemma B.1 (Asymptotic Linear Representation) *Assume γ^* belongs to the interior of Θ^{J+1}. Let Assumptions 1, 2, B.1, and B.3 hold. Assume (a) $\|\hat{p} - p^*\|_\infty = o_p(n^{-1/4})$.*

(i) Assume (b) $M_{sn}^{IPW}(\gamma^*, \hat{P}, \hat{p}) = M_{sn}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2})$. Then
$$\hat{\gamma}^{IPW} - \gamma^* = -\left(\Gamma'_{*|c} W \Gamma_{*|c}\right)^{-1} \Gamma'_{*|c} W M_{sn}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2}).$$

(ii) Assume (c) $\sup_{|\theta - \gamma^*| < \delta} \|\hat{e}(\theta) - e^*(\theta)\|_\infty = o_p(1)$, for some $\delta > 0$. (d) $M_{sn}^{EIF}(\gamma^*, \hat{P}, \hat{p}, \hat{e}(\gamma^*)) = M_{sn}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2})$. Then
$$\hat{\gamma}^{EIF} - \gamma^* = -\left(\Gamma'_{*|c} W \Gamma_{*|c}\right)^{-1} \Gamma'_{*|c} W M_{sn}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2}).$$

Lemma B.2 (Nonparametric Estimation) *Let Assumptions B.2 and B.5 hold. Then the conditions (a) to (d) in Lemma B.1 hold.*

Proof of Lemma B.1 The consistency $\hat{\gamma}^{IPW} = \gamma^* + o_p(1)$ and $\hat{\gamma}^{EIF} = \gamma^* + o_p(1)$ is directly implied by the proofs of Theorems 2 and 3 in Cattaneo (2010). We only note the main difference in the following proof for $\hat{\gamma}^{IPW}$ and $\hat{\gamma}^{EIF}$. Denote $m_i(\gamma_t) \equiv m(Y_i; \gamma_t)$, $P_{ti} \equiv P_t(X_i)$, and $e_{ti}(\gamma_t) \equiv e_t(X_i; \gamma_t)$.

For $\hat{\gamma}^{IPW}$, in the proof of Theorem 4 in Cattaneo (2010), $\theta = \gamma, \theta_0 = \gamma^*$, and the t-th element of M_{sn}^{IPW} is $M_{[t],n}^{IPW}$. The main difference is in

$$\Delta_{[t],n}(\gamma, P - P^*, p - p^*) = \frac{1}{n} \sum_{i=1}^{n} D_{ti} m_i(\gamma_t) \Lambda_i, \text{ where}$$
$$\Lambda_i \equiv \Lambda_n(X_i) \equiv -\frac{\left(P_{ti} - P_{ti}^*\right)}{P_{ti}^{*2}} \frac{P_{si}^*}{p_c^*} + \frac{1}{P_{ti}^* p_c^*}\left(P_{si} - P_{si}^*\right) - \frac{P_{si}^*}{P_{ti}^* p_c^*}\left(p_c - p_c^*\right).$$

The last two terms are from estimating the ratio for adjusting for the treated $P_c(X_i)/p_c$. We modify

$$R_{3n} = \sup_{|\gamma_t - \gamma_t^*| \leq \delta_n} \frac{\left|M_{[t],n}^{IPW}(\gamma^*, \hat{P}, \hat{p}) - M_{[t],n}^{IPW}(\gamma^*, P^*, p^*) - \Delta_{[t],n}(\gamma^*, \hat{P} - P^*, \hat{p} - p^*)\right|}{1 + C\sqrt{n}|\gamma_t - \gamma_t^*|}$$
$$R_{4n} = \sup_{|\gamma_t - \gamma_t^*| \leq \delta_n} \frac{\left|\Delta_{[t],n}(\gamma^*, \hat{P} - P^*, \hat{p} - p^*) - \Delta_{[t],n}(\gamma, \hat{P} - P^*, \hat{p} - p^*)\right|}{1 + C\sqrt{n}|\gamma_t - \gamma_t^*|}.$$

R_{1n} and R_{2n} are the same.

For $\hat{\gamma}^{EIF}$, Equation (A.2) in Cattaneo (2010) becomes

$$\sup_{|\gamma_t-\gamma_t^*|\le\delta_n} \frac{\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n \left(\hat{e}_{ti}(\gamma_t)-e_{ti}(\gamma_t^*)\right)\left(\frac{D_{ti}}{\hat{P}_{ti}}-\frac{D_{ci}}{\hat{P}_{si}}\right)\frac{\hat{P}_{si}}{\hat{p}_c}\right|}{1+C\sqrt{n}|\gamma_t-\gamma_t^*|} \le R_{1n}+R_{2n}.$$

Define $\hat{\Upsilon}_i \equiv \left(\frac{D_{ti}}{\hat{P}_{ti}}-\frac{D_{ci}}{\hat{P}_{si}}\right)\frac{\hat{P}_{si}}{\hat{p}_c}$ to be approximated by $\Upsilon_i + D_{ti}\Lambda_i - D_{ci}\frac{p_c^*-\hat{p}_c}{p_c^{*2}}$ and $\Upsilon_i \equiv \left(\frac{D_{ti}}{P_{ti}^*}-\frac{D_{ci}}{P_{si}^*}\right)\frac{P_{si}^*}{p_c^*}$.

For some convex linear combination between γ_t and γ_t^*, $\tilde{\gamma}_t$,

$$\begin{aligned}
R_{1n} &= \sup_{|\gamma_t-\gamma_t^*|\le\delta_n,\|e_t-e_t^*\|_\infty\le\delta_n} \frac{\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n \left(\frac{\partial}{\partial\gamma}e_i(\tilde{\gamma}_t)-\frac{\partial}{\partial\gamma}e_i^*(\gamma_t^*)\right)\left(\gamma_t-\gamma_t^*\right)\hat{\Upsilon}_i\right|}{1+C\sqrt{n}|\gamma_t-\gamma_t^*|} \\
&\le C\sup_{|\gamma_t-\gamma_t^*|\le\delta_n,\|e_t-e_t^*\|_\infty\le\delta_n} \frac{1}{n}\sum_{i=1}^n\left|\frac{\partial}{\partial\gamma}e_i(\gamma_t)-\frac{\partial}{\partial\gamma}e_i^*(\gamma_t)\right| \\
&\quad + C\sup_{|\gamma_t-\gamma_t^*|\le\delta_n}\left|\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial}{\partial\gamma}e_i(\gamma_t)-\frac{\partial}{\partial\gamma}e_i^*(\gamma_t)\right)\Upsilon_i\right| \\
&\quad + \frac{C}{n}\sum_{i=1}^n\sup_{|\gamma_t-\gamma_t^*|\le\delta_n}\left|\frac{\partial}{\partial\gamma}e_i^*(\gamma_t)\right|\left|\hat{\Upsilon}_i-\Upsilon_i\right| \\
R_{2n} &= \sup_{|\gamma_t-\gamma_t^*|\le\delta_n}\frac{\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\gamma}e_i^*(\gamma_t^*)(\gamma_t-\gamma_t^*)\hat{\Upsilon}_i\right|}{1+C\sqrt{n}|\gamma_t-\gamma_t^*|}.
\end{aligned}$$

$\square$

Proof of Lemma B.2 We verify the condition (b) for $\hat{\gamma}_t^{IPW}$ in Lemma B.1 by showing the followings are $o_p(1)$:

$$R_{1n} = \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\frac{D_{ti}m_i(\gamma_t^*)}{\hat{P}_{ti}}\frac{\hat{P}_{si}}{\hat{p}_c} - \frac{D_{ti}m_i(\gamma_t^*)}{P_{ti}^*}\frac{P_{si}^*}{p_s^*} - D_{ti}m_i(\gamma_t^*)\Lambda_i\Big\}\Big|$$

$$R_{2n} = \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\frac{D_{ti}m_i(\gamma_t^*)}{P_{ti}^*}\Lambda_i P_{ti}^* - e_i^*(\gamma_t^*)\Lambda_i P_{ti}^*\Big\}\Big|$$

$$R_{3n} = \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{e_i^*(\gamma_t^*)\Lambda_i P_{ti}^* + e_i^*(\gamma_t^*)\Big(\frac{D_{ti}}{P_{ti}^*} - \frac{D_{ci}}{P_{si}^*}\Big)\frac{P_{si}^*}{p_s^*}\Big\}\Big|$$

$$\leq \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{-e_i^*(\gamma_t^*)\frac{\hat{P}_{ti} - P_{ti}^*}{P_{ti}^*} + e_i^*(\gamma_t^*)\frac{D_{ti} - P_{ti}^*}{P_{ti}^*}\Big\}\frac{P_{si}^*}{p_s^*}\Big| \tag{1}$$

$$+\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{e_i^*(\gamma_t^*)\frac{\hat{P}_{si} - P_{si}^*}{P_{si}^*} - e_i^*(\gamma_t^*)\frac{D_{ci} - P_{si}^*}{P_{si}^*}\Big\}\frac{P_{si}^*}{p_s^*}\Big| \tag{2}$$

$$+\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i^*(\gamma_t^*)\frac{P_{si}^*}{p_s^*}\big(\hat{p}_c - p_s^*\big)\Big| \tag{3}$$

R_{1n}, R_{2n}, (1), and (2) are $o_p(1)$. (3) is $O_p(1)\times o_p(1) = o_p(1)$ because $\mathbb{E}[e_i^*(\gamma_t^*)P_{si}^*/p_s^*] = 0$. For the condition (d) for $\hat{\gamma}_t^{EIF}$ in Lemma B.1,

$$R_{4n} = \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{D_{ti}\big(m_i(\gamma_t^*) - e_i(\gamma_t^*)\big)\Lambda_i\Big\}\Big|$$

$$R_{5n} = \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{-\big(\hat{e}_i(\gamma_t^*) - e_i^*(\gamma_t^*)\big)\Big(\frac{D_{ti}}{P_{ti}^*} - \frac{D_{ci}}{P_{si}^*}\Big)\frac{P_{si}^*}{p_s^*}\Big\}\Big|$$

$$R_{6n} = \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}D_{ci}e_i(\gamma_t^*)\frac{\hat{p}_c - p_c^*}{p_c^{*2}}\Big|.$$

R_{4n} and R_{5n} are $o_p(1)$ following the same arguments as the case for $\hat{\beta}_t^{EIF}$. $R_{6n} = o_p(1)$ follows the same reasoning as (3) above. □

B.2 Proof of Theorem 2

We calculate the semiparametric efficiency bound for the parameter λ_t defined by

$$\int\int\int m(y;\lambda_t)f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(X)}{p_t}\ W_{X_1c}((x_1,x_2))\ dydx_1dx_2 = 0.$$

The pathwise derivative w.r.t. θ is

$$\int\int\int \frac{d}{d\theta}m(y;\lambda_t(\theta))f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(x)}{p_t}\ W_{X_1c}((x_1,x_2))\ dydx_1dx_2$$
$$+\int\int\int m(y;\lambda_t)\frac{d}{d\theta}\Big(f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(x)}{p_t}\Big)\ W_{X_1c}((x_1,x_2))\ dydx_1dx_2 \tag{4}$$
$$+\int\int\int m(y;\lambda_t)f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(x)}{p_t}\ \frac{d}{d\theta}W_{X_1c}((x_1,x_2))\ dydx_1dx_2 = 0. \tag{5}$$

The result for the decomposition parameter $\gamma_{t|t}$ is directly applied to (4) where the moment function is replaced by $m(y;\lambda_t)W_{X_1c}((x_1,x_2))$. That is, (4) contributes

$$\frac{P_t(X)}{p_t}\frac{D_t}{P_t(X)}m(Y;\lambda_t)W_{X_1c}((X_1,X_2)) = \frac{D_t}{P_t(X)}m(Y;\lambda_t)\frac{P_c(X)}{\mathbb{P}(T=c|X_2)}\frac{\mathbb{P}(T=t|X_2)}{p_t}.$$

For (5),

$$\mathbb{E}\left[e_t(X;\lambda_t)\ \frac{P_t(X)}{p_t}\ \frac{d}{d\theta}W_{X_1c}((X_1,X_2))\right]$$
$$=\mathbb{E}\left[e_t(X;\lambda_t)\ \frac{P_t(X)}{p_t}\ W_{X_1c}((X_1,X_2))\ \left(\frac{\dot{P}_c(X)}{P_c(X)}-\frac{\dot{p}_t(X)}{p_t(X)}+\frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}-\frac{\dot{\mathbb{P}}(T=c|X_2)}{\mathbb{P}(T=c|X_2)}\right)\right].$$

The proof of Theorem 2 in Lee (2018a) implies the first part containing $\dot{P}_t(X)$ and $\dot{P}_c(X)$ contributes

$$e_t(X;\lambda_t)\frac{P_t(X)}{p_t}\ W_{X_1c}((X_1,X_2))\left(\frac{D_c}{P_c(X)}-\frac{D_t}{P_t(X)}\right)$$

to the efficient influence function.

For the rest part containing $\dot{\mathbb{P}}(T=t|X_2)$ and $\dot{\mathbb{P}}(T=c|X_2)$, we define the score as

$$S(y,t,x;\theta_0) = S_y(y,t,x)+S_1(x_1,x_2,t)+S_{p2}(t,x_2)+S_{x_2}(x_2)$$

where $S_1(x_1,x_2,T)\equiv\sum_{j\in\mathcal{T}}D_js_{xj}(x_1,x_2)$, $s_{xj}(x_1,x_2)\equiv\frac{d}{d\theta}\log f_{X_1|X_2T}(x_1|x_2,j;\theta)\big|_{\theta_0}$, $S_{p2}(T,x_2)\equiv\sum_{j\in\mathcal{T}}D_j\dot{\mathbb{P}}(T=j|X_2=x_2)/\mathbb{P}(T=j|X_2=x_2)$, $\dot{\mathbb{P}}(T=j|X_2)\equiv\frac{d}{d\theta}\mathbb{P}(T=j|X_2;\theta)\big|_{\theta_0}$, and $S_{x_2}(x_2)\equiv\frac{d}{d\theta}\log f_{X_2}(x_2;\theta)\big|_{\theta_0}$. The tangent space is characterized $\mathcal{H}_y+\mathcal{H}_1+\mathcal{H}_{p2}+\mathcal{H}_{x_2}$, where $\mathcal{H}_1\equiv\{S_1(X_1,X_2,T): s_{xj}(X_1,X_2)\in L_0^2(F_{X_1|X_2T}(X_1|X_2,j)),\forall j\in\mathcal{T}\}$, $\mathcal{H}_{p2}\equiv\{S_{p2}(T,X_2): S_{p2}(T,X_2)\in L_0^2(F_{T|X_2})\}$, and $\mathcal{H}_{x_2}\equiv\{S_{x2}(X_2): S_{x2}(X_2)\in L_0^2(F_{X_2})\}$.

Similar to Equation (12) in Lee (2018a),

$$
\begin{aligned}
&\mathbb{E}\left[\frac{D_t - \mathbb{P}(T=t|X_2)}{\mathbb{P}(T=t|X_2)} S(Z;\theta_0)\middle| X_2\right] \\
&= \mathbb{E}\left[\frac{D_t - \mathbb{P}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}\left(\sum_{j\in\mathcal{T}} D_j s_j(Y,X) + D_j s_{xj}(X_1,X_2) + D_j \frac{\dot{\mathbb{P}}(T=j|X_2)}{\mathbb{P}(T=j|X_2)} + S_{x2}(X_2)\right)\middle| X_2\right] \\
&= \mathbb{E}\left[\frac{D_t}{\mathbb{P}(T=t|X_2)}\left(s_t(Y,X) + s_{xt}(X_1,X_2) + \frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}\right)\middle| X_2\right] \\
&\quad - \mathbb{E}\left[\sum_{j\in\mathcal{T}}\left(D_j s_j(Y,X) + D_j s_{xj}(X_1,X_2) + D_j\frac{\dot{\mathbb{P}}(T=j|X_2)}{\mathbb{P}(T=j|X_2)}\right)\middle| X_2\right] \\
&= \mathbb{E}[s_t(Y,X)|T=t,X_2] + \mathbb{E}[s_{xt}(X_1,X_2)|T=t,X_2] + \frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)} \\
&\quad - \sum_{j\in\mathcal{T}}\left(\mathbb{P}(T=j|X_2)\mathbb{E}[s_j(Y,X)|T=j,X] + \dot{\mathbb{P}}(T=j|X_2)\right) = \frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}
\end{aligned}
$$

by the low of iterated expectations, $\mathbb{E}[s_j(Y,X)|T=j,X]=0$, and $\mathbb{E}[s_{xj}(X_1,X_2)|T=j,X_2]=0$, $\forall j \in \mathcal{T}$. We first calculate

$$
\mathbb{E}[e_t(X;\lambda_t)P_c(X)|X_2] = \mathbb{E}\left[m(y;\lambda_t)\frac{P_c(X)}{P_t(X)}\middle| T=t, X_2\right]\mathbb{P}(T=t|X_2).
$$

Then by the law of iterated expectations,

$$
\begin{aligned}
&\mathbb{E}\left[e_t(X;\lambda_t)\frac{P_t(X)}{p_t}W_{X_1s}(X)\frac{\dot{\mathbb{P}}(T=c|X_2)}{\mathbb{P}(T=c|X_2)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e_t(X;\lambda_t)\frac{P_t(X)}{p_t}W_{X_1c}(X)\middle| X_2\right]\mathbb{E}\left[\frac{D_c - \mathbb{P}(T=c|X_2)}{\mathbb{P}(T=c|X_2)}S(Z;\theta_0)\middle| X_2\right]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{P}(T=t|X_2)}{p_t}\mathbb{E}\left[m(Y;\lambda_t)W_{X_1c}(X)\middle| T=t,X_2\right]\frac{D_c - \mathbb{P}(T=c|X_2)}{\mathbb{P}(T=c|X_2)}S(Z;\theta_0)\right].
\end{aligned}
$$

We obtain the main component of the efficient influence function[3]

$$
\begin{aligned}
&\psi_{X_1c}(Z;\lambda_t,p,e(\lambda_t)) \equiv \Bigg(\psi_s(Z;\lambda_t,p,e(\lambda_t))\frac{p_c}{\mathbb{P}(T=c|X_2)} \\
&+ \mathbb{E}\big[m(Y;\lambda_t)W_{X_1c}((X_1,X_2))\big|T=t,X_2\big]\left(\frac{D_t}{\mathbb{P}(T=t|X_2)} - \frac{D_c}{\mathbb{P}(T=c|X_2)}\right)\Bigg)\frac{\mathbb{P}(T=t|X_2)}{p_t}.
\end{aligned}
$$

[3]Alternatively we may calculate the efficient bound following Jacho-Chávez (2009) for the inverse conditional density-weighted functions.

$\square$

B.3 Proof of Theorem 3

We decompose the estimator as follows

$$\sqrt{n}(\hat{F}^{EIF}_{Y\langle t|c\rangle} - F_{Y\langle t|c\rangle}) = \sqrt{n}\big(\frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_2 - \varphi_2\big) = G_n\big[\hat{\varphi}_2 - \varphi_2\big] + G_n\big[\varphi_2\big] + \sqrt{n}\,\mathbb{E}\big[\hat{\varphi}_2 - \varphi_2\big]. \tag{6}$$

The second term $G_n[\varphi_2]$ is $\bar{O}_p(1)$ by the Donsker property in Assumption B.6 (a). The first term $G_n[\hat{\varphi}_2 - \varphi_2] = \bar{o}_p(1)$ by a stochastic equicontinuity argument using Lemma A.1 in Lee (2018b). Assumption B.6 (a) and (b) ensure the conditions of Lemma A.1 in Lee (2018b). That is, the estimators $\hat{P}_t(X)$ and $\hat{F}_{Y|TX}(y|t,X)$ belongs to $\mathcal{C}^{\alpha}_{M}(\mathcal{X})$ and satisfies the Hölder continuity with probability approaching to one. We calculate the third term in the following.

$$\begin{aligned}
\mathbb{E}\big[\hat{\varphi}_2 - \varphi_2\big] &= \mathbb{E}\Bigg[\Bigg(\frac{D_t}{\hat{P}_t(X)}\mathbf{1}\{Y\le y\} + \hat{F}_{Y|TX}(y|t,X)\,\Big(\frac{D_c}{\hat{P}_c(X)} - \frac{D_t}{\hat{P}_t(X)}\Big)\Bigg)\frac{\hat{P}_c(X)}{\hat{p}_c} - \varphi_2\,\Bigg]\\
&= \mathbb{E}\Bigg[\Bigg(\frac{P_t(X)}{\hat{P}_t(X)}F_{Y|TX}(y|t,X) + \hat{F}_{Y|TX}(y|t,X)\,\Big(\frac{P_c(X)}{\hat{P}_c(X)} - \frac{P_t(X)}{\hat{P}_t(X)}\Big)\Bigg)\frac{\hat{P}_c(X)}{\hat{p}_c} - F_{Y|TX}(y|t,X)\frac{P_c(X)}{p_c}\,\Bigg]\\
&= \mathbb{E}\Bigg[\Bigg(\frac{P_t(X)}{\hat{P}_t(X)}F_{Y|TX}(y|t,X) + F_{Y|TX}(y|t,X)\,\Big(\frac{P_c(X)}{\hat{P}_c(X)} - \frac{P_t(X)}{\hat{P}_t(X)}\Big)\Bigg)\frac{\hat{P}_c(X)}{\hat{p}_c} - F_{Y|TX}(y|t,X)\frac{P_c(X)}{p_c}\,\Bigg]\\
&\quad + \mathbb{E}\Bigg[\Big(\hat{F}_{Y|TX}(y|t,X) - F_{Y|TX}(y|t,X)\Big)\,\Big(\frac{P_c(X)}{\hat{P}_c(X)} - \frac{P_t(X)}{\hat{P}_t(X)}\Big)\frac{\hat{P}_c(X)}{\hat{p}_c}\Bigg]\\
&= \mathbb{E}\Bigg[F_{Y|TX}(y|t,X)\,\Big(\frac{P_c(X)}{\hat{p}_c} - \frac{P_c(X)}{p_c}\Big)\,\Bigg] + \bar{o}_p(n^{-1/2})\\
&= \mathbb{E}\Bigg[F_{Y|TX}(y|t,X)\frac{P_c(X)}{p_c}\Bigg]\Big(1 - \frac{1}{n}\sum_{i=1}^{n}D_{ci}/p_c\Big) + \bar{o}_p(n^{-1/2})
\end{aligned}$$

where the fourth equality is implied by Assumption B.6 (c)(i). Together with the second term $G_n[\varphi_2]$, $\sqrt{n}(\hat{F}^{EIF}_{Y\langle t|c\rangle} - F_{Y\langle t|c\rangle}) = G_n[\psi_{t|c}(Z;y)] + \bar{o}_p(1)$.

We decompose $\sqrt{n}(\hat{F}^{IPW}_{Y\langle t|c\rangle} - F_{Y\langle t|c\rangle})$ similarly as in (6). The third term

$$
\begin{aligned}
&\mathbb{E}\big[\hat{\varphi}_1 - \varphi_1\big] = \mathbb{E}\Bigg[\frac{D_t}{\hat{P}_t(X)}\mathbf{1}\{Y \le y\}\frac{\hat{P}_c(X)}{\hat{p}_c} - \varphi_1\Bigg] \\
&= \mathbb{E}\Bigg[\frac{P_t(X)}{\hat{P}_t(X)}F_{Y|TX}(y|t,X)\frac{\hat{P}_c(X)}{\hat{p}_c} - F_{Y|TX}(y|t,X)\frac{P_c(X)}{p_c}\Bigg] \\
&= \mathbb{E}\Bigg[F_{Y|TX}(y|t,X)\,\bigg(-\frac{P_c(X)}{p_c}\Big(\frac{\hat{p}_c}{p_c} - 1\Big) - \frac{P_c(X)}{p_c}\Big(\frac{\hat{P}_t(X)}{P_t(X)} - 1\Big) + \frac{\hat{P}_c(X) - P_c(X)}{p_c}\bigg)\Bigg] + \bar{o}_p(n^{-1/2}) \\
&= \gamma - \gamma\frac{1}{n}\sum_{i=1}^{n}\frac{D_{ci}}{p_c} + \mathbb{E}\Bigg[F_{Y|TX}(y|t,X)\frac{P_c(X)}{p_c}\Big(\frac{\hat{P}_c(X)}{P_c(X)} - \frac{\hat{P}_t(X)}{P_t(X)}\Big)\Bigg] + \bar{o}_p(n^{-1/2})
\end{aligned}
$$

The third term in the last equation is by Assumption B.6 (c)(ii). Together with the third term $G_n[\varphi_1]$, $\sqrt{n}(\hat{F}^{IPW}_{Y\langle t|c\rangle} - F_{Y\langle t|c\rangle}) = G_n[\psi_{t|c}(Z;y)] + \bar{o}_p(1)$.

Define the class of measurable functions $\mathcal{H} = \{(\mathcal{Y} \times \mathcal{T} \times \mathcal{X}) \to \psi(Y,T,X;y) : y \in \mathcal{Y}\}$. By Lemma A2 in Donald and Hsu (2014) and the Assumptions in the Appendix, $\mathcal{H}$ is P-Donsker. The weak convergence is implied by Donsker's Theorem in Section 2.8.2 in van der Vaart and Wellner (1996). □

B.4 Proof of Corollary 1

By the functional delta method (e.g., Theorem 3.9.4 in van der Vaart and Wellner (1996)) and the linearity of the Hadamard derivative, the weak convergence to a Gaussian process is implied. □

B.5 Proof of Corollary 2

The results follow Theorem 3 and the proof of Lemma 4.2 in Donald and Hsu (2014). In comparison with the influence function in Donald and Hsu (2014), $\psi_{t|c}(Z;y)$ in (11) contains the additional terms $F_{Y\langle t|c\rangle}(y)\left(\frac{D_c}{P_c(X)} - \frac{D_t}{P_t(X)}\right)\frac{P_c(X)}{p_c}$. Accounting for the estimation error of these additional terms is similar to the proof for showing that the process in (11) in the proof of Lemma 4.2 in Donald and Hsu (2014) weakly converges to a zero process conditional on the sample path $\{Z_1, Z_2, ..., Z_n\}$ with probability approaching one. This shows the validity of the multiplier bootstrap method in our setup. □

B.6 Proof of Corollary A.1

Let $\sqrt{n}\big(\hat{\theta}(\cdot)-\theta_0(\cdot)\big) = n^{-1/2}\sum_{i=1}^{n}\psi_{tin}(\cdot)+o_p(1) \Rightarrow \mathbb{G}_t(\cdot)$ from Theorem 3. Assume θ_t is continuously differentiable with strictly positive derivative $(\partial/\partial y)\theta_0(y)\big|_{y=Q_\tau} \equiv \theta_0'(Q_\tau)$. The Hadamard derivative is shown in Example 3.9.24 in van der Vaart and Wellner (1996). Then the influence function for estimating the quantile process is $\psi^{Q}_{t|c}(Z_i;\tau) \equiv -\psi_{tin}(Q_\tau)/\theta_0'(Q_\tau)$. The result follows. $\square$

C The Workforce Investment Act Programs

In this section we discuss in more detail the institutional background of the Workforce Investment Act (WIA) programs and some additional details on our estimation approach.

The Workforce Investment Act (WIA) was passed in 1998 to replace the Job Training Partnership Act (JTPA). Its main goal was to "...consolidate, coordinate, and improve employment, training, literacy, and vocational rehabilitation programs in the United States..." by reforming the previous workforce programs that had become "fragmented" and "uncoordinated." The Act established the largest network of public-financed career service programs in the United States. The WIA Adult program, WIA Dislocated Workers, and WIA Youth programs are the three flagship programs under this Act.

An individual is eligible for WIA Adult if he or she is age 18 and older who are unemployed at time of application or who are under-employed or whose family meets adult low income guidelines. Dislocated workers are officially defined by meeting one of the following criteria: (i) has been laid off or terminated, or received notice of termination or lay off and is unlikely to return to previous industry of occupation, (ii) has been terminated or laid off, or has received a notice of termination or lay off, as a result of permanent closure of, or substantial layoff at a plant or facility, (iii) was self-employed and now unemployed because of a natural disaster, (iv) was self-employed (including farmer, rancher, or fisherman), but is unemployed as a result of general economic conditions in the community in which he or she resides or because of a natural disaster, or (v) is a displaced homemaker.

WIA services are offered through 3,000 one-stop career centers across the country. There are three levels of services. *Core services* include labor market information, job search, and placement assistance. *Intensive services* include counseling, comprehensive assessments, and individual career planning. *Training services* connect participants to job opportunities in their local communities and provide training in both basic and vocational skills toward specific occupations. Participants use an "individual training account" to select an appropriate training program from a qualified training provider. Although participation in WIA is voluntary, access is restricted; program staffs must admit participants and authorize any services that are provided.

In 2013, the Workforce Innovation and Opportunity Act (WIOA) law was passed, replacing the previous WIA of 1998 as the primary federal workforce development legislation to bring about increased coordination among federal workforce development and related programs. While WIOA made changes to the workforce system, it did not significantly change the basic set of services that the local areas offered, nor who was eligible to receive them. In addition, many of the important changes explicitly introduced by WIOA reflected changes the local areas were already making under WIA.

C.1 Additional Details on the Estimation Approach

We define the *earnings outcome* to be the difference in average earnings between four quarters after leaving the program and the three quarters prior to entering. First, we estimate nonparametrically the probability of treatment for each individual i, given their characteristics, i.e., the propensity scores $P_t(X_i) \equiv \mathbb{P}(T = t|X = X_i)$ for $t \in \{1, 2, 3, 4\}$ and $i = 1, .., n$. We use a multinomial logistic series estimator, where the order of the polynomial is selected using the Akaike Information Criterion (AIC). Given the estimated propensity scores $\hat{P}_t(X_i)$, we compute the common support region for estimation following Flores et al. (2012): for each group-t, we find the minimum and maximum estimated propensity scores: $p_t^{min} \equiv \min_{\{i:T_i=t\}} \hat{P}_t(X_i)$ and $p_t^{max} \equiv \max_{\{i:T_i=t\}} \hat{P}_t(X_i)$. Then we define the *support region* for t to be the subpopulation whose $\hat{P}_t(X_i)$ bounded between p_t^{min} and p_t^{max}: $S_t \equiv \{i : \hat{P}_t(X_i) \in [p_t^{min}, p_t^{max}]\}$. The *common support region* is the intersection of the support regions for all $t \in \mathcal{T}$: $CS \equiv \cap_{t\in\mathcal{T}} S_t$. Observations that fall outside of the common support region are dropped.

Next, we estimate the means $\mathbb{E}[Y\langle t\rangle]$ and $\mathbb{E}[Y\langle t|c\rangle]$ using the EIF estimator:

$$\hat{\beta}_t : \hat{\mathbb{E}}[Y(t)] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{ti}}{\hat{P}_t(X_i)}Y_i - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - 1\Big)\hat{e}_t(X_i)\right)$$
$$\hat{\gamma}_{t|c} : \hat{\mathbb{E}}[Y(t)|T = c] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{ti}}{\hat{P}_t(X_i)}Y_i - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - \frac{D_{ci}}{\hat{P}_c(X_i)}\Big)\hat{e}_t(X_i)\right)\frac{\hat{P}_c(X_i)}{\hat{p}_c},$$

where $e_t(X_i) \equiv \mathbb{E}[Y|T = t, X = X_i]$ is also estimated using polynomial-regression series estimators with AIC to select the order of the polynomial. Finally, $p_t \equiv \mathbb{P}(T = t)$ is obtained from the sample analogue $\hat{p}_t = n^{-1}\sum_{i=1}^{n} D_{ti}$.

We estimate the τ-quantiles $Q_\tau(Y\langle t\rangle)$ and $Q_\tau(Y\langle t|c\rangle)$ with:

$$\hat{\beta}_t : \hat{Q}_\tau(Y(t)) = \arg\min_{q\in\Theta} \left| \frac{1}{n}\sum_{i=1}^{n} \left(\frac{D_{ti}}{\hat{P}_t(X_i)}(\mathbf{1}\{Y_i \leq q\} - \tau) - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - 1 \Big) \hat{e}_t(X_i; q) \right) \right|$$

$$\hat{\gamma}_{t|c} : \hat{Q}_\tau(Y(t)|T = c) = \arg\min_{q\in\Theta} \left| \frac{1}{n}\sum_{i=1}^{n} \left(\frac{D_{ti}}{\hat{P}_t(X_i)}(\mathbf{1}\{Y_i \leq q\} - \tau) - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - \frac{D_{ci}}{\hat{P}_c(X_i)} \Big) \hat{e}_t(X_i; q) \right) \frac{\hat{P}_c(X_i)}{\hat{p}_c} \right|,$$

where $\hat{e}_t(X_i; q) = \hat{\mathbb{E}}[\mathbf{1}\{Y \leq q\}|T = t, X = X_i] - \tau$.

References

Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics 137*(2), 674–707.

Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics 155*(2), 138–154.

Donald, S. G. and Y.-C. Hsu (2014). Estimation and inference for distribution functions and quantile functions in treatment effect models. *Journal of Econometrics 178*(3), 383–397.

Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann (2012). Estimating the effects of length of exposure to instruction in a training program: The case of job corps. *The Review of Economics and Statistics 94*(1), 153–171.

Jacho-Chávez, D. T. (2009). Efficiency bounds for semiparametric estimation of inverse conditional-density-weighted functions. *Econometric Theory 25*(3), 847–855.

Lee, Y.-Y. (2018a). Efficient propensity score regression estimators of multivalued treatment effects for the treated. *Journal of Econometrics 204*(2), 207–222.

Lee, Y.-Y. (2018b). Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models. *SSRN working paper No. 3250485*.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: with Application to Statistics*. New York: Springer-Verlag.