# The changes-in-changes model with covariates

Blaise Melly, Giulia Santangelo

Bern University, *European Commission - JRC*

First version: January 2015

Last changes: January 23, 2015

Preliminary!

**Abstract:**

Athey and Imbens (2006) suggest the changes-in-changes model as an alternative to difference-in-differences. Their approach does not depend on the scale of the dependent variable and recovers the whole distribution of the counterfactual outcome. Estimation is relatively straightforward in the absence of covariates but is not well developed in the presence of covariates. We suggest a flexible semiparametric estimator that does not impose any separability assumption. We estimate the conditional outcome distributions for both groups and both periods using quantile regression. We then apply the changes-in-changes transformations conditionally on the covariates and integrate the estimated conditional distribution over the empirical distribution of the covariates to obtain unconditional effects. We derive joint functional central limit theorems and bootstrap validity results for the estimators of the potential outcome distributions as well as their functionals. We apply the suggested estimators to estimate the impact of food stamps on the birthweight distribution. We find positive effects, especially at the bottom and at the top of the distribution.

Keywords: Difference-in-differences, heterogenous treatment effects, counterfactual distribution, quantile regression, unconditional quantile and distribution effects

JEL classification: C12, C13, C14, C21

*Blaise Melly, Bern University, Department of Economics, Bern, Switzerland, blaise.melly@vwi.unibe.ch; Giulia Santangelo, European Commission - Joint Research Center, giulia.santangelo@jrc.ec.europa.eu.*

# 1 Introduction

Differences-in-differences is a quasi-experimental technique used to estimate the effects of a treatment that is not affecting everyone at the same time. The common trend assumption is the main restriction imposed on the data generating process. Athey and Imbens (2006, AI thereafter) suggest the changes-in-changes (CIC) model as an alternative to differences-in-differences (DID). Their approach does not depend on the scale of the dependent variable and recovers the whole distribution of the counterfactual outcome. The main identifying restriction is the time invariance of the distribution of the unobservables within each group (treated and control).

Estimation is relatively straightforward in the absence of covariates. In the presence of covariates, AI suggest either a nonparametric strategy or a parametric strategy based on a separability assumption. The former naturally suffers from the curse of dimensionality and is not practicable in many applications. The later justifies a simple two step estimation procedure (OLS regression followed by the unconditional CIC transformation applied to the residuals) but imposes restrictions that limit the appeal of the method to analyze heterogeneous effects. The main reason for using the CIC model is that the effects may not be additive; it would therefore be contradictory to impose additive separability for the effects of covariates.

While the AI paper has been quite influential in the theoretical literature and is well cited, the number of empirical applications of this approach is limited so far. According to the survey by Lechner (2011) the lack of tractable estimator in the presence of covariates is one of the main reasons for the relative low number of applications. Almost all applications incorporate covariates in the estimation, at the very least as robustness checks. Time-constant covariates are useful when the trends differ by covariates. Time-varying covariates are particularly useful to capture time-varying differences between groups. They also allow including linear trends as regressors, which relax the time invariance assumption. And even when the time invariance assumption is satisfied unconditionally, there may be efficiency gain of including covariates.

In this paper we suggest a flexible semiparametric estimator that does not impose any separability assumption. We estimate the whole conditional outcome distributions for both groups and both periods using quantile regression. We apply the quantile-quantile and probability-probability transformations at the heart of the AI procedure conditionally on the covariates. Finally we integrate the conditional distributions over the empirical distribution of the covariates to obtain unconditional estimates. We have implemented this estimator in Stata and hope that this will contribute to the popularization of the CIC model.[1]

Under standard regularity conditions we obtain uniformly consistent and asymptotically Gaussian estimators for functionals of conditional and unconditional potential outcome distributions. Examples of these functionals include distribution functions, quantile functions, quantile effects, distribution effects, average effects, Lorenz curves, and Gini coefficients. We derive functional central limit theorems for the estimators of the conditional and unconditional potential outcome distributions as well as of the functionals of interest. We also show that the exchangeable bootstrap is valid for estimating the limit laws of the estimators. This allows us to construct confidence sets that are uniform in the sense that they cover the entire functionals with prespecified probability and can be used to test functional hypotheses such as no-effect, positive effect, or stochastic dominance. We also suggest a new, consistent test of the validity of the identifying assumption that can be used when several periods are observed during which no group is treated.

The proof of these results starts from the existing functional central limit theorem and the validity of the exchangeable bootstrap for the (entire) empirical coefficient process of quantile regression. Then, we show that all the operators defining the estimators (the quantile-quantile transformation, the probability-probability transformation, and the counterfactual operator) are Hadamard differentiable with respect to their arguments. The results follow from the functional delta method.

We apply our estimation and inference procedures to study the effects of food stamps on the distribution of the birthweights. We use the same data and a similar strategy as in Almond, Hoynes, and Schanzenbach (2011). In this application there is a particular interest in knowing the effects at the lower tail of the distribution (probability of a low birthweight). It is, therefore, a natural application for the CIC approach but no estimator was available because Almond, Hoynes, and Schanzenbach (2011) included many covariates in all their specifications. We find a positive average effect of food stamps on birthweight with a U-shaped pattern for the quantile treatment effects. The effects are particularly large at the lower tail of the distribution. Nevertheless, these effects are relatively small in absolute value and cannot be the main motivation for the food stamp program.

The rest of the paper is organized as follows. Section 2 describes our model and shows that

[1] The codes are available by email from the authors.

the treatment effects of interest are identified. Section 3 describes our estimators. Section 4 gives the econometric results and their proofs. Section 5 discusses the results of the application. Finally section 6 concludes.

# 2 Model and identification

We consider a CIC model with covariates. We use as much as possible the same notation as AI. To simplify the notation we consider the simplest possible setup in Sections 2, 3, and 4: two periods and two groups. We explain in Section 5 how we deal with a more complex setup like the one we have in our application.

## 2.1 Model

An individual belongs to a group $G \in \{0,1\}$ (where group 1 is the treatment group) and is observed in period $T \in \{0,1\}$. We assume that only individuals in group 1 in period 1 are treated. We also observe a vector of covariates $X$ whose distribution has support $\mathbb{X}$. We use the potential outcome notation and let $Y^N$ denote the outcome if the individual does not receive the treatment, and let $Y^I$ be the outcome if she does receive the treatment. Thus, if $I$ is an indicator for the treatment, the realized outcome is

$$Y = Y^N \cdot (1-I) + Y^I \cdot I$$

In the two-groups two-periods model $I = G \cdot T$. The observed data are $(Y, G, T, X)$.

We assume that the control potential outcome is a nonseparable function of the covariates, the time period and an unobservable term.

**Assumption 1 (potential outcome)** *The outcome of an individual in the absence of intervention satisfies the relationship* $Y^N = h(X, T, U)$.

The random variable $U$ represents the unobservable component of $Y$. Without the other assumptions defined below Assumption 1 does not restrict the data generating process because any random variable $Y^N$ can be represented as $h(X, T, U)$ for a non-restricted function $h(\cdot)$ and a non-restricted random variable $U$.

To ease the notational burden, we introduce the shorthands

$$Y^N_{gtx} \overset{d}{\sim} Y^N|G=g,T=t,X=x,\; Y^I_{gtx} \overset{d}{\sim} Y^I|G=g,T=t,X=x$$
$$Y_{gtx} \overset{d}{\sim} Y|G=g,T=t,X=x,\; U_{gx} \overset{d}{\sim} U|G=g,X=x$$

where $\overset{d}{\sim}$ is a shorthand for "is distributed as." The corresponding conditional distribution functions are $F_{Y^N|gtx}$, $F_{Y^I|gtx}$, $F_{Y|gtx}$, and $F_{U|gx}$, with supports $\mathbb{Y}^N_{gtx}$, $\mathbb{Y}^I_{gtx}$, $\mathbb{Y}_{gtx}$, and $\mathbb{U}_{gx}$, respectively.

The distribution of $Y^I_{11x}$ is identified by the observed distribution of $Y_{11x}$. In order to identify the treatment effect for this population we need to identify $Y^N_{11x}$.[2] We make the following assumptions:

**Assumption 2 (strict monotonicity)** *The production function $h(t,x,u)$ is strictly increasing in $u$ for $t \in \{0,1\}$ and $\forall x \in \mathbb{X}$.*

**Assumption 3 (time invariance)** *We have $U \perp T|G,X$.*

**Assumption 4 (support)** $\mathbb{U}_{1x} \subset \mathbb{U}_{0x}$ for $\forall x \in \mathbb{X}$.

Assumptions 1 and 2 imply that the $\tau$ quantile of $Y^N$ given $X=x$ and $T=t$ is given by the $h(\cdot)$ function evaluated at the $\tau$ quantile of $U$ given $X=x$ and $T=t$:

$$F^{-1}_{Y^N}(\tau|X=x,T=t) = h\left(x,t,F^{-1}_U(\tau|X=x,T=t)\right)$$

The function $h(\cdot)$ is not identified without further normalization. A natural normalization consists in imposing $U|X=x,T=t \sim U(0,1)$. Under this normalization, the $h(\cdot)$ is simply the quantile function of $Y^N$. This normalization is not needed because we are not interested in $h(\cdot)$ directly but it helps getting the intuition. Matzkin (2003) discusses the identification of this type of models.

Assumption 3 is the key restriction that allows us to extrapolate the ranks from one period to the other. It is the CIC counterpart of the common trend assumption in the DID model. The stronger assumption that $U$ is constant over time for all individuals is called rank preservation or rank invariance. Here, we need only equality in distribution over time, which is sometimes called rank similarity.[3]

[2] We focus on group 1 because the treated outcome is only observed for this group. We focus on period 1 because the treated outcome is only observed in period 1. Identifying treatment effects for group 0 or period 0 therefore requires much stronger assumptions.

[3] See for instance the discussion in Chernozhukov and Hansen (2005).

Assumptions 1-3 are not testable with only 2 periods and 2 groups. The possibility to test the time invariance assumption when more than 2 periods is discussed in Section 4.4.

The identification strategies will consist of finding observations in group 0 that have similar $U$ as the observations in group 1. Under Assumption 4 these control observations exist. Note that Assumption 4 is testable because it implies $\mathbb{Y}_{10x} \subset \mathbb{Y}_{00x}$ for $\forall x \in \mathbb{X}$. The consequence of its violation is that only a part of the counterfactual distribution is identified. Our results would apply only to the part of $\mathbb{U}_{0x}$ that overlaps with $\mathbb{U}_{1x}$.

In principle, the model allows for discrete outcomes. However, the assumption of strict monotonicity is very restrictive for discrete outcomes. It means that the discreteness of the outcome comes from the discreteness of the unobserved term. A model like an ordered logit or rounded outcomes are excluded. Therefore, we do not advise using this model for discrete outcomes.[4] We assume that the conditional distribution of $Y$ is continuous in Assumption 6.

## 2.2 Identification

Proposition 1 trivially extends Theorem 3.1 in AI to the case with covariates.[5]

**Proposition 1** *(identification of the conditional distribution) Suppose that Assumptions 1–4 hold and let* $0 < \tau < 1$*. Then the* $\tau$ *quantile of* $Y_{11x}^N$ *is identified for* $\forall x \in \mathbb{X}$ *with*

$$F_{Y^N|11x}^{-1}(\tau) = F_{Y|01x}^{-1}\left(F_{Y|00x}\left(F_{Y|10x}^{-1}(\tau)\right)\right).$$

[4] AI also consider the case of a weakly increasing $h(\cdot)$ and show that the treatment effects are only partially identified in that case. We briefly discuss this case in the conclusion and let it for future work.

[5] See also the work of Fortin and Lemieux (1998) and Altonji and Blank (1999), who derive similar expressions as AI in a different context.

**Proof.** Let $x$ be any point in $\mathbb{X}$. By assumption $h(t,x,u)$ is invertible in $u$; denote this inverse by $h^{-1}(t,x,y)$. Consider the distribution of $Y^N_{gtx}$:

$$\begin{aligned}
F_{Y^N|gtx}(y) &= \Pr\left(h(t,x,U) \leq y | G=g, T=t, X=x\right) \\
&= \Pr\left(U \leq h^{-1}(t,x,y) | G=g, T=t, X=x\right) \\
&= \Pr\left(U \leq h^{-1}(t,x,y) | G=g, X=x\right) \\
&= \Pr\left(U_{gx} \leq h^{-1}(t,x,y)\right) \\
&= F_{U|gx}\left(h^{-1}(t,x,y)\right) \qquad (1)
\end{aligned}$$

The first line is true by definition, the second line by inverting $h$, the third line by the time invariance assumption, the fourth and fifth lines are definitions.

The preceding equation is central to the proof and will be applied to all four combinations $(g,t)$. First, letting $(g,t)=(0,0)$ and substituting $y=h(0,x,u)$,

$$\begin{aligned}
F_{Y|00x}(h(0,x,u)) &= F_{U|0x}\left(h^{-1}(0,x,h(0,x,u))\right) \\
&= F_{U|0x}(u).
\end{aligned}$$

Then applying $F^{-1}_{Y|00x}$ to each side, we have, for all $u \in \mathbb{U}_{0x}$,

$$h(0,x,u) = F^{-1}_{Y|00x}\left(F_{U|0x}(u)\right). \qquad (2)$$

Second, applying (1) with $(g,t)=(0,1)$, using the fact that $h^{-1}(1,x,y) \in \mathbb{U}_{0x}$ for all $y \in \mathbb{Y}_{01x}$, and applying the transformation $F^{-1}_{U|0x}(\cdot)$ to both sides, we have

$$F_{Y|01x}(y) = F_{U|0x}\left(h^{-1}(1,x,y)\right)$$

such that

$$F^{-1}_{U|0x}\left(F_{Y|01x}(y)\right) = h^{-1}(1,x,y). \qquad (3)$$

Combining (2) and (3) yields, for all $y \in \mathbb{Y}_{01x}$,

$$\begin{aligned}
h\left(0,x,h^{-1}(1,x,y)\right) &= F^{-1}_{Y|00x}\left(F_{U|0x}\left(h^{-1}(1,x,y)\right)\right) \\
&= F^{-1}_{Y|00x}\left(F_{U|0x}\left(F^{-1}_{U|0x}\left(F_{Y|01x}(y)\right)\right)\right) \\
&= F^{-1}_{Y|00x}\left(F_{Y|01x}(y)\right). \qquad (4)
\end{aligned}$$

Note that $h\left(0, x, h^{-1}(1, x, y)\right)$ is the period 0 outcome for an individual with characteristic $x$ and the realization of $u$ that corresponds to outcome $y$ in group 0 and period 1. Equation (4) shows that this outcome can be determined from the observable distributions.

Third, apply (1) with $(g,t) = (1,0)$ and substitute $y = h(0, x, u)$ to get

$$\begin{aligned} F_{Y|10x}(y) &= F_{U|1x}\left(h^{-1}(0, x, h(0, x, u))\right) \\ &= F_{U|1x}(u). \end{aligned} \tag{5}$$

Combining (4) and (5), and substituting into (1) with $(g,t) = (1,1)$, we obtain, for all $y \in \mathbb{Y}_{01}$,

$$\begin{aligned} F_{Y^N|11x}(y) &= F_{U|1x}\left(h^{-1}(1, x, y)\right) \\ &= F_{Y|10x}\left(h\left(0, x, h^{-1}(1, x, y)\right)\right) \\ &= F_{Y|10x}\left(F^{-1}_{Y|00x}\left(F_{Y|01x}(y)\right)\right). \end{aligned}$$

By Assumption 4, it follows that $\mathbb{Y}^N_{11} \subset \mathbb{Y}_{01}$. Thus, the directly observable distributions $F_{Y|10x}$, $F_{Y|00x}$, and $F_{Y|01x}$ identify the distribution of $Y^N_{11x}$. By inverting this distribution we obtain the representation of the quantile function given as the result of Proposition 1. ■

The intuition for the result is relatively simple. An individual at the $\tau$ quantile of the outcome distribution in period 0 and group 0 would be at the $F_{Y|00x}\left(F^{-1}_{Y|10x}(\tau)\right)$ quantile of the outcome distribution in the same period but in group 1. By the time invariance assumption this relative rank does not change over time such that the $\tau$ quantile of the outcome distribution in period 1 and group 0 is the same as the $F_{Y|00x}\left(F^{-1}_{Y|10x}(\tau)\right)$ quantile of the outcome distribution in the same period but in group 1.

In period 1, group 1 is treated such that $F^{-1}_{Y|11x}(\tau) = F^{-1}_{Y^I|11x}(\tau)$. From Proposition 1, $F^{-1}_{Y^N|11x}(\tau)$ is also identified. Since this is true for all quantiles, it follows that the conditional quantile treatment effect (QTE) process is identified for group 1 in period 1 and for the covariates value $x$:

$$\begin{aligned} \Delta^{QE}(\cdot|x) &= F^{-1}_{Y^I|11x}(\cdot) - F^{-1}_{Y^N|11x}(\cdot) \\ &= F^{-1}_{Y|11x}(\cdot) - F^{-1}_{Y|01x}\left(F_{Y|00x}\left(F^{-1}_{Y|10x}(\cdot)\right)\right). \end{aligned}$$

Of course, all functionals of the distribution of $Y_{11x}$ are also identified; for instance the average treatment effect and the distribution treatment effect:

$$\Delta^{AE}(|x) = E\left[Y^I_{11x}\right] - E\left[Y^N_{11x}\right]$$
$$\Delta^{DE}(\cdot|x) = F_{Y^I|11x}(\cdot) - F_{Y^N|11x}(\cdot).$$

These conditional treatment effects are identified for all $x$ in the support. Thus, it is possible to estimate the whole conditional treatment effects functionals. This allows analyzing the heterogeneity of the effects with respect to the observable characteristics. However, these high-dimensional functions are difficult to convey to the policy makers and the public. Therefore, we are often more interested in the unconditional treatment effects. The unconditional distribution of the treated outcome for period 1 and group 1 is

$$F_{Y^I|11}(y) = F_{Y|11}(y) = \int_{\mathbb{X}} F_{Y|11x}(y)\, dF_{X|11}(x)$$

where $F_{X,g}(x) \equiv F_X(x|G=g)$. Similarly, for the control outcome,

$$\begin{aligned} F_{Y^N|11}(y) &= \int_{\mathbb{X}} F_{Y^N|11x}(y)\, dF_{X|11}(x) \\ &= \int_{\mathbb{X}} F_{Y|10x}\left(F^{-1}_{Y|00x}\left(F_{Y|01x}(y)\right)\right) dF_{X|11}(x). \end{aligned}$$

All the elements are observable in the last expression. Again, all functionals of the marginal distribution of $Y_{11}$ are identified. We are in particular interested in the unconditional quantile effect process

$$\Delta^{QE}(\cdot) = F^{-1}_{Y^I|11}(\cdot) - F^{-1}_{Y^N|11}(\cdot).$$

The quantile and distribution treatment effects describe the causal effects of the treatment on the distribution of the potential outcome. The distribution of the individual treatment effects is also of high interest but is in general more difficult to identify.[6] If we have panel data and we strengthen the time invariance assumption (also called rank similarity in some models) to rank invariance (also called rank preservation), then we can identify the distribution of the individual treatment effect:

$$\begin{aligned} F_{\Delta|11}(\delta) &\equiv F_{Y^I-Y^N|11}(\delta) \\ &= \int_{\mathbb{X}} F_{Y^I-Y^N|11x}(\delta)\, dF_{X|11}(x) \\ &= \int_{\mathbb{X}} \int_0^1 1\left(\Delta^{QE}(\cdot|x) \leq \delta\right) dF_{X|11}(x). \end{aligned} \tag{6}$$

[6] See the interesting discussion in Heckman, Smith, and Clements (1997).

This allows, for instance, to identify the proportion of the population that benefits from the treatment. We do not further discuss this estimand in this paper because we only have repeated cross-sections in our application. The limiting distribution of the plug-in estimator based on (6) can be derived by combining the results in this paper with the results in the Appendix C in Chernozhukov, Fernández-Val, and Melly (2009).

# 3 Estimation

In this section we suggest estimators for the distributions and treatment effects identified in the previous section. The identification results are constructive in the sense that they show that the estimands are functions of observable quantities. We follow the plug-in principle and replace all the observable elements by consistent estimators. At least three different conditional distribution or quantile functions must be estimated. To make estimation both practical and realistic, we impose a flexible semiparametric restriction on the functional form of the conditional distribution or quantile function. We discuss an estimator based on linear quantile regression estimator but other estimators, such as distribution regression, could be used as well. Quantile regression is flexible in that by considering rich enough transformations of the original regressors, one could approximate the true conditional quantile function arbitrarily well when $Y$ has a smooth conditional density.

We apply the linear quantile regression estimator of Koenker and Bassett (1978) separately in the four samples defined by the values of $(g, t)$ to estimate four conditional quantile regression processes:

$$\hat{\beta}_{gt}(u) = \underset{b \in \mathbb{R}^{k+1}}{\arg\min} \sum_{i: G_i = g, T_i = t} \left(u - 1\left(Y_i \leq X_i' b\right)\right) \cdot \left(Y_i - X_i' b\right)$$

These estimator directly provide estimates of the conditional quantile functions

$$\hat{F}^{-1}_{Y|gtx}(u) = x' \hat{\beta}_{gt}(u).$$

These estimated conditional quantile functions may be non-monotonic in the sense that $\bar{u} \geq \tilde{u}$ does not necessarily implies $\hat{F}^{-1}_{Y|gtx}(\bar{u}) > \hat{F}^{-1}_{Y|gtx}(\tilde{u})$. Thus, they cannot be directly inverted. Instead, we use the sample analog of the following alternative representation of the conditional distribution

$$F_{Y|gtx}(y) = \int_0^1 1\left(F^{-1}_{Y|gtx}(u) \leq y\right) du$$

to obtain an estimator

$$\hat{F}_{Y|gtx}(y) = \int_0^1 1\left(\hat{F}^{-1}_{Y|gtx}(u) \leq y\right) du$$

The statistical properties of this estimator are studied in Chernozhukov, Fernández-Val, and Galichon (2010).

The formulas above assume that the whole quantile regression process has been estimated.[7] In practice, it may not be computationally possible to estimate the whole process. The quantile regression coefficients can be estimated on a fine mesh $u_1 < \cdots < u_S$, with mesh width $\delta$ such that $\delta\sqrt{n} \to 0$. In this case, the estimator of the conditional distribution is computed as

$$\hat{F}_{Y|gtx}(y) = \delta \cdot \sum_{s=1}^{S} 1\left(x'\hat{\beta}_{gt}(u_s) \leq y\right).$$

Even with this simplification the computational burden may be too high because a large number of quantile regressions must be estimated in four different groups. In addition, we propose inference procedures based on the bootstrap, which multiplies the computational burden. In our application, the number of observations is above 5,000,000. In order to reduce the computational burden, we use the new algorithms developed in Melly (2014). These algorithms compute more quickly the discretized quantile regression process by using the information contained in the already estimated quantile regressions. This allows dividing by about 10 the computational burden.

Tail trimming seems unavoidable in practice because . For notational simplification we abstract from this issue in this paper. The formula can be easily adapted to tail trimming as in Chernozhukov, Fernández-Val, and Melly (2013).

Using these tools based on quantile regression, we estimate $F^{-1}_{Y^N|11x}(y) = F^{-1}_{Y|01x}\left(F_{Y|00x}\left(F^{-1}_{Y|10x}(\tau)\right)\right)$ by

$$\hat{F}^{-1}_{Y^N|11x}(\tau) = x'\hat{\beta}_{01}\left(\int_0^1 1\left(x'\hat{\beta}_{00}(u) \leq x'\hat{\beta}_{10}(\tau)\right) du\right).$$

$F^{-1}_{Y^I|11x}(\tau)$ can be simply estimated by

$$\hat{F}^{-1}_{Y^I|11x}(\tau) = x'\hat{\beta}_{11}(\tau)$$

[7] It is indeed possible to estimate *all* quantile regressions because the value of the estimates changes only at a finite number of points in any finite samples. The number of distinct quantile regressions is of order $\mathcal{O}(n \cdot \log n)$, see Portnoy (1991).

and the estimator for the conditional QTE is

$$\hat{\Delta}^{QE}(\cdot|x) = x'\left(\hat{\beta}_{11}(\tau) - \hat{\beta}_{01}\left(\int_0^1 1\left(x'\hat{\beta}_{00}(u) \leq x'\hat{\beta}_{10}(\tau)\right) du\right)\right). \tag{7}$$

The unconditional counterfactual distribution that we would observe if the group 1 in period 1 was non treated can be estimated by

$$\begin{aligned}\hat{F}_{Y^N|11}(y) &= \int_{\mathbb{X}} \hat{F}_{Y^N|11x}(y)\, dF_{X|11}(x) \\ &= \frac{1}{n_{11}} \sum_{i:G_i=1,T_i=1} \sum_{j=1}^{S} 1\left(x_i'\hat{\beta}_{01}\left(\int_0^1 1\left(x_i'\hat{\beta}_{00}(u_s) \leq x_i'\hat{\beta}_{10}(\tau)\right)\right) \leq y\right).\end{aligned}$$

$F_{Y^I|11}(y)$ can be estimated by the sample distribution of $Y_i$ in group 1 and period 1. Alternatively, we can integrate the conditional distribution estimated by quantile regression over the distribution of the covariates

$$\hat{F}_{Y^I|11}(y) = \frac{1}{n_{11}} \sum_{i:G_i=1,T_i=1} \sum_{j=1}^{S} 1\left(x_i'\hat{\beta}_{11}(u_s) \leq y\right).$$

In the application we use this second alternative and obtain the following estimator for the unconditional QTE

$$\hat{\Delta}^{QE}(\cdot) = \hat{F}^{-1}_{Y^I|11}(y) - \hat{F}^{-1}_{Y^N|11}(y). \tag{8}$$

# 4 Asymptotic results

## 4.1 Sampling and estimation assumptions

In addition to the identification assumptions made in section 2, we make the following assumptions regarding the sampling process of $\{(Y_i, T_i, G_i, X_i)\}_{i=1}^n$ :

**Assumption 5 (Data generating process)** (i) *Conditional on* $T_i = t$ and $G_i = g$, $(Y_i, X_i)$ is a random draw from the subpopulation with $G_i = g$ during period $t$ that has probability law $P_{gt}$. (ii) For all $t$, $g \in \{0, 1\}$, $\alpha_{gt} \equiv \Pr(T_i = t, G_i = g) > 0$.

This sampling process corresponds to repeated cross-sectional sampling. This corresponds to our application in section 5. The theoretical results below must be modified if we observe a panel of individuals over time. As in Section 5.3 of AI, the only difference would the presence of additional terms taking into account the correlation of the estimated conditional distributions over time within groups.

We also impose the following conditions for all $t$, $g \in \{0, 1\}$.

**Assumption 6 (Quantile regression regularity conditions)** (i) The conditional quantile function takes the form

$$F_{Y|G,T,X}^{-1}(u|g,t,x) = x'\beta_{gt}(u)$$

for all $u \in (0,1)$ and $x \in \mathbb{X}$. (ii) The conditional density function $f_{Y|G,T,X}(y|x)$ exists, is uniformly bounded, and is uniformly continuous in $(y,x)$ in the support $\mathbb{Y}_{gt}\mathbb{X}$, which is a compact subset of $\mathbb{R}^{K+1}$. (iii) The minimal eigenvalue of $J_{gt}(u) \equiv E\left[f_{Y|gtX}(X'\beta(u))XX'\right]$ is bounded away from zero uniformly over $u$. (iv) $E\|X\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

These are standard regularity in quantile regression models.

## 4.2 Asymptotic distribution: ingredients

We use four main ingredients in our derivation of the limiting distribution of our estimators. We then combine these ingredients using the functional delta method.

**1. Conditional quantile and distribution functions** Under Assumptions 5 and 6, Corollary 5.2 in Chernozhukov, Fernández-Val, and Melly (2013) implies that, for all $t$, $g \in \{0,1\}$, as $n \to \infty$ in $l^{\infty}(\mathbb{U}_g\mathbb{X})$

$$\sqrt{n}\left(\hat{F}_{Y|gtx}^{-1}(u) - F_{Y|gtx}^{-1}(u)\right) \rightsquigarrow \mathbb{Z}_{gt}^{Q}(u,x) \tag{9}$$

as stochastic process indexed by $(u,x) \in \mathbb{U}_g\mathbb{X}$. The limit processes $\mathbb{Z}_{gt}^{Q}(u,x)$ are independent tight Gaussian processes with mean zero and covariance function

$$V_{gt}^{Q}(u,x,\tilde{u},\tilde{x}) = \alpha_{gt}^{-1} \cdot x' J_{gt}(u)^{-1}\left(\min(u,\tilde{u}) - u \cdot \tilde{u}\right) E\left[XX'|G=g,T=t\right] J_{gt}(u)^{-1}\tilde{x}.$$

The same corollary also provides the limiting distribution of the conditional distribution process. For all $t$, $g \in \{0,1\}$, as $n \to \infty$ in $l^{\infty}(\mathbb{Y}_{gt}\mathbb{X})$

$$\sqrt{n\alpha_{gt}}\left(\hat{F}_{Y|gtx}(y) - F_{Y|gtx}(y)\right) \rightsquigarrow \mathbb{Z}_{gt}^{F}(u,x) \equiv -f_{Y|gtx}(y)\,\mathbb{Z}_{gt}\left(F_{Y|gtx}(y),x\right) \tag{10}$$

as stochastic process indexed by $(y,x) \in \mathbb{Y}_{gt}\mathbb{X}$. The limit processes $\mathbb{Z}_{gt}^{F}(u,x)$ are independent tight Gaussian processes with mean zero and covariance function

$$V_{gt}^{F}(y,x,\tilde{y},\tilde{x}) = f_{Y|gtx}(y)\,f_{Y|gtx}(\tilde{y})\,V_{gt}^{Q}\left(F_{Y|gtx}(y),x,F_{Y|gtx}(\tilde{y}),\tilde{x}\right).$$

**2. Quantile-quantile transformation** The following lemma shows that the quantile-quantile transformation is Hadamard differentiable.

**Lemma 2** *Let $F$ and $G$ be two distribution functions that have a compact support and are continuously differentiable on their support with strictly positive derivatives $f$ and $g$. The quantile-quantile map*

$$\phi^{QQ}(F,G) = F \circ G^{-1}$$

*is Hadamard-differentiable at $F$ and $G$ tangentially to the set of functions $h_1$ and $h_2$ with derivative map*

$$\phi_{F,G}(h_1,h_2) = h_1 \circ G^{-1} - \frac{f \circ G^{-1}}{g \circ G^{-1}} h_2 \circ G^{-1}.$$

**Proof.** The proof follows from the Hadamard differentiability of the inverse map $G^{-1}$ (see Lemma 3.9.23(ii) in Van Der Vaart and Wellner (1996)) and from the chain rule for the Hadamard derivative (see Lemma 3.9.27 in Van Der Vaart and Wellner (1996)).[8] ■

**3. Probability-probability transformation** The following lemma shows that the probability-probability transformation is Hadamard differentiable.

**Lemma 3** *Let $F$ and $G$ be two distribution functions. $F$ has a compact support and is continuously differentiable on its support with strictly positive derivative $f$. The probability-probability map*

$$\varphi^{PP}(F,G) = F^{-1} \circ G.$$

*is Hadamard-differentiable at $F$ and $G$ tangentially to the set of functions $h_1$ and $h_2$ with derivative map*

$$\varphi_{F,G}(h_1,h_2) = -\frac{h_1}{f} \circ F^{-1} \circ G - \frac{h_2}{f \circ F^{-1}} \circ G.$$

**Proof.** The proof follows from the Hadamard differentiability of the inverse map $G^{-1}$ (see Lemma 3.9.23(ii) in Van Der Vaart and Wellner (1996)) and from the chain rule for the Hadamard derivative (see Lemma 3.9.27 in Van Der Vaart and Wellner (1996)). ■

[8] See also problem 4 on page 398 of Van Der Vaart and Wellner (1996).

**4. Counterfactual operator** Lemma D.1 in Chernozhukov, Fernández-Val, and Melly (2013) establishes that the counterfactual operator

$$\phi^{C}\left(F,G\right)=\int F\left(y,x\right)dG\left(x\right)$$

is Hadamard differentiable with the derivative map

$$\phi^{C}_{F_{Y|X},F_X}\left(\gamma,\pi\right)=\int \gamma\left(y,x\right)dF_X\left(x\right)+\pi\left(F_{Y|X}\left(y|x\right)\right).$$

the Hadamard differentiability of the counterfactual operator used to integrate the conditional distribution.

## 4.3 Limiting distribution of the CIC estimator

In this section we state and prove the main theoretical results of this paper. Theorem 4 provides the asymptotic distribution of the estimator of the conditional

**Theorem 4** *Suppose that Assumptions 1 to 6 hold. Then, (i)*

$$\sqrt{n}\left(\hat{F}^{-1}_{Y^I|11x}\left(\tau\right)-F^{-1}_{Y^I|11x}\left(\tau\right)\right)\rightsquigarrow \mathbb{Z}^{Q}_{11}\left(\tau,x\right)$$

*and*

$$\sqrt{n}\left(\hat{F}^{-1}_{Y^N|11x}\left(\tau\right)-F^{-1}_{Y^N|11x}\left(\tau\right)\right)\rightsquigarrow \mathbb{Z}^{Q}_{N}\left(\tau,x\right)$$

*as stochastic processes indexed by $(\tau,x)\in(0,1)\times\mathbb{X}$ and where $\mathbb{Z}^{Q}_{11}\left(\tau,x\right)$ and $\mathbb{Z}^{Q}_{N}\left(\tau,x\right)$ are independent tight zero-mean Gaussian process defined in (9) and (11).*

**Proof.** The results for $\hat{F}^{-1}_{Y^I|11x}\left(\tau\right)$ were already shown in (9). For the non-treated outcome, remember the definition of the estimator: $\hat{F}^{-1}_{Y^N|11x}\left(y\right)=\hat{F}^{-1}_{Y|01x}\left(\hat{F}_{Y|00x}\left(\hat{F}^{-1}_{Y|10x}\left(\tau\right)\right)\right)$. From (9), we know that the inside component converges to a Gaussian process

$$\sqrt{n}\left(\hat{F}^{-1}_{Y^I|10x}\left(\tau\right)-F^{-1}_{Y^I|10x}\left(\tau\right)\right)\rightsquigarrow \mathbb{Z}^{Q}_{10}\left(\tau,x\right).$$

By the functional delta method and the Hadamard differentiability of the quantile-quantile transformation shown in Lemma (2), we obtain

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y|00x}\left(\hat{F}^{-1}_{Y|10x}\left(\tau\right)\right)-F_{Y|00x}\left(F^{-1}_{Y|10x}\left(\tau\right)\right)\right)&\rightsquigarrow \mathbb{Z}^{R}\left(\tau,x\right)\\ &\equiv \mathbb{Z}^{F}_{00}\left(F^{-1}_{Y|10x}\left(\tau\right),x\right)+\frac{f_{Y|00x}\left(F^{-1}_{Y|10x}\left(\tau\right)\right)}{f_{Y|10x}\left(F^{-1}_{Y|10x}\left(\tau\right)\right)}\mathbb{Z}^{Q}_{10}\left(\tau,x\right).\end{aligned}$$

By the functional delta method and the Hadamard differentiability of the probability-probability transformation shown in Lemma (3), we obtain

$$\sqrt{n}\left(\hat{F}_{Y|01x}^{-1}\left(\hat{F}_{Y|00x}\left(\hat{F}_{Y|10x}^{-1}(\tau)\right)\right)-F_{Y|01x}^{-1}\left(F_{Y|00x}\left(F_{Y|10x}^{-1}(\tau)\right)\right)\right)\rightsquigarrow\mathbb{Z}_N^Q(\tau,x)$$

with

$$\mathbb{Z}_N^Q(\tau,x)\equiv\mathbb{Z}_{01}^Q\left(F_{Y|00x}\left(F_{Y|10x}^{-1}(\tau)\right),x\right)+\frac{\mathbb{Z}^R(\tau,x)}{f_{Y|01x}\left(F_{Y|00x}\left(F_{Y|10x}^{-1}(\tau)\right)\right)}. \tag{11}$$

■

**Corollary 5** *Suppose that Assumptions 1 to 6 hold. Then,*

$$\sqrt{n}\left(\hat{\Delta}^{QE}(\tau|x)-\Delta^{QE}(\tau|x)\right)\rightsquigarrow\mathbb{Z}^{QTE}(\tau,x)$$

*as stochastic processes indexed by* $(\tau,x)\in(0,1)\times\mathbb{X}$ *and where* $\mathbb{Z}^{QTE}(\tau,x)=\mathbb{Z}_{11}^Q(\tau,x)-\mathbb{Z}_N^Q(\tau,x)$ *is a tight zero-mean Gaussian process defined.*

**Proof.** Trivial by the functional delta method. ■

**Corollary 6** *Suppose that Assumptions 1 to 6 hold and* $\phi\left(F_{Y^I|11x}^{-1},F_{Y^N|11x}^{-1},w\right)$*, a functional of interest indexed by* $w$*, is Hadamard differentiable with derivatives* $\phi_I'$ *and* $\phi_N'$*.*

$$\sqrt{n}\left(\phi\left(\hat{F}_{Y^I|11x}^{-1},\hat{F}_{Y^N|11x}^{-1},w\right)-\phi\left(F_{Y^I|11x}^{-1},F_{Y^N|11x}^{-1},w\right)\right)\rightsquigarrow\phi_I'\left(\mathbb{Z}_{11}^Q(\cdot,x),w\right)+\phi_N'\left(\mathbb{Z}_N^Q(\cdot,x),w\right)$$

*as a stochastic process indexed by* $w$*.*

**Proof.** Trivial by the functional delta method. ■

**Theorem 7** *Suppose that Assumptions 1 to 6 hold. Then,*

$$\sqrt{n}\left(\hat{F}_{Y^I|11}^{-1}(\tau)-F_{Y^I|11}^{-1}(\tau)\right)\rightsquigarrow\mathbb{Z}_{11}^Q(\tau)$$

*and*

$$\sqrt{n}\left(\hat{F}_{Y^N|11}^{-1}(\tau)-F_{Y^N|11}^{-1}(\tau)\right)\rightsquigarrow\mathbb{Z}_N^Q(\tau)$$

*jointly as stochastic processes indexed by* $\tau\in(0,1)$ *and where* $\mathbb{Z}_{11}^Q(\tau)$ *and* $\mathbb{Z}_N^Q(\tau)$ *are tight zero-mean Gaussian process defined in (13) and (12).*

**Proof.** For the control outcome, we apply Corollary 6 with $\phi(\cdot) = F_{Y^N|11x}(y)$ to obtain the the weak convergence of the conditional distribution process

$$\sqrt{n}\left(\hat{F}_{Y^N|11x}(y) - F_{Y^N|11x}(y)\right) \rightsquigarrow \mathbb{Z}_N^F(y,x)$$

where

$$\mathbb{Z}_N^F(y,x) = f_{Y^N|11x}(y)\,\mathbb{Z}_N^Q\left(F_{Y^N|11x}(y), x\right).$$

In addition, by the Donsker Theorem, the empirical distribution of the covariates in group 1 during period 1

$$\frac{1}{\sqrt{n}}\sum_{i:G_i=T_i=1}\left(f(Y_i,X_i) - \int f(Y_i,X_i)\,dP_{11}\right) \rightsquigarrow \mathbb{Z}_{11}^X(f(y,x))$$

as stochastic process indexed by $f \in \mathcal{F}$, where $\mathcal{F}$ is a universal Donsker class. The limit processes $\mathbb{Z}_{11}^X(f)$ are tight $P_{11}$-Brownian bridges with covariance function

$$V_{11}^X(y,x,\tilde{y},\tilde{x}) = \alpha_{11}^{-1}\cdot\left(\int f(y,x)\,f(\tilde{y},\tilde{x})\,dP_{11} - \int f(y,x)\,dP_{11}\int f(y,x)\,dP_{11}\right).$$

Thus, by the functional delta method and the Hadamard differentiability of the counterfactual operator

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y^N|11}(y) - F_{Y^N|11}(y)\right) &= \sqrt{n}\left(\int_{\mathbb{X}}\hat{F}_{Y^N|11x}(y)\,d\hat{F}_{X|11}(x) - \int_{\mathbb{X}}F_{Y^N|11x}(y)\,dF_{X|11}(x)\right)\\ &\rightsquigarrow \int_{\mathbb{X}}\mathbb{Z}_N^F(y,x)\,dF_{X|11}(x) + \mathbb{Z}_{11}^X\left(F_{Y^N|11x}(y)\right) \equiv \mathbb{Z}_N^F(y)\end{aligned}$$

Finally by the functional delta method

$$\sqrt{n}\left(\hat{F}_{Y^N|11}^{-1}(\tau) - F_{Y^N|11}^{-1}(\tau)\right) \rightsquigarrow \frac{-\mathbb{Z}_N^F\left(F_{Y^N|11}^{-1}(\tau)\right)}{f_{Y^N|11}\left(F_{Y^N|11}^{-1}(\tau)\right)} \equiv \mathbb{Z}_N^Q(\tau). \tag{12}$$

The results for the treated outcome follows directly from Theorem 4.1(ii) in Chernozhukov, Fernández-Val, and Melly (2013):

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y^I|11}(y) - F_{Y^I|11}(y)\right) &\rightsquigarrow \int_{\mathbb{X}} f_{Y|11x}(y)\,\mathbb{Z}_{11}^Q\left(F_{Y|11x}(y), x\right)dF_{X|11}(x) + \mathbb{Z}_{11}^X\left(F_{Y|11x}(y)\right)\\ &\equiv \mathbb{Z}_{11}^F(y)\end{aligned}$$

Finally by the functional delta method

$$\sqrt{n}\left(\hat{F}_{Y^I|11}^{-1}(\tau) - F_{Y^I|11}^{-1}(\tau)\right) \rightsquigarrow \frac{-\mathbb{Z}_{11}^F\left(F_{Y|11}^{-1}(\tau)\right)}{f_{Y|11}\left(F_{Y|11}^{-1}(\tau)\right)} \equiv \mathbb{Z}_{11}^Q(\tau). \tag{13}$$

■

**Corollary 8** *Suppose that Assumptions 1 to 6 hold. Then,*

$$\sqrt{n}\left(\hat{\Delta}^{QE}(\tau) - \Delta^{QE}(\tau)\right) \rightsquigarrow \mathbb{Z}^{QTE}(\tau)$$

*as stochastic processes indexed by* $\tau \in (0,1)$ *and where* $\mathbb{Z}^{QTE}(\tau) = \mathbb{Z}_{11}^{Q}(\tau) - \mathbb{Z}_{N}^{Q}(\tau)$ *is a tight zero-mean Gaussian process.*

**Proof.** Trivial by the functional delta method. ■

**Corollary 9** *Suppose that Assumptions 1 to 6 hold and* $\phi\left(F_{Y^I|11}^{-1}, F_{Y^N|11}^{-1}, w\right)$*, a functional of interest indexed by* $w$*, is Hadamard differentiable with derivatives* $\phi_I'$ *and* $\phi_N'$*.*

$$\sqrt{n}\left(\phi\left(\hat{F}_{Y^I|11}^{-1}, \hat{F}_{Y^N|11}^{-1}, w\right) - \phi\left(F_{Y^I|11}^{-1}, F_{Y^N|11}^{-1}, w\right)\right) \rightsquigarrow \phi_I'\left(\mathbb{Z}_{11}^{Q}(\cdot), w\right) + \phi_N'\left(\mathbb{Z}_{N}^{Q}(\cdot), w\right)$$

*as a stochastic process indexed by* $w$*.*

**Proof.** Trivial by the functional delta method. ■

A few remarks are worth making. First, these results naturally imply that the pointwise estimators (QTE at a single quantile for instance) converge to normal distributions. The asymptotic variances of these estimators are given by the results above; these formulas can be used to develop analytical estimators of the standard errors. However, all asymptotic variances contain terms that are difficult to estimate such as the conditional density of the dependent variable given the covariates. Therefore, we suggest using resampling methods to estimate the standard errors of the estimates. This method also allows performing inference on the whole processes.

Second, while we focus mostly on the quantile treatment effects the corollaries provide the results for all Hadamard differentiable functionals of the conditional and marginal distributions of the potential outcomes. A simple special case of interest is the average treatment effect. A second, more involved example is the Lorenz curve and the Gini coefficient. Our results apply to all these examples.

Third, our results simplify to the results in AI in the special case where the covariates vector $X$ contains only a constant and we consider only the pointwise results. Even in the case, we contribute to the literature by providing the limiting distribution of the whole quantile and distribution processes. Our method of proof is also simpler because we use explicitly tools from the empirical process theory, which are perfectly suited to study the CIC estimator.

### 4.4 Inference

In this section we prove the validity of a general resampling procedure called the exchangeable bootstrap. This procedure incorporates many popular forms of resampling as special cases, namely the empirical bootstrap, weighted bootstrap, $m$ out of $n$ bootstrap, and subsampling. It is quite useful for applications to have all of these schemes covered by our theory. For example, in small samples with categorical covariates, we might want to use the weighted bootstrap to gain accuracy and robustness to "small cells," whereas in large samples, where computational tractability can be an important consideration, we might prefer subsampling.

We draw realizations of the vectors of weights that follow the following condition:

**Condition BW (bootstrap weights)** *For each $t$, $g \in \{0,1\}$, let $(w_{gt1},...,w_{gtn_{gt}})$ be an exchangeable,*[9] *nonnegative random vector, which is independent of the data, such that for some $\epsilon > 0$*

$$\sup_{n_k} E[w_{gt1}^{2+\epsilon}] < \infty$$

$$n_{gt}{}^{-1} \sum_{i=1}^{n_{gt}} (w_{gti} - \bar{w}_{gt})^2 \underset{p}{\rightarrow} 1$$

$$\bar{w}_{gt} = n_{gt}^{-1} \sum_{i=1}^{n_{gt}} w_{gti} \underset{p}{\rightarrow} 1.$$

*Moreover, the vectors $(w_{gt1},...,w_{gtn_{gt}})$ are independent across $(g,t)$.*

For example, $(w_{k1},...,w_{kn_k})$ are multinomial vectors with dimension $n_k$ and probabilities $(1/n_k,\ldots,1/n_k)$ in the empirical bootstrap. The exchangeable bootstrap uses the components of $(w_{k1},...,w_{kn_k})$ as random sampling weights in the construction of the bootstrap version of the estimators. Thus, for instance, the bootstrap version of the quantile regression estimator is

$$\hat{\beta}^*_{gt}(u) = \arg\min_{b\in\mathbb{R}^{k+1}} \sum_{i:G_i=g,T_i=t} w_{gti} \cdot \left(u - 1\left(Y_i \le X_i'b\right)\right) \cdot \left(Y_i - X_i'b\right),$$

the bootstrap version of the estimator of the conditional quantile function is

$$\hat{F}^{*-1}_{Y|gtx}(u) = x'\hat{\beta}^*_{gt}(u),$$

[9] A sequence of random variables $X_1, X_2, ..., X_n$ is exchangeable if for any finite permutation $\sigma$ of the indices $1, 2, ..., x$ the joint distribution of the permuted sequence $X_{\sigma(1)}, X_{\sigma(2)}, ..., X_{\sigma(n)}$ is the same as the joint distribution of the original sequence.

and the bootstrap version of the unconditional distribution of the non-treated outcome is

$$\hat{F}^{*}_{Y^N|11}(y) = \frac{1}{n_{11}} \sum_{i:G_i=1,T_i=1} \sum_{j=1}^{S} w_{11i} \cdot 1\left( x_i' \hat{\beta}^{*}_{01} \left( \int_0^1 1\left( x_i' \hat{\beta}^{*}_{00}(u_s) \leq x_i' \hat{\beta}^{*}_{10}(\tau) \right) \right) \leq y \right).$$

The exchangeable bootstrap distributions are useful to perform asymptotically valid inference on the counterfactual effects of interest. We focus on uniform methods that cover standard pointwise methods for real-valued parameters as special cases, and that also allow us to consider richer functional parameters and hypotheses. For example, an asymptotic simultaneous $(1-\alpha)$-confidence band for the unconditional quantile treatment effect process $\hat{\Delta}^{QE}(\tau)$ is defined by the end-point functions

$$\hat{\Delta}^{QE}(\tau)^{\pm} = \hat{\Delta}^{QE}(\tau) \pm \widehat{t}_{1-\alpha} \widehat{\Sigma}^{QE}(\tau)^{1/2}/\sqrt{n}, \tag{14}$$

such that

$$\lim_{n\to\infty} \Pr\left\{ \Delta^{QE}(\tau) \in [\hat{\Delta}^{QE}(\tau)^{-}, \hat{\Delta}^{QE}(\tau)^{+}] \text{ for all } \tau \in (0,1) \right\} = 1-\alpha. \tag{15}$$

Here, $\widehat{\Sigma}^{QE}(\tau)$ is a uniformly consistent estimator of the asymptotic variance function of $\sqrt{n}(\hat{\Delta}^{QE}(\tau) - \Delta^{QE}(\tau))$. In order to achieve the coverage property (15), we set the critical value $\widehat{t}_{1-\alpha}$ as a consistent estimator of the $(1-\alpha)$-quantile of the Kolmogorov-Smirnov maximal $t$-statistic:

$$t = \sup_{\tau\in(0,1)} \sqrt{n} \widehat{\Sigma}^{QE}(\tau)^{-1/2} |\hat{\Delta}^{QE}(\tau) - \Delta^{QE}(\tau)|.$$

To do that we draw $B$ bootstrap samples and set $\widehat{t}_{1-\alpha}$ to the $(1-\alpha)$-sample quantile of $\{\widehat{t}_b : 1 \leq b \leq B\}$ where $\widehat{t}_b = \sup_{\tau\in(0,1)}^{-1/2} \widehat{\Sigma}^{QE}(\tau) |\sqrt{n}(\hat{\Delta}^{QE}(\tau)^{*} - \hat{\Delta}^{QE}(\tau))|$ for $1 \leq b \leq B$.

Theorem 10 is the third main result of this paper. It verifies the validity of the bootstrap for all the estimators considered in this paper.

**Theorem 10** *Assumptions 1 to 6 hold and the bootstrap weights follow the condition BW. Then, the exchangeable boostrap consistently estimates the law of the limit stochastic processes in the Theorems and Corollaries 4 to 9. The confidence bands have a correct coverage probability.*

**Proof.** By Corollary 5.2(ii) in Chernozhukov, Fernández-Val, and Melly (2013) the exchangeable bootstrap is valid for the quantile regression-based estimators of the conditional distribution and quantile functions. The result then follows by the functional delta method for the bootstrap (see chapter 3.9 in Van Der Vaart and Wellner (1996)) and the Hadamard differentiability of all the functionals involved. ■

Our confidence bands can be used to test functional hypotheses about counterfactual effects. For example, it is straightforward to test no-effect, positive effect or stochastic dominance hypotheses by verifying whether the entire null hypothesis falls within the confidence band of the relevant counterfactual functional. In addition to the traditional null hypotheses of interest, this result also provides a test for the validity of the time invariance assumption when there are several periods during which both groups are non-treated. To consider this case, let's assume that we also observe $(Y, X)$ in period $-1$ and that none of the group is treated during that period. Our assumptions 1 to 4 imply

$$F_{Y,0-1x}\left(F_{Y,1-1x}^{-1}(\tau)\right) = F_{Y,00x}\left(F_{Y,10x}^{-1}(\tau)\right) \text{ for all } \tau \in (0,1) \text{ and } x \in \mathbb{X}.$$

All quantiles and covariate values must be considered in order to detect any deviation from the null hypothesis (validity of the assumptions 1 to 5, in particular of the time invariance assumption). Kolmogorov-Smirnov or a Cramer-von-Misses types of tests are the natural way to test this hypothesis. Our Theorem 10 justifies using the exchangeable bootstrap to estimate the critical values for these tests.

## 5 The impact of food stamps on the birthweight distribution

In this section, we apply our methods to estimate the impact of food stamps on the birthweight distribution. We use the same data as Almond, Hoynes, and Schanzenbach (2011) and complement their DID analysis by using the CIC approach. The Food Stamp Program (FSP) provided during the 1960s and early1970s a sizable improvement in the resources available to America's poorest. Because of data availability, Almond, Hoynes, and Schanzenbach (2011) analyzed the effects starting from 1968 when already about 40% of the counties had introduced food stamps. They estimated the effects until 1977, two years after all counties implemented the FSP. The sharp timing of the county-by-county rollout of the program can be exploited in the identification strategy to analyze the impact of the FSP considering the month in which the program began operating in each of the roughly 3100 U.S. counties. For our analysis we have to exclude the counties that were already treated at the beginning of the period because we cannot identify the counterfactual non-treated outcome for them. We also have to stop the analysis in June 1974 because the effects are not identified for periods when everyone is treated.

The econometric analysis is performed using data combined from several sources. The administrative FSP data from USDA annual reports on county food stamp caseloads contain information on the month and year that each county implemented a food stamp program, which is used as key treatment or policy variable. *Vital Statistics Natality Data* are available beginning in 1968 for about 2 million observations per year (a 100% or 50% sample of births depending on the state-year). They contain birth outcomes including birthweight, gender, and race. All the models estimated by Almond, Hoynes, and Schanzenbach (2011) include covariates. Almond, Hoynes, and Schanzenbach (2011) present results separately for white and black mothers and consider four different specifications with different controls. To keep the size of this paper reasonable we present only results for white mothers based on the first specification in Almond, Hoynes, and Schanzenbach (2011). This specification controls include 1960 county variables (log of population, percentage of land in farming, percentage of population black, urban, age below 5, age above 65, and with income less than $3,000), each interacted with a linear time trend, per capita county transfer income (public assistance, medical care, and retirement and disability benefits), and county real per capita income. These covariates come from the 1960 Census of Population and Census of Agriculture and the Bureau of Economic Analysis, Regional Economic Information System (REIS) data.

There is a clear interest in knowing the effect of food stamps on the distribution of birthweights. Low-birthweights are particularly detrimental to the future health of the newborn. For this reason, Almond, Hoynes, and Schanzenbach (2011) do not only report the estimated average effect but they also estimate the effects on the distributions using the DID estimator with $1(Y \leq y)$ for several thresholds (see in particular the right-hand side of Table 1 and Figure 3). Even in the absence of covariates,[10] this procedure is inconsistent except in three cases: (i) There is no time effect. (ii) There is no group effect. (iii) The dependent variable is uniformly distributed.

The reason for the asymptotic bias of the DID applied to the distribution is that the time effect and the group effect cannot be additive if $Y$ is not uniformly distributed. To illustrate this point we consider a very simple example. The outcome is drawn from the following normal distribution:

$$Y_i \sim N(0.5 \cdot T + 0.5 \cdot G, 1).$$

The treatment has no effect (there is no interaction between $T$ and $G$) uniformly over the distrib-

[10] There are naturally additional issues with the linear regression of a binary dependent variable on a non-saturated set of covariates, especially when we need to estimate the conditional probabilities that $Y = 1$.

ution. We can calculate analytically the probability limit of the DID estimator for $1(Y \leq -0.5)$:

$$\begin{aligned} F_{Y_{11}}(-0.5) - F_{Y_{10}}(-0.5) - F_{Y_{01}}(-0.5) + F_{Y_{00}}(-0.5) &= \Phi(-1.5) - \Phi(-1) - \Phi(-1) + \Phi(-0.5) \\ &\approx 0.058 \end{aligned}$$

This implies that the DID estimator is biased and this bias is large relative to the observe probability of $1(Y \leq -0.5)$ which is 0.067. When we do the same calculation at all possible cutoffs, we obtain Figure 1. We can see that the bias is positive at the lower tail and negative at the upper tail.

Since we mostly focus on quantile treatment in this paper, Figure 2 shows the asymptotic bias of the QTE implied by the DID estimator applied to the distribution function. These results are again very misleading given that the true effects are zero uniformly over $(0,1)$. Therefore, the results reported by Almond, Hoynes, and Schanzenbach (2011) for the effects on the distribution should be taken with caution.

We now apply our estimators to the food stamp application. This application is more complicated than our theoretical framework because we have 26 time periods (quarters) and, thus, 26 different groups of counties that were treated starting from a different point in time. We could in principle use any pair of time periods and estimate the effect for the group of counties that were not treated in the first but treated in the second period. These results would be almost impossible to present and highly imprecise. Therefore, we present the average results for all the possible pairs weighted in such a way that they are representative of the treated counties in our data.

A second complication is that some of the covariates are interacted with linear time trends. The coefficients on these covariates are not identified if we estimate the quantile regression separately for each period. Therefore, we estimate a single quantile regression process for all the non-treated observations and include indicator variables for the time periods and the groups.

A third practical issue is the size of the sample. In the original paper, the authors could collapse the data to averages per county-quarter cells. We cannot do that if we want to use the CIC estimator because we really need the whole distribution. Thus, we have to keep above 5,000,000 observations and estimate the whole quantile regression process for all these observations. This computational task becomes insurmountable if we need to bootstrap the results. Our solution was to use the newly developed quantile regression algorithms developed in Melly (2014) and subsampling instead of the bootstrap.[11]

Our main results are shown in Figure 3. As Almond, Hoynes, and Schanzenbach (2011) we find a positive effect of food stamps on birthweight. The absolute value of our estimated effect is larger but the standard errors are also larger. As Almond, Hoynes, and Schanzenbach (2011) we also find that the effects are stronger at the lower tail of the distribution, which is certainly a good outcome for the policy because this is where it matters the most. On the other hand, contrary to the original article we find also a larger effect at the upper end of the distribution.

Figure 4 shows both potential outcome distributions. While the effects we find are strongly significant, this figure makes clear that they are not large in absolute value. The two quantile functions are almost impossible to distinguish. Finally Figure 5 shows that in this application the results depend on the presence of covariates. Without covariates the effects are very close to zero for most of the distribution. On the other hand, the U-shaped pattern is the same with and without covariates.

## 6 Conclusion

The CIC model of Athey and Imbens (2006) is an elegant and coherent way to identify treatment effects when different groups received the treatment at different point in time. The objective of our paper is to increase the applicability of this approach by offering new estimators that incorporate covariates in the estimation. Our estimators are based on the estimation of the conditional distribution function by families of linear quantile regressions. Quantile regression is flexible in that by considering rich enough transformations of the original regressors, one could approximate the true conditional quantile function arbitrarily well when $Y$ has a smooth conditional density. Nevertheless, the suggested estimators are easy to use, in particular because we provide codes that implement the estimation and inference procedures developed in this paper.

When compliance is only partial our results may be useful to develop estimators for the instrumental CIC model of de Chaisemartin and D'Haultfoeuille (2014). They are also useful for applications outside of the difference-in-differences research design. For example, Strittmatter (2014) and Bitler, Domina, and Hoynes (2014) suggest to report a new estimand, the translated quantile treatment effects. This estimand corresponds to the counterfactual non-treated distrib-

[11] In each bootstrap draw we sample 500 counties from above 3000 counties. This allows for arbitrary correlation between the observations within the counties.

ution in the CIC model. Wüthrich (2014) provides a closed-form expression for the instrumental variable quantile regression estimator of Chernozhukov and Hansen (2006) when the instrument is binary. This closed-form is, again, very similar to the expressions for which we have developed estimators. Finally, in Melly and Santangelo (2014), we suggest a new correction for sample selection that is based on panel data and exploits the information provided by the periods during which the now non-employed women worked. We make a time invariance assumption that allows us to "impute" the conditional ranks in the wage distribution from one year to another.

In this paper we have focused exclusively on continuous outcomes. The reason is that AI shows that the effects are only partially identified with discrete outcomes. It would nevertheless be interesting to develop estimators for these bounds in the presence of covariates. Incorporating covariates is especially attractive because it allows tightening the bounds. The estimators based on quantile regression are not well suited for discrete outcomes. On the other hand, it seems possible to develop similar estimators based on distribution regression because this method accommodates discrete and mixed dependent variables. We hope to extend the analysis to discrete outcomes in future work.

# 7 Figures

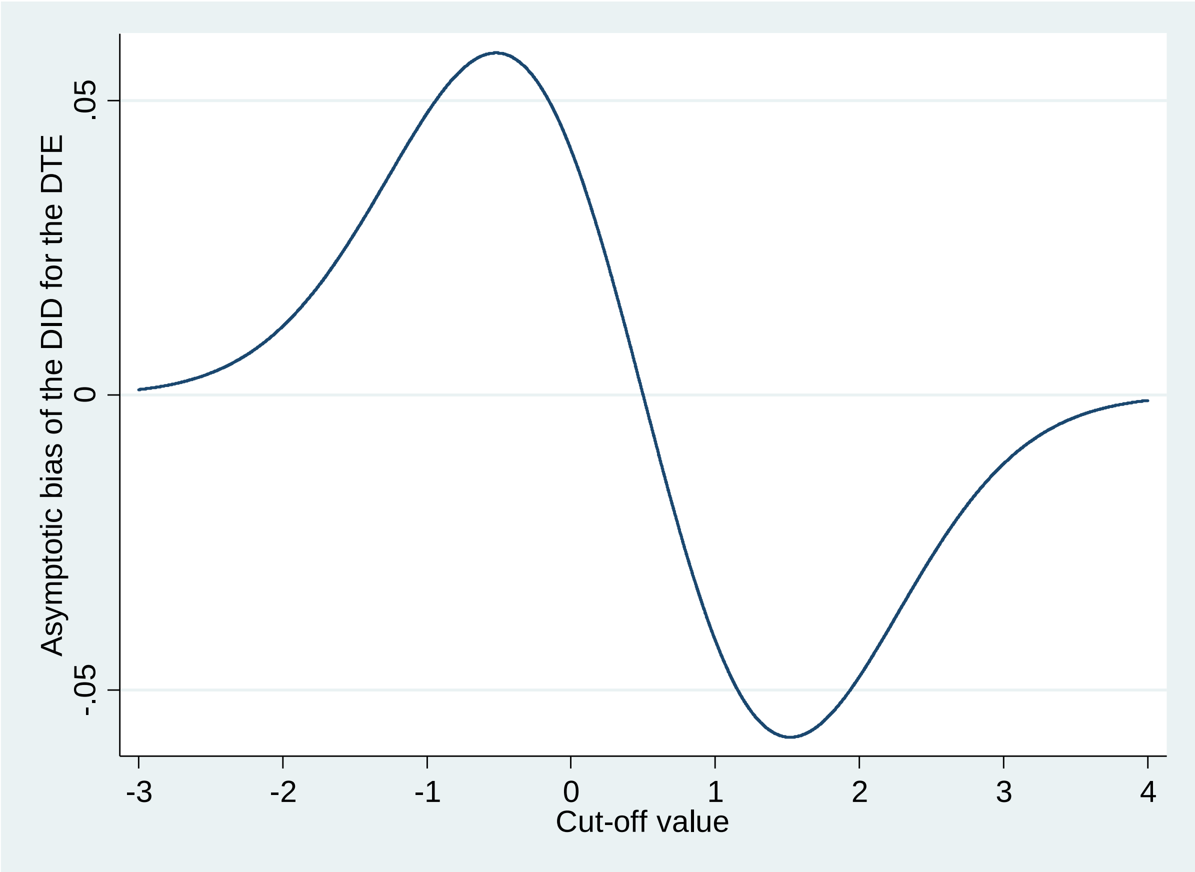


Figure 1: The bias of the DID estimator for the distribution effects.

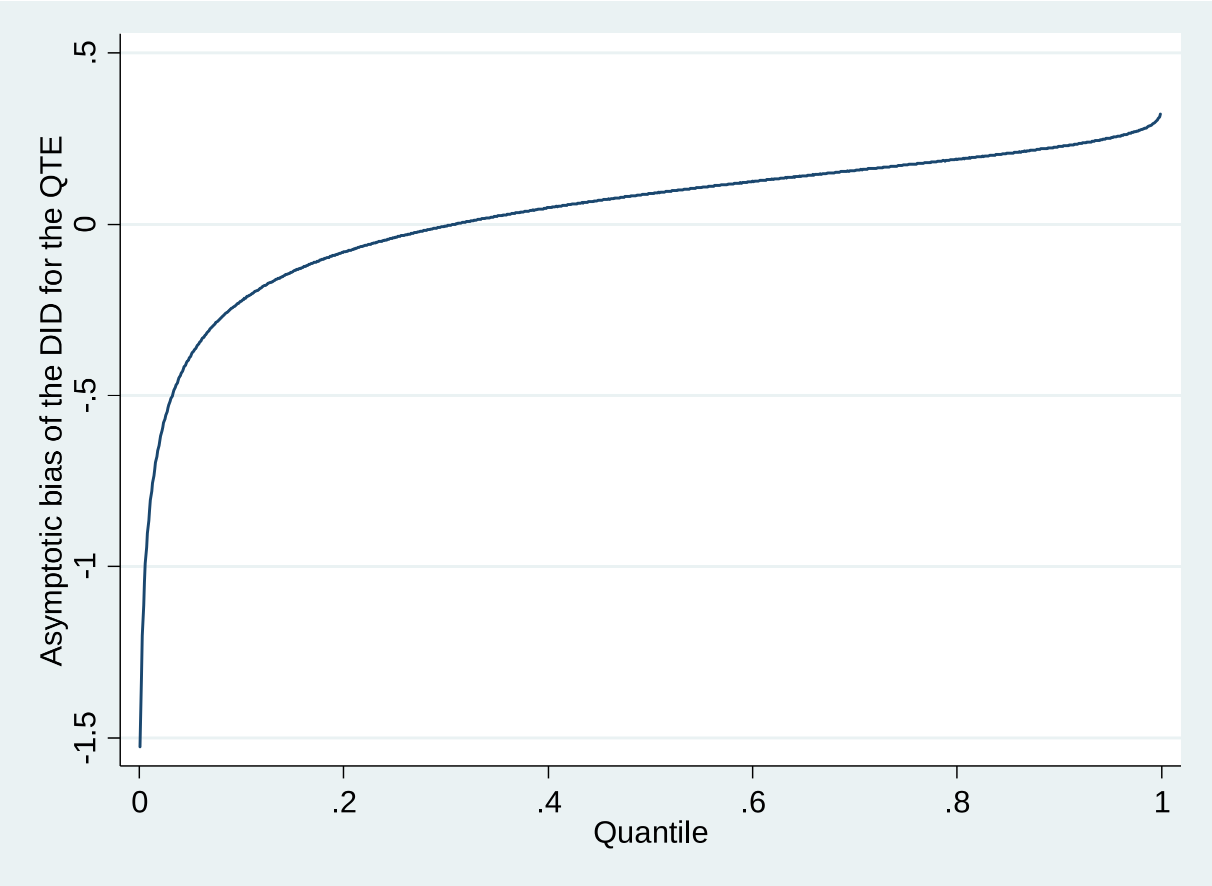


Figure 2: The bias of the DID estimator for the QTE.

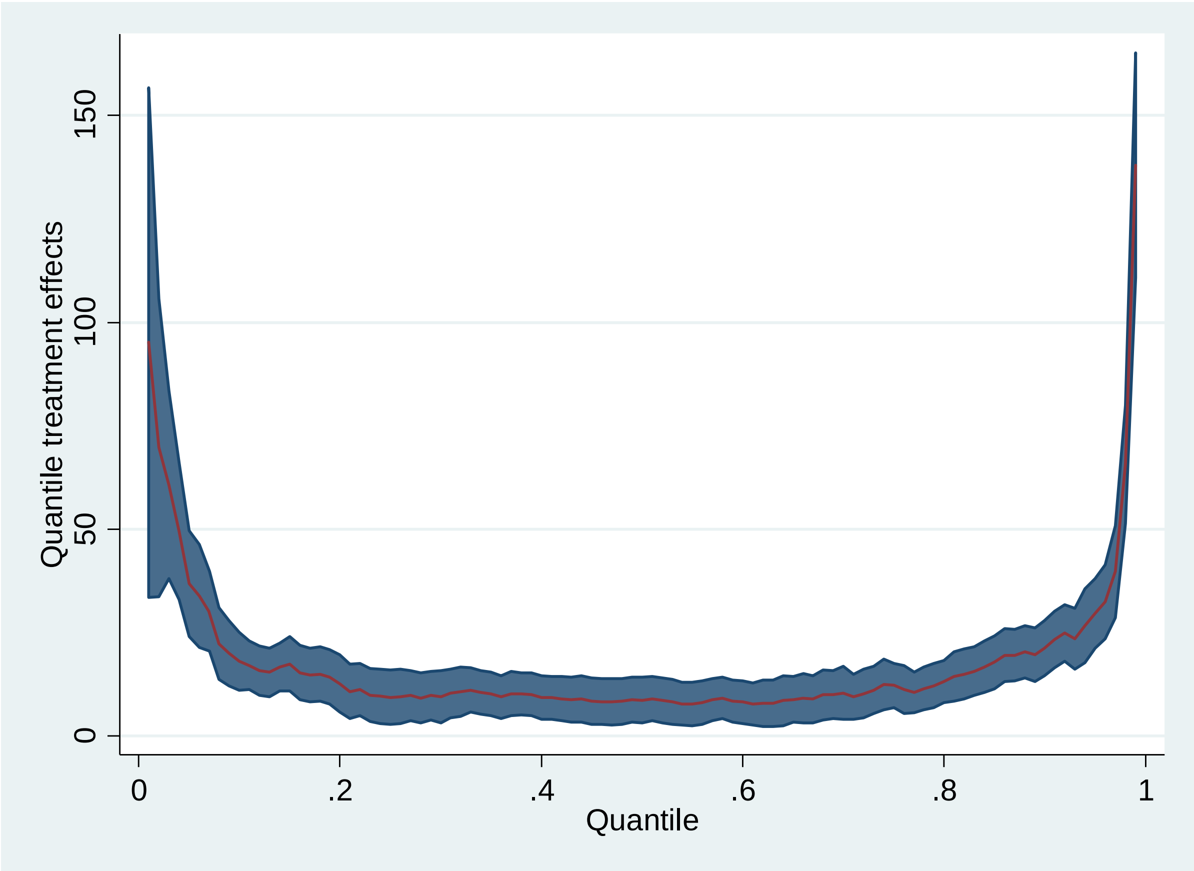


Figure 3: Quantile treatment effect of food stamps on birthweight

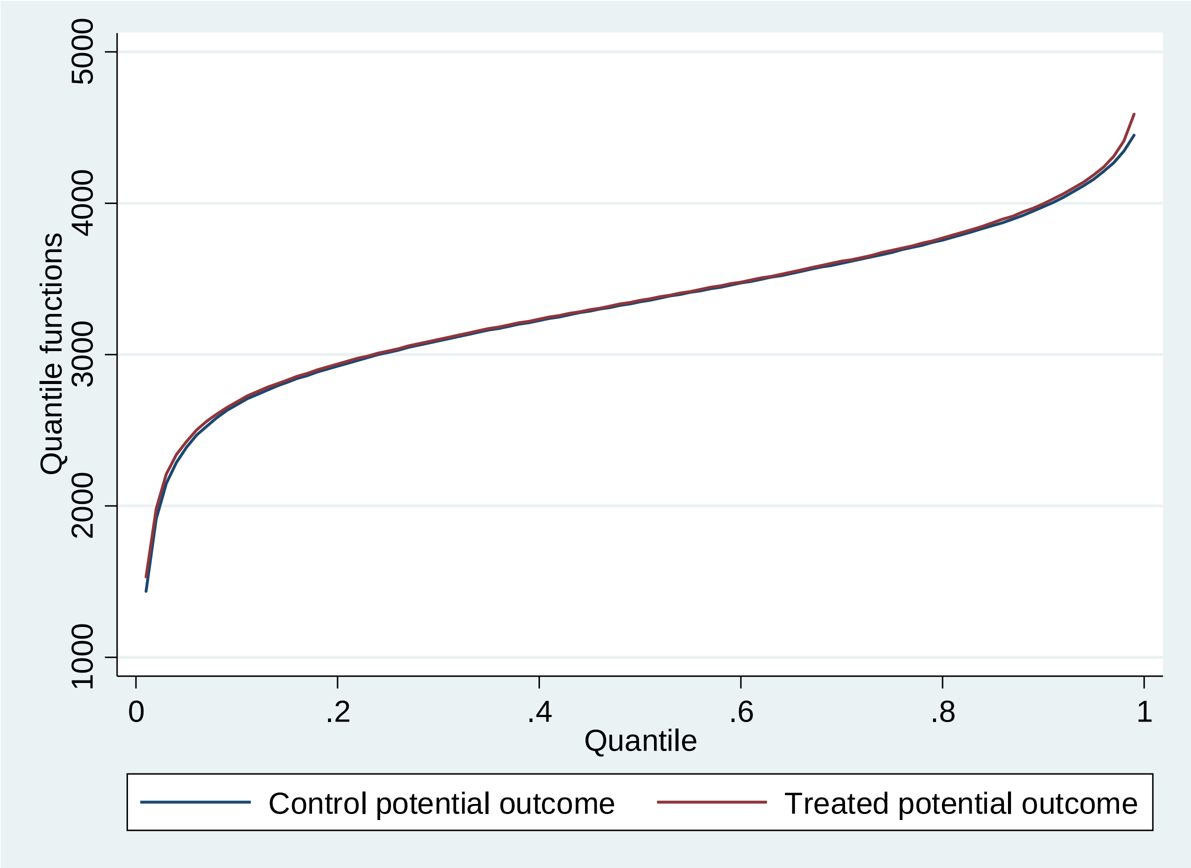


Figure 4: Quantile functions of the control and treated potential outcomes

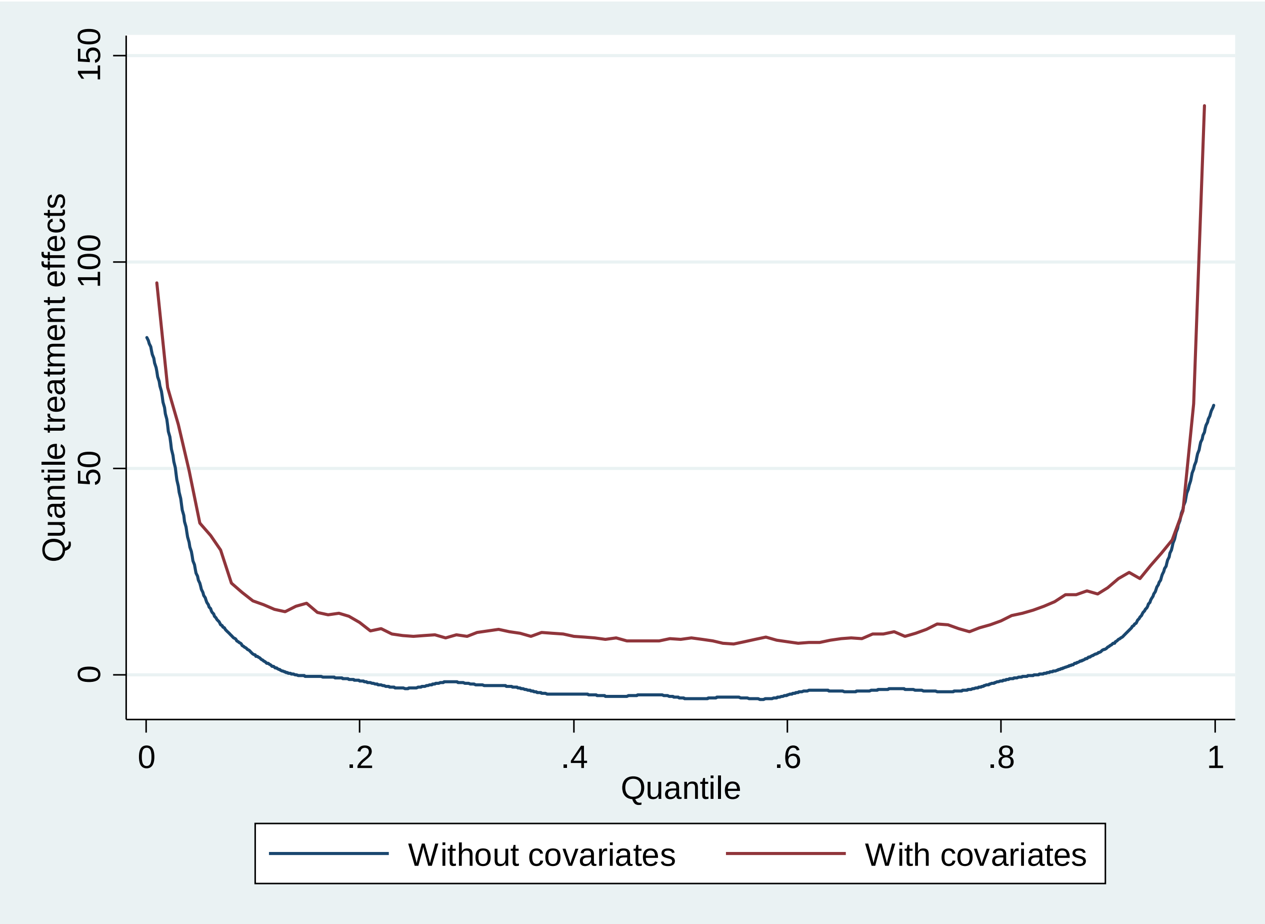


Figure 5: Quantile treatment effects estimated without and with covariates

# References


Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011): "Inside the war on poverty: the impact of food stamps on birth outcomes," *The Review of Economics and Statistics*, 93(2), 387–403.

Altonji, J., and R. Blank (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, pp. 3143–3259. North-Holland, New York.

Athey, S., and G. W. Imbens (2006): "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 74(2), 431–497.

Bitler, M., T. Domina, and H. Hoynes (2014): "Experimental Evidence On Distributional Effects of Head Start," working paper.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and probability curves without crossing," *Econometrica*, 78(3), 1093–1125.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2009): "Inference on counterfactual distributions," Cemmap working paper 09/09.

——— (2013): "Inference on counterfactual distributions," *Econometrica*, 81(6), 2205–2268.

Chernozhukov, V., and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

de Chaisemartin, C., and X. D'Haultfoeuille (2014): "Fuzzy Changes-in Changes," Discussion paper, Competitive Advantage in the Global Economy (CAGE).

Fortin, N. M., and T. Lemieux (1998): "Rank regressions, wage distributions, and the gender gap," *Journal of Human Resources*, pp. 610–643.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.

Koenker, R., and G. Bassett (1978): "Regression quantiles," *Econometrica*, 46, 33–50.

Lechner, M. (2011): *The estimation of causal effects by difference-in-difference methods*. Now.

Matzkin, R. L. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71(5), 1339–1375.

Melly, B. (2014): "Quantile regression in Stata: algorithms, pointwise and functional inference," mimeo, University of Bern.

Melly, B., and G. Santangelo (2014): "The evolution of the gender wage gap: 1968-2008," working paper.

Portnoy, S. (1991): "Asymptotic behavior of the number of regression quantile breakpoints," *SIAM journal on scientific and statistical computing*, 12(4), 867–883.

Strittmatter, A. (2014): "Comparing Distributional Policy Parameters between Populations with Different Outcome Structures," Discussion paper, University of St. Gallen, School of Economics and Political Science.

Van Der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence*. Springer.

Wüthrich, K. (2014): "A Comparison of two Quantile Models with Endogeneity," Discussion paper, Universitaet Bern, Departement Volkswirtschaft.