Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2014-18

Fuzzy Changes-in-Changes

C. DE CHAISEMARTIN[1]
X. D'HAULTFOEUILLE[2]

June, 2014

1 Warwick University. clement.de-chaisemartin@warwick.ac.uk
2 CREST-LMI. xavier.dhaultfoeuille@ensae.fr

Fuzzy Changes-in-Changes*

Clément de Chaisemartin† Xavier D'Haultfœuille‡

June 24, 2014

Abstract

The changes-in-changes model extends the widely used difference-in-differences to situations where outcomes may evolve heterogeneously. Contrary to difference-in-differences, this model is invariant to the scaling of the outcome. This paper develops an instrumental variable changes-in-changes model, to allow for situations in which perfect control and treatment groups cannot be defined, so that some units may be treated in the "control group", while some units may remain untreated in the "treatment group". This is the case for instance with repeated cross sections, if the treatment is not tied to a strict rule. Under a mild strengthening of the changes-in-changes model, treatment effects in a population of compliers are point identified when the treatment rate does not change in the control group, and partially identified otherwise. We show that simple plug-in estimators of treatment effects are asymptotically normal and that the bootstrap is valid. Finally, we use our results to reanalyze findings in Field (2007) and Duflo (2001).

Keywords: differences-in-differences, changes-in-changes, imperfect compliance, instrumental variables, quantile treatment effects, partial identification.

JEL Codes: C21, C23

*We are very grateful to Esther Duflo and Erica Field for sharing their data with us. We also want to thank Alberto Abadie, Joshua Angrist, Stéphane Bonhomme, Marc Gurgand, Guido Imbens, Thierry Magnac, Blaise Melly, Roland Rathelot, Bernard Salanié, Frank Vella, Fabian Waldinger, participants at the 7th IZA Conference on Labor Market Policy Evaluation, North American and European Summer Meetings of the Econometric Society, 11th Doctoral Workshop in Economic Theory and Econometrics and seminar participants at Boston University, Brown University, Columbia University, CREST, MIT, Paris School of Economics and St Gallen University for their helpful comments.

†Warwick University, clement.de-chaisemartin@warwick.ac.uk

‡CREST, xavier.dhaultfoeuille@ensae.fr

1 Introduction

Difference-in-differences (DID) is a popular method to evaluate the effect of a treatment in the absence of experimental data. In its basic version, a "control group" is untreated at two dates, whereas a "treatment group" becomes treated at the second date. If the effect of time is the same in both groups, the so-called common trends assumption, one can measure the effect of the treatment by comparing the evolution of the outcome in both groups. DID only require repeated cross section data, which may explain why this method is so pervasive.

Notwithstanding, the common trends assumption raises a number of concerns. First, it is unlikely to hold if the effect of time is heterogenous. Suppose for instance that one studies the effect of job training on wages, using data where low-wage workers benefit from job training after a given date. If high wages increase more than low wages, the common trends assumption fails to hold. Second, the common trends assumption is not invariant to monotonic transformations of the outcome. It requires that the effect of time and group on the outcome be additively separable, which cannot be true for both the outcome and its logarithm.

To deal with these problems, Athey & Imbens (2006) consider a nonlinear extension of difference-in-differences, the changes-in-changes (CIC) model.[1] It relies on the assumption that a control and a treatment unit with the same outcome at the first period would also have had the same outcome at the second period if the treatment unit had then not been treated. Hereafter, we refer to this condition as the common change assumption. This condition allows for heterogeneous effects of time, and it is invariant to monotonic transforms of the outcome.

However, many natural experiments cannot be analyzed within the standard DID or CIC framework. They do not lead to a sharp change in treatment rate for any group defined by a set of observable characteristics, but only to a larger increase of the treatment rate in some groups than in others. With panel data at hand, the analyst could define the treatment group as units going from non treatment to treatment between the two periods, and the control group as units remaining untreated at the two periods. But this definition of groups might violate the common trends assumption. Units choosing to go from non treatment to treatment between the two periods might do so because they experience different trends in outcomes.

In such settings, the standard practice is to use linear instrumental variable (IV) regressions to estimate treatment effects. A good example is Duflo (2001), who uses a school construction program in Indonesia to measure returns to education. Many schools were constructed in districts where there were few schools previous to the program, while few schools were constructed in districts which already had many schools. She uses the first group of districts as a treatment group, and the second as a control group. Because more schools were constructed

[1] Their estimator is closely related to an estimator proposed by Juhn et al. (1993) and Altonji & Blank (2000) to decompose the Black-White wage differential into changes in the returns to skills and changes in the relative skill distribution.

in treatment districts, years of schooling increased more there. The author then estimates returns to schooling through an IV regression in which time and group fixed effects are used as included instruments for treatment, while the excluded instrument is the interaction of time and group. The resulting coefficient for treatment in this "IV-DID" regression is the ratio of the DID on the outcome and on treatment, which is sometimes referred to as the Wald-DID. Other examples of papers estimating "IV-DID" regressions include Burgess & Pande (2005), Lochner & Moretti (2004), Field (2007), and Akerman et al. (2013), among many others.

We start showing that IV-DID relies on substantially stronger assumptions than DID. On top of standard common trends assumptions, IV-DID also requires that the effect of the treatment be homogeneous in the treatment and control groups. Assume for instance that the effect of the treatment strictly positive in both groups, but twice as large in the control than in the treatment group. Assume also that the treatment rate increased twice as much in the treatment than in the control group. Then, the Wald-DID is equal to 0: the lower increase of the treatment rate in the control group is exactly compensated by the fact that the effect of the treatment is higher in this group. The Wald-DID does not estimate the effect of the treatment in any of the two groups, or a weighted average of the two, because the effect of the treatment is different in the two groups.

Therefore, we study an instrumental variable changes-in-changes (IV-CIC) model which circumvents the shortcomings of IV-DID. This model does not require common trend assumptions, is invariant to monotonic transforms of the outcome, and does not impose homogeneity of the effect of the treatment in the treatment and in the control groups. It combines an increasing production function for the outcome as in Athey & Imbens (2006), and a latent index model for treatment choice in the spirit of Vytlacil (2002). Relative to Athey & Imbens (2006), the main supplementary ingredient we impose is a strengthening of the common change assumption. Formally, we impose that the unobserved terms in the outcome and treatment equations jointly satisfy the common change assumption. Importantly, this allows for endogenous selection, including Roy models where potential outcomes evolve heterogeneously.

In this framework, we show that the marginal distributions of potential outcomes for compliers are point identified if the treatment rate remains constant in the control group, and partially identified otherwise. The intuition for this result goes as follows. When the treatment rate is constant in the control group, any change in the distribution of the outcome of this group can be attributed to time. By the common change assumption, time has the same effect in both groups among individuals with the same outcome. We can therefore use the control group to identify the effect of time, and remove this effect in the treatment group. Any remaining change in the distribution of the outcome in the treatment group can then be attributed to the increase in treatment rate it experienced over time. Thus, the marginal distributions of potential outcomes for compliers are identified. But when the treatment rate is not constant in the control group, the evolution of the outcome in this group may stem both from the

effect of time and from the change in the treatment rate. Therefore, the effect of time is only partially identified, which in turn implies that the marginal distributions of potential outcomes for compliers are partially identified as well. We exhibit bounds on these distributions, and show that they are sharp under testable monotonicity conditions. The smaller the change of the treatment rate in the control group, the tighter the bounds.

We also develop inference on average and quantile treatment effects. Using the functional delta method, we show that simple plug-in estimators of treatment effects in the fully identified case, and of the bounds in the partially identified one, are asymptotically normal under mild conditions. Because the variance takes a complicated form, the bootstrap is convenient to use here, and we prove that it is consistent.

Finally, we apply our results to three different data sets. We first revisit Field (2007), who studies the effect of granting property titles to urban squatters on their labor supply. As the treatment rate is stable in the comparison group used by the author, we are in the point identified case. Our IV-CIC model allows us to study distributional effects of the treatment. We show that property rights have a stronger relative effect on households with a low initial labor supply. We then study the effect of a new medical treatment to ease smoking cessation. The treatment rate slightly increases in our comparison group. Therefore, we are in the partially identified case but our bounds are tight. We show that this new treatment reduces the share of smokers who fail to quit and remain heavy smokers. Finally, we revisit results in Duflo (2001) on returns to education. The treatment rate substantially changes in the comparison group used by the author, so our bounds are wide and uninformative. Our IV-CIC model does not allow us to draw informative conclusions on returns to education from the natural experiment studied by the author.

Researchers must therefore find a control group in which the treatment rate is stable over time to point identify treatment effects under our non linear IV-CIC model. This might be possible to achieve when a group is excluded from treatment at both dates, when a policy is extended to a previously ineligible group, or when a program or a technology previously available in some geographic areas is extended to others (see e.g. Field, 2007, or Akerman et al., 2013). When exposure to treatment slightly changes in the control group, researchers can still use our model to derive tight bounds for treatment effects. When exposure to treatment substantially changes in the control group, using our IV-CIC model will result in wide and uninformative bounds. In such instances, point identification can still be achieved using IV-DID, at the expense of imposing more stringent conditions.

Besides Athey & Imbens (2006), our paper is related to several papers in the literature. Blundell et al. (2004) and Abadie (2005) consider a conditional version of the common trend assumption, and adjust for covariates using propensity score methods. Donald & Lang (2007) and Manski & Pepper (2012) allow for some variations in the way time affects the control and treatment groups, provided these variations satisfy some restrictions. Bonhomme & Sauder

(2011) consider a linear model allowing for heterogeneous effects of time, and show how it can be identified using an instrument. D'Haultfoeuille et al. (2013) study the possibly nonlinear effects of a continuous treatment using repeated cross sections. de Chaisemartin (2013) studies the identifying assumptions underlying IV-DID regressions.

The remainder of the paper is organized as follows. In section 2 we consider a toy model to convey the point that IV-DID might not capture the effect of the treatment if this effect is not the same in the treatment and in the control groups. In Section 3, we introduce our IV-CIC model. Section 4 is devoted to identification. Section 5 deals with inference. In section 6 we apply our results to the three aforementioned data sets. Section 7 concludes. The appendix gathers the main proofs. Due to a concern for brevity, a number of extensions and proofs are deferred to a web appendix (see de Chaisemartin & D'Haultfoeuille, 2014). In this appendix, we show how our framework can accommodate for covariates, how it can be extended to settings with many periods and many groups, and how our model can be tested.

2 A cautionary tale of IV-DID

We are interested in the effect of a binary treatment D on some continuous outcome. $Y(1)$ and $Y(0)$ denote the two potential outcomes of the same individual with and without treatment. The observed outcome is $Y = DY(1) + (1-D)Y(0)$. Let $T \in \{0,1\}$ denote time and let G be a dummy equal to 1 for subjects in the treatment group. We consider a fuzzy setting, where $D \neq G \times T$. Some units may be treated in the control group or at period 0, and all units are not necessarily treated in the treatment group at period 1. But we assume that at period 1, individuals in the treatment group receive extra incentives to get treated. We model this by introducing the binary instrument $Z = T \times G$. The two corresponding potential treatments, $D(1)$ and $D(0)$, stand for the treatment an individual would choose to receive with and without this supplementary incentive. The observed treatment is $D = ZD(1) + (1-Z)D(0)$.

In such settings, a first strategy to estimate the effect of the treatment is to run an IV regression of the outcome on the treatment with time and group as included instruments, and the interaction of the two as the excluded instrument. However, this estimation strategy relies on strong assumptions. It requires common trend assumptions for both potential outcomes and treatments, and it also requires that the effect of the treatment be homogeneous in the treatment and in the control groups (see de Chaisemartin, 2013).

To convey this last point, let us consider a simple model in which the coefficient arising from the aforementioned regression might not have any causal interpretation if the effect of the treatment differs in the two groups. For any random variable X, let

$$DID_X = E(X|T=1,G=1) - E(X|T=0,G=1) - (E(X|T=1,G=0) - E(X|T=0,G=0)).$$

The estimand arising from the "IV-DID" regression described above is the Wald-DID, which

is equal to $\frac{DID_Y}{DID_D}$. Now, assume that the causal model generating the potential outcomes is

$$Y(d) = \alpha + \beta T + \gamma G + \delta_0 d(1-G) + \delta_1 dG + U, \tag{1}$$

where U is a random variable with mean 0, supposed to be mean independent of (T, G):

$$E(U|T, G) = 0. \tag{2}$$

In this model, the effect of time is assumed to be constant, while the effect of the treatment is allowed to vary across groups. If potential outcomes follow the model defined by Equations (1) and (2), it is easy to show that

$$\begin{aligned} DID_Y &= \delta_1 \left(P(D=1|T=1,G=1) - P(D=1|T=0,G=1)\right) \\ &- \delta_0 \left(P(D=1|T=1,G=0) - P(D=1|T=1,G=0)\right). \end{aligned} \tag{3}$$

If the effect of the treatment is the same in the two groups, i.e. if $\delta_0 = \delta_1 = \delta$, Equation (3) implies that the Wald-DID is equal to δ. But if $\delta_0 \neq \delta_1$, the Wald-DID might not have any causal interpretation. Assume for instance that

$$\begin{aligned} &P(D=1|T=1,G=1) - P(D=1|T=0,G=1) > 0, \\ &P(D=1|T=1,G=0) - P(D=1|T=1,G=0) > 0, \\ &\delta_0 > 0, \text{ and } \delta_1 = \delta_0 \times \frac{P(D=1|T=1,G=0) - P(D=1|T=1,G=0)}{P(D=1|T=1,G=1) - P(D=1|T=0,G=1)}. \end{aligned}$$

The Wald-DID is then equal to 0 while every observation in the population has a strictly positive treatment effect. de Chaisemartin (2013) shows that this result extends to more general models with heterogeneous effects of time and of the treatment. In these models, the Wald-DID captures a causal effect if the average effect of the treatment is the same in the treatment and in the control groups, at least among subjects whose treatment status changes over time.

This homogeneity condition might sometimes be a strong assumption. In Duflo (2001), it requires that returns to schooling be the same in districts in which many schools were constructed as in districts in which few schools were constructed. Districts in which many schools were constructed are those where few schools were available previous to the program. Returns to education might be higher in those districts, for instance if they suffer from a shortage of qualified labor. But those districts are probably less developed, so returns to education might also be lower if there are no jobs available for educated workers in these areas.

3 The instrumental variable Changes-in-Changes model

IV-DID relies on strong assumptions. As an alternative, we propose an IV-CIC model that does not require common trends or treatment effect homogeneity assumptions, and whose assumptions have a more straightforward economic interpretation.

Let us first introduce more notations. For any random variables R and S, $R \sim S$ means that R and S have the same probability distribution. $\mathcal{S}(R)$ and $\mathcal{S}(R|S)$ denote respectively the support of R and the support of R conditional on S. As Athey & Imbens (2006), for any random variable R we introduce the corresponding random variables R_{gt} such that

$$R_{gt} \sim R|\, G = g, T = t.$$

Let F_R and $F_{R|S}$ denote the cumulative distribution function (cdf) of R and its cdf conditional on S. For any event A, $F_{R|A}$ is the cdf of R conditional on A. With a slight abuse of notation, $P(A)F_{R|A}$ should be understood as 0 when $P(A) = 0$. For any increasing function F on the real line, we denote by F^{-1} its generalized inverse:

$$F^{-1}(q) = \inf\{x \in \mathbb{R}/F(x) \geq q\}.$$

In particular, F_X^{-1} is the quantile function of X. We adopt the convention that $F_X^{-1}(q) = \inf \mathcal{S}(X)$ for $q < 0$, and $F_X^{-1}(q) = \sup \mathcal{S}(X)$ for $q > 1$. We let $\lambda_d = P(D_{01} = d)/P(D_{00} = d)$ be the ratio of the shares of people receiving treatment d in period 1 and period 0 in the control group. For instance, $\lambda_0 > 1$ when the share of untreated observations increases in the control group between period 0 and 1. $\lambda_0 > 1$ implies that $\lambda_1 < 1$ and conversely. $\mu_d = P(D_{11} = d)/P(D_{10} = d)$ is the equivalent of λ_d for the treatment group.

As in Athey & Imbens (2006), we consider the following model for the potential outcomes:

$$Y(d) = h_d(U_d, T), \quad d \in \{0; 1\}. \tag{4}$$

We also consider the following assumptions.

Assumption 1 *(Monotonicity)*

$h_d(u, t)$ is strictly increasing in u for all $(d, t) \in \{0, 1\}^2$.

Assumption 2 *(Latent index model for potential treatments)*

$D(z) = 1\{V \geq v_z(T)\}$ with $v_0(t) > v_1(t)$ for $t \in \{0; 1\}$.

Assumption 3 *(Time invariance within groups)*

For $d \in \{0, 1\}$, $(U_d, V) \perp\!\!\!\perp T \,|\, G$.

Remarks on these assumptions are in order. U_d can be interpreted as an ability index. V represents taste for treatment. Our latent index model for potential treatments is the same as in Vytlacil (2002), except that the threshold can depend on time. As shown by Vytlacil (2002), this model implies that the instrument must have a monotonic effect on treatment, as in Imbens & Angrist (1994). As it is formulated in Assumption 2, it also implies that time can affect treatment in only one direction. Actually, all our theorems would remain valid if U_d

and V were indexed by time (and then denoted by U_d^t and V^t), except that we would have to rewrite Assumption 3 as follows: for $d \in \{0,1\}$, $\left(U_d^0, V^0\right)|G \sim \left(U_d^1, V^1\right)|G$. This would allow individual ability and taste for treatment to change over time, provided their distribution remains the same in each group. In this model, time could induce some observations to go from non-treatment to treatment, while having the opposite effect on other observations. In what follows, we do not index U_d and V by time to alleviate the notational burden.

Assumption 3 requires that the joint distribution of ability and propensity for treatment remains stable in each group over time. It implies $U_d \perp\!\!\!\perp T \,|\, G$ and $V \perp\!\!\!\perp T \,|\, G$, which correspond to the time invariance assumption in Athey & Imbens (2006). As a result, Assumptions 1-3 impose a standard CIC model both on Y and D. But Assumption 3 also implies $U_d \perp\!\!\!\perp T \,|\, G, V$, which means that in each group, the distribution of ability among people with a given taste for treatment should not change over time. This is the key supplementary ingredient with respect to the standard CIC model that we are going to use for identification.

This supplementary ingredient is compatible with Roy selection in a model where time has heterogeneous effects on the outcome, provided the treatment does not affect this pattern of heterogeneity. Assume potential treatments follow a Roy model: $D(z) = \mathbb{1}\{Y(1) - Y(0) \geq c(z)\}$. Assume also that

$$Y(d) = U_d + \eta_d T + \gamma U_d T, \tag{5}$$

and that the standard CIC assumption is verified:

$$(U_0, U_1) \perp\!\!\!\perp T|G. \tag{6}$$

Equation (5) allows for different trends in potential outcomes across ability levels. It satisfies Assumption 1 provided $\gamma > -1$. One can then rewrite

$$D(z) = \mathbb{1}\left\{U_1 - U_0 \geq \frac{c(z) - (\eta_1 - \eta_0)T}{1 + \gamma T}\right\}.$$

Assumption 2 is satisfied with $V = U_1 - U_0$ and $v_z(T) = [c(z) - (\eta_1 - \eta_0)T]/(1 + \gamma T)$. $(U_0, U_1) \perp\!\!\!\perp T|G$ then implies that Assumption 3 is satisfied as well. On the contrary, with γ_d instead of γ in Equation (5), Assumption 3 may be violated because then $V = U_1 - U_0 + T(\gamma_1 U_1 - \gamma_0 U_0)$. Therefore, Assumption 3 is compatible with a Roy model in which time can have heterogeneous effects on the outcome across ability levels, provided these heterogeneous effects are not affected by the treatment. This is not an innocuous assumption, but this is an improvement relative to IV-DID which is incompatible with Roy selection and Equation (5).[2]

Hereafter, we refer to Assumptions 1-3 as to the IV-CIC model. Finally, we impose the two following assumptions, which are directly testable in the data.

[2]IV-DID relies on common trend assumptions on potential outcomes and treatments (see de Chaisemartin, 2013). Together with Roy selection and Equation (5), these assumptions imply $U_1 - U_0 \perp\!\!\!\perp G$. This amounts to assuming that groups are as good as randomly assigned, in which case we do not need to resort to a longitudinal analysis to capture treatment effects.

Assumption 4 *(Data restrictions)*

1. $\mathcal{S}(Y_{gt}|D=d) = \mathcal{S}(Y) = [\underline{y}, \overline{y}]$ *with* $(\underline{y}, \overline{y}) \in \overline{\mathbb{R}}^2$*, for* $(g,t,d) \in \{0;1\}^3$.

2. $F_{Y_{gt}|D=d}$ *is strictly increasing and continuous on* $\mathcal{S}(Y)$*, for* $(g,t,d) \in \{0;1\}^3$.

Assumption 5 *(Rank condition)*

$P(D_{11}=1) - P(D_{10}=1) > 0$.

The first condition of Assumption 4 is a common support condition. Athey & Imbens (2006) take a similar assumption and show how to derive partial identification results when it is not verified. Point 2 is satisfied if the distribution of Y is continuous with positive density in each of the eight groups $\times$ period $\times$ treatment status cells. Assumption 5 is a rank condition. Our IV-CIC model requires that the treatment rate changes in at least one group. If it decreases in the two groups we can just consider $1-D$ as the treatment variable.

Before getting to the identification results, it is useful to define five subpopulations of interest. Assumptions 2 and 3 imply that

$$\begin{aligned} P(D_{10}=1) &= P(V \geq v_0(0)|G=1) \\ P(D_{11}=1) &= P(V \geq v_1(1)|G=1). \end{aligned}$$

Therefore, under Assumption 5 $v_0(0) > v_1(1)$. Similarly, if the treatment rate increases (resp. decreases) in the control group, $v_0(0) > v_0(1)$ (resp. $v_0(0) < v_0(1)$). Finally, Assumption 2 implies $v_1(1) \leq v_0(1)$. Let always takers be such that $V \geq v_0(0)$, and let never takers be such that $V < v_1(1)$. Always takers are units who get treated in period 0 even without receiving any incentive for treatment. Never takers are units who do not get treated in period 1 even after receiving an incentive for treatment. Let $TC = \{V \in [\min(v_0(0), v_0(1)), \max(v_0(0), v_0(1)))\}$. TC stands for "time compliers," and represents observations whose treatment status switches between the two periods because of the effect of time. Let $IC = \{V \in [v_1(1), v_0(1))\}$.[3] IC stands for instrument compliers and corresponds to observations which become treated through the effect of Z only. Finally, let $C = \{V \in [v_1(1), v_0(0))\}$. C stands for compliers and corresponds to untreated observations at period 0 who become treated at period 1, through both the effect of Z and time. If the treatment rate increases in the control group, we have $C = IC \cup TC$, while if it decreases we have $C = IC \setminus TC$.

Our identification results focus on compliers. Our parameters of interest are the cdf of $Y(1)$ and $Y(0)$ within this population, as well as their Local Average Treatment Effect (LATE) and Quantile Treatment Effects (QTE). Their LATE and QTE are respectively defined by

$$\begin{aligned} \Delta &= E\left(Y_{11}(1) - Y_{11}(0)|C\right), \\ \tau_q &= F^{-1}_{Y_{11}(1)|C}(q) - F^{-1}_{Y_{11}(0)|C}(q), \quad q \in (0,1). \end{aligned}$$

[3] IC is defined to be empty when $v_0(1) = v_1(1)$.

4 Identification

4.1 Point identification results

We first show that when the treatment rate does not change between the two periods in the control group, the cdf of $Y(1)$ and $Y(0)$ among compliers are identified. Consequently, the LATE and QTE are also point identified. Let $Q_d(y) = F^{-1}_{Y_{01}|D=d} \circ F_{Y_{00}|D=d}(y)$ be the quantile-quantile transform of Y from period 0 to 1 in the control group conditional on $D = d$. This transform maps y at rank q in period 0 into the corresponding y' at rank q as well in period 1. Also, let $Q_D = DQ_1 + (1 - D)Q_0$. Finally, let $H_d(q) = F_{Y_{10}|D=d} \circ F^{-1}_{Y_{00}|D=d}(q)$ be the inverse quantile-quantile transform of Y from the control to the treatment group in period 0 conditional on $D = d$. This transform maps rank q in the control group into the corresponding rank q' in the treatment group with the same value of y.

Theorem 4.1 *If Assumptions 1-5 hold and for $d \in \{0, 1\}$ $P(D_{00} = d) = P(D_{01} = d) > 0$, $F_{Y_{11}(d)|C}(y)$ is identified by*

$$\begin{aligned} F_{Y_{11}(d)|C}(y) &= \frac{P(D_{10} = d)F_{Q_d(Y_{10})|D=d}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)} \\ &= \frac{P(D_{10} = d)H_d \circ F_{Y_{01}|D=d}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)}. \end{aligned}$$

This implies that Δ and τ_q are also identified. Moreover,

$$\Delta = \frac{E(Y_{11}) - E(Q_D(Y_{10}))}{E(D_{11}) - E(D_{10})}.$$

This theorem combines ideas from Imbens & Rubin (1997) and Athey & Imbens (2006). We seek to recover the distribution of, say, $Y(1)$ among compliers in the treatment $\times$ period 1 cell. When the treatment rate does not change in the control group, $v_0(0) = v_0(1)$. As a result, there are no time compliers, and compliers are merely instrument compliers. To recover the distribution of $Y(1)$ among them, we start from the distribution of Y among all treated observations of this cell. As shown in Table 1, those include both compliers and always takers. Consequently, we must "withdraw" from this distribution the cdf of $Y(1)$ among always takers, exactly as in Imbens & Rubin (1997). But this last distribution is not observed. To reconstruct it, we adapt the ideas in Athey & Imbens (2006) and apply the quantile-quantile transform from period 0 to 1 among treated observations in the control group to the distribution of $Y(1)$ among always takers in the treatment group in period 0.

Intuitively, the quantile-quantile transform uses a double-matching to reconstruct the unobserved distribution. Consider an always taker in the treatment $\times$ period 0 cell. She is first matched to an always taker in the control $\times$ period 0 cell with same y. Those two always takers are observed at the same period of time and are both treated. Therefore, under Assumption 1 they must have the same u_1. Second, the control $\times$ period 0 always taker is matched to

her rank counterpart among always takers of the control × period 1 cell. We denote y^* the outcome of this last observation. Because $U_1 \perp\!\!\!\perp T|G, V \geq v_0(0)$, those two observations must also have the same u_1. Consequently, $y^* = h_1(u_1, 1)$, which means that y^* is the outcome that the treatment × period 0 cell always taker would have obtained in period 1.

<table>
<tr><th></th><th>Period 0</th><th>Period 1</th></tr>
<tr><td rowspan="2">Control Group</td><td>30% treated: Always Takers</td><td>30% treated: Always Takers</td></tr>
<tr><td>70% untreated: Never Takers and Compliers</td><td>70% untreated: Never Takers and Compliers</td></tr>
<tr><td rowspan="2">Treatment Group</td><td>20% treated: Always Takers</td><td rowspan="2">65% treated: Always Takers and Compliers</td></tr>
<tr><td>80% untreated: Never Takers and Compliers</td></tr>
<tr><td></td><td></td><td>35% Untreated: Never Takers</td></tr>
</table>

Table 1: Populations of interest when $P(D_{00} = 0) = P(D_{01} = 0)$.

Note that our LATE estimand is similar to the LATE estimand in Imbens & Angrist (1994).

$$\Delta = \frac{E(Y|G = 1, T = 1) - E(Q_D(Y)|G = 1, T = 0)}{E(D|G = 1, T = 1) - E(D|G = 1, T = 0)}.$$

This is the standard Wald ratio in the treatment group with T as the instrument, except that we have $Q_D(Y)$ instead of Y in the second term of the numerator. $Q_D(Y)$ accounts for the fact time is not a standard instrument. It influences selection into treatment, a condition all instruments must satisfy, but it is also directly included in the potential outcome equations, meaning that it violates the standard exclusion restriction. When the treatment rate is stable in the control group, we can identify this direct effect by looking at how the distribution of the outcome evolves in this group. We can then net out this direct effect in the treatment group, so as to recover the effect of time on the outcome which only goes through the effect of time on treatment. This is exactly what $Q_D(.)$ does.

Under Assumptions 1-5, the LATE and QTE for compliers are point identified when $0 < P(D_{00} = 0) = P(D_{01} = 0) < 1$, but not in the extreme cases where $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$. For instance, when $P(D_{00} = 1) = P(D_{01} = 1) = 1$, $F_{Y_{11}(1)|C}$ is identified by Theorem 4.1, but $F_{Y_{11}(0)|C}$ is not. Such situations are likely to arise in practice, for instance when a policy is extended to a previously ineligible group, or when a program or a technology previously available in some geographic areas is extended to others (see Subsection 6.1 below). We therefore consider a mild strengthening of our assumptions under which both $F_{Y_{11}(0)|C}$ and $F_{Y_{11}(1)|C}$ are point identified in those instances.

Assumption 6 *(Common effect of time on both potential outcomes)* $h_0(u,t) = h_1(u,t) = h(u,t)$ *for every* $(u,t) \in \mathcal{S}(U) \times \{0,1\}$.

Assumption 6 requires that the effect of time be the same on both potential outcomes. It implies that two observations with the same outcome in period 0 will also have the same outcome in period 1 if they do not switch treatment between the two periods, even if they do not share the same treatment at period 0. Under this assumption, if $P(D_{00} = 1) = P(D_{01} = 1) = 1$, changes in the distribution of Y in the control group over time allow us to identify the effect of time both on $Y(0)$ and $Y(1)$, hence allowing us to recover $F_{Y_{11}(0)|C}$ and $F_{Y_{11}(1)|C}$.

Theorem 4.2 *If Assumptions 1-6 hold and* $P(D_{00} = d) = P(D_{01} = d) = 0$ *for some* $d \in \{0,1\}$, $F_{Y_{11}(d)|C}(y)$ *and* $F_{Y_{11}(1-d)|C}(y)$ *are identified by*

$$\begin{aligned} F_{Y_{11}(d)|C}(y) &= \frac{P(D_{10} = d)F_{Q_{1-d}(Y)_{10}|D=d}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)} \\ F_{Y_{11}(1-d)|C}(y) &= \frac{P(D_{10} = 1-d)F_{Q_{1-d}(Y)_{10}|D=1-d}(y) - P(D_{11} = 1-d)F_{Y_{11}|D=1-d}(y)}{P(D_{10} = 1-d) - P(D_{11} = 1-d)}. \end{aligned}$$

This implies that Δ *and* τ_q *are also identified. Moreover,*

$$\Delta = \frac{E(Y_{11}) - E(Q_{1-d}(Y_{10}))}{E(D_{11}) - E(D_{10})}.$$

A last situation worth noting is when the treatment rate is equal to 0 at both dates in the control group, and is also equal to 0 in the first period in the treatment group. This is a special case of Theorem 4.2, but in such instances we can actually identify the model under fewer assumptions. To see this, note that in such situations,

$$F_{Y_{11}(1)|C} = F_{Y_{11}|D=1} \tag{7}$$

because there are no always takers in the treatment group. Therefore, we only need to recover $F_{Y_{11}(0)|C}$. But since the distribution of $Y_{11}(0)$ among never takers is identified by $F_{Y_{11}|D=0}$, under Assumption 2 we only need to recover $F_{Y_{11}(0)}$. This can be achieved under the standard changes-in-changes assumptions, as the control group remains fully untreated at both dates.

4.2 Partial identification

When $P(D_{00} = d) = P(D_{01} = d)$, $F_{Y_{11}(d)|C}$ is identified under Assumptions 1-5 or Assumptions 1-6. We show now that if this condition is not verified, the functions $F_{Y_{11}(d)|C}$ are partially identified. For that purpose, we must distinguish between two cases.

First, when $P(D_{00} = d) \neq P(D_{01} = d)$ and $P(D_{00} = d) > 0$, the second matching described in the previous section collapses, because treated (resp. untreated) observations in the control group are no longer comparable in period 0 and 1. For instance, when the treatment rate

increases in the control group, treated observations in the control group include only always takers in period 0. In period 1 they also include time compliers, as is shown in Table 2. Therefore, we cannot match period 0 and period 1 observations on their rank anymore. However, under Assumption 3 the respective weights of time compliers and always takers in period 1 are known. We can therefore derive best and worst case bounds for the distribution of the outcome for always takers in period 1, and match period 0 observations to their best and worst case rank counterparts.

<table>
<tr><th></th><th>Period 0</th><th>Period 1</th></tr>
<tr><td rowspan="2">Control Group</td><td>30% treated: Always Takers</td><td>35% treated: Always Takers and Time Compliers</td></tr>
<tr><td>70% untreated: Never Takers, Instrument Compliers and Time Compliers</td><td>65% untreated: Never Takers and Instrument Compliers</td></tr>
<tr><td rowspan="2">Treatment Group</td><td>25% treated: Always Takers</td><td rowspan="2">60% treated: Always Takers, Instrument Compliers and Time Compliers</td></tr>
<tr><td>75% untreated: Never Takers, Instrument Compliers and Time Compliers</td></tr>
<tr><td></td><td></td><td>40% Untreated: Never Takers</td></tr>
</table>

Table 2: Populations of interest when $P(D_{01} = 1) > P(D_{00} = 1)$.

Second, when $P(D_{00} = d) = 0$ the first matching collapses. For instance, if $P(D_{00} = 1) = 0$, there are no treated observations in the control group in period 0 to which treated observations in the treatment group in period 0 can be matched. Still, the cdf of Y among treated observations in the treatment $\times$ period 1 cell writes as a weighted average of the cdf of $Y(d)$ among compliers and always or never takers. We can use this fact to bound $F_{Y_{11}(d)|C}$.

The derivation of our bounds relies on the following lemma which relates $F_{Y_{11}(d)|C}$ to observed distributions and one unidentified cdf.

Lemma 4.1 *If Assumptions 1-5 hold, then:*

- *If* $P(D_{00} = d) > 0$,

$$F_{Y_{11}(d)|C}(y) = \frac{P(D_{10} = d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y) + (1 - \lambda_d)F_{Y_{01}(d)|TC}(y)) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)}.$$

- *If* $P(D_{00} = d) = 0$,

$$F_{Y_{11}(d)|C}(y) = \frac{P(D_{10} = d)F_{Y_{11}(d)|(2d-1)V>(2d-1)v_0(0)}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)}.$$

When $P(D_{00}=d)>0$, we need to bound $F_{Y_{01}(d)|TC}$ to derive bounds on $F_{Y_{11}(d)|C}$. To do so, we must take into account the fact that $F_{Y_{01}(d)|TC}$ is related to two other cdf. To alleviate the notational burden, let $T_d = F_{Y_{01}(d)|TC}$, $C_d(T_d) = F_{Y_{11}(d)|C}$, $G_0(T_0) = F_{Y_{01}(0)|V<v_0(0)}$ and $G_1(T_1) = F_{Y_{01}(1)|V\geq v_0(0)}$. With those notations, we have

$$\begin{aligned} G_d(T_d) &= \lambda_d F_{Y_{01}|D=d} + (1-\lambda_d)T_d \\ C_d(T_d) &= \frac{P(D_{10}=d)H_d \circ G_d(T_d) - P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}. \end{aligned}$$

Let $M_0(x)=\max(0,x)$ and $m_1(x)=\min(1,x)$, and let

$$\begin{aligned} \underline{T}_d &= M_0\left(m_1\left(\frac{\lambda_d F_{Y_{01}|D=d} - H_d^{-1}(\mu_d F_{Y_{11}|D=d})}{\lambda_d - 1}\right)\right), \\ \overline{T}_d &= M_0\left(m_1\left(\frac{\lambda_d F_{Y_{01}|D=d} - H_d^{-1}(\mu_d F_{Y_{11}|D=d} + (1-\mu_d))}{\lambda_d - 1}\right)\right). \end{aligned}$$

$\underline{T}_d$ (resp. $\overline{T}_d$) is the lowest (resp. highest) possible value of T_d compatible with the fact that T_d, $G_d(T_d)$ and $C_d(T_d)$ should all be included between 0 and 1. We can therefore bound $F_{Y_{11}(d)|C}$ by $C_d(\underline{T}_d)$ and $C_d(\overline{T}_d)$, but these bounds can be further improved by remarking that $F_{Y_{11}(d)|C}$ must be increasing. Therefore, we define our bounds as:

$$\begin{aligned} \underline{B}_d(y) &= \sup_{y'\leq y} C_d\left(\underline{T}_d\right)(y'), \\ \overline{B}_d(y) &= \inf_{y'\geq y} C_d\left(\overline{T}_d\right)(y'). \end{aligned} \tag{8}$$

If the support of the outcome is unbounded, $\underline{B}_0$ and $\overline{B}_0$ are proper cdf when $\lambda_0>1$, but they are defective when $\lambda_0<1$. When $\lambda_0<1$, time compliers belong to the group of treated observations in the control $\times$ period 1 cell (cf. Table 2). Their $Y(0)$ is not observed in period 1, so the data does not impose any restriction on $F_{Y_{01}(0)|TC}$: it could be equal to 0 or to 1, hence the defective bounds. On the contrary, when $\lambda_0>1$, time compliers belong to the group of untreated observations in the control $\times$ period 1 cell. Moreover, under Assumption 3, we know that they account for $100(1-1/\lambda_0)\%$ of this group. Consequently, the data imposes some restrictions on $F_{Y_{01}(0)|TC}$. For instance, we must have

$$F_{Y_{01}|D=0,Y\geq\alpha} \leq F_{Y_{01}(0)|TC} \leq F_{Y_{01}|D=0,Y\leq\beta},$$

where $\alpha = F^{-1}_{Y_{01}|D=0}(1/\lambda_0)$ and $\beta = F^{-1}_{Y_{01}|D=0}(1-1/\lambda_0)$. $\underline{B}_0$ and $\overline{B}_0$ are trimming bounds in the spirit of Horowitz & Manski (1995) when $\lambda_0>1$, but not when $\lambda_0<1$, which is the reason why they are defective then. On the contrary, $\underline{B}_1$ and $\overline{B}_1$ are always proper cdf, while we could have expected them to be defective when $\lambda_0>1$. This asymmetry stems from the fact that when $\lambda_0>1$, time compliers do not belong to our population of interest ($C = IC \setminus TC$), while when $\lambda_0<1$, they belong to it ($C = IC \cup TC$).

When $P(D_{00} = d) = 0$, our bounds are much simpler. We simply bound $F_{Y_{11}(1)|(2d-1)V \geq (2d-1)v_0(0)}$ by 0 and 1. For $d = 1$, this yields

$$\begin{aligned} \underline{B}_1(y) &= M_0 \left(\frac{P(D_{11} = d)F_{Y_{11}|D=d} - P(D_{10} = d)}{P(D_{11} = d) - P(D_{10} = d)} \right) \\ \overline{B}_1(y) &= m_1 \left(\frac{P(D_{11} = d)F_{Y_{11}|D=d}}{P(D_{11} = d) - P(D_{10} = d)} \right). \end{aligned} \tag{9}$$

For $d = 0$, this yields trivial bounds: $\underline{B}_0(y) = 0$ and $\overline{B}_0(y) = 1$.

In the following theorem, we show that $\underline{B}_d$ and $\overline{B}_d$ are indeed bounds for $F_{Y_{11}(d)|C}$. We also consider whether these bounds are sharp or not. Hereafter, we say that $\underline{B}_d$ is sharp if there exists a sequence of cdf $(G_k)_{k\in\mathbb{N}}$ such that supposing $F_{Y_{11}(d)|C} = G_k$ is compatible with both the data and the model, and for all y, $\lim_{k\to\infty} G_k(y) = \underline{B}_d(y)$. We establish that $\underline{B}_d$ and $\overline{B}_d$ are sharp under Assumption 7 below. Note that this assumption is testable from the data.

Assumption 7 *(Increasing bounds)*

For $(d, g, t) \in \{0, 1\}^3$, $F_{Y_{gt}|D=d}$ is continuously differentiable, with positive derivative on $\overset{\circ}{\mathcal{S}}(Y)$. Moreover, either (i) $P(D_{00} = d) = 0$ or (ii) $\underline{T}_d, \overline{T}_d, G_d(\underline{T}_d)$ and $G_d(\overline{T}_d)$ are increasing and $C_d(\underline{T}_d)$ and $C_d(\overline{T}_d)$ are strictly increasing.

We can finally state the theorem summarizing the discussion on partial identification.

Theorem 4.3 *If Assumptions 1-5 hold, we have*

$$\underline{B}_d(y) \leq F_{Y_{11}(d)|C}(y) \leq \overline{B}_d(y).$$

Moreover, if Assumption 7 holds, $\underline{B}_d(y)$ and $\overline{B}_d(y)$ are sharp.

A consequence of Theorem 4.3 is that QTE and LATE are partially identified when $P(D_{00} = 0) \neq P(D_{01} = 0)$ or $P(D_{00} = 0) \in \{0, 1\}$. To ensure that the bounds on the LATE are well defined, we impose the following technical condition.

Assumption 8 *(Existence of moments)*

$\int |y| d\overline{B}_1(y) < +\infty$ *and* $\int |y| d\underline{B}_1(y) < +\infty$.

Corollary 4.4 *If Assumptions 1-5 and 8 hold and $P(D_{00} = 0) \neq P(D_{01} = 0)$, Δ and τ_q are partially identified, with*

$$\int y d\overline{B}_1(y) - \int y d\underline{B}_0(y) \leq \Delta \leq \int y d\underline{B}_1(y) - \int y d\overline{B}_0(y),$$

$$\max(\overline{B}_1^{-1}(q), \underline{y}) - \min(\underline{B}_0^{-1}(q), \overline{y}) \leq \tau_q \leq \min(\underline{B}_1^{-1}(q), \overline{y}) - \max(\overline{B}_0^{-1}(q), \underline{y}).$$

Moreover, these bounds are sharp under Assumption 7.

When $\mathcal{S}(Y)$ is unbounded and $\lambda_0 < 1$, our bounds on Δ will be infinite because our bounds for the cdfs of $Y(1)$ and $Y(0)$ of compliers are defective. Our bounds on τ_q will also be infinite for low and high values of q. On the contrary, when $\lambda_0 > 1$ our bounds on τ_q will be finite for every $q \in (0,1)$. Our bounds on Δ will also be finite provided $\underline{B}_0$ and $\overline{B}_0$ admit an expectation.

5 Inference

In this section, we develop inference on LATE and QTE in the point and partially identified cases. In both cases, we impose the following conditions.

Assumption 9 *(Independent and identically distributed observations)*

$(Y_i, D_i, G_i, T_i)_{i=1,...,n}$ *are i.i.d.*

Assumption 10 *(Technical conditions for inference 1)*

$\mathcal{S}(Y)$ *is a bounded interval* $[\underline{y}, \overline{y}]$*. Moreover, for all* $(d,g,t) \in \{0,1\}^3$*,* $F_{dgt} = F_{Y_{gt}|D=d}$ *and* $F_{Y_{11}(d)|C}$ *are continuously differentiable with strictly positive derivatives on* $[\underline{y}, \overline{y}]$*.*

Athey & Imbens (2006) impose a condition similar to Assumption 10 when studying the asymptotic properties of their estimator.

We first consider the point identified case, which corresponds either to $0 < P(D_{00} = 0) = P(D_{01} = 0) < 1$ under Assumptions 1-5, or to $P(D_{00} = 0) = P(D_{01} = 0) \in \{0,1\}$ under Assumptions 1-6. For simplicity, we focus hereafter on the first case but the asymptotic properties of the estimator are similar in the second case. Let $\widehat{F}_{dgt}$ (resp. $\widehat{F}^{-1}_{dgt}$) denote the empirical cdf (resp. quantile function) of Y on the subsample $\{i : D_i = d, G_i = g, T_i = t\}$ and $\widehat{Q}_d = \widehat{F}^{-1}_{d01} \circ \widehat{F}_{d00}$. We also let $\mathcal{I}_{gt} = \{i : G_i = g, T_i = t\}$ and n_{gt} denote the size of $\mathcal{I}_{gt}$ for all $(d,g) \in \{0,1\}^2$. Our estimator of the LATE is

$$\widehat{\Delta} = \frac{\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} Y_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} \widehat{Q}_{D_i}(Y_i)}{\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} D_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} D_i}$$

Let $\widehat{P}(D_{gt} = d)$ be the proportion of subjects with $D = d$ in the sample $\mathcal{I}_{gt}$, let $\widehat{H}_d = \widehat{F}_{d10} \circ \widehat{F}^{-1}_{d00}$, and let

$$\widehat{F}_{Y_{11}(d)|C} = \frac{\widehat{P}(D_{01} = d)\widehat{H}_d \circ \widehat{F}_{d01} - \widehat{P}(D_{11} = d)\widehat{F}_{d11}}{\widehat{P}(D_{10} = d) - \widehat{P}(D_{11} = d)}.$$

Our estimator of the QTE of order q for compliers is

$$\widehat{\tau}_q = \widehat{F}^{-1}_{Y_{11}(1)|C}(q) - \widehat{F}^{-1}_{Y_{11}(0)|C}(q).$$

We say hereafter that an estimator $\widehat{\theta}$ of θ is root-n consistent and asymptotically normal if there exists Σ such that $\sqrt{n}(\widehat{\theta}-\theta) \xrightarrow{L} \mathcal{N}(0,\Sigma)$. Theorem 5.1 below shows that both $\widehat{\Delta}$ and $\widehat{\tau}_q$ are root-n consistent and asymptotically normal. Because the asymptotic variances take complicated expressions, we consider the bootstrap for inference. For any statistic T, we let T^* denote its bootstrap counterpart. For any root-n consistent statistic $\widehat{\theta}$ estimating consistently θ, we say that the bootstrap is consistent if with probability one and conditional on the sample, $\sqrt{n}(\widehat{\theta}^*-\widehat{\theta})$ converges to the same distribution as the limit distribution of $\sqrt{n}(\widehat{\theta}-\theta)$.[4] Theorem 5.1 shows that bootstrap confidence intervals are asymptotically valid.

Theorem 5.1 *Suppose that Assumptions 1-5, 9-10 hold and* $0 < P(D_{00}=0) = P(D_{01}=1) < 1$. *Then* $\widehat{\Delta}$ *and* $\widehat{\tau}_q$ *are root-n consistent and asymptotically normal. Moreover, the bootstrap is consistent for both* $\widehat{\Delta}$ *and* $\widehat{\tau}_q$.

In contrast with Athey & Imbens (2006), our proof of Theorem 5.1 is based on the weak convergence of the empirical cdfs of the different subgroups, and on a repeated use of the functional delta method. This approach can be readily applied to other functionals of $(F_{Y_{11}(0)|C}, F_{Y_{11}(1)|C})$, and to settings with several groups and periods considered in the web appendix.

We also follow this approach in the partially identified case. First, suppose that $0 < \widehat{P}(D_{00}=0) < 1$ and $0 < \widehat{P}(D_{10}=0) < 1$. Let $\widehat{\lambda}_d = \frac{\widehat{P}(D_{01}=d)}{\widehat{P}(D_{00}=d)}$, $\widehat{\mu}_d = \frac{\widehat{P}(D_{11}=d)}{\widehat{P}(D_{10}=d)}$ and define

$$\begin{aligned}
\underline{\widehat{T}}_d &= M_0\left(m_1\left(\frac{\widehat{\lambda}_d \widehat{F}_{Y_{01}|D=d} - \widehat{H}_d^{-1}(\widehat{\mu}_d \widehat{F}_{Y_{11}|D=d})}{\widehat{\lambda}_d - 1}\right)\right),\\
\overline{\widehat{T}}_d &= M_0\left(m_1\left(\frac{\widehat{\lambda}_d \widehat{F}_{Y_{01}|D=d} - \widehat{H}_d^{-1}(\widehat{\mu}_d \widehat{F}_{Y_{11}|D=d} + (1-\widehat{\mu}_d))}{\widehat{\lambda}_d - 1}\right)\right),\\
\widehat{G}_d(T) &= \widehat{\lambda}_d \widehat{F}_{Y_{01}|D=d} + (1-\widehat{\lambda}_d)T,\\
\widehat{C}_d(T) &= \frac{\widehat{P}(D_{10}=d)\widehat{H}_d \circ \widehat{G}_d(T) - \widehat{P}(D_{11}=d)\widehat{F}_{Y_{11}|D=d}}{\widehat{P}(D_{10}=d) - \widehat{P}(D_{11}=d)}.
\end{aligned}$$

To estimate bounds for $F_{Y_{11}(d)|C}$, we use

$$\underline{\widehat{B}}_d(y) = \sup_{y' \le y} \widehat{C}_d\left(\underline{\widehat{T}}_d\right)(y'), \quad \overline{\widehat{B}}_d(y) = \inf_{y' \ge y} \widehat{C}_d\left(\overline{\widehat{T}}_d\right)(y').$$

Therefore, to estimate bounds for the LATE and QTE, we use

$$\underline{\widehat{\Delta}} = \int y d\overline{\widehat{B}}_1(y) - \int y d\underline{\widehat{B}}_0(y), \quad \overline{\widehat{\Delta}} = \int y d\underline{\widehat{B}}_1(y) - \int y d\overline{\widehat{B}}_0(y),$$

$$\underline{\widehat{\tau}}_q = \overline{\widehat{B}}_1^{-1}(q) - \underline{\widehat{B}}_0^{-1}(q), \quad \overline{\widehat{\tau}}_q = \underline{\widehat{B}}_1^{-1}(q) - \overline{\widehat{B}}_0^{-1}(q).$$

When $\widehat{P}(D_{00}=0) \in \{0,1\}$ or $\widehat{P}(D_{10}=0) \in \{0,1\}$, the bounds on Δ and τ_q are defined similarly, but instead of $\underline{\widehat{B}}_d$ and $\overline{\widehat{B}}_d$, we use the empirical counterparts of the bounds on $F_{Y_{11}(d)|C}$ given by Equation (9).

[4] See, e.g., van der Vaart (2000), Section 23.2.1, for a formal definition of conditional convergence.

Let $B_\Delta = (\underline{\Delta}, \overline{\Delta})$ and $B_{\tau_q} = (\underline{\tau}_q, \overline{\tau}_q)'$, and let $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$ be the corresponding estimators. Theorem 5.2 below establishes the asymptotic normality and the validity of the bootstrap for both $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$, for $q \in \mathcal{Q} \subset (0,1)$, where $\mathcal{Q}$ is defined as follows. First, when $P(D_{00} = 0) \in \{0,1\}$, $P(D_{10} = 0) = 1$,[5] or $\lambda_0 > 1$, we merely let $\mathcal{Q} = (0,1)$. When $\lambda_0 < 1$, we have to exclude small and large q from $\mathcal{Q}$. This is because the (true) bounds put mass at the boundaries $\underline{y}$ or $\overline{y}$ of the support of Y. Similarly, the estimated bounds put mass on the estimated boundaries, which must be estimated. Because estimated boundaries typically have non-normal limit distribution, the asymptotic distribution of the bounds of the estimated QTE will also be non-normal. We thus restrict ourselves to $(\underline{q}, \overline{q})$, with $\underline{q} = \overline{B}_0(\underline{y})$ and $\overline{q} = \underline{B}_0(\overline{y})$. Another issue is that the bounds might be irregular at some $q \in (0,1)$, because they include in their definitions the kinked functions M_0 and m_1.[6] Let

$$q_1 = \frac{\mu_1 F_{Y_{11}|D=1} \circ F^{-1}_{Y_{01}|D=1}(\frac{1}{\lambda_1}) - 1}{\mu_1 - 1}, \quad q_2 = \frac{\mu_1 F_{Y_{11}|D=1} \circ F^{-1}_{Y_{01}|D=1}(1 - \frac{1}{\lambda_1})}{\mu_1 - 1}$$

denote the two points at which the bounds can be kinked. When $\lambda_0 < 1$, we restrict ourselves to $\mathcal{Q} = (\underline{q}, \overline{q}) \backslash \{q_1, q_2\}$. Note that q_1 and q_2 may not belong to $(\underline{q}, \overline{q})$, depending on λ_1 and μ_1, so that $\mathcal{Q}$ may in fact be equal to $(\underline{q}, \overline{q})$.

Theorem 5.2 relies on the following technical assumption, which involves the bounds rather than the true cdf since we are interested in estimating these bounds. Note that the strict monotonicity requirement is only a slight reinforcement of Assumption 7.

Assumption 11 *(Technical conditions for inference 2)*

For $d \in \{0,1\}$, the sets $\underline{\mathcal{S}}_d = [\underline{B}_d^{-1}(\underline{q}), \underline{B}_d^{-1}(\overline{q})] \cap \mathcal{S}(Y)$ and $\overline{\mathcal{S}}_d = [\overline{B}_d^{-1}(\underline{q}), \overline{B}_d^{-1}(\overline{q})] \cap \mathcal{S}(Y)$ are not empty. The bounds $\underline{B}_d$ and $\overline{B}_d$ are strictly increasing on $\underline{\mathcal{S}}_d$ and $\overline{\mathcal{S}}_d$. Their derivative, whenever they exist, are strictly positive.

Theorem 5.2 *Suppose that Assumptions 1-5, 7, 9-11 hold and $q \in \mathcal{Q}$. Then $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$ are root-n consistent and asymptotically normal. Moreover, the bootstrap is consistent for both.*

To construct confidence intervals of level $1-\alpha$ for Δ (resp. τ_q), one can use the lower bound of the two-sided confidence interval of level $1-\alpha$ of $\underline{\Delta}$ (resp. $\underline{\tau_q}$), and the upper bound of the two-sided confidence interval of $\overline{\Delta}$ (resp. $\overline{\tau_q}$). Such confidence intervals are asymptotically valid but conservative. Alternatively, one could use one-sided confidence intervals of level $1-\alpha$ on $\underline{\Delta}$ and $\overline{\Delta}$ (resp $\underline{\tau_q}$ and $\overline{\tau_q}$).[7]

[5] Assumption 4 rules out $P(D_{10} = 0) = 0$.

[6] This problem does not arise when $\lambda_0 > 1$. Kinks can arise only at 0 or 1 in this case.

[7] As shown in Imbens & Manski (2004), such confidence intervals are not uniformly valid. The solutions to this problem suggested by Imbens & Manski (2004) or Stoye (2009) require that bounds converge uniformly towards normal distributions. In Theorem 5.2, we only show pointwise convergence. Uniform convergence is likely to fail for QTE because of the kinks of $\underline{B}_d$ and $\overline{B}_d$ at the points q_1 or q_2.

So far, we have implicitly considered that we know whether point or partial identification holds, which is not the case in practice. This is an important issue, since the estimators and the way confidence intervals are constructed differ in the two cases. Abstracting from extreme cases where $P(D_{gt} = d) = 0$, testing point identification is equivalent to testing $\lambda_0 = 1$. One can conduct asymptotically valid inference by using point identification results if the t-statistic $\left|\frac{\widehat{\lambda}_0 - 1}{\widehat{\sigma}_{\lambda_0}}\right|$ is lower than a sequence c_n satisfying $c_n \to +\infty$ and $\frac{c_n}{\sqrt{n}} \to 0$. This pretest ensures that the probability of conducting inference under the wrong maintained assumption vanishes to 0 asymptotically. This procedure is similar to those recently developed in moment inequality models, which guarantee uniformly valid inference (see for instance Andrews & Soares, 2010). In the moment inequality literature, the choice of $c_n = \sqrt{2\ln(\ln(n))}$ has been advocated (see Andrews & Soares, 2010), so we stick to it in our applications.

6 Applications

6.1 Property rights and labor supply

Between 1996 and 2003, the Peruvian government issued property titles to 1.2 million urban households, the largest titling program targeted to squatters in the developing world. Field (2007) examines the labor market effects of increases in tenure security resulting from the program. Tenure insecurity encompasses fear of eviction by the government and fear of property theft by other residents. Such concerns might remove individuals from the labor force.

To isolate the causal effect of property rights, the author uses a survey conducted in 2000, and exploits two sources of variation in exposure to the titling program at that time. Firstly, this program took place at different dates in different neighborhoods. In 2000, it had approximately reached 50% of targeted neighborhoods. Secondly, it only impacted squatters, i.e. households without a property title prior to the program. The author can therefore construct four groups of households: squatters in neighborhoods reached by the program before 2000, squatters in neighborhoods reached by the program after 2000, non- squatters in neighborhoods reached by the program before 2000, and non- squatters in neighborhoods reached by the program after 2000. The share of households with a property title in each group is shown in Table 4.

Table 3: Share of households with a property right

	Reached after 2000	Reached before 2000
Squatters	0%	71%
Non-squatters	100%	100%

In Table 5 of her paper, the author estimates IV-DID regressions to capture the effect of having a property right on hours worked per week by the household. Whether the neighborhood was

reached before or after 2000 plays the role of time, while squatters and non-squatters are the two groups. In what follows, we apply our IV-CIC model to the same data. $P(D_{10} = 1) = 0$, so $F_{Y_{11}(1)|C}(y)$ is identified by $F_{Y_{11}|D=1}(y)$ as shown in Equation (7). Moreover, as $P(D_{00} = 1) = P(D_{10} = 1) = 1$, we can use Theorem 4.2 to identify $F_{Y_{11}(0)|C}(y)$. The resulting estimates are displayed in Figure 1. $\widehat{F}_{Y_{11}(0)|C}$ stochastically dominates $\widehat{F}_{Y_{11}(1)|C}$, which indicates that property rights have a positive impact on the number of hours worked over the entire distribution of hours. We test for stochastic dominance by adapting to our context the bootstrap test of Abadie (2002). We reject the null hypothesis that the two distributions are equal hypothesis at the 10% level (p-value=0.09). Details on the construction of the test are provided in de Chaisemartin & D'Haultfoeuille (2014).

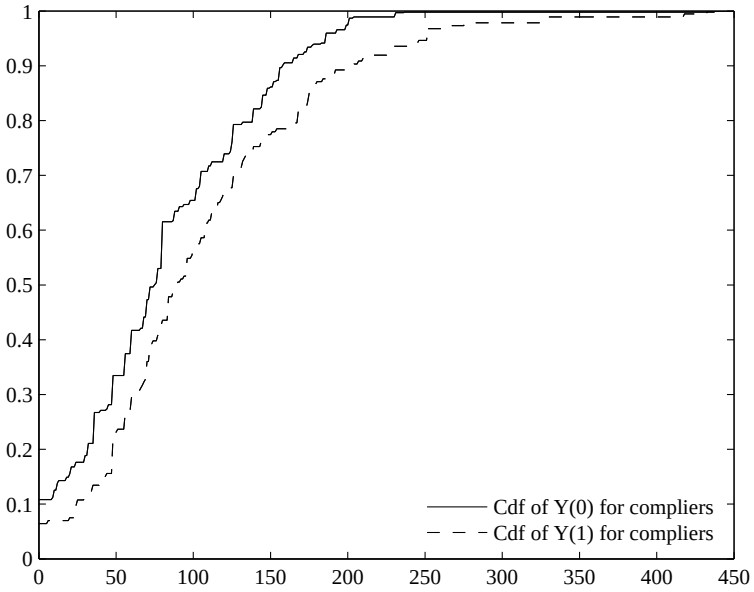

Figure 1: Estimated cdf of $Y(0)$ and $Y(1)$ for compliers.

As per our IV-CIC model, the LATE of titling on hours of work is equal to 23.3. This LATE is 17% lower than the one we would have obtained through an IV-DID regression (27.2), and the difference between the two is statistically significant (p-value=0.02).[8] Quantile treatment effects are shown in Figure 2. They are fairly constant, most of them being close to +20 hours of work per week. This implies that the effect of the treatment is highly heterogeneous in relative terms. We compute that being granted a property title increases labor supply by more than 40% for households at the 25th percentile of the distribution of hours worked per

[8]Our IV-DID LATE does not match exactly the "Tilted" coefficient in the second column of Table 5 in Field (2007). We estimated the same regression as the author, but without control variables. This is to ensure that the resulting coefficient is comparable with our IV-CIC LATE, which is estimated without controls. Our IV-CIC model allows for discrete controls, but here the sample size is too small to include as many as the author did.

week, and by 10% only for households at the 75th percentile. The difference between the two is marginally significant (p-value=0.12). An explanation for this pattern could go as follows. The main source of variation in hours worked per week at the household level is presumably the size of the household. In every household, only one household member has to stay home to look after the household's residence, irrespective of the household size. Being granted a property title therefore allows this household member to increase her labor supply, but has no effect on the labor supply of other members.

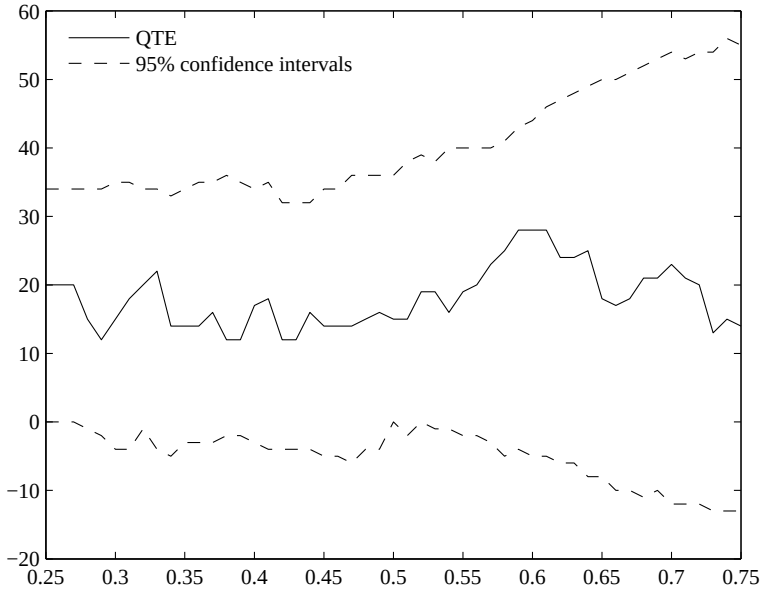

Figure 2: Estimated QTE on the number of hours worked.

6.2 Effectiveness of a smoking cessation treatment

Smoking rate among the adult population in France is around 30%. This is much higher than in most western countries (see e.g. Beck et al., 2007). Varenicline, a pharmacotherapy for smoking cessation, has been marketed in France since February 2007. Randomized controlled trials (RCT) have shown Varenicline to be more efficient than other pharmacotherapies used in smoking cessation (see e.g. Jorenby et al., 2006). However, there have been few studies based on non experimental data to confirm the efficacy of this new drug in real life settings.

We use a database from 17 French smoking cessation clinics, in which doctors, nurses, and psychologists help smokers quit. When a patient comes for the first time, the clinic staff measure the number of carbon monoxide (CO) parts per million in the air she expires. CO is a biomarker for recent tobacco use. After collecting those measures and discussing with the patient, they may advise her treatments, such as nicotine replacement therapies, to help her quit. Patients then come back for follow-up visits. During those visits, CO measures are

made to validate tobacco abstinence. This measure is much more reliable than daily cigarettes smoked, because it is not self-reported. Below 5 parts per million, a patient is regarded as a non smoker by clinics staff. She is regarded as a light smoker when her CO is between 5 and 10, as a smoker when it is between 10 and 20, and as a heavy smoker when it is above 20.[9]

The rate of prescription of Varenicline after February 2007 ranges from 0% to 37% across clinics. A very strong predictor of clinics propensity to prescribe varenicline is the share of their staff holding the "diplome universitaire de tabacologie" (DUT) in 2005-2006, i.e. before varenicline was released. The DUT is a university degree awarded to staff who followed a 60 hours course on how to help smokers quit. The share of staff holding it ranges from 0 to 100% across clinics, with a median equal to 60%. The correlation between prescription rate and share of staff holding this degree is equal to 0.63. Staff who took this training a few years before varenicline got market approval might have then been told that preliminary RCT showed this new drug to be promising. They might also have a stronger taste for learning than those who do not take this training, and might be more prone to adopting medical innovations.

We use the share of staff holding this degree in 2005-2006 to construct two groups of "control" and "treatment" clinics. Control clinics are those belonging to the first tercile of this measure, while treatment clinics are those belonging to the third tercile. Period 0 covers the 2 years before the release of Varenicline (February 2005 to January 2007), while period 1 extends over the 2 years following it (February 2007 to January 2009). Our sample is made up of the 7,468 patients who attended control and treatment clinics over these two periods and who came to at least one follow-up visit. By construction, the prescription rate of Varenicline is 0% in control and treatment clinics at period 0. At period 1, it is equal to 4.9% in control clinics and 25.4% in treatment clinics. This differential evolution of treatment rates over time across those two groups of clinics is the source of variation we use to measure the effect of varenicline.

Table 4: Share of patients prescribed varenicline

	Before February 2007	After February 2007
Treatment clinics	0%	25.4%
Control clinics	0%	4.9%

$P(D_{10} = 1) = 0$, so $F_{Y_{11}(1)|C}(y)$ is identified by $F_{Y_{11}|D=1}(y)$ as shown in Equation (7). On the other hand, $\widehat{P}(D_{00} = 1) \neq \widehat{P}(D_{01} = 1)$ and our pretest rejects $\lambda_0 = 1$. We therefore use partial identification results of Theorem 4.3 to estimate bounds for $F_{Y_{11}(0)|C}(y)$.

Figure 3 shows estimated bounds for QTE controlling for quartiles of expired CO at baseline.[10]

[9] Typically, light smokers smoke less than 10 cigarettes a day, smokers smoke between 10 and 20 cigarettes a day, and heavy smokers smoke more than 20 cigarettes.

[10] In de Chaisemartin & D'Haultfoeuille (2014) we discuss how our model can incorporate discrete controls.

The two bounds are very close to 0 up to the 50th percentile, but τ_q is significantly different from 0 for $q \in (0.62, 0.82)$.[11] Varenicline does not have an impact on low and median values of CO at follow-up, but it has a strong effect on high values.

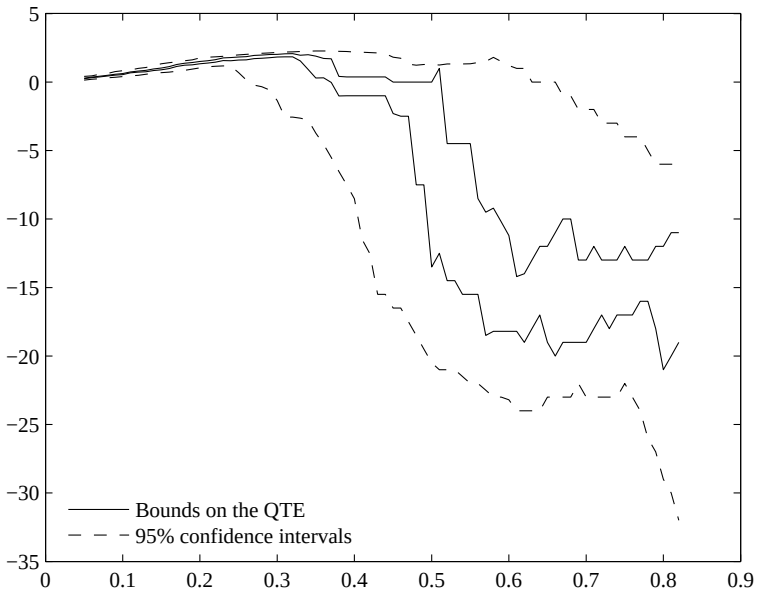

Figure 3: Estimated bounds for QTE, including baseline CO as a control.

Finally, we use our results to analyze the effect of varenicline on the probability of being a non smoker, a light smoker, a smoker, or an heavy smoker at follow-up. Results are displayed in Table 5. Varenicline does not significantly increase the share of non smokers at follow-up, even though our bounds point towards a small, positive effect. However, it has a large and significant negative effect on the share of heavy smokers: even as per our worst case upper bound, it decreases this share by 13.4 percentage points.

QTE without controls are similar to those displayed in Figure 3, but they are less precisely estimated.

[11] We show results up to the 82% percentile only, because $\overline{q} = 0.82$.

Table 5: Effect of Varenicline on shares of non, light, medium, & heavy smokers.

	$P(Y_{11}(1) \in I\|C) - P(Y_{11}(0) \in I\|C)$		
I	Lower bound	Upper bound	p-value
[0;5]	0.4%	19.5%	0.48
(5;10]	-19.1%	8.2%	0.93
(10;20]	4.5%	15.1%	0.27
(20;+∞)	-24.0%	-13.4%	0.02

Notes: the p-value corresponds to the test of the null hypothesis that $P(Y_{11}(1) \in I|C) - P(Y_{11}(0) \in I|C) = 0$.

6.3 Returns to education

In 1973, the Indonesian government launched a major school construction program, the so-called Sekolah Dasar INPRES program. It led to the construction of more than 61,000 primary schools between 1973-1974 and 1978-1979. Duflo (2001) uses the 1995 SUPAS census to measure the effect of this program on completed years of education in a first step, and returns to education in a second step. In what follows, we only consider the latter set of results.

There was substantial variation in treatment intensity across regions, as the government tried to allocate more schools to districts with low initial enrolment. The author thus constructs two groups of high and low program regions, by regressing the number of schools constructed on the number of children in each region. High treatment regions are those with a positive residual in that regression, as they received more schools than what their population predicts. Exposure to treatment also varied by cohort: children between 2 and 6 in 1974 were exposed to the treatment as they were to enter primary school after the program was launched, while children between 12 and 17 in 1974 were not exposed as they were to have finished primary school by that time.

Number of years of education is larger for the second cohort in the two groups of regions, as schools were constructed in both groups. But the difference is larger in high treatment regions because more schools were constructed there. The author exploits this pattern to measure returns to education. She uses first a simple IV-DID regression in which birth cohort plays the role of the time variable, while low and high treatment regions are the two groups. The resulting coefficient, which we can infer from Table 3 in the paper, is imprecisely estimated, so the author turns to richer specifications. All of them include cohort and region of birth fixed effects, so one can show that the resulting coefficient is a weighted average of Wald-DID across all possible pairs of regions and birth cohorts.

In what follows, we apply our IV-CIC model to the same data. As it does not allow for a

multivariate treatment, we consider a dummy for primary school completion as our treatment variable. The program was a primary school construction program. The larger increase in completed years of education in high program regions mostly comes from a larger increase in the share of individuals completing primary school. For instance, the share of individuals completing middle school did not evolve differently in the two groups. Our binary treatment should capture most of the variation in educational attainment induced by the program.

Table 6: Share of individuals completing primary school

	Older cohort	Younger cohort
High treatment regions	81.2%	90.0%
Low treatment regions	89.8%	94.3%

Table 6 shows that the share of individuals completing primary school increased more in high than in low treatment regions. Still, it increased in low treatment regions as well, and our pretest rejects $\lambda_0 = 1$. We therefore use Theorem 4.3 and estimate bounds for $F_{Y_{11}(0)|C}(y)$ and $F_{Y_{11}(1)|C}(y)$. The resulting estimates are displayed in Figure 4. The bounds are wide. The bounds for QTE are uninformative, as 0 always lies between $\widehat{\underline{\tau}}_q$ and $\widehat{\overline{\tau}}_q$. This is because the increase in treatment rate was not much larger in high than in low treatment regions.

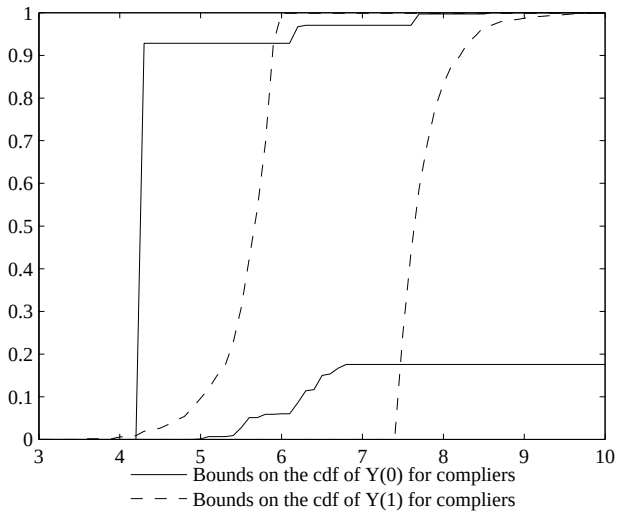

Figure 4: Estimated bounds on the cdf of $Y(0)$ and $Y(1)$ for compliers.

This application shows that when exposition to treatment substantially changes in the control group as well, using our IV-CIC model may result in wide and uninformative bounds. In such instances, point identification can still be achieved using IV-DID, but this strategy relies on more stringent conditions than our IV-CIC model, as discussed in Section 2 and de Chaise-

martin (2013). Besides common trend conditions on potential outcomes and treatments, IV-DID requires that returns to education be homogeneous across high and low treatment regions. Low treatment regions had fewer schools previous to the program. Returns to education could be higher in those areas if they face a shortage of qualified labor. They could also be lower if they are less developed and few qualified jobs are available there.

Finally, another strategy to recover point identification would be to look for another control group in which educational attainment did not change over time, and then use our IV-CIC model. One could for instance use regions in which primary school completion rate changed the least across the two cohorts.

7 Conclusion

In this paper, we develop an IV-CIC model to identify treatment effects when the treatment rate increases more in some groups than in others, for instance following a legislative change. Our model brings several improvements to IV-DID, the model currently used in the literature in such settings. It does not require common trend assumptions, it is invariant to monotonic transforms of the outcome, and it does not impose that some subgroups of observations in the treatment and in the control groups have the same treatment effects.

We show that when the treatment rate is stable between period 0 and 1 in the control group, a LATE and QTE among compliers are point identified under our IV-CIC assumptions. When the treatment rate also changes between period 0 and 1 in the control group, the same LATE and QTE are partially identified. The smaller the change in the treatment rate in the control group, the tighter the bounds. We conduct inference on treatment effects and sharp bounds estimators by proving their asymptotic normality and showing the validity of the bootstrap.

Applied researchers must therefore find a control group in which the treatment rate does not evolve too much over time to derive informative conclusions under our non linear IV-CIC model. If such a control group is not available, point identification can still be achieved using IV-DID, but results will rely on stronger assumptions.

References

Abadie, A. (2002), 'Bootstrap tests for distributional treatment effects in instrumental variable models', *Journal of the American Statistical Association* **97**(457), pp. 284–292.

Abadie, A. (2005), 'Semiparametric difference-in-differences estimators', *Review of Economic Studies* **72**(1), 1–19.

Akerman, A., Gaarder, I. & Mogstad, M. (2013), The skill complementarity of broadband internet, Technical report, Stockholm University, Department of Economics.

Altonji, J. & Blank, R. (2000), Race and gender in the labor market, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Amsterdam: Elsevier, pp. 3143–3259.

Andrews, D. W. K. & Soares, G. (2010), 'Inference for parameters defined by moment inequalities using generalized moment selection', *Econometrica* **78**(1), 119–157.

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

Beck, F., Guilbert, P. & Gautier, A. (2007), Baromètre Santé [French health barometer], Saint-Denis, INPES.

Blundell, R., Dias, M. C., Meghir, C. & Reenen, J. V. (2004), 'Evaluating the employment impact of a mandatory job search program', *Journal of the European Economic Association* **2**(4), 569–606.

Bonhomme, S. & Sauder, U. (2011), 'Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling', *Review of Economics and Statistics* **93**(2), 479–494.

Burgess, R. & Pande, R. (2005), 'Do rural banks matter? evidence from the indian social banking experiment', *American Economic Review* **95**(3), 780–795.

de Chaisemartin, C. (2013), A note on the assumptions underlying instrumented difference in differences., Working paper.

de Chaisemartin, C. & D'Haultfoeuille, X. (2014), Web appendix to "fuzzy changes-in-changes", Technical report.

D'Haultfoeuille, X., Hoderlein, S. & Sasaki, Y. (2013), Nonlinear difference-in-differences in repeated cross sections with continuous treatments. CEMMAP Working Paper CWP40/13.

Donald, S. G. & Lang, K. (2007), 'Inference with difference-in-differences and other panel data', *The Review of Economics and Statistics* **89**(2), 221–233.

Duflo, E. (2001), 'Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment', *American Economic Review* **91**(4), 795–813.

Field, E. (2007), 'Entitled to work: Urban property rights and labor supply in Peru', *The Quarterly Journal of Economics* **122**(4), 1561–1602.

Horowitz, J. L. & Manski, C. F. (1995), 'Identification and robustness with contaminated and corrupted data', *Econometrica* **63**(2), 281–302.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–75.

Imbens, G. W. & Manski, C. F. (2004), 'Confidence intervals for partially identified parameters', *Econometrica* **72**(6), 1845–1857.

Imbens, G. W. & Rubin, D. B. (1997), 'Estimating outcome distributions for compliers in instrumental variables models', *Review of Economic Studies* **64**(4), 555–574.

Jorenby, D. E., Hays, J. T., Rigotti, N. A., Azoulay, S. & Watsky, E. J. (2006), 'Efficacy of varenicline, an $\alpha 4\beta 2$ nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation', *Journal of the American Medical Association* **306**(17).

Juhn, C., Murphy, K. M. & Pierce, B. (1993), 'Wage inequality and the rise in returns to skill', *Journal of Political Economy* **101**(3), 410–442.

Lochner, L. & Moretti, E. (2004), 'The effect of education on crime: Evidence from prison inmates, arrests, and self-reports', *The American Economic Review* **94**(1), 155–189.

Manski, C. & Pepper, J. (2012), Partial identification of treatment response with data on repeated cross sections, Working papers.

Stoye, J. (2009), 'More on confidence intervals for partially identified parameters', *Econometrica* **77**(4), 1299–1315.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak convergence and Empirical Processes*, Springer.

Vytlacil, E. (2002), 'Independence, monotonicity, and latent index models: An equivalence result', *Econometrica* **70**(1), 331–341.

A Main proofs

The proofs of Theorems 5.1 and 5.2 rely on technical lemmas appearing in our web appendix (see de Chaisemartin & D'Haultfoeuille (2014)).

Lemma 4.1

We only prove the formula for $d = 0$, the reasoning being similar for $d = 1$. We first show that

$$F_{Y_{11}(0)|C}(y) = \frac{P(D_{10} = 0)F_{Y_{11}(0)|V<v_0(0)}(y) - P(D_{11} = 0)F_{Y_{11}|D=0}(y)}{P(D_{10} = 0) - P(D_{11} = 0)}. \tag{10}$$

To this aim, note first that

$$\begin{aligned} P(C|G = 1, T = 1, V < v_0(0)) &= \frac{P(V \in [v_1(1), v_0(0))|G = 1, T = 1)}{P(V < v_0(0)|G = 1, T = 1)} \\ &= \frac{P(V < v_0(0)|G = 1, T = 1) - P(V < v_1(1)|G = 1, T = 1)}{P(V < v_0(0)|G = 1, T = 1)} \\ &= \frac{P(V < v_0(0)|G = 1, T = 0) - P(V < v_1(1)|G = 1, T = 1)}{P(V < v_0(0)|G = 1, T = 0)} \\ &= \frac{P(D_{10} = 0) - P(D_{11} = 0)}{P(D_{10} = 0)}. \end{aligned}$$

The third equality stems from Assumption 3, and $P(D_{10} = 0) > 0$ because of Assumption 5. Then

$$\begin{aligned} F_{Y_{11}(0)|V<v_0(0)}(y) &= P(V \in [v_1(1), v_0(0))|G = 1, T = 1, V < v_0(0))F_{Y_{11}(0)|V\in[v_1(1),v_0(0))}(y) \\ &\quad + P(V < v_1(1)|G = 1, T = 1, V < v_0(0))F_{Y_{11}|V<v_1(1)}(y) \\ &= \frac{P(D_{10} = 0) - P(D_{11} = 0)}{P(D_{10} = 0)}F_{Y_{11}(0)|C}(y) + \frac{P(D_{11} = 0)}{P(D_{10} = 0)}F_{Y_{11}|D=0}(y) \end{aligned}$$

This proves (10), and thus the second point of the lemma.

To prove the first point of the lemma, we show that for all $y \in \mathcal{S}(Y_{11}(0)|V < v_0(0))$,

$$F_{Y_{11}(0)|V<v_0(0)} = F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0} \circ F_{Y_{01}(0)|V<v_0(0)}. \tag{11}$$

By Assumption 3, $(U_0, \mathbb{1}\{V < v_0(0)\}) \perp\!\!\!\perp T|G$, which implies

$$U_0 \perp\!\!\!\perp T|G, V < v_0(0).$$

As a result, for all $(g, t) \in \{0, 1\}^2$,

$$\begin{aligned} F_{Y_{gt}(0)|V<v_0(0)}(y) &= P(h_0(U_0, t) \leq y|G = g, T = t, V < v_0(0)) \\ &= P(U_0 \leq h_0^{-1}(y, t)|G = g, T = t, V < v_0(0)) \\ &= P(U_0 \leq h_0^{-1}(y, t)|G = g, V < v_0(0)) \\ &= F_{U_0|G=g,V<v_0(0)}(h_0^{-1}(y, t)). \end{aligned}$$

The second point of Assumption 4 combined with Assumptions 1 and 3 implies that $F_{U_0|G=g,V<v_0(0)}$ is strictly increasing. Hence, its inverse exists and for all $q \in (0,1)$,

$$F^{-1}_{Y_{gt}(0)|V<v_0(0)}(q) = h_0\left(F^{-1}_{U_0|G=g,V<v_0(0)}(q), t\right).$$

This implies that for all $y \in \mathcal{S}(Y_{g1}(0)|V < v_0(0))$,

$$F^{-1}_{Y_{g0}(0)|V<v_0(0)} \circ F_{Y_{g1}(0)|V<v_0(0)}(y) = h_0(h_0^{-1}(y,1),0), \tag{12}$$

which is independent of g.

Now, we have

$$\begin{aligned}
&\mathcal{S}(Y_{10}|D=0) = \mathcal{S}(Y_{00}|D=0)\\
&\Rightarrow \mathcal{S}(Y_{10}(0)|V<v_0(0)) = \mathcal{S}(Y_{00}(0)|V<v_0(0))\\
&\Rightarrow \mathcal{S}(h_0(U_0,0)|V<v_0(0),G=1,T=0) = \mathcal{S}(h_0(U_0,0)|V<v_0(0),G=0,T=0)\\
&\Rightarrow \mathcal{S}(U_0|V<v_0(0),G=1) = \mathcal{S}(U_0|V<v_0(0),G=0)\\
&\Rightarrow \mathcal{S}(h_0(U_0,1)|V<v_0(0),G=1,T=1) = \mathcal{S}(h_0(U_0,1)|V<v_0(0),G=0,T=1)\\
&\Rightarrow \mathcal{S}(Y_{11}(0)|V<v_0(0)) = \mathcal{S}(Y_{01}(0)|V<v_0(0)),
\end{aligned}$$

where the third and fourth implications are obtained combining Assumptions 1 and 3. Therefore, for all $y \in \mathcal{S}(Y_{11}(0)|V < v_0(0))$,

$$F^{-1}_{Y_{10}(0)|V<v_0(0)} \circ F_{Y_{11}(0)|V<v_0(0)}(y) = F^{-1}_{Y_{00}(0)|V<v_0(0)} \circ F_{Y_{01}(0)|V<v_0(0)}(y).$$

This proves (11), because $V < v_0(0)$ is equivalent to $D = 0$ when $T = 0$, and because the second point of Assumption 4 implies that $F^{-1}_{Y_{10}|D=0}$ is strictly increasing on $(0,1)$.

Finally, we show that

$$F_{Y_{01}(0)|V<v_0(0)}(y) = \lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0)F_{Y_{01}(0)|TC}(y). \tag{13}$$

Suppose first that $\lambda_0 \leq 1$. Then, $v_0(1) \leq v_0(0)$ and TC is equivalent to the event $V \in [v_0(1), v_0(0))$. Moreover, reasoning as for $P(C|G=1, V<v_0(0))$, we get

$$\lambda_0 = \frac{P(V<v_0(1)|G=0)}{P(V<v_0(0)|G=0)} = P(V<v_0(1)|G=0, V<v_0(0)).$$

Then

$$\begin{aligned}
F_{Y_{01}(0)|V<v_0(0)}(y) &= P(V<v_0(1)|G=0,V<v_0(0))F_{Y_{01}(0)|V<v_0(1)}(y)\\
&\quad + P(V\in[v_0(1),v_0(0))|G=0,V<v_0(0))F_{Y_{01}|V\in[v_0(1),v_0(0))}(y)\\
&= \lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0)F_{Y_{01}(0)|TC}(y).
\end{aligned}$$

If $\lambda_0 > 1$, $v_0(1) > v_0(0)$ and TC is equivalent to the event $V \in [v_0(0), v_0(1))$.

$$\frac{1}{\lambda_0} = P(V<v_0(0)|G=0, V<v_0(1))$$

and

$$F_{Y_{01}|D=0}(y) = \frac{1}{\lambda_0} F_{Y_{01}(0)|V<v_0(0)}(y) + \left(1 - \frac{1}{\lambda_0}\right) F_{Y_{01}(0)|TC}(y),$$

so that we also get (13).

Finally, the first point of the lemma follows by combining (10), (11) and (13).

Theorem 4.1

The proof follows from Lemma 4.1: $\lambda_0 = \lambda_1 = 1$ when $P(D_{00} = d) = P(D_{01} = d) > 0$.

Theorem 4.2

Assume that $P(D_{00} = 0) = P(D_{01} = 0) = 0$ (the proof is symmetric when $P(D_{00} = 1) = P(D_{01} = 1) = 0$). This implies that $P(D_{00} = 1) = P(D_{01} = 1) > 0$, so for $F_{Y_{11}(1)|C}(y)$ the proof directly follows from Lemma 4.1, by noting that $\lambda_1 = 1$.

For $F_{Y_{11}(0)|C}(y)$, one can use the same steps as in the proof of Lemma 4.1 to show that Equation (10) also holds here:

$$F_{Y_{11}(0)|C}(y) = \frac{P(D_{10} = 0) F_{Y_{11}(0)|V<v_0(0)}(y) - P(D_{11} = 0) F_{Y_{11}|D=0}(y)}{P(D_{10} = 0) - P(D_{11} = 0)}. \tag{14}$$

Then, let $\underline{v}$ denote the lower bound of $\mathcal{S}(V|G = 0)$. Following similar steps as those used to establish Equation (12), one can show that for all $y \in \mathcal{S}(Y_{01}(0)|V < v_0(0)) = \mathcal{S}(Y_{00}(0)|V \geq v_0(0)) = \mathcal{S}(Y)$,

$$\begin{aligned} F^{-1}_{Y_{10}(0)|V<v_0(0)} \circ F_{Y_{11}(0)|V<v_0(0)}(y) &= h_0(h_0^{-1}(y,1),0), \\ F^{-1}_{Y_{00}(1)|V\geq\underline{v}} \circ F_{Y_{01}(1)|V\geq\underline{v}}(y) &= h_1(h_1^{-1}(y,1),0). \end{aligned}$$

Under Assumption 6, this implies that for all $y \in \mathcal{S}(Y)$,

$$\begin{aligned} F_{Y_{11}(0)|V<v_0(0)}(y) &= F_{Y_{10}(0)|V<v_0(0)} \circ F^{-1}_{Y_{00}(1)|V\geq\underline{v}} \circ F_{Y_{01}(1)|V\geq\underline{v}}(y) \\ &= F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=1} \circ F_{Y_{01}|D=1}(y), \end{aligned} \tag{15}$$

where the second equality follows from the fact that $P(D_{00} = 1) = P(D_{01} = 1) = 1$. Combining Equations (14) and (15) yields the result for $F_{Y_{11}(0)|C}(y)$.

Theorem 4.3

We focus on the case where $P(D_{00} = d) > 0$. When $P(D_{00} = d) = 0$, the proofs are immediate.

1. Construction of the bounds.

We only establish the validity of the bounds for $d = 0$, the reasoning being similar for $d = 1$. We start by considering the case where $\lambda_0 < 1$. We first show that in such instances, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ if and only if

$$\underline{T_0} \leq T_0 \leq \overline{T_0}. \tag{16}$$

Indeed, $G_0(T_0)$ is included between 0 and 1 if and only if

$$\frac{-\lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0} \leq T_0 \leq \frac{1-\lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0},$$

while $C_0(T_0)$ is included between 0 and 1 if and only if

$$\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0}) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0} \leq T_0 \leq \frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0)) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}.$$

Since $-\lambda_0 F_{Y_{01}|D=0}/(1-\lambda_0) \leq 0$ and $(1-\lambda_0 F_{Y_{01}|D=0})/(1-\lambda_0) \geq 1$, T_0, $G_0(T_0)$ and $C_0(T_0)$ are all included between 0 and 1 if and only if

$$M_0\left(\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0}) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}\right) \leq T_0 \leq m_1\left(\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0)) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}\right). \quad (17)$$

By composing each term of these inequalities by $M_0(.)$ and then by $m_1(.)$, we obtain (16) since $M_0(T_0) = m_1(T_0) = T_0$ and $M_0 \circ m_1 = m_1 \circ M_0$.

Now, when $\lambda_0 < 1$, $G_0(T_0)$ is increasing in T_0, so $C_0(T_0)$ as well is increasing in T_0. Combining this with (16) implies that for every y',

$$C_0(\underline{T_0})(y') \leq C_0(T_0)(y') \leq C_0(\overline{T_0})(y').$$

Because $C_0(T_0)(y)$ is a cdf,

$$C_0(T_0)(y) = \inf_{y' \geq y} C_0(T_0)(y') \leq \inf_{y' \geq y} C_0(\overline{T_0})(y'). \quad (18)$$

The lower bound follows similarly.

Let us now turn to the case where $\lambda_0 > 1$. Using the same reasoning as above, we get that $G_0(T_0)$ and $C_0(T_0)$ are included between 0 and 1 if and only if

$$\begin{aligned} \frac{\lambda_0 F_{Y_{01}|D=0} - 1}{\lambda_0 - 1} &\leq T_0 \leq \frac{\lambda_0 F_{Y_{01}|D=0}}{\lambda_0 - 1}, \\ \frac{\lambda_0 F_{Y_{01}|D=0} - H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0))}{\lambda_0 - 1} &\leq T_0 \leq \frac{\lambda_0 F_{Y_{01}|D=0} - H_0^{-1}(\mu_0 F_{Y_{11}|D=0})}{\lambda_0 - 1}. \end{aligned}$$

The inequalities in the first line are not binding since they are implied by those on the second line. Thus, we also get (17). Hence, using the same argument as previously,

$$\overline{T_0} \leq T_0 \leq \underline{T_0}. \quad (19)$$

Besides, when $\lambda_0 > 1$, $G_0(T_0)$ is decreasing in T_0, so that $C_0(T_0)$ as well is decreasing in T_0. Combining this with (19) implies that for every y, (18) holds as well. This proves the result.

2. Sketch of the proof of sharpness.

The full proof is in our web appendix (see de Chaisemartin & D'Haultfoeuille, 2014). We only consider the sharpness of $\underline{B}_0$, the reasoning being similar for the upper bound. The proof is

also similar and actually simpler for $d = 1$. The corresponding bounds are indeed proper cdf, so we do not have to consider converging sequences of cdf as we do in case b) below.

a. $\lambda_0 > 1$**.** We show that if Assumptions 4-7 hold, then $\underline{B}_0$ is sharp. For that purpose, we construct $\widetilde{h}_0, \widetilde{U}_0, \widetilde{V}$ such that:

(i) $Y = \widetilde{h}_0(\widetilde{U}_0, T)$ when $D = 0$ and $D = 1\{\widetilde{V} \geq v_Z(T)\}$;

(ii) $\widetilde{h}_0(., t)$ is strictly increasing for $t \in \{0, 1\}$;

(iii) $(\widetilde{U}_0, \widetilde{V}) \perp\!\!\!\perp T|G$;

(iv) $F_{\widetilde{h}_0(\widetilde{U}_0,1)|G=0,T=1,\widetilde{V}\in[v_0(0),v_0(1))} = \underline{T}_0$.

(i) ensures that Equation (4) and Assumption 2 are satisfied on the observed data. Because we can always define $\widetilde{Y}(0)$ as $\widetilde{h}_0(\widetilde{U}_0, T)$ when $D = 1$ and $\widetilde{D}(z) = 1\{\widetilde{V} \geq v_z(T)\}$ when $Z \neq z$ without contradicting the data and the model, (i) is actually sufficient for Equation (4) and Assumption 2 to hold globally, not only on observed data. (ii) and (iii) ensure that Assumptions 1 and 3 hold. Finally, (iv) ensures that the DGP corresponding to $(\widetilde{h}_0, \widetilde{U}_0, \widetilde{V})$ rationalizes the bound. If $(\widetilde{h}_0, \widetilde{U}_0, \widetilde{V})$ satisfy Assumptions 1-5 and are such that $\widetilde{T}_0 = \underline{T}_0$, we can apply Lemma 4.1 to show that the bound is attained.

The construction of $\widetilde{h}_0$, $\widetilde{U}_0$, and $\widetilde{V}$ is long, so its presentation is deferred to our web appendix.

b. $\lambda_0 < 1$**.** The idea is similar as in the previous case. A difference, however, is that when $\lambda_0 < 1$, $\underline{T}_0$ is not a proper cdf, but a defective one, since $\lim_{y\to\overline{y}} \underline{T}_0(y) < 1$. As a result, we cannot define a DGP such that $\widetilde{T}_0 = \underline{T}_0$, However, by Lemma 3 (see de Chaisemartin & D'Haultfoeuille (2014)), there exists a sequence $(\underline{T}_0^k)_k$ of cdf such that $\underline{T}_0^k \to \underline{T}_0$, $G_0(\underline{T}_0^k)$ is an increasing bijection from $\mathcal{S}(Y)$ to $(0, 1)$ and $C_0(\underline{T}_0^k)$ is increasing and onto $(0, 1)$. We can then construct a sequence of DGP $(\widetilde{h}_0^k(., 0), \widetilde{h}_0^k(., 1), \widetilde{U}_0^k, \widetilde{V}^k)$ such that Points (i) to (iii) listed above hold for every k, and such that $\widetilde{T}_0^k = \underline{T}_0^k$. Since $\underline{T}_0^k(y)$ converges to $\underline{T}_0(y)$ for every y in $\overset{\circ}{\mathcal{S}}(Y)$, we thus define a sequence of DGP such that $\widetilde{T}_0^k$ can be arbitrarily close to $\underline{T}_0$ on $\overset{\circ}{\mathcal{S}}(Y)$ for sufficiently large k. Since $C_0(.)$ is continuous, this proves that $\underline{B}_0$ is sharp on $\overset{\circ}{\mathcal{S}}(Y)$. This construction is long, so its exposition is deferred to our web appendix.

Corollary 4.4

See de Chaisemartin & D'Haultfoeuille (2014).

Theorem 5.1

Hereafter, we let $\mathcal{C}^0$ and $\mathcal{C}^1$ denote respectively the set of continuous functions and the set of continuously differentiable functions with strictly positive derivative on $\mathcal{S}(Y)$.

We first show that $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C})$ tends to a continuous gaussian process. Let $\widetilde{\theta} = (F_{000}, F_{001}, ..., F_{111}, \mu_0, \mu_1)$. By Lemma 4, $\widehat{\widetilde{\theta}} = (\widehat{F}_{000}, \widehat{F}_{001}, ..., \widehat{F}_{111}, \widehat{\mu}_0, \widehat{\mu}_1)$ converges to a continuous gaussian process. Let

$$\pi_d : (F_{000}, F_{001}, ..., F_{111}, \mu_0, \mu_1) \mapsto (F_{d10}, F_{d00}, F_{d01}, F_{d11}, 1, \mu_d), \quad d \in \{0, 1\},$$

so that $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C}) = \left(R_1 \circ \pi_0(\widetilde{\theta}), R_1 \circ \pi_1(\widetilde{\theta})\right)$, where R_1 is defined as in Lemma 5. π_d is Hadamard differentiable as a linear continuous map. Because $F_{d10}, F_{d00}, F_{d01}, F_{d11}$ are continuously differentiable with strictly positive derivative by Assumption 10, $\mu_d > 0$, and $\mu_d \neq 1$ under Assumption 4, R_1 is also Hadamard differentiable at $(F_{d10}, F_{d00}, F_{d01}, F_{d11}, 1, \mu_d)$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}$. By the functional delta method (see, e.g., van der Vaart & Wellner, 1996, Lemma 3.9.4), $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C})$ tends to a continuous gaussian process.

Now, by integration by parts for Lebesgue-Stieljes integrals,

$$\Delta = \int_{\underline{y}}^{\overline{y}} F_{Y_{11}(0)|C}(y) - F_{Y_{11}(1)|C}(y) dy.$$

Moreover, the map $\varphi_1 : (F_1, F_2) \mapsto \int_{\mathcal{S}(Y)} (F_2(y) - F_1(y)) dy$, defined on the domain of bounded càdlàg functions, is linear. Because $\mathcal{S}(Y)$ is bounded by Assumption 10, φ_1 is also continuous with respect to the supremum norm. It is thus Hadamard differentiable. Because $\widehat{\Delta} = \varphi_1 \left(\widehat{F}_{Y_{11}(1)|C}, \widehat{F}_{Y_{11}(0)|C}\right)$, $\widehat{\Delta}$ is asymptotically normal by the functional delta method. The asymptotic normality of $\widehat{\tau}_q$ follows along similar lines. By Assumption 10, $F_{Y_{11}(d)|C}$ is differentiable with strictly positive derivative on its support. Thus, the map $(F_1, F_2) \mapsto F_2^{-1}(q) - F_1^{-1}(q)$ is Hadamard differentiable at $(F_{Y_{11}(0)|C}, F_{Y_{11}(1)|C})$ tangentially to the set of functions that are continuous at $(F^{-1}_{Y_{11}(0)|C}(q), F^{-1}_{Y_{11}(1)|C}(q))$ (see Lemma 21.3 in van der Vaart, 2000). By the functional delta method, $\widehat{\tau}_q$ is asymptotically normal.

The validity of the bootstrap follows along the same lines. By Lemma 4, the bootstrap is consistent for $\widehat{\theta}$. Because both the LATE and QTE are Hadamard differentiable functions of $\widehat{\theta}$, as shown above, the result simply follows by the functional delta method for the bootstrap (see, e.g., van der Vaart, 2000, Theorem 23.9).

Theorem 5.2

Let $\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_0, \mu_0, \lambda_1, \mu_1)$. By Lemma 6, for $d \in \{0, 1\}$ and $q \in \mathcal{Q}$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \underline{B}_d(y) dy$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \overline{B}_d(y) dy$, $\theta \mapsto \overline{B}_d^{-1}(q)$, and $\theta \mapsto \underline{B}_d^{-1}(q)$ are Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. Because $\underline{\Delta} = \int_{\mathcal{S}(Y)} \underline{B}_0(y) - \overline{B}_1(y) dy$, $\underline{\Delta}$ is also a Hadamard differentiable function of θ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. The same reasoning applies for $\overline{\Delta}$, and for $\underline{\tau_q}$ and $\overline{\tau_q}$ for every $q \in \mathcal{Q}$. The theorem then follows from Lemma 4, the functional delta method, and the functional delta method for the bootstrap.