Fuzzy Change in Change
Secondary Job Market Paper *

Clément de Chaisemartin† Xavier D'Haultfœuille‡

November 9, 2012

Abstract

Fuzzy Differences in Differences (DID) have often been used to measure treatment effects, but they rely on strong assumptions. They require common trends across groups, both for potential outcomes and for potential treatments. Therefore, they are even more sensitive than standard DID to logs versus levels problems; estimated treatment effects might substantially change when the scale of the outcome is changed. Moreover, quantile treatment effects are not identified in this model. In this paper, we extend the change in change (CIC) model of Athey and Imbens (2006) to tackle those two shortcomings. Under a mild strengthening of their assumptions, we show that 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 also define a stronger set of assumptions under which treatment effects are point identified, even when the treatment rate changes in the control group. We show that simple plug-in estimators of treatment effects are asymptotically normal, and that the bootstrap is valid. Finally, we apply our framework to measure the effectiveness of a treatment for giving up smoking.

Keywords: difference in differences, change in change, imperfect compliance, instrumental variable, quantile treatment effect, partial identification.

JEL Codes: C21, C23

*We are very grateful to Alberto Abadie, Joshua Angrist, Marc Gurgand, Guido Imbens, Thierry Magnac, Bernard Salanié, Frank Vella, Eric Verhoogen, 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 the Paris School of Economics, CREST, Columbia University and MIT for their helpful comments.

†Paris School of Economics, chaisemartin@pse.ens.fr

‡CREST, xavier.dhaultfoeuille@ensae.fr

1 Introduction

Differences in the evolution of treatment rates across groups are one possible source of variation to measure treatment effects. This method has sometimes been referred to as a fuzzy difference in differences (DID). It has often been used in empirical work, even though it relies on strong functional form assumptions, and only identifies average treatment effects. This paper develops an alternative model which tackles those two issues.

Fuzzy DID have been extensively used in empirical research. The best-known example is Duflo (2001). She studies a school construction program in Indonesia, which led to the construction of many schools in districts where few schools were previously available, while few schools were constructed in districts where many schools were already available. Consequently, educational attainment increased more in the former group of districts than in the latter. She uses this to estimate returns to education. Other examples include papers using differential evolution of exposure to some treatment over time across US states (see Evans & Ringel, 1999, among many others). All those papers build upon the some kind of fuzzy DID intuition.

But fuzzy DID relies on strong assumptions. Duflo uses an instrumental variable regression to estimate returns to education, in which the excluded instrument for educational attainment is a group-time interaction. In her most basic model, the resulting coefficient is equal to the DID of the outcome Y, divided by the DID of the treatment D. This is merely the standard Wald ratio, but where Y and D are replaced by their first differences ΔY and ΔD, and where the instrument is group. This quantity is sometimes referred to as the Wald-DID. Imbens & Angrist (1994) have shown that the Wald ratio identifies a local average treatment effect if, among other conditions, the instrument is independent of potential outcomes and treatments. Here, this means that group should be independent of the first differences of both potential outcomes and treatments. This implies that fuzzy DID relies on two common trend assumptions, on potential outcomes and treatments.

Common trend assumptions have an important drawback: they are not invariant to monotonous transformations of the outcome. As shown in Athey & Imbens (2002), they require that the Data Generating Process (DGP) be additively separable in time and group, which can not be true for both the outcome and its logarithm. In practice, this leads to a problem, which is well known in the literature as the logs versus levels problem; when considering the level of the outcome or its rate of growth, treatment effects estimated through DID might considerably change. For instance, Meyer et al. (1995) find no significant effect of injury benefits on injury duration, while they find strong effects on the logarithm of injury duration.

To handle this issue, Athey & Imbens (2006) developed a non-linear DID model, the Change in Change (CIC) model. But the CIC model requires that one group gets fully treated in period 1, while the other group remains fully untreated, so that it does not apply to "fuzzy situations", where the treatment rate only increases more in one group.

Another issue with fuzzy DID is that quantile treatment effects (QTE) are not identified. Indeed, standard quantile-IV models (see Abadie et al., 2002 and Chernozhukov & Hansen, 2005) require that the instrument be independent of potential outcomes and treatments, while here, the most we can reasonably assume is that the instrument (group) is independent of trends on potential outcomes and treatments.

Consequently, we develop a model which tackles those two issues: its assumptions are invariant to a monotonous transformation of the outcome, and it identifies QTE. Our model combines ideas from Imbens & Rubin (1997) and Athey & Imbens (2006), which is the reason why we call it the fuzzy Change in Change model.

The basic version of the change in change model of Athey & Imbens (2006) is a two groups-two periods model, with one treatment group getting fully treated in period 1, while the control group remains fully untreated. They posit a "production function" type of DGP, in which potential outcomes are a strictly increasing function of some unobserved ability index. Their identifying assumption is that the distribution of ability is stable over time in each group. This implies that if the treatment had not been released in period 1, the relative ranking of treatment and control group subjects would have remained unchanged between the two periods. Assume for instance that in period 0, among subjects belonging to the first decile of the distribution of the outcome, 30% belonged to the treatment group, while 70% belonged to the control group. Their assumption merely states that if the treatment had not been released in period 1, the first decile of the outcome would still have comprised 30% of treatment group subjects, and 70% of control group subjects. As a result, treatment effects are identified out of the changes in the relative ranking of the two groups over the two periods.

Our model is a mild strengthening of theirs. Just as them, we posit a "production function" type of DGP for the potential outcomes. But since we want to tackle situations in which treatment is not a deterministic function of time and group, we also posit a rational choice model for potential treatments, in which a subject gets treated if her propensity for treatment is above some threshold. This model is inspired from Vytlacil (2002). This implies that it is equivalent to a "no-defier" type of condition (see Imbens & Angrist, 1994). Then, we slightly strengthen their identifying assumption by assuming that the distribution of ability and propensity for treatment is stable over time in each group.

In our model, treatment effects are point identified when the treatment rate is stable over time in the control group, and partially identified otherwise. Our assumptions imply that if the treatment rate had not changed in any of the two groups between the two periods, the relative ranking of treatment and control subjects would have remained unchanged between period 0 and 1. As a result, when the treatment rate does not change in the control group, treatment effects among a population of compliers are point identified. In such instances, changes in the relative ranking of treatment and control group subjects are solely attributable to the increase of the treatment rate in the treatment group. On the contrary, when the treatment rate changes in the control group, the same LATE and QTE are partially identified, since changes in the relative ranking of treatment and control group subjects can be attributed to changes of the treatment rate in both groups. The smaller the change of the treatment rate in the control group, the tighter our bounds.

Then, we derive several other identification results. We start giving sufficient conditions under which our bounds are sharp. We also present a stronger set of assumptions, under which treatment effects are always point identified, even when the treatment rate changes in the control group. Finally, we show that our fuzzy CIC model is testable. When it is point identified, its testable implication is very similar to the testable implication of Angrist et al. (1996) model, which has been studied by Kitagawa (2008). But when the model is partially identified, the testable implication is somewhat different.

We also develop inference on LATE and QTE. 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 evaluate the effectiveness of Varenicline, a pharmacotherapy for smoking cessation. Our results show evidence of highly heterogeneous treatment effects. We also find a significantly lower average effect of Varenicline compared to what would be obtained through a standard Fuzzy DID approach.

The remainder of the paper is organized as follows. Section 2 is devoted to identification. Section 3 deals with inference. In section 4 we apply our results to measure the efficacy of Varenicline. Section 5 concludes.

2 Identification of the Fuzzy CIC model

2.1 The model

Let $T \in \{0;1\}$ denote time and $G \in \{0;1\}$ denote treatment (1) and control (0) groups. Treatment status D is supposed to be binary. In our fuzzy change-in-change model, some individuals may be treated in the control group or at period 0, and all individuals are not necessarily treated in the treatment group at period 1. However, we assume that at period 1, individuals from the treatment group receive extra incentives to get treated. We model this by introducing the binary instrument $Z = T \times G$ indicating whether people receive an extra incentive to get treated or not. For instance, in Duflo (2001) Z corresponds to the fact that in period 1 an intensive school construction program was implemented in treatment districts and not in control ones, giving supplementary incentives to go to school to children living in those districts. The two corresponding potential treatments, $D(1)$ and $D(0)$, stand for treatment status with and without this supplementary incentive. Observed treatment is $D = ZD(1) + (1-Z)D(0)$. $Y(d)$ corresponds to the potential outcome for an individual with treatment status d. Implicit in this notation is the exclusion restriction that the instrument does not affect the outcome directly. We only observe the outcome $Y = DY(1) + (1-D)Y(0)$.

For any random variables (R, 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), we introduce, for any random variable R, 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 an 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 every $q < 0$, and $F_X^{-1}(q) = \sup \mathcal{S}(X)$ for every $q > 1$.

Let $\lambda_d = \frac{P(D_{01}=d)}{P(D_{00}=d)}$. 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. Let $\mu_d = \frac{P(D_{11}=d)}{P(D_{10}=d)}$ be the equivalent of λ_d for the treatment group.

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

$$\forall d \in \{0;1\}, \; Y(d) = h_d(U_d, T)$$

U_d can be seen as an ability index which determines potential outcomes. Then, we consider the following assumptions.

A.1 Monotonicity

$\forall d \in \{0;1\}$, $h_d(u,t)$ is strictly increasing in u for every $t \in \{0,1\}$.

A.2 No defiers

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

A.3 Time invariance within groups

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

Remarks on these assumptions are in order. Firstly, our results would also hold if U_d were indexed by T, allowing observations to have a different ability level at each period. We suppress this dependence only for the ease of exposition. But this means that A.1 alone is a mere representation. We could set $h_d(u,t) = F^{-1}_{Y(d)|T=t}(u)$ and $U_d^t = F_{Y(d)|T=t}(Y(d))$. A.2 is the same as in Imbens & Angrist (1994). As shown by Vytlacil (2002), it is equivalent to supposing the threshold model $D(z) = 1\{V \geq v_z(T)\}$ for potential treatments, with $v_0(t) \geq v_1(t)$ for every $t \in \{0;1\}$.[1] V can be interpreted as the individual propensity to be treated. The threshold model representation shows that we are imposing a (non-strict) monotonicity on $D(z)$, just as we do for $Y(d)$ in A.1.

Basically, A.3 states that groups remain stable over time. It implies that $U_d \perp\!\!\!\perp T \,|\, G$ and $V \perp\!\!\!\perp T \,|\, G$. As a result, A.1-A.3 impose a standard CIC model both on Y and on D. But A.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. In the standard CIC model, $U_d \perp\!\!\!\perp T \,|\, G$ implies that the relative ranking of treatment and control observations would not have changed from period 0 to 1 if the treatment had not been released, so that treatment effects are identified out of those changes in relative rankings. Here, our joint independence assumption implies that this relative ranking would have remained unchanged if treatment rates had not changed in the control and treatment groups. Putting it in other words, if treatment group observations accounted for 70% of the top decile of Y in period 0, they would also have made up 70% of this top decile in period 1 if the treatment rate had not changed in any of the two groups. Therefore, we will be able to attribute any change in the relative ranking of control and treatment group observations over time to a change in the treatment rate either in the control or in the treatment group.

This implies that when the treatment rate does not change in the control group, changes

[1] V as well should be indexed by T, but we drop this notation to alleviate the notational burden.

in the relative ranking of treatment and control group observations over time can only come from the increase of the treatment rate in the treatment group. Therefore, treatment effects are point identified in this special case. On the contrary, when the treatment rate also changes in the control group, changes in this relative ranking can either be due to the change of the treatment rate in the control or in the treatment group. Consequently, we end up with one equation with two unknowns, so that treatment effects are partially identified. This is the key intuition underlying the paper. Hereafter, we refer to assumptions A.1 to A.3 as to the fuzzy CIC model. On top of those 3 Assumptions, we also take two assumptions testable from the data.

A.4 Data restrictions

1. $\mathcal{S}(Y_{tg}|D=d) = \mathcal{S}(Y) = [\underline{y}, \overline{y}]$ with $(\underline{y}, \overline{y}) \in \overline{\mathbb{R}}^2$, for $(t, g, d) \in \{0; 1\}^3$.
2. $F_{Y_{tg}|D=d}$ is strictly increasing and continuous on $\mathcal{S}(Y)$, for $(g, t, d) \in \{0; 1\}^3$.

A.5 Changes in the treatment rates

1. $P(D_{11} = 1) - P(D_{10} = 1) > 0$.
2. When $P(D_{10} = 1) > 0$ and $P(D_{00} = 1) > 0$, $\mu_1 > \lambda_1$.

The first point of A.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.

Finally, to simplify the exposition, we also assume that the treatment rate increases in the treatment group. This is without loss of generality: if the treatment rate decreases in the treatment group, it suffices to invert the definition of treatment. We also assume that the treatment rate increases relatively more in the treatment than in the control group, which is also without loss of generality. If this is not satisfied, it suffices to switch the two groups.

Overall, our fuzzy CIC model appears as a very natural extension of the CIC analysis to the fuzzy case. Actually, one can show that the fuzzy CIC model reduces to the assumptions of Athey & Imbens (2006) when $P(D_{11} = 1) = 1$ and $P(D_{10} = 1) = P(D_{01} = 1) = P(D_{00} = 1) = 0$, namely when their "sharp" setting holds.

Before getting to the identification results, it is useful to define five subpopulations of interest. Assumption A.3 implies that $P(D_{10} = 1) = P(V \geq v_0(0)|G = 1)$, and similarly $P(D_{11} = 1) = P(V \geq v_1(1)|G = 1)$. Therefore, under assumption A.5 we have that $v_0(0) >$

$v_1(1)$. Similarly, if the treatment rate increases (resp. decreases) in the control group, this implies that $v_0(0) > v_0(1)$ (resp. $v_0(0) < v_0(1)$). Finally, assumption A.2 implies $v_1(1) \leq v_0(1)$. Now, we denote AT and NT the events $V \geq v_0(0)$ and $V < v_1(1)$. AT corresponds to always takers in period 0, i.e. to observations who get treated in period 0 even without receiving any incentive for treatment. NT corresponds to never takers in period 1, that is to say to observations who do not get treated in period 1 even after receiving an incentive for treatment. Let us denote $TC = V \in [\min(v_0(0), v_0(1)), \max(v_0(0), v_0(1)))$. TC stands for "time compliers" and represents observations whose treatment status switches because of the effect of time. Finally, let $IC = V \in [v_1(1), v_0(1))$.[2] IC stands for instrument compliers. This population corresponds to compliers as per the definition of Imbens & Angrist (1994), i.e. to observations that become treated through the effect of Z only. However, in our Fuzzy CIC model, we cannot learn anything on this population because we do not know its size. Instead, our identification results focus on observations that satisfy $V \in [v_1(1), v_0(0))$. This corresponds to untreated observations at period 0 that become treated at period 1, through both the effect of Z and time. We refer to those observations as compliers to simplify the exposition, and we let hereafter C denote the event $V \in [v_1(1), v_0(0))$. If the treatment rate increases in the control group (i.e. if $v_0(1) < v_0(0)$), we merely have $C = IC \cup TC$, while if it decreases we have $C = IC \setminus TC$.

Our parameters of interest are the cdf of $Y(1)$ and $Y(0)$ among compliers, as well as the Local Average Treatment Effet (LATE) and Quantile Treatment Effects (QTE) within this population, which are respectively defined by

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

2.2 Point identification results

We first show that when the treatment rate does not change between the two periods, the cdf of $Y(1)$ and $Y(0)$ among compliers are identified in general. Consequently, the LATE and QTE are also point identified. Define $Q_d(y) = F^{-1}_{Y_{01}|D=d} \circ F_{Y_{00}|D=d}(y)$. Q_d is the quantile-quantile tranform of Y from period 0 to 1 in the control group conditional on $D = d$. To each y with rank q in period 0, it associates y' such that y' is 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 a transform which we refer to as the "pseudo" quantile-quantile tranform of Y from period 0 to 1 in the control group conditional on $D = d$. To each rank q in period 0, it associates the rank q' in period 1 such that the corresponding values of the outcome y and y' in period 0 and 1 are the same.

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

Theorem 2.1 *If A.1-A.5 hold and, for $d \in \{0,1\}$, $P(D_{00}=d)=P(D_{01}=d)>0$,*

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

Hence, Δ and τ_q are identified when $1 > P(D_{00}=0)=P(D_{01}=0)>0$. Moreover,

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

Let us explain the intuition underlying this theorem in the general case. We seek to recover the distribution of $Y(1)$ and $Y(0)$ 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 Figure 1, those observations include both compliers and always takers. Consequently, we must "withdraw" from this observed distribution the distribution of $Y(1)$ among always takers, exactly as in Imbens & Rubin (1997). But this last distribution is not observed, hence the need to reconstruct it. For that purpose, we adapt the ideas in Athey & Imbens (2006). As is shown in Figure 1, all treated observations in the control group or in period 0 are always takers, so that the distribution of $Y(1)$ among always takers is identified within those three cells. Since A.3 implies

$$U_1 \perp\!\!\!\perp T|G, V \geq v_0(0),$$

the distribution of U_1 is the same in periods 0 and 1 among always takers in the control group. Similarly, this equation implies that U_1 also has the same distribution in period 0 and 1 among always takers of the treatment group. This implies that the quantile-quantile transform among always takers is the same in the treatment and control groups. As a result, we can identify the distribution of $Y(1)$ among treatment $\times$ period 1 always takers, applying the quantile-quantile transform from period 0 to 1 among treated observations identified from the control group to the distribution of $Y(1)$ among always takers in the treatment group in period 0. Identification of the distribution of $Y(0)$ among compliers of the period 1 $\times$ test group cell is obtained through similar steps.

	Period 0	Period 1
Control Group	30% treated: Always Takers	30% treated: Always Takers
	70% untreated: Never Takers and Compliers	70% untreated: Never Takers and Compliers
Treatment Group	20% treated: Always Takers	65% treated: Always Takers and Compliers
	80% untreated: Never Takers and Compliers	35% Untreated: Never Takers

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

Another way to understand the transform we use to reconstruct the cdf of $Y(1)$ among always takers is to regard it as a double matching. Consider an always taker in the treatment $\times$ period 0 cell. He 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 have the same treatment status. Therefore, under assumption A.1 they must have the same u_1. Second, the control $\times$ period 0 always taker is matched to its rank counterpart among always takers of the control $\times$ period 1 cell (this is merely the quantile-quantile transform). 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 $\times$ period 0 cell always taker would have obtained in period 1. Therefore, to recover the whole distribution of $Y(1)$ in period 1 among test group always takers, we translate the whole distribution of always takers in the period 0 $\times$ test group cell from y to the corresponding y^* for each value of y.

Note that the formula of the LATE shows an interesting parallel with the standard Wald parameter in the Imbens-Angrist IV model. Once noted that conditional on $G = 1$, $Z = T$, we have

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

The standard Wald parameter has the same expression, except that here Y is replaced by $Q_D(Y)$ in the second term of the numerator. The reason why the standard Wald parameter does not identify a causal effect here is that conditional on $G = 1$, Z (i.e. T) is not independent of $Y(d)$ because the distributions of potential outcomes might evolve with time. To take into account the effect of time on the distribution of potential outcomes, we apply the quantile-quantile transform observed between period 0 and 1 in the control group to the distribution of

Y in period 0 in the treatment group. Indeed, Assumptions A.1 and A.3 ensure that quantile-quantile transforms are the same in the two groups. Likewise, the formulae of the cdf of $Y(1)$ and $Y(0)$ among compliers are very similar to those obtained in Imbens & Rubin (1997). For instance, the cdf of $Y(1)$ rewrites as

$$\frac{P(D=1|G=1,Z=1)F_{Y|D=1,G=1,Z=1}(y)-P(D=1|G=1,Z=0)F_{Q_1(Y)|D=1,G=1,Z=0}(y)}{P(D=1|G=1,Z=1)-P(D=1|G=1,Z=0)}.$$

The cdf of $Y(1)$ in the Imbens and Angrist IV model has the same expression except that $Q_1(Y)$ is replaced by Y in the second term of the numerator. Here again, this is to account for the fact that conditional on $G=1$, the instrument T is not independent of potential outcomes.

Point identification of the LATE, QTE and the cdf of compliers holds when $1>P(D_{00}=0)=P(D_{01}=0)>0$, but not in the extreme cases where $P(D_{00}=0)=P(D_{01}=0)\in\{0,1\}$. In such situations, $F_{Y_{11}(0)|C}$ or $F_{Y_{11}(1)|C}$ is point identified, but not both in general. The other cdf is partially identified only, the bounds being given in the following subsection. An important exception worth mentioning is when $P(D_{00}=0)=P(D_{01}=0)=P(D_{10}=0)=0$, corresponding to treatments appearing at period 1, and for which the control group is not eligible. In such situations, $F_{Y_{11}(0)|C}$ is identified by Theorem 2.2, while $F_{Y_{11}|C}$ is also identified by

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

since there is no always takers in the treatment group. This case generalizes the one considered by Athey & Imbens (2006) to situations with never takers in the treatment group.

We now consider a supplementary assumption under which treatment effects are point identified even when the treatment rate changes in the control group. A.3 implies $U_d \perp\!\!\!\perp T|G=0,1\{V<v_0(0)\}$. When the treatment rate does not change in the control group, $v_0(0)=v_0(1)$, so that $U_d \perp\!\!\!\perp T|G=0,1\{V<v_0(0)\}$ rewrites as $U_d \perp\!\!\!\perp T|G=0,D(0)$. Therefore, when the treatment rate does not change in the control group, A.3 means that treated (resp. untreated) observations in the control group are "similar" in period 0 and 1: the distribution of U_1 (resp. U_0) should be the same in those two populations. When the treatment rate changes in the control group, it seems more difficult to argue that those two groups are similar. However, it might be reasonable to claim that conditional on some covariates, they are similar. Under this supplementary assumption, the cdf of $Y(1)$ and $Y(0)$ are point identified even when the treatment rate changes in the control group. Identification is obtained reweighting period 1 observations to account for the fact that their distribution of X is not necessarily the same as in period 0, exactly as in a standard matching. Let us consider the two following assumptions:

A.6 Conditional time invariance

$U_d \perp\!\!\!\perp T|G=0, D(0)=d, X.$

A.7 Strictly increasing cdf and invariant support

1. $F_{U_d|G=0,D(0)=d,X}$ is strictly increasing almost surely.

2. $\mathcal{S}(U_d|G,T,D(0),X) = \mathcal{S}(U_d|D(0))$ almost surely.

3. $\mathcal{S}(X|G=0,T=0,D=d) = \mathcal{S}(X|G=0,T=1,D=d)$.

Theorem 2.2 *If A.1-A.7 hold and* $0 < P(D_{00}=d)$*,*

$$F_{Y_{11}(d)|C}(y) = \frac{P(D_{10}=d)H_d \circ E\left(F_{Y_{01}|D=d,X}(y)|D_{00}=d\right) - P(D_{11}=d)F_{Y_{11}|D=d}(y)}{P(D_{10}=d)-P(D_{11}=d)}.$$

The only difference between the formulae for $F_{Y_{11}(d)|C}$ in Theorems 2.2 and 2.1 is that $F_{Y_{01}|D=d}(y)$ has been replaced by $F_{Y_{01}|D=d,X}(y)$ integrated over the distribution of X among observations such that $D_{00}=d$. This accounts for the fact that the distribution of X might not be the same among untreated (resp. treated) observations of the control group in period 0 and 1.

Assumption A.7 is mostly technical. Assumption A.6, on the other hand, is a strong assumption. To understand it better, suppose for instance that $P(D_{01}=0) > P(D_{00}=0)$. Then, a sufficient condition for assumption A.6 to hold is

$$U_d \perp\!\!\!\perp V|G=0, T=1-d, D(0)=d, X, \quad d \in \{0,1\}.$$

To some extent, it has a flavor of the ignorability condition in the matching model of Rosenbaum & Rubin (1983) even though we believe it is weaker. Ignorability states that selection is exogenous once controlling for X. Here we only posit that the propensity to be treated is exogenous, once controlling for both X and $D(0)$.

2.3 Partial identification

When $P(D_{00}=d) = P(D_{01}=d) > 0$, $F_{Y_{11}(d)|C}$ is identified under A.1-A.5. We show now that if this condition is not verified, the functions $F_{Y_{11}(d)|C}$ are still partially identified. To derive bounds on these cdf and then on the LATE and QTE, we first relate $F_{Y_{11}(d)|C}$ with the observed distributions and one unidentified cdf.

First, when $P(D_{00}=d) \neq P(D_{01}=d)$ and $P(D_{00}=d) > 0$, the first matching works as in the previous section but the second one collapses. Indeed, in such instances, we no longer have $v_0(1) = v_0(0)$. Among treated observations in the control $\times$ period 0 cell, U_1 is distributed conditional on $G=0, V \geq v_0(0)$, while it is distributed conditional on $G=0, V \geq v_0(1)$ in period 1. This implies that we cannot match period 0 and period 1 observations on their rank. For instance, when the treatment rate increases in the control group, treated observations in the control group include only always takers in period 0, while in period 1 they also include time compliers, as is shown in Figure 2. However, under Assumption A.3 the share of time compliers among treated observations in the control group in period 1 is known. Therefore, as shown in the following Lemma, under Assumptions A.1-A.5, the distributions of potential outcomes among compliers can be written as functions of observed distributions and of $F_{Y_{01}(d)|TC}$, in a formula where $F_{Y_{01}(d)|TC}$ enters with a weight identified from the data.

	Period 0	Period 1
Control Group	30% treated: Always Takers 70% untreated: Never Takers, Instrument Compliers and Time Compliers	35% treated: Always Takers and Time Compliers 65% untreated: Never Takers and Instrument Compliers
Treatment Group	25% treated: Always Takers 75% untreated: Never Takers, Instrument Compliers and Time Compliers	60% treated: Always Takers, Instrument Compliers and Time Compliers 40% Untreated: Never Takers

$P(D_{01} = 1) \geq P(D_{00} = 1)$

	Period 0	Period 1
Control Group	35% treated: Always Takers and Time Compliers 65% untreated: Never Takers and Instrument Compliers	30% treated: Always Takers 70% untreated: Never Takers, Instrument Compliers and Time Compliers
Treatment Group	25% treated: Always Takers and Time Compliers 75% untreated: Never Takers and Instrument Compliers.	60% treated: Always Takers and Instrument Compliers 40% Untreated: Never Takers and Time Compliers

$P(D_{01} = 1) < P(D_{00} = 1)$

Figure 2: Populations of interest.

The other case we have to consider is when $P(D_{00} = d) = 0$ (whether $P(D_{00} = d) =$

$P(D_{01} = d)$ or not). In this case, the first step of the aforementioned double matching collapses for the distribution of $Y(d)$. Indeed, 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, we can use the fact that 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 to bound $F_{Y_{11}(d)|C}$. The following lemma summarizes these results.

Lemma 2.1 *If A.1-A.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* $P(D_{00} = d) = 0$,

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

From this Lemma, it appears that when $P(D_{00} = d) > 0$, we merely need to bound $F_{Y_{01}(d)|TC}$ to derive bounds for $F_{Y_{11}(d)|C}$. In order 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

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

The fact that T_d, $G_d(T_d)$ and $C_d(T_d)$ should all be included between 0 and 1 imposes several restrictions on T_d from which we derive our bounds.

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

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

When $P(D_{00} = d) > 0$, we can bound $F_{Y_{11}(d)|C}$ by $C_d(\underline{T}_d)$ and $C_d(\overline{T}_d)$. These bounds can however be improved by remarking that $F_{Y_{11}(d)|C}$ is increasing. Therefore, we finally define our bounds as:

$$\begin{aligned} \underline{B}_d(y) &= \sup_{y' \leq y} C_d\left(\underline{T}_d\right)(y'), \\ \overline{B}_d(y) &= \inf_{y' \geq y} C_d\left(\overline{T}_d\right)(y'). \end{aligned} \tag{2}$$

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

$$\underline{B}_d(y)=M_0\left(\frac{P(D_{10}=d)-P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}\right),\quad \overline{B}_d(y)=m_1\left(\frac{-P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}\right).$$

When $d=0$, the bounds are actually trivial since $\underline{B}_0(y)=0$ and $\overline{B}_1(y)=1$.

Another important case where the bounds take a simple form is when $P(D_{10}=1)=0$. In this case, one can check that

$$\underline{B}_1=\overline{B}_1=F_{Y_{11}|D=1}.$$

This was expected, because as explained in the previous section, $F_{Y_{11}|C}$ is identified by (1).

Theorem 2.3 proves that $\underline{B}_d$ and $\overline{B}_d$ are indeed bounds for $F_{Y_{11}(d)|C}$. We also consider the issue of whether these bounds are sharp or not. Hereafter, we say that $\underline{B}_d$ is sharp (and similarly for $\overline{B}_d$) if there exists a sequence of cdf $(G_n)_{n\in\mathbb{N}}$ such that supposing $F_{Y_{11}(d)|C}=G_n$ is compatible with both the data and the model, and for all y, $\lim_{n\to\infty}G_n(y)=\underline{B}_d(y)$. We establish that $\underline{B}_d$ and $\overline{B}_d$ are sharp under Assumption A.8 below. Note that this assumption is testable from the data.

A.8 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) $\underline{T}_d$, $G_d(\underline{T}_d)$ and $C_d(\underline{T}_d)$ (resp. $\overline{T}_d$, $G_d(\overline{T}_d)$ and $C_d(\overline{T}_d)$) are increasing.

Theorem 2.3 *If A.1-A.5 hold, we have*

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

Moreover, if A.8 holds, $\underline{B}_d(y)$ and $\overline{B}_d(y)$ are sharp.

To understand the sharpness result, let us define $\mathcal{T}_d$, the set of all functions T_d increasing and included between 0 and 1 such that $G_d(T_d)$ and $C_d(T_d)$ are also increasing and included between 0 and 1. $\mathcal{T}_0$ is the set of all cdf $F_{Y_{01}(0)|TC}$ such that $F_{Y_{01}(0)|V<v_0(0)}$ and $F_{Y_{11}(0)|C}$ are cdf. Putting it in other words, $\mathcal{T}_d$ is the set of all candidates for $F_{Y_{01}(d)|TC}$ which can be rationalized by the data.

Assume that there exists T_0^- and T_0^+ in $\mathcal{T}_0$ such that $T_0^-\leq T_0\leq T_0^+$ for every $T_0\in\mathcal{T}_0$. When $\lambda_0>1$, $G_0(.)$ is decreasing in T_0, which implies that $C_0(.)$ is also decreasing in T_0. Therefore, in such instances the sharp lower bound is equal to $C_0(T_0^+)$, while the sharp upper bound is equal to $C_0(T_0^-)$. Moreover, when $\lambda_0>1$, it appears after some algebra that T_0, $G_0(T_0)$ and $C_0(T_0)$ are all included between 0 and 1 if and only if $\overline{T}_0\leq T_0\leq\underline{T}_0$. Therefore, if $\underline{T}_0$, $G_0(\underline{T}_0)$ and $C_0(\underline{T}_0)$ are increasing, $\underline{T}_0$ is in $\mathcal{T}_0$ and $T_0^+ = \underline{T}_0$. This suggests that $\underline{B}_0 = C_0(\underline{T}_0)$ is sharp.

To establish this rigorously, we have to build a DGP rationalizing $\underline{B}_0$. When $\lambda_0 > 1$, time compliers are untreated in period 1 as shown in Figure 2. Therefore, the data imposes several constraints on $F_{Y_{01}(0)|TC}$. Indeed, untreated observations in the control group in period 1 are either time compliers or observations which verify $V < v_0(0)$. For each value of y, we need to allocate the total weight of the density $F'_{Y_{01}|D=0}$ across those two populations, since

$$F'_{Y_{01}|D=0} = \frac{1}{\lambda_0} F'_{Y_{01}(0)|V<v_0(0)} + (1 - \frac{1}{\lambda_0}) F'_{Y_{01}(0)|TC}.$$

For low values of y, we allocate as much weight as possible to the population of time compliers and as little weight as possible to the other population by setting $F'_{Y_{01}(0)|V<v_0(0)}(y)$ at the lowest possible value which still ensures that $F_{Y_{11}(0)|C}$ is constant and equal to 0, while $F_{Y_{01}(0)|TC}(y)$ grows at the fastest possible pace. For some value of y, $F_{Y_{01}(0)|TC}(y)$ reaches 1, which means that we have allocated a sufficient weight to the population of time compliers. Let us denote by y^* this value. For all $y \geq y^*$, we set $F'_{Y_{01}(0)|V<v_0(0)}(y) = \lambda_0 F'_{Y_{01}|D=0}(y)$, which means that we consider that all untreated observations in the control group in period 1 with a Y above y^* are such that $V < v_0(0)$.

When $\lambda_0 < 1$, $G_0(.)$ is increasing in T_0, which implies that $C_0(.)$ is also increasing in T_0. Therefore, in such instances the sharp lower bound is equal to $C_0(T_0^-)$, while the sharp upper bound is equal to $C_0(T_0^+)$. When $\lambda_0 < 1$, T_0, $G_0(T_0)$ and $C_0(T_0)$ are all included between 0 and 1 if and only if $\underline{T}_0 \leq T_0 \leq \overline{T}_0$. Therefore, if $\underline{T}_0$, $G_0(\underline{T}_0)$ and $C_0(\underline{T}_0)$ are increasing, $\underline{T}_0$ is in $\mathcal{T}_0$ and $T_0^- = \underline{T}_0$. This suggests that $\underline{B}_0 = C_0(\underline{T}_0)$ is sharp. In this case, defining a DGP rationalizing $\underline{B}_0$ essentially amounts to set $F_{Y_{01}(0)|TC}$ to 0. Indeed, when $\lambda_0 < 1$, Time Compliers in the control group are treated in period 1, so that the data does not impose any restriction on $F_{Y_{01}(0)|TC}$. However, setting $F_{Y_{01}(0)|TC} = 0$ is not compatible with the fact that the function $h_0(., 1)$ must be strictly increasing. Therefore, we construct a sequence of DGP such that T_d^n converges towards $\underline{T}_d$ on $\overset{\circ}{\mathcal{S}(Y)}$, which proves sharpness on $\overset{\circ}{\mathcal{S}(Y)}$.

A consequence of Theorem 2.3 is that QTE and LATE are partially identified when $P(D_{00} = 0) \neq P(D_{01} = 0)$ or $P(D_{00} = 0) \in \{0, 1\}$. The bounds are given in the following corollary. To ensure that the bounds on the LATE are well defined, we impose the following technical condition.

A.9 Existence of moments We have $\int |y| d\overline{B}_1(y) < +\infty$ and $\int |y| d\underline{B}_1(y) < +\infty$.[3]

[3] $\int |y| d\overline{B}_1(y)$ is the integral of the absolute value function with respect to the probability measure ν defined on $[\underline{y}, \overline{y}]$ and generated by $\overline{B}_1$. The same holds for $\int |y| d\underline{B}_1(y)$, $\int |y| d\overline{B}_0(y)$ and $\int |y| d\underline{B}_0(y)$. Because we may have $\lim_{y \to \underline{y}} \overline{B}_0(y) > 0$ or $\lim_{y \to \overline{y}} \overline{B}_0(y) < 1$, ν may admit a mass at $\underline{y}$ or $\overline{y}$.

Corollary 2.4 *If A.1-A.5 and A.9 hold,* $P(D_{00}=0) \neq P(D_{01}=0)$ *or* $P(D_{00}=0) \in \{0,1\}$, 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, suppose that A.8 holds. Then

- *If* $\lambda_0 > 1$ *or* $E(|Y_{11}(0)|\,|C) < +\infty$, *the bounds on* Δ *are sharp.*

- *If* $\lambda_0 > 1$ *or for* $d \in \{0,1\}$, $\underline{B}_d(y) = q$ *and* $\overline{B}_d(y) = q$ *admit a unique solution, the bounds on* τ_q *are sharp.*

Interestingly, $\underline{B}_0$ and $\overline{B}_0$ are defective when $\lambda_0 < 1$: $\lim_{y \to \overline{y}} \underline{B}_0(y) < 1$ and $\lim_{y \to \underline{y}} \overline{B}_0(y) > 0$. On the contrary, $\underline{B}_1$ and $\overline{B}_1$ are proper cdf. The reason for this asymmetry is that when $\lambda_0 < 1$, time compliers belong to the group of treated observations in the control $\times$ period 1 cell. Moreover, under assumption A.3, we know that they represent $\frac{1}{\lambda_1}$ % of this group. Consequently, the data imposes some restrictions on $F_{Y_{01}(1)|TC}$. For instance, $F_{Y_{01}(1)|TC} \geq F_{Y_{01}|D=1,Y \geq \alpha}$, where $\alpha = F^{-1}_{Y_{01}|D=1}(\frac{1}{\lambda_1})$, which means that time compliers' cannot have higher Y than the $1 - \frac{1}{\lambda_1}$ % of observations with highest Y of this group. On the contrary, their $Y(0)$ is not observed. Therefore, the data does not impose any restriction on $F_{Y_{01}(0)|TC}$: it could be arbitrarily close to 0, which corresponds to the case where their $Y(0)$ tends to $+\infty$, hence the defective bounds for $F_{Y_{11}(0)|C}$.

When $\lambda_0 > 1$, $\underline{B}_0$, $\overline{B}_0$, $\underline{B}_1$ and $\overline{B}_1$ are all proper cdf. We could have expected $\underline{B}_1$ and $\overline{B}_1$ to be defective because when $\lambda_0 > 1$, time compliers are untreated in period 1, so that their $Y(1)$ is unobserved. This second asymmetry stems from the fact that when $\lambda_0 > 1$, time compliers do not belong to our population of compliers ($C = IC \setminus TC$), while when $\lambda_0 < 1$, time compliers are included within our population of interest ($C = IC \cup TC$). Setting $F_{Y_{01}(1)|TC}(y) = 0$ does not imply that $\lim_{y \to +\infty} F_{Y_{11}(1)|C}(y) < 1$ when $TC \cap C$ is empty, while setting $F_{Y_{01}(0)|TC}(y) = 0$ yields $\lim_{y \to +\infty} F_{Y_{11}(0)|C}(y) < 1$ when $TC \subset C$.[4]

As a result, when $\lambda_0 < 1$ and $\mathcal{S}(Y)$ is unbounded, the bounds on Δ are infinite, and some bounds on τ_q are also infinite. On the contrary, when $\lambda_0 > 1$ the bounds on τ_q are always finite, for every $q \in (0,1)$. The bounds on the LATE may also be finite in this case.[5] Table 1

[4] This situation is similar to the one encountered by Horowitz & Manski (1995) and Lee (2009), who show in a related missing data problem that their bounds are not defective, contrary to what usually happens in missing data problems.

[5] $\underline{B}_0$ and $\overline{B}_0$ may still have fat tails and do not admit an expectation.

summarizes the situation. Interestingly, our bounds on Δ and τ_q are sharp in general under A.8, even though they may be infinite when the support is unbounded.

Table 1: Finiteness of the bounds when $\underline{y} = -\infty$, $\overline{y} = +\infty$.

	$\lambda_0 < 1$ $(\lambda_1 > 1)$	$\lambda_0 > 1$ $(\lambda_1 < 1)$
$\underline{\tau}_q$, q small	finite	finite
$\overline{\tau}_q$, q small	$+\infty$	finite
$\underline{\tau}_q$, q large	$-\infty$	finite
$\overline{\tau}_q$, q large	finite	finite
$\underline{\Delta}$	$-\infty$	finite in general
$\overline{\Delta}$	$+\infty$	finite in general

q small means $0 < q < \underline{q}$ for a well chosen $\underline{q}$. Similarly, q large means $\overline{q} < q < 1$ for a well chosen $\overline{q}$.

2.4 Testability

We now show that our fuzzy CIC model is testable. For every $y \leq y'$ in $\mathcal{S}(Y)^2$, let

$$I_d(y, y') = [\min(\underline{T}_d(y), \overline{T}_d(y)), \max(\underline{T}_d(y'), \overline{T}_d(y'))],$$

with the convention that $I_d(y, y') = \emptyset$ if

$$\min(\underline{T}_d(y), \overline{T}_d(y)) > \max(\underline{T}_d(y'), \overline{T}_d(y')).$$

Theorem 2.5 *Testability*

1. *If A.4 holds, we reject A.1-A.3 together if for some $d \in \{0; 1\}$, there exists $y_0 < y_1$ in $\mathcal{S}(Y)^2$ such that one of the two following statements holds:*

   (a) $I_d(y_0, y_1) = \emptyset$

   (b) $I_d(y_0, y_1) \neq \emptyset$, *and for every $t_0 \leq t_1$ in $I_d(y_0, y_1)^2$,*

$$\frac{P(D_{10} = d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_1) + (1 - \lambda_d)t_1) - P(D_{11} = d)F_{Y_{11}|D=d}(y_1)}{P(D_{10} = d) - P(D_{11} = d)}$$
$$< \quad \frac{P(D_{10} = d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_0) + (1 - \lambda_d)t_0) - P(D_{11} = d)F_{Y_{11}|D=d}(y_0)}{P(D_{10} = d) - P(D_{11} = d)}.$$

2. *If A.4 holds, we reject A.1-A.3 and A.6-A.7 together if for some $d \in \{0;1\}$, there exists $y_0 < y_1$ in $\mathcal{S}(Y)^2$ such that*

$$\begin{aligned} & \frac{P(D_{10}=d)H_0 \circ E\left(F_{Y_{01}|D=d,X}(y_1)|D_{00}=d\right) - P(D_{11}=d)F_{Y_{11}|D=d}(y_1)}{P(D_{10}=d)-P(D_{11}=d)} \\ < \quad & \frac{P(D_{10}=d)H_0 \circ E\left(F_{Y_{01}|D=d,X}(y_0)|D_{00}=d\right) - P(D_{11}=d)F_{Y_{11}|D=d}(y_0)}{P(D_{10}=d)-P(D_{11}=d)}. \end{aligned}$$

Theorem 2.5 provides a simple (theoretical) test of the model. When the treatment rate does not change in the control group, i.e. when $\lambda_0 = \lambda_1 = 1$, the testable implication of the fuzzy CIC model is merely that

$$\frac{P(D_{10}=d)H_d \circ F_{Y_{01}|D=d}(y) - P(D_{11}=d)F_{Y_{11}|D=d}(y)}{P(D_{10}=d)-P(D_{11}=d)}$$

must be increasing. Therefore, in such instances, our test is similar to the one developed by Kitagawa (2008) for the treatment effect model of Angrist et al. (1996) with binary treatment and instrument. Indeed, in their model, the cdf of compliers can also be written as the difference between two increasing functions, and thus may not be increasing (see Imbens & Rubin, 1997). Similarly, as shown in point 2 of Theorem 2.5, our test of Assumptions A.1-A.3 and A.6-A.7 is also similar to the one developed by Kitagawa (2008).

On the contrary, when the treatment rate changes in the control group, our test of the fuzzy CIC model is slightly different. In such instances, despite the fact that the functions $F_{Y_{11}(d)|C}$ are not identified, we can still reject the model when either $\mathcal{T}_0$ or $\mathcal{T}_1$ is empty. For instance, $\mathcal{T}_0$ is empty if there is no function $F_{Y_{01}(0)|TC}$ such that $F_{Y_{01}(0)|TC}$, $F_{Y_{01}(0)|V<v_0(0)}$ and $F_{Y_{11}(0)|C}$ are all cdf, while such a function should exist under A.1-A.3 as shown in Lemma 2.1. We give two sufficient conditions under which $\mathcal{T}_0$ is empty. The condition presented in point a) ensures that it is not possible to find a function $F_{Y_{01}(0)|TC}$ increasing and included between 0 and 1, such that $F_{Y_{01}(0)|V<v_0(0)}$ and $F_{Y_{11}(0)|C}$ are also included between 0 and 1. Similarly, if the condition presented in point b) holds, then it is not possible to find a function $F_{Y_{01}(0)|TC}$ increasing and included between 0 and 1, such that $F_{Y_{01}(0)|V<v_0(0)}$ and $F_{Y_{11}(0)|C}$ are also increasing.

Note that the test presented in point b) is not exactly the same when $\lambda_0 > 1$ than when $\lambda_0 < 1$. Indeed, when $\lambda_0 < 1$, $\lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0)t$ is increasing in t. Therefore, one sufficient condition to reject A.1-A.3 together in this case is that there exists $y_0 \leq y_1$ in $\mathcal{S}(Y)^2$ such that

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_d F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)\overline{T}_0(y_1)) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0)-P(D_{11}=0)} \\ < \quad & \frac{P(D_{10}=0)H_0 \circ (\lambda_d F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)\underline{T}_0(y_0)) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0)-P(D_{11}=0)}. \end{aligned}$$

On the contrary, when $\lambda_0 > 1$, $\lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0)t$ is decreasing in t. Then, one sufficient condition to reject A.1-A.3 together is that there exists $y_0 \leq y_1$ in $\mathcal{S}(Y)^2$ such that for every t in $\left[\overline{T}_0(y_1), \underline{T}_0(y_0)\right]$,

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_d F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)t) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0) - P(D_{11}=0)} \\ < \; & \frac{P(D_{10}=0)H_0 \circ (\lambda_d F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)t) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0) - P(D_{11}=0)}. \end{aligned}$$

Therefore, the test will be much simpler to implement when $\lambda_0 < 1$ than when $\lambda_0 > 1$, since for every $y_0 \leq y_1$, it is sufficient to check whether the inequality is verified for only one value of (t_1, t_0). On the contrary, when $\lambda_0 > 1$, for every $y_0 \leq y_1$ we must check whether the inequality is verified for an infinity of t.

3 Inference

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

A.10 $(Y_i, D_i, G_i, T_i)_{i=1,\dots,n}$ are i.i.d.

A.11 $\mathcal{S}(Y)$ 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 $[\underline{y}, \overline{y}]$.

A similar assumption as A.11 is also made by Athey & Imbens (2006) when studying the asymptotic properties of their estimator.

We first consider the point identified case corresponding to either $1 > P(D_{00}=0) = P(D_{01}=0) > 0$, $P(D_{00}=0) = P(D_{01}=0) = P(D_{10}=0) = 0$ or A.7 holds with $X = 1$, because the three cases lead to the same form for the cdf of compliers and the LATE. Let $\widehat{F}_{dgt}$ (resp. $\widehat{F}^{-1}_{dgt}$) denote the empirical cdf 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 $(d, g, t) \in \{0,1\}^3$. Then our estimator of the LATE is defined by:

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

To estimate the QTE, let us define $\widehat{P}(D_{gt} = d)$ as the proportion of $D = d$ in the sample $\mathcal{I}_{gt}$ and

$$\widehat{F}_{Y_{11}(d)|C} = \frac{\widehat{P}(D_{11}=d)\widehat{F}_{d11} - \widehat{P}(D_{01}=1)\widehat{H}_1 \circ \widehat{F}_{d01}}{\widehat{P}(D_{11}=1) - \widehat{P}(D_{10}=1)}.$$

Then the estimator of the QTE is the simple plug-in estimator:

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

We say hereafter that an estimator $\widehat{\theta}$ of a parameter θ_0 is root-n consistent and asymptotically normal if there exists Σ such that $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{L} \mathcal{N}(0,\Sigma)$. Theorem 3.1 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 the parameter θ_0, we say that the bootstrap is consistent if with probability one and conditional on the sample (see, e.g., van der Vaart & Wellner, 1996, chapter , for a formal definition of conditional convergence), $\sqrt{n}(\widehat{\theta}^*-\widehat{\theta})$ converges to the same distribution as the limit distribution of $\sqrt{n}(\widehat{\theta}-\theta_0)$. Theorem 3.1 ensures that bootstrap confidence intervals will be asymptotically valid.

Theorem 3.1 *Suppose that Assumptions A1-A5, A10-A11 hold and either* $1 > P(D_{00}=0) = P(D_{01}=1) > 0$, $P(D_{00}=0) = P(D_{01}=0) = P(D_{10}=0) = 0$ *or A6-A7 holds with* $X=1$. *Then* $\widehat{\Delta}$ *and* $\widehat{\tau}_q$ *are root-n consistent and asymptotically normal. Moreover, the bootstrap is consistent for both* $\widehat{\Delta}$ *and* $\widehat{\tau}_q$.

We now turn to the partially identified case. First, suppose that $0 < \widehat{P}(D_{00}=0) < 1$ and $\widehat{P}(D_{10}=0) < 1$. 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

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

The estimated bounds on $F_{Y_{11}(d)|C}$ are then defined by

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

Finally, the bounds on the average and quantile treatment effects are estimated by

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

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

When $\widehat{P}(D_{00}=0) \in \{0,1\}$ or $\widehat{P}(D_{10}=0)=1$, the bounds on Δ and τ_q are defined similarly, but instead of $\widehat{\underline{B}}_d$ and $\widehat{\overline{B}}_d$, we use the empirical counterparts of the bounds on $F_{Y_{11}(d)|C}$ given by Equation (2).

We let $B_\Delta = (\underline{\Delta}, \overline{\Delta})$ and $B_{\tau_q} = (\underline{\tau}_q, \overline{\tau}_q)'$, $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$ being the corresponding estimators. The following theorem 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 below.

When $P(D_{00} = 0) \in \{0,1\}$, $P(D_{10} = 0) = 1$ $\lambda_0 > 1$, we merely let $\mathcal{Q} = (0,1)$.

When $\lambda_0 < 1$, we should exclude small and large q from $\mathcal{Q}$. The limit distribution of $\widehat{B}_{\tau_q}$ is degenerated for small or large q because the measure corresponding to $\underline{B}_d$ and $\overline{B}_d$ put mass at $\underline{y}$ or $\overline{y}$. 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 because the bounds include in their definitions the kinked functions M_0 and m_1, they may also be irregular at some $q \in (\underline{q}, \overline{q})$.[6] Formally, let

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

Therefore, when $\lambda_0 < 1$, we let $\mathcal{Q} = (\underline{q}, \overline{q}) \backslash \{q_1, q_2\}$. 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 be equal to $(\underline{q}, \overline{q})$. However, we expect in general λ_1 to be close to one and μ_1 to be large, in which case q_1 and q_2 do belong to $(\underline{q}, \overline{q})$.

Theorem 3.2 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 A.8.

A.12 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 3.2 *Suppose that Assumptions A.1-A.5, A.8, A.10-A.12 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.*

A confidence interval of level $1 - \alpha$ on Δ (and similarly for τ_q) can be based on the lower bound of the two-sided (bootstrap) confidence interval of level $1 - \alpha$ of $\underline{\Delta}$, and the upper bound of the two-sided (bootstrap) confidence interval of $\overline{\Delta}$. Such confidence intervals are asymptotically valid but conservative. Because $\underline{\Delta} < \overline{\Delta}$, a confidence interval on Δ could alternatively can be based on one-sided confidence intervals of level $1 - \alpha$ on $\underline{\Delta}$ and $\overline{\Delta}$.[7]

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

[7] As remarked by Imbens & Manski (2004), such confidence intervals suffer however from a lack of uniformity, since their coverage rate falls below the nominal level when one gets close to point identification, i.e. when $\lambda_d \to 1$. We could use instead the proposals of Imbens & Manski (2004) or Stoye (2009). Uniform validity of such confidence intervals would require to strengthen the pointwise normality result of Theorem 3.2 to a

Finally, note that we have implicitly considered that we knew whether point identification or partial identification holds, but this is seldom the case in practice, except if one is ready to posit A.6 and A.7. This issue is important since the estimators and the way confidence intervals are constructed differ in the two cases. Actually, and abstracting from extreme cases where $P(D_{gt} = d) = 0$, testing point identification is simply equivalent to testing $\lambda_0 = 1$ versus $\lambda_0 \neq 1$. $\widehat{\lambda}_0$ is a root-n consistent estimator. Therefore, valid inference can be made by first checking whether $|\widehat{\lambda}_0 - 1| \leq c_n$, with $(c_n)_{n\in\mathbb{N}}$ a sequence satisfying $c_n \to 0$, $\sqrt{n}c_n \to \infty$, and then applying either the point or partial identified framework.

4 Application to the evaluation of a smoking cessation treatment

4.1 Data and methods

In this section, we apply our framework to assess the effect of Varenicline, a pharmacotherapy for smoking cessation support which has been marketed in France since February 2007. Smoking rate among the adult population is around 30% in France, which is much higher than in most western countries (see e.g. Beck et al., 2007). Randomized trials 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 very few studies based on non experimental data to confirm the efficacy of this new drug in real life settings. Moreover, studies on this new drug only investigated its average effect on cessation rate, and none considered potentially heterogeneous effects for instance looking at quantile treatment effects.

We use the database of the French smoking cessation clinics program. During patients' first visit, smoking status is evaluated according to daily cigarettes smoked and a measure of expired carbon monoxide (CO), which is a biomarker for recent tobacco use. At the end of this first visit, treatments may be prescribed to patients (nicotine replacement therapies…). Follow-up visits are offered during which CO measures are usually made to validate tobacco abstinence.

Doctors often discretize this continuous outcome in order to ease its interpretation. Below 5 parts per million, a patient is regarded as a non smoker. She is considered a light smoker between 5 and 10, a smoker between 10 and 20, a heavy smoker between 20 and 30 and a

uniform one, in order to satisfy Assumption 1.i) in Stoye (2009). This is likely to fail for the QTE because of the possible kinks of $\underline{B}_d$ and $\overline{B}_d$ at q_1 or q_2. Because q_1 and q_2 change with the underlying distributions, it is likely that $q = q_1$ or $q = q_2$ for some underlying distribution, violating uniform asymptotic normality.

very heavy smoker above 30.[8] Therefore, a patient with a CO of 25 at last follow-up not only failed to quit but is still a heavy smoker.

Our data consist of 55 clinics for which at least one patient was recorded before and after the release of Varenicline, and which followed at least 50% of the patients they consulted. The outcome is patients expired CO at their last follow-up visit. 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).

To define the control and treatment groups, we rely on the heterogeneity of the rate of prescription of Varenicline among clinics. The kernel density estimate of this rate of prescription is shown in Figure 3. It has three modes, with a first peak at very low rates of prescription, a second much lower peak around 25% and a third peak around 35-40%. In de Chaisemartin (2011), some elements are given to explain this high degree of heterogeneity in prescription patterns across clinics. It appears that among professionals working in low prescription rate clinics, there are many psychologists while in high prescription rate clinics professionals are mostly physicians. Contrary to nicotine replacement therapies, Varenicline must be prescribed by a doctor. Therefore a first and somewhat mechanical explanation for this difference in prescription rates is that patients in control clinics are less likely to be consulted by a doctor. Still, approximately 47% of them were consulted by a doctor and only 1.7% were prescribed Varenicline. Another difference across professionals working in the two groups of clinics is that most professionals working in low prescription rate clinics are trained to cognitive and behavioral therapies, while few of them received such training in high prescription rate clinics. Therefore, a complementary explanation for this sharp difference in prescription rates is that there might be two approaches to smoking cessation among professionals. The first approach, which seems to be more prominent in high prescription rate clinics, puts the emphasis on providing patients pharmacotherapies to reduce the symptoms of withdrawal. The second approach, which seems more prominent in the other group of clinics, lays more the emphasis on giving them intensive psychological support.

[8] Typically, light smokers smoke less than 10 cigarettes a day, smokers between 10 and 20, heavy smokers between 20 and 40 and very smokers more than 40.

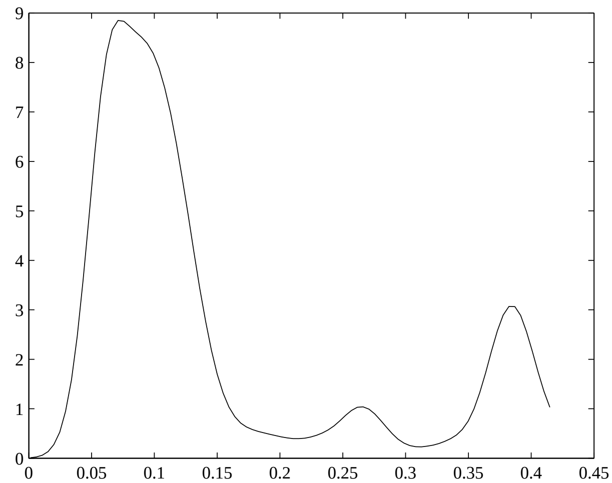

Figure 3: Density of the prescription rate of Varenicline across clinics.

Based on these observations, we define the control group as patients registered by the 15 clinics whose prescription rate is below 3%. Similarly, the treatment group consists of patients recorded by the 13 clinics whose prescription rate is above 20% prescription rate. 15,561 patients consulted those 28 clinics from February 2005 to January 2009. However, many patients never came back for follow-up visits, and we have to exclude them from the analysis. At the end, we rely on 9,324 patients for whom follow-up CO measures were available. Patients consulted in those cessation clinics are middle-aged, educated and very heavy smokers. They have been classified as "hardcore" addicts in the medical literature (Faou et al., 2009). Descriptive statistics on the sample can be found in de Chaisemartin (2011).

By construction, the prescription rate of Varenicline is 0% in control and treatment clinics at period 0. At period 1, it is equal to 1.7% in control clinics and 33.8% in treatment clinics. This sharp rise in Varenicline prescription in treatment clinics entails a strong decrease in the prescription of other treatments such as nicotine patch (de Chaisemartin, 2011).

Table 2: Prescription rate of Varenicline

	$T = 0$	$T = 1$
Control clinics	0.0%	1.7%
Treatment clinics	0.0%	33.8%

4.2 Results

Since $P(D_{10} = 0) = 1$, $F_{Y_{11}(1)|C}(y)$ is point identified under A.1-A.5 by $F_{Y_{11}|D=1}$. On the contrary, since $P(D_{00} = 0) > P(D_{01} = 0)$, $F_{Y_{11}(0)|C}(y)$ is partially identified and we use Theorem 2.3 to estimate bounds for it. The resulting bounds on QTE are displayed in Figure 4, as well as the upper (resp. lower) bound of the 90% confidence interval of the upper (resp. lower) bound, which yields an asymptotically 95% confidence interval for QTE (see Imbens & Manski, 2004). It appears that QTE are not significantly different from 0 below the 50th percentile. This corresponds to the fact that approximately 50% of compliers would have quitted even without taking Varenicline. QTE are significantly negative from the 50th to the 90th percentile and no longer significant after. This means that Varenicline has a significant impact up to the 90th percentile of the distribution of CO but may leave the right tail unaffected.

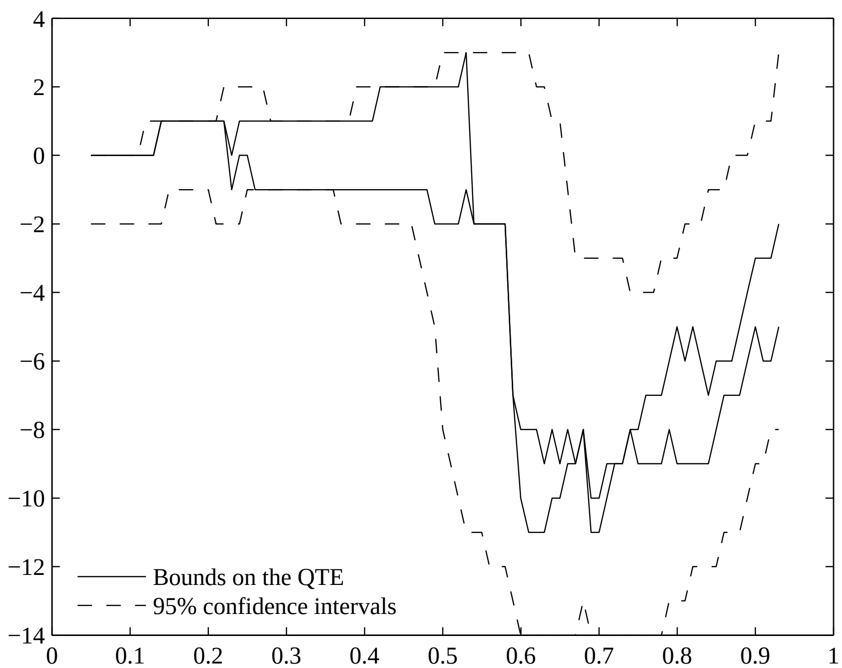

Figure 4: Bounds for QTE of Varenicline on CO at follow-up

We use Theorem 2.2 to identify $F_{Y_{11}(0)|C}(y)$ under A.1-A.7 with $X = 1$. When the set of covariates is empty, Assumption A.7 merely means that time compliers do not differ from never takers. This is a strong Assumption but it seems suitable here as a first order approximation since we have a very small number of time compliers. To further ease the interpretation, we also discretize the outcome. The distributions of $Y(1)$ and $Y(0)$ among compliers estimated through our fuzzy CIC model are displayed in Table 4.2.

Table 4.2: Distributions of Y(0) and Y(1) among compliers

I	$P(Y_{11}(0) \in I \| C)$	$P(Y_{11}(1) \in I \| C)$	P-value
[0;5]	53.5%	63.5%	0.08
(5;10]	3.5%	9.7%	0.22
(10;20]	23.2%	14.0%	0.08
(20;30]	16.1%	7.1%	0.03
(30;+∞)	3.8%	5.7%	0.50

Varenicline increases the share of patients considered as successful quitters (i.e. with a CO at last follow-up below 5 ppm) by 10 percentage points. This increase comes from substantial decreases of the share of patients still considered as smokers (i.e. with a CO included between 10 and 20 ppm) and heavy smokers (i.e. with a CO included between 20 and 30 ppm) at last follow-up. On the contrary, Varenicline had no impact on the share of patients whose cessation attempt failed and remained very heavy smokers.

An alternative way of studying the effect of this treatment on cessation rate is to run a fuzzy DID on $Y^* = 1\{CO \leq 5\}$. The resulting LATE as per this analysis is equal to 0.165, meaning that according to a fuzzy DID model, varenicline increased cessation rate by 16.5 pp among compliers (P-value = 0.01). The difference with our fuzzy CIC estimate is large and almost statistically significant (P-value = 0.08).

4.3 Robustness checks

In de Chaisemartin (2011), many robustness checks are conducted. For instance, the importance of the attrition bias is evaluated computing a placebo DID on the share of patients included in the analysis. The sensitivity of the results to the somewhat arbitrary choice of the 3% and 20% thresholds is investigated conducting the whole analysis again with 8 different pairs of thresholds. All those tests prove conclusive.

Here, we develop another type of robustness test, intended at testing the crucial identifying assumption specific to our fuzzy CIC model. This assumption states that if Varenicline had not been released, the relative ranking of treatment and control patients in terms of their expired CO at follow-up would have remained unchanged between period 0 and 1. In this particular example, this means that if control patients were easier to treat (i.e. less addicted) than treatment patients in period 0, then they should remain equally easier to treat in period 1. This assumption could be violated, for instance if the most addicted patients came to treatment clinics in period 1 because they were aware that they could get a prescription of Varenicline in those clinics and not in control clinics.

Therefore, to assess the credibility of this assumption, we conduct the exact same fuzzy CIC analysis but taking patients' initial CO instead of their last follow-up CO as our outcome. Initial CO level is indeed a good proxy for patients' "toughness" to be treated since it reflects their degree of addiction on the day when they initiated their cessation attempt. If the relative ranking of treatment and control observations on this measure has remained unchanged between period 0 and 1, we should find that placebo QTE on initial CO are not significantly different from 0. The resulting bounds for those "placebo" QTE bounds are displayed in Figure 5 along with the 95% confidence intervals on the QTE. Except on a very small interval, 0 always lies within the confidence interval. This strongly supports the identifying assumption of our fuzzy CIC model.

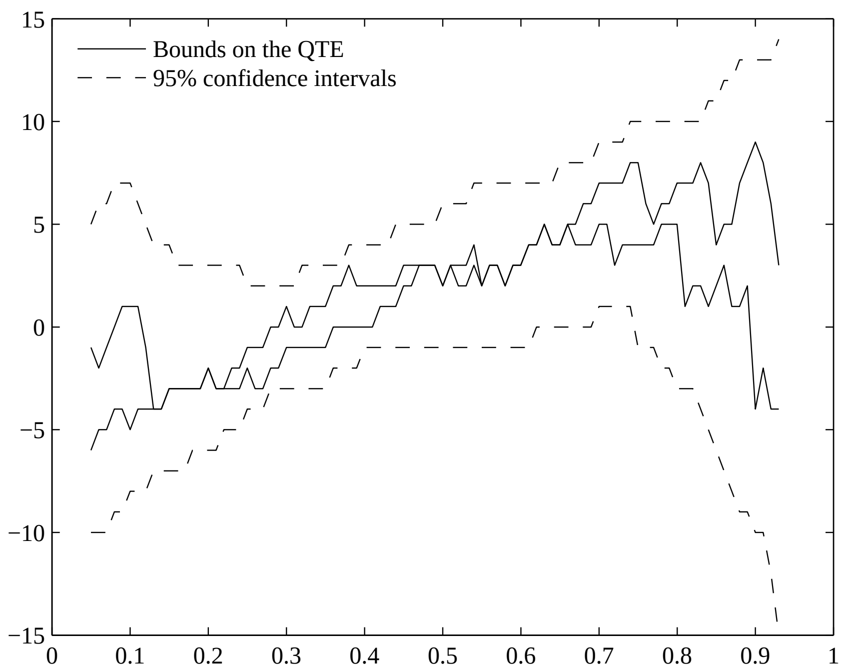

Figure 5: Bounds for QTE of Varenicline on initial CO among compliers

5 Conclusion

In this paper, we develop a fuzzy CIC model to identify treatment effects when the treatment rate increases more in one group than in others. Our model brings two improvements to fuzzy DID, the model currently used in the literature to identify treatment effects in such settings. Its assumptions (which are extremely similar to those of Athey & Imbens (2006)) are invariant to a monotone transformation of the outcome so that results will not suffer from the logs versus levels problem mentioned in the introduction. Moreover, it allows studying

QTE among compliers, while fuzzy DID only identifies a LATE.

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 fuzzy CIC assumptions. Second, we show that 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 give conditions testable from the data under which our bounds are sharp. We derive testable implications of our model in the spirit of Kitagawa (2008). We conduct inference on treatment effects and sharp bounds estimators by proving their asymptotic normality and showing the validity of the bootstrap. Finally, we apply our results to evaluate the effectiveness of a treatment for smoking cessation. Our results significantly differ from those obtained through the standard fuzzy DID approach, and the magnitude of this difference is large. This shows that the restrictive assumptions underlying the standard fuzzy DID approach can have significant policy implications.

Appendix A: proofs

Theorem 2.1

The proof follows directly from Lemma 2.1, by noting that $\lambda_0 = \lambda_1$ when $P(D_{00} = d) = P(D_{01} = d)$.

Theorem 2.2

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

For $t \in \{0, 1\}$, let $h_0^{-1}(., t)$ denote the inverse of $h_0(., t)$. We have almost surely

$$\begin{aligned} F_{Y_{0t}(0)|D(0)=0,X}(y) &= P(h_0(U_0, t) \leq y | G = 0, T = t, D(0) = 0, X) \\ &= P(U_0 \leq h_0^{-1}(y, t) | G = 0, T = t, D(0) = 0, X) \\ &= P(U_0 \leq h_0^{-1}(y, t) | G = 0, D(0) = 0, X) \\ &= F_{U_0|G=0,D(0)=0,X}(h_0^{-1}(y, t)), \end{aligned}$$

where the first equality stems from A.1 and the third from A.6 and A.7. By A.7, $F_{U_0|G=0,D(0)=0,X}$ is strictly increasing. Hence, the inverse of $F_{Y_{0t}(0)|D(0)=0,X}$ exists and for all $q \in (0, 1)$, almost surely

$$F^{-1}_{Y_{0t}(0)|D(0)=0,X}(q) = h_0\left(F^{-1}_{U_0|G=0,D(0)=0,X}(q), t\right).$$

This implies that almost surely,

$$F^{-1}_{Y_{00}(0)|D(0)=0,X} \circ F_{Y_{01}(0)|D(0)=0,X}(y) = h_0(h_0^{-1}(y, 1), 0) \quad \forall y \in \mathcal{S}(Y_{01}(0)|D(0) = 0, X). \tag{3}$$

Besides, Equation (9) below implies that for all $y \in \mathcal{S}(Y_{11}(0)|V < v_0(0))$,

$$F^{-1}_{Y_{10}(0)|V<v_0(0)} \circ F_{Y_{11}(0)|V<v_0(0)}(y) = h_0(h_0^{-1}(y, 1), 0) \quad \forall y \in \mathcal{S}(Y_{11}(0)|V < v_0(0)). \tag{4}$$

Now, almost surely,

$$\begin{aligned} \mathcal{S}(Y_{11}(0)|V < v_0(0)) &= \mathcal{S}(h_0(U_0, 1)|G = 1, T = 1, V < v_0(0)) \\ &= \mathcal{S}(h_0(U_0, 1)|G = 1, T = 0, V < v_0(0)) \\ &= \mathcal{S}(h_0(U_0, 1)|G = 1, T = 0, D(0) = 0) \\ &= \mathcal{S}(h_0(U_0, 1)|G = 0, T = 0, D(0) = 0, X) \\ &= \mathcal{S}(h_0(U_0, 1)|G = 0, T = 1, D(0) = 0, X) \\ &= \mathcal{S}(Y_{01}(0)|D(0) = 0, X), \end{aligned} \tag{5}$$

where the first equality holds by A.1 and A.2, the second by A.3, the third by A.2, the fourth by A.7, the fifth by A.6 and the last by A.1.

Combining (3), (4) and (5), we obtain, for all $y \in \mathcal{S}(Y_{11}(0)|V < v_0(0))$,

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

This rewrites as

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

Integrating this last equation over the distribution of X conditional on $D_{00} = 0$, we get

$$F_{Y_{00}|D=0} \circ F^{-1}_{Y_{10}|D=0} \circ F_{Y_{11}(0)|V<v_0(0)}(y) = E\left(F_{Y_{01}|D=0,X}(y)|D_{00} = 0\right).$$

Composing again by $F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0}$, we obtain

$$F_{Y_{11}(0)|V<v_0(0)}(y) = F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0} \circ E\left(F_{Y_{01}|D=0,X}(y)|D_{00} = 0\right).$$

The result follows by plugging this expression into Equation (7) below.

Lemma 2.1

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

Let us first note that under Assumptions A.1 and A.3, the first point of A.4 implies

$$\mathcal{S}(Y_{11}(0)|V < v_0(0)) = \mathcal{S}(Y_{01}(0)|V < v_0(0)). \tag{6}$$

Indeed,

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

where the third and fourth implications are obtained combining Assumptions A.1 and A.3.

Then, we show that

$$F_{Y_{11}(0)|C}(y) = \frac{P(D_{10} = 0)F_{Y_{11}(0)|V<v_0(0)}(y) - P(D_{11} = 0)F_{Y_{11}|D=0}(y)}{P(D_{10} = 0) - P(D_{11} = 0)}. \tag{7}$$

First, note 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 A.3., and $P(D_{10}=0)>0$ because of A.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 (7), and thus the second point of the lemma.

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

$$F_{Y_{11}(0)|V<v_0(0)}=F_{Y_{10}|D=0}\circ F^{-1}_{Y_{00}|D=0}\circ F_{Y_{01}(0)|V<v_0(0)}. \tag{8}$$

By A.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 $(g,t)\in\{0,1\}^2$,

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

The second point of A.4 combined with Assumptions A.1 and A.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 $y\in\mathcal{S}(Y_{g1}(0)|V<v_0(0))$,

$$F^{-1}_{Y_{g0}(0)|V<v_0(0)}\circ F_{Y_{g1}(0)|V<v_0(0)}(y)=h_0(h_0^{-1}(y,1),0), \tag{9}$$

which is independent of g. As shown in (6), $\mathcal{S}(Y_{11}(0)|V < v_0(0)) = \mathcal{S}(Y_{01}(0)|V < v_0(0))$ under Assumptions A.1-A.4. Therefore, we have that 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 (8), because $V < v_0(0)$ is equivalent to $D = 0$ when $T = 0$, and because the second point of A.4 implies that $F^{-1}_{Y_{10}|D=0}$ is strictly increasing on $(0, 1)$.

Now, we show that

$$F_{Y_{01}(0)|V<v_0(0)}(y) = \lambda_0 F_{001}(y) + (1 - \lambda_0) F_{Y_{01}(0)|TC}(y). \tag{10}$$

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

$$\begin{aligned} F_{Y_{01}(0)|V<v_0(0)}(y) &= P(V < v_0(1)|G = 0, V < v_0(0)) F_{Y_{01}(0)|V<v_0(1)}(y) \\ &\quad + P(V \in [v_0(1), v_0(0))|G = 0, V < v_0(0)) F_{Y_{01}|V\in[v_0(1),v_0(0))}(y) \\ &= \lambda_0 F_{001}(y) + (1 - \lambda_0) F_{Y_{01}(0)|TC}(y). \end{aligned}$$

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

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

so that we also get (10).

Finally, the first point of the lemma follows by combining (7), (8) and (10).

Theorem 2.3

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

1. Construction of the bounds.

We only establish the bounds for $d = 0$, the reasoning being similar for $d = 1$. We start considering the case where $\lambda_0 < 1$. We first show that in such instances, $0 \leq T_d, G_d(T_d), C_d(T_d) \leq 1$ implies

$$\underline{T_d} \leq T_d \leq \overline{T_d}. \tag{11}$$

Indeed, $G_d(T_d)$ is included between 0 and 1 if and only if

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

while $C_d(T_d)$ is included between 0 and 1 if and only if

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

Since

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

and

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

T_d, $G_d(T_d)$ and $C_d(T_d)$ are all included between 0 and 1 if and only if

$$M_0\left(\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0}) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}\right) \leq T_d \leq m_1\left(\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0)) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}\right). \tag{12}$$

$M_0(.)$ and $m_1(.)$ are increasing. Therefore, we can compose each term of this inequality by $M_0(.)$ and then by $m_1(.)$ while keeping the ordering unchanged. Because $M_0(T_d) = m_1(T_d) = T_d$ and $M_0 \circ m_1 = m_1 \circ M_0$, we obtain (11).

Moreover, when $\lambda_0 < 1$, $G_d(T_d)$ is increasing in T_d, so that $C_d(T_d)$ as well is increasing in T_d. Combining this with (11) implies that for every y',

$$C_d(\underline{T_d})(y') \leq C_d(T_d)(y') \leq C_d(\overline{T_d})(y'). \tag{13}$$

Because $C_d(T_d)(y)$ is a cdf,

$$C_d(T_d)(y) = \inf_{y' \geq y} C_d(T_d)(y') \leq \inf_{y' \geq y} C_d(\overline{T_d})(y').$$

The lower bound follows similarly.

Let us now turn to the case where $\lambda_0 > 1$. Using the same reasoning as above, we get

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

The inequalities in the first line are not binding since they are implied by those on the second line. Thus, we also get (12). Hence, using the same argument as previously,

$$\overline{T_d} \leq T_d \leq \underline{T_d}. \tag{14}$$

Besides, when $\lambda_0 > 1$, $G_d(T_d)$ is decreasing in T_d, so that $C_d(T_d)$ as well is decreasing in T_d. Combining this with (14) implies that for every y, (13) holds as well. This proves the result since $C_d(T_d)(y)$ is a cdf and should be increasing.

2. Sharpness.

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 that we do not have to consider converging sequences of cdf as we do in the case b below .

a. $\lambda_0 > 1$**.** We show that if A.4 holds and $\underline{T}_0$, $G_0(\underline{T}_0)$ and $C_0(\underline{T}_0)$ are increasing, then $\underline{B}_0$ is sharp. On that purpose, we construct $(\widetilde{U}_0, \widetilde{V}, \widetilde{h}_0(.,.))$, possibly different from the true ones, rationalizing the data, for which $F_{Y_{11}(0)|C}$ is equal to the lower bound and which satisfy A.1-A.3.

By Lemma 5.1 below, to show that $\underline{B}_0$ is sharp, it suffices to check that

1. $\left(\widetilde{U}_0, \widetilde{h}_0(.,.)\right)$ rationalize the data.

2. $T_0 = \underline{T}_0$.

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

4. $\widetilde{U}_0$ verifies

$$\widetilde{U}_0 \perp\!\!\!\perp T \,|\, G = 0, V < v_t(z), \forall (t,z) \in \{(0,0),(1,0)\}, \tag{15}$$

$$\widetilde{U}_0 \perp\!\!\!\perp T \,|\, G = 1, V < v_t(z), \forall (t,z) \in \{(0,0),(1,1)\}. \tag{16}$$

The ideas behind Lemma 5.1 are as follows. When $C_0(\underline{T}_0)$ is increasing, $\underline{B}_0$ is merely equal to $C_0(\underline{T}_0)$. Therefore, $\underline{B}_0$ is attained if it is possible to construct a DGP such that $T_0 = \underline{T}_0$. Consequently, Points 1 and 2 ensure that the DGP rationalizing $\underline{B}_0$ is consistent with the data. Point 3 corresponds to A.1. Attaining $\underline{B}_0$ involves $h_0(u,t)$ and U_0 but not V. Therefore, it is not necessary to check A.2. For the same reason, we do not have to check $V \perp\!\!\!\perp T|G$. Moreover, we only observe D rather than V, so that Point 4 is sufficient to get A.3.

We first consider untreated observations in the control group in period 1. We set $\widetilde{h}_0(.,1) = F^{-1}_{Y_{01}|D=0}$. $\widetilde{h}_0(.,1)$ is strictly increasing on $[0,1]$ by A.4. As we will see below, combining this choice with the constraints imposed by Points 1-4 above determines $\widetilde{h}_0(.,0)$ and all conditional distributions of $\widetilde{U}_0$. First, to attain $\underline{B}_0$ we must have that $T_0 = \underline{T}_0$. We therefore set

$$F_{\widetilde{U}_0|G=0,T=1,TC} = \underline{T}_0 \circ F^{-1}_{Y_{01}|D=0},$$

which is a valid cdf on $[0,1]$ since (i) $\underline{T}_0$ is increasing by A.8 and $F^{-1}_{Y_{01}|D=0}$ is also increasing, (ii) $\lim_{y\to\underline{y}} \underline{T}_0(y) = 0$ and $\lim_{y\to\overline{y}} \underline{T}_0(y) = 1$ when $\lambda_0 > 1$.

Moreover, given $\widetilde{h}_0(.,1)$, we must have $F_{\widetilde{U}_0|G=0,T=1,V<v_0(1)} = I$ to rationalize the data. Reasoning as for the proof of (10), one can show that

$$F_{\widetilde{U}_0|G=0,T=1,V<v_0(0)} = \lambda_0 F_{\widetilde{U}_0|G=0,T=1,V<v_1(0)} + (1-\lambda_0) F_{\widetilde{U}_0|G=0,T=1,TC}.$$

This imposes to set

$$F_{\widetilde{U}_0|G=0,T=1,V<v_0(0)} = \lambda_0 I + (1-\lambda_0)\underline{T}_0 \circ F^{-1}_{Y_{01}|D=0} = G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}.$$

As previously, this is a valid cdf since (i) $G_0(\underline{T}_0)$ is increasing by A.8 and $F^{-1}_{Y_{01}|D=0}$ is increasing and (ii) $\lim_{y\to\underline{y}} G_0(\underline{T}_0)(y) = 0$ and $\lim_{y\to\overline{y}} G_0(\underline{T}_0)(y) = 1$, as shown in the first point of Lemma 5.2.

Now, let us consider untreated observations in the control group in period 0. For (15) to hold, we define $F_{\widetilde{U}_0|G=0,T=0,V<v_0(1)} = I$ and $F_{\widetilde{U}_0|G=0,T=0,V<v_0(0)} = G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}$. This implies that we must set

$$\widetilde{h}_0(.,0) = F^{-1}_{Y_{00}|D=0} \circ G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}$$

to rationalize the data. As shown in the first point of Lemma 5.2, $G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}$ is strictly increasing on $[0,1]$ and included between 0 and 1. $F^{-1}_{Y_{00}|D=0}$ is also strictly increasing on $[0,1]$ by A.4. Therefore, $\widetilde{h}_0(.,0)$ is strictly increasing on $[0,1]$. As a result, Point 3 holds.

Now, let us consider untreated observations in the treatment group. To ensure compatibility between the data and $\widetilde{h}_0(.,0)$, we set

$$\begin{aligned} F_{\widetilde{U}_0|G=1,T=0,V<v_0(0)} &= F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0} \circ G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}, \\ F_{\widetilde{U}_0|G=1,T=1,V<v_1(1)} &= F_{Y_{11}|D=0} \circ F^{-1}_{Y_{01}|D=0}. \end{aligned}$$

Both are valid cdf as compositions of increasing functions and a cdf. Using the fact that

$$F_{\widetilde{U}_0|G=1,T=t,V<v_0(0)} = \mu_0 F_{\widetilde{U}_0|G=1,T=t,V<v_1(1)} + (1-\mu_0) F_{\widetilde{U}_0|G=1,T=t,C},$$

one can see that it is sufficient to set, for $t \in \{0,1\}$,

$$\begin{aligned} F_{\widetilde{U}_0|G=1,T=t,C} &= \frac{P(D_{10}=0)F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0} \circ G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0} - P(D_{11}=0)F_{Y_{11}|D=0} \circ F^{-1}_{Y_{01}|D=0}}{P(D_{10}=0) - P(D_{11}=0)} \\ &= C_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0} \end{aligned}$$

to have that (16) holds. $F_{\widetilde{U}_0|G=1,T=t,C}$ is a valid cdf since (i) $C_0(\underline{T}_0)$ is increasing by A.8 and $F^{-1}_{Y_{01}|D=0}$ is increasing by A.4 and (ii) $\lim_{y\to\underline{y}} C_0(\underline{T}_0)(y) = 0$ and $\lim_{y\to\overline{y}} C_0(\underline{T}_0)(y) = 1$ as shown in the second point of Lemma 5.2. As a result, Points 1-4 are all satisfied.

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 $T_0 = \underline{T}_0$. On the other hand, by Lemma 5.3 below, there exists a sequence $(\underline{T}_0^n)_n$ of cdf such that $\underline{T}_0^n \to \underline{T}_0$, $G_0(\underline{T}_0^n)$ is an increasing bijection from $\mathcal{S}(Y)$ to $[0,1]$ and $C_0(\underline{T}_0^n)$ is increasing and onto $[0,1]$. We thus define a sequence of DGP $(\widetilde{U}_0^n, \widetilde{h}_0^n(.,t))$ such that Points 1 to 4 listed above hold for every n, and such that $T_0^n = \underline{T}_0^n$. Since $\underline{T}_0^n(y)$ converges to $\underline{T}_0(y)$ for every y in $\overset{\circ}{\mathcal{S}(Y)}$, we thus define a sequence of DGP such that T_0^n is arbitrarily close from $\underline{T}_0$ on $\overset{\circ}{\mathcal{S}(Y)}$ for sufficiently large n. Since $C_0(.)$ is continuous, this proves that $\underline{B}_0$ is sharp on $\overset{\circ}{\mathcal{S}(Y)}$.

Let us first consider untreated observations in the control group in period 1. Let

$$\widetilde{h}_0^n(.,1) = \left(\lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0)\underline{T}_0^n\right)^{-1} = G_0(\underline{T}_0^n)^{-1}.$$

$\widetilde{h}_0^n(.,1)$ is strictly increasing on $[0,1]$ since $G_0(\underline{T}_0^n)$ is an increasing bijection on $[0,1]$. To have that $T_0^n = \underline{T}_0^n$, we must set

$$F_{\widetilde{U}_0^n|G=0,T=1,TC} = \underline{T}_0^n \circ G_0(\underline{T}_0^n)^{-1},$$

which is possible since $\underline{T}_0^n$ is a cdf and $G_0(\underline{T}_0^n)^{-1}$ is increasing and onto $\mathcal{S}(Y)$. Moreover, given $\widetilde{h}_0^n(.,1)$, we must set

$$F_{\widetilde{U}_0^n|G=0,T=1,V<v_0(1)} = F_{Y_{01}|D=0} \circ G_0(\underline{T}_0^n)^{-1}$$

to rationalize the data. Here as well this yields a valid cdf since $F_{Y_{01}|D=0}$ is a cdf. Since

$$F_{\widetilde{U}_0^n|G=0,T=1,V<v_0(0)} = \lambda_0 F_{\widetilde{U}_0^n|G=0,T=1,V<v_0(1)} + (1-\lambda_0)F_{\widetilde{U}_0^n|G=0,T=1,TC},$$

we set $F_{\widetilde{U}_0^n|G=0,T=1,V<v_0(0)} = I$.

Now, let us consider untreated observations in the control group in period 0. For (15) to hold, we must have that $F_{\widetilde{U}_0^n|G=0,T=0,V<v_0(0)} = I$ and $F_{\widetilde{U}_0^n|G=0,T=0,V<v_0(1)} = F_{Y_{01}|D=0} \circ G_0(\underline{T}_0^n)^{-1}$. This implies that we must set $\widetilde{h}_0^n(.,0) = F^{-1}_{Y_{00}|D=0}$ to rationalize the data. $\widetilde{h}_0^n(.,0)$ is strictly increasing on $[0,1]$ under A.4, therefore, Point 3 is verified.

Finally, let us consider untreated observations in the treatment group. From the definition of $\widetilde{h}_0^n(.,0)$ and $\widetilde{h}_0^n(.,1)$, and some algebra, we obtain that

$$F_{\widetilde{U}_0^n|G=1,T=0,V<v_0(0)} = F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0},$$
$$F_{\widetilde{U}_0^n|G=1,T=1,V<v_1(1)} = F_{Y_{11}|D=0} \circ G_0(\underline{T}_0^n)^{-1},$$

both being proper cdf. Using the fact that

$$F_{\widetilde{U}_0^n|G=1,T=t,V<v_0(0)} = \mu_0 F_{\widetilde{U}_0^n|G=1,T=t,V<v_1(1)} + (1-\mu_0)F_{\widetilde{U}_0^n|G=1,T=t,C},$$

one can see that it is sufficient to set, for $t \in \{0,1\}$,

$$\begin{aligned} F_{\widetilde{U}_0^n|G=1,T=t,C} &= \frac{P(D_{10}=0)F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0} - P(D_{11}=0)F_{Y_{11}|D=0} \circ G_0(\underline{T}_0^n)^{-1}}{P(D_{10}=0)-P(D_{11}=0)} \\ &= C_0(\underline{T}_0^n) \circ G_0(\underline{T}_0^n)^{-1} \end{aligned}$$

to have that (16) holds. That $F_{\widetilde{U}_0^n|G=1,T=t,C}$ is a cdf follows by Points 2 and 3 of Lemma 5.3. This completes the proof when $\lambda_0 < 1$.

Theorem 2.5

We prove point 1 for $d=0$ only. Assume first that there exists $y_0 < y_1$ in $\mathcal{S}(Y)^2$ such that $I_d(y_0,y_1)=\emptyset$. Assume also that $\mathcal{T}_0$ is not empty. Then there exists a function T_0 increasing and included between 0 and 1 such that $G_0(T_0)$ and $C_0(T_0)$ are also increasing and included between 0 and 1. As shown in (11), when $\lambda_0 \leq 1$, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ implies that we must have

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

Conversely, as shown in (14), when $\lambda_0 > 1$, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ implies

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

Therefore, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ always implies

$$\min(\overline{T_0}, \underline{T_0}) \leq T_0 \leq \max(\overline{T_0}, \underline{T_0}). \tag{17}$$

Moreover, T_0 increasing implies

$$T_0(y_0) \leq T_0(y_1). \tag{18}$$

Combining Equations (17) and (18) implies that we must have

$$\min(\overline{T_0}(y_0), \underline{T_0}(y_0) \leq T_0(y_0)) \leq T_0(y_1) \leq \max(\overline{T_0}(y_1), \underline{T_0}(y_1)). \tag{19}$$

This contradicts the fact that $I_d(y_0,y_1)=\emptyset$, which proves a).

Now assume that there exists $y_0 < y_1$ in $\mathcal{S}(Y)^2$ such that $I_d(y_0,y_1) \neq \emptyset$ and for every $t_0 \leq t_1$ in $I_0(y_0,y_1)^2$,

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)t_1) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0)-P(D_{11}=0)} \\ < \; & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)t_0) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0)-P(D_{11}=0)}. \end{aligned} \tag{20}$$

Assume also that $\mathcal{T}_0$ is not empty. $C_0(T_0)$ increasing implies

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)T_0(y_1)) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0)-P(D_{11}=0)} \\ \geq \; & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)T_0(y_0)) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0)-P(D_{11}=0)}. \end{aligned} \tag{21}$$

As shown above in Equation (19), the fact that we must have $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ and T_0 increasing implies that we must have $T_0(y_0) \leq T_0(y_1)$ and $(T_0(y_0), T_0(y_1)) \in I_0(y_0, y_1)^2$, which is not empty by assumption. Combining this with Equation (21) proves that there exists $t_0 = T_0(y_0) \leq t_1 = T_0(y_1)$ in $I_0(y_0, y_1)^2$ such that

$$\begin{array}{cc} & \dfrac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)t_1) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0) - P(D_{11}=0)} \\ \geq & \dfrac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)t_0) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0) - P(D_{11}=0)}. \end{array}$$

This contradicts (20) and proves b). The proof of point 2 is straightforward.

Corollary 2.4

The bounds on Δ and τ_q are a direct consequence of Theorem 2.3. Note that the bounds on the LATE are well defined under A.9. We now prove that these bounds are sharp under A.8. We only focus on the lower bound, the result being similar for the upper bound. First remark that because the model and data impose no condition on the joint distribution of (U_0, U_1). Hence, by the proof Theorem 2.3, we can rationalize the fact that $(F_{Y_{11}(0)|C}, F_{Y_{11}(1)|C}) = (\underline{B}_0, \overline{B}_1)$ when $\lambda_0 > 1$. The results on the sharpness of Δ and τ_q follows directly. When $\lambda_0 < 1$, on the other hand, we only rationalize the fact that $(F_{Y_{11}(0)|C}, F_{Y_{11}(1)|C}) = (G_{0n}, F_{Y_{11}(1)|C})$, where G_{0n} converges pointwise to $\underline{B}_0$. To show the sharpness of the LATE and QTE, we thus have to prove that $\lim_{n\to\infty} \int y dG_{0n}(y) = \int y d\underline{B}_0(y)$ and $\lim_{n\to\infty} G_{0n}^{-1}(q) = \underline{B}_0^{-1}(q)$. As for the first point, we have, by integration by parts for Lebesgue-Stieljes integrals,

$$\begin{aligned} \int y dG_{0n}(y) &= \overline{y} - \int_{\underline{y}}^{\overline{y}} G_{0n} dy \\ &= -\int_{\underline{y}}^{0} G_{0n}(y) dy + \int_0^{\overline{y}} \left[1 - G_{0n}(y)\right] dy. \end{aligned} \tag{22}$$

We now prove convergence of each integral. As shown by Lemma 5.3, G_{0n} can be defined as $G_{0n} = C_0(\underline{T}_0^n)$ with $\underline{T}_0^n \leq T_0$, T_0 denoting the true cdf associated to $Y_{11}(0)$ on compliers, i.e. $C_0(T_0) = F_{Y_{11}(0)|C}$. Because $C_0(.)$ is increasing when $\lambda_0 < 1$, $G_{0n} \leq F_{Y_{11}(0)|C}$. $E(|Y_{11}(0)|\ |C) < +\infty$ implies that $\int_{\underline{y}}^0 F_{Y_{11}(0)|C}(y) dy < +\infty$. Thus, by the dominated convergence theorem,

$$\lim_{n\to\infty} \int_{\underline{y}}^0 G_{0n} dy = \int_{\underline{y}}^0 \underline{B}_0(y) dy.$$

Now consider the second integral in (22). If $\overline{y} < +\infty$, first, we can also apply the dominated convergence theorem, remarking that $1 - G_{0n} \leq 1$, to yield $\int_0^{\overline{y}} \left[1 - G_{0n}(y)\right] dy \to \int_0^{\overline{y}} \left[1 - \underline{B}_0(y)\right] dy$. Finally, if $\overline{y} = +\infty$, $\lim_{y\to+\infty} \underline{B}_0(y) = \ell < 1$ so that

$$\int_0^{\overline{y}} \left[1 - \underline{B}_0(y)\right] dy = +\infty.$$

By Fatou's lemma,

$$\liminf \int_0^{\overline{y}} [1 - G_{0n}(y)]\, dy \geq \int_0^{\overline{y}} [1 - \underline{B}_0(y)]\, dy,$$

yielding convergence of the second integral in (22). Hence, the lower bound of Δ is sharp.

Now, let us turn to τ_q. Let y_0 be the unique solution to $\underline{B}_0(y) = q$. Fix $\varepsilon > 0$. By pointwise convergence, A.8 and unicity of the solution of $\underline{B}_0(y) = q$,

$$\lim_{n\to\infty} G_{0n}(y_0 + \varepsilon) = \underline{B}_0(y_0 + \varepsilon) > \underline{B}_0(y_0) = q.$$

Thus, there exists n_0 such that for all $n \geq n_0$, $G_{0n}(y_0 + \varepsilon) > q$. As a result, by definition of G_{0n}^{-1}, $y_0 + \varepsilon > G_{0n}^{-1}(q)$ for all $n \geq n_0$. Similarly, there exists n_1 such that for all $n \geq n_1$, $y_0 - \varepsilon < G_{0n}^{-1}(q)$. Hence, for all $n \geq \max(n_0, n_1)$, $|G_{0n}^{-1}(q) - \overline{B}_0^{-1}(q)| < \varepsilon$. The result follows.

Theorem 3.1

We first show that $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C})$ tends to a continuous gaussian process. By Lemma 5.4, $\widetilde{\theta} = (\widehat{F}_{000}, \widehat{F}_{001}, ..., \widehat{F}_{111}, \widehat{\mu}_0, \widehat{\mu}_1)$ converges to a continuous gaussian process. Let

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

so that $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C}) = \left(T_1 \circ \pi_0(\widetilde{\theta}), T_1 \circ \pi_1(\widetilde{\theta})\right)$, where T_1 is defined as in Lemma 5.5. 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 A.10 and $\mu_d > 0$, T_1 is also Hadamard differentiable at $(F_{d10}, F_{d00}, F_{d01}, F_{d11}, \mu_d)$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}$ (where $\mathcal{C}^0$ denotes the set of continuous functions on $\mathcal{S}(Y)$). 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 A.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 A.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 5.4, 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 3.2

As in the previous proof, the idea is to prove, using Lemma 5.5, that the bounds are Hadamard differentiable functions of $\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_0, \mu_0, \lambda_1, \mu_1)$. The theorem then follows from Lemma 5.4 and the functional delta method. The proof is however complicated by the fact that even if the primitive cdf are smooth, the bounds $\underline{B}_d$ and $\overline{B}_d$ may admit kinks, so that Hadamard differentiability is not trivial to derive. The proof is also lengthy as $\underline{B}_d$ and $\overline{B}_d$ take different forms depending on $d \in \{0,1\}$ and whether $\lambda_0 < 1$ or $\lambda_0 > 1$. Note that although both $\underline{B}_d$ and $\overline{B}_{1-d}$ appear in the bounds of Δ or τ_q, we can consider them separately because Hadamard differentiability of $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \overline{B}_d(y)dy$ and $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \underline{B}_d(y)dy$ for $d \in \{0,1\}$ imply (since, as in the previous proof, $\underline{\Delta} = \int_{\mathcal{S}(Y)} \underline{B}_0(y) - \overline{B}_1(y)dy$ and similarly for $\overline{\Delta}$) Hadamard differentiability of $\theta \mapsto (\overline{\Delta}, \underline{\Delta})$. The same reasoning applies for the bounds of the QTE. Note also that by A.8, $\underline{B}_d = C_d(\underline{T}_d)$.

1. Lower bound $\underline{B}_d$

For $d \in \{0,1\}$, let $U_d = \frac{\lambda_d F_{d01} - H_d^{-1}(m_1(\mu_d F_{d11}))}{\lambda_d - 1}$, so that

$$\begin{aligned}\underline{T}_d &= M_0\left(m_1\left(U_d\right)\right),\\ C_d(\underline{T}_d) &= \frac{\mu_d F_{d11} - H_d\left(\lambda_d F_{d01} + (1-\lambda_d)\underline{T}_d\right)}{\mu_d - 1}.\end{aligned}$$

Also, let

$$y^u_{0d} = \inf\{y : U_d(y) > 0\} \text{ and } y^u_{1d} = \inf\{y : U_d(y) > 1\}.$$

When y^u_{0d} and y^u_{1d} are in $\mathbb{R}$, we have, by continuity of U_d, $U_d(y^u_{0d}) = 0$ and $U_d(y^u_{1d}) = 1$. Consequently, $\underline{T}_d(y^u_{0d}) = U_d(y^u_{0d})$ and $\underline{T}_d(y^u_{1d}) = U_d(y^u_{1d})$.

Case 1: $\lambda_0 < 1$ and $d = 0$.

In this case, $U_0 = \frac{H_0^{-1}(\mu_0 F_{011}) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}$. We first prove by contradiction that $y^u_{00} = +\infty$. First, because $\lim_{y\to+\infty} U_0(y) < 1$, we have

$$\lim_{y\to+\infty} \underline{T}_0(y) = M_0(\lim_{y\to+\infty} U_0(y)) < 1.$$

Thus, by A.8, $\underline{T}_0(y) < 1$ for all y, implying $U_0(y) = \underline{T}_0(y) < 1$ for all y. Hence, $y^u_{10} = +\infty$.

Therefore, when $y_{00}^u < +\infty$, there exists y such that $0 < U_0(y) < 1$. Assume that there exists $y' \geq y$ such that $U_0(y') < 0$. By continuity and the intermediate value theorem, this would imply that there exists $y'' \in (y, y')$ such that $U_0(y'') = 0$. But since both $U_0(y)$ and $U_0(y'')$ are included between 0 and 1, this would imply that $\underline{T}_0$ is strictly decreasing between y and y'', which is not possible under A.8. This proves that when $y_{00}^u < +\infty$, there exists y such that for every $y' \geq y$, $0 \leq U_0(y') < 1$.

Consequently, $\underline{T}_0 = U_0$ for every $y' \geq y$. This in turn implies that $C_0(\underline{T}_0) = 0$ for every $y' \geq y$. Moreover, $C_0(\underline{T}_0)$ is increasing under A.8, which implies that $C_0(\underline{T}_0) = 0$ for every y. As a result, this proves that when $y_{00}^u < +\infty$, $C_0(\underline{T}_0) = 0$. This implies that $\mathcal{S}_0$ is empty, which violates A.8. Therefore, under A.8, we cannot have $y_{00}^u < +\infty$ when $\lambda_0 < 1$.

Because $y_{00}^u = +\infty$, $\underline{T}_0 = 0$, so that

$$C_0(\underline{T}_0)(y) \quad = \quad \frac{\mu_0 F_{011}(y) - H_0\left(\lambda_0 F_{001}(y)\right)}{\mu_0 - 1}.$$

Remark that by A.12, Lemma 5.5, points 1 and 2 and the chain rule, $\theta \mapsto \int_{\mathcal{S}(Y)} C_0(\underline{T}_0)(y)dy$ is Hadamard differentiable. Moreover, by A.12 and A.8, $C_0(\underline{T}_0)$ is also increasing and differentiable with strictly positive derivative on $\mathcal{S}(Y)$. Thus, by the chain rule and Hadamard differentiability of $F \mapsto F^{-1}(q)$ at $F = C_0(\underline{T}_0)$ (see, e.g., van der Vaart, 2000, Lemma 21.2) and for $q \in \mathcal{Q}$, $\theta \mapsto B_0(\underline{T}_0)^{-1}(q)$ is Hadamard differentiable.

Case 2: $\lambda_0 > 1$ and $d = 0$.

In this case,

$$U_0 \quad = \quad \frac{\lambda_0 F_{001} - H_0^{-1}(\mu_0 F_{011})}{\lambda_0 - 1}.$$

Therefore, $\lim_{y \to \underline{y}} U_0(y) = 0$, and $\lim_{y \to \overline{y}} U_0(y) > 1$. As a result, $-\infty < y_{10}^u < +\infty$, and $\underline{T}_0(y_{10}^u) = U_0(y_{10}^u) = 1$. This in turn implies $C_0(\underline{T}_0)(y_{10}^u) = 0$. Combining this with Assumption A.8 implies that $C_0(\underline{T}_0)(y) = 0$ for every $y \leq y_{10}^u$. Moreover, A.8 also implies that $\underline{T}_d(y) = 1$ for every $y \geq y_{10}^u$. Therefore,

$$C_0(\underline{T}_0)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y_{10}^u \\ \frac{\mu_0 F_{011}(y) - H_0(\lambda_0 F_{001}(y) + (1-\lambda_0))}{\mu_0 - 1} & \text{if } y > y_{10}^u \end{array} \right.$$

Thus, $C_0(\underline{T}_0)(y) = M_0(T_2(F_{011}, F_{010}, F_{000}, F_{001}, \lambda_0, \mu_0))$, where T_2 is defined as in Lemma 5.5. Hadamard differentiability of $\int_{\underline{y}}^{\overline{y}} C_0(\underline{T}_0)(y)dy$ thus follows by points 1 and 2 of Lemma 5.5, the chain rule and the fact that by Assumption A.11, $(F_{011}, F_{010}, F_{000}, F_{001}, \lambda_0, \mu_0) \in (\mathcal{C}^1)^4 \times [0, \infty) \times ([0, \infty) \backslash \{1\})$. As for the QTE, note that by point 1 of Lemma 5.5 applied to $[a, b] = [y_{10}^u, \overline{y}]$, $\theta \mapsto C_0(\underline{T}_0)$ is still Hadamard differentiable as a function on $(y_{10}^u, \overline{y})$. By A.12,

$C_0(\underline{T}_0)$ is also strictly increasing and differentiable with positive derivative on $(y_{10}^u, \overline{y})$. Thus, by the chain rule and Hadamard differentiability of $F \mapsto F^{-1}(q)$ at $(C_0(\underline{T}_0), q)$ with $q \in \mathcal{Q}$, $\theta \mapsto C_0(\underline{T}_0)^{-1}(q)$ is Hadamard differentiable.

Case 3: $\lambda_0 < 1$ and $d = 1$.

In this case,

$$U_1 \quad = \quad \frac{\lambda_1 F_{100} - H_1^{-1}(\mu_1 F_{111})}{\lambda_1 - 1}.$$

$\mu_1 > 1$ implies that $\frac{1}{\mu_1} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{1}{\mu_1})$ is in $\overset{\circ}{\mathcal{S}}(Y)$ under A.4.

Case 3.a: $\lambda_0 < 1$, $d = 1$ and $y_{01}^u < y^*$.

We have $U_1(y^*) = \frac{\lambda_1 F_{100}(y^*)-1}{\lambda_1 - 1} < 1$. Assume that $U_1(y^*) < 0$. Since $y_{01}^u < y^*$, this implies that there exists $y < y^*$ such that $0 < U_1(y)$. Since U_1 is continuous, there also exists $y^{'} < y^*$ such that $0 < U_1(y') < 1$. By continuity and the intermediate value theorem, this finally implies that there exists $y^{''}$ such that $y^{'} < y^{''}$ and $U_1(y^{''}) = 0$. This contradicts A.8 since this would imply that $\underline{T}_1$ is decreasing between $y^{'}$ and $y^{''}$. This proves that

$$0 \leq U_1(y^*) < 1.$$

Therefore, $\underline{T}_1(y^*) = U_1(y^*)$, which in turn implies that $C_1(\underline{T}_1)(y^*) = 0$. By A.8, this implies that for every $y \leq y^*$, $C_1(\underline{T}_1)(y) = 0$.

For every y greater than y^*,

$$U_1(y) \quad = \quad \frac{\lambda_1 F_{100}(y) - 1}{\lambda_1 - 1}.$$

$U_1(y) < 1$. Since $U_1(y^*) \geq 0$ and $y \mapsto \frac{\lambda_1 F_{100}(y)-1}{\lambda_1 - 1}$ is increasing, $U_1(y) \geq 0$. Consequently, for $y \geq y^*$, $\underline{T}_1(y) = U_1(y)$.

Finally, we obtain

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y^* \\ \frac{\mu_1 F_{111}(y)-1}{\mu_1 - 1} & \text{if } y > y^* \end{array} \right.$$

The result follows as in case 2 above.

Case 3.b: $\lambda_0 < 1$, $d = 1$ and $y_{01}^u \geq y^*$.

For all $y \geq y^*$, $U_1(y) = \frac{\lambda_1 F_{100}(y)-1}{\lambda_1 - 1}$. This implies that $y_{01}^u = F_{100}^{-1}(1/\lambda_1) < +\infty$ and $U_1(y_{01}^u) = 0$. Because $y \mapsto \frac{\lambda_1 F_{100}(y)-1}{\lambda_1 - 1}$ is increasing, $U_1(y) \geq 0$ for every $y \geq y_{01}^u$. Moreover,

$U_1(y) \leq 1$. Therefore, $\underline{T}_1(y) = U_1(y)$ for every $y \geq y_{01}^u$. Beside, for every y lower than y_{01}^u, $\underline{T}_1(y) = 0$. As a result,

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{101}(y))}{\mu_1 - 1} & \text{if } y \leq y_{01}^u \\ \frac{\mu_1 F_{111}(y) - 1}{\mu_1 - 1} & \text{if } y > y_{01}^u. \end{array} \right.$$

This implies that

$$\int_{\underline{y}}^{\overline{y}} C_1(\underline{T}_1)(y) dy = \frac{1}{\mu_1 - 1} \left[T_3(F_{111}) - T_4(F_{101}, H_1) \right],$$

where T_3 and T_4 are defined in Lemma 5.5. By this lemma and the fact that F_{111} is a cdf, T_3 is Hadamard differentiable at F_{111}. As shown in the proof of Lemma 5.5, $H_1 = F_{110} \circ F_{100}^{-1}$ is a Hadamard differentiable function of (F_{110}, F_{100}). Thus, by Lemma 5.5 and the chain rule, $T_4(F_{101}, H_1)$ is a Hadamard differentiable function of $(F_{101}, F_{110}, F_{100})$. The result follows for $\int_{\underline{y}}^{\overline{y}} C_1(\underline{T}_1)(y) dy$. The expression above also shows that $C_1(\underline{T}_1)$ is Hadamard differentiable as a function of $(F_{100}, F_{101}, F_{110}, F_{111}, \lambda_1, \mu_1)$ when considering the restriction of these functions to intervals such that $C_1(\underline{T}_1)$ is defined on $(\underline{y}, y_{01}^u)$ only. By A.12, $C_1(\underline{T}_1)$ is also a differentiable function with positive derivative on $(\underline{y}, y_{01}^u)$. Therefore, $\theta \mapsto C_1(\underline{T}_1)^{-1}(q)$ is Hadamard differentiable for $q \in (C_1(\underline{T}_1)(\underline{y}), C_1(\underline{T}_1)(y_{01}^u) = (0, q_1)$. The same holds when considering the interval $(y_{01}^u, \overline{y})$ instead of $(\underline{y}, y_{01}^u)$. Hence, $\theta \mapsto C_1(\underline{T}_1)^{-1}(q)$ is Hadamard differentiable for $q \in (0,1) \backslash \{q_1\}$.

Case 4: $\lambda_0 > 1$ and $d = 1$.

In this case,

$$U_1 \quad = \quad \frac{H_1^{-1}(\mu_1 F_{111}) - \lambda_1 F_{100}}{1 - \lambda_1}.$$

Therefore, $\lim_{y \to \underline{y}} U_1(y) = 0$, which implies that $-\infty < y_{11}^u$. As above, $\mu_1 > 1$ implies that y^* is in $\overset{\circ}{\mathcal{S}}(Y)$ under A.4. $U_1(y^*) = \frac{1 - \lambda_1 F_{100}(y^*)}{1 - \lambda_1} > 1$, which implies that $y_{11}^u < +\infty$. Therefore, reasoning as for case 2, we obtain

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y_{11}^u \\ \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{100}(y) + (1 - \lambda_1))}{\mu_1 - 1} & \text{if } y > y_{11}^u. \end{array} \right.$$

The result follows as in case 2 above.

2. Upper bound $\overline{B}_d$.

Let $V_d = \frac{\lambda_d F_{d01} - H_d^{-1}(M_0(\mu_d F_{d11} + (1 - \mu_d)))}{\lambda_d - 1}$, so that

$$\begin{aligned} \overline{T}_d &= M_0 \left(m_1 \left(V_d \right) \right), \\ C_d(\overline{T}_d) &= \frac{\mu_d F_{d11} - H_d \left(\lambda_d F_{d01} + (1 - \lambda_d) \overline{T}_d \right)}{\mu_d - 1}. \end{aligned}$$

Also, let

$$y^v_{0d} = \inf\{y : V_d(y) > 0\}, \quad y^v_{1d} = \inf\{y : V_d(y) > 1\}.$$

Note that when y^v_{0d} and y^v_{1d} are in $\mathbb{R}$, by continuity of V_d we have $V_d(y^v_{0d}) = 0$ and $V_d(y^v_{1d}) = 1$. Consequently, $\overline{T}_d(y^v_{0d}) = V_d(y^v_{0d})$ and $\overline{T}_d(y^v_{1d}) = V_d(y^v_{1d})$.

Case 1: $\lambda_0 < 1$ and $d = 0$

In this case,

$$V_0 \quad = \quad \frac{H_0^{-1}(\mu_0 F_{011} + (1-\mu_0)) - \lambda_0 F_{001}}{1-\lambda_0}.$$

Since $\mu_0 < 1$, $\lim_{y\to\underline{y}} V_0(y) > 0$ and can even be greater than 1.

First, let us prove by contradiction that $y^v_{10} = -\infty$. $V_0(y) \leq 1$ for every $y \leq y^v_{10}$. Using the fact that $\lim_{y\to\underline{y}} V_0(y) > 0$ and that $\overline{T}_0$ must be increasing under A.8, one can also show that $0 \leq V_0(y)$ for every $y \leq y^v_{10}$. This implies that $\overline{T}_0(y) = V_0(y)$ which in turn implies that $C_0(\overline{T}_0)(y) = 1$ for every $y \leq y^v_{10}$. Since $C_0(\overline{T}_0)$ must be increasing under A.8, this implies that for every $y \in \mathcal{S}(Y)$,

$$C_0(\overline{T}_0)(y) \quad = \quad 1.$$

This implies that $\mathcal{S}_0$ is empty, which violates A.8. Therefore, $y^v_{10} = -\infty$.

$y^v_{10} = -\infty$ implies that $\lim_{y\to\underline{y}} \overline{T}_0(y) = 1$. This combined with Assumption A.8 implies that $\overline{T}_0(y) = 1$ for every $y \in \mathcal{S}(Y)$. Therefore,

$$C_0(\overline{T}_0)(y) \quad = \quad \frac{\mu_0 F_{011}(y) - H_0\left(\lambda_0 F_{001}(y) + (1-\lambda_0)\right)}{\mu_0 - 1}.$$

The result follows as in case 1 of the lower bound.

Case 2: $\lambda_0 > 1$ and $d = 0$.

In this case,

$$V_0 \quad = \quad \frac{\lambda_0 F_{001} - H_0^{-1}(\mu_0 F_{011} + (1-\mu_0))}{\lambda_0 - 1}.$$

Since $\mu_0 < 1$, $\lim_{y\to\underline{y}} V_0(y) < 0$. Therefore, $y^v_{00} > -\infty$.

Case 2.a): $\lambda_0 > 1$, $d = 0$ and $y^v_{00} < +\infty$.

If $y^v_{00} \in \mathbb{R}$, $\overline{T}_0(y^v_{00}) = V_0(y^v_{00})$ which in turn implies that $C_0(\overline{T}_0)(y^v_{00}) = 1$. By A.8, this implies that for every $y \geq y^v_{00}$, $C_0(\overline{T}_0)(y) = 1$. For every $y \leq y^v_{00}$, $\overline{T}_0(y) = 0$, so that

$$C_0(\overline{T}_0) \quad = \quad \frac{\mu_0 F_{011} - H_0\left(\lambda_0 F_{001}\right)}{\mu_0 - 1}.$$

As a result,

$$C_0(\overline{T}_0)(y) = \left| \begin{array}{ll} \frac{\mu_0 F_{011}(y) - H_0(\lambda_0 F_{001}(y))}{\mu_0 - 1} & \text{if } y \leq y_{00}^v \\ 1 & \text{if } y > y_{00}^v. \end{array} \right.$$

The result follows as in case 2 of the lower bound.

Case 2.b): $\lambda_0 > 1$, $d = 0$ and $y_{00}^v = +\infty$

If $y_{00}^v = +\infty$, $\overline{T}_0(y) = 0$ for every $y \in \mathcal{S}(Y)$, so that

$$C_0(\overline{T}_0)(y) = \frac{\mu_0 F_{011}(y) - H_0\left(\lambda_0 F_{001}(y)\right)}{\mu_0 - 1}.$$

The result follows as in case 1 of the lower bound.

Case 3: $\lambda_0 < 1$ and $d = 1$

In this case,

$$V_1 = \frac{\lambda_1 F_{100} - H_1^{-1}(\mu_1 F_{111} - (\mu_1 - 1))}{\lambda_1 - 1}.$$

Therefore, $\lim_{y \to \underline{y}} V_1(y) = 0$, which implies that $-\infty < y_{11}^v$. $\mu_1 > 1$ implies that $\frac{\mu_1 - 1}{\mu_1} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{\mu_1 - 1}{\mu_1})$ is in $\mathring{\mathcal{S}}(Y)$ under A.4.

Case 3.a): $\lambda_0 < 1$, $d = 1$ and $y_{11}^v > y^*$

$$V_1(y^*) = \frac{\lambda_1 F_{100}(y^*)}{\lambda_1 - 1} > 0.$$

If $y^* < y_{11}^v$, $V_1(y^*) < 1$. Therefore, $0 < \overline{T}_1(y^*) = V_1(y^*) < 1$. This implies that $C_1(\overline{T}_1)(y^*) = 1$ which in turn implies that $C_1(\overline{T}_1)(y) = 1$ for every $y \geq y^*$ under A.8.

For every y lower than y^*,

$$V_1(y) = \frac{\lambda_1 F_{100}(y)}{\lambda_1 - 1}.$$

$V_1(y) > 0$. Since by assumption $y_{11}^v > y^*$, $V_1(y) < 1$. Consequently, for $y \leq y^*$, we have $\overline{T}_1(y) = V_1(y)$.

As a result,

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y)}{\mu_1 - 1} & \text{if } y \leq y^* \\ 1 & \text{if } y > y^*. \end{array} \right.$$

The result follows as in case 2 of the lower bound.

Case 3.b): $\lambda_0 < 1$, $d = 1$, and $y_{11}^v \leq y^*$

First, $V_1(y_{11}^v) = 1$, implying $\overline{T}_1(y_{11}^v) = 1$. By A.8, $\overline{T}_1(y) = 1$ for all $y \geq y_{11}^v$. Second, if $y \leq y_{11}^v \leq y^*$, $V_1(y) = \frac{\lambda_1 F_{100}(y)}{\lambda_1 - 1}$. Thus V_1 is increasing on $(-\infty, y_{11}^v)$. Moreover $V_1(y_{11}^v) = 1$.

Hence, $V_1(y) \leq 1$ for every $y \leq y_{11}^v$. Because we also have $V_1(y) \geq 0$, $\overline{T}_1(y) = V_1(y)$ for every $y \leq y_{11}^v$.

As a result,

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y)}{\mu_1 - 1} & \text{if } y \leq y_{11}^v \\ \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{Y_{01|D=1}}(y) + 1 - \lambda_1)}{\mu_1 - 1} & \text{if } y > y_{11}^v. \end{array} \right.$$

The result follows as in case 3.b) of the lower bound. Note that here, $C_1(\overline{T}_1)(y)$ is kinked at y_{11}^v, with $C_1(\overline{T}_1)(y_{11}^v) = q_2$. Hence, we have to exclude this point when Hadamard differentiability of $\theta \mapsto C_1(\overline{T}_1)^{-1}(q)$.

Case 4: $\lambda_0 > 1$ *and* $d = 1$

In this case,

$$V_1 \quad = \quad \frac{H_1^{-1}(\mu_1 F_{111} - (\mu_1 - 1)) - \lambda_1 F_{100}}{1 - \lambda_1}.$$

$\lim_{y \to \overline{y}} V_0(y) = 1$, which implies that $y_{00}^v < +\infty$. As above, $\mu_1 > 1$ implies that $\frac{\mu_1 - 1}{\mu_1} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{\mu_1 - 1}{\mu_1})$ is in $\overset{\circ}{\mathcal{S}}(Y)$ under A.4. $V_1(y^*) = \frac{-\lambda_1 F_{100}(y^*)}{1 - \lambda_1} < 0$. Since $\overline{T}_1$ is increasing under A.8, one can show that this implies that $y_{01}^v > y^*$. Therefore, reasoning as for case 2, we obtain that

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{100}(y))}{\mu_1 - 1} & \text{if } y \leq y_{01}^v \\ 1 & \text{if } y > y_{01}^v. \end{array} \right.$$

The result follows as in case 2 of the lower bound.

Appendix B: technical lemmas

Lemma 5.1 *When A.4 and A.8 hold, to show that $\underline{B}_0$ is sharp, it suffices to check that*

1. $\left(\widetilde{U}_0, \widetilde{h}_0(.,.)\right)$ *rationalize the data.*

2. $T_0 = \underline{T}_0$.

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

4. $\widetilde{U}_0$ *verifies*

$$\begin{aligned}
\widetilde{U}_0 \quad &\perp\!\!\!\perp \quad T \,|\, G = 0, V < v_t(z), \forall (t,z) \in \{(0,0),(1,0)\} \\
\widetilde{U}_0 \quad &\perp\!\!\!\perp \quad T \,|\, G = 1, V < v_t(z), \forall (t,z) \in \{(0,0),(1,1)\}.
\end{aligned}$$

Proof:

When $C_0(\underline{T}_0)$ is increasing, $\underline{B}_0$ is merely equal to $C_0(\underline{T}_0)$. Therefore, $\underline{B}_0$ is attained if it is possible to construct a DGP such that $T_0 = \underline{T}_0$. Consequently, Points 1 and 2 ensure that the DGP rationalizing $\underline{B}_0$ is consistent with the data. Point 3 corresponds to A.1. Attaining $\underline{B}_0$ involves $h_0(u,t)$ and U_0 but not V. Therefore, it is not necessary to check A.2. For the same reason, we do not have to check $V \perp\!\!\!\perp T|G$. What we show now is that if $V \perp\!\!\!\perp T|G$ and Point 4 is verified, then it is possible to construct $\widetilde{V}$ such that A.3 is verified. We prove the result for $G = 0$ only.

On that purpose, we start by proving that if

$$\widetilde{U}_0 \perp\!\!\!\perp T \,|\, G = 0, 1\{V < v_0(0)\}, \tag{23}$$

then it is possible to construct $\widetilde{V}$ such that A.3 holds for $G = 0$. Set $\widetilde{V} = V$ for $T = 0$, and

$$\begin{aligned}
f_{\widetilde{V}|\widetilde{U}_0, G=0, T=1, V<v_0(0)}(v|u) \quad &= \quad \frac{f_{\widetilde{U}_0|V, G=0, T=0, V<v_0(0)}(u|v) f_{V|G=0, T=0, V<v_0(0)}(v)}{f_{\widetilde{U}_0|G=0, T=0, V<v_0(0)}(u)} \\
f_{\widetilde{V}|\widetilde{U}_0, G=0, T=1, V\geq v_0(0)}(v|u) \quad &= \quad \frac{f_{\widetilde{U}_0|V, G=0, T=0, V\geq v_0(0)}(u|v) f_{V|G=0, T=0, V\geq v_0(0)}(v)}{f_{\widetilde{U}_0|G=0, T=0, V\geq v_0(0)}(u)}.
\end{aligned}$$

It is easy to check that this defines valid densities. Let us now check that $\widetilde{V}$ verifies A.3.

$$\begin{aligned}
& f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1}(u,v) \\
= \; & f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1,V<v_0(0)}(u,v)P(V<v_0(0)|G=0,T=1) \\
+ \; & f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1,V\geq v_0(0)}(u,v)P(V\geq v_0(0)|G=0,T=1) \\
= \; & f_{\widetilde{V}|\widetilde{U}_0,G=0,T=1,V<v_0(0)}(v|u)f_{\widetilde{U}_0|G=0,T=1,V<v_0(0)}(u)P(V<v_0(0)|G=0,T=1) \\
+ \; & f_{\widetilde{V}|\widetilde{U}_0,G=0,T=1,V\geq v_0(0)}(v|u)f_{\widetilde{U}_0|G=0,T=1,V\geq v_0(0)}(u)P(V\geq v_0(0)|G=0,T=1) \\
= \; & \frac{f_{\widetilde{U}_0|V,G=0,T=0,V<v_0(0)}(u|v)f_{V|G=0,T=0,V<v_0(0)}(v)}{f_{\widetilde{U}_0|G=0,T=0,V<v_0(0)}(u)}f_{\widetilde{U}_0|G=0,T=1,V<v_0(0)}(u)P(V<v_0(0)|G=0,T=1) \\
+ \; & \frac{f_{\widetilde{U}_0|V,G=0,T=0,V\geq v_0(0)}(u|v)f_{V|G=0,T=0,V\geq v_0(0)}(v)}{f_{\widetilde{U}_0|G=0,T=0,V\geq v_0(0)}(u)}f_{\widetilde{U}_0|G=0,T=1,V\geq v_0(0)}(u)P(V\geq v_0(0)|G=0,T=1) \\
= \; & f_{\widetilde{U}_0|V,G=0,T=0,V<v_0(0)}(u|v)f_{V|G=0,T=0,V<v_0(0)}(v)P(V<v_0(0)|G=0,T=1) \\
+ \; & f_{\widetilde{U}_0|V,G=0,T=0,V\geq v_0(0)}(u|v)f_{V|G=0,T=0,V\geq v_0(0)}(v)P(V\geq v_0(0)|G=0,T=1) \\
= \; & f_{\widetilde{U}_0|V,G=0,T=0,V<v_0(0)}(u|v)f_{V|G=0,T=0,V<v_0(0)}(v)P(V<v_0(0)|G=0,T=0) \\
+ \; & f_{\widetilde{U}_0|V,G=0,T=0,V\geq v_0(0)}(u|v)f_{V|G=0,T=0,V\geq v_0(0)}(v)P(V\geq v_0(0)|G=0,T=0) \\
= \; & f_{\widetilde{U}_0,V|G=0,T=0}(u,v) \\
= \; & f_{\widetilde{U}_0,\widetilde{V}|G=0,T=0}(u,v)
\end{aligned}$$

The first, second and sixth equalities follow from Bayes law. The third arises from the definition of $\widetilde{V}$ in period 1. The fourth follows from point 4). The fifth is obtained using the fact that $V \perp\!\!\!\perp T|G$. The sixth comes from Bayes law. Finally, the seventh comes from the fact that in period 0, $\widetilde{V}=V$.

To conclude the proof, we show that if we can construct $\widetilde{U}_0$ such that (15) is verified, then we can also construct $\widetilde{U}_0$ such that (23) holds. First, note that (15) combined with $V \perp\!\!\!\perp T|G$ implies

$$\begin{aligned}
\widetilde{U}_0 \quad & \perp\!\!\!\perp \quad T\,|\,G=0, V<\min(v_0(0),v_1(0)) \\
\widetilde{U}_0 \quad & \perp\!\!\!\perp \quad T\,|\,G=0, \min(v_0(0),v_1(0))\leq V<\max(v_0(0),v_1(0)).
\end{aligned}$$

We never observe $Y(0)$ for observations in the control group such that $V\geq \max(v_0(0),v_1(0))$ since they are treated both in period 0 and in period 1. Therefore, when constructing $\widetilde{U}_0$, we can merely set $\widetilde{U}_0=U_0$ within that population without violating Point 1. This does not modify the bounds either, so that Point 2 still holds. As a result,

$$\widetilde{U}_0 \quad \perp\!\!\!\perp \quad T\,|\,G=0, V\geq \max(v_0(0),v_1(0)),$$

since U_0 verifies A.3. Combining those three equations yields (23).

Lemma 5.2 *Assume A.4 and A.8 hold, and $\lambda_d>1$. Then:*

1. $G_d(\underline{T}_d)$ *is a bijection from* $\mathcal{S}(Y)$ *to* $[0,1]$*;*

2. $C_d(\underline{T}_d)\left(\mathcal{S}(Y)\right) = [0,1]$.

Proof: we only prove the result for $d = 0$, the reasoning being similar otherwise. Let I denote the identity function on $[0,1]$. One can show that when $\lambda_0 > 1$,

$$G_0(\underline{T}_0) = \min\left(\lambda_0 F_{Y_{01}|D=0}, \max\left(\lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0), H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right)\right)\right). \tag{24}$$

By Assumption A.4, $\mu_0 F_{Y_{11}|D=0}$ is strictly increasing. Moreover, $\mathcal{S}(Y_{10}|D=0) = \mathcal{S}(Y_{00}|D=0)$ implies that H_0^{-1} is strictly increasing on $[0,1]$. Consequently, $H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right) \circ F^{-1}_{Y_{01}|D=0}$ is strictly increasing since $\mu_0 < 1$. Therefore, $G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}$ is strictly increasing on $[0,1]$ as a composition of the max and min of strictly increasing functions. It is also continuous by A.4, as a composition of continuous function. Moreover, it is easy to see that since $\mathcal{S}(Y_{1t}|D=0) = \mathcal{S}(Y_{0t}|D=0)$,

$$\begin{aligned}
\lim_{y \to \underline{y}} H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right) \circ F^{-1}_{Y_{01}|D=0}(y) &= 0, \\
\lim_{y \to \overline{y}} H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right) \circ F^{-1}_{Y_{01}|D=0}(y) &\leq 1.
\end{aligned}$$

Hence, by Equation (24),

$$\lim_{y \to \underline{y}} G_0(\underline{T}_0)(y) = 0, \ \lim_{y \to \overline{y}} G_0(\underline{T}_0)(y) = 1. \tag{25}$$

Point 1 finally follows, by the intermediate value theorem.

Now, we have

$$C_0(\underline{T}_0) = \frac{P(D_{10}=0) F_{Y_{10}|D=0} \circ F^{-1}_{Y_{01}|D=0} \circ G_0(\underline{T}_0) - P(D_{11}=0) F_{Y_{11}|D=0}}{P(D_{10}=0) - P(D_{11}=0)}.$$

(25) implies that $G_0(\underline{T}_0)$ is a cdf. Hence, by A.4

$$\lim_{y \to \underline{y}} C_0(\underline{T}_0)(y) = 0, \ \lim_{y \to \overline{y}} C_0(\underline{T}_0)(y) = 1.$$

Moreover, $C_0(\underline{T}_0)$ is increasing by A.8. Combining this with A.4 yields Point 2, since $C_0(\underline{T}_0)$ is continuous by A.4 once more.

Lemma 5.3 *Assume A.4 and A.8 hold,* $0 < P(D_{g0} = 0)$ *for* $g \in \{0;1\}$ *and* $\lambda_0 < 1$*. Then there exists a sequence of cdf* $\underline{T}_0^n$ *such that*

1. $\underline{T}_0^n(y) \to \underline{T}_d(y)$ *for all* $y \in \overset{\circ}{\mathcal{S}}(Y)$*;*

2. $G_0(\underline{T}_0^n)$ *is an increasing bijection from* $\mathcal{S}(Y)$ *to* $[0,1]$*;*

3. $C_0(\underline{T}_0^n)$ *is increasing and onto* $[0,1]$.

The same holds for the upper bound.

Proof: we consider a sequence $(y_n)_{n\in\mathbb{N}}$ converging to $\overline{y}$ and such that $y_n < \overline{y}$. Since $y_n < \overline{y}$, we can also define a strictly positive sequence $(\eta_n)_{n\in\mathbb{N}}$ such that $y_n + \eta_n < \overline{y}$. By A.8, H_0 is continuously differentiable. Moreover,

$$H_0' = \frac{F'_{Y_{10}|D=0} \circ F'_{Y_{00}|D=0}}{F'_{Y_{00}|D=0} \circ F'_{Y_{00}|D=0}}$$

is strictly positive on $\mathcal{S}(Y)$ under A.8. $F'_{Y_{11}|D=0}$ is also strictly positive on $\mathcal{S}(Y)$ under A.8. Therefore, using Taylor expansion of H_0 and $F_{Y_{11}|D=0}$, it is easy to show that there exists constants $k_n > 0$ and $K_n > 0$ such that for all $y < y' \in [y_n, y_n + \eta_n]^2$,

$$\begin{aligned} H_0(y') - H_0(y) &\geq k_n(y' - y), &(26)\\ F_{Y_{11}|D=0}(y') - F_{Y_{11}|D=0}(y) &\leq K_n(y' - y). &(27)\end{aligned}$$

We also define a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ by

$$\varepsilon_n = \min\left(\eta_n, \frac{k_n(1-\lambda_0)\left(T_0(y_n) - \underline{T}_0(y_n)\right)}{\mu_0 K_n}\right). \quad (28)$$

Note that as shown in (11), since $\lambda_0 < 1$, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ implies that we must have

$$\underline{T_0} \leq T_0,$$

which implies in turn that $\varepsilon_n \geq 0$. Consequently, since $0 \leq \varepsilon_n \leq \eta_n$, inequalities (26) and (27) also hold for $y < y' \in [y_n, y_n + \varepsilon_n]^2$.

We now define $\underline{T}_0^n$. For every n such that $\varepsilon_n > 0$, let

$$\underline{T}_0^n(y) = \left| \begin{array}{ll} \underline{T}_0(y) & \text{if } y < y_n \\ \underline{T}_0(y_n) + \frac{T_0(y_n+\varepsilon_n) - \underline{T}_0(y_n)}{\varepsilon_n}(y - y_n) & \text{if } y \in [y_n, y_n + \varepsilon_n] \\ T_0(y) & \text{if } y > y_n + \varepsilon_n. \end{array} \right.$$

For every n such that $\varepsilon_n = 0$, let

$$\underline{T}_0^n(y) = \left| \begin{array}{ll} \underline{T}_0(y) & \text{if } y < y_n \\ T_0(y) & \text{if } y \geq y_n \end{array} \right.$$

Then, we verify that $\underline{T}_0^n$ defines a sequence of cdf which satisfy Points 1, 2 and 3. Under A.8, $\underline{T}_0(y)$ is increasing, which implies that $\underline{T}_0^n(y)$ is increasing on $(\underline{y}, y_n)$. Since $T_0(y)$ is a cdf, $\underline{T}_0^n(y)$ is also increasing on $(y_n + \varepsilon_n, \overline{y})$. Finally, it is easy to check that when $\varepsilon_n > 0$,

$\underline{T}_0^n(y)$ is increasing on $[y_n, y_n+\varepsilon_n]$. $\underline{T}_0^n$ is continuous on $(\underline{y}, y_n)$ and $(y_n+\varepsilon_n, \overline{y})$ under A.4. It is also continuous at y_n and $y_n+\varepsilon_n$ by construction. This proves that $\underline{T}_0^n(y)$ is increasing on $\mathcal{S}(Y)$. Moreover,

$$\begin{aligned}\lim_{y\to\underline{y}} \underline{T}_0^n(y) &= \lim_{y\to\underline{y}} \underline{T}_0(y) = 0\\ \lim_{y\to\overline{y}} \underline{T}_0^n(y) &= \lim_{y\to\overline{y}} T_0(y) = 1.\end{aligned}$$

Hence, $\underline{T}_0^n$ is a cdf. Point 1 also holds by construction of $\underline{T}_0^n(y)$.

$G_0(\underline{T}_0^n) = \lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0)\underline{T}_0^n$ is strictly increasing and continuous as a sum of the strictly increasing and continuous function $\lambda_0 F_{Y_{01}|D=0}$ and an increasing and continuous function. Moreover, $G_0(\underline{T}_0^n)$ tends to 0 (resp. 1) when y tends to $\underline{y}$ (resp. to $\overline{y}$). Point 2 follows by the intermediate value theorem.

Finally, let us show Point 3. Because $G_0(\underline{T}_0^n)$ is a continuous cdf, $C_0(\underline{T}_0^n)$ is also continuous and converges to 0 (resp. 1) when y tends to $\underline{y}$ (resp. to $\overline{y}$). Thus, the proof will be completed if we show that $C_0(\underline{T}_0^n)$ is increasing. By A.8, $C_0(\underline{T}_0^n)$ is increasing on $(\underline{y}, y_n)$. Moreover, since $F_{Y_{11}(0)|C} = C_0(T_0)$, $C_0(\underline{T}_0^n)$ is also increasing on $(y_n+\varepsilon_n, \overline{y})$. Finally, when $\varepsilon_n > 0$, we have that for all $y < y' \in [y_n, y_n+\varepsilon_n]^2$,

$$\begin{aligned}& H_0(\lambda_0 F_{Y_{01}|D=0}(y') + (1-\lambda_0)\underline{T}_0^n(y')) - H_0(\lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0)\underline{T}_0^n(y))\\ \geq\ & k_n(1-\lambda_0)\left(\underline{T}_0^n(y') - \underline{T}_0^n(y)\right)\\ \geq\ & \frac{k_n(1-\lambda_0)\left(T_0(y_n) - \underline{T}_0(y_n)\right)}{\varepsilon_n}(y'-y)\\ \geq\ & \mu_0 K_n(y'-y)\\ \geq\ & \mu_0\left(F_{Y_{11}|D=0}(y') - F_{Y_{11}|D=0}(y)\right),\end{aligned}$$

where the first inequality follows by (26) and $F_{Y_{01}|D=0}(y') \geq F_{Y_{01}|D=0}(y)$, the second by the definition of $\underline{T}_0^n$ and $T_0(y_n+\varepsilon_n) \geq T_0(y_n)$, the third by (28) and the fourth by (27). This implies that $C_0(\underline{T}_0^n)$ is increasing on $[y_n, y_n+\varepsilon_n]$, since

$$C_0(\underline{T}_0^n) = \frac{H_0(\lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0)\underline{T}_0^n) - \mu_0 F_{Y_{11}|D=0}}{1-\mu_0}.$$

It is easy to check that under A.4 $C_0(\underline{T}_0^n)$ is continuous on $\mathcal{S}(Y)$. This completes the proof.

Lemma 5.4 *Suppose that $P(D_{g0} = d) > 0$ for $(d,g) \in \{0,1\}^2$ and let*

$$\theta_0 = (F_{000}, F_{001}, ..., F_{111}, \lambda_0, \mu_0, \lambda_1, \mu_1)$$

and

$$\widehat{\theta} = (\widehat{F}_{000}, \widehat{F}_{001}, ..., \widehat{F}_{111}, \widehat{\lambda}_0, \widehat{\mu}_0, \widehat{\lambda}_1, \widehat{\mu}_1).$$

Then

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\Longrightarrow G,$$

where G denotes a continuous gaussian process. Moreover, the bootstrap is consistent for $\widehat{\theta}$.

Proof: we prove the result for $\eta=(F_{000},F_{001},...,F_{111},p_{000},...,p_{011})$. The result follows since $(\lambda_0,\mu_0,\lambda_1,\mu_1)$ is a smooth function of $(p_{000},...,p_{011})$. let $\mathbb{G}_n$ denote the standard empirical process, $p_{dgt}=P(D_{gt}=d)$ and let, for any $(y,d,g,t)\in(\mathcal{S}(Y)\cup\{+\infty\})\times\{0,1\}^3$,

$$f_{y,d,g,t}(Y,D,G,T)=\frac{\mathbb{1}\{D=d\}\mathbb{1}\{G=g\}\mathbb{1}\{T=t\}\mathbb{1}\{Y\leq y\}}{p_{dgt}}.$$

We have, for all $(y,d,g,t)\in(\mathcal{S}(Y)\cup\{-\infty,+\infty\})\times\{0,1\}^3$,

$$\begin{aligned}\sqrt{n}\left(\widehat{F}_{dgt}(y)-F_{dgt}(y)\right) &= \frac{\sqrt{n}}{n_{dgt}}\sum_{i=1}^{n}\mathbb{1}\{D_i=d\}\mathbb{1}\{G_i=g\}\mathbb{1}\{T_i=t\}\mathbb{1}\{Y_i\leq y\}-F_{dgt}(y)\\ &= \frac{\sqrt{n}}{n_{dgt}}\sum_{i=1}^{n}\mathbb{1}\{D_i=d\}\mathbb{1}\{G_i=g\}\mathbb{1}\{T_i=t\}\left[\mathbb{1}\{Y_i\leq y\}-F_{dgt}(y)\right].\\ &= \frac{np_{dgt}}{n_{dgt}}\mathbb{G}_n f_{y,d,g,t}.\end{aligned}$$

Note that letting $d=0$ and $y=+\infty$ yields a similar result on p_{0gt}. Now, $\mathcal{F}=\{f_{y,d,g,t}$ $(y,d,g,t)\in(\mathcal{S}(Y)\cup\{+\infty\})\times\{0,1\}^3\}$ is Donsker (see, e.g., van der Vaart, 2000, Example 19.6). Besides, $np_{dgt}/n_{dgt}\xrightarrow{\mathbb{P}}1$. Thus, by Slutski's lemma (see, e.g., van der Vaart, 2000, Theorem 18.10 (v)), $\widehat{\eta}$, the empirical counterpart of η, converges to a gaussian process.

Now let us turn to the bootstrap. Observe that

$$\sqrt{n}\left(\widehat{F}^*_{dgt}(y)-F_{dgt}(y)\right)=\frac{np_{dgt}}{n^*_{dgt}}\mathbb{G}^*_n f_{y,d,g,t},$$

where $\mathbb{G}^*_n$ denote the bootstrap empirical process. Because $np_{dgt}/n^*_{dgt}\xrightarrow{\mathbb{P}}1$ and by consistency of the bootstrap empirical process (see, e.g., van der Vaart, 2000, Theorem 23.7), the bootstrap is consistent for $\widehat{\eta}$.

Lemma 5.5 *The following functions are Hadamard differentiable tangentially to the set $\mathcal{C}^0$ of continuous functions on a given bounded interval $[a,b]$:*

1. $T_1(F_1,F_2,F_3,F_4,\mu)=\frac{\mu F_4-F_1\circ F_2^{-1}\circ F_3}{\mu-1}$ *and* $T_2(F_1,F_2,F_3,F_4,\lambda,\mu)=\frac{\mu F_1\circ F_2^{-1}\circ q(F_3,\lambda)-F_4}{\mu-1}$ *with either $q(F_4,\lambda)=\lambda F_4$ or $q(F_4,\lambda)=\lambda F_4+1-\lambda$, at any $(F_{10},F_{20},F_{30},F_{40},\lambda_0,\mu_0)\in(\mathcal{C}^1)^4\times[0,\infty)\times([0,\infty)\backslash\{1\})$, where $\mathcal{C}^1$ denotes the set of continuously differentiable functions with strictly positive derivative on $[a,b]$.*

2. $T_3(F_1) = \int_a^b m_1(F_1)(y)dy$ *and* $T_4(F_1, F_2) = \int_a^b F_2(m_1(F_1))(y)dy$*, at any* (F_{10}, F_{20}) *such that (i)* F_{10} *is increasing on* $[a, b]$ *and the equation* $F_{10}(y) = 1$ *admits at most one solution on* (a, b)*, (ii)* F_{20} *is continuously differentiable on* $[0, 1]$*. The same holds if we replace* m_1 *(and the equation* $F_{10}(y) = 1$*) by* M_0 *(and* $F_{10}(y) = 0$*).*

Proof: 1. We first prove that $\phi_1(F_1, F_2, F_3) = F_1 \circ F_2^{-1} \circ F_3$ is Hadamard differentiable at (F_{10}, F_{20}, F_{30}). Let $\mathcal{D}$ denote the set of bounded càdlàg functions on $[a, b]$. Because $(F_{10}, F_{20}) \in \mathcal{C}^1 \times \mathcal{C}^1$, the function $\phi_2 : (F_1, F_2, F_3) \mapsto (F_1 \circ F_2^{-1}, F_3)$ defined on $\mathcal{D} \times \mathcal{D} \times \mathcal{D}$ is Hadamard differentiable at (F_{10}, F_{20}, F_{30}) tangentially to $\mathcal{D} \times \mathcal{C}^0 \times \mathcal{D}$ (see, e.g., van der Vaart & Wellner, 1996, Problem 3.9.4). Moreover computations show that its derivative at (F_{10}, F_{20}, F_{30}) satisfies

$$d\phi_2(h_1, h_2, h_3) = (h_1 \circ F_{20}^{-1} - \frac{F'_{10} \circ F_{20}^{-1}}{F'_{20} \circ F_{20}^{-1}} h_2 \circ F_{20}^{-1}, h_3). \tag{29}$$

This shows that $d\phi_2(\mathcal{D} \times \mathcal{C}^0 \times \mathcal{C}^0) \subset \mathcal{D} \times \mathcal{C}^0$.

Moreover, the composition function $\phi_3 : (U, V) \mapsto U \circ V$ is Hadamard differentiable at any $(U_0, V_0) \in \mathcal{C}^1 \times \mathcal{D}$, tangentially to the set $\mathcal{C}^0 \times \mathcal{D}$ (see, e.g., van der Vaart & Wellner, 1996, Lemma 3.9.27). It is thus Hadamard differentiable at $(F_{10} \circ F_{20}^{-1}, F_{30}) \in \mathcal{C}^1 \times \mathcal{D}$. Thus, by the chain rule (see van der Vaart & Wellner, 1996, Lemma 3.9.3), $\phi_1 = \phi_3 \circ \phi_2$ is also Hadamard differentiable at (F_{10}, F_{20}, F_{30}) tangentially to $\mathcal{D} \times \mathcal{C}^0 \times \mathcal{C}^0$, and thus also tangentially to $(\mathcal{C}^0)^3$.

Finally, because T_1 (and similarly T_2) is a smooth function of $(\phi_1(F_1, F_2, F_3), F_4, \mu)$, it is also Hadamard differentiable at $(F_{10}, F_{20}, F_{30}, F_{40})$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.

2. we only prove the result for T_4 and m_1, the reasoning being similar (and more simple) for T_3 and M_0. We have

$$\begin{aligned} \frac{T_4(F_{10} + th_{tF}, F_{20} + th_{tH}) - T_4(F_{10}, F_{20})}{t} &= \int_a^b h_{tH} \circ m_1(F_{10} + th_{tF})(y)dy \\ &+ \int_a^b \frac{F_{20} \circ m_1(F_{10} + th_{tF}) - F_{20} \circ m_1(F_{10})}{t}(y)dy. \end{aligned}$$

Consider the first integral I_1.

$$\begin{aligned} & |h_{t2} \circ m_1(F_{10} + th_{t1})(y) - h_2 \circ m_1(F_{10})(y)| \\ \leq \; & |h_{t2} \circ m_1(F_{10} + th_{t1})(y) - h_2 \circ m_1(F_{10} + th_{t1})(y)| \\ + \; & |h_2 \circ m_1(F_{10} + th_{t1})(y) - h_2 \circ m_1(F_{10})(y)| \\ \leq \; & ||h_{t2} - h_2||_\infty + |h_2 \circ m_1(F_{10} + th_{t1})(y) - h_2 \circ m_1(F_{10})(y)| \end{aligned}$$

By uniform convergence of h_{t2} towards h_2, the first term in the last inequality converges to 0 when t goes to 0. By convergence of $m_1(F_{10} + th_{t1})$ towards $m_1(F_{10})$ and continuity of h_2, the second term also converges to 0. As a result,

$$h_{t2} \circ m_1(F_{10} + th_{t1})(y) \to h_2 \circ m_1(F_{10})(y).$$

Moreover, for t small enough,

$$|h_{t2} \circ m_1(F_{10} + th_{t1})(y)| \leq ||h_2|| + 1.$$

Thus, by the dominated convergence theorem, $I_1 \to \int_a^b h_2 \circ m_1(F_{10})(y)dy$, which is linear in h_2 and continuous since the integral is taken over a bounded interval.

Now consider the second integral I_2. Let us define $\underline{y}_1$ as the solution to $F_{10}(y) = 1$ on (a, b) if there is one such solution, $\underline{y}_1 = b$ otherwise. We prove that almost everywhere,

$$\frac{F_{20} \circ m_1(F_{10}(y) + th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t} \to H_0'(F_{10}(y))h_1(y)\mathbb{1}\{y < \underline{y}_1\}. \tag{30}$$

For $y < \underline{y}_1$, $F_{10}(y) < 1$, so that for t small enough, $F_{10}(y) + th_{t1}(y) < 1$. Therefore, for t small enough,

$$\begin{aligned}
\frac{F_{20} \circ m_1(F_{10}(y) + th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t} &= \frac{H_0y \circ (F_{10}(y) + th_{t1}(y)) - F_{20} \circ F_{10}(y)}{t} \\
&= \frac{(H_0'(F_{10}(y)) + \varepsilon(t))(F_{10}(y) + th_{t1}(y) - F_{10}(y))}{t} \\
&= (H_0'(F_{10}(y)) + \varepsilon(t))h_{t1}(y)
\end{aligned}$$

for some function $\varepsilon(t)$ converging towards 0 when t goes to 0. Therefore,

$$\frac{H_0y \circ m_1(F_{10}(y) + th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t} \to H_0'(F_{10}(y))h_1(y),$$

so that (30) holds. For $b > y > \underline{y}_1$, $F_{10}(y) > 1$, so that for t small enough, $F_{10}(y) + th_{t1}(y) > 1$. Therefore, for t small enough,

$$\frac{F_{20} \circ m_1(F_{10}(y) + th_{t1}(y)) - H_0y \circ m_1(F_{10}(y))}{t} = 0,$$

so that (30) holds as well. Thus, (30) holds almost everywhere.

Now, remark that m_1 is 1-Lipschitz. As a result,

$$\begin{aligned}
\left|\frac{F_{20} \circ m_1(F_{10}(y) + th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t}\right| &\leq ||F_{20}'|||h_{t1}(y)| \\
&\leq ||F_{20}'|| \left(|h_1(y)| + ||h_{t1} - h_1||_\infty\right).
\end{aligned}$$

Because $||h_{t1} - h_1||_\infty \to 0$, $|h_1(y)| + ||h_{t1} - h_1||_\infty \leq |h_1(y)| + 1$ for t small enough. Thus, by the dominated convergence theorem,

$$\int_a^b \frac{F_{20} \circ m_1(F_{10} + th_{t1}) - F_{20} \circ m_1(F_{10})}{t}(y)dy \to \int_a^{\underline{y}_1} H_0'(F_{10}(y))h_1(y)dy.$$

The right-hand side is linear with respect to h_1. It is also continuous since the integral is taken over a bounded interval and F_{20} is continuously differentiable. The second result follows.

References

Abadie, A., Angrist, J. & Imbens, G. (2002), 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings', *Econometrica* **70**(1), 91–117.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association* **91**(434), pp. 444–455.

Athey, S. & Imbens, G. W. (2002), Identification and inference in nonlinear difference-in-differences models, NBER Technical Working Papers 0280, National Bureau of Economic Research, Inc.

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.

Chernozhukov, V. & Hansen, C. (2005), 'An iv model of quantile treatment effects', *Econometrica* **73**(1), 245–261.

de Chaisemartin, C. (2011), Fuzzy differences in differences, Working Papers 2011-10, Centre de Recherche en Economie et Statistique.

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.

Evans, W. N. & Ringel, J. S. (1999), 'Can higher cigarette taxes improve birth outcomes?', *Journal of Public Economics* **72**(1), 135–154.

Faou, A.-L. L., Baha, M., Rodon, N., Lagrue, G. & Ménard, J. (2009), 'Trends in the profile of smokers registered in a national database from 2001 to 2006: changes in smoking habits', *Public Health* **123**, 6–11.

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).

Kitagawa, T. (2008), A bootstrap test for instrument validity in the heterogeneous treatment effect model. Working Paper.

Lee, D. S. (2009), 'Training, wages, and sample selection: Estimating sharp bounds on treatment effects', *Review of Economic Studies* **76**(3), 1071–1102.

Meyer, B. D., Viscusi, W. K. & Durbin, D. L. (1995), 'Workers' compensation and injury duration: Evidence from a natural experiment', *American Economic Review* **85**(3), 322–40.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), pp. 41–55.

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.