Fuzzy Difference in Differences & Fuzzy Change in Change*

Clément de Chaisemartin[†‡]

May 13, 2011

Abstract

Difference in Differences (DID) and Change in Change (CIC) require "perfect compliance": treatment rate should be 0% in the control group and during period 0 (no "always takers") and 100% in the treatment group in period 1 (no "never takers"). In many instances, the treatment rate increases more in the treatment than in the control group but there are never or always takers. This paper derives identification results which apply to such "fuzzy DID" and "fuzzy CIC" settings. Its first contribution is that its fuzzy DID identification results only require one common trend assumption on the outcome (Y) whereas the standard instrumental variable (IV) result usually invoked in such settings relies on a supplementary common trend assumption on treatment rate and on a "no defiers" assumption. When there are never takers but no always takers, common trend on Y is sufficient to identify an ATT as with standard DID. When there are always takers, it is no longer sufficient but partial identification is still possible provided Y is bounded. It is also possible to derive a second and narrower identification region under the supplementary assumption that treatment effects do not change between the two periods in the control group. I use those findings to measure the efficacy of a new pharmacotherapy for smoking cessation. Its second contribution is that it is the first paper which considers extending the CIC model to applications with imperfect compliance. The CIC assumptions are not sufficient for identification when the perfect compliance assumption is violated. One important exception is when there are no always takers, in which case

*I am extremely thankful to Xavier d'Haultfoeuille for his numerous advice and encouragements. I am very grateful to Andrew Clark, Joseph Doyle, Pauline Givord, Marc Gurgand, Edwin Leuven, Thierry Magnac, Eric Maurin, Thomas Piketty, Roland Rathelot, participants at the European Conference in Econometrics (EC^2), seminar participants at the Paris School of Economics and at CREST for their helpful comments.

[†]Paris School of Economics, 48 boulevard Jourdan 75014 Paris France, chaisemartin@pse.ens.fr

[‡]CREST, 15 boulevard Gabriel Péri 92245 Malakoff France

CIC assumptions are sufficient. Otherwise, only partial identification is obtained. Moreover, when there are always takers but the share of treated observations in the control group remained stable between period 0 and 1, point identification can be recovered through a slight strengthening of the CIC assumptions which amounts to defining an IV-CIC model.

Keywords: Difference in Differences, heterogeneous treatment effect, imperfect compliance, partial identification, smoking cessation, Change in Change, quantile treatment effects

JEL Codes: C21, C23, I19

Introduction

Since the seminal work by Ashenfelter and Card [1985], differences in differences (DID) are commonly used to estimate average treatment effects on the treated (ATT) when treatment D is not randomly allocated. DID compare the evolution of some mean outcome Y between two periods (0 and 1) and across two groups of individuals (control and treatment). In Rubin's causal model where potential outcomes with and without treatment ($Y(1)$ and $Y(0)$) are introduced, and where treatment effects ($Y(1) - Y(0)$) are allowed to be heterogeneous across observations, it has been shown that a DID identifies an ATT under two assumptions (see Abadie [2005]). The first one is a common trend assumption which states that if all observations had remained untreated the mean of Y would have followed parallel trends from period 0 to 1 in the two groups. The second one, which is implicit, is a perfect compliance assumption: the treatment rate should be equal to 0% in the control group and during period 0 (no "always takers") and to 100% in the treatment group in period 1 (no "never takers").[1] In many instances, this last assumption is violated: the treatment rate (or treatment intensity if treatment is multivariate) increases more in the treatment than in the control group but there are "never" or "always" takers.[2] This differential change in treatment rate / intensity across the control and the treatment group might still be used to identify an ATT. This is what I refer to as a fuzzy DID identification strategy.

When compliance is imperfect, common trend alone is not sufficient for identification in a model allowing for heterogeneous treatment effect. Under common trend on $Y(0)$, if no observation is treated in any group, trends are parallel in the two groups and the DID is merely equal to 0. In a standard DID, the only reason why trends might diverge across groups is that observations in the treatment group $\times$ period 1 cell get treated, so that the DID measures the effect of the treatment on them. A DID computation will therefore yield one equation with only one unknown. In a fuzzy DID, since there might be treated observations in each of the four time $\times$ group cells, diverging trends can potentially

[1] By never takers, I merely refer to untreated observations in the treatment group in period 1. Always takers are treated observations in the three other groups.

[2] This might not be an issue when panel data is available. In this case, researchers can indeed *choose* observations making up the treatment and the control group. They can for instance keep only observations of the control group who were untreated in period 0 and 1, and observations of the treatment group untreated in period 0 and treated in period 1 (see for instance Field [2005]). Despite its arbitrariness, which definition of groups ensures that the perfect compliance assumption is met. But when only pooled cross-sections are available, it is no longer possible to select observations thus.

arise from the effect of the treatment within each of those four subgroups and a DID computation will yield one equation with up to four unknowns. The identification problem arises because $Y(1)-Y(0)$ is allowed to vary across observations, implying that the effect of the treatment might vary across cells. Assuming $Y(1)-Y(0)$ to be constant across observations would solve the issue: the unknowns in the DID equation would all be equal to each other so that we would be back to one equation with one unknown.

Therefore, the starting point of the paper is to show that in a fuzzy DID, when treatment effect is allowed to be heterogeneous, a common trend assumption on $Y(0)$ is generally not sufficient to identify some ATT. However, in the special case where there are never takers but no always takers (a situation I henceforth refer to as the "no always takers" special case), this assumption is sufficient for identification, as in the standard DID model. Indeed, in such a situation, even though not all observations of the test group are treated in period 1, there are still treated observations in one group only, so that a DID computation will yield one equation with one unknown. When there are always takers, common trend on $Y(0)$ does not allow for point identification, but partial identification of some ATT is still possible provided Y is bounded. I derive explicit sharp bounds in this case. The identification region is likely to be narrow enough to identify the sign of this ATT when there are "few" always takers. Whether there are "many" or "few" never takers does not matter. It is also possible to derive a second and narrower identification region for the same ATT under the supplementary assumption that treatment effects do not change between the two periods in the control group. This second identification region will be narrow when there are few treated observations in the treatment group in period 0, and when the change in the the treatment rate from period 0 to 1 is small in the control group.

Actually, fuzzy DID has already been used often in the applied economics literature. Up to now, researchers who implemented it estimated the impact of the treatment through an instrumental variable (IV) regression using the interaction of time and group as an instrument for treatment. The resulting coefficient is the DID on Y divided by the DID on D. Duflo [2001] uses this strategy to estimate the impact of educational attainment on wages. Papers which use differential evolution of exposure to treatment across US states to estimate treatment effects build upon the same intuition. A good example is Evans and Ringel [1999] who use changes in cigarette taxes across US states as an

instrument for smoking prevalence among pregnant women, in order to estimate the impact of smoking during pregnancy on children's weight. Because their regressions include state and year fixed effects, their estimate arises from the comparison of the evolution of children's weight in states with changes in tax to the same evolution in states with no changes in tax. However, the underlying assumptions of this identification strategy have not been clarified so far. Imbens and Angrist [1994] have shown that IV coefficients can be interpreted as a local average treatment effect (LATE) in a model allowing for heterogeneous treatment effect. I put forward in a companion note (de Chaisemartin [2011]) that when applied to fuzzy DID their result holds under two common trend assumptions, on Y and on D, and a monotonicity assumption (no "defiers"). Common trend on Y allows recovering the intention to treat effect of the policy, whereas common trend on D allows recovering the share of compliers.

Consequently, my fuzzy DID results contribute to the literature because they require only one common trend assumption on Y. Thus, I remove the monotonicity condition. Even though it is often thought of as an innocuous assumption, it may be restrictive in some instances as discussed in Small and Tan [2007]. Above all, they do not require common trend on D. One might argue that the marginal cost of this second common trend assumption is weaker than for the first: if one is ready to believe that without the program trends would have been parallel on Y, one should be ready to take the same assumption on D. However, this might not always be true. For instance, in Evans and Ringel [1999], it may be the case that states which choose to rise taxes on cigarettes do so because they face an increasing trend in smoking, whereas there is no reason to suspect that this decision is related to trends on babies weight at birth. Moreover, even in applications where there is no obvious reason to suspect that trends on Y or on D would have strongly diverged, there is no reason why they should have been **exactly** parallels neither because assignment to treatment is not random. The most one can reasonably expect is that trends in the untreated group provide a fairly good first order approximation of what would have happened in the treated group. Results requiring one first order approximation might therefore be more reliable than results requiring two. The combination of two small errors in the numerator and in the denominator of the Wald-DID could indeed lead to a large difference between the Wald-DID and the true treatment effect. Therefore, the first contribution of this paper is to bring new fuzzy DID identification results which rely on weaker assumptions than the

standard Imbens and Angrist IV result.

Those fuzzy DID results might be useful in applications with no or few always takers. To illustrate this, I measure the efficacy of a new pharmacotherapy for smoking cessation. Varenicline is a drug which was made available to French cessation clinics in February 2007 as one possible pharmacotherapy for smoking cessation support. In 15 services, less than 3% of all patients consulted have been prescribed varenicline during the year following its release. In 13 services, more than 20% of patients were prescribed varenicline. Because in this application there are some but few always takers, I derive bounds for the ATT which are narrow enough to infer its sign. Had there been more always takers, 0 would lie within the identification region. Therefore, in a fuzzy DID, common trend on Y is sufficient to obtain accurate information on an ATT when there are few always takers, even if there are many never takers. My results might also be useful in applications considering the extension of a policy, that is to say when the control group was already eligible in period 0 and the test group became eligible in period 1 (see for instance Bach [2009]). Indeed, in such situations the share of treated observations in the treatment group in period 1 is by definition equal to 0. Consequently, the second identification region I derive will be narrow provided the share of treated observations did not change too much between period 0 and 1 in the control group.

The main limitation of DID and fuzzy DID is that they identify only the average effect of the treatment within specific populations whereas one might be interested in other parameters, such as quantile treatment effects. In the DID literature, two approaches already exist to estimate quantile treatment effects. The most common approach is the quantile-DID, in which the transformation used to reconstruct the counterfactual distribution of the outcome is to add the change over time at the qth quantile in the control group to the qth quantile of the first-period treatment group (see Meyer et al. [1995] and Poterba et al. [1995]). This amounts to matching treatment and control observations in period 0 on their quantile, and period 0 and period 1 observations in the control group on their quantile as well. Athey and Imbens [2002] show that the model rationalizing this transformation has several unattractive features: it assumes that time and group effects are additively separable, its assumptions are not robust to a monotonous transform of Y and it places restrictions on the data. Therefore, Athey and Imbens [2006] suggested another transformation to reconstruct the counterfactual distribution of

the outcome. It amounts to matching period 0 treatment and control observations on their value of Y, and period 0 and period 1 control observations on their quantile. They show that this transformation can be rationalized by a model, the Change in Change (CIC) which has several important advantages with respect to the quantile-DID model: it does not rely on additive separability in time and group, its assumptions are robust to a monotonous transform of Y and it does not place restrictions on the data.

In fuzzy applications, that is to say in situations where the treatment rate increases more in one group but where there are never and / or always takers, there is not yet a well-established procedure to study quantile treatment effects. A first solution could be quantile IV regressions in which treatment is instrumented by a time and group interaction. However, we still lack a clear consensus on how to do quantile IV-regressions (see Abadie et al. [2002] and Chernozukhov and Hansen [2005]). Moreover, to the best of my knowledge no paper considered whether the existing quantile IV regressions are adapted to fuzzy DID. An extension of the CIC model to situations of imperfect compliance could provide a second solution. The second contribution of this paper is to develop such an extension.

Fuzzy CIC results are strikingly close to fuzzy DID results. Indeed, my first fuzzy CIC result is a non identification result: when compliance is imperfect, the mechanics of the CIC model collapse. To understand why, one needs first to understand how the CIC model works. Think of Y as wages, and of treatment as whether an individual completed highschool. Potential wages with and without completing highschool are denoted $Y(1)$ and $Y(0)$. The perfect compliance assumption states that all observations in the test group $\times$ period 1 cell completed highschool and no observations in the three remaining cells completed it. We seek to reconstruct the counterfactual distribution of wages without completing highschool in the period 1 $\times$ test group cell. On that purpose, period 0 observations with same wages in the test and in the control group are matched. To rationalize this matching, it is assumed that $Y(0)$ is a function of time and of an unobserved heterogeneity index U. Since those observations did not complete highschool, are observed at the same period and have the same wages, they must have the same U. Then period 0 and period 1 observations in the control group are matched on their quantile in the distribution of wages. It is indeed assumed that groups are stable over time so that the distribution of U is time invariant within group: since those observations belong to the

same group and have same rank they must have the same U. Finally, combining those two matching, period 0 observations in the treatment group are matched to period 1 observations in the control group with same U. Wages of the latter observations are counterfactual wages that the former would have obtained in period 1 if they had not completed highschool.

But when compliance is imperfect, this double matching collapses. Two control and test group observations with same wages in period 0 might no longer have the same U, because one might have completed highschool whereas the other might not. Despite the fact they have same wages, the observation which did not complete highschool probably has greater unobserved ability, which enabled it to compensate for its lower education. Similarly, two control group observations with same rank in period 0 and 1 might not have the same U. Assume that more observations completed highschool in period 1. Then, the period 1 observation probably has greater unobserved ability: it has the same rank despite the fact that more people have completed highschool in its cell.

Despite this general non identification result, there is one special case of imperfect compliance to which the CIC model readily extends, exactly as with fuzzy DID. When there are no always takers, the counterfactual distribution of wages without completing highschool in the period 1 $\times$ treatment group cell is identified under the exact same assumptions as in Athey and Imbens [2006]. This is because in this special case no observations are treated in period 0, so that period 0 treatment and control observations can be matched on their wages as in the standard CIC model. Moreover, since by assumption no observations are treated in period 0 and 1 in the control group, control observations in the two periods can be matched based on their quantile: their ranking cannot have been disrupted by a change in treatment rate. Therefore, the double matching process works. One just needs to account for the fact that not all observations in the period 1 $\times$ treatment group cell are treated when computing quantile treatment effects. Moreover, I also show that even when there are always takers, CIC assumptions are still sufficient to place bounds on the distribution of $Y(0)$ among treated observations of the period 1 $\times$ treatment group cell so that quantile treatment effects are partially identified. The resulting bounds will be tight when the shares of treated observations within the three remaining cells are small as what happens with fuzzy DID.

When there are large shares of always takers, quantile treatment effects can still be identified

through a strengthening of Athey and Imbens's assumptions. It requires introducing an instrument for treatment which should be in a one to one relationship with the time and group interaction term. This could for instance be a policy which gives supplementary incentives to complete highschool such as a new benefit, which was released in period 1 and is available only to test group individuals. It also requires assuming that treated (resp. untreated) observations in the control group have the same distribution of U in period 0 and 1. For this last assumption to be credible, the share of treated observations should be stable between period 0 and 1 in the control group. Under those supplementary assumptions, distributions of $Y(1)$ and $Y(0)$ among compliers of the period 1 $\times$ treatment group cell are identified, as well as various parameters of interest such as a Local Average Treatment Effect (LATE) and quantile treatment effects within this population. This result will prove particularly useful in applications considering extensions of a program to a new group previously not eligible and using a previously eligible group as a control.

This result can be seen as a combination of ideas in Abadie [2003] and in Athey and Imbens [2006]. To recover the distribution of $Y(1)$ among compliers of the period 1 $\times$ treatment group cell, I consider the distribution of wages among all treated observations in this cell. Since those observations include both compliers and always takers, I need to "withdraw" from it the distribution of wages among always takers. This is the same idea as in the "weighting" scheme suggested by Abadie to recover statistical characteristics of compliers (see Abadie [2003], and Frölich and Melly [2008] for an application to unconditional quantiles). But Abadie and Frölich and Melly have in mind applications to randomized experiments, where the distribution of wages among always takers can be recovered from the distribution of wages among treated observations in the control group due to random assignment. Here, groups are not random. Therefore, I use distributions of wages among always takers in the three remaining cells to reconstruct the distribution of $Y(1)$ among always takers of the period 1 $\times$ treatment group cell through the same double-matching process as in Athey and Imbens. Those three distributions are observed since by definition observations which completed highschool in the three remaining cells must be always takers. Period 0 always takers in the treatment and in the control group are matched on their wages. Since they have the same treatment status, are observed at the same period and have the same wage, they must have the same U. Then, period 0 and 1 always

takers in the control group are matched on their quantile. This will yield couples of observations with same U since the distribution of U among always takers in the control group is the same in period 0 and 1, hence the need to assume that the share of treated observations is stable across periods in the control group. Finally, the period 0 $\times$ treatment group always takers is matched to his period 1 $\times$ control group counterpart. To recover the distribution of $Y(0)$ among compliers of the period 1 $\times$ treatment group cell, I also proceed in two steps. First, I reconstruct the distribution of $Y(0)$ among compliers and never takers of this cell from distributions of $Y(0)$ among untreated observations in the three remaining cells. Then, I "withdraw" from this reconstructed distribution the distribution of $Y(0)$ among never takers (i.e. untreated observations) of the period 1 $\times$ treatment group cell.

Finally, I show that in applications where the share of treated observations also increases in the control group, partial identification of the distributions of $Y(1)$ and $Y(0)$ among compliers of the period 1 $\times$ treatment group cell is obtained through a strengthening of IV-CIC assumptions. This result will yield tight bounds in applications where the change in the treatment rate in the control group is small.

Therefore, the second contribution of this paper is to derive identification results inspired from the CIC model which allow computing or bounding quantile treatment effects in fuzzy applications. In applications with never takers but no always takers, and in applications with potentially large shares of always takers but where the share of treated observations remained stable between period 0 and 1 in the control group, exact identification is obtained. In applications with few always takers or where the share of treated observations does not increase much in the control group, tight bounds on quantile treatment effects are obtained.

The remainder of the paper is divided into two parts. The first part deals with fuzzy DID. Section 1 is devoted to fuzzy DID identification results. Section 2 considers inference. Section 3 is devoted to the application. Then, the second part deals with fuzzy CIC. Section 4 presents fuzzy CIC identification results. The last section concludes.

Part I

Fuzzy Difference in Differences

1 Identification

I place myself in the pooled cross-section case: each individual is observed only at one period. Let $T \in \{t_0; t_1\}$ denote time and $G \in \{g_c; g_t\}$ denote treatment (g_t) and control (g_c) groups. I assume that treatment status is binary and is denoted by an indicator D (results can easily be extended when treatment is discrete).

Throughout the paper it is implicitly assumed that the stable unit treatment value assumption holds. Under this assumption I define $Y(1)$ and $Y(0)$ as the potential outcomes of an individual with and without the treatment. Only the actual outcome $Y = Y(1) \times D + Y(0) \times (1 - D)$ is observed. The treatment effect is $Y(1) - Y(0)$. Average treatment effects are the corresponding expectations. $X \sim Y$ means that X and Y have the same probability distribution. $\mathbb{X}$ is the support of X. To alleviate the notational burden, I introduce several shorthands following Athey and Imbens [2006]:

$Y_{ij}(k) \sim Y(k) \,|\, t = i,\, g = j \;\forall (k, i, j) \in \{0; 1\} \times \{t_0; t_1\} \times \{g_c; g_t\}$

$Y_{ij} \sim Y \,|\, t = i,\, g = j \;\forall (i, j) \in \{t_0; t_1\} \times \{g_c; g_t\}$

$D_{ij} \sim D \,|\, t = i,\, g = j \;\forall (i, j) \in \{t_0; t_1\} \times \{g_c; g_t\}$

Under those notations, the standard DID parameter is:

$$DID = \mathbb{E}(Y_{t_1,g_t}) - \mathbb{E}(Y_{t_0,g_t}) - [\mathbb{E}(Y_{t_1,g_t}) - \mathbb{E}(Y_{t_0,g_c})]\,.$$

I denote by DID^P the DID on treatment rate from period 0 to 1 across the two groups. I assume that $DID^P \neq 0$: the definition of a fuzzy DID is that exposure to treatment should have evolved differentially in the two groups. Without loss of generality, I assume that $DID^P > 0$. The no always takers special case is met when $\mathbb{P}(D_{t_0,g_t} = 1) = \mathbb{P}(D_{t_1,g_c} = 1) = \mathbb{P}(D_{t_0,g_c} = 1) = 0$. It is likely to arise for instance when a new social program is implemented with only a specific group eligible

to it (unemployed...) and take-up is below 100%. $ATT_{i,j} = \mathbb{E}(Y_{i,j}(1) - Y_{i,j}(0)|\, D = 1), \forall (i,j) \in \{t_0; t_1\} \times \{g_c; g_t\}$ is the average treatment effect on treated individuals of group j in period i. $ATT = \mathbb{E}(Y(1) - Y(0)|\, D = 1)$ is the average treatment effect on the treated. I denote $\mathbb{P}_{AT} = \mathbb{P}(D_{t_0,g_t} = 1) + \mathbb{P}(D_{t_1,g_c} = 1) + \mathbb{P}(D_{t_0,g_c} = 1)$ the sum of the three shares of always takers.

I take a common trend assumption which is at the basis of the DID approach (see for instance Abadie [2005]):

Assumption DID 1: Common trend for the outcome variable

$$\mathbb{E}(Y_{t_1,g_t}(0)) - \mathbb{E}(Y_{t_0,g_t}(0)) = \mathbb{E}(Y_{t_1,g_c}(0)) - \mathbb{E}(Y_{t_0,g_c}(0)).$$

Lemma DID 1: Non-identification.

Under Assumption DID 1, none of the $ATT_{i,j}$ is identified and

$$DID = ATT_{t_1,g_t} \times \mathbb{P}(D_{t_1,g_t} = 1) - ATT_{t_0,g_t} \times \mathbb{P}(D_{t_0,g_t} = 1)$$

$$-ATT_{t_1,g_c} \times \mathbb{P}(D_{t_1,g_c} = 1) + ATT_{t_0,g_c} \times \mathbb{P}(D_{t_0,g_c} = 1). \quad (1)$$

According to Lemma DID 1, under Assumption DID 1, if compliance is imperfect, the DID on Y can be written as a weighted DID of four average treatment effects on four different populations. This is the equation with several unknowns mentioned in the introduction. Because two ATT enter the equation with positive sign and two enter with negative sign, the DID cannot be given any causal interpretation. It might for instance be positive whereas the four ATT are negative. The intuition for this result is that under common trend on $Y(0)$, if no observations had been treated in any of the four time $\times$ group cells, trends would have been parallel in the two groups and the DID would have merely been equal to 0. In a standard DID, the only reason why trends might diverge across groups is that observations in the treatment group get treated in period 1, so that the DID measures the effect of the treatment on them. In a fuzzy DID, since there might be treated observations in several time $\times$ group cells, diverging trends can potentially arise from the effect of the treatment in each of those cells. Then, if no restrictions are placed on how heterogeneous the treatment effect can be across these

four subgroups, it is not possible to identify any of the $ATT_{i,j}$ from a standard DID computation, since it yields one equation with several unknowns.

Proposition DID 1: Point identification.

i) In the no always takers special case, Assumption DID 1 is sufficient for ATT_{t_1,g_t} to be identified and

$$ATT_{t_1,g_t} = \frac{DID}{\mathbb{P}(D_{t_1,g_t} = 1)}$$

ii) Under Assumption DID 1 and the supplementary assumption that $\forall (i,j) \in \{t_0; t_1\} \times \{g_c; g_t\}$, $ATT_{i,j} = ATT$, the $ATT_{i,j}$ and the ATT are identified:

$$\forall (i,j) \in \{t_0; t_1\} \times \{g_c; g_t\}, \; ATT_{i,j} = ATT = \frac{DID}{DID^P}$$

In the no always takers special case, common trend is sufficient to identify ATT_{t_1,g_t} as in a standard DID because there are treated observations in one group only. Therefore, there is only one unknown left in (1). This result is strikingly similar to Battistin and Rettore's [2008] result on regression discontinuity (RDD). They indeed show that in a fuzzy RDD, when treatment rate is equal to 0 below the eligibility threshold, so that fuzziness arises only because of never takers (i.e. untreated individuals above the threshold), identification is obtained under the same assumptions than in a sharp RDD. Estimation of ATT_{t_1,g_t} still requires being able to estimate $\mathbb{P}(D_{t_1,g_t} = 1)$. Sometimes treatment status is not observed, making it impossible to estimate $\mathbb{P}(D_{t_1,g_t} = 1)$ (see e.g. Eissa and Leibman [1996]). Since ATT_{t_1,g_t} and DID have the same sign and $|DID| \leq |ATT_{t_1,g_t}|$, it is at least possible to estimate a lower bound of ATT_{t_1,g_t} by computing the DID. For instance, Eissa and Leibman's 1.4 percentage points DID is a lower bound on the true effect of the EITC extension on lone mothers' participation to the labor market.

Then in part ii) of Proposition DID 1, I show that it is enough to restrict the heterogeneity of the treatment effect, assuming that it does not vary across time and group, to identify exactly the ATT. This is because under this assumption the four unknowns in (1) are actually equal to each other. But this is fairly restrictive an assumption. The underlying assumption to a fuzzy DID is indeed that treatment rate increased more from period 0 to 1 in the treatment group than in the control group.

This might for instance be the case because treatment group individuals were more incentivized to receive the treatment in period 1 than in period 0. Inside the treatment group, treated individuals during period 1 are therefore likely to differ from those treated during period 0 so that the average treatment effect could arguably be different in these two groups.

Before stating Proposition DID 2, I define three quantities:

$$B^0(u,v) = \frac{DID + \big(\mathbb{E}(Y_{t_0,g_t}|\,D=1) - u\big)\times\mathbb{P}(D_{t_0,g_t}=1) + \big(\mathbb{E}(Y_{t_1,g_c}|\,D=1) - u\big)\times\mathbb{P}(D_{t_1,g_c}=1) - \big(\mathbb{E}(Y_{t_0,g_c}|\,D=1) - v\big)\times\mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_1,g_t}=1)},$$

$$B^1 = \frac{DID + \big(\mathbb{E}(Y_{t_0,g_t}|\,D=1) - M\big)\times\mathbb{P}(D_{t_0,g_t}=1) + \big(max\big(\mathbb{E}(Y_{t_1,g_c}|\,D=1); \mathbb{E}(Y_{t_0,g_c}|\,D=1)\big) - M\big)\times\big(\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)\big)}{\mathbb{P}(D_{t_1,g_t}=1)}$$

and

$$B^2 = \frac{DID + \big(\mathbb{E}(Y_{t_0,g_t}|\,D=1) - m\big)\times\mathbb{P}(D_{t_0,g_t}=1) + \big(min\big(\mathbb{E}(Y_{t_1,g_c}|\,D=1); \mathbb{E}(Y_{t_0,g_c}|\,D=1)\big) - m\big)\times\big(\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)\big)}{\mathbb{P}(D_{t_1,g_t}=1)}.$$

Proposition DID 2: Partial Identification.

i) Under Assumption DID 1 and the supplementary assumption that $\exists(m,\, M) \in \mathbb{R}^2 / \mathbb{P}(m \leq Y(0) \leq M) = 1$,

$$B_- \leq ATT_{t_1,g_t} \leq B_+.$$

$B_- = max\left(B^0(M,m)\,;\, \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M\right)$ *and* $B_+ = min\left(B^0(m,M)\,;\, \mathbb{E}(Y_{t_1,g_t}|\,D=1) - m\right)$,

B_- *and* B_+ *are sharp.*

$\mathbb{P}_{AT} \leq \mathbb{P}(D_{t_1,g_t}=1)$ *is a sufficient condition to have that either* $B_- = B^0(M,m)$ *or* $B_+ = B^0(m,M)$.

ii) Under Assumption DID 1 and the supplementary assumptions that $\exists(m,\, M) \in \mathbb{R}^2 / \mathbb{P}(m \leq Y(0) \leq M) = 1$ *and that* $ATT_{t_1,g_c} = ATT_{t_0,g_c}$,

$$B_-^{'} \leq ATT_{t_1,g_t} \leq B_+^{'}$$

$B_-^{'} = max\left(min(B^1\,;\, B^2)\,;\, \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M\right)$ *and*

$B_+^{'} = min\left(max(B^1\,;\, B^2)\,;\, \mathbb{E}(Y_{t_1,g_t}|\,D=1) - m\right)$.

$B_-^{'}$ *and* $B_+^{'}$ *are sharp.*

$\mathbb{P}(D_{t_0,g_t}=1) + |\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)| \leq \mathbb{P}(D_{t_1,g_t}=1)$ *is a sufficient condition to have that either* $B_-^{'} = min(B^1\,;\, B^2)$ *or* $B_+^{'} = max(B^1\,;\, B^2)$.

If $Y(0)$ is bounded, it is possible to find bounds for ATT_{t_1,g_t} which can be non-parametrically estimated from the sample in the spirit of Manski [1990]. This comes from the fact that the only three

quantities appearing in (1) which are not observed and do not belong to ATT_{t_1,g_t} are $\mathbb{E}(Y_{t_0,g_t}(0)|\, D=1)$, $\mathbb{E}(Y_{t_1,g_c}(0)|\, D=1)$ and $\mathbb{E}(Y_{t_0,g_c}(0)|\, D=1)$. Therefore, it suffices to build up worst-case scenarii for each of them to derive bounds for ATT_{t_1,g_t}. But those worst case scenarii might not be compatible with the common trend assumption and might therefore yield values lower (resp. higher) than the lowest (resp. highest) possible value for ATT_{t_1,g_t} compatible with the data, i.e. $\mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$ (resp. $\mathbb{E}(Y_{t_1,g_t}|\, D=1) - m$). Hence the need to ensure that $B_- \geq \mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$ and $B_+ \leq \mathbb{E}(Y_{t_1,g_t}|\, D=1) - m$. If $B_- = \mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$ and $B_+ = \mathbb{E}(Y_{t_1,g_t}|\, D=1) - m$, the bounds are uninformative. If $\mathbb{P}_{AT} \leq \mathbb{P}(D_{t_1,g_t}=1)$, that is to say if the share of treated observations in the period 1 $\times$ treatment group cell is greater than the shares of always takers, at least one of the bounds is informative. Conversely, when $\mathbb{P}_{AT} > \mathbb{P}(D_{t_1,g_t}=1)$, at least one of the bounds in uninformative. There is no sufficient condition on $\mathbb{P}_{AT}$ which ensures that the two bounds are informative (except $\mathbb{P}_{AT}=0$), because even when $\mathbb{P}_{AT}$ is very small, it is still possible to build up a DGP such that one of the bounds is uninformative, for instance setting $\mathbb{E}(Y_{t_0,g_t}(0)|\, D=1) = M$. Apart from such extreme cases, if $\mathbb{P}_{AT} \leq \mathbb{P}(D_{t_1,g_t}=1)$, it is likely that the two bounds will be informative. This condition appears because $\mathbb{P}_{AT}$ is the "size" of the three subgroups for which $Y(0)$ is not observed, which enter into (1), and for which worst case scenarii must be constructed. $\mathbb{P}(D_{t_1,g_t}=1)$ is the size of the only subgroup for which $Y(0)$ is not observed, which enters the common trend equation and does not enter into (1), that is to say the size of the only degree of freedom left to verify common trend once worst case scenarii have been constructed for the three groups of always takers. When the two bounds are informative, the length of $[B_-; B_+]$ is equal to $(M-m) \times \frac{\mathbb{P}_{AT}}{\mathbb{P}(D_{t_1,g_t}=1)}$. It is increasing with $\mathbb{P}_{AT}$, and decreasing with $\mathbb{P}(D_{t_1,g_t}=1)$. However, whether 0 belongs to $[B_-; B_+]$ does not depend on $\mathbb{P}(D_{t_1,g_t}=1)$ but on the size of DID with respect to $M-m$, , $\mathbb{P}(D_{t_0,g_t}=1)$, $\mathbb{P}(D_{t_1,g_c}=1)$, and $\mathbb{P}(D_{t_0,g_c}=1)$.

In part ii) of Proposition DID 2 I show that narrower bounds for ATT_{t_1,g_t} can be derived under the supplementary assumption that the ATT is constant over time in the control group.[3] Such an assumption might be credible for instance when the treatment rate does not significantly change between period 0 and 1 in the control group, when observable characteristics of treated individuals

[3] I am very grateful to Roland Rathelot for suggesting this result.

in the control group do not change much over the two periods, or when $\mathbb{E}(Y_{t_1,g_c}|\,D=1)$ is close from $\mathbb{E}(Y_{t_0,g_c}|\,D=1)$. Under this hypothesis, (1) becomes an equation with only three unknowns, and worst case analysis must be conducted on only two expectations. Those worst case scenarii might also not be compatible with common trend and may therefore yield lower and upper bounds outside the range of values of ATT_{t_1,g_t} compatible with the data, hence the need to ensure that $B^{'}_{-} \geq \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$ and $B^{'}_{+} \leq \mathbb{E}(Y_{t_1,g_t}|\,D=1) - m$ for the bounds to be sharp. If $\mathbb{P}(D_{t_0,g_t}=1) + |\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)| \leq \mathbb{P}(D_{t_1,g_t}=1)$, at least one of the two bounds will be informative. The sign of ATT_{t_1,g_t} will be identified if both $\mathbb{P}(D_{t_0,g_t}=1)$ and $|\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)|$ are small. With respect to part i) of the Proposition, $\mathbb{P}(D_{t_1,g_c}=1) + \mathbb{P}(D_{t_0,g_c}=1)$ has been replaced by $|\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)|$: what matter are no longer the shares of treated observations in the control group but the change in this share from period 0 to 1. This is somewhat similar to the change in the size of the identification region when Lee bounds (see Lee [2009] and Horowitz and Manski [1995]) are used to deal with attrition instead of Manski bounds. This result is of particular interest to place narrow bounds on the ATT in applications considering the extension of policy to a group which was previously not eligible to it and which use a group previously eligible as the control group. Indeed, in such cases, $\mathbb{P}(D_{t_0,g_t}=1)=0$. Consequently, if the change in the treatment rate from period 0 to 1 in the control group is not too large, $\left[B^{'}_{-}; B^{'}_{+}\right]$ will be narrow. Point identification can even be obtained if $\mathbb{P}(D_{t_1,g_c}=1) = \mathbb{P}(D_{t_0,g_c}=1)$.

2 Inference

The objective of this section is to build up confidence intervals (CI) for ATT_{t_1,g_t} based upon the identification results of section 2. I denote LB^{θ}_{x} and UB^{θ}_{x} the lower and upper bounds of the CI of a parameter θ with $x\%$ asymptotic coverage. A first candidate is $CI^1 = \left[LB^{\frac{DID}{DID^P}}_{(1-\alpha)}; UB^{\frac{DID}{DID^P}}_{(1-\alpha)}\right]$. In the no always takers special case, common trend is enough for CI^1 to be a consistent CI for ATT_{t_1,g_t}, since $ATT_{t_1,g_t} = \frac{DID}{DID^P}$. But when there are always takers, CI^1 is a CI for ATT_{t_1,g_t} (i.e. $ATT_{t_1,g_t} = \frac{DID}{DID^P}$) only under the very strong assumption that ATT do not vary across time $\times$ group cells. In such cases, partial identification results might allow deriving CI for ATT_{t_1,g_t} under weaker assumptions. This is the purpose of Proposition DID 3.

Proposition DID 3: CI for ATT_{t_1,g_t}based on partial identification results

i) Under Assumption DID 1 and the supplementary assumption that $\exists(m,\, M) \in \mathbb{R}^2/\mathbb{P}(m \leq Y(0) \leq M) = 1$, $CI^2 = \left[LB^{B_-}_{(1-\alpha)}; UB^{B_+}_{(1-\alpha)}\right]$ *and* $CI^3 = \left[LB^{B_-}_{(1-2\alpha)}; UB^{B_+}_{(1-2\alpha)}\right]$ *are CI for* ATT_{t_1,g_t} *with asymptotic coverage of* $(1-\alpha)\%$.

ii) Under Assumption DID 1 and the supplementary assumptions that $\exists(m,\, M) \in \mathbb{R}^2/\mathbb{P}(m \leq Y(0) \leq M) = 1$ *and that* $ATT_{t_1,g_c} = ATT_{t_0,g_c}$, *if either* $\mathbb{P}(D_{t_1,g_c} = 1) - \mathbb{P}(D_{t_0,g_c} = 1) \neq 0$ *or* $\mathbb{P}(D_{t_0,g_t} = 1) \neq 0$, *then*
$CI^4 = \left[LB^{B'_-}_{(1-\alpha)}; UB^{B'_+}_{(1-\alpha)}\right]$ *and* $CI^5 = \left[LB^{B'_-}_{(1-2\alpha)}; UB^{B'_+}_{(1-2\alpha)}\right]$ *are CI for* ATT_{t_1,g_t} *with asymptotic coverage of* $(1-\alpha)\%$.

Based on the first partial identification result in Proposition DID 2, one can build a CI for ATT_{t_1,g_t} with $(1-\alpha)\%$ asymptotic coverage using the lower bound of the $(1-\alpha)\%$ CI of B_- and the upper bound of the $(1-\alpha)\%$ CI of B_+. This yields CI^2. As shown in Imbens and Manski [2004], using $(1-2\alpha)\%$ lower and upper bounds will also yield a CI for ATT_{t_1,g_t} with $(1-\alpha)\%$ asymptotic coverage. This is CI^3. However, it suffers from uniform convergence issues: when we get close to point identification ($\mathbb{P}_{AT} \to 0$), CI^3 will be narrower than CI^1 despite the fact that it is based on a partial identification result whereas CI^1 relies on point identification and stronger assumptions. To circumvent this issue, Imbens and Manski introduce a third CI lying in-between the $(1-\alpha)\%$ and the $(1-2\alpha)\%$ CI. It accounts for the fact that because the parameter is partially identified, the $(1-\alpha)\%$ CI is too conservative and also avoids the above mentioned uniform convergence issue. Stoye [2009] shows that this third CI relies on a superefficiency condition which is verified when by construction $\widehat{B_-} \leq \widehat{B_+}$ and when

$$\sqrt{n}\begin{pmatrix} \widehat{B_-} - B_- \\ \widehat{B_+} - B_+ \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

uniformly in $\mathcal{P}$. While the former is true here, the latter is not as shown in Proposition DID 4. Therefore, this third CI cannot be used here.

Finally, based on the second identification result in Proposition DID 2 which relies on stronger identifying assumptions, one can use B'_- and B'_+ instead of B_- and B_+ to build up CI for ATT_{t_1,g_t}. Using the lower bound of the $(1-\alpha)\%$ CI of B'_- and the upper bound of the $(1-\alpha)\%$ CI of B'_+ yields

CI^4. Using the corresponding $(1-2\alpha)\%$ bounds yields CI^5.

Proposition DID 3 shows how to build up CI for ATT_{t_1,g_t} based upon CI for B_-, B_+, $B_-^{'}$ and $B_+^{'}$. I show now how to construct such CI for B_- and B_+. Let $(Y_i,\, D_i,\, T_i,\, G_i)_{1\leq i\leq n}$ be an iid sample of size n drawn from the distribution of $(Y,\, D,\, T,\, G)$. I assume that $\mathbb{P}(T=i,\, G=j)>0\ \forall (i,\, j)\in \{t_0;t_1\}\times\{g_c;g_t\}$ and that Y is bounded, meaning that $\exists (m,\, M)\in\mathbb{R}^2/\mathbb{P}(m\leq Y(0)\leq M)=1$, where m and M are known by the econometrician. Empirical counterparts are used to estimate B_- and B_+. I consider the asymptotic behavior of $\widehat{B_-}$ and $\widehat{B_+}$. On that purpose, I define a variance matrix $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\\ \rho & \sigma_2^2\end{pmatrix}$ whose explicit expression is given in Appendix B and which can be consistently estimated by $\widehat{\Sigma}$.

Proposition DID 4: $\sqrt{n}$-consistency of $\widehat{B_-}$ and $\widehat{B_+}$.

If $B^0(M,m)>\mathbb{E}(Y_{t_1,g_t}|\,D=1)-M$,

$$\sqrt{n}\left(\widehat{B_-}-B_-\right)\xrightarrow{d}\mathcal{N}(0,\sigma_1^2).$$

If $B^0(M,m)=\mathbb{E}(Y_{t_1,g_t}|\,D=1)-M$,

$$\sqrt{n}\left(\widehat{B_-}-B_-\right)\xrightarrow{d}S^1$$

where $S^1=max\left(N^1;N^2\right)$ *with* $\begin{pmatrix}N^1 & N^2\end{pmatrix}^{'}\sim\mathcal{N}\left(0,\Sigma\right)$.

If $B^0(M,m)<\mathbb{E}(Y_{t_1,g_t}|\,D=1)-M$,

$$\sqrt{n}\left(\widehat{B_-}-B_-\right)\xrightarrow{d}\mathcal{N}(0,\sigma_2^2).$$

Similarly one can show that $\widehat{B_+}$ *is* $\sqrt{n}$*-consistent with three possible limiting distributions depending on the respective positions of* $B^0(m,M)$ *and* $\mathbb{E}(Y_{t_1,g_t}|\,D=1)-m$.

B_- is not differentiable at $B^0(M,m)=\mathbb{E}(Y_{t_1,g_t}|\,D=1)-M$ and B^+ is not differentiable at $B^0(m,M)=\mathbb{E}(Y_{t_1,g_t}|\,D=1)-m$. Therefore, $\sqrt{n}\left(\widehat{B_-}-B_-\right)$ and $\sqrt{n}\left(\widehat{B_+}-B_+\right)$ do not converge to a normal distribution uniformly in $\mathcal{P}$. If $B^0(M,m)>\mathbb{E}(Y_{t_1,g_t}|\,D=1)-M$, $\sqrt{n}\left(\widehat{B_-}-B_-\right)$

converges to a normal distribution. If $B^0(M,m) < \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$, it converges to another normal distribution. If $B^0(M,m) = \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$, its limiting distribution is non standard.

In all cases, it is possible to build CI for B_- and B_+. Let us consider B_- (the reasoning follows the same steps for B_+). If $B^0(M,m) > \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$, a CI for B_- is $CI^A = \left[\widehat{B^0(M,m)} - \frac{q_{1-\frac{\alpha}{2}} \times \widehat{\sigma_1^2}}{\sqrt{n}}; \widehat{B^0(M,m)} + \frac{q_{1-\frac{\alpha}{2}} \times \widehat{\sigma_1^2}}{\sqrt{n}}\right]$, where $q_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}^{th}$ quantile of a $\mathcal{N}(0,1)$ distribution. If $B^0(M,m) = \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$, a CI for B_- is $CI^B = \left[\widehat{B_-} + \frac{\widetilde{q}_{\frac{\alpha}{2}}}{\sqrt{n}}; \widehat{B_-} + \frac{\widetilde{q}_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right]$, where $\widetilde{q}_{\frac{\alpha}{2}}$ and $\widetilde{q}_{1-\frac{\alpha}{2}}$ are the $\frac{\alpha}{2}^{th}$ and $1-\frac{\alpha}{2}^{th}$ quantiles of $\widetilde{S^1} = max\left(N^1; N^2\right)$ with $\begin{pmatrix} N^1 & N^2 \end{pmatrix}' \sim \mathcal{N}\left(0, \widehat{\Sigma}\right)$. Finally, if $B^0(M,m) < \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$, a CI for B_- is

$$CI^C = \left[\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M - \frac{q_{1-\frac{\alpha}{2}} \times \widehat{\sigma_2^2}}{\sqrt{n}}; \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{q_{1-\frac{\alpha}{2}} \times \widehat{\sigma_2^2}}{\sqrt{n}}\right].$$

But $B^0(M,m)$ and $\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$ are unknown, hence the need to find CI with $(1-\alpha)\%$ asymptotic coverage irrespective of their respective position. This is achieved by choosing CI^A when $\widehat{B^0(M,m)}$ is more than $\frac{ln(n)}{\sqrt{n}}$ above $\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M$, CI^B when $\widehat{B^0(M,m)}$ is less than $\frac{ln(n)}{\sqrt{n}}$ away from $\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M$, and CI^C when $\widehat{B^0(M,m)}$ is more than $\frac{ln(n)}{\sqrt{n}}$ below $\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M$[4]. The reason why this decision rule yields a CI with $(1-\alpha)\%$ asymptotic coverage uniformly in $B^0(M,m)$ and $\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$ is that since $\frac{1}{\sqrt{n}} = o\left(\frac{ln(n)}{\sqrt{n}}\right)$, the probability to pick the "wrong" CI converges to 0.

Proposition DID 5: CI for $\widehat{B_-}$ and $\widehat{B_+}$ with uniform asymptotic coverage

$$CI = CI^A \times 1_{\left\{\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}} < \widehat{B^0(M,m)}\right\}}$$

$$+CI^B \times 1_{\left\{\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M - \frac{ln(n)}{\sqrt{n}} \leq \widehat{B^0(M,m)} \leq \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}}\right\}}$$

$$+CI^C \times 1_{\left\{\widehat{B^0(M,m)} + \frac{ln(n)}{\sqrt{n}} < \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M\right\}}$$

is a CI for B_- with $(1-\alpha)\%$ asymptotic coverage uniformly in $B^0(M,m)$ and $\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$. A CI for B_+ with $(1-\alpha)\%$ asymptotic coverage uniformly in $B^0(m,M)$ and $\mathbb{E}(Y_{t_1,g_t}|\,D=1) - m$ can be constructed following the same steps.

[4] Instead of $ln(n)$, one can choose whatever sequence u_n such that $u_n \to +\infty$ and $\frac{u_n}{\sqrt{n}} \to 0$

Let us now consider B'_- and B'_+. As in Proposition DID 4, one can show that whatever the value of $\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)$, $\mathbb{E}(Y_{t_1,g_c}|\,D=1)-\mathbb{E}(Y_{t_0,g_c}|\,D=1)$, $min(B^1\ ;\ B^2)-\mathbb{E}(Y_{t_1,g_t}|\,D=1)-M$ and $max(B^1\ ;\ B^2)-\mathbb{E}(Y_{t_1,g_t}|\,D=1)-m$, $\widehat{B'_-}$ and $\widehat{B'_+}$ are $\sqrt{n}$-consistent, with standard normal limiting distributions when those four quantities are different from 0, and with non standard limiting distributions when one of them quantities is equal to 0. It is also possible to derive CI for B_- and B_+ with $(1-\alpha)\%$ asymptotic coverage irrespective of the value of those four unknown quantities. Because both B'_- and B'_+ are not differentiable at 3 points, careful analysis of their limiting distribution requires distinguishing 27 cases. Similarly, the construction of uniform CI for B'_- and B'_+ involves 27 auxiliary CI. Due to a concern for brevity, the two corresponding propositions are not presented here.

3 Application to the impact of varenicline on smoking cessation.

3.1 Data and methods

I use the data base of French smoking cessation clinics participating in the "Consultation Dépendance Tabagique" program (hereafter referred to as CDT). This program started in 2001 and led to the progressive implementation of smoking cessation services nationwide. 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.

Varenicline is a pharmacotherapy for smoking cessation support which was made available to these centers in February 2007. 59 services recorded at least one patient per year in 2006 and 2007 and followed at least 50% of their patients. The kernel density estimate of the rate of prescription of varenicline per center is shown in Figure 1. It is bimodal, with a first peak at very low rates of prescription, and a second smaller peak around 35-40%. In 15 services, less than 3% of all patients consulted have been prescribed varenicline during the year following its release. In 13 services, more than 20% of patients were prescribed varenicline. I exploit this to estimate the impact of varenicline on smoking cessation through a fuzzy DID identification strategy. The control group is made up

of patients registered by the 15 "below 3% prescription rate" services, whereas the treatment group consists of patients recorded by "above 20% prescription rate" centers. Period 0 goes from February 2006 to January 2007, and period 1 from February 2007 to January 2008.

[Figure 1 inserted here]

8 581 patients consulted those 28 services over period 0 and 1. Because many patients never came back for follow-up visits, there are only 5 299 patients (62% of the initial sample) for whom follow-up CO measures are available. I exclude patients for whom no such measures are available from the analysis. Among remaining patients, which I refer to as the included sample, I compute a point prevalence abstinence rate, that is to say the share of patients whose last follow-up CO determination was inferior or equal to 5 parts per million (ppm).

3.2 Results

In Table 1, I provide descriptive statistics on patients per group of centers and per period of time. Patients consulted in those cessation services are middle-aged, rather educated and the majority of them are employed. They are very heavy smokers since they smoke more than 21.6 cigarettes per day on average, which corresponds to the 90th percentile in the French distribution of smokers (Beck et al. [2007]). 17% of them suffer from chronic obstructive pulmonary diseases (COPD) and more than 30% suffer from tobacco related diseases (lung cancer, COPD...). They have therefore been classified as "hardcore" addicts in the medical literature.

[Table 1 inserted here]

In period 0, the prescription rate of varenicline was equal to 0% in control centers and to 0.01% in treatment centers (varenicline was prescribed to 6 patients recorded in the last week of January 2007, that is to say right before the release of varenicline). In period 1, it was equal to 1.6% in control centers and to 38.2% in treatment centers. This sharp rise in varenicline prescription in treatment centers entailed a strong decrease in the prescription of other treatments such as nicotine patch. Finally, from period 0 to 1, the point prevalence abstinence rate increased (from 53.7% to 56.9%) in

treatment centers, whereas it decreased (from 46.6% to 41.6%) in control centers. Among treatment patients prescribed varenicline in period 0, abstinence rate was equal to 50.0%. Among control patients prescribed varenicline in period 1, abstinence rate was equal to 58.3%. Applying the formulas of section 2, I compute that $\widehat{B_-} = 19.1\%$ (P-value = 0.008) and $\widehat{B_+} = 24.5\%$ (P-value = 0.001). Finally, $\frac{\widehat{DID}}{\widehat{DID^P}} = 22.7\%$ (P-value=0.003).

$\widehat{B^0(M,m)}$ is higher than $\widehat{\mathbb{E}}(Y_{t_1,g_t}|\, D=1) - 1 - \frac{ln(5299)}{\sqrt{5299}}$, and $\widehat{B^0(m,M)} + \frac{ln(5299)}{\sqrt{5299}}$ is lower than $\widehat{\mathbb{E}}(Y_{t_1,g_t}|\, D=1)$. Consequently, the CI to be used for B_- and B_+ are CI^A (see Proposition DID 5). Then, using Proposition DID 3, I construct 3 CI for ATT_{t_1,g_t}: $CI^1 = \left[LB_{95}^{\frac{DID}{DID^P}}; UB_{95}^{\frac{DID}{DID^P}}\right] = [7.8\%; 37.5\%]$, $CI^2 = \left[LB_{95}^{B_-}; UB_{95}^{B_+}\right] = [5.0\% : 38.6\%]$, $CI^3 = \left[LB_{90}^{B_-}; UB_{90}^{B_+}\right] = [7.3\% : 36.3\%]$. The uniform convergence issue mentioned in Imbens and Manski [2004] shows up here since CI^3 is shorter than CI^1. But here even CI^2 is enough to infer the sign of ATT_{t_1,g_t}.

Point identification of ATT_{t_1,g_t} relies on a strong constant treatment effect assumption whereas identification of $[B_-;\, B_+]$ is obtained under much weaker assumptions. Moreover, inference on $\widehat{B_-}$ is sufficient to draw inference on the sign of ATT_{t_1,g_t}. Finally, even using CI^2, inference on $\widehat{B_-}$ and $\widehat{B_+}$ yields a 95% CI for ATT_{t_1,g_t} which is only slightly broader than the one obtained when drawing inference on $\frac{\widehat{DID}}{\widehat{DID^P}}$. Therefore, one might consider that here, the parameters which achieve the best trade-off between the accuracy of the information they deliver and the identifying assumptions on which they rely are $\widehat{B_-}$ and $\widehat{B_+}$ and not $\frac{\widehat{DID}}{\widehat{DID^P}}$.

3.3 Robustness checks

The only substantial assumption which is needed to identify $[B_-;\, B_+]$ is the common trend assumption. To "test" it, I use the fact that I have several years of data available and I compute placebo DID from 2003 to 2008. They are displayed in the top panel of Table 2 along with their P-values. Only the 2006-2007 DID is significant, which gives some credit to the common trend assumption. I also compute 2006-2007 placebo DID on 9 patients' observable characteristics. They are also displayed in Table 2. This test is less conclusive since 2 DID out of 9 are significantly different from 0 at the 95% level. For instance, daily cigarettes smoked increased by 1.45 more among treatment centers' than among control centers' patients from 2006 to 2007. Similarly, the percentage of patients suffering from

COPD increased by 4.4 percentage points more in treatment than in control services. This might cast some doubt on the validity of the common trend assumption. However, the P-value obtained on the DID of percentage of successful quits from 2006 to 2007 is still the lowest by far out of the 14 DID computed in Table 2. Moreover, high number of daily cigarettes smoked and COPD are predictors of unsuccessful quits. Since my fuzzy DID identification strategy does not correct for diverging trends on those variables, it might underestimate the true effect of varenicline.

Attrition seems orthogonal to the interaction of period 1 and treatment centers, since the DID computed on the percentage of patients included is low and insignificant (+2.2%, P-value = 0.30). Therefore, estimates do not seem contaminated by attrition bias. However, the delay between patients' first visit and the last CO measure available increased more in treatment than in control clinics. This is very likely to be because varenicline being a newly released drug with more severe secondary effects than nicotine patch, doctors put more effort in following their patients over a longer period of time to ensure they tolerate it well. Anyway, since smoking cessation is known to be a "duration" type of process, observing patients over a longer period of time in period 1 than in period 0 in treatment clinics can only bias downward my estimate.

Finally, one might worry about the arbitrariness of the definition of my treatment and control groups which is not based on some objective characteristic of cessation services. I investigate the sensitivity of the results to the 3%-20% rule as a robustness check. I ran the same analysis with 9 different pairs of thresholds and always got $\widehat{B_-} \geq 0$ with 6 P-values lower than 0.05. The results of this last robustness check are displayed in the bottom panel of Table 2.

[Table 2 inserted here]

Part II

Fuzzy Change in Change

4 Identification

4.1 Identification under CIC assumptions

I place myself in the pooled cross-section case: each individual is observed only at one period. Let $T \in \{t_0; t_1\}$ denote time and $G \in \{g_c; g_t\}$ denote treatment (g_t) and control (g_c) groups. I assume that treatment status is binary and is denoted by an indicator D. Throughout the paper it is implicitly assumed that the stable unit treatment value assumption holds. Under this assumption I define $Y(1)$ and $Y(0)$ as the potential outcomes of an individual with and without the treatment. Only the actual outcome $Y = Y(1) \times D + Y(0) \times (1 - D)$ is observed. The treatment effect is $Y(1) - Y(0)$. $X \sim Y$ means that X and Y have the same probability distribution. $\mathbb{X}$ is the support of X. To alleviate the notational burden, I introduce several shorthands following Athey and Imbens [2006]:

$Y_{t,g}(k) \sim Y(k) \,|\, T = t,\, G = g \;\; \forall (t, g, k) \in \{t_0; t_1\} \times \{g_c; g_t\} \times \{0; 1\}$

$Y_{t,g} \sim Y \,|\, T = t,\, G = g \;\; \forall (t, g) \in \{t_0; t_1\} \times \{g_c; g_t\}$

$D_{t,g} \sim D \,|\, T = t,\, G = g \;\; \forall (t, g) \in \{t_0; t_1\} \times \{g_c; g_t\}$

Let F_X and $F_{X|Y}$ denote respectively the cumulative distribution function (cdf) of a random variable X and the cdf of X conditional on Y. Let F_X^{-1}denote the inverse cdf of X. The standard definition of $F_X^{-1}(q)$ is $\forall q \in [0; 1]\,,\; F_X^{-1}(q) = inf\,\{x \in \mathbb{X} / F_X(x) \geq q\}$.

Athey and Imbens take the following assumptions:

Assumption CIC 1: Model

$\forall k \in \{0; 1\}\,,\; Y(k) = h_k(U, T)$ where U represents individuals unobserved characteristics.

Assumption CIC 2: Strict monotonicity

$\forall k \in \{0; 1\}\,,\; \forall t \in \{t_0; t_1\}\,,\; h_k(u, t)$ is strictly increasing in u.

Assumption CIC 3: Time invariance within groups

$U \perp\!\!\!\perp T \,|\, G$

Assumption CIC 4: Support

$\mathbb{U}|G = g_t \subseteq \mathbb{U}|G = g_c$

Assumption CIC 5: Perfect compliance

$D = 1 \Longleftrightarrow T \times G = \{t_1; g_t\}$

Under those assumptions, they prove the following result:

Theorem 3.1, Athey and Imbens [2006]:

i) Under Assumption CIC 1 to Assumption CIC 4, if U is either continuous or discrete, then

$$F_{Y_{t_1,g_t}(0)}(y) = F_{Y_{t_0,g_t}(0)}\left(F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right)\right).$$

Under Assumption CIC 1 to Assumption CIC 5, if U is either continuous or discrete,

ii) $F_{Y_{t_1,g_t}(0)}(y)$ *is identified:*

$$F_{Y_{t_1,g_t}(0)}(y) = F_{Y_{t_0,g_t}}\left(F^{-1}_{Y_{t_0,g_c}}\left(F_{Y_{t_1,g_c}}(y)\right)\right).$$

iii) $\mathbb{E}\left(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0)\right)$ *is identified:*

$$\mathbb{E}\left(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0)\right) = \mathbb{E}\left(Y_{t_1,g_t}\right) - \mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}}\left(F_{Y_{t_0,g_c}}(Y_{t_0,g_t})\right)\right).$$

iv) $\forall q \in [0;1]\,,\ F^{-1}_{Y_{t_1,g_t}(1)}(q) - F^{-1}_{Y_{t_1,g_t}(0)}(q)$ *is identified:*

$$F^{-1}_{Y_{t_1,g_t}(1)}(q) - F^{-1}_{Y_{t_1,g_t}(0)}(q) = F^{-1}_{Y_{t_1,g_t}}(q) - F^{-1}_{Y_{t_1,g_c}}\left(F_{Y_{t_0,g_c}}\left(F^{-1}_{Y_{t_0,g_t}}(q)\right)\right).$$

Under Assumption CIC 5, the cdf of $Y(0)$ in the period 1 $\times$ test group cell is not observed because all observations of this cell are treated. But Theorem 3.1 states that it can be recovered from three observable functions ($F_{Y_{t_0,g_t}}(.)$, $F^{-1}_{Y_{t_0,g_c}}(.)$ and $F_{Y_{t_1,g_c}}(.)$), making it possible to compute the average treatment effect as well as quantile treatment effects within this cell. The intuition of this

theorem is as follows. Take an observation in the period 0 $\times$ control group cell. Denote y its observed outcome, q the quantile corresponding to y in the distribution of this cell, and u its realization of U. By Assumption CIC 1 and Assumption CIC 5, $y = h_0(u, t_0)$. Now consider the observation at the qth quantile of the distribution of Y in the period 1 $\times$ control group cell, and denote y^* its observed outcome. Since the distribution of U is time invariant within group (Assumption CIC 3), those two observations must have the same u. Therefore, $y^* = h_0(u, t_1)$, which means that the qth quantile of the period 1 $\times$ control group cell identifies the period 1 $Y(0)$ of an observation with unobserved heterogeneity u. Now consider an observation in the period 0 $\times$ test group cell with observed outcome y. Since $Y(0)$ only depends on time and unobserved heterogeneity (Assumption CIC 1) and since $h_0(., t_0)$ is invertible (Assumption CIC 2), it must have the same unobserved heterogeneity u than the period 0 $\times$ control group observation. Thus for an observation of the test group with observed outcome y in period 0 and unobserved heterogeneity u, it is possible to recover its period 1 $Y(0)$: it is merely equal to y^*. Therefore, to recover the whole counterfactual distribution of $Y(0)$ in the period 1 $\times$ test group cell, it suffices to translate the whole distribution of Y in the period 0 $\times$ test group cell from y to the corresponding y^* for each value of y.

Part i) of their Theorem does not rely on Assumption CIC 5 whereas part ii) does, hence my non-identification Lemma:

Lemma CIC 1: Non-identification.

Under Assumption CIC 1 to Assumption CIC 4, if U is either continuous or discrete, $F_{Y_{t_1,g_t}(0)}(y)$ is not necessarily equal to $F_{Y_{t_0,g_t}}\left(F^{-1}_{Y_{t_0,g_c}}\left(F_{Y_{t_1,g_c}}(y)\right)\right)$.

Assumption CIC 1 to Assumption CIC 4 are sufficient for part i) of Athey and Imbens's Theorem to hold. It states that the cdf of $Y_{t_1,g_t}(0)$ can be recovered from the cdf of $Y_{t_0,g_t}(0)$ and $Y_{t_1,g_c}(0)$ and from the inverse cdf of $Y_{t_0,g_c}(0)$. However, when compliance is imperfect, this might prove useless since those functions may not be fully observed: only $F_{Y_{t_0,g_t}(0)|D=0}$, $F_{Y_{t_1,g_c}(0)|D=0}$ and $F^{-1}_{Y_{t_0,g_c}(0)|D=0}$ are observed. This is because two steps in the intuition of Theorem 3.1 collapse. First, two control and test group observations with same y in period 0 might no longer have the same u, because one might be untreated which means that $y = h_0(u_0, t_0)$, whereas the other is treated meaning that $y = h_1(u_1, t_0)$. $h_0(u_0, t_0) = h_1(u_1, t_0)$ does not imply $u_0 = u_1$. Then, two observations at the qth

quantile of the distribution of Y in the control group in period 0 and in period 1 do not necessarily have the same u. Indeed, if the treatment rate changed between the two periods in the control group this could disrupt the ranking of observations. Assume for instance that Y are wages, D is secondary education completion. If the share of observations completing secondary education increased in the control group, then a period 1 untreated observation with the same rank in the wage distribution as a period 0 untreated observation has probably a higher rank in the unobserved ability distribution. Indeed, despite the fact that more of its counterparts are educated and therefore have an advantage on the labor market, it has the same rank in the wage distribution. Despite this general non identification result I now state an identification result. It holds in one important special case which is when there are no always takers, that is to say when $\mathbb{P}(D_{t_0,g_t}=1)=\mathbb{P}(D_{t_1,g_c}=1)=\mathbb{P}(D_{t_0,g_c}=1)=0$.

Theorem CIC 1: Point identification in the no always takers special case

Let $G(y)=\frac{F_{Y_{t_0,g_t}}\left(F^{-1}_{Y_{t_0,g_c}}\left(F_{Y_{t_1,g_c}}(y)\right)\right)-F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=1)}$. *In the no always takers special case, under Assumption CIC 1 to Assumption CIC 4:*

i) $F_{Y_{t_1,g_t}(0)|D=1}$ *is identified:*

$$F_{Y_{t_1,g_t}(0)|D=1}(y)=G(y)$$

ii) $\mathbb{E}\left(Y_{t_1,g_t}(1)-Y_{t_1,g_t}(0)|D=1\right)$ *is identified:*

$$\tau^{CIC}=\mathbb{E}\left(Y_{t_1,g_t}(1)-Y_{t_1,g_t}(0)|D=1\right)=\frac{\mathbb{E}\left(Y_{t_1,g_t}\right)-\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}}\left(F_{Y_{t_0,g_c}}(Y_{t_0,g_t})\right)\right)}{\mathbb{P}(D_{t_1,g_t}=1)}$$

iii) $\forall q\in[0;1]$, $F^{-1}_{Y_{t_1,g_t}(1)|D=1}(q)-F^{-1}_{Y_{t_1,g_t}(0)|D=1}(q)$ *is identified:*

$$\tau^{CIC}_q=F^{-1}_{Y_{t_1,g_t}(1)|D=1}(q)-F^{-1}_{Y_{t_1,g_t}(0)|D=1}(q)=F^{-1}_{Y_{t_1,g_t}|D=1}(q)-inf\left\{y\in\mathbb{Y}_{t_1,g_c}/G(y)\geq q\right\}$$

Theorem CIC 1 holds because in the no always takers special case, $F_{Y_{t_0,g_t}}=F_{Y_{t_0,g_t}(0)}$, $F_{Y_{t_1,g_c}}=F_{Y_{t_1,g_c}(0)}$ and $F^{-1}_{Y_{t_0,g_c}}=F^{-1}_{Y_{t_0,g_c}(0)}$. Consequently, all the functions required to recover the cdf of $Y_{t_1,g_t}(0)$ from Athey and Imbens's Theorem 3.1 are fully observed. Indeed, in the previous lemma, the fact that some period 0 observations could potentially be treated made it impossible to match treatment and control observations on their Y. Moreover, a change in the treatment rate in the control group

could have disrupted the ranking of control observations between the two periods. But since in the no always takers case, no one gets treated in period 0 and in the two control cells, the CIC double matching process is sufficient to reproduce $F_{Y_{t_1,g_t}(0)}(y)$, the distribution of $Y(0)$ within the period 1 $\times$ test group cell. Since $F_{Y_{t_1,g_t}(0)|D=0}(y)$ is observed, subtracting it to $F_{Y_{t_1,g_t}(0)}(y)$ allows recovering $F_{Y_{t_1,g_t}(0)|D=1}(y)$. Consequently, since $F_{Y_{t_1,g_t}(1)|D=1}(y)$ is observed, it is possible to identify the average treatment effect on the treated and quantile treatment effects.

Even though when there are always takers CIC assumptions are not sufficient for identification, I show now that they are still sufficient to place bounds on $F_{Y_{t_1,g_t}(0)|D=1}(y)$ and therefore on τ^{CIC} and τ_q^{CIC} (due to a concern for brevity I give explicit formulas for bounds on $F_{Y_{t_1,g_t}(0)|D=1}(y)$ only). Those bounds will be tight in applications with small shares of observations treated within the three remaining cells.

Proposition CIC 1: Partial Identification in the CIC model.

Under Assumption CIC 1 to Assumption CIC 4,

$$B_-^{CIC} \leq F_{Y_{t_1,g_t}(0)|D=1}(y) \leq B_+^{CIC}$$

with

$$B_-^{CIC} = \frac{F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)-F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=1)}$$

and

$$B_+^{CIC} = \frac{F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)+\mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)+\mathbb{P}(D_{t_0,g_t}=1)-F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=1)}.$$

The bounds are obtained as follows. Athey and Imbens's Theorem 3.1 states that the cdf of $Y_{t_1,g_t}(0)$ can be recovered from the cdf of $Y_{t_0,g_t}(0)$ and $Y_{t_1,g_c}(0)$, and from the inverse cdf of $Y_{t_0,g_c}(0)$. When there are always takers, those three functions are not fully observed: only $F_{Y_{t_0,g_t}(0)|D=0}$, $F_{Y_{t_1,g_c}(0)|D=0}$ and $F^{-1}_{Y_{t_0,g_c}(0)|D=0}$ are observed. But the cdf of $Y_{t_1,g_t}(0)$, can be bounded placing bounds on $F_{Y_{t_0,g_t}(0)|D=1}$, $F_{Y_{t_1,g_c}(0)|D=1}$ and $F^{-1}_{Y_{t_0,g_c}(0)|D=1}$. The length of $\left[B_-^{CIC};B_+^{CIC}\right]$ is increasing with $\mathbb{P}(D_{t_0,g_c}=1)$, $\mathbb{P}(D_{t_1,g_c}=1)$, and $\mathbb{P}(D_{t_0,g_t}=1)$. It is decreasing with $\mathbb{P}(D_{t_1,g_t}=1)$. This is is very similar to the result obtained with fuzzy DID.

4.2 Identification in the IV Change in Change model

To deal with applications with potentially large share of always takers, I give a second identification result. It requires that the share of treated observations remained approximately constant between period 0 and 1 in the control group. Under this last assumption, one can indeed assume that unobserved heterogeneity of treated and untreated observations in the control group did not change between period 0 and 1 since within that group selection into treatment seems to have remained constant over time. This result also requires assuming that within the test group, the share of treated observations increased because of a policy change, or because of supplementary incentives for treatment given to that group only. This amounts to introducing a binary instrument for treatment Z which is such that $Z = 1 \iff T \times G = \{t_1; g_t\}$. I also introduce the two corresponding potential treatment statuses, $D(0)$ and $D(1)$, which stand for treatment without and with the policy. Observed treatment is $D = Z \times D(1) + (1 - Z) \times D(0)$.

I introduce a slightly more restrictive set of assumptions than Athey and Imbens's. In particular, I replace Assumption CIC 3 by

Assumption CIC 3': Time invariance of U and $D(0)$ within groups

$(U,\, D(0)) \perp\!\!\!\perp T \,|\, G$

Note that Assumption CIC 3' is equivalent to $D(0) \perp\!\!\!\perp T \,|\, G$ and $U \perp\!\!\!\perp T \,|\, G,\, D(0)$. The first assumption means that selection into treatment would have remained time invariant if no policy had been implemented. It has one testable implication which is that within the control group, the share of treated individuals should have remained constant between period 0 and 1. The second assumption states that in the control group, treated observations should remain "the same" across time, meaning that the distribution of their unobserved heterogeneity should not change. This will be all the more credible that observable characteristics of treated observations do not change much over time in the control group.

I also replace Assumption CIC 4 by Assumption CIC 4':

Assumption CIC 4': Support

$\mathbb{U}|G = g_t, D(0) = 1 \subseteq \mathbb{U}|G = g_c, D(0) = 1$

$\mathbb{U}|G = g_t, D(0) = 0 \subseteq \mathbb{U}|G = g_c, D(0) = 0$

Finally I introduce a monotonicity assumption:

Assumption CIC 6: Monotonicity

$D(1) \geq D(0)$

Assumption CIC 6 means that there should be no defiers, that is to say observations who get treated without the policy and do not get treated with it.

Theorem CIC 2: Identification in the IV-CIC model

Let $H^1(y) = \frac{F_{Y_{t_1,g_t}|D=1}(y)\times\mathbb{P}(D_{t_1,g_t}=1) - F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(F_{Y_{t_1,g_c}|D=1}(y)\right)\right)\times\mathbb{P}(D_{t_0,g_t}=1)}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)}$

and $H^0(y) = \frac{F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(F_{Y_{t_1,g_c}|D=0}(y)\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0) - F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_0,g_t}=0)-\mathbb{P}(D_{t_1,g_t}=0)}$.

Under Assumption CIC 1, Assumption CIC 2, Assumption CIC 3', Assumption CIC 4' and Assumption CIC 6:

i) $F_{Y_{t_1,g_t}(1)|D(1)>D(0)}(y)$ *and* $F_{Y_{t_1,g_t}(0)|D(1)>D(0)}(y)$ *are identified:*

$$F_{Y_{t_1,g_t}(1)|D(1)>D(0)}(y) = H^1(y)$$

and

$$F_{Y_{t_1,g_t}(0)|D(1)>D(0)}(y) = H^0(y).$$

ii) $\mathbb{E}\left(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0)|D(1) > D(0)\right)$ *is identified:*

$$\tau^{IV-CIC} = \mathbb{E}\left(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0)|D(1) > D(0)\right) =$$

$$\frac{\mathbb{E}\left(Y_{t_1,g_t}|D=1\right)\times\mathbb{P}(D_{t_1,g_t}=1) - \mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=1}\left(F_{Y_{t_0,g_c}|D=1}(Y_{t_0,g_t})\right)|D=1\right)\times\mathbb{P}(D_{t_0,g_t}=1)}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)}$$

$$-\frac{\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=0}\left(F_{Y_{t_0,g_c}|D=0}(Y_{t_0,g_t})\right)|D=0\right)\times\mathbb{P}(D_{t_0,g_t}=0) - \mathbb{E}\left(Y_{t_1,g_t}|D=0\right)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_0,g_t}=0)-\mathbb{P}(D_{t_1,g_t}=0)}.$$

iii) $\forall q \in [0;1]$, $F^{-1}_{Y_{t_1,g_t}(1)|D(1)>D(0)}(q) - F^{-1}_{Y_{t_1,g_t}(0)|D(1)>D(0)}(q)$ *is identified:*

$$\tau^{IV-CIC}_q = F^{-1}_{Y_{t_1,g_t}(1)|D(1)>D(0)}(q) - F^{-1}_{Y_{t_1,g_t}(0)|D(1)>D(0)}(q)$$

$$= inf\left\{y \in \mathbb{Y}_{t_1,g_t} | D = 1/H^1(y) \geq q\right\} - inf\left\{y \in \mathbb{Y}_{t_1,g_c} | D = 0/H^0(y) \geq q\right\}.$$

Theorem CIC 2 states that it is possible to recover the distribution of both $Y(1)$ and $Y(0)$ among compliers of the period 1 $\times$ test group cell from observable distributions. This makes it possible to identify a LATE, as well as quantile treatment effects inside this population. To recover the distribution of $Y(1)$, I consider the distribution of Y among all treated observations of this cell. Since those observations include both compliers and always takers, I need to "withdraw" from it the distribution of Y among always takers. But this last distribution is not observed (because I cannot distinguish compliers from always takers), hence the need to reconstruct it. On that purpose, I use distributions of Y among always takers in the three remaining cells. Those three distributions are observed since by definition treated observations in the three remaining cells must be always takers. Then, identification relies on a similar double-matching process than Athey and Imbens's theorem 3.1 except that it requires considering period $\times$ group cells among treated observations only. Take an always taker in the period 0 $\times$ control group $\times$ $D = 1$ cell. Denote y its observed outcome, q the quantile corresponding to y in the distribution of this cell, and u the realization of U for this observation. By Assumption CIC 1, $y = h_1(u, t_0)$. Now consider the observation at the qth quantile of the period 1 $\times$ control group $\times$ $D = 1$ cell, and denote y^* its observed outcome. Since the distribution of U is time invariant within group $\times$ D cells (Assumption CIC 4'), those two always takers must have the same u. Therefore, $y^* = h_1(u, t_1)$, which means that the qth quantile of the period 1 $\times$ control group $\times$ $D = 1$ cell identifies the period 1 $Y(1)$ of an always taker with unobserved heterogeneity u. Now consider an always taker in the period 0 $\times$ test group $\times$ $D = 1$ cell with observed outcome y. Since Y only depends on time, treatment status and unobserved heterogeneity (Assumption CIC 1) and since $h_1(., t_0)$ is invertible (Assumption CIC 2), she must have the same unobserved heterogeneity u than the first always taker. Thus for an always taker of the period 0 $\times$ test group $\times D = 1$ cell with observed outcome y, it is possible to recover its period 1 $Y(1)$: it is merely equal to y^*. Therefore, to recover the whole distribution of $Y(1)$ in period 1 among test group always takers, it suffices to translate the whole distribution of the period 0 $\times$ test group cell $\times$ $D = 1$ from y to the corresponding y^* for each value of y.

Identification of the distribution of $Y(0)$ among compliers of the period 1 $\times$ test group cell is

obtained as follows. The distribution of $Y(0)$ in period 1 among observations which were untreated in period 0 in the treatment group is reconstructed through a double-matching process across untreated observations. But those observations include both compliers and never takers, hence the need to "withdraw" the distribution of $Y(0)$ among never takers. This is achieved easily since never takers in this cell are all untreated observations so that this distribution is observed.

Finally, to deal with applications with potentially large share of always takers and where the treatment rate also increases in the control group (see for instance Duflo [2001]), I give a second partial identification result. I show that in such applications partial identification of quantile treatment effects can be obtained through a strengthening of CIC-IV assumptions. First, the instrument Z should now take three values, reflecting the fact that in period 1 the two groups received supplementary incentives for treatment, but that the increase in incentives for treatment was stronger in the treatment group. Consequently, $Z = 0 \Longleftrightarrow T = t_0$, $Z = 1 \Longleftrightarrow T = t_1$ and $G = g_c$ and $Z = 2 \Longleftrightarrow T = t_1$ and $G = g_t$. There are now three potential treatment statuses: $D(0)$, $D(1)$ and $D(2)$, hence the need to modify Assumption CIC 6 into

Assumption CIC 6': Monotonicity

$D(2) \geq D(1) \geq D(0)$.

Moreover, I also need to take a common trend assumption on the treatment rate, which states that if the treatment group had also received a low amount of supplementary incentives for treatment in period 1, the share of observations treated would have followed the same evolution in the treatment and in the control group between the two periods:

Assumption CIC 7: Common trend on treatment rate

$\mathbb{P}(D_{t_1,g_t}(1) = 1) - \mathbb{P}(D_{t_0,g_t}(0) = 1) = \mathbb{P}(D_{t_1,g_c}(1) = 1) - \mathbb{P}(D_{t_0,g_c}(0) = 1)$.

Under this modified set of assumptions, the distributions of $Y(1)$ and $Y(0)$ within a specific population of compliers, i.e. those who get treated if and only if they receive "strong" incentives for treatment, are partially identified. Consequently, the average treatment effect and quantile treatment effects are also partially identified but due to a concern for brevity I give explicit bounds for distributions only.

Proposition CIC 2: Partial Identification in the IV-CIC model

Under Assumption CIC 1, Assumption CIC 2, Assumption CIC 3', Assumption CIC 4', Assumption

CIC 6' and Assumption CIC 7, $F_{Y_{t_1,g_t}(1)|D(2)>D(1)}(y)$ and $F_{Y_{t_1,g_t}(0)|D(2)>D(1)}(y)$ are partially identified:

$$B_-^{1,CIC-IV} \leq F_{Y_{t_1,g_t}(1)|D(2)>D(1)}(y) \leq B_+^{1,CIC-IV}$$

and

$$B_-^{0,CIC-IV} \leq F_{Y_{t_1,g_t}(0)|D(2)>D(1)}(y) \leq B_+^{0,CIC-IV}$$

with

$$B_-^{1,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=1}(y)\times\mathbb{P}(D_{t_1,g_t}=1) - F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(\frac{F_{Y_{t_1,g_c}|D=1}(y)\times\mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=1)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=1) - \left[\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right]}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)-\left[\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right]},$$

$$B_+^{1,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=1}(y)\times\mathbb{P}(D_{t_1,g_t}=1) - F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(\frac{F_{Y_{t_1,g_c}|D=1}(y)\times\mathbb{P}(D_{t_1,g_c}=1) - \left(\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right)}{\mathbb{P}(D_{t_0,g_c}=1)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=1)}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)-\left[\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right]},$$

$$B_-^{0,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0) - F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(F_{Y_{t_1,g_c}|D=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_c}=0)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0) - \left[\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right]}{\mathbb{P}(D_{t_1,g_t}=0)-\mathbb{P}(D_{t_0,g_t}=0)-\left[\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right]}$$

and

$$B_+^{0,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0) - F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0) - \left(\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=0)-\mathbb{P}(D_{t_0,g_t}=0)-\left[\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right]}.$$

The bounds are obtained as follows. Under common trend, the DID on treatment rate ($\mathbb{P}(D_{t_1,g_t} = 1) - \mathbb{P}(D_{t_0,g_t} = 1) - [\mathbb{P}(D_{t_1,g_c} = 1) - \mathbb{P}(D_{t_0,g_c} = 1)]$) identifies the size of a population of compliers. Because the share of treated observations increases in the control group as well, it is no longer possible to assume that treated (resp. untreated) observations are the same in the control group in period 0 and 1, i.e. that their distribution of U is the same. Period 0 observations can not be matched to their rank counterpart in period 1, because the fact that some "compliers" got treated in period 1 might have disrupted the distribution of U. But thanks to monotonicity, the size of this population of compliers is known: it is equal to the change in the treatment rate between those two periods, and therefore the maximum and minimum impact of those compliers on the rank of observations is known as well, hence the partial identification result. Bounds will be tight in applications where the change in the treatment rate in the control group across the two periods is small.

Summary and conclusions

This paper provides new identification results applying to fuzzy DID and fuzzy CIC. Most of the fuzzy DID results hold under a common trend assumption on the outcome only, whereas the IV result commonly invoked in such settings holds under two common trends (on the outcome and on the treatment) and a monotonicity assumption. This single common trend assumption is sufficient to identify an ATT when there are no always takers, or at least its sign when there are "few" of them. When the shares of always takers are "large", supplementary assumptions must be taken. For instance, identification of an ATT can be obtained under the assumption that ATT do not vary across time and group. The milder assumption that the ATT in the control group did not change from period 0 to 1 substantially improves partial identification. This last result is of particular interest in applications considering the extension of a policy because in such situations it is likely to yield a narrow identification region. I present an application in which the bounds I derive allow drawing inference on the sign of an ATT. This is because in this example, there are few always takers. Had there been more of them, the identification region would have been too large to infer the sign of the ATT. Consequently, in a fuzzy DID, common trend on Y is sufficient to obtain accurate information on the ATT when there are few always takers, even if there are many never takers.

Similarly, in a fuzzy CIC, assumptions of the standard CIC model are sufficient for identification when there are no always takers. When there are always takers, they are no longer sufficient. However, it is possible to recover identification in applications with always takers but where the share of treated observations remained stable between the two periods in the control group, even though this is at the expense of slightly stronger assumptions.

References

[1] Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263, April 2003.

[2] Alberto Abadie. Semiparametric difference-in-differences estimators. *Review of Economic Studies*, 72(1):1–19, 01 2005.

[3] Alberto Abadie, Joshua Angrist, and Guido Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117, January 2002.

[4] Orley Ashenfelter and David Card. Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67(4):648–60, November 1985.

[5] Susan Athey and Guido W. Imbens. Identification and inference in nonlinear difference-in-differences models. NBER Technical Working Papers 0280, National Bureau of Economic Research, Inc, September 2002.

[6] Susan Athey and Guido W. Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 03 2006.

[7] Laurent Bach. Are small-and-medium-sized firms really credit constrained ? evidence from a french targeted credit programme. Working papers, Centre de Recherche en Economie et Statistique.

[8] Erich Battistin and Enrico Rettore. Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. *Journal of Econometrics*, 142(2):715–730, February 2008.

[9] François Beck and Philippe Guilbert. Baromètre santé [french health barometer]. Technical report, INPES, 2007.

[10] Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, February 2004.

[11] Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261, 01 2005.

[12] Small D and Tan Z. A stochastic monotonicity assumption for the instrumental variables method. Technical report, Department of Statistics, Wharton School, University of Pennsylvania., 2007.

[13] Clément de Chaisemartin. Instrumented differences in differences. Working paper, 2011.

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

[15] William N. Evans and Jeanne S. Ringel. Can higher cigarette taxes improve birth outcomes? *Journal of Public Economics*, 72(1):135–154, April 1999.

[16] Erica Field. Property rights and investment in urban slums. *Journal of the European Economic Association*, 3(2-3):279–290, 04/05 2005.

[17] Markus Frölich and Blaise Melly. Unconditional quantile treatment effects under endogeneity. IZA Discussion Papers 3288, Institute for the Study of Labor (IZA), January 2008.

[18] Joel L Horowitz and Charles F Manski. Identification and robustness with contaminated and corrupted data. *Econometrica*, 63(2):281–302, March 1995.

[19] Guido W Imbens and Joshua D Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–75, March 1994.

[20] Guido W. Imbens and Charles F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, November 2004.

[21] David S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102, 07 2009.

[22] Charles F Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2):319–23, May 1990.

[23] Bruce D Meyer, W Kip Viscusi, and David L Durbin. Workers' compensation and injury duration: Evidence from a natural experiment. *American Economic Review*, 85(3):322–40, June 1995.

[24] James M. Poterba, Steven F. Venti, and David A. Wise. Do 401(k) contributions crowd out other personal saving? *Journal of Public Economics*, 58(1):1–32, September 1995.

[25] Jorg Stoye. More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315, 07 2009.

Tables

Table 1: Descriptive Statistics

	Whole sample	Test Centers			Control Centers		
		2006	2007	P-value	2006	2007	P-value
Patients' characteristics							
% males	48.8%	47.9%	47.9%	0.98	48.5%	50.4%	0.30
Age	44.1	44.6	43.7	0.08	44.0	44.3	0.52
% employed	67.3%	65.3%	68.3%	0.11	65.3%	69.8%	0.01
% with no degree	17.0%	19.2%	21.0%	0.25	14.2%	14.1%	0.98
Daily cigarettes smoked	21.6	21.7	21.93	0.60	22.1	20.9	<0.01
FTND	5.9	5.8	5.8	0.29	6.0	5.9	0.11
% with AHAD>=11	39.8%	40.3%	39.1%	0.54	42.2%	37.7%	0.01
% with DHAD>=11	11.9%	13.1%	11.7%	0.29	11.6%	11.2%	0.72
% with chronic obstructive pulmonary diseases	16.7%	16.2%	18.1%	0.19	17.5%	15.1%	0.09
Treatment prescribed							
% prescribed nicotine patch	53.4%	75.0%	45.5%	<0.001	45.9%	49.7%	0.05
% prescribed varenicline	10.0%	0.01%	38.2%	<0.001	0%	1.6%	<0.001
Cessation Outcome							
Number of days between the first visit and the last CO measure	86.7	89.3	96.7	0.05	84.8	77.6	0.03
% of successful quits	49.3%	53.7%	56.9%	0.11	46.6%	41.6%	<0.01
N	5 299	1 195	1 303		1 300	1 501	

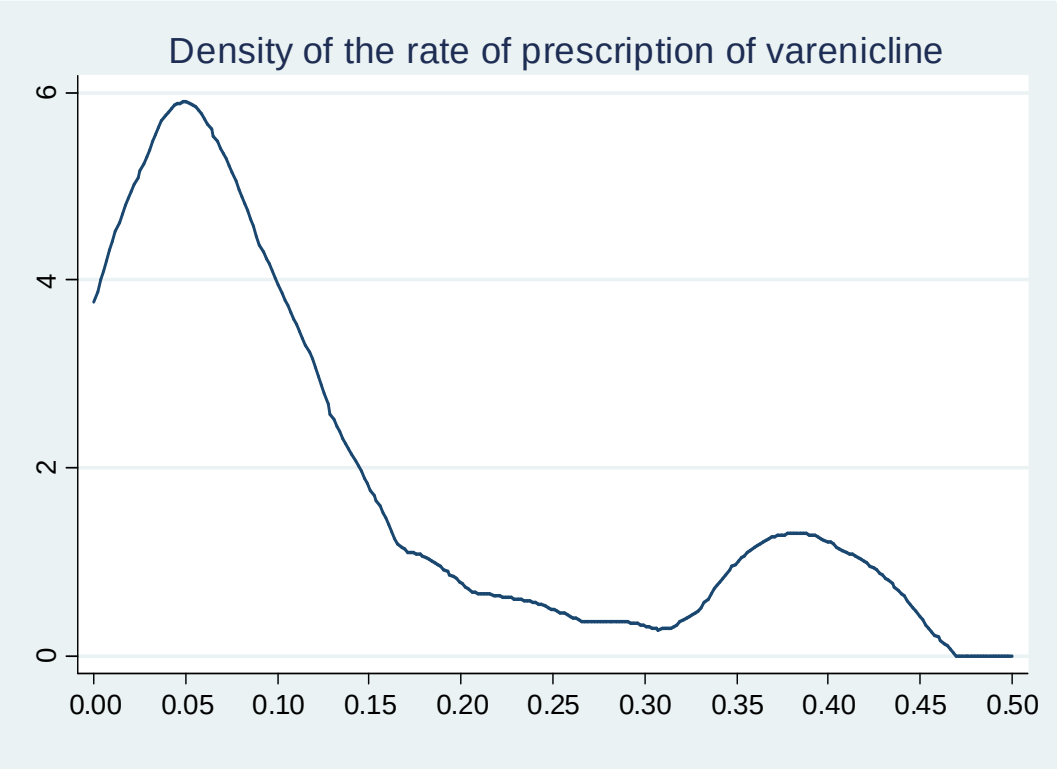
[1]FTND stands for Fagerström Test for Nicotine Dependence and is a measure of patients' degree of addiction.
[2]AHAD is the anxiety scale in the Hospital Anxiety Depression (HAD) scale, scored from 0 to 21, which is used to identify individuals with anxio-depressive disorders, with a threshold score of 11 (see Zigmond et al. [1983]).
[3]DHAD is the depression scale in the Hospital Anxiety Depression (HAD) scale, scored from 0 to 21, which is used to identify individuals with anxio-depressive disorders, with a threshold score of 11 (see Zigmond et al. [1983]).
[4]CO stands for carbon monoxide which is a biomarker for tobacco use.

Table 2: Robustness Checks

		Common Trend		
		Diff in diff	P-value	N
2003-2004		0.045	0.36	1 580
2004-2005		0.032	0.46	2 499
2005-2006		0.042	0.19	4 136
2006-2007		0.082	0.003	5 299
2007-2008		-0.043	0.17	4 400
		Placebo DID		
		Diff in diff	P-value	N
Patients' observable characteristics				
% Males		-0.020	0.46	5 299
Age		-1.153	0.08	5 298
% employed		-0.015	0.57	5 299
% with no degree		0.019	0.36	5 299
Daily cigarettes smoked		1.454	0.02	5 299
FTND		0.237	0.06	5 299
% with AHAD>=11		0.033	0.22	5 299
% with DHAD>=11		-0.010	0.59	5 299
% with chronic obstructive pulmonary diseases		0.043	0.04	5 299
Measurement of smoking status				
Number of days between the first visit and the last CO measure		14.653	0.004	5 299
% included		0.022	0.30	8 581
		P-value of B_ according to inclusion threshold		
		Test centers thresholds		
		Threshold 1: 15%	Threshold 2: 20%	Threshold 3: 25%
Control centers thresholds	Threshold 1: 2%	0.04	0.03	0.04
	Threshold 2: 3%	0.01	0.01	0.02
	Threshold 3: 4%	0.11	0.08	0.14

Figures

Figure 1: Density of the prescription rate of varenicline



Appendix A: Explicit expression of Σ

Let

$$X = \Big(\begin{array}{cccccccc} YTG & Y(1-T)G & YT(1-G) & Y(1-T)(1-G) & YD(1-T)G & YDT(1-G) & YD(1-T)(1-G) \\ DTG & D(1-T)G & DT(1-G) & D(1-T)(1-G) & TG & (1-T)G & T(1-G) & (1-T)(1-G) \end{array} \Big)'$$

Let us denote $\theta = \mathbb{E}(X)$, $V = \mathbb{V}(X)$, $\widehat{\theta}$ the sample counterpart of θ and $\widehat{V}$ the sample counterpart of V.

Since Y is bounded, all the coordinates of X have a variance. Therefore, according to the central limit Theorem,

$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V)$.

Let us denote

$$h(x) = \begin{pmatrix} \frac{\frac{x_1}{x_{12}} - \frac{x_2}{x_{13}} - \frac{x_3}{x_{14}} + \frac{x_4}{x_{15}} + \left(\frac{x_5}{x_9} - M\right) \times \frac{x_9}{x_{13}} + \left(\frac{x_6}{x_{10}} - M\right) \times \frac{x_{10}}{x_{14}} - \left(\frac{x_7}{x_{11}} - m\right) \times \frac{x_{11}}{x_{15}}}{\frac{x_8}{x_{12}}} \\ \frac{x_1}{x_{12}} - M \end{pmatrix},$$

which I define $\forall x = (x_1, x_2, x_3, x_4 x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}) \in \mathbb{R}^7 \times (\mathbb{R}^*)^8$.

$\theta \in \mathbb{R}^7 \times (\mathbb{R}^*)^8$ and h is continuously differentiable over $\mathbb{R}^7 \times (\mathbb{R}^*)^8$ with jacobian $H(x) \in M_{2,15}$.

I can therefore apply the delta method to state that:

$$\sqrt{n} \begin{pmatrix} \widehat{B^0(M,m)} - B^0(M,m) \\ \widehat{\mathbb{E}}(Y_{t_1,g_t} | \, D = 1) - M - (\mathbb{E}(Y_{t_1,g_t} | \, D = 1) - M) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0; \Sigma)$$

where $\Sigma = H(\theta) V H(\theta)'$.

A consistent estimator of Σ is $\widehat{\Sigma} = H(\widehat{\theta}) \widehat{V} H(\widehat{\theta})'$.

Appendix B: proofs

Proof of Lemma DID 1:

$\forall (i,j) \in \{t_0; t_1\} \times \{g_c; g_t\}$, $Y_{i,j} = Y_{i,j}(1) \times D + Y_{i,j}(0) \times (1 - D) = (Y_{i,j}(1) - Y_{i,j}(0)) \times D + Y_{i,j}(0)$,

then,

$$DID = \mathbb{E}\left[(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0))D\right] - \mathbb{E}\left[(Y_{t_0,g_t}(1) - Y_{t_0,g_t}(0))D\right]$$

$$-\mathbb{E}\left[(Y_{t_1,g_c}(1) - Y_{t_1,g_c}(0))D\right] + \mathbb{E}\left[(Y_{t_0,g_c}(1) - Y_{t_0,g_c}(0))D\right]$$

$$+\mathbb{E}(Y_{t_1,g_t}(0)) - \mathbb{E}(Y_{t_0,g_t}(0)) - \mathbb{E}(Y_{t_1,g_c}(0)) + \mathbb{E}(Y_{t_0,g_c}(0)).$$

Under Assumption DID 1,

$$\mathbb{E}(Y_{t_1,g_t}(0)) - \mathbb{E}(Y_{t_0,g_t}(0)) - \mathbb{E}(Y_{t_1,g_c}(0)) + \mathbb{E}(Y_{t_0,g_c}(0)) = 0.$$

Thus

$$DID = \mathbb{E}(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0)|\, D=1) \times \mathbb{P}(D_{t_1,g_t}=1) - \mathbb{E}(Y_{t_0,g_t}(1) - Y_{t_0,g_t}(0)|\, D=1) \times \mathbb{P}(D_{t_0,g_t}=1)$$

$$- \left[\mathbb{E}(Y_{t_1,g_c}(1) - Y_{t_1,g_c}(0)|\, D=1) \times \mathbb{P}(D_{t_1,g_c}=1) - \mathbb{E}(Y_{t_0,g_c}(1) - Y_{t_0,g_c}(0)|\, D=1) \times \mathbb{P}(D_{t_0,g_c}=1)\right], \quad (2)$$

hence the result.

QED.

Proof of Proposition DID 1:

Proof of i)

In the "no always takers" special case, $\mathbb{P}(D_{t_0,g_t}=1)$, $\mathbb{P}(D_{t_1,g_c}=1)$ and $\mathbb{P}(D_{t_0,g_c}=1)$ are all equal to 0. Therefore, (2) can be rewritten as

$$DID = \mathbb{E}(Y_{t_1,g_t}(1) - Y_{t_1,g_t}(0)|\, D=1) \times \mathbb{P}(D_{t_1,g_t}=1),$$

hence the result.

Proof of ii)

From (2),

$$DID = ATT_{t_1,g_t} \times \mathbb{P}(D_{t_1,g_t}=1) - ATT_{t_0,g_t} \times \mathbb{P}(D_{t_0,g_t}=1) - ATT_{t_1,g_c} \times \mathbb{P}(D_{t_1,g_c}=1) + ATT_{t_0,g_c} \times \mathbb{P}(D_{t_0,g_c}=1).$$

If $\forall(i,j) \in \{t_0; t_1\} \times \{g_c; g_t\}$, $ATT_{i,j} = ATT$, then,

$$DID = ATT \times DID^P,$$

hence the result.

QED.

Proof of Proposition DID 2:

Proof of i)

Assume that $\exists(m,\, M) \in \mathbb{R}^2 / \, \mathbb{P}(m \leq Y(0) \leq M) = 1$. I denote

$$A = \mathbb{E}(Y_{t_0,g_t}(0)|\, D=1) \times \mathbb{P}(D_{t_0,g_t}=1) + \mathbb{E}(Y_{t_1,g_c}(0)|\, D=1) \times \mathbb{P}(D_{t_1,g_c}=1) - \mathbb{E}(Y_{t_0,g_c}(0)|\, D=1) \times \mathbb{P}(D_{t_0,g_c}=1).$$

This is the only quantity appearing in (2) which cannot be estimated from the sample and therefore needs to be bounded.

Since $m \leq Y(0) \leq M$, $A_1^- \leq A \leq A_1^+$, with

$$A_1^- = m \times \mathbb{P}(D_{t_0,g_t}=1) + m \times \mathbb{P}(D_{t_1,g_c}=1) - M \times \mathbb{P}(D_{t_0,g_c}=1)$$

and

$$A_1^+ = M \times \mathbb{P}(D_{t_0,g_t}=1) + M \times \mathbb{P}(D_{t_1,g_c}=1) - m \times \mathbb{P}(D_{t_0,g_c}=1).$$

But for bounds to be sharp, the common trend assumption should hold, which implies:

$$0 = \mathbb{E}(Y_{t_1,g_t}(0)|\, D=1) \times \mathbb{P}(D_{t_1,g_t}=1) + \mathbb{E}(Y_{t_1,g_t}|\, D=0) \times (1 - \mathbb{P}(D_{t_1,g_t}=1))$$

$$- \mathbb{E}(Y_{t_0,g_t}(0)|\, D=1) \times \mathbb{P}(D_{t_0,g_t}=1) - \mathbb{E}(Y_{t_0,g_t}|\, D=0) \times (1 - \mathbb{P}(D_{t_0,g_t}=1))$$

$$- \mathbb{E}(Y_{t_1,g_c}(0)|\, D=1) \times \mathbb{P}(D_{t_1,g_c}=1) - \mathbb{E}(Y_{t_1,g_c}|\, D=0) \times (1 - \mathbb{P}(D_{t_1,g_c}=1))$$

$$+ \mathbb{E}(Y_{t_0,g_c}(0)|\, D=1) \times \mathbb{P}(D_{t_0,g_c}=1) + \mathbb{E}(Y_{t_0,g_c}|\, D=0) \times (1 - \mathbb{P}(D_{t_0,g_c}=1)).$$

The only quantity in this equation which is both unobserved and does not enter into (2) is $\mathbb{E}(Y_{t_1,g_t}(0)|\, D = 1)$. For common trend to hold, it should be equal to

$$\frac{A + \mathbb{E}(Y_{t_0,g_t}|\, D = 0) \times (1 - \mathbb{P}(D_{t_0,g_t} = 1)) + \mathbb{E}(Y_{t_1,g_c}|\, D = 0) \times (1 - \mathbb{P}(D_{t_1,g_c} = 1))}{\mathbb{P}(D_{t_1,g_t} = 1)}$$

$$- \frac{\mathbb{E}(Y_{t_1,g_t}|\, D = 0) \times (1 - \mathbb{P}(D_{t_1,g_t} = 1)) + \mathbb{E}(Y_{t_0,g_c}|\, D = 0) \times (1 - \mathbb{P}(D_{t_0,g_c} = 1))}{\mathbb{P}(D_{t_1,g_t} = 1)}.$$

Since $m \leq \mathbb{E}(Y_{t_1,g_t}(0)|\, D = 1) \leq M$, this implies that we should have $A_2^- \leq A \leq A_2^+$, with

$$A_2^- = m \times \mathbb{P}(D_{t_1,g_t} = 1) - \mathbb{E}(Y_{t_0,g_t}|\, D = 0) \times (1 - \mathbb{P}(D_{t_0,g_t} = 1)) - \mathbb{E}(Y_{t_1,g_c}|\, D = 0) \times (1 - \mathbb{P}(D_{t_1,g_c} = 1))$$

$$+ \mathbb{E}(Y_{t_1,g_t}|\, D = 0) \times (1 - \mathbb{P}(D_{t_1,g_t} = 1)) + \mathbb{E}(Y_{t_0,g_c}|\, D = 0) \times (1 - \mathbb{P}(D_{t_0,g_c} = 1))$$

and

$$A_2^+ = M \times \mathbb{P}(D_{t_1,g_t} = 1) - \mathbb{E}(Y_{t_0,g_t}|\, D = 0) \times (1 - \mathbb{P}(D_{t_0,g_t} = 1)) - \mathbb{E}(Y_{t_1,g_c}|\, D = 0) \times (1 - \mathbb{P}(D_{t_1,g_c} = 1))$$

$$+ \mathbb{E}(Y_{t_1,g_t}|\, D = 0) \times (1 - \mathbb{P}(D_{t_1,g_t} = 1)) + \mathbb{E}(Y_{t_0,g_c}|\, D = 0) \times (1 - \mathbb{P}(D_{t_0,g_c} = 1)).$$

Consequently, we should have

$$max(A_1^-; A_2^-) \leq A \leq min(A_1^+; A_2^+). \tag{3}$$

Combining (2) and (3) and rearranging yields B_- and B_+, which are sharp by construction.

I show now that if none of the two bounds is informative then $\mathbb{P}_{AT} > \mathbb{P}(D_{t_1,g_t} = 1)$. If B_- and B_+ are uninformative we have $B^0(M, m) < \mathbb{E}(Y_{t_1,g_t}|\, D = 1) - M$ and $B^0(m, M) > \mathbb{E}(Y_{t_1,g_t}|\, D = 1) - m$. Subtracting those two inequalities yields $\mathbb{P}_{AT} > \mathbb{P}(D_{t_1,g_t} = 1)$. This implies that $\mathbb{P}_{AT} \leq \mathbb{P}(D_{t_1,g_t} = 1)$ is a sufficient condition to have that at least one of the two bounds is informative.

To show that this condition is not sufficient to have that the two bounds are informative, it suffices to consider the following DGP. $M = 1$, $m = 0$, $\mathbb{P}(D_{t_1,g_t} = 1) = 1$, $\mathbb{P}(D_{t_0,g_t} = 1) = \mathbb{P}(D_{t_1,g_c} = 1) = 0.1$, $\mathbb{P}(D_{t_0,g_c} = 1) = 0$, $\mathbb{E}(Y_{t_1,g_t}(1)|\, D = 1) = \mathbb{E}(Y_{t_1,g_t}(0)|\, D = 1) = 1$, $\mathbb{E}(Y_{t_0,g_t}(0)|\, D = 1) = \mathbb{E}(Y_{t_1,g_c}(0)|\, D =$

1) = 0.5, $\mathbb{E}(Y_{t_0,g_t}(0)|\, D=0) = \mathbb{E}(Y_{t_1,g_c}(0)|\, D=0) = 1$, $\mathbb{E}(Y_{t_0,g_c}(0)|\, D=0) = 0.9$. Those are all the quantities which are needed to compute B_- since the remaining expectations cancel out in the calculation. $\mathbb{P}_{AT} = 0.2 \leq \mathbb{P}(D_{t_1,g_t}=1) = 1$, the common trend assumption holds ($1\times1-0.5\times0.1-1\times0.9-0.5\times0.1-1\times0.9+0.9=0$), and B_- is not informative since it is equal to $\mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$.

Proof of ii)

If $\exists(m,\, M) \in \mathbb{R}^2 / \mathbb{P}(m \leq Y(0) \leq M) = 1$,

$$\mathbb{E}(Y_{t_1,g_c}|\, D=1) - M \leq ATT_{t_1,g_c} \leq \mathbb{E}(Y_{t_1,g_c}|\, D=1) - m$$

and

$$\mathbb{E}(Y_{t_0,g_c}|\, D=1) - M \leq ATT_{t_0,g_c} \leq \mathbb{E}(Y_{t_0,g_c}|\, D=1) - m.$$

If $ATT_{t_1,g_c} = ATT_{t_0,g_c} = ATT_{g_c}$, these two inequalities imply that

$$max\left(\mathbb{E}(Y_{t_1,g_c}|\, D=1); \mathbb{E}(Y_{t_0,g_c}|\, D=1)\right) - M \leq ATT_{g_c}$$

and

$$ATT_{g_c} \leq min\left(\mathbb{E}(Y_{t_1,g_c}|\, D=1); \mathbb{E}(Y_{t_0,g_c}|\, D=1)\right) - m.$$

Moreover from (2) we get:

$$ATT_{t_1,g_t} = \frac{DID + ATT_{t_0,g_t} \times \mathbb{P}(D_{t_0,g_t}=1) + ATT_{g_c} \times (\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1))}{\mathbb{P}(D_{t_1,g_t}=1)}.$$

Therefore, combining this last equality with the two preceding inequalities yields B^1 and B^2 as lower or upper bounds to ATT_{t_1,g_t} depending on the sign of $\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)$. For some DGP, $min(B^1; B^2)$ might be smaller than $\mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$, which means that $min(B^1; B^2)$ is not a sharp lower bound, hence the need to set $B_-^{'} = max\left(min(B^1; B^2); \mathbb{E}(Y_{t_1,g_t}|\, D=1) - M\right)$ to ensure sharpness.

Finally, I show that $\mathbb{P}(D_{t_0,g_t}=1) + |\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1)| \leq \mathbb{P}(D_{t_1,g_t}=1)$ is a sufficient condition to have that at least one of the two bounds is informative. Assume $\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1) \geq 0$. None of the two bounds is informative if $B^1 < \mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$ and $B^2 > \mathbb{E}(Y_{t_1,g_t}|\, D=$

$1) - m$. Subtracting those two inequalities yields

$$(M-m) \times (\mathbb{P}(D_{t_0,g_t}=1) + \mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1))$$

$$+ (min\,(\mathbb{E}(Y_{t_1,g_c}|\,D=1); \mathbb{E}(Y_{t_0,g_c}|\,D=1)) - max\,(\mathbb{E}(Y_{t_1,g_c}|\,D=1); \mathbb{E}(Y_{t_0,g_c}|\,D=1))) \times (\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1))$$

$$> (M-m) \times \mathbb{P}(D_{t_1,g_t}=1) \tag{4}$$

Since (4) is a necessary condition to have that none of the bounds is informative, the converse inequality is sufficient to have that at least one of the two bounds is informative. But $\mathbb{P}(D_{t_0,g_t}=1) + \mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1) \leq \mathbb{P}(D_{t_1,g_t}=1)$ implies the converse inequality, hence the result. The proof is symmetric if $\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1) < 0$.

QED.

Proof of Proposition DID 3:

Proof of i)

Under Assumption DID 1 and the supplementary assumptions that $\exists(m,\, M) \in \mathbb{R}^2 / \mathbb{P}(m \leq Y(0) \leq M) = 1$, $ATT_{t_1,g_t} \in [B_-; B_+]$ according to the first part of Proposition DID 2.

$$\lim_{n\to+\infty} \mathbb{P}(ATT_{t_1,g_t} \geq LB^{B_-}_{(1-\alpha)}) \geq \lim_{n\to+\infty} \mathbb{P}(B_- \geq LB^{B_-}_{(1-\alpha)}) = 1 - \frac{\alpha}{2}.$$

Similarly,

$$\lim_{n\to+\infty} \mathbb{P}(ATT_{t_1,g_t} \leq UB^{B_+}_{(1-\alpha)}) \geq 1 - \frac{\alpha}{2}$$

which implies that

$$\lim_{n\to+\infty} \mathbb{P}(LB^{B_-}_{(1-\alpha)} \leq ATT_{t_1,g_t} \leq UB^{B_+}_{(1-\alpha)}) \geq 1 - \alpha.$$

Therefore, $CI^2 = \left[LB^{B_-}_{(1-\alpha)}; UB^{B_+}_{(1-\alpha)}\right]$ is a CI for ATT_{t_1,g_t} with $(1-\alpha)\%$ asymptotic coverage.

Then, consider $\mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq ATT_{t_1,g_t} \leq UB^{B_+}_{(1-2\alpha)})$.

If $ATT_{t_1,g_t} = B_-$,

$$\lim_{n\to+\infty} \mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq B_- \leq UB^{B_+}_{(1-2\alpha)}) = \lim_{n\to+\infty} \mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq B_-) - \lim_{n\to+\infty} \mathbb{P}(B_- > UB^{B_+}_{(1-2\alpha)}) = 1 - \alpha$$

since the second term converges to 0.

If $ATT_{t_1,g_t} = B_+$, the same argument holds and $\lim_{n\to+\infty} \mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq ATT_{t_1,g_t} \leq UB^{B_+}_{(1-2\alpha)}) = 1-\alpha$ as well.

If $B_- < ATT_{t_1,g_t} < B_+$,

$$\lim_{n\to+\infty} \mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq ATT_{t_1,g_t} \leq UB^{B_+}_{(1-2\alpha)})$$

$$= \lim_{n\to+\infty} \mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq ATT_{t_1,g_t}) - \lim_{n\to+\infty} \mathbb{P}(ATT_{t_1,g_t} > UB^{B_+}_{(1-2\alpha)}) = 1.$$

Therefore, $\lim_{n\to+\infty} \mathbb{P}(UB^{B_-}_{(1-2\alpha)} \leq ATT_{t_1,g_t} \leq UB^{B_+}_{(1-2\alpha)}) \geq 1-\alpha$ whatever the value of ATT_{t_1,g_t} so that $CI^3 = \left[LB^{B_-}_{(1-2\alpha)}; UB^{B_-}_{(1-2\alpha)}\right]$ is also a CI for ATT_{t_1,g_t} with $(1-\alpha)\%$ asymptotic coverage.

Proof of ii)

The proof follows the same steps as in i), once noted that under Assumption DID 1 and the supplementary assumptions that $\exists (m,\, M) \in \mathbb{R}^2 / \forall k \in \{0;1\}\ \mathbb{P}(m \leq Y(k) \leq M) = 1$ and that $ATT_{t_1,g_c} = ATT_{t_0,g_c}$, $ATT_{t_1,g_t} \in \left[B'_-; B'_+\right]$ as per the second part of Proposition DID 2.

QED.

Proof of Proposition DID 4:

By the delta method,

$$\sqrt{n}\left(\widehat{B^0(M,m)} - B^0(M,m)\right) \xrightarrow{d} \mathcal{N}(0;\sigma_1^2).$$

By the central limit theorem,

$$\sqrt{n}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\, D=1) - M - (\mathbb{E}(Y_{t_1,g_t}|\, D=1) - M)\right) \xrightarrow{d} \mathcal{N}(0;\sigma_2^2).$$

If $B^0(M,m) > \mathbb{E}(Y_{t_1,g_t}|\, D=1) - M$,

$$\sqrt{n}\left(\widehat{B_-} - B_-\right) = \sqrt{n}\left(max\left(\widehat{B^0(M,m)}; \widehat{\mathbb{E}}(Y_{t_1,g_t}|\, D=1) - M\right) - max\left(B^0(M,m); \mathbb{E}(Y_{t_1,g_t}|\, D=1) - M\right)\right)$$

$$= \sqrt{n}\left(\widehat{B^0(M,m)} - B^0(M,m)\right) + \sqrt{n}\left(max\left(\widehat{B^0(M,m)}; \widehat{\mathbb{E}}(Y_{t_1,g_t}|\, D=1) - M\right) - \widehat{B^0(M,m)}\right).$$

The second term is $o_p(1)$ because $max\left(\widehat{B^0(M,m)}; \widehat{\mathbb{E}}(Y_{t_1,g_t}|\, D=1) - M\right) = \widehat{B^0(M,m)}$ with probabil-

ity approaching 1. This implies the result.

If $B^0(M,m) < \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$, the proof is symmetric.

If $B^0(M,m) = \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$,

$$\sqrt{N}\left(\widehat{B_-} - B_-\right) = max\left(\sqrt{N}\left(\widehat{B^0(M,m)} - B^0(M,m)\right); \sqrt{N}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M - (\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M)\right)\right).$$

Due to the continuous mapping Theorem,

$$max\left(\sqrt{N}\left(\widehat{B^0(M,m)} - B^0(M,m)\right); \sqrt{N}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M - (\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M)\right)\right) \hookrightarrow S^1 = \left(max\left(N^1; N^2\right)\right)$$

where $\begin{pmatrix} N_1 & N_2 \end{pmatrix}' \sim \mathcal{N}(0,\Sigma)$.

QED.

Proof of Proposition DID 5:

$$\mathbb{P}(B_- \in CI) = \mathbb{P}\left(B_- \in CI^A \mid \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}} < \widehat{B^0(M,m)}\right) \times \mathbb{P}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}} < \widehat{B^0(M,m)}\right)$$

$$+\mathbb{P}\left(B_- \in CI^B \mid \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M - \frac{ln(n)}{\sqrt{n}} \leq \widehat{B^0(M,m)} \leq \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}}\right)$$

$$\times\mathbb{P}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M - \frac{ln(n)}{\sqrt{n}} \leq \widehat{B^0(M,m)} \leq \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}}\right)$$

$$+\mathbb{P}\left(B_- \in CI^C \mid \widehat{B^0(M,m)} + \frac{ln(n)}{\sqrt{n}} < \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M\right) \times \mathbb{P}\left(\widehat{B^0(M,m)} + \frac{ln(n)}{\sqrt{n}} < \widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M\right).$$

If $B^0(M,m) > \mathbb{E}(Y_{t_1,g_t}|\,D=1) - M$,

$$\mathbb{P}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - M + \frac{ln(n)}{\sqrt{n}} < \widehat{B^0(M,m)}\right)$$

$$= \mathbb{P}\left(\sqrt{n}\left(\widehat{B^0(M,m)} - B^0(M,m)\right) - \sqrt{n}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t}|\,D=1) - \mathbb{E}(Y_{t_1,g_t}|\,D=1)\right) > ln(n) - \left(B^0(M,m) - (\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M)\right)\sqrt{n}\right)$$

I denote V_n this sequence.

$$\lim_{n\to+\infty} ln(n) - \left(B^0(M,m) - (\mathbb{E}(Y_{t_1,g_t}|\,D=1) - M)\right)\sqrt{n} = -\infty.$$

Consequently, $\forall x \in \mathbb{R}$, $\exists n_0 \in \mathbb{N} / n \geq n_0 \Rightarrow$

$$\mathbb{P}\left(\sqrt{n}\left(\widehat{B^0(M,m)} - B^0(M,m)\right) - \sqrt{n}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M - (\mathbb{E}(Y_{t_1,g_t} | D=1) - M)\right) > x\right) \leq V_n$$

Therefore,

$$\lim_{n\to+\infty} \mathbb{P}\left(\sqrt{n}\left(\widehat{B^0(M,m)} - B^0(M,m)\right) - \sqrt{n}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M - (\mathbb{E}(Y_{t_1,g_t} | D=1) - M)\right) > x\right) \leq \lim_{n\to+\infty} V_n$$

A delta method and the central limit theorem imply that

$$\lim_{n\to+\infty} \mathbb{P}\left(\sqrt{n}\left(\widehat{B^0(M,m)} - B^0(M,m)\right) - \sqrt{n}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M - (\mathbb{E}(Y_{t_1,g_t} | D=1) - M)\right) > x\right) = 1 - F(x),$$

where $F(.)$ is the cdf of a random variable with a normal distribution.

Since this holds $\forall x \in \mathbb{R}$, we can let x go to $-\infty$ which yields $1 \leq \lim_{n\to+\infty} V_n$. Therefore,

$$\lim_{n\to+\infty} \mathbb{P}\left(\widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M + \frac{ln(n)}{\sqrt{n}} < \widehat{B^0(M,m)}\right) = 1,$$

which implies that

$$\lim_{n\to+\infty} \mathbb{P}\left(B_- \in CI^B \mid \widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M - \frac{ln(n)}{\sqrt{n}} \leq \widehat{B^0(M,m)} \leq \widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M + \frac{ln(n)}{\sqrt{n}}\right) = 0$$

and

$$\lim_{n\to+\infty} \mathbb{P}\left(\widehat{B^0(M,m)} + \frac{ln(n)}{\sqrt{n}} < \widehat{\mathbb{E}}(Y_{t_1,g_t} | D=1) - M\right) = 0.$$

Consequently,

$$\lim_{n\to+\infty} \mathbb{P}(B_- \in CI) = \lim_{n\to+\infty} \mathbb{P}\left(B^0(M,m) \in CI^A\right) = 1 - \alpha.$$

If $B^0(M,m) = \mathbb{E}(Y_{t_1,g_t} | D=1) - M$ or $B^0(M,m) < \mathbb{E}(Y_{t_1,g_t} | D=1) - M$, the same type of reasoning yields $\lim_{n\to+\infty} \mathbb{P}(B_- \in CI) = 1 - \alpha$ which completes the proof.

QED.

Proof of Lemma CIC 1:

The proof is fairly straightforward. Assume for instance that Assumption CIC 6 does not hold because $\mathbb{P}(D_{t_1,g_c}=0)<1$. Under Assumption CIC 1 to Assumption CIC 4, $F_{Y_{t_1,g_t}(0)}(y)=F_{Y_{t_0,g_t}(0)}\left(F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right)\right)$ $F_{Y_{t_1,g_c}}(y)=F_{Y_{t_1,g_c}(0)}(y)\times\mathbb{P}(D_{t_1,g_c}=0)+F_{Y_{t_1,g_c}(1)}(y)\times(1-\mathbb{P}(D_{t_1,g_c}=0))$. Since $\mathbb{P}(D_{t_1,g_c}=0)<1$, this is not necessarily equal to $F_{Y_{t_1,g_c}(0)}(y)$. Therefore, $F_{Y_{t_1,g_t}(0)}(y)$ is not necessarily equal to $F_{Y_{t_0,g_t}}\left(F^{-1}_{Y_{t_0,g_c}}\left(F_{Y_{t_1,g_c}}(y)\right)\right)$.

QED.

Proof of Theorem CIC 1:

Proof of i)

In the no always takers special case,

$$F_{Y_{t_0,g_t}}\left(F^{-1}_{Y_{t_0,g_c}}\left(F_{Y_{t_1,g_c}}(y)\right)\right)=F_{Y_{t_0,g_t}(0)}\left(F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right)\right).$$

According to part i) of Athey and Imbens's Theorem 3.1, this is equal to $F_{Y_{t_1,g_t}(0)}(y)$, which can be rewritten as

$$F_{Y_{t_1,g_t}(0)|D=1}(y)\times\mathbb{P}(D_{t_1,g_t}=1)+F_{Y_{t_1,g_t}(0)|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0).$$

Rearranging this last equation yields the result.

Proof of ii)

In the no always takers special case,

$$\mathbb{E}\left(Y_{t_1,g_t}\right)-\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}}\left(F_{Y_{t_0,g_c}}(Y_{t_0,g_t})\right)\right)=\mathbb{E}\left(Y_{t_1,g_t}\right)-\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}(0)}\left(F_{Y_{t_0,g_c}(0)}(Y_{t_0,g_t}(0))\right)\right)=\mathbb{E}\left(Y_{t_1,g_t}\right)-\mathbb{E}\left(Y_{t_1,g_t}(0)\right)$$

according to part iii) of Athey and Imbens's Theorem 3.1. This last expression can be rewritten as

$$\mathbb{E}(Y_{t_1,g_t}(1)|\,D_{t_1,g_t}=1)\times\mathbb{P}(D_{t_1,g_t}=1)+\mathbb{E}(Y_{t_1,g_t}(0)|\,D_{t_1,g_t}=0)\times\mathbb{P}(D_{t_1,g_t}=0)$$

$$-\mathbb{E}(Y_{t_1,g_t}(0)|\,D_{t_1,g_t}=1)\times\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{E}(Y_{t_1,g_t}(0)|\,D_{t_1,g_t}=0)\times\mathbb{P}(D_{t_1,g_t}=0)$$

which is equal to

$$\mathbb{E}(Y_{t_1,g_t}(1)-Y_{t_1,g_t}(0)|\,D_{t_1,g_t}=1)\times\mathbb{P}(D_{t_1,g_t}=1),$$

hence the result.

Proof of iii)

$$\tau_q^{CIC} = F^{-1}_{Y_{t_1,g_t}(1)|D=1}(q) - F^{-1}_{Y_{t_1,g_t}(0)|D=1}(q)$$

$$= F^{-1}_{Y_{t_1,g_t}|D=1}(q) - inf\left\{y \in \mathbb{Y}_{t_1,g_t}(0)/F_{Y_{t_1,g_t}(0)|D=1}(y) \geq q\right\}$$

$$= F^{-1}_{Y_{t_1,g_t}|D=1}(q) - inf\left\{y \in \mathbb{Y}_{t_1,g_c}/F_{Y_{t_1,g_t}(0)|D=1}(y) \geq q\right\}$$

because of Assumption CIC 4 and because of the definition of the no always takers special case,

$$= F^{-1}_{Y_{t_1,g_t}|D=1}(q) - inf\left\{y \in \mathbb{Y}_{t_1,g_c}/G(y) \geq q\right\}$$

because of i).

QED.

Proof of Proposition CIC 1:

Under Assumption CIC 1 to Assumption CIC 4,

$$F_{Y_{t_1,g_t}(0)}(y) = F_{Y_{t_0,g_t}(0)}\left(F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right)\right). \tag{5}$$

$$F_{Y_{t_1,g_c}(0)}(y) = F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c} = 0) + F_{Y_{t_1,g_c}(0)|D=1}(y) \times \mathbb{P}(D_{t_1,g_c} = 1).$$

Since $0 \leq F_{Y_{t_1,g_c}(0)|D=1}(y) \leq 1$,

$$F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c} = 0) \leq F_{Y_{t_1,g_c}(0)}(y) \leq F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c} = 0) + \mathbb{P}(D_{t_1,g_c} = 1).$$

Since $F^{-1}_{Y_{t_0,g_c}(0)}$ is weakly increasing,

$$F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c} = 0)\right) \leq F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right)$$

and

$$F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right) \leq F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c}=0) + \mathbb{P}(D_{t_1,g_c}=1)\right) \tag{6}$$

Then,

$$F_{Y_{t_0,g_c}(0)}(y) = F_{Y_{t_0,g_c}|D=0}(y) \times \mathbb{P}(D_{t_0,g_c}=0) + F_{Y_{t_0,g_c}(0)|D=1}(y) \times \mathbb{P}(D_{t_0,g_c}=1).$$

Since $0 \leq F_{Y_{t_0,g_c}(0)|D=1}(y) \leq 1$,

$$F_{Y_{t_0,g_c}|D=0}(y) \times \mathbb{P}(D_{t_0,g_c}=0) \leq F_{Y_{t_0,g_c}(0)}(y) \leq F_{Y_{t_0,g_c}|D=0}(y) \times \mathbb{P}(D_{t_0,g_c}=0) + \mathbb{P}(D_{t_0,g_c}=1).$$

Therefore,

$$\left\{y/F_{Y_{t_0,g_c}|D=0}(y) \geq \frac{q}{\mathbb{P}(D_{t_0,g_c}=0)}\right\} \subseteq \left\{y/F_{Y_{t_0,g_c}(0)}(y) \geq q\right\} \subseteq \left\{y/F_{Y_{t_0,g_c}|D=0}(y) \geq \frac{q-\mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right\}$$

which implies:

$$F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{q-\mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right) \leq F^{-1}_{Y_{t_0,g_c}(0)}(q) \leq F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{q}{\mathbb{P}(D_{t_0,g_c}=0)}\right). \tag{7}$$

Combining (6) and (7) yields:

$$F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c}=0) - \mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right) \leq F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right)$$

and

$$F^{-1}_{Y_{t_0,g_c}(0)}\left(F_{Y_{t_1,g_c}(0)}(y)\right) \leq F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y) \times \mathbb{P}(D_{t_1,g_c}=0) + \mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right). \tag{8}$$

Then,

$$F_{Y_{t_0,g_t}(0)}(y) = F_{Y_{t_0,g_t}|D=0}(y) \times \mathbb{P}(D_{t_0,g_t}=0) + F_{Y_{t_0,g_t}(0)|D=1}(y) \times \mathbb{P}(D_{t_0,g_t}=1).$$

Since $0 \leq F_{Y_{t_0,g_t}(0)|D=1}(y) \leq 1$,

$$F_{Y_{t_0,g_t}|D=0}(y) \times \mathbb{P}(D_{t_0,g_t}=0) \leq F_{Y_{t_0,g_t}(0)}(y) \leq F_{Y_{t_0,g_t}|D=0}(y) \times \mathbb{P}(D_{t_0,g_t}=0) + \mathbb{P}(D_{t_0,g_t}=1). \tag{9}$$

Combining (5), (8) and (9) yields:

$$F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)\leq F_{Y_{t_1,g_t}(0)}(y)$$

and

$$F_{Y_{t_1,g_t}(0)}(y)\leq F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)+\mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)+\mathbb{P}(D_{t_0,g_t}=1),$$

which implies

$$B^{CIC}_{-}\leq F_{Y_{t_1,g_t}(0)|D=1}(y)\leq B^{CIC}_{+}$$

with

$$B^{CIC}_{-}=\frac{F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)-F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=1)}$$

and

$$B^{CIC}_{+}=\frac{F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)+\mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)+\mathbb{P}(D_{t_0,g_t}=1)-F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=1)}.$$

QED.

Proof of Theorem CIC 2:

Proof of i)

To alleviate the notational burden, I introduce $U_g \overset{d}{\sim} U|G=g$.

By assumption Assumption CIC 2, $h_j(u,t)$ is invertible with respect to u. Denote $h_j^{-1}(u;t)$ its inverse.

$\forall(t,g,j,k)\in\{t_0;t_1\}\times\{g_c;g_t\}\times\{0;1\}^2$,

$$F_{Y_{t,g}(j)|D(0)=k}(y)=\mathbb{P}\left(h_j(U,t)\leq y|G=g,T=t,D(0)=k\right)$$

$$=\mathbb{P}\left(U\leq h_j^{-1}(y;t)|G=g,T=t,D(0)=k\right)$$

$$=\mathbb{P}\left(U\leq h_j^{-1}(y;t)|G=g,D(0)=k\right)$$

by assumption Assumption CIC 3'.

Therefore,

$$F_{Y_{t,g}(j)|D(0)=k}(y) = F_{U_g|D(0)=k}(h_j^{-1}(y;t)). \tag{10}$$

Let j and k be equal to 1.

I apply (10) to all four combinations $(t,g) \in \{t_0;t_1\} \times \{g_c;g_t\}$.

First, letting $(t,g) = (t_0,g_c)$ and substituting $y = h_1(u,t_0)$ yields $F_{Y_{t_0,g_c}(1)|D(0)=1}(h_1(u,t_0)) = F_{U_{g_c}|D(0)=1}(u)$. Then applying $F^{-1}_{Y_{t_0,g_c}(1)|D(0)=1}(.)$ to each side, we have, $\forall u \in \mathbb{U}|G = g_c, D(0) = 1$,

$$h_1(u,t_0) = F^{-1}_{Y_{t_0,g_c}(1)|D(0)=1}\left(F_{U_{g_c}|D(0)=1}(u)\right). \tag{11}$$

Second, letting $(t,g) = (t_1,g_c)$ and using the fact that $\forall y \in \mathbb{Y}_{t_1,g_c}|D(0) = 1$, $h_1^{-1}(y;1) \in \mathbb{U}_{g_c}|D(0) = 1$, and applying the transformation $F^{-1}_{U_{g_c}|D(0)=1}(.)$ to both sides of (10), yields

$$F^{-1}_{U_{g_c}|D(0)=1}(F_{Y_{t_1,g_c}(1)|D(0)=1}(y)) = h_1^{-1}(y;t_1). \tag{12}$$

Combining (11) and (12) yields, $\forall y \in \mathbb{Y}_{t_1,g_c}|D(0) = 1$,

$$h_1(h_1^{-1}(y;t_1),t_0) = F^{-1}_{Y_{t_0,g_c}(1)|D(0)=1}\left(F_{Y_{t_1,g_c}(1)|D(0)=1}(y)\right). \tag{13}$$

Third, apply (10) with $(t,g) = (t_0,g_t)$ and substitute to get $y = h_1(u,t_0)$ to get

$$F_{Y_{t_0,g_t}(1)|D(0)=1}(h_1(u,t_0)) = F_{U_{g_t}|D(0)=1}(u). \tag{14}$$

Fourth, apply (10) with $(t,g) = (t_1,g_t)$ to get

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) = F_{U_{g_t}|D(0)=1}(h_1^{-1}(y;t_1)). \tag{15}$$

Therefore, combining (14) and (15) yields

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) = F_{Y_{t_0,g_t}(1)|D(0)=1}(h_1(h_1^{-1}(y;t_1),t_0)). \tag{16}$$

Since by Assumption CIC 4' $\mathbb{Y}_{t_1,g_t}|D(0)=1 \subseteq \mathbb{Y}_{t_1,g_c}|D(0)=1$, substituting (13) in (16) we finally get

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) = F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(F_{Y_{t_1,g_c}|D=1}(y)\right)\right) \tag{17}$$

once noted that $F_{Y_{t_0,g_t}(1)|D(0)=1}(.) = F_{Y_{t_0,g_t}|D=1}(.)$, $F^{-1}_{Y_{t_0,g_t}(1)|D(0)=1}(.) = F^{-1}_{Y_{t_0,g_t}|D=1}(.)$ and $F_{Y_{t_1,g_c}(1)|D(0)=1}(.) = F_{Y_{t_1,g_c}|D=1}(.)$.

Letting j and k be equal to 0 and can show that

$$F_{Y_{t_1,g_t}(0)|D(0)=0}(y) = F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(F_{Y_{t_1,g_c}|D=0}(y)\right)\right). \tag{18}$$

Then,

$$F_{Y_{t_1,g_t}|D=1}(y) = F_{Y_{t_1,g_t}(1)|D(1)=1}(y)$$

$$= F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=1}(y)\times\mathbb{P}(D_{t_1,g_t}(0)=1|\,D(1)=1)+F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y)\times\mathbb{P}(D_{t_1,g_t}(0)=0|\,D(1)=1).$$

By Assumption CIC 6, this is equal to

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}+F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_1,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}.$$

By Assumption CIC 3', this can be rewritten as

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y)\times\frac{\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}+F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}.$$

According to (21), this is equal to

$$F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(F_{Y_{t_1,g_c}|D=1}(y)\right)\right)\times\frac{\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}$$

$$+F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}.$$

Rearranging this last equation yields the first part of i).

Now, I prove the second part of i). According to (18),

$$F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(F_{Y_{t_1,g_c}|D=0}(y)\right)\right) = F_{Y_{t_1,g_t}(0)|D(0)=0}(y)$$

$$= F_{Y_{t_1,g_t}(0)|D(1)=0,D(0)=0}(y)\times\mathbb{P}(D_{t_1,g_t}(0)=0|\,D(1)=0)+F_{Y_{t_1,g_t}(0)|D(1)=1,D(0)=0}(y)\times\mathbb{P}(D_{t_1,g_t}(1)=1|\,D(0)=0).$$

By Assumption CIC 6, this is equal to

$$F_{Y_{t_1,g_t}(0)|D(1)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=0)}{\mathbb{P}(D_{t_1,g_t}(0)=0)}+F_{Y_{t_1,g_t}(0)|D(1)=1,D(0)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_1,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(0)=0)}.$$

By Assumption CIC 3', this can be rewritten as

$$F_{Y_{t_1,g_t}(0)|D(1)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=0)}{\mathbb{P}(D_{t_0,g_t}(0)=0)}+F_{Y_{t_1,g_t}(0)|D(1)=1,D(0)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_0,g_t}(0)=0)}.$$

Rearranging this last equation yields the second part of i).

Proof of ii)

From (21), one can show that $Y_{t_1,g_t}(1)|D(0)=1$ and $F^{-1}_{Y_{t_1,g_c}|D=1}\left(F_{Y_{t_0,g_c}|D=1}(Y_{t_0,g_t})\right)|D=1$ have the same cdf. Similarly, one can show from (18) that $Y_{t_1,g_t}(0)|D(0)=0$ and $F^{-1}_{Y_{t_1,g_c}|D=0}\left(F_{Y_{t_0,g_c}|D=0}(Y_{t_0,g_t})\right)|D=0$ also have the same cdf. Therefore, taking expectations yields

$$\mathbb{E}\left(Y_{t_1,g_t}(1)|D(0)=1\right)=\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=1}\left(F_{Y_{t_0,g_c}|D=1}(Y_{t_0,g_t})\right)|D=1\right) \tag{19}$$

and

$$\mathbb{E}\left(Y_{t_1,g_t}(0)|D(0)=0\right)=\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=0}\left(F_{Y_{t_0,g_c}|D=0}(Y_{t_0,g_t})\right)|D=0\right) \tag{20}$$

Then,

$$\mathbb{E}\left(Y_{t_1,g_t}|D=1\right)=\mathbb{E}\left(Y_{t_1,g_t}|D(1)=1\right)$$

$$=\mathbb{E}\left(Y_{t_1,g_t}(1)|D(0)=D(1)=1\right)\times\mathbb{P}(D_{t_1,g_t}(0)=1|\,D(1)=1)$$

$$+\mathbb{E}\left(Y_{t_1,g_t}(1)|D(1)>D(0)\right)\times\mathbb{P}(D_{t_1,g_t}(0)=1|\,D(1)=1).$$

By Assumption CIC 6 this is equal to

$$\mathbb{E}\left(Y_{t_1,g_t}(1)|D(0)=1\right)\times\frac{\mathbb{P}(D_{t_1,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}+\mathbb{E}\left(Y_{t_1,g_t}(1)|D(1)>D(0)\right)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_1,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}.$$

By Assumption CIC 3' this can be rewritten as

$$\mathbb{E}\left(Y_{t_1,g_t}(1)|D(0)=1\right)\times\frac{\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}+\mathbb{E}\left(Y_{t_1,g_t}(1)|D(1)>D(0)\right)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}.$$

By (19) this is equal to

$$\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=1}\left(F_{Y_{t_0,g_c}|D=1}(Y_{t_0,g_t})\right)|D=1\right)\times\frac{\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}$$

$$+\mathbb{E}\left(Y_{t_1,g_t}(1)|D(1)>D(0)\right)\times\frac{\mathbb{P}(D_{t_1,g_t}(1)=1)-\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(1)=1)}.$$

Rearranging this last equation yields:

$$\mathbb{E}\left(Y_{t_1,g_t}(1)|D(1)>D(0)\right)=\frac{\mathbb{E}\left(Y_{t_1,g_t}|D=1\right)\times\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=1}\left(F_{Y_{t_0,g_c}|D=1}(Y_{t_0,g_t})\right)|D=1\right)\times\mathbb{P}(D_{t_0,g_t}=1)}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)}.$$

Similarly, one can show that

$$\mathbb{E}\left(Y_{t_1,g_t}(0)|D(1)>D(0)\right)=\frac{\mathbb{E}\left(F^{-1}_{Y_{t_1,g_c}|D=0}\left(F_{Y_{t_0,g_c}|D=0}(Y_{t_0,g_t})\right)|D=0\right)\times\mathbb{P}(D_{t_0,g_t}=0)-\mathbb{E}\left(Y_{t_1,g_t}|D=0\right)\times\mathbb{P}(D_{t_1,g_t}=0)}{\mathbb{P}(D_{t_0,g_t}=0)-\mathbb{P}(D_{t_1,g_t}=0)}.$$

Combining these last two equations yields the result.

Proof of iii)

$$\tau_q^{IV-CIC}=F^{-1}_{Y_{t_1,g_t}(1)|D(1)>D(0)}(q)-F^{-1}_{Y_{t_1,g_t}(0)|D(1)>D(0)}(q)$$

$$=inf\left\{y\in\mathbb{Y}_{t_1,g_t}(1)|D(1)>D(0)/F_{Y_{t_1,g_t}(1)|D(1)>D(0)}(y)\geq q\right\}$$

$$-inf\left\{y\in\mathbb{Y}_{t_1,g_t}(0)|D(1)>D(0)/F_{Y_{t_1,g_t}(0)|D(1)>D(0)}(y)\geq q\right\}.$$

$$=inf\left\{y\in\mathbb{Y}_{t_1,g_t}|D=1/F_{Y_{t_1,g_t}(1)|D(1)>D(0)}(y)\geq q\right\}-inf\left\{y\in\mathbb{Y}_{t_1,g_c}|D=0/F_{Y_{t_1,g_t}(0)|D(1)>D(0)}(y)\geq q\right\}.$$

The first change in support holds because $\mathbb{Y}_{t_1,g_t}(1)|D(1)>D(0)\subseteq\mathbb{Y}_{t_1,g_t}|D=1$: $D(1)>D(0)\Rightarrow$

$D(1) = 1$, and $D = D(1)$ in the treatment group $\times$ period 1 cell.

The second change in support holds since $\mathbb{Y}_{t_1,g_t}(0)|D(1) > D(0) \subseteq \mathbb{Y}_{t_1,g_c}(0)|D(1) > D(0)$ because of Assumption CIC 4'. Moreover, $D(1) > D(0) \Rightarrow D(0) = 0$ and $D = D(0)$ in the treatment group $\times$ period 1 cell, so that $\mathbb{Y}_{t_1,g_c}(0)|D(1) > D(0) \subseteq \mathbb{Y}_{t_1,g_c}|D = 0$. Combining those two steps yields$\mathbb{Y}_{t_1,g_t}(0)|D(1) > D(0) \subseteq \mathbb{Y}_{t_1,g_c}|D = 0$.

Finally, using the formulas in i), we get that this last expression is equal to

$$inf\left\{y \in \mathbb{Y}_{t_1,g_t}|D = 1/H^1(y) \geq q\right\} - inf\left\{y \in \mathbb{Y}_{t_1,g_c}|D = 0/H^0(y) \geq q\right\}.$$

QED.

Proof of Proposition CIC 2:

As shown in the proof of Theorem CIC 2,

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) = F_{Y_{t_0,g_t}|D(0)=1}\left(F^{-1}_{Y_{t_0,g_c}|D(0)=1}\left(F_{Y_{t_1,g_c}|D(0)=1}(y)\right)\right) \tag{21}$$

$F_{Y_{t_0,g_t}(1)|D(0)=1}(.) = F_{Y_{t_0,g_t}|D=1}(.)$, $F^{-1}_{Y_{t_0,g_c}(1)|D(0)=1}(.) = F^{-1}_{Y_{t_0,g_c}|D=1}(.)$, but $F_{Y_{t_1,g_c}(1)|D(0)=1}(.)$ is not observed.

$$F_{Y_{t_1,g_c}|D=1}(y) = F_{Y_{t_1,g_c}(1)|D(1)=1,D(0)=1}(y) \times \mathbb{P}(D_{t_1,g_c}(1) = 1, D_{t_1,g_c}(0) = 1|\, D(1) = 1)$$

$$+F_{Y_{t_1,g_c}(1)|D(1)=1,D(0)=0}(y) \times \mathbb{P}(D_{t_1,g_c}(1) = 1, D_{t_1,g_c}(0) = 0|\, D(1) = 1)$$

By Assumption CIC 6' and Assumption CIC 3', this is equal to

$$F_{Y_{t_1,g_c}(1)|D(0)=1}(y) \times \frac{\mathbb{P}(D_{t_0,g_c}(0) = 1)}{\mathbb{P}(D_{t_1,g_c}(1) = 1)}$$

$$+F_{Y_{t_1,g_c}(1)|D(1)=1,D(0)=0}(y) \times \frac{\mathbb{P}(D_{t_1,g_c}(1) = 1) - \mathbb{P}(D_{t_0,g_c}(0) = 1)}{\mathbb{P}(D_{t_1,g_c}(1) = 1)}.$$

Since $0 \leq F_{Y_{t_1,g_c}(1)|D(1)=1,D(0)=0}(y) \leq 1$ and $\mathbb{P}(D_{t_1,g_c}(1)=1) - \mathbb{P}(D_{t_0,g_c}(0)=1) \geq 0$,

$$\frac{F_{Y_{t_1,g_c}|D=1}(y) \times \mathbb{P}(D_{t_1,g_c}=1) - (\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1))}{\mathbb{P}(D_{t_0,g_c}=1)} \leq F_{Y_{t_1,g_c}(1)|D(0)=1}(y)$$

and

$$F_{Y_{t_1,g_c}(1)|D(0)=1}(y) \leq \frac{F_{Y_{t_1,g_c}|D=1}(y) \times \mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=1)}. \tag{22}$$

Plugging (22) into (21) finally yields

$$F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(\frac{F_{Y_{t_1,g_c}|D=1}(y) \times \mathbb{P}(D_{t_1,g_c}=1) - (\mathbb{P}(D_{t_1,g_c}=1) - \mathbb{P}(D_{t_0,g_c}=1))}{\mathbb{P}(D_{t_0,g_c}=1)}\right)\right) \leq F_{Y_{t_1,g_t}(1)|D(0)=1}(y)$$

and

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) \leq F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(\frac{F_{Y_{t_1,g_c}|D=1}(y) \times \mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=1)}\right)\right). \tag{23}$$

Then,

$$F_{Y_{t_1,g_t}|D=1}(y) = F_{Y_{t_1,g_t}(1)|D(2)=1}(y)$$

$$= F_{Y_{t_1,g_t}(1)|D(2)=1,D(1)=1,D(0)=1}(y) \times \mathbb{P}(D_{t_1,g_t}(2)=1, D_{t_1,g_t}(1)=1, D_{t_1,g_t}(0)=1|\, D(2)=1)$$

$$+F_{Y_{t_1,g_t}(1)|D(2)=1,D(1)=1,D(0)=0}(y) \times \mathbb{P}(D_{t_1,g_t}(2)=1, D_{t_1,g_t}(1)=1, D_{t_1,g_t}(0)=0|\, D(2)=1)$$

$$+F_{Y_{t_1,g_t}(1)|D(2)=1,D(1)=0,D(0)=0}(y) \times \mathbb{P}(D_{t_1,g_t}(2)=1, D_{t_1,g_t}(1)=0, D_{t_1,g_t}(0)=0|\, D(2)=1)$$

By Assumption CIC 6', this is equal to

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) \times \frac{\mathbb{P}(D_{t_1,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(2)=1)}$$

$$+F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y) \times \frac{\mathbb{P}(D_{t_1,g_t}(1)=1, D_{t_1,g_t}(0)=0)}{\mathbb{P}(D_{t_1,g_t}(2)=1)}$$

$$+F_{Y_{t_1,g_t}(1)|D(2)=1,D(1)=0}(y) \times \frac{\mathbb{P}(D_{t_1,g_t}(2)=1, D_{t_1,g_t}(1)=0)}{\mathbb{P}(D_{t_1,g_t}(2)=1)}.$$

By Assumption CIC 6', Assumption CIC 3' and Assumption CIC 7, this can be rewritten as

$$F_{Y_{t_1,g_t}(1)|D(0)=1}(y) \times \frac{\mathbb{P}(D_{t_0,g_t}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(2)=1)}$$

$$+F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y) \times \frac{\mathbb{P}(D_{t_1,g_c}(1)=1)-\mathbb{P}(D_{t_0,g_c}(0)=1)}{\mathbb{P}(D_{t_1,g_t}(2)=1)}$$

$$+F_{Y_{t_1,g_t}(1)|D(2)=1,D(1)=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_t}(2)=1)-[\mathbb{P}(D_{t_0,g_t}(0)=1)+\mathbb{P}(D_{t_1,g_c}(1)=1)-\mathbb{P}(D_{t_0,g_c}(0)=1)]}{\mathbb{P}(D_{t_1,g_t}(2)=1)}.$$

Since $0 \leq F_{Y_{t_1,g_t}(1)|D(1)=1,D(0)=0}(y) \leq 1$ and using (23) we finally get

$$B_{-}^{1,CIC-IV} \leq F_{Y_{t_1,g_t}(1)|D(2)=1,D(1)=0}(y) \leq B_{+}^{1,CIC-IV} \tag{24}$$

with

$$B_{-}^{1,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=1}(y)\times\mathbb{P}(D_{t_1,g_t}=1)-F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(\frac{F_{Y_{t_1,g_c}|D=1}(y)\times\mathbb{P}(D_{t_1,g_c}=1)}{\mathbb{P}(D_{t_0,g_c}=1)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=1)-\left[\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right]}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)-\left[\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right]}$$

and

$$B_{+}^{1,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=1}(y)\times\mathbb{P}(D_{t_1,g_t}=1)-F_{Y_{t_0,g_t}|D=1}\left(F^{-1}_{Y_{t_0,g_c}|D=1}\left(\frac{F_{Y_{t_1,g_c}|D=1}(y)\times\mathbb{P}(D_{t_1,g_c}=1)-\left(\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right)}{\mathbb{P}(D_{t_0,g_c}=1)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=1)}{\mathbb{P}(D_{t_1,g_t}=1)-\mathbb{P}(D_{t_0,g_t}=1)-\left[\mathbb{P}(D_{t_1,g_c}=1)-\mathbb{P}(D_{t_0,g_c}=1)\right]}.$$

Similarly, one can show that

$$B_{-}^{0,CIC-IV} \leq F_{Y_{t_1,g_t}(0)|D(2)=1,D(1)=0}(y) \leq B_{+}^{0,CIC-IV} \tag{25}$$

with

$$B_{-}^{0,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)-F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(F_{Y_{t_1,g_c}|D=0}(y)\times\frac{\mathbb{P}(D_{t_1,g_c}=0)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)-\left[\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right]}{\mathbb{P}(D_{t_1,g_t}=0)-\mathbb{P}(D_{t_0,g_t}=0)-\left[\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right]}$$

and

$$B_{+}^{0,CIC-IV} = \frac{F_{Y_{t_1,g_t}|D=0}(y)\times\mathbb{P}(D_{t_1,g_t}=0)-F_{Y_{t_0,g_t}|D=0}\left(F^{-1}_{Y_{t_0,g_c}|D=0}\left(\frac{F_{Y_{t_1,g_c}|D=0}(y)\times\mathbb{P}(D_{t_1,g_c}=0)-\left(\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right)}{\mathbb{P}(D_{t_0,g_c}=0)}\right)\right)\times\mathbb{P}(D_{t_0,g_t}=0)}{\mathbb{P}(D_{t_1,g_t}=0)-\mathbb{P}(D_{t_0,g_t}=0)-\left[\mathbb{P}(D_{t_1,g_c}=0)-\mathbb{P}(D_{t_0,g_c}=0)\right]}.$$

QED.