Quantile Treatment Effects in Difference in Differences Models with Panel Data

Brantly Callaway*

Tong Li†

Supplementary Appendix

September 24, 2018

This Supplementary Appendix contains additional results for "Quantile Treatment Effects in Difference in Differences Models with Panel Data." The first section contains Monte Carlo simulations. The second section contains additional asymptotic results for the case where the Distributional DID Assumption holds after conditioning on covariates as in Section 4 in the main text. The third section contains supplementary figures.

1 Monte Carlo Simulations

This section contains results from several Monte Carlo simulations. In particular, we consider the model from Example 1: $Y_{0s} = \theta_s + \eta + v_s$ for $s = t, t-1, t-2$. We compare the results using our approach to the results from the Change in Changes model (Athey and Imbens (2006)). Because our results for the QTT do not depend on modeling outcomes for the treated group, we assess the performance of each method on generating the distribution of counterfactual outcomes for the treated group in the last period, i.e. $F_{Y_{0t}|D=1}$. We report results for the 0.1, 0.5, and 0.9 quantiles.

Throughout, we assume $v_s \sim N(0,1)$ and are independent from η and independent of each other. We also set $(\eta|D=0) \sim N(0,1)$, $\theta_{t-1} = 1$, and $\theta_{t-2} = 0$ for each DGP below. The only remaining objects to set are the distribution of η for the treated group, and the value of θ_3. In this setup, the method proposed in the current paper should be valid (as discussed in Example 1) and the Change in Changes model should be valid. The possible issue with the Change in Changes method here is that it requires more restrictive support conditions than our approach. However, because everything is normally distributed here, the conditions for the Change in Changes model are satisfied though there may be performance issues in finite samples. We consider the following data generating processes.

DGP 1: $\eta|D=1 \sim N(1,1), \theta_t = 1$
DGP 2: $\eta|D=1 \sim N(2,1), \theta_t = 1$
DGP 3: $\eta|D=1 \sim N(4,1), \theta_t = 1$
DGP 4: $\eta|D=1 \sim N(1,1), \theta_t = 4$

*Department of Economics, Temple University, 1301 Cecil B. Moore Avenue, Ritter Annex 867, Philadelphia, PA 19122. Phone: (215) 204-8881, Email: brantly.callaway@temple.edu

†Department of Economics, Vanderbilt University, VU Station B #351819, Nashville, TN 37235-1819. Phone: (615) 322-3582, Fax: (615) 343-8495, Email: tong.li@vanderbilt.edu

Under each of these setups, the true counterfactual distribution is $N(\mu_1 + \theta_3, 2)$ where μ_1 is the mean of η for the treated group. Moving from DGP 1 to DGP 2 and DGP 3, the location of the distribution of the unobserved heterogeneity for the treated group moves further away from the location of the distribution of the unobserved heterogeneity for the untreated group. DGP 4 introduces a relatively large time effect that is common across both groups. We consider each DGP with n (the number of observations) equal to 100 and then $n = 1000$. We set the probability of being to treated to 0.5. We run each simulation 1000 times and compute the mean bias, median absolute deviation, and root mean squared error and report them in Table 1.

Results The performance of our method and the Change in Changes method is very similar for DGP 1. Our method performs somewhat better at the 0.9 quantile. And both methods perform better moving from 100 observations to 1000 observations.

Our method performs substantially better than the Change in Changes method in DGP 2 and DGP 3. To give an example, the root mean squared error of the Change in Changes method for the 0.9 quantile with 100 observations is six times larger than our method. The difference in performance between the two methods is essentially negligible at the 0.1 quantile, large at the 0.5 quantile, and extremely large at the 0.9 quantile. This is due to the distribution of η for the treated group being shifted to the right in these experiments and would be reversed if it were shifted to the left. Also, notice that this is a finite sample issue as the performance of the Change in Changes method improves going from 100 to 1000 observations (though in the latter case, our approach still performs substantially better).

Finally, in DGP 4 where the time effects are relatively large, there is little difference in the performance of the two methods. Overall, the biggest differences in performance between our method and the Change in Changes method occurred when the distribution of unobserved heterogeneity was quite different between the treated and untreated groups; in this case, our method performed substantially better. Importantly, this case is a leading case for using DID-type methods. A final takeaway is that the performance of our method did not vary much across different DGPs – in particular, the performance did not decline with large time effects or large differences in the distributions of time invariant unobserved heterogeneity between the treated and untreated groups.

2 Asymptotic Results when Distributional Difference in Differences Assumption holds conditional on covariates

This section develops the asymptotic properties of our estimator in the case where the Distributional Difference in Differences Assumption holds conditional on covariates. In this setup, the propensity score needs to be estimated. Throughout this section, we denote the true propensity score by $p_0(x)$ and the counterfactual distribution given in Theorem 2 (in the main text) by $F^p_{Y_{0t}|D=1}$ for notational clarity. We impose the following condition on the estimator of the propensity score.

Assumption SA1 (High Level Condition on the Propensity Score)**.**

$$\sup_{x \in \mathcal{X}} |\hat{p}(x) - p_0(x)| = o_p\left(n^{-1/4}\right)$$

Below, we provide primitive conditions for Assumption SA 1 to hold when the propensity score is estimated parametrically or nonparametrically, noting that other (e.g. semiparametric)

Table 1: Monte Carlo Simulations

	N=100			**N=1000**		
	10%	50%	90%	10%	50%	90%
DGP 1						
PQTT Bias	0.182	0.089	0.124	0.033	0.009	0.004
PQTT MAD	0.351	0.234	0.319	0.106	0.068	0.096
PQTT RMSE	0.480	0.347	0.499	0.148	0.103	0.142
CIC Bias	−0.044	−0.063	−0.226	−0.002	−0.008	−0.011
CIC MAD	0.308	0.278	0.502	0.097	0.085	0.171
CIC RMSE	0.481	0.413	0.687	0.148	0.127	0.263
DGP 2						
PQTT Bias	0.172	0.082	0.087	0.037	0.015	0.005
PQTT MAD	0.337	0.228	0.284	0.097	0.071	0.098
PQTT RMSE	0.480	0.330	0.458	0.149	0.104	0.143
CIC Bias	−0.066	−0.060	−0.737	0.002	−0.014	−0.071
CIC MAD	0.309	0.341	0.858	0.092	0.114	0.304
CIC RMSE	0.465	0.533	0.998	0.142	0.163	0.462
DGP 3						
PQTT Bias	0.229	0.118	0.094	0.031	0.011	0.010
PQTT MAD	0.367	0.235	0.299	0.097	0.068	0.096
PQTT RMSE	0.503	0.345	0.472	0.150	0.103	0.147
CIC Bias	−0.073	−0.909	−2.637	−0.011	−0.106	−1.526
CIC MAD	0.394	0.992	2.689	0.132	0.363	1.593
CIC RMSE	0.606	1.135	2.727	0.194	0.506	1.618
DGP 4						
PQTT Bias	0.195	0.094	0.039	0.028	0.016	0.005
PQTT MAD	0.320	0.229	0.312	0.099	0.070	0.100
PQTT RMSE	0.463	0.338	0.474	0.145	0.105	0.145
CIC Bias	−0.060	−0.061	−0.238	−0.005	−0.004	−0.012
CIC MAD	0.313	0.296	0.502	0.099	0.093	0.170
CIC RMSE	0.480	0.428	0.678	0.146	0.133	0.257

Notes: This table presents results from Monte Carlo experiments under DGP 1 - DGP 4. The rows labeled by PQTT provide the results developed using the approach of the current paper. The results labeled CIC come from the Change in Changes method. The columns are for the 0.1 quantile, 0.5 quantile, and 0.9 quantile. We use 1000 Monte Carlo simulations.

estimators can satisfy this condition as well. Before that, we state some preliminary results used for deriving the asymptotic distribution of our estimator.

Lemma SA1. Let $F^p_{\Delta Y_{0t}|D=1}(\delta,\bar{p}) = E\left[\frac{1-D}{p}\frac{\bar{p}(X)}{1-\bar{p}(X)}\mathbb{1}\{\Delta Y_t \leq \delta\}\right]$ denote the propensity score reweighted distribution of the change in untreated potential outcomes for the treated group for a particular propensity score $\bar{p}$. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3 and 4.1 and Assumption SA 1, the pathwise derivative $\Gamma(p_0)(\hat{p}-p_0)$ exists and is given by

$$\Gamma(\delta,p_0)(\hat{p}-p_0) = E\left[\frac{1-D}{p}\frac{\mathbb{1}\{\Delta Y_t \leq \delta\}}{(1-p_0(X))^2}(\hat{p}(X)-p_0(X))\right]$$

Proof.

$$\begin{aligned}
&\frac{F^p_{\Delta Y_{0t}|D=1}(\delta,p_0+h(\bar{p}-p_0)) - F^p_{\Delta Y_{0t}|D=1}(\delta,p_0)}{h} \\
&\quad = E\left[\frac{1-D}{p}\mathbb{1}\{\Delta Y_t \leq \delta\}\left(\frac{p_0(X)+h(\bar{p}(X)-p_0(X))}{1-p_0(X)-h(\bar{p}(X)-p_0(X))} - \frac{p_0(X)}{1-p_0(X)}\right)\right]\Big/ h \\
&\quad = E\left[\frac{1-D}{p}\mathbb{1}\{\Delta Y_t \leq \delta\}\frac{(\bar{p}(X)-p_0(X))}{(1-p_0(X))^2 - h(\bar{p}(X)-p_0(X)) + p_0(X)h(\bar{p}(X)-p_0(X))}\right] \\
&\quad \to E\left[\frac{1-D}{p}\mathbb{1}\{\Delta Y_t \leq \delta\}\frac{(\bar{p}(X)-p_0(X))}{(1-p_0(X))^2}\right] \qquad \text{as } h \to 0
\end{aligned}$$

□

Lemma SA2. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3 and 4.1 and Assumption SA 1,

$$\sqrt{n}|F^p_{\Delta Y_{0t}|D=1}(\delta,\hat{p}) - F^p_{\Delta Y_{0t}|D=1}(\delta,p_0) - \Gamma(\delta,p_0)(\hat{p}-p_0)|_\infty = o_p(1)$$

Proof.

$$\begin{aligned}
&\sqrt{n}|F_{\Delta Y_{0t}|D=1}(\delta,\hat{p}) - F_{\Delta Y_{0t}|D=1}(\delta,p_0) - \Gamma(\delta,p_0)(\hat{p}-p_0)|_\infty \\
&\quad = \sqrt{n}\left|E\left[\frac{1-D}{p}\mathbb{1}\{\Delta Y_t \leq \delta\}\left(\frac{\hat{p}(X)}{1-\hat{p}(X)} - \frac{p_0(X)}{1-p_0(X)} - \frac{(\hat{p}(X)-p_0(X)}{(1-p_0(X))^2}\right)\right]\right|_\infty \\
&\quad = \sqrt{n}\left|E\left[\frac{1-D}{p}\mathbb{1}\{\Delta Y_t \leq \delta\}\left(\frac{(\hat{p}(X)-p_0(X))^2}{(1-\hat{p}(X))(1-p_0(X))^2}\right)\right]\right|_\infty \\
&\quad \leq C\sqrt{n}\sup_{x\in\mathcal{X}}|\hat{p}(x)-p_0(x)|^2 \to 0
\end{aligned}$$

where the last line holds because p is bounded away from 0 and 1, $p_0(x)$ is uniformly bounded away from 1, and $\hat{p}(x)$ converges uniformly to $p_0(x)$. Then, the result holds under Assumption SA 1. □

Next, we state a high level condition on the pathwise derivative given above. Let $W = (D, X, \Delta Y_t)$ and $W_i = (D_i, X_i, \Delta Y_{it})$.

Assumption SA2 (High Level Conditions on Pathwise Derivative)**.**
(i) $\sqrt{n}(\Gamma(\delta,p_0)(\hat{p}-p_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi_{\delta,p_0}(W_i) + o_p(1)$ uniformly in $\delta \in \Delta\mathcal{Y}_{0t|D=1}$.
(ii) $\{\varphi_{\delta,p_0}(W) : \delta \in \Delta\mathcal{Y}_{0t|D=1}\}$ is a uniformly bounded Donsker class of functions.

We will show the validity of both of these conditions in the case where the propensity score is estimated parametrically and nonparametrically using series logit below and note that these conditions potentially allow other estimators of the propensity score as well. The next result establishes an equivalent result for Proposition 2 (in the main text) in the case where the Distributional DID Assumption holds conditional on covariates.

Proposition SA 1. *Let $\hat{G}^p_{\Delta Y_{0t}|D=1}(\delta) = \sqrt{n}\left(\hat{F}^p_{\Delta Y_{0t}|D=1}(\delta) - F^p_{\Delta Y_{0t}|D=1}(\delta)\right)$ where $F^p_{\Delta Y_{0t}|D=1}(\delta)$ is given in Equation (5) in the main text. Let $\tilde{Y}^p_{it} = F^{p,-1}_{\Delta Y_{0t}|D=1}(F_{\Delta Y_{t-1}|D=1}(\Delta Y_{it-1})) + F^{-1}_{Y_{t-1}|D=1}(F_{Y_{t-2}|D=1}(Y_{it-2}))$, let $\tilde{F}^p_{Y_{0t}|D=1}(y) = \frac{1}{n_D}\sum_{i\in\mathcal{T}} \mathbb{1}\{\tilde{Y}^p_{it} \leq y\}$, and let $\tilde{G}^p_{Y_{0t}|D=1}(y) = \sqrt{n}\left(\tilde{F}^p_{Y_{0t}|D=1}(y) - F^p_{Y_{0t}|D=1}(y)\right)$. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3 and 4.1 and Assumptions SA 1 and 2*

$$(\hat{G}^p_{\Delta Y_{0t}|D=1}, \hat{G}_{\Delta Y_{t-1}|D=1}, \tilde{G}^p_{Y_{0t}|D=1}, \hat{G}_{Y_t|D=1}, \hat{G}_{Y_{t-1}|D=1}, \hat{G}_{Y_{t-2}|D=1}) \rightsquigarrow (\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$$

in the space $\mathcal{S} = l^\infty(\Delta\mathcal{Y}_{0t|D=1}) \times l^\infty(\Delta\mathcal{Y}_{t-1|D=1}) \times l^\infty(\mathcal{Y}_{0t|D=1}) \times l^\infty(\mathcal{Y}_{t|D=1}) \times l^\infty(\mathcal{Y}_{t-1|D=1}) \times l^\infty(\mathcal{Y}_{t-2|D=1})$ where $(\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$ is a tight Gaussian process with mean 0 and covariance $V(y, y') = E[\psi^p(y)\psi^p(y')']$ for $y = (y_1, y_2, y_3, y_4, y_5, y_6) \in \mathcal{S}$ and $y' = (y'_1, y'_2, y'_3, y'_4, y'_5, y'_6) \in \mathcal{S}$ and with $\psi^p(y)$ given by

$$\psi^p(y) = \begin{pmatrix} \varphi_{y_1,p_0}(W) + \frac{1-D}{p}\frac{p_0(X)}{1-p_0(X)}\mathbb{1}\{\Delta Y_t \leq y_1\} - F^p_{\Delta Y_{0t}|D=1}(y_1) \\ \frac{D}{p}\mathbb{1}\{\Delta Y_{t-1} \leq y_2\} - F_{\Delta Y_{t-1}|D=1}(y_2) \\ \frac{D}{p}\mathbb{1}\{\tilde{Y}^p_t \leq y_3\} - F^p_{Y_{0t}|D=1}(y_3) \\ \frac{D}{p}\mathbb{1}\{Y_t \leq y_4\} - F_{Y_t|D=1}(y_4) \\ \frac{D}{p}\mathbb{1}\{Y_{t-1} \leq y_5\} - F_{Y_{t-1}|D=1}(y_5) \\ \frac{D}{p}\mathbb{1}\{Y_{t-2} \leq y_6\} - F_{Y_{t-2}|D=1}(y_6) \end{pmatrix}$$

Proof. First, we claim that the class of functions $\{(d, x, \Delta y_t, y_1) \mapsto \varphi_{y_1,p_0}(w) + \frac{1-d}{p}\frac{p_0(x)}{1-p_0(X)}\mathbb{1}\{\delta y_t \leq y_1\} | y_1 \in \Delta Y_{0t|D=1}\}$ is Donsker. To show this, first notice that, $\varphi_{y_1,p_0}(w)$ is Donsker by Assumption SA 2(ii). Next, let $\mathcal{K} = \left\{\frac{1-d}{p}\frac{p_0(x)}{1-p_0(x)}\mathbb{1}\{\Delta y_t \leq y_1\} | y_1 \in \Delta\mathcal{Y}_{0t|D=1}\right\}$. $\mathcal{K}$ is Donsker because $\mathbb{1}\{\Delta Y_t \leq y_1\} | y_1 \in \Delta\mathcal{Y}_{0t|D=1}\}$ is Donsker, and $\frac{1-d}{p}\frac{p_0(x)}{1-p_0(x)}$ is a uniformly bounded and measurable function so that we can apply Van der Vaart and Wellner (1996, Example 2.10.10). Then, the result holds by Van der Vaart and Wellner (1996, Example 2.10.7).

Finally, the result follows from Lemma SA 2 and from the functional central limit theorem for empirical distribution functions. □

The next result establishes an analogous result to Proposition 3 for the case where identification depends on covariates.

Proposition SA 2. *Let $\hat{G}^p_0(y) = \sqrt{n}(\hat{F}^p_{Y_{0t}|D=1}(y) - F^p_{Y_{0t}|D=1}(y))$ and let $\hat{G}^p_1(y) = \sqrt{n}(\hat{F}_{Y_t|D=1}(y) - F_{Y_t|D=1}(y))$. Under the Conditional Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 3.2, 3.3 and 4.1 and Assumptions SA 1 and 2*

$$(\hat{G}^p_0, \hat{G}^p_1) \rightsquigarrow (\mathbb{G}^p_0, \mathbb{G}^p_1)$$

where $\mathbb{G}^p_0$ and $\mathbb{G}^p_1$ are tight Gaussian processes with mean 0 with almost surely uniformly continuous paths on the space $\mathcal{Y}_{0t|D=1} \times \mathcal{Y}_{t|D=1}$ given by

$$\mathbb{G}^p_1 = \mathbb{V}^p_1$$

and

$$\mathbb{G}_0^p = \mathbb{V}_0^p + \int \Big\{ \mathbb{W}_1^p \circ K_2(y,v) - f^p_{\Delta Y_{0t}|D=1}\left(y - F^{-1}_{Y_{t-1}|D=1} \circ F_{Y_{t-2}|D=1} \circ \frac{\mathbb{W}_4^p - \mathbb{W}_3^p \circ K_1(v)}{f_{Y_{t-1}|D=1} \circ K_1(v)} \right)$$

$$- \mathbb{W}_2^p \circ K_3(y,v) \Big\} \times \frac{f_{\Delta Y_{t-1}|Y_{t-2},D=1}(K_3(y,v)|v)}{f_{\Delta Y_{t-1}|D=1}(K_3(y,v))}\, dF_{Y_{t-2}|D=1}(v)$$

where $K_1(v)$, $K_2(y,v)$ are defined in Proposition 3 in the main text and $K_3(y,v) := F^{-1}_{\Delta Y_{t-1}|D=1} \circ F^p_{\Delta Y_{0t}|D=1}(K_2(y,v))$.

Proof. The result follows immediately from Proposition SA 1, by the Hadamard differentiability of the map ϕ established in Lemma C.4 in the main text, and similar arguments as in the proof of Proposition 3 in the main text. □

Theorem SA 1. *Suppose $F^p_{Y_{0t}|D=1}$ admits a positive continuous density $f^p_{Y_{0t}|D=1}$ on an interval $[a,b]$ containing an ε-enlargement of the set $\{F^{p,-1}_{Y_{0t}|D=1}(\tau) : \tau \in \mathcal{T}\}$. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, and Assumptions 3.2, 3.3 and 4.1 and Assumptions SA 1 and 2*

$$\sqrt{n}(\widehat{QTT}^p(\tau) - QTT^p(\tau)) \rightsquigarrow \bar{\mathbb{G}}_1^p(\tau) - \bar{\mathbb{G}}_0^p(\tau)$$

where $(\bar{\mathbb{G}}_0^p(\tau), \bar{\mathbb{G}}_0^p(\tau))$ is a stochastic process in the metric space $(l^\infty(\mathcal{T}))^2$ with

$$\bar{\mathbb{G}}_0^p(\tau) = \frac{\mathbb{G}_0^p(F^{p,-1}_{Y_{0t}|D=1}(\tau))}{f^p_{Y_{0t}|D=1}(F^{p,-1}_{Y_{0t}|D=1}(\tau))} \quad \text{and} \quad \bar{\mathbb{G}}_1^p(\tau) = \frac{\mathbb{G}_1^p(F^{-1}_{Y_t|D=1}(\tau))}{f_{Y_t|D=1}(F^{-1}_{Y_t|D=1}(\tau))}$$

Proof. The result follows from the Hadamard differentiability of the quantile map (Van der Vaart and Wellner (1996, Lemma 3.9.23(ii)) and by Proposition SA 2. □

Primitive Conditions for the Propensity Score

The next set of assumptions guarantees that Assumption SA 1 holds under parametric and nonparametric models for the propensity score; we also give explicit expressions for the pathwise derivative in Assumption SA 2 for parametric and nonparametric models for the propensity score.

Assumption P1 (Parametric Model)**.**

(i) $p_0(x) = \Lambda(x'\beta)$ for a known function $\Lambda : \mathbb{R} \to [0,1]$ with $\beta \in \text{int}(\mathcal{B})$ where $\mathcal{B}$ is a compact subset of $\mathbb{R}^k$.

(ii) Let $\mathcal{U} = \{x'\beta : x \in \mathcal{X}, \beta \in \mathcal{B}\}$. For all $u \in \mathcal{U}$, there exists an $\epsilon > 0$ such that $\Lambda(u) \in [\epsilon, 1-\epsilon]$.

(iii) $\Lambda(u)$ is strictly increasing and twice continuously differentiable with first derivatives bounded away from 0 and infinity and bounded second derivatives.

Assumption P2 (Distribution of X)**.**

(i) The support $\mathcal{X}$ of X is a subset of a compact set.

(ii) $E[XX']$ is positive definite.

These are standard conditions for parametrically estimating the propensity score and will hold, for example, in logit and probit models. Under Assumption P 1 and Assumption P 2, β can be estimated using maximum likelihood, and it is straightforward to show that

$$\sqrt{n}(\hat{\beta}-\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\left[\frac{\lambda(X'\beta)^2}{\Lambda(X'\beta)(1-\Lambda(X'\beta))}XX'\right]^{-1}\frac{(D_i-\Lambda(X_i'\beta))\lambda(X_i'\beta)}{\Lambda(X_i'\beta)(1-\Lambda(X_i'\beta))}X_i + o_p(1)$$

and

$$\sqrt{n}(\hat{p}(x)-p_0(x)) = \lambda(x'\beta)x'\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\left[\frac{\lambda(X'\beta)^2}{\Lambda(X'\beta)(1-\Lambda(X'\beta))}XX'\right]^{-1}\frac{(D_i-\Lambda(X_i'\beta))\lambda(X_i'\beta)}{\Lambda(X_i'\beta)(1-\Lambda(X_i'\beta))}X_i + o_p(1)$$

where $\lambda(u)$ is the derivative of $\Lambda(u)$. This implies Assumption SA 1 holds when the propensity score is estimated parametrically. Also, it immediately follows that in the case where the propensity score is estimated parametrically,

$$\Gamma(\delta, p_0)(\hat{p}-p_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi^p_{\delta,p_0}(W_i) + o_p(1)$$

with

$$\varphi^p_{\delta,p_0}(W_i) = E\left[\frac{(1-D)}{p}\frac{\mathbb{1}\{\Delta Y_t \leq \delta\}}{(1-p_0(X))^2}\lambda(X'\beta)X'\right]E\left[\frac{\lambda(X'\beta)^2}{\Lambda(X'\beta)(1-\Lambda(X'\beta))}XX'\right]^{-1}\frac{(D_i-\Lambda(X_i'\beta))\lambda(X_i'\beta)}{\Lambda(X_i'\beta)(1-\Lambda(X_i'\beta))}X_i$$

The second part of Assumption SA 2 also follows immediately.

Next, we consider the case where the propensity score is estimated nonparametrically by using series logit methods. Following Hirano, Imbens, and Ridder (2003), we make the following assumptions on the propensity score

Assumption NP1 (Differentiability of Conditional Expectation)**.** $E[\mathbb{1}\{\Delta Y_{0t} \leq y\}|X, D=0]$ is continuously differentiable for all $x \in \mathcal{X}$.

Assumption NP2. (Distribution of X)

(i) The support $\mathcal{X}$ of X is a Cartesian product of compact intervals; that is, $\mathcal{X} = \prod_{j=1}^{r}[x_{lj}, x_{uj}]$ where r is the dimension of X and x_{lj} and x_{uj} are the smallest and largest values in the support of the j-th dimension of X.

(ii) The density of X, $f_X(\cdot)$, is bounded away from 0 on $\mathcal{X}$.

Assumption NP3. (Assumptions on the propensity score)

(i) $p_0(x)$ is continuously differentiable of order $s \geq 7r$ where r is the dimension of X.

(ii) There exist $\underline{p}$ and $\bar{p}$ such that $0 < \underline{p} \leq p_0(x) \leq \bar{p} < 1$.

Assumption NP4. (Series Logit Estimator)

For nonparametric estimation of the propensity score, $p_0(x)$ is estimated by series logit where the power series of the order $K = n^{\nu}$ for some $\frac{1}{4(s/r-1)} < \nu < \frac{1}{9}$.

Assumptions NP 1 to 4 are standard assumptions in the literature which depends on first step estimation of the propensity score. Hirano, Imbens, and Ridder (2003) developed the properties of the series logit estimator under the same set of assumptions. Similar assumptions have been used in, for example, Firpo (2007) and Donald and Hsu (2014). Assumption NP 2 says that X is continuously distributed though our setup can easily handle discrete covariates as well by splitting the sample based on the discrete covariates. Assumption NP 3(i) is a standard assumption on differentiability of the propensity score and guarantees the existence of ν that satisfies the conditions of Assumption NP 4. Assumption NP 3(ii) is a standard overlap condition.

Let

$$\varphi^{np}_{\delta,p_0}(W) = \frac{\mathbb{1}\{\Delta Y \leq \delta | X\}}{p(1-p_0(X))}(D - p_0(X))$$

Using arguments similar to Hirano, Imbens, and Ridder (2003) and Donald and Hsu (2014), we can show that Assumption SA 1 holds under the above assumptions; we can also show that

$$\Gamma(\delta, p_0)(\hat{p} - p_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi^{np}_{\delta,p_0}(W_i) + o_p(1)$$

which corresponds to Assumption SA 2(i). Finally, to show that Assumption SA 2(ii) holds, we note that the class of functions $\{(d, x, \Delta y_t, \delta) \mapsto \frac{\mathbb{1}\{\Delta y_t \leq \delta | x\}}{p(1-p_0(x))}(d - p_0(x)) | \delta \in \Delta\mathcal{Y}_{0t|D=1}\}$ is Donsker which holds by Donald and Hsu (2014, Lemma A.2). This implies that our results will hold in the case where the propensity score is estimated nonparametrically using series logit.

Results for the Bootstrap

Finally, we show that the empirical bootstrap can be used to construct asymptotically valid confidence bands. The steps for the bootstrap are the same as for the case without covariates – only the $F_{\Delta Y_{0t}|D=1}(\delta)$ should be calculated using the result on re-weighting rather than replacing it directly with $F_{\Delta Y_t|D=0}(\delta)$. Below, terms indexed by * are estimated using a bootstrapped sample. The same series terms used to estimate the propensity score can be reused in each bootstrap iteration. Our results in this section essentially follow using the same arguments as in Chen, Linton, and Van Keilegom (2003) and Ferreira, Firpo, and Galvao (2017).

Lemma SA3. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3 and 4.1 and Assumptions SA 1 and 2,

$$\begin{aligned}\sqrt{n}\left(\hat{F}^{p,*}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*) - \hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p})\right) \\ = \sqrt{n}\left(\hat{F}^{p,*}_{\Delta Y_{0t}|D=1}(\delta;p_0) - \hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;p_0) + \Gamma(\delta,\hat{p})(\hat{p}^* - \hat{p})\right) + o_{p^*}(1)\end{aligned}$$

uniformly in $\delta \in \Delta\mathcal{Y}_{0t|D=1}$.

Proof. Uniformly in $\delta \in \Delta\mathcal{Y}_{0t|D=1}$ and by adding and subtracting terms,

$$\begin{aligned}
\sqrt{n}&\left(\hat{F}^{p,*}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*) - \hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p})\right) \\
&= \sqrt{n}\left\{\left(\hat{F}^{p,*}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*) - \hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*)\right)\right. \\
&\qquad\qquad \left. - \sqrt{n}\left(\hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;p_0) - \hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;p_0)\right)\right\} \\
&\quad + \sqrt{n}\left(\hat{F}^{p,*}_{\Delta Y_{0t}|D=1}(\delta;p_0) - \hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;p_0) + \Gamma(\delta,\hat{p})(\hat{p}^* - \hat{p})\right) \\
&\quad + \sqrt{n}\left\{\left(\hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*) - F^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*)\right)\right. \\
&\qquad\qquad \left. - \sqrt{n}\left(\hat{F}^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}) - F^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p})\right)\right\} \\
&\quad + \sqrt{n}\left(F^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}^*) - F^{p}_{\Delta Y_{0t}|D=1}(\delta;\hat{p}) - \Gamma(\delta,\hat{p})(\hat{p}^* - \hat{p})\right)
\end{aligned}$$

The first, third, and fourth terms in the first equality converge uniformly to 0. These hold by Assumptions SA 1 and 2 and by arguments similar to those in Lemma SA 2. This implies the result. ☐

Lemma SA4. For some random variable X, let $\hat{G}^*_X(x) = \sqrt{n}\left(\hat{F}^*_X(x) - \hat{F}_X(x)\right)$ and let $\tilde{G}^{p,*}_{Y_{0t}|D=1}(\delta) = \sqrt{n}\left(\hat{F}^{p,*}_{Y_{0t}|D=1}(\delta) - \hat{F}^{p}_{Y_{0t}|D=1}(\delta)\right)$. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3 and 4.1 and Assumptions SA 1 and 2.

$$(\hat{G}^{p,*}_{\Delta Y_{0t}|D=1}, \hat{G}^{*}_{\Delta Y_{t-1}|D=1}, \tilde{G}^{p,*}_{Y_{0t}|D=1}, \hat{G}^{*}_{Y_{t}|D=1}, \hat{G}^{*}_{Y_{t-1}|D=1}, \hat{G}^{*}_{Y_{t-2}|D=1}) \rightsquigarrow_* (\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$$

where $(\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$ is the tight Gaussian process given in Proposition SA 1.

Proof. The result follows from Lemma SA 3 and by Van der Vaart and Wellner (1996, Theorem 3.6.1). ☐

Theorem SA 2. *Under the Conditional Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 3.2, 3.3 and 4.1 and Assumptions SA 1 and 2,*

$$\sqrt{n}\left(\widehat{QTT}^{p*}(\tau) - \widehat{QTT}^{p}(\tau)\right) \rightsquigarrow_* \bar{\mathbb{G}}^p_1(\tau) - \bar{\mathbb{G}}^p_0(\tau)$$

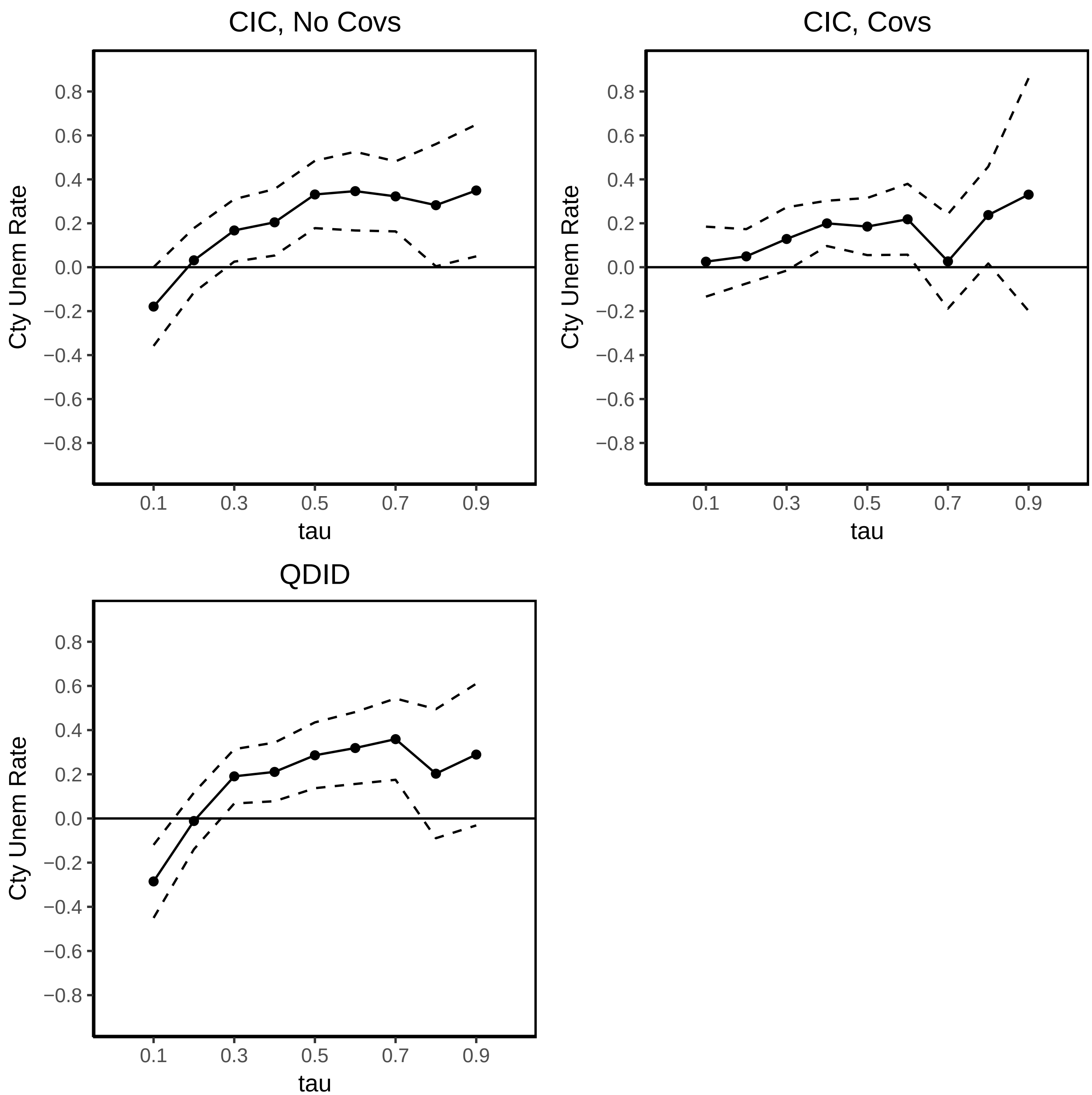
where $(\bar{\mathbb{G}}^p_0, \bar{\mathbb{G}}^p_1)$ *are as in Theorem SA 1.*

Proof. The result holds by Lemma SA 4, by the Hadamard Differentiability of our estimator of the QTT, and by the Delta method for the bootstrap (Van der Vaart and Wellner (1996, Theorem 3.9.11)). ☐

3 Supplementary Figures

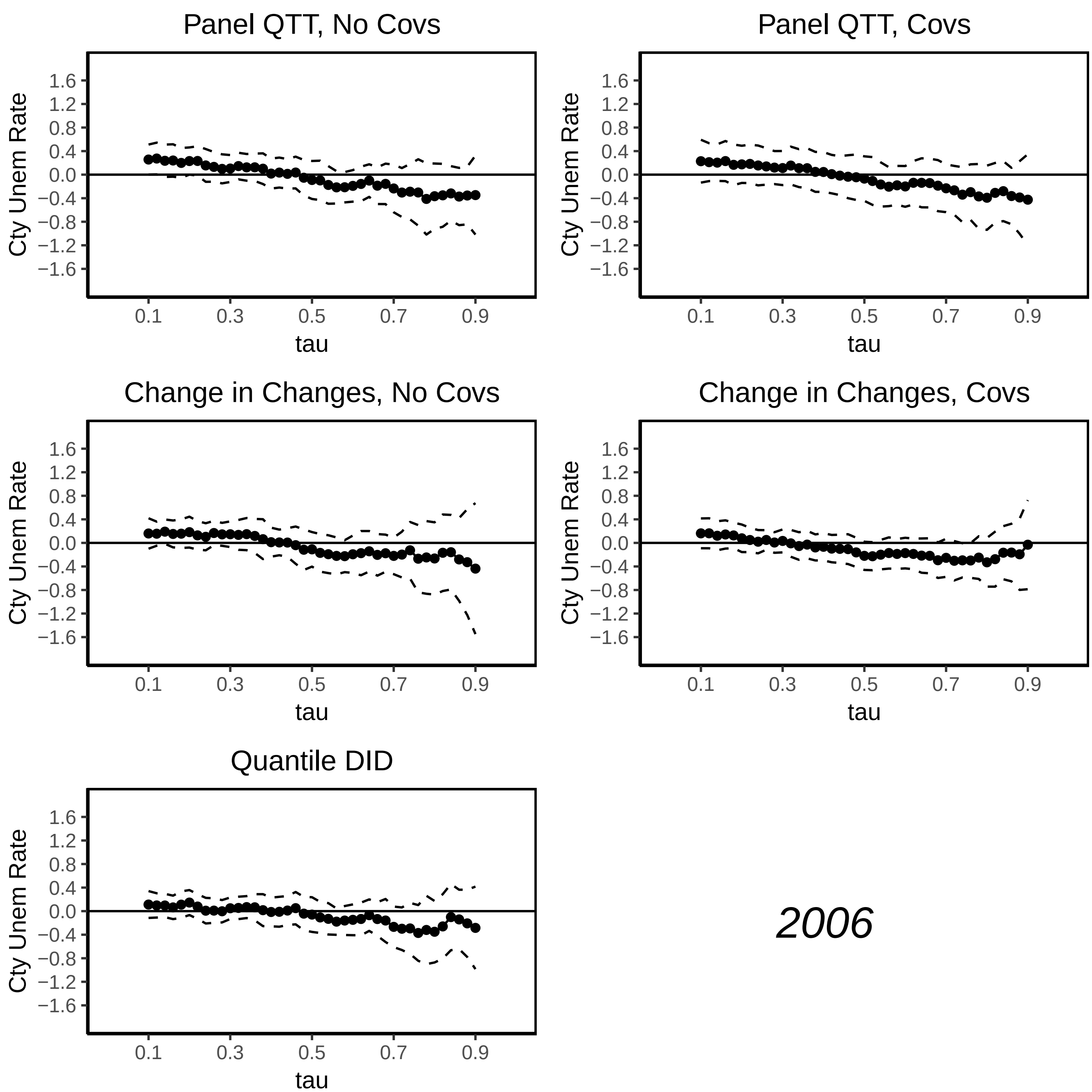
Change in Changes and Quantile Difference in Differences Estimates for 2007

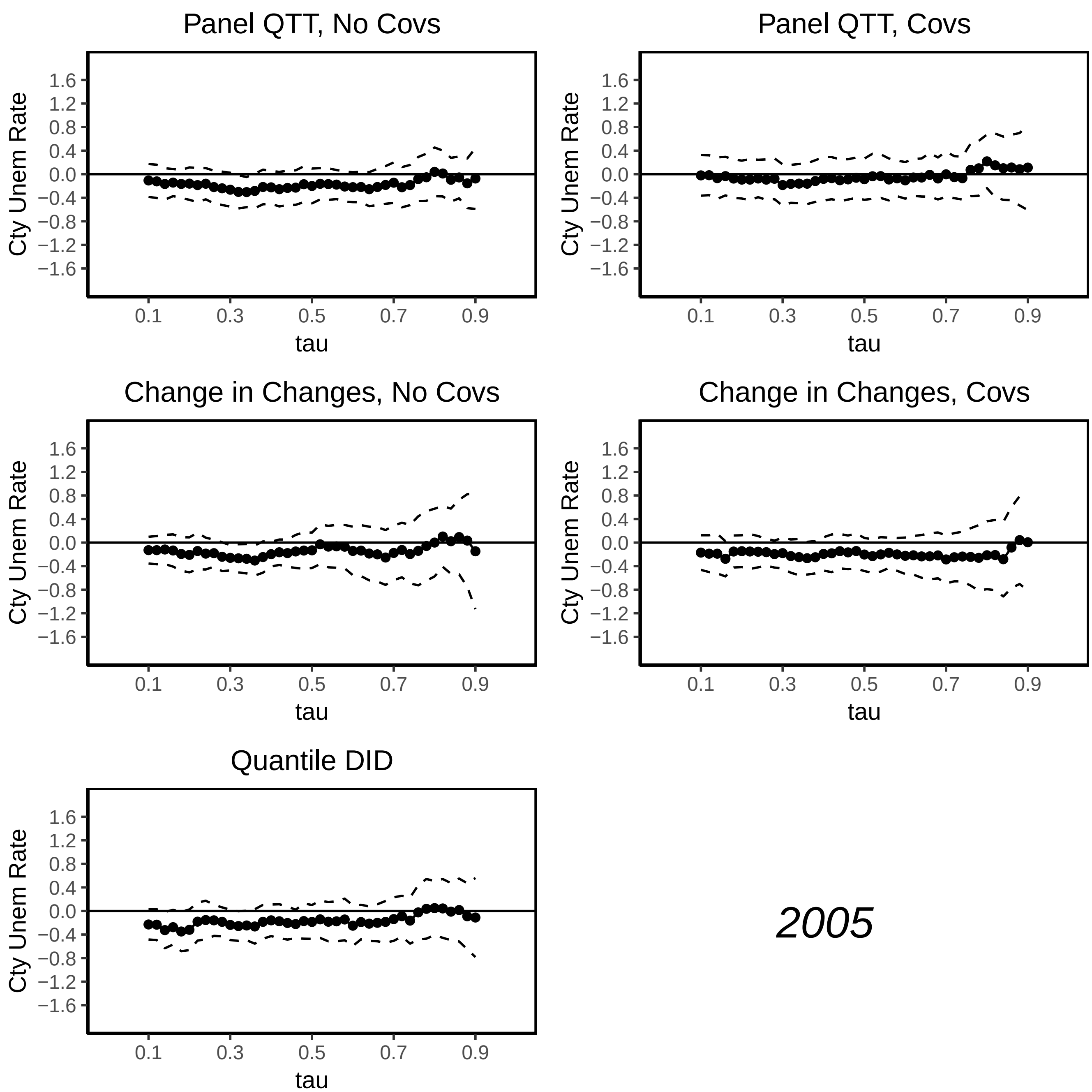
The plots in this section are for 2007 using alternative methods. The figure computes the QTT using the Change in Changes method (both with and without covariates) and the Quantile Difference in Differences method. Here, we report pointwise 95% confidence intervals computed using 1000 bootstrap iterations.

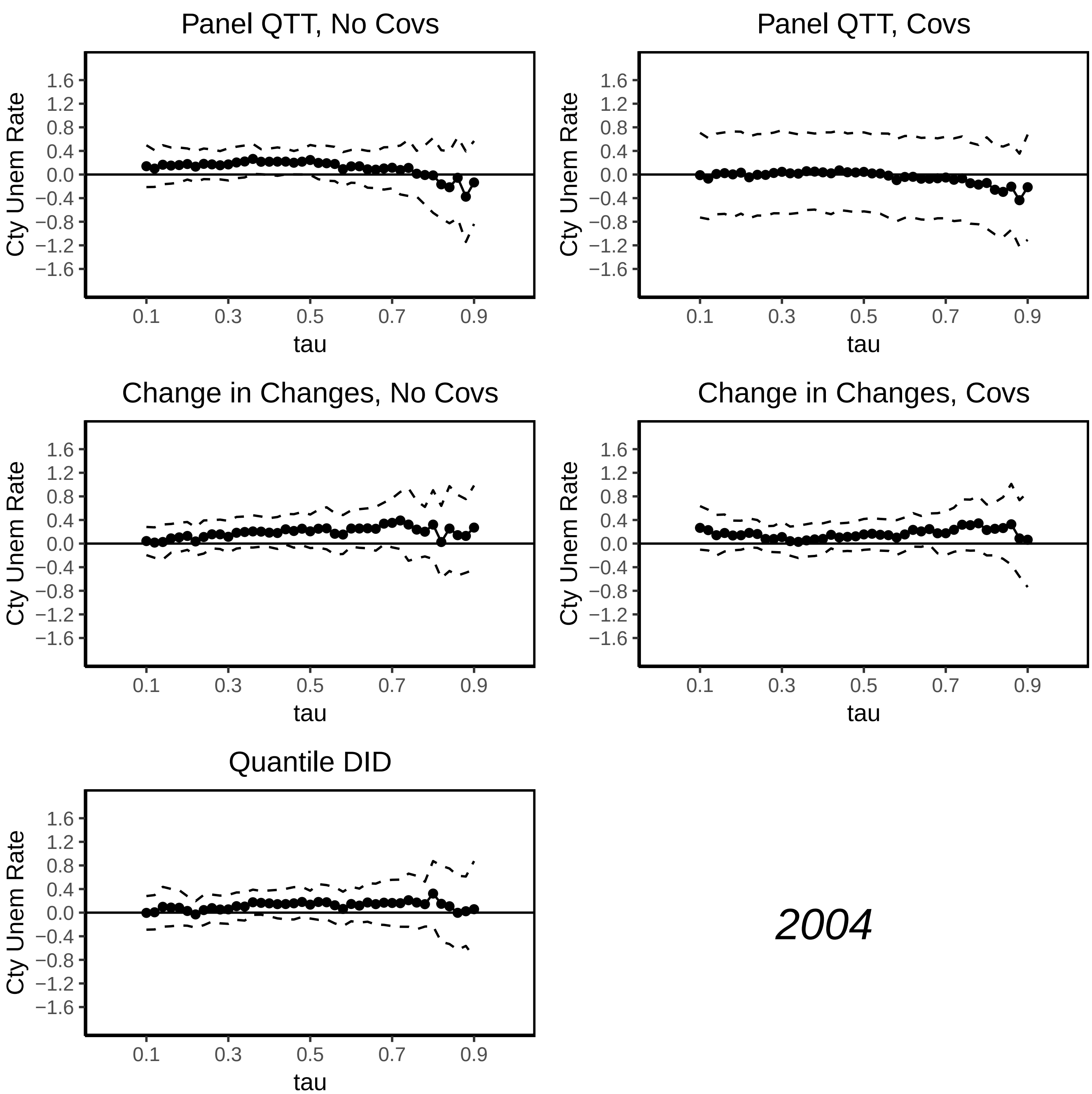


Pre-Treatment QTT Estimates

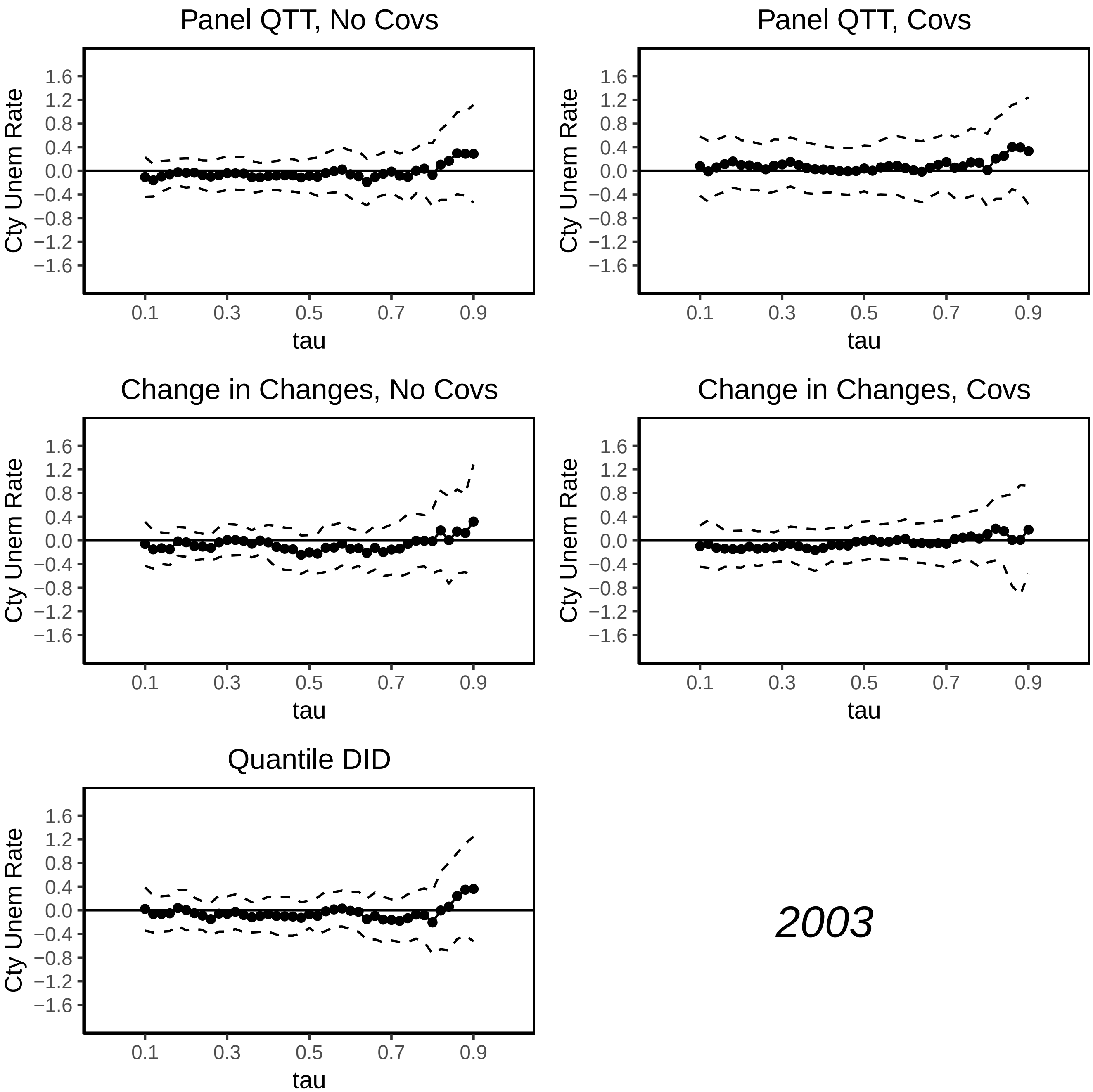
This section plots results from 2002-2006 using our method (both with and without covariates), the Change in Changes method (both with and without covariates) and the Quantile Difference in Differences method. These are all pre-treatment periods, so we are interested in "pre-testing" each method using these pre-treatment periods. In other words, in pre-treatment periods the QTT should be 0 at all values of τ. Here, we use a finer grid of possible values for τ (τ ranges from 0.1 to 0.9 in increments of 0.02), and we report uniform (rather than pointwise) 95% confidence bands using the bootstrap with 100 iterations.



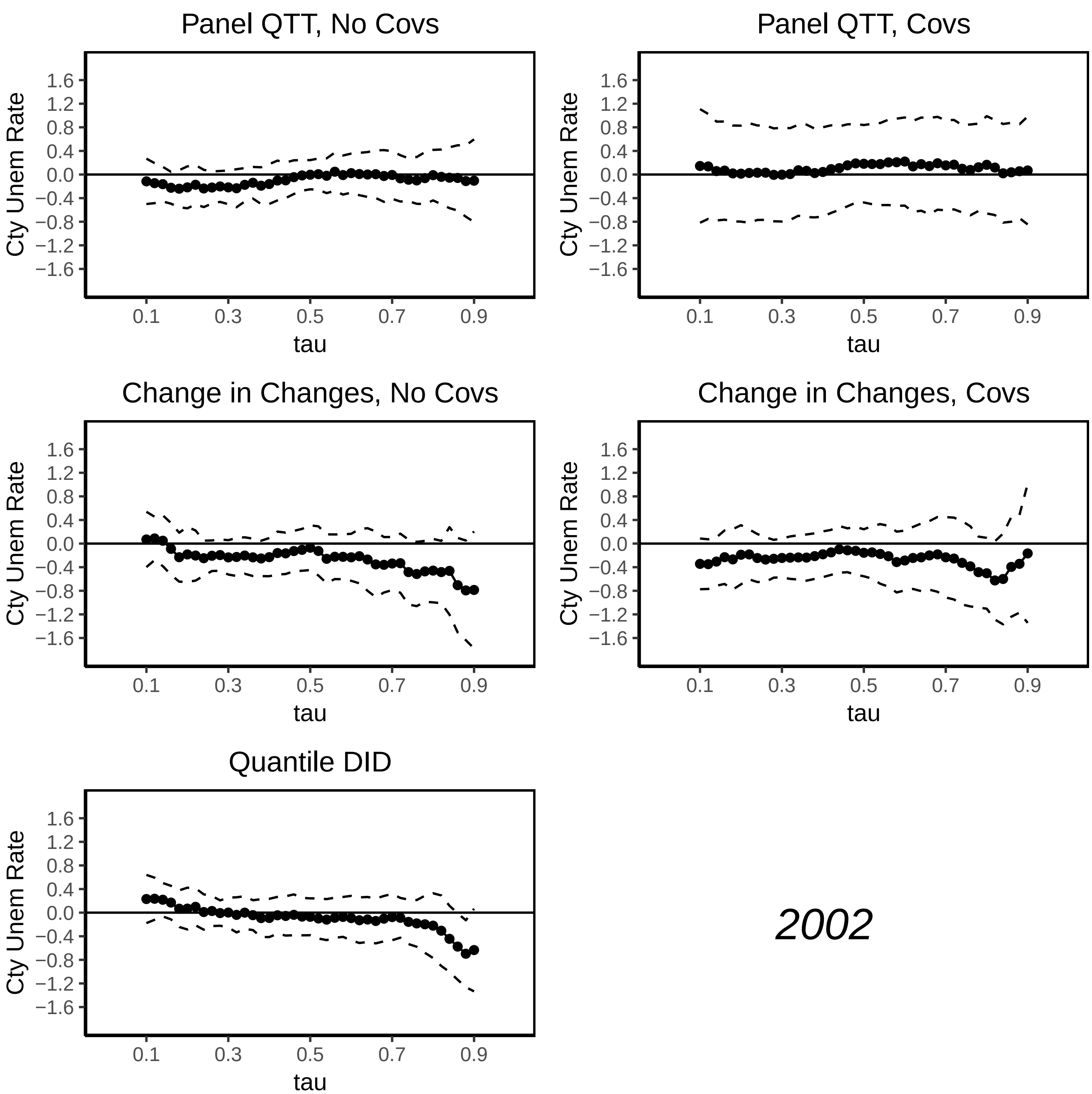

Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
1.6
1.2
0.8
0.4
0.0
−0.4
−0.8
−1.2
−1.6
0.1
0.3
0.5
0.7
0.9
tau
2005



Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
tau
1.6
1.2
0.8
0.4
0.0
−0.4
−0.8
−1.2
−1.6
0.1
0.3
0.5
0.7
0.9
2004



Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
tau
1.6
1.2
0.8
0.4
0.0
−0.4
−0.8
−1.2
−1.6
0.1
0.3
0.5
0.7
0.9
2003


Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
1.6
1.2
0.8
0.4
0.0
−0.4
−0.8
−1.2
−1.6
0.1
0.3
0.5
0.7
0.9
tau
2002

References

[1] Athey, Susan and Guido Imbens. "Identification and inference in nonlinear difference-in-differences models". *Econometrica* 74.2 (2006), pp. 431–497.

[2] Chen, Xiaohong, Oliver Linton, and Ingrid Van Keilegom. "Estimation of semiparametric models when the criterion function is not smooth". *Econometrica* 71.5 (2003), pp. 1591–1608.

[3] Donald, Stephen G and Yu-Chin Hsu. "Estimation and inference for distribution functions and quantile functions in treatment effect models". *Journal of Econometrics* 178 (2014), pp. 383–397.

[4] Ferreira, Francisco HG, Sergio Firpo, and Antonio F Galvao. "Estimation and inference for actual and counterfactual growth incidence curves". Working Paper. 2017.

[5] Firpo, Sergio. "Efficient semiparametric estimation of quantile treatment effects". *Econometrica* 75.1 (2007), pp. 259–276.

[6] Hirano, Keisuke, Guido Imbens, and Geert Ridder. "Efficient estimation of average treatment effects using the estimated propensity score". *Econometrica* 71.4 (2003), pp. 1161–1189.

[7] Van der Vaart, Aad W and Jon A Wellner. *Weak Convergence*. Springer, 1996.