Warwick Economics Research Paper Series

# Supplement to Fuzzy Differences-in-Differences

Clement de Chaisemartin and Xavier D'Haultfoeuille

October, 2015
Series Number: 1066
ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

# Supplement to “Fuzzy Differences-in-Differences”

Clément de Chaisemartin* Xavier D’Haultfœuille†

October 6, 2015

**Abstract**

This paper gathers the supplementary material to de Chaisemartin & D’Haultfœuille (2015). First, we show that two commonly used IV and OLS regressions with time and group fixed effects estimate weighted averages of Wald-DIDs. It then follows from Theorem 3.1 in de Chaisemartin & D’Haultfœuille (2015) that these regressions estimate weighted sums of LATEs, with potentially many negative weights as we illustrate through two applications. We review all papers published in the American Economic Review between 2010 and 2012 and find that 10.1% of these papers estimate one or the other regression. Second, we consider estimators of the bounds on average and quantile treatment effects derived in Theorems 3.2 and 3.3 in de Chaisemartin & D’Haultfœuille (2015) and we study their asymptotic behavior. Third, we revisit Gentzkow et al. (2011) and Field (2007) using our estimators. Finally, we present all the remaining proofs not included in the main paper.

## 1 Fuzzy DID regressions, and their pervasiveness in economics

### 1.1 Fuzzy DID regressions...

Researchers using fuzzy DID designs usually do not estimate simple regressions with two groups and two periods, but more complex specifications with multiple groups and periods. Practices are not unified so details of their specifications can vary. In this section, we study two regression specifications which have often been used. We show that in both cases, the coefficient of treatment is equal to a weighted sum of Wald-DIDs. Following the result of the first point of Theorem 3.1, it then easy to show that this weighted sum can be rewritten as a weighted sum of the LATEs of switchers in the different groups, with potentially many negative weights, as we illustrate through two examples. Therefore, these coefficients could lie far from the LATE of switchers in any group.

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

†CREST, xavier.dhaultfoeuille@ensae.fr

First, we study the coefficient of a treatment variable $D$ in a 2SLS regression of $Y$ on a constant, group dummies $(1\{G=g\})_{1\leq g\leq \overline{g}}$, time dummies $(1\{T=t\})_{1\leq t\leq \overline{t}}$, and $D$, with a first stage fully saturated in $(T,G)$. As the first stage is fully saturated, the second stage is a regression of $Y$ on a constant, group dummies $(1\{G=g\})_{1\leq g\leq \overline{g}}$, time dummies $(1\{T=t\})_{1\leq t\leq \overline{t}}$, and $E(D|T,G)$. This 2SLS regression is therefore algebraically equivalent to an OLS regression at the group $\times$ period level of $Y$ on time and group dummies and a measure of treatment intensity in each group $\times$ period cell. As shown in the next subsection, such OLS regressions are pervasive in applied work.

Assume that for every $1\leq t\leq \overline{t}$ the mean of treatment does not follow a parallel evolution in any pair of groups between $t-1$ and $t$.[1] For every $(g,g',t)\in\{0,...,\overline{g}\}^2\times\{1,...,\overline{t}\}$, let

$$\begin{aligned} DID_D(g,g',t) &= E(D_{gt})-E(D_{gt-1})-\left(E(D_{g't})-E(D_{g't-1})\right), \\ W_{DID}(g,g',t) &= \frac{E(Y_{gt})-E(Y_{gt-1})-\left(E(Y_{g't})-E(Y_{g't-1})\right)}{E(D_{gt})-E(D_{gt-1})-\left(E(D_{g't})-E(D_{g't-1})\right)}. \end{aligned}$$

For $(g,t)\in\{1,...,\overline{g}\}\times\{1,...,\overline{t}\}$, let

$$w^a_{gt}=\frac{DID_D(g,g-1,t)P(G\geq g)P(T\geq t)\left(E\left(D|G\geq g,T\geq t\right)-E\left(D|G\geq g\right)-E\left(D|T\geq t\right)+E(D)\right)}{\sum_{t=1}^{\overline{t}}\sum_{g=1}^{\overline{g}}DID_D(g,g-1,t)P(G\geq g)P(T\geq t)\left(E\left(D|G\geq g,T\geq t\right)-E\left(D|G\geq g\right)-E\left(D|T\geq t\right)+E(D)\right)}.$$

For $(g,t)\in\{0,...,\overline{g}\}\times\{1,...,\overline{t}\}$, let

$$w^b_{gt} = \frac{\left[E(D_{gt})-E(D_{gt-1})\right]P(G=g)P(T\geq t)(E(D|G=g,T\geq t)-E(D|G=g)-E(D|T\geq t)+E(D))}{\sum_{t=1}^{\overline{t}}\sum_{g=0}^{\overline{g}}\left[E(D_{gt})-E(D_{gt-1})\right]P(G=g)P(T\geq t)(E(D|G=g,T\geq t)-E(D|G=g)-E(D|T\geq t)+E(D))}.$$

**Theorem S1** *Let $\beta$ denote the coefficient of $D$ in a 2SLS regression of $Y$ on a constant, $(1\{G=g\})_{1\leq g\leq \overline{g}}$, $(1\{T=t\})_{1\leq t\leq \overline{t}}$, and $D$, with a first stage fully saturated in $(T,G)$.*

1. *If $T \perp\!\!\!\perp G$,*

$$\beta = \sum_{t=1}^{\overline{t}}\sum_{g=1}^{\overline{g}}W_{DID}(g,g-1,t)w^a_{gt}.$$

2. *Morever, if $D$ is binary and Model (1) and Assumptions 1-4 are satisfied,*

$$\beta=\sum_{t=1}^{\overline{t}}\sum_{g=0}^{\overline{g}}\Delta_{gt}w^b_{gt},$$

*where $\Delta_{gt}$ is the LATE of units in group $g$ switching treatment between $t-1$ and $t$.*

[1] If for some $t$, there are groups which experience a parallel evolution of their mean treatment between $t-1$ and $t$, the formula in the first point of Theorem 1 remains valid after grouping together these groups. The formula in the second point of Theorem 1 remains valid as is.

The first statement of the theorem shows that if $T \perp\!\!\!\perp G$, $\beta$ is a weighted average of Wald-DIDs between $t-1$ and $t$ and across pairs of groups, for all consecutive dates $t-1$ and $t$. With only two dates, one can order groups according to their increase in treatment between the two dates, thus ensuring that all the weights $w^a_{gt}$ are positive. With more than two dates, some of the weights $w^a_{gt}$ might be negative.

Then, it follows from Theorem 3.1 that when $D$ is binary and under appropriate common trends assumptions, each of these Wald-DIDs is equal to a weighted difference of the LATE of switchers of both groups. Rearranging this sum of weighted differences yields the second result. A similar result with the same weights holds if treatment is not binary but ordered and with a finite support. A difference though is that in such instances, $\beta$ is not equal to a weighted sum of LATEs but to a weighted sum of the ACRs parameters we introduced in Section 4.3. Some of the weights $w^b_{gt}$ might be negative. With two periods, that will be the case for instance if the distribution of the changes in treatment between period 0 and 1 across groups is not symmetric around 0.

Many papers estimate the regression studied in Theorem 1 with aggregate data at the group $\times$ period level. Results of Theorem 1 still apply to these regressions. We now review three cases of such group-level regressions which frequently arise in practice. First, when the group level variables are constructed from micro-level variables (e.g.: average wage in county $c$ and year $t$) and the OLS regression is weighted by the population in each group $\times$ period, the first and second statements of Theorem 1 apply as is. Second, when the group level variables are constructed from micro-level variables but the regressions are not weighted, the first and second statements of Theorem 1 also apply as is, except that now $P(G = g) = \frac{1}{\overline{g}}$ for every group. Note that with unweighted regressions, $G$ is automatically independent of $T$ unless some groups appear or disappear, which is unlikely to be the case when groups are counties, states, or regions. Third, there are instances where all units in each group $\times$ period share the same value of the treatment. This is for instance the case in Gentzkow et al. (2011). When that is the case, the second statement of Theorem 1 actually gets simpler. In such settings, when treatment changes in one group, all units switch treatment. Therefore, $\Delta_{gt}$ is equal to the average effect of changing the treatment from its value in period $t-1$ to its value in period $t$ across all units, normalized by the change in treatment from period $t-1$ to $t$.

We use Theorem 1 to revisit an empirical application. Enikolopov et al. (2011) study the effect of having access to an independent TV channel on the share of people voting for opposition parties in Russia. They regress the share of votes for opposition parties in the 1995 and 1999 elections in region $r$ on region dummies, an indicator for the 1999 election, and on the share of people having access to the independent TV channel in region $r$ at the time of the election. Figure 1 below presents the weights $w^d_{g1999}$ for the 1938 regions in their sample. Regions are ordered according to the increase in the share of people watching the independent TV channel they experienced between the two elections, from the lowest to the largest increase. 1020 weights are negative, and the negative weights sum up to -2.26 (against 3.26 for positive weights: negative weights therefore account for 41% of the sum of the absolute value of weights). If the effect of gaining access to an independent TV channel is heterogeneous across regions where few / many voters gained access to it between 1995 and 1999, the regression coefficients in Enikolopov et al. (2011) could lie far from the LATE in any region.

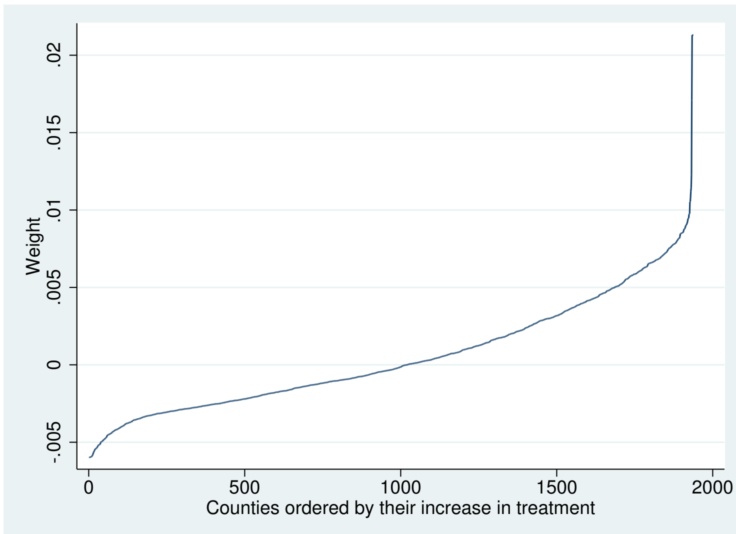


*Figure S1*: $w^b_{gt}$ in Enikolopov et al. (2011).

Second, we study the coefficient of a treatment variable $D$ in a 2SLS regression of $Y$ on a constant, group dummies $(1\{G=g\})_{1\leq g\leq \overline{g}}$, time dummies $(1\{T=t\})_{1\leq t\leq \overline{t}}$, and $D$, where the instrument for $D$ is equal to $f(G)1\{T\geq t_0\}$ for some $t_0\geq 1$. This specification corresponds exactly to the one estimated in the first column and third line of Table 7 in Duflo (2001): there $f(G)$ is the number of schools constructed during the INPRES program in one's district of birth, and $1\{T\geq t_0\}$ is a dummy for being born late enough to enter school after the program completion.

Let $T^{**}=1\{T\geq t_0\}$. For any random variable $R$ and for any $(g,t)\in\{0,...,\overline{g}\}\times\{0,1\}$, let $R^{**}_{gt}\sim R|G=g,T^{**}=t$. Assume that there are no groups where treatment follows a parallel evolution before and after $t_0$,[2] and let groups be ordered according to their increase of treatment before and after $t_0$:

$$E(D^{**}_{01})-E(D^{**}_{00})<E(D^{**}_{11})-E(D^{**}_{10})<...<E(D^{**}_{\overline{g}1})-E(D^{**}_{\overline{g}0}).$$

[2]If there are groups which experience a parallel evolution of their mean treatment, the formula in the first point of Theorem 2 remains valid after grouping together these groups. The formula in the second point remains valid as is.

For any $(g, g') \in \{0, ..., \overline{g}\}^2$, let

$$\begin{aligned} DID^{**}_R(g, g') &= E(R^{**}_{g1}) - E(R^{**}_{g0}) - (E(R^{**}_{g'1}) - E(R^{**}_{g'0})), \\ W^{**}_{DID}(g, g') &= \frac{DID^{**}_Y(g, g')}{DID^{**}_D(g, g')}. \end{aligned}$$

Let also

$$\begin{aligned} w^c_g &= \frac{DID^{**}_D(g, g-1)P(G \geq g)(E(f(G)|G \geq g) - E(f(G)))}{\sum_{g'=1}^{\overline{g}} DID^{**}_D(g', g'-1)P(G \geq g')(E(f(G)|G \geq g') - E(f(G)))} \text{ for } 1 \leq g \leq \overline{g}, \\ w^d_g &= \frac{\left[E(D^{**}_{g1}) - E(D^{**}_{g0})\right] P(G = g)(f(g) - E(f(G)))}{\sum_{g=0}^{\overline{g}} \left[E(D^{**}_{g1}) - E(D^{**}_{g0})\right] P(G = g)(f(g) - E(f(G)))} \text{ for } 0 \leq g \leq \overline{g}. \end{aligned}$$

**Theorem S2** *Let $\beta$ denote the coefficient of $D$ in a 2SLS regression of $Y$ on a constant, $(1\{G = g\})_{1 \leq g \leq \overline{g}}$, $(1\{T = t\})_{1 \leq t \leq \overline{t}}$, and $D$, where the instrument for $D$ writes as $f(G)1\{T \geq t_0\}$ for some $1 \leq t_0 \leq \overline{t}$.*

1. *If $T \perp\!\!\!\perp G$,*

$$\beta = \sum_{g=1}^{\overline{g}} W^{**}_{DID}(g, g-1) w^c_g.$$

2. *Morever, if $D$ is binary and Model* (1) *and Assumptions 1-4 are satisfied with $T^{**}$ instead of $T$,*

$$\beta = \sum_{g=0}^{\overline{g}} \Delta_g w^d_g,$$

*where $\Delta_g$ is the LATE of the switchers of group $g$.*

The first statement of the theorem shows that if $T \perp\!\!\!\perp G$, $\beta$ is a weighted average of Wald-DIDs before and after $t_0$ and across groups with consecutive evolutions of their mean treatment. Then, it follows from Theorem 3.1 that when $D$ is binary and under appropriate common trends assumptions, each of these Wald-DIDs is equal to a weighted difference of the LATE of switchers of both groups. Rearranging this sum of weighted differences yields the second result. Here as well, a similar result with the same weights holds if treatment is not binary but ordered and with a finite support. Note that the weights $w^d_g$ are all positive if and only if all groups where treatment increases (resp. decreases) have a value of $f(G)$ greater (resp. lower) than the mean of $f(G)$ in the population.

We illustrate this result by estimating the weights $w^d_g$ for the 284 districts in Duflo (2001). Districts are ordered according to the increase in years of schooling they experienced between the two cohorts, from the lowest to the largest increase. 132 weights out of 284 are negative, and the negative weights sum up to -3.28 (against 4.28 for positive weights). If switchers' ACRs are heterogeneous across districts with positive and negative weights, the regression coefficient in Duflo (2001) could lie far from the ACR of switchers in any district.

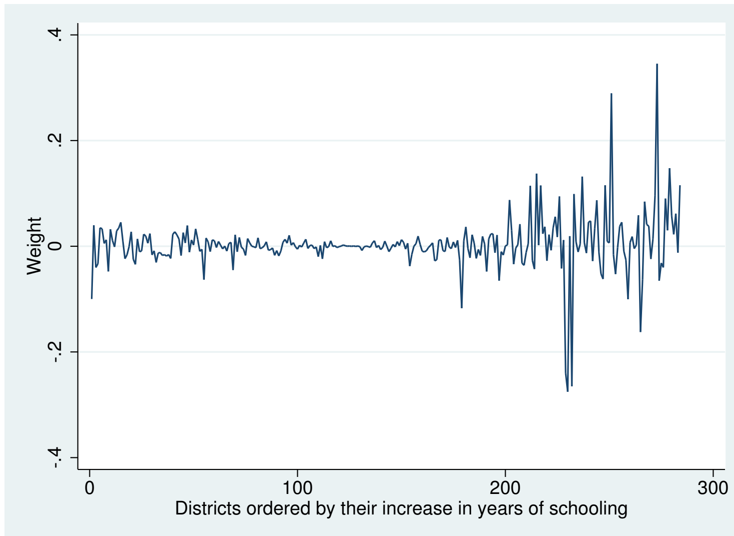


*Figure S2*: $w_g^d$ in Duflo (2001).

### 1.2 ... and their pervasiveness in economics.

We assess the pervasiveness of the fuzzy DID method in economics by conducting a review of all papers relying partly or fully on this method that were published in the American Economic Review (AER) between 2010 and 2012. We choose this journal because among the top journals in economics, it was the first which started posting online the data used in the empirical papers it publishes, thus enabling us to reanalyze some of the fuzzy DID papers published there.

Over these three years, the AER published 337 papers. This excludes papers and proceedings, comments, replies, and presidential addresses. Out of these 337 papers, 34 papers estimate either ratios of DIDs on their outcome and treatment variables, or the regressions studied in Theorems S2 and S1, or regressions very close to one of these two regressions. Therefore, their main coefficient of interest is equal to a Wald-DID, or to a weighted average of Wald-DIDs. When one withdraws from the denominator theory papers and lab experiments, the proportion of papers using the fuzzy DID method raises to 19.5%. Fuzzy DID is therefore a very popular method among economists using real world data to study empirical questions.

*Table S1*: Fuzzy DID papers published in the AER between 2010 and 2012

| | 2010 | 2011 | 2012 | Total |
|---|---|---|---|---|
| # papers using the fuzzy DID method | 5 | 15 | 14 | 34 |
| % of published papers | 5.2% | 13.0% | 11.2% | 10.1% |
| % of empirical papers, excluding lab experiments | 12.8% | 24.6% | 19.2% | 19.7% |

We now review each of the 34 papers published by the AER between 2010 and 2012 and which we included in our fuzzy DID count, and carefully justify why their methodology qualifies as a fuzzy DID. For each paper, we use the following presentation:
**Title of the paper.** *Where the fuzzy DID method is used in the paper.*
Why the method used in the paper qualifies as a fuzzy DID.

1. **Patient Cost-Sharing and Hospitalization Offsets in the Elderly.** *Elasticities of care use to co-payment estimated after Tables 2 and 3.*
   The elasticity discussed after Table 2 is estimated as the ratio of the effect of the Medicare reform on utilization, divided by the effect of the Medicare reform on co-payment. Both effects are estimated through standard sharp DID specifications in Table 2. Therefore, the elasticity estimate is a Wald-DID. Note that even though elasticities do not appear in regression tables, estimating them is one of the main goals of the paper: elasticity estimates are referred to in the abstract.

2. **The Effect of Medicare Part D on Pharmaceutical Prices and Utilization.** *Tables 2 and 3.*
   In regression equation (1), the dependent variable is the change in the price of drug j between 2003 and 2006, and the explanatory variable is the Medicare market share for drug j in 2003. This regression is equivalent to that studied in Theorem S1, with two periods (2003 and 2006), drug dummies, and a treatment equal to 0 in 2003 and to the Medicare market share of drug j in 2003 in 2006.

3. **The Gender Wage Gap and Domestic Violence.** *Table 2.*
   In regression equation (2), the dependent variable is the log of female assaults among females of race r in county c in year t, and the explanatory variables are race, year, county, race × year, race × county, and county × year dummies, as well as the gender wage gap in county c, year t, and race r. Differencing this equation with respect to one race (say white people) yields the same regression as that considered in Theorem S1.

4. **Inherited Trust and Growth.** *Figure 4 and Table 6.*
   Figure 4 presents a regression of changes in income per capita from 1935 to 2000 on changes in inherited trust over the same period and a constant. This regression is equivalent to that in Theorem S1 with 2 periods, country dummies, and inherited trust as the treatment variable.

5. **Inheritance Law and Investment in Family Firms.** *Table 7.*
   In the regressions presented in Table 7, the dependent variable is the capital expenditure of firm j in year t, and the explanatory variables are firm dummies, a dummy for whether year t is a succession period for firm j, and the interaction of this dummy with the level of investor protection in the country where firm j is located. This specification is similar to that studied in Theorem S1 with two periods (succession and no succession).

6. **Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms.** *Tables 3 to 12.*
   In regression equation (11), the dependent variable is the change in exporting status of firm i in sector j between 1992 and 1996, and the explanatory variable is the change in trade tariffs in Brasil for products in sector j over the same period. This regression is equivalent to that studied in Theorem S1.

7. **Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards.** *Table 5 column 2.*
   In the regression in Table 5 column (2), the dependent variable is a dummy for whether a car sold is a flexible fuel vehicle, and the explanatory variables are state and month dummies, and the percent ethanol availability in each month × state. This regression is the same as that considered in Theorem S1.

8. **What Do Trade Negotiators Negotiate About? Empirical Evidence from the World Trade Organization.** *Table 3, OLS columns.*
   In regression equations (15a) and (15b), the dependent variable is the ad valorem tariff level bound by country c on product g, while the explanatory variables are country and product fixed effects, and two treatment variables which vary at the country × product level. These regressions are therefore the same as that considered in Theorem S1, except that they have two treatment variables.

9. **Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia.** *Tables 3 and 4, columns 4-6.*
   In the regression in, say, Table 3 column (4), the dependent variable is the total number of contributions to Wikipedia by individual i at period t, regressed on individual fixed effects, a dummy for whether period t is after the Wikipedia block, and the interaction of this

dummy and a measure of social participation by individual i. This regression is the same as that considered in Theorem S1 (treatment is equal to 0 before the block, and to social participation after it).

10. **Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks.** *Table 2, Panel C, Columns 3-4.*
In regression equation (7), the dependent variable is change in crime rates between week t and the same week one year ago in borough b, and the explanatory variables are a dummy for whether week t is around the terrorist attacks in London, and the number of police forces in borough b in week t. The interaction of the time dummy and of whether borough b belongs to Theseus operation is used as the excluded instrument for police forces. This regression is equivalent to that studied in Theorem S2 (borough fixed effects disappear because of the first differencing with respect to the previous year, something the authors do to control for seasonality).

11. **The Impact of Regulations on the Supply and Quality of Care in Child Care Markets.** *Table 7, Columns 4 and 5.*
In Regression Equation (1), the dependent variable is the outcome for market m in state s in year t, and the explanatory variables are state and year fixed effects and a measure of regulations in state s in year t. This regression is the same as that considered in Theorem S1.

12. **House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis.** *Tables 2 and 3.*
Regression equations (1) and (2) are respectively equivalent to the second and first stages of the 2SLS regression studied in Theorem S2. Here, everything is in first differences between 2006 and 2002. In Theorem S2 we consider the regression in levels but with fixed effects so the two specifications are equivalent. In levels, the instrument would be the elasticity interacted with the year 2006.

13. **State Misallocation and Housing Prices: Theory and Evidence from China.** *Table 5, Panel A.*
In regression equation (15), the dependent variable is a measure of the quantity of housing services in household i's residence in year t, while the explanatory variables are a dummy for period t being after the reform, a measure of mismatch in household i, and the interaction of the measure of mismatch and the time dummy. This specification almost perfectly coincides with that studied in Theorem S1, except that it has a measure a mismatch in household i instead of household fixed effects. If the mismatch measure can take only two values, it is easy to show that the coefficient of interest $\alpha_1$ is equal to the DID of the

outcome before and after the reform and across the two groups of households, divided by the difference between the value of mismatch in these two groups.

14. **The Fundamental Law of Road Congestion: Evidence from US Cities.** *Table 5.* In regression equation (4), the dependent variable is the change in vehicle kilometers traveled in MSA s between periods t and t-1, and the explanatory variable is the change in kilometers of roads in MSA s between periods t and t-1. This regression is almost equivalent to that considered in Theorem S1, except that it does not have time specific constants. The two regressions are equivalent if the effect of time on the outcome is linear. If that is the case, the first differences of the time effects in the equation in levels $\delta_t - \delta_{t-1}$ are equal to each other, so estimating the first-difference regression with just one constant or with time specific constants will yield the same result.

15. **The Consequences of Radical Reform: The French Revolution.** *Table 3.* In Equation (1), the dependent variable is urbanization in polity j at time t, while the explanatory variables are time and polity dummies, and the number of years of French presence in polity j interacted with the time effects. This regression is equivalent to that studied in Theorem S1.

16. **School Desegregation, School Choice, and Changes in Residential Location Patterns by Race.** *Table 6.* In the regression presented in, say, the first column of Table 6, the dependent variable is enrolment in schools of MSA j in year t, while the explanatory variables are time and MSA effects and the value of the dissimilarity index of schools in MSA j in year t. The excluded instrument for the dissimilarity index is a dummy for whether in period t, the MSA was desegregated. This regression is the same as that studied in Theorem S2.

17. **The Effects of Rural Electrification on Employment: New Evidence from South Africa.** *Tables 4 and 5 columns 5-8, Table 8 columns 3-4, Table 9 column 2, and Table 10 columns 2, 4, and 6.* Regression equations (3) and (4) are respectively equivalent to the second and first stages of the 2SLS regression studied in Theorem S2. Here, everything is in first differences between the first and the second wave of the panel. In Theorem S2 we consider the regression in levels but with fixed effects so the two specifications are equivalent. In Theorem S2, the instrument would be the land gradient $Z_j$ interacted with a dummy for the second wave of the panel.

18. **Media and Political Persuation: Evidence from Russia.** *Table 3.* In regression equation (5), the dependent variable is the share of votes for party j in year t and subregion s, and the explanatory variables are subregion and time effects, and the

NTV audience in subregion s in period t. This regression is the same as that studied in Theorem S1.

19. **Dynamic Inefficiencies in an Employment-Based Health Insurance System: Theory and Evidence.** *Tables 2, 3, 5, and 6, Column 3.*
    In regression equation (7), the dependent variable is the health expenditures of individual j working in industry i in period t and region r, and the explanatory variables are individual effects, region specific time effects, and the job tenure of individual j. The death rate of establishments in industry i in period t and region r is used as an instrument for the job tenure of individual j. Within each region, the regression has time effects and individual effects, and an instrument varying only across industry × periods cells. Even though this instrument does not have the exact same form as that in the regression studied in Theorem S2, these two regressions are close.

20. **The Effect of Newspaper Entry and Exit on Electoral Politics.** *Tables 2 and 3.*
    In regression equation (1), the dependent variable is, say, voter turnout in county c in election year t, and the explanatory variables are county fixed effects, state-year effects, and the number of newspapers in county c in year t. Within each state, this regression is the same as that studied in Theorem S1 (within each state, state-year effects become year effects).

21. **Americans Do IT Better: US Multinationals and the Productivity Miracle.** *Table 2, Columns 6-8.*
    In the regression in, say, column 6 of Table 2, the dependent variable is the log of output per worker in firm i in period t, while the explanatory variables are firms and time fixed effects, and the log of the amount of IT capital per employee ($\ln(C/L)$) as well as the interaction of $\ln(C/L)$ and a dummy for whether the firm is owned by a US multinational. The coefficient of $\ln(C/L)$ is equal to the same weighted average of Wald-DIDs as the coefficient considered in Theorem S1, within the sample of firms which are not owned by a US multinational. The coefficient of the interaction is equal to the difference between this weighted average in the sample of firms owned by a US multinational, and in the sample of those not owned by a US multinational.

22. **Standard Setting Committees: Consensus Governance for Shared Technology Platforms.** *Table 4, columns 1-3.*
    In regression equation (5), the dependent variable is a measure of time to consensus for project i submitted to committee j, while the explanatory variables are a dummy for projects submitted to the standards track, a measure of distributional conflict, and the interaction of the standards track and distributional conflict. This specification almost coincides with that studied in Theorem S1, except that it has a measure of distributional

conflict instead of committee fixed effects. If the measure of distributional conflict can take only two values, it is easy to show that the coefficient of interest $\tau$ is equal to the DID of the outcome across the standards and non-standards track and the low and high value of distributional conflict, divided by the difference between the value of distributional conflict in these two groups.

23. **Compulsory Licensing: Evidence from the Trading with the Enemy Act.** *Table 2, columns 3-8.*
    In the regression equation in the beginning of Section III, the dependent variable is the number of patents by US inventors in patent class c at period t, and the explanatory variables are patent class and time fixed effects, and the interaction of period t being after the trading with the enemy act and a measure of treatment intensity. Therefore, this regression is the same as that in Theorem S1 (treatment is equal to 0 before the act).

24. **The Internet and Local Wages: A Puzzle.** *Tables 2 and 4.*
    In regression equation (1), the dependent variable is the difference between log wages in 2000 and 1995 in county i, and the explanatory variable is the extent of advanced Internet investment by businesses in county i in 2000. This regression is equivalent to that in Theorem S1. Table 4 presents regressions where advanced internet investment is instrumented by a county level variable. These regressions are equivalent to that in Theorem S2.

25. **Estimating the Peace Dividend: The Impact of Violence on House Prices in Northern Ireland.** *Table 1, columns 3 and 5-7.*
    In regression equation (1), the dependent variable is the price of houses in region r at time t, while the explanatory variables include region and time fixed effects, and the numbers of people killed because of the civil war in region r at time t-1. This regression is the same as that studied in Theorem S1.

26. **Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry.** *Tables 2 and 5.*
    In regression equation (1), the dependent variable is the change of the log premium for employer e in market m in year t, and explanatory variables are time and market effects, and the change in various treatment variables (change in the fraction of self-insured employees...). This regression is the same as that studied in Theorem S1, except that it has several treatment variables. In regression equation (3), the treatment variables are instrumented by a dummy for period $t$ being after the merger of two insurers and a market level-variable. This regression is similar to that studied in Theorem S2.

27. **The Enduring Impact of the American Dust Bowl: Short- and Long-Run Ad-**

**justments to Environmental Catastrophe.** *Table 2.* In regression equation (1), the dependent variable is, say, the change in log land value in county c between period t and 1930, and the explanatory variables are state × year effects, the share of county c in high erosion, and the share of county c in medium erosion. Within each state, this regression is equivalent to that in Theorem S1, except that it has two treatment variables.

28. **A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution.** *Table 5.*
In, say, the first regression equation in the bottom of page 1915, the dependent variable is the change in the price of house j between sales 2 and 3, and the explanatory variables are the change in various pollutants in the area around house j between sales 2 and 3. This regression is equivalent to that in Theorem S1, except that it has several treatment variables.

29. **The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit.** *Table 3.*
In the reduced form of regression equation (4), the dependent variable is the change in test scores for child i between years a and a-1, while the explanatory variable is the change in the expected EITC income of her family based on her family income in year a-1. This regression is equivalent to that considered in Theorem S1, except that it does not have years specific intercepts. The first stage is the same regression but with the change in the income of the family of student i between years a and a-1. Overall, the 2SLS coefficient arising from regression equation (4) is a ratio of 2 weighted averages of Wald-DIDs.

30. **Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees.** *Tables 3-6.*
In regression equation (1), the dependent variable is the test score of student i in school j in grade g in year t, and the explanatory variables are grade, school, year, and grade × year effects, and the fraction of Katrina students received by school j in grade g and year t. Within each grade, this regression is the same as that considered in Theorem S1 (within each grade, grade × year effects become simple year effects).

31. **The Collateral Channel: How Real Estate Shocks Affect Corporate Investment.** *Table 5.*
In regression equation (1), the dependent variable is the value of investment in firm i and year t divided by the lagged book value of properties, plants, and equipments (PPE), and the explanatory variables are firm and time dummies and the market value of firm i in year t divided by its lagged PPE. This regression is the same as that studied in Theorem S1.

32. **The Spending and Debt Response to Minimum Wage Hikes.** *Tables 1, 2, and 5.*
In regression equation (1), the outcome variable is, say, income of household i at period t, and the explanatory variables include household and time dummies, and the minimum wage in the state where household i lives in period t. This regression is the same as that considered in Theorem S1.

33. **Exports, Export Destinations, and Skills.** *Table 5.*
In regression equation (7), the dependent variable is a measure of skills in the labor force employed by company i in industry j at period t, and the explanatory variables are firm and industry × time dummies, the ratio of exports to sales in firm i at period t, and the share of firm exports to high income destinations over total exports. To instrument this variable, the authors use a dummy for the years 1999 or 2000 (a large devaluatation happened in Brazil in 1999) interacted with the share of exports of firm i to Brazil in 1998. This specification is very similar to that studied in Theorem S2.

34. **Political Aid Cycles.** *Table 3, columns 4 and 5, and Tables 4 and 5.*
In regression equation (2), the dependent variable is the amount of donations received by receiver r from donor d in year t, and the explanatory variables are donor × receiver dummies, a dummy for whether there is an election in country r in year t, a measure of alignment between the ruling political parties in countries r and d, and the interaction of the election dummy and the measure of alignment. This specification is very close to that studied in Theorem S1, with units of observation being pairs of donors and receivers.

## 2 Inference in the partially identified case

In this section, we show how to draw inference on the bounds given in the second statements of Theorems 3.2 and 3.3 in de Chaisemartin & D'Haultfœuille (2015). We adopt the same notations hereafter. In order for the bounds to be finite, we assume that $\mathcal{S}(Y) = [\underline{y}, \overline{y}]$ with $-\infty < \underline{y} < \overline{y} < +\infty$. We also suppose for simplicity that $\underline{y}$ and $\overline{y}$ are known by the researcher.[3] If not, they can respectively be estimated by $\min_{i=1\dots n} Y_i$ and $\max_{i=1\dots n} Y_i$, and Theorem S3 below remains valid under regularity conditions on $F_{Y_{d01}}$ at these boundaries.

First, let us consider the Wald-TC bounds. Let $\widehat{\lambda}_{0d} = \frac{\widehat{P}(D_{01}=d)}{\widehat{P}(D_{00}=d)}$, $\widehat{\lambda}_{1d} = \frac{\widehat{P}(D_{11}=d)}{\widehat{P}(D_{10}=d)}$, and

$$\widehat{\underline{F}}_{d01}(y) = M_0\left[1 - \widehat{\lambda}_{0d}(1 - \widehat{F}_{Y_{d01}}(y))\right] - M_0(1 - \widehat{\lambda}_{0d})\mathbb{1}\{y < \overline{y}\},$$
$$\widehat{\overline{F}}_{d01}(y) = m_1\left[\widehat{\lambda}_{0d}\widehat{F}_{Y_{d01}}(y)\right] + (1 - m_1(\widehat{\lambda}_{0d}))\mathbb{1}\{y \geq \underline{y}\}.$$

[3] In particular, we estimate $F^{-1}_{Y_{dgt}}(0)$ and $F^{-1}_{Y_{dgt}}(1)$ by $\underline{y}$ and $\overline{y}$ respectively. The definition of $\widehat{F}^{-1}_{Y_{dgt}}(\tau)$ for $\tau \in (0, 1)$ remains the same as in Section 5 of de Chaisemartin & D'Haultfœuille (2015).

Then define

$$\widehat{\underline{\delta}}_d = \int y d\widehat{\overline{F}}_{d01}(y) - \frac{1}{n_{d00}} \sum_{i \in \mathcal{I}_{d00}} Y_i, \quad \widehat{\overline{\delta}}_d = \int y d\widehat{\underline{F}}_{d01}(y) - \frac{1}{n_{d00}} \sum_{i \in \mathcal{I}_{d00}} Y_i.$$

Finally, we estimate the bounds by

$$\widehat{\underline{W}}_{TC} = \frac{\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} Y_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} \left[ Y_i + \widehat{\overline{\delta}}_{D_i} \right]}{\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} D_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} D_i}, \ \widehat{\overline{W}}_{TC} = \frac{\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} Y_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} \left[ Y_i + \widehat{\underline{\delta}}_{D_i} \right]}{\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} D_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} D_i}.$$

Now let us turn to the Wald-CIC bounds. For $d \in \{0, 1\}$, let

$$\widehat{\underline{T}}_d = M_{01} \left( \frac{\widehat{\lambda}_{0d} \widehat{F}_{Y_{d01}} - \widehat{H}_d^{-1}(\widehat{\lambda}_{1d} \widehat{F}_{Y_{d11}})}{\widehat{\lambda}_{0d} - 1} \right), \widehat{\overline{T}}_d = M_{01} \left( \frac{\widehat{\lambda}_{0d} \widehat{F}_{Y_{d01}} - \widehat{H}_d^{-1}(\widehat{\lambda}_{1d} \widehat{F}_{Y_{d11}} + (1 - \widehat{\lambda}_{1d}))}{\widehat{\lambda}_{0d} - 1} \right),$$

$$\widehat{G}_d(T) = \widehat{\lambda}_{0d} \widehat{F}_{Y_{d01}} + (1 - \widehat{\lambda}_{0d}) T, \ \widehat{C}_d(T) = \frac{\widehat{\lambda}_{1d} \widehat{F}_{Y_{d11}} - \widehat{H}_d \circ \widehat{G}_d(T)}{\widehat{\lambda}_{1d} - 1}.$$

We then estimate the bounds on $F_{Y_{11}(d)|S_1}$ by

$$\widehat{\underline{F}}_{CIC,d}(y) = \sup_{y' \leq y} \widehat{C}_d \left( \widehat{\underline{T}}_d \right)(y'), \quad \widehat{\overline{F}}_{CIC,d}(y) = \inf_{y' \geq y} \widehat{C}_d \left( \widehat{\overline{T}}_d \right)(y').$$

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

$$\widehat{\underline{W}}_{CIC} = \int y d\widehat{\overline{F}}_{CIC,1}(y) - \int y d\widehat{\underline{F}}_{CIC,0}(y), \quad \widehat{\overline{W}}_{CIC} = \int y d\widehat{\underline{F}}_{CIC,1}(y) - \int y d\widehat{\overline{F}}_{CIC,0}(y),$$

$$\widehat{\underline{\tau}}_q = \widehat{\overline{F}}_{CIC,1}^{-1}(q) - \widehat{\underline{F}}_{CIC,0}^{-1}(q), \quad \widehat{\overline{\tau}}_q = \widehat{\underline{F}}_{CIC,1}^{-1}(q) - \widehat{\overline{F}}_{CIC,0}^{-1}(q).$$

Hereafter, we define $\underline{q} = \overline{F}_{CIC,0}(\underline{y})$, $\overline{q} = \underline{F}_{CIC,0}(\overline{y})$, $q_1 = [\lambda_{11} F_{Y_{111}} \circ F_{Y_{101}}^{-1}(\frac{1}{\lambda_{01}}) - 1]/[\lambda_{11} - 1]$ and $q_2 = [\lambda_{11} F_{Y_{111}} \circ F_{Y_{101}}^{-1}(1 - 1/\lambda_{01})]/[\lambda_{11} - 1]$. Our results rely on the following assumptions.

**Assumption S1** *(Technical conditions for inference with TC bounds)*

1. $\mathcal{S}(Y) = [\underline{y}, \overline{y}]$ *with* $-\infty < \underline{y} < \overline{y} < +\infty$.

2. $\lambda_{00} \neq 1$ *and for* $d \in \{0, 1\}$, *the equation* $F_{d01}(y) = 1/\lambda_{d0}$ *admits at most one solution.*

Assumption S1 allows for continuous or discrete outcome variables. In the case of a discrete variable, the equation $F_{d01}(y) = 1/\lambda_{d0}$ will have no solution, except if there is a point in the support of $Y_{d01}$ at which $F_{d01}(y)$ is exactly equal to $1/\lambda_{d0}$. Therefore, Assumption 1 rules out only very rare scenarios. In the continuous case, the equation $F_{d01}(y) = 1/\lambda_{d0}$ will have a unique solution if, e.g., $F_{d01}$ is strictly increasing on its support.

**Assumption S2** *(Technical conditions for inference with CIC bounds)*

1. $\lambda_{00} \neq 1$ *and* $\underline{q} < \overline{q}$.

2. $\underline{F}_{CIC,d}$ *and* $\overline{F}_{CIC,d}$ *are strictly increasing on* $\underline{\mathcal{S}}_d = [\underline{F}^{-1}_{CIC,d}(\underline{q}), \underline{F}^{-1}_{CIC,d}(\overline{q})]$ *and* $\overline{\mathcal{S}}_d = [\overline{F}^{-1}_{CIC,d}(\underline{q}), \overline{F}^{-1}_{CIC,d}(\overline{q})]$ *respectively. Their derivatives are strictly positive whenever they exist.*

The condition $\underline{q} < \overline{q}$ in Assumption S2 is automatically satisfied when $\lambda_{00} > 1$, because then the bounds are proper cdfs so $\underline{q} = 0$ and $\overline{q} = 1$. When $\lambda_{00} < 1$ and Assumption 9 holds, one can show that it is satisfied when $\lambda_{10} < H_0(\lambda_{00}) - H_0(1 - \lambda_{00})$. The larger the increase of the treatment rate in the treatment group and the smaller the increase in the control group, the more this condition is likely to hold.[4] The strict monotonicity requirement is only a slight reinforcement of Assumption 9. When $\lambda_{00} < 1$, $\underline{F}_{CIC,0}$ and $\overline{F}_{CIC,0}$ satisfy Assumption S2 when $H_0(\lambda_{00}F_{001}) - \lambda_{10}F_{011}$ and $H_0(\lambda_{00}F_{001} + 1 - \lambda_{00}) - \lambda_{10}F_{011}$ have positive derivatives on $\mathcal{S}(Y)$. If $H_0$ is equal to the identity function, this will hold if the ratio of the derivatives of $F_{011}$ and $F_{001}$ is strictly lower than $\frac{\lambda_{00}}{\lambda_{10}}$. Hence, here as well, the larger the increase of the treatment rate in the treatment group and the smaller the increase in the control group, the more this condition is likely to hold. It is possible to derive similar sufficient conditions for Assumption S2 to hold in the three other possible cases ($\underline{F}_{CIC,0}$ and $\overline{F}_{CIC,0}$ when $\lambda_{00} > 1$, $\underline{F}_{CIC,1}$ and $\overline{F}_{CIC,1}$ when $\lambda_{00} < 1$, and $\underline{F}_{CIC,1}$ and $\overline{F}_{CIC,1}$ when $\lambda_{00} > 1$). We refer the reader to the proof of Lemma S6 for more details.

Theorem S3 establishes the asymptotic normality of the estimated bounds of $\Delta$ and $\tau_q$ for $q \in \mathcal{Q}$, with $\mathcal{Q} = (\underline{q}, \overline{q})\backslash\{q_1, q_2\}$ when $\lambda_{00} > 1$ and $\mathcal{Q} = (0, 1)$ when $\lambda_{00} < 1$.

**Theorem S3** *Assume that Model* (1) *and Assumptions 1-2 and 12 hold.*

- *If Assumptions 5 and S1 also hold, then* $(\widehat{\underline{W}}_{TC} - \underline{W}_{TC}, \widehat{\overline{W}}_{TC} - \overline{W}_{TC})$ *are asymptotically normal. Moreover, the bootstrap is consistent for both.*

- *If Assumptions 6-7, 9, 13 and S2 hold, then* $(\widehat{\underline{W}}_{CIC} - \underline{W}_{CIC}, \widehat{\overline{W}}_{CIC} - \overline{W}_{CIC})$ *and* $(\widehat{\underline{\tau}}_q - \underline{\tau}_q, \widehat{\overline{\tau}}_q - \overline{\tau}_q)$*, for* $q \in \mathcal{Q}$*, are asymptotically normal. Moreover, the bootstrap is consistent for both.*

For the CIC bounds, we restrict $q$ to $\mathcal{Q}$ when $\lambda_{00} < 1$ because the estimated bounds on $\tau_q$ are not root-n consistent and asymptotically normal for every $q$. First, the estimated bounds are equal to the true bounds with probability approaching one for $q < \underline{q}$ or $q > \overline{q}$, because basically, the true bounds put mass at the boundaries $\underline{y}$ or $\overline{y}$.[5] Second, the bounds may exhibit kinks at $q_1$ and $q_2$, which also leads to asymptotic non-normality of $\widehat{\underline{\tau}}_q$ and $\widehat{\overline{\tau}}_q$. On the other hand,

[4] Note that this equation is automatically satisfied when $\lambda_{00} = 1$.

[5] A similar conclusion holds if $\underline{y}$ or $\overline{y}$ are estimated rather than known by the researcher: the estimators are n consistent and not asymptotically normal.

when $\lambda_{00} > 1$, asymptotic normality holds for every $q \in (0, 1)$: the bounds on $F_{Y_{11}(d)|S_1}$ are not defective cdfs, and they do not exhibit kinks, except possibly at the boundaries of their support.

Theorem S3 can be used to construct confidence intervals on $\Delta$ and $\tau_q$ as follows. Let us focus on the Wald-TC bounds on $\Delta$, the reasoning being similar for other bounds and parameters. If we know ex-ante that partial identification holds or, equivalently, that $\lambda_{00} \neq 1$, we can follow Imbens & Manski (2004) and use the lower bound of the one-sided confidence interval of level $1-\alpha$ on $\underline{W}_{TC}$ and the upper bound of the one-sided confidence interval of level $1-\alpha$ on $\overline{W}_{TC}$. However, in practice we rarely know ex-ante whether $\lambda_{00} = 1$ or not. This is an important issue, since the estimators and the way confidence intervals are constructed differ in the two cases. To address this issue, we propose a procedure which yields confidence intervals with desired asymptotic coverage in both cases. Let $\widehat{\sigma}_{\lambda_{00}}$ denote an estimator of the variance of $\widehat{\lambda}_{00}$. Our procedure has three steps:

1. Compare $t_{\lambda_{00}} = \left|\frac{\widehat{\lambda}_{00}-1}{\widehat{\sigma}_{\lambda_{00}}}\right|$ to some sequence $(c_n)_{n\in\mathbb{N}}$ satisfying $c_n \to +\infty$ and $\frac{c_n}{\sqrt{n}} \to 0$.
2. If $t_{\lambda_{00}} \leq c_n$, form confidence intervals for $\Delta$ using the point identification results.
3. If $t_{\lambda_{00}} > c_n$, form confidence intervals for $\Delta$ using the partial identification results.

This procedure yields pointwise valid confidence intervals, because comparing $|t_{\lambda_{00}}|$ to $c_n$ instead of a fixed critical value ensures that asymptotically, the probability of conducting inference under the wrong maintained assumption vanishes to 0. An inconvenient of this procedure is that it relies on the choice of a tuning parameter, the sequence $(c_n)_{n\in\mathbb{N}}$. Note that many procedures recently suggested in the moment inequality literature also share this inconvenient (see Andrews & Soares, 2010 or Chernozhukov et al., 2013). Also, it is unclear whether the confidence interval $CI_{1-\alpha}$ resulting from that procedure is uniformly valid, i.e. whether it satisfies

$$\lim_{n\to\infty} \inf_{P\in\mathcal{P}_0} \inf_{\Delta\in[\underline{W}_{TC},\overline{W}_{TC}]} P(\Delta \in CI_{1-\alpha}) \geq 1-\alpha,$$

where $\mathcal{P}_0$ denotes a set of distributions of $(D, G, T, Y)$. Uniformly valid confidence intervals on partially identified parameters have for instance been proposed by Imbens & Manski (2004), Andrews & Soares (2010), Andrews & Barwick (2012), Chernozhukov et al. (2013), and Romano et al. (2014). However, to the best of our knowledge none of the existing procedure applies to our context. The solutions suggested by Imbens & Manski (2004) or Stoye (2009) require that the bounds converge uniformly towards normal distributions. But as our bounds involve the kinked functions $m_1(\lambda_{0d})$ and $M_0(1-\lambda_{0d})$, their estimator is not asymptotically normal when $\lambda_{00} = 1$. The literature on moment inequality models does not apply either. One can for instance show that under Assumptions 1 , 2, and 5, our parameter of interest $\Delta$ satisfies a moment inequality model with four moment inequalities. However, the moments depend on preliminary estimated parameters that once again, do not have an asymptotically normal distribution when $\lambda_{00} = 1$, thus violating the requirements of, e.g., Andrews & Soares (2010) and Andrews & Barwick (2012).

# 3 Supplementary applications

## 3.1 Effects of newspapers on electoral participation in the US

Gentzkow et al. (2011) study the effect of newspapers on electoral participation in the US. They estimate OLS regressions of the change in turnout between consecutive elections in county $c$ on election dummies and the change of the number of daily newspapers available in county $c$. In column 2 of their Table 2, they find that one additional newspaper increases turnout by 0.26 percentage points in US presidential elections from 1872 to 1928. Their regression specification is exactly equivalent to that studied in Theorem S1. We estimate the weights $w_{gt}^b$ in this application, and find that treatment effects in 32% of county $\times$ election cells receive a negative weight, and that negative weights sum up to -0.27. The validity of their coefficient therefore relies on the assumption that the effect of newspapers on turnout is constant over time and across counties.

To avoid relying on that assumption, we use a first estimator inspired from the weighted sum of Wald-DIDs in the first point of Theorem 4.1. As the authors include state-year effects in their specifications, we slightly modify our estimator to also allow for differential trends across states. Our estimator is obtained in five steps. First, for each election the sets of counties $\mathcal{G}_{st}$, $\mathcal{G}_{it}$, and $\mathcal{G}_{dt}$ are respectively defined as counties where the number of newspapers remains stable, increases, and decreases between elections $t-1$ and $t$. Second, we restrict the sample to counties in $\mathcal{G}_{st}$ or $\mathcal{G}_{it}$ and estimate a 2SLS regression of the change in turnout between elections $t-1$ and $t$ on state dummies and the change in the number of newspapers. The instrument for the change in newspapers is a dummy for counties in $\mathcal{G}_{it}$. Let $\beta_{DID}(1,0,t)$ denote the coefficient of the change in newspapers in this regression. Without the state dummies, we would have $\beta_{DID}(1,0,t) = W^*_{DID}(1,0,t)$. Therefore, $\beta_{DID}(1,0,t)$ is a modified version of $W^*_{DID}(1,0,t)$ allowing for state-specific trends. Third, we restrict the sample to counties in $\mathcal{G}_{st}$ or $\mathcal{G}_{dt}$ and estimate a 2SLS regression of the change in turnout between elections $t-1$ and $t$ on state dummies and the change in the number of newspapers. The instrument for the change in newspapers is a dummy for counties in $\mathcal{G}_{dt}$. Here as well, the coefficient of the change in newspapers $\beta_{DID}(-1,0,t)$ is a modified version of $W^*_{DID}(-1,0,t)$ allowing for state-specific trends. Fourth, we estimate the weights $w_t$ and $w_{10|t}$ allowing for state-specific trends. We repeat these steps for each election and our estimator is finally equal to

$$\beta_{DID} = \sum_{t=0}^{16} w_{1872+4t}(w_{10|1872+4t}\beta_{DID}(1,0,1872+4t) + (1-w_{10|1872+4t})\beta_{DID}(-1,0,1872+4t)).$$

This estimator does not rely on any constant treatment effect assumption, because it only uses counties where the number of newspapers is stable as controls.

However, this estimator still requires that the effect of newspapers on turnout do not vary over time (Assumption 4 in the main paper). In this context, this assumption is not warranted. Historians have shown that in the end of the 19th century, alternative ways of communicating information such as radio stations, telegraphic lines, and telephonic lines quickly developed in the US, thus ending the print monopoly of mass media (see White, 2003). This might have reduced the effects of newspapers. In their Table 5, the authors give suggestive evidence of this by showing that their regression coefficients diminish over time. To avoid relying on that assumption, we use a second estimator $\beta_{TC}$. $\beta_{TC}$ closely resembles the weighted sum of Wald-TCs we introduced in the second point of Theorem 4.1, except that we allow for state-specific trends in each of the regressions we estimate to compute this weighted sum.[6][7] On the other end, estimating a Wald-CIC type of estimator while controlling for state-specific trends appears difficult. For each pairs of consecutive elections, there are many states where only few counties had, say, 2 newspapers at both elections. This makes it impossible to estimate the quantile-quantile transforms $Q_d$ within-state.[8] We could estimate a weighted average of Wald-CIC estimators without controlling for state-specific trends, but we prefer to remain as close as possible to the authors' original specification

Results are presented in Table 2 below. $\beta_{DID}$ is close to the estimator in Gentzkow et al. (2011). On the other hand, $\beta_{TC}$ is almost twice as large and is significantly different from their estimator (t-stat=2.05). It is also significantly different from $\beta_{DID}$ at the 10% level (t-stat=1.72). To reconstruct the change in turnout that a county in $\mathcal{G}_{it}$ or $\mathcal{G}_{dt}$ would have experienced if its number of newspapers had not changed, $\beta_{DID}$ uses all counties in the same state and in $\mathcal{G}_{st}$. To reconstruct this counterfactual trend, $\beta_{TC}$ only uses counties in the same state, in $\mathcal{G}_{st}$, and with the same number of newspapers in period $t-1$ as the county in $\mathcal{G}_{it}$ or $\mathcal{G}_{dt}$. The fact that $\beta_{TC}$ and $\beta_{DID}$ substantially differ indicates that among counties in $\mathcal{G}_{st}$, those with different numbers of newspapers experience different evolutions of their turnouts. $\beta_{DID}$ and $\beta_{TC}$ rely on different "common trends" assumptions between counties. But challenging one while defending the other seems difficult as these two assumptions are substantively very close. On the other hand, $\beta_{TC}$ does not require that the effect of newspapers on turnout be constant over time, an assumption

[6] Using directly the two weighted sums we introduced in the first and second points of Theorem 4.1 increases even further the difference between our estimators and that of Gentzkow et al. (2011).

[7] Only 18% of county × election cells have 3 newspapers or more, and only 9% have 4 or more. To estimate the numerators of our Wald-TCs, we group the number of newspapers into 4 categories: 0, 1, 2, and more than 3. Results remain unchanged if we instead group the number of newspapers into 5 categories: 0, 1, 2, 3, and more than 4.

[8] On the other hand, this does not prevent us from estimating the additive shifts $\delta_d$ within-state, which we use in our Wald-TC type of estimator.

which is not warranted in this context as we explained above. We therefore choose $\beta_{TC}$ as our preferred estimator.

*Table S2*: Effect of one additional newspaper on turnout

| | Gentzkow et al. (2011) | $\beta_{DID}$ | $\beta_{TC}$ | OLS |
|---|---|---|---|---|
| Effect of newspapers on turnout | 0.0026 | 0.0031 | 0.0047 | -0.0079 |
| | (0.0009) | (0.0012) | (0.0014) | (0.0007) |
| N | 15627 | 15627 | 15627 | 15627 |

*Notes.* This table reports estimates of the effect of one additional newspaper on turnout. Standard errors are clustered at the district level. For $\beta_{DID}$ and $\beta_{TC}$, clustered standard errors are obtained by block bootstrap.

This application also illustrates that our Wald-TC estimator can be used when only aggregate data are available, provided all units in each group $\times$ period cell share the same value of the treatment, as is the case in Gentzkow et al. (2011). In such instances, our Wald-CIC estimator can also be used if one is ready to assume that Assumptions 1-2 and 6-7 are satisfied with $\overline{Y}_{gt}$ instead of $Y$. On the other hand, when units in the same group $\times$ period cell can have different values of the treatment, one cannot use our Wald-TC and Wald-CIC estimators, because $\delta_d$ and $Q_d$ cannot be estimated from aggregate data. This is for instance the case in Enikolopov et al. (2011). In such instances, authors can still follow our recommendation of finding a control group where treatment is stable and then estimate the Wald-DID.

### 3.2 Effects of a titling program in Peru on labour supply

Between 1996 and 2003, the Peruvian government issued property titles to 1.2 million urban households, the largest titling program targeted to squatters in the developing world. Field (2007) examines the labor market effects of increases in tenure security resulting from the program. To isolate the effect of property rights, the author uses a survey conducted in 2000, and exploits two sources of variation in exposure to the titling program. Firstly, this program took place at different dates in different neighborhoods. In 2000, it had approximately reached 50% of targeted neighborhoods. Secondly, it only impacted squatters, i.e. households without a property title prior to the program. The author can therefore construct four groups of households: squatters in neighborhoods reached by the program before 2000, squatters in neighborhoods reached by the program after 2000, non-squatters in neighborhoods reached by the program before 2000, and non-squatters in neighborhoods reached by the program after 2000. Table 3 presents the share of households with a property title in 2000 in each group.

*Table S*3: Share of households with a property right

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

In Table 5 of her paper, the author uses 2SLS regressions to estimate the effect of having a property right on househods' labor supply. Her dependent variable is the number of hours worked per week by each household. Her explanatory variables are a dummy for squatters, a dummy for neighbourhoods reached before 2000, a dummy for whether the household has a property right, and a rich set of 62 control variables. Her instrument for property rights is the interaction of the squatters and reached before 2000 dummies. Therefore, her estimator is a Wald-DID accounting linearly for the effect of covariates. We revisit her results and compute instead the estimator $\widehat{W}_{CIC}^X$ introduced in Section 5.2 of the main paper, with the same set of covariates. $\widehat{W}_{CIC}^X$ also accounts linearly for the effect of covariates so this estimator is comparable to the author's. As all units in the control group are treated, we cannot estimate exactly $\widehat{W}_{CIC}^X$ but we follow Theorem 3.5 and apply the quantile-quantile transform of treated units in the control group to untreated units in the treatment group. On top of Assumptions 1X-2X and 6X-7X, the validity of this estimator also requires a conditional version of Assumption 10. Her Wald-DID and our Wald-CIC estimator with covariates are respectively equal to 18.07 and 16.17, thus implying that being granted a property title increases the number of hours worked by 16 to 18 hours. The two point estimates are not significantly different (t-stat=1.29). Quantile treatment effects are shown in Figure 3. They are negative and insignificant in the bottom of the distribution of the outcome, and positive and significant in the top. As per our estimates, being granted a property title decreases the first decile of labour supply by 5 hours and increases the 9th decile by 53 hours. These two estimates are significantly different (t-stat=2.21). The best affine approximation to the QTE function has a slope of 74.6 with a standard error of 25.8.[9] Overall, our reanalysis yields a point estimate very similar to the author's for the average effect of property titles, but it also unveils an interesting pattern of heterogeneous effects along the distribution of the outcome.

[9] We estimate the standard error of this slope by bootstrap: in each bootstrap sample, we estimate the QTE and the slope of the best affine approximation to the QTE function.

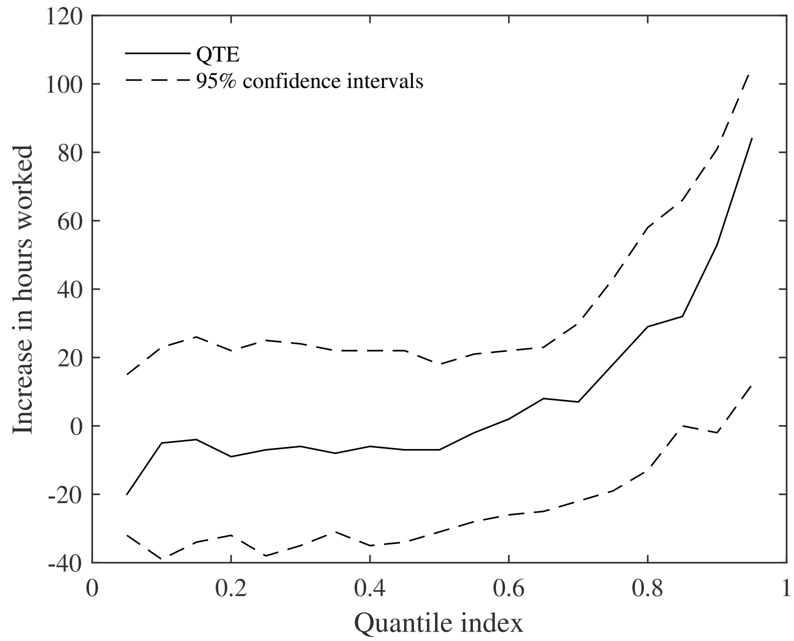


*Figure S*3: Estimated LQTEs on the number of hours worked in Field (2005).

## 4 Supplementary proofs

In this section and in the next, we use the same notations as those used in the proofs of de Chaisemartin & D'Haultfœuille (2015).

**Theorem 3.3 (sharpness of the bounds)**

**Sharpness of the bounds for** $F_{Y_{11}(d)|S_1}(y)$

We only consider the sharpness of $\underline{F}_{CIC,0}$, the reasoning being similar for the upper bound. The proof is also similar and actually simpler for $d = 1$. The corresponding bounds are proper cdf, so we do not have to consider converging sequences of cdf as we do in case b) below.

**a.** $\lambda_{00} > 1$. We show that if Assumptions 2, 7, and 9 hold, then $\underline{F}_{CIC,0}$ is sharp. For that purpose, we construct $\widetilde{h}_0, \widetilde{U}_0, \widetilde{V}$ such that:

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

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

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

(iv) $F_{\widetilde{h}_0(\widetilde{U}_0,1)|G=0,T=1,\widetilde{V}\in[v_{00},v_{01})} = \underline{T}_0$.

First, let

$$\begin{aligned}\widetilde{h}_0(.,0) &= F_{000}^{-1} \circ G_0(\underline{T}_0) \circ F_{001}^{-1},\\ \widetilde{h}_0(.,1) &= F_{001}^{-1}.\end{aligned}$$

Second, let

$$\begin{aligned}\widetilde{U}_0 &= (1-D)\widetilde{h}_0^{-1}(Y,T)\\ &+D(1-T)(1-G)1\{V\in[v_{00},v_{01})\}\widetilde{U}_0^1\\ &+DTG1\{V\in[v_{11},v_{00})\}\widetilde{U}_0^2\\ &+D\left[1-(1-T)(1-G)1\{V\in[v_{00},v_{01})\}-TG1\{V\in[v_{11},v_{00})\}\right]U_0,\end{aligned}$$

where $\widetilde{U}_0^1$ and $\widetilde{U}_0^2$ are two random variables such that $\mathcal{S}(\widetilde{U}_0^1)=\mathcal{S}(\widetilde{U}_0^2)=(0,1)$, and

$$\begin{aligned}F_{\widetilde{U}_0^1|G=0,T=0,V\in[v_{00},v_{01})} &= \underline{T}_0 \circ F_{001}^{-1},\\ F_{\widetilde{U}_0^2|G=1,T=1,V\in[v_{11},v_{00})} &= C_0(\underline{T}_0)\circ F_{001}^{-1}.\end{aligned}$$

$F_{\widetilde{U}_0^1|G=0,T=0,V\in[v_{00},v_{01})}$ is a valid cdf on $(0,1)$ since (i) $\underline{T}_0$ is increasing by Assumption 9 and $F_{001}^{-1}$ is also increasing, (ii) $\lim_{y\to\underline{y}}\underline{T}_0(y)=0$ and $\lim_{y\to\overline{y}}\underline{T}_0(y)=1$ when $\lambda_{00}>1$. $F_{\widetilde{U}_0^2|G=1,T=1,V\in[v_{11},v_{00})}$ is also a valid cdf on $(0,1)$ since (i) $C_0(\underline{T}_0)$ is increasing by Assumption 9 and $F_{001}^{-1}$ is also increasing, (ii) $C_0(\underline{T}_0)(\mathcal{S}(Y))=(0,1)$ when $\lambda_{00}>1$, as per the second point of Lemma S1.

Third, for every $u\in(0,1)$, let

$$\begin{aligned}P_0(u) &= \underline{T}_0\circ F_{001}^{-1}(u),\\ P_1(u) &= C_0(\underline{T}_0)\circ F_{001}^{-1}(u),\\ P_2(u) &= H_0\circ G_0(\underline{T}_0)\circ F_{001}^{-1}(u).\end{aligned}$$

As shown in the proof of Lemma S6 (lower bound, case 2), Assumption 9 ensures that $P_0(u)$, $P_1(u)$, and $P_2(u)$ are non differentiable at only one point. Moreover, using the fact that

$$F_{001}=\frac{1}{\lambda_{00}}G_0\left(\underline{T}_0\right)+\left(1-\frac{1}{\lambda_{00}}\right)\underline{T}_0, \tag{20}$$

$$H_0\circ G_0(\underline{T}_0)=\lambda_{10}F_{011}+(1-\lambda_{10})C_0(\underline{T}_0), \tag{21}$$

and $\underline{T}_0$, $G\left(\underline{T}_0\right)$, and $C_0(\underline{T}_0)$ are increasing under Assumption 9, one can show that

$$\begin{aligned}0\le \left(1-\frac{1}{\lambda_{00}}\right)P_0'(u)\le 1,\\ 0\le \frac{(1-\lambda_{10})\,P_1'(u)}{P_2'(u)}\le 1,\end{aligned}$$

for any $u$ at which $P_0(.)$, $P_1(.)$, and $P_2(.)$ are differentiable, and $P_2'(u) > 0$. Then, let $B_{S_0}$ and $B_{S_1}$ be two Bernoulli random variables such that for every $u \in (0,1)$,

$$P(B_{S_0} = 1|\widetilde{U}_0 = u, D = 0, G = 0, T = 1) = \left(1 - \frac{1}{\lambda_{00}}\right) P_0'(u),$$
$$P(B_{S_1} = 1|\widetilde{U}_0 = u, D = 0, G = 1, T = 0) = \frac{(1 - \lambda_{10}) P_1'(u)}{P_2'(u)},$$

with the convention that $P(B_{S_0} = 1|\widetilde{U}_0 = u, D = 0, G = 0, T = 1)$ and $P(B_{S_1} = 1|\widetilde{U}_0 = u, D = 0, G = 1, T = 0)$ are equal to 0 at the point at which $P_0(u)$, $P_1(u)$, and $P_2(u)$ are not differentiable, and $P(B_{S_1} = 1|\widetilde{U}_0 = u, D = 0, G = 1, T = 0) = 0$ when $P_2'(u) = 0$. The first convention is innocuous as it applies to a 0 Lebesgue measure set. As we shall see later, the second convention is also innocuous, because when $P_2'(u) = 0$, Equation (21) implies that $P_1'(u) = 0$ as well.

Finally, let

$$\begin{aligned}\widetilde{V} &= (1-D)(1-G)T\left[B_{S_0}\widetilde{V}^1 + (1-B_{S_0})\widetilde{V}^2\right] \\ &\quad +(1-D)G(1-T)\left[B_{S_1}\widetilde{V}^3 + (1-B_{S_1})\widetilde{V}^4\right] \\ &\quad + (1-(1-D)\left[(1-G)T + G(1-T)\right])\, V,\end{aligned}$$

where $\widetilde{V}^1$, $\widetilde{V}^2$, $\widetilde{V}^3$, and $\widetilde{V}^4$ are such that $\mathcal{S}(\widetilde{V}^1) = \mathcal{S}(V) \cap [v_{00}, v_{01})$, $\mathcal{S}(\widetilde{V}^2) = \mathcal{S}(V) \cap (-\infty, v_{00})$, $\mathcal{S}(\widetilde{V}^3) = \mathcal{S}(V) \cap [v_{11}, v_{00})$, $\mathcal{S}(\widetilde{V}^4) = \mathcal{S}(V) \cap (-\infty, v_{11})$, and

$$\begin{aligned}f_{\widetilde{V}^1|G=0,T=1,D=0,B_{S_0}=1,\widetilde{U}_0}(v|u) &= f_{V|G=0,T=0,V\in[v_{00},v_{01}),\widetilde{U}_0}(v|u), \\ f_{\widetilde{V}^2|G=0,T=1,D=0,B_{S_0}=0,\widetilde{U}_0}(v|u) &= f_{V|G=0,T=0,V<v_{00},\widetilde{U}_0}(v|u), \\ f_{\widetilde{V}^3|G=1,T=0,D=0,B_{S_1}=1,\widetilde{U}_0}(v|u) &= f_{V|G=1,T=1,V\in[v_{11},v_{00}),\widetilde{U}_0}(v|u), \\ f_{\widetilde{V}^4|G=1,T=0,D=0,B_{S_1}=0,\widetilde{U}_0}(v|u) &= f_{V|G=1,T=1,V<v_{11},\widetilde{U}_0}(v|u).\end{aligned}$$

We shall now show that $(\widetilde{h}_0(.,0), \widetilde{h}_0(.,1), \widetilde{U}_0, \widetilde{V})$ satisfies (i), (ii), (iii), and (iv). By construction, Point (i) is satisfied. Moreover, it follows from Assumption 7 that $\widetilde{h}_0(.,1)$ is strictly increasing on $(0,1)$. Besides, $G_0(\underline{T}_0) \circ F_{001}^{-1}$ is strictly increasing on $(0,1)$ and included between 0 and 1 as shown in the first point of Lemma S1. $F_{000}^{-1}$ is also strictly increasing on $(0,1)$ by Assumption 7. Therefore, $\widetilde{h}_0(.,0)$ is also strictly increasing on $(0,1)$, and Point (ii) is satisfied.

Then, we check Point (iii). We show that it holds in the control group. For that purpose, we use Bayes law to write

$$\begin{aligned}&f_{\widetilde{U}_0,\widetilde{V}|G=0,T=t}(u,v) \\ =\ & P(\widetilde{V} < v_{01}|G=0,T=t)[P(\widetilde{V} < v_{00}|G=0,T=t,\widetilde{V} < v_{01}) f_{\widetilde{U}_0|G=0,T=t,\widetilde{V}<v_{00}}(u) f_{\widetilde{V}|G=0,T=t,\widetilde{V}<v_{00},\widetilde{U}_0}(v|u) \\ &+P(\widetilde{V} \in [v_{00},v_{01})|G=0,T=t,\widetilde{V} < v_{01}) f_{\widetilde{U}_0|G=0,T=t,\widetilde{V}\in[v_{00},v_{01})}(u) f_{\widetilde{V}|G=0,T=t,\widetilde{V}\in[v_{00},v_{01}),\widetilde{U}_0}(v|u)] \\ &+P(\widetilde{V} \geq v_{01}|G=0,T=t) f_{\widetilde{U}_0,\widetilde{V}|G=0,T=t,\widetilde{V}\geq v_{01}}(u,v), \qquad (22)\end{aligned}$$

and we show that all elements in the right-hand side of the previous display are equal for $t = 0$ and $t = 1$.

We first evaluate all of these quantities when $T = 1$. First, it follows from the definition of $\widetilde{V}$ that

$$P(\widetilde{V} < v_{01}|G = 0, T = 1) = p_{0|01}. \tag{23}$$

Then,

$$\begin{aligned} P(\widetilde{U}_0 \leq u|G = 0, T = 1, \widetilde{V} < v_{01}) &= P(\widetilde{U}_0 \leq u|G = 0, T = 1, D = 0) \\ &= P(\widetilde{h}_0^{-1}(Y, 1) \leq u|G = 0, T = 1, D = 0) \\ &= P(Y \leq F_{001}^{-1}(u)|G = 0, T = 1, D = 0) \\ &= u. \end{aligned}$$

Therefore,

$$f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_{01}}(u) = 1.$$

Then, we have, almost everywhere,

$$\begin{aligned} & f_{\widetilde{U}_0,1\{\widetilde{V}\in[v_{00},v_{01})\}|G=0,T=1,\widetilde{V}<v_{01}}(u, 1) \\ = \; & P(\widetilde{V} \in [v_{00}, v_{01})|G = 0, T = 1, \widetilde{V} < v_{01}, \widetilde{U}_0 = u) f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_{01}}(u) \\ = \; & P(B_{S_0} = 1|G = 0, T = 1, D = 0, \widetilde{U}_0 = u) \\ = \; & \left(1 - \frac{1}{\lambda_{00}}\right) P_0'(u). \end{aligned} \tag{24}$$

The second equality follows from the definition of $\widetilde{V}$, and from $f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_{01}}(u) = 1$. Equation (24) and the fact that $P_0'$ is a density imply that

$$P(\widetilde{V} \in [v_{00}, v_{01})|G = 0, T = 1, \widetilde{V} < v_{01}) = 1 - \frac{1}{\lambda_{00}}, \tag{25}$$

$$f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}\in[v_{00},v_{01})}(u) = P_0'(u), \tag{26}$$

and

$$P(\widetilde{V} < v_{00}|G = 0, T = 1, \widetilde{V} < v_{01}) = \frac{1}{\lambda_{00}}, \tag{27}$$

$$f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_{00}}(u) = \lambda_{00} - (\lambda_{00} - 1) P_0'(u). \tag{28}$$

Next, we have

$$\begin{aligned} f_{\widetilde{V}|G=0,T=1,\widetilde{V}\in[v_{00},v_{01}),\widetilde{U}_0}(v|u) &= f_{\widetilde{V}^1|G=0,T=1,D=0,B_{S_0}=1,\widetilde{U}_0}(v|u), \\ &= f_{V|G=0,T=0,V\in[v_{00},v_{01}),\widetilde{U}_0}(v|u), \end{aligned} \tag{29}$$

and

$$\begin{aligned} f_{\widetilde{V}|G=0,T=1,\widetilde{V}<v_{00},\widetilde{U}_0}(v|u) &= f_{\widetilde{V}^2|G=0,T=1,D=0,B_{S_0}=0,\widetilde{U}_0}(v|u) \\ &= f_{V|G=0,T=0,V<v_{00},\widetilde{U}_0}(v|u). \end{aligned} \tag{30}$$

Then, we evaluate all of these quantities when $T = 0$. First, notice that

$$\begin{aligned} P(\widetilde{V} < v_{01}|G = 0, T = 0) &= P(V < v_{01}|G = 0, T = 0) \\ &= P(V < v_{01}|G = 0, T = 1) \\ &= p_{0|01}. \end{aligned} \tag{31}$$

The first equality follows from the definition of $\widetilde{V}$ and the second from the fact $V$ satisfies Assumption 1. One can use similar arguments to show that

$$P(\widetilde{V} \in [v_{00}, v_{01})|G = 0, T = 0, \widetilde{V} < v_{01}) = 1 - \frac{1}{\lambda_{00}}, \tag{32}$$

$$P(\widetilde{V} < v_{00}|G = 0, T = 0, \widetilde{V} < v_{01}) = \frac{1}{\lambda_{00}}. \tag{33}$$

Then, it follows from the definition of $\widetilde{V}$ and $\widetilde{U}_0$ that

$$f_{\widetilde{U}_0|G=0,T=0,\widetilde{V}\in[v_{00},v_{01})}(u) = f_{\widetilde{U}_0^1|G=0,T=0,V\in[v_{00},v_{01})}(u) = P_0'(u). \tag{34}$$

Next,

$$\begin{aligned} P(\widetilde{U}_0 \leq u|G = 0, T = 0, \widetilde{V} < v_{00}) &= P(\widetilde{U}_0 \leq u|G = 0, T = 0, D = 0) \\ &= P(\widetilde{h}_0^{-1}(Y, 0) \leq u|G = 0, T = 0, D = 0) \\ &= P(Y \leq F_{000}^{-1} \circ G_0(\underline{T}_0) \circ F_{001}^{-1}(u)|G = 0, T = 0, D = 0) \\ &= G_0(\underline{T}_0) \circ F_{001}^{-1}(u) \\ &= \lambda_{00}u - (\lambda_{00} - 1)\, P_0(u), \end{aligned}$$

where the last equality follows from (20). This implies that

$$f_{\widetilde{U}_0|G=0,T=0,\widetilde{V}<v_{00}}(u) = \lambda_{00} - (\lambda_{00} - 1)\, P_0'(u). \tag{35}$$

Then, it follows from the definition of $\widetilde{V}$ that

$$f_{\widetilde{V}|G=0,T=0,\widetilde{V}\in[v_{00},v_{01}),\widetilde{U}_0}(v|u) = f_{V|G=0,T=0,V\in[v_{00},v_{01}),\widetilde{U}_0}(v|u), \tag{36}$$

$$f_{\widetilde{V}|G=0,T=0,\widetilde{V}<v_{00},\widetilde{U}_0}(v|u) = f_{V|G=0,T=0,V<v_{00},\widetilde{U}_0}(v|u). \tag{37}$$

Finally,

$$\begin{aligned} f_{\widetilde{U}_0,\widetilde{V}|G=0,T=0,\widetilde{V}\geq v_{01}}(u, v) &= f_{U_0,V|G=0,T=0,V\geq v_{01}}(u, v) \\ &= f_{U_0,V|G=0,T=1,V\geq v_{01}}(u, v) \\ &= f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1,\widetilde{V}\geq v_{01}}(u, v), \end{aligned} \tag{38}$$

where the first and last equality follow from the definition of $(\widetilde{U}_0, \widetilde{V})$, while the second equality follows from the fact $(U_0, V)$ satisfies Assumption 1.

Finally, combining Equation (22) with Equations (23) and (31), (25) and (32), (27) and (33), (26) and (34), (28) and (35), (29) and (36), (30) and (37), and (38), we get that

$$f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1}(u,v) = f_{\widetilde{U}_0,\widetilde{V}|G=0,T=0}(u,v).$$

This shows that (iii) holds in the control group. Showing that it also holds in the treatment group relies on a very similar reasoning, so we skip this part of the proof due to a concern for brevity.

**b.** $\lambda_{00} < 1$**.** The idea is similar as in the previous case. A difference, however, is that when $\lambda_{00} < 1$ and $\overline{y} = +\infty$, $\underline{T}_0$ is not a proper cdf, but a defective one, since $\lim_{y\to+\infty} \underline{T}_0(y) < 1$. As a result, we cannot define a DGP such that $\widetilde{T}_0 = \underline{T}_0$, However, by Lemma S2, there exists a sequence $(\underline{T}_0^k)_k$ of cdf such that $\underline{T}_0^k \to \underline{T}_0$, $G_0(\underline{T}_0^k)$ is an increasing bijection from $\mathcal{S}(Y)$ to $(0,1)$ and $C_0(\underline{T}_0^k)$ is increasing and onto $(0,1)$. We can then construct a sequence of DGP $(\widetilde{h}_0^k(.,0), \widetilde{h}_0^k(.,1), \widetilde{U}_0^k, \widetilde{V}^k)$ such that Points (i) to (iii) listed above hold for every $k$, and such that $\widetilde{T}_0^k = \underline{T}_0^k$. Since $\underline{T}_0^k(y)$ converges to $\underline{T}_0(y)$ for every $y$ in $\mathring{\mathcal{S}}(Y)$, we thus define a sequence of DGP such that $\widetilde{T}_0^k$ can be arbitrarily close to $\underline{T}_0$ on $\mathring{\mathcal{S}}(Y)$ for sufficiently large $k$. Since $C_0(.)$ is continuous, this proves that $\underline{F}_{CIC,0}$ is sharp on $\mathring{\mathcal{S}}(Y)$.

In what follows, we exhibit $\widetilde{h}_0^k(.,0)$ and $\widetilde{h}_0^k(.,1)$ satisfying (i), as well as distributions of $\widetilde{U}_0^k$ for all relevant subpopulations which are a) compatible with the data, b) satisfy (iii), and c) reach the bound. We do not not exhibit $(\widetilde{U}_0^k, \widetilde{V}^k)$ as we did in the previous proof, to avoid repeating twice similar arguments.

Let

$$\widetilde{h}_0^k(.,1) = G_0(\underline{T}_0^k)^{-1}$$
$$\widetilde{h}_0^k(.,0) = F_{000}^{-1}$$

$\widetilde{h}_0^k(.,1)$ is strictly increasing on $(0,1)$ since $G_0(\underline{T}_0^k)$ is an increasing bijection on $(0,1)$ as shown in Lemma S2. $\widetilde{h}_0^k(.,0)$ is strictly increasing on $(0,1)$ under Assumption 7. Therefore, (i) is verified.

Let us consider first the distribution of $\widetilde{U}_0^k$ among untreated observations in the control group in period 1. It follows from Bayes rule that

$$F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_{00}} = \lambda_{00} F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_{01}} + (1-\lambda_{00}) F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}\in[v_{01},v_{00})} \tag{39}$$

Given $\widetilde{h}_0^k(.,1)$, to have $\widetilde{T}_0^k = \underline{T}_0^k$, we must have

$$F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}\in[v_{01},v_{00})} = \underline{T}_0^k \circ G_0(\underline{T}_0^k)^{-1}.$$

This defines a valid cdf since $\underline{T}_0^k$ is a cdf and $G_0(\underline{T}_0^k)^{-1}$ is increasing and onto $\mathcal{S}(Y)$. It can be achieved by constructing $\widetilde{V}$ using an appropriate Bernoulli random variable to split untreated observations in the control group in period 0 between some for which $\widetilde{V} \in [v_{01}, v_{00})$, and some for which $\widetilde{V} < v_{01}$, exactly as we did for $\lambda_{00} > 1$.

Given $\widetilde{h}_0^k(.,1)$, and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k, 1) = Y$ for all observations such that $G = 0, T = 1, \widetilde{V} < v_{01}$, a few computations yield

$$F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_{01}} = F_{001} \circ G_0(\underline{T}_0^k)^{-1}.$$

Plugging the last two equations into (39) finally yields $F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_{00}} = I$, where $I$ denotes the identity function on $[0,1]$.

Now, let us turn to untreated observations in the control group in period 0. Given $\widetilde{h}_0^k(.,0)$, and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k, 0) = Y$ for all observations such that $G = 0, T = 0, \widetilde{V} < v_{00}$, a few computations yield $F_{\widetilde{U}_0^k|G=0,T=0,\widetilde{V}<v_{00}} = I$. Since $Y(0)$ is not observed for observations such that $G = 0, T = 1, \widetilde{V} \in [v_{01}, v_{00})$, the data does not impose any constraint on their $U_0$, so we can set

$$F_{\widetilde{U}_0^k|G=0,T=0,\widetilde{V}\in[v_{01},v_{00})} = \underline{T}_0^k \circ G_0(\underline{T}_0^k)^{-1}.$$

Therefore, the distributions of $\widetilde{U}_0^k|G = 0, T = t, \widetilde{V} < v_{01}$ and $\widetilde{U}_0^k|G = 0, T = t, \widetilde{V} \in [v_{01}, v_{00})$ satisfy (iii).

Then, let us consider untreated observations in the treatment group in period 1. Using the definition of $\widetilde{h}_0^k(.,1)$ and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k, 1) = Y$ for all observations such that $G = 1, T = 1, \widetilde{V} < v_{11}$, one can show after a few computations that

$$F_{\widetilde{U}_0^k|G=1,T=1,\widetilde{V}<v_{11}} = F_{011} \circ G_0(\underline{T}_0^k)^{-1}.$$

Since $Y(0)$ is not observed for observations such that $G = 1, T = 1, \widetilde{V} \in [v_{11}, v_{00})$, the data does not impose any constraint on their $U_0$, so we can set

$$F_{\widetilde{U}_0^k|G=1,T=1,\widetilde{V}\in[v_{11},v_{00})} = C_0(\underline{T}_0^k) \circ G_0(\underline{T}_0^k)^{-1}.$$

This defines a valid cdf, as shown in Points 2 and 3 of Lemma S2.

Finally, let us consider untreated observations in the treatment group in period 0. It follows from Bayes rule that we must have

$$F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_{00}} = \lambda_{10} F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_{11}} + (1 - \lambda_{10}) F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}\in[v_{11},v_{00})}. \tag{40}$$

To satisfy point (iii), we must have

$$F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_{11}} = F_{011} \circ G_0(\underline{T}_0^k)^{-1}.$$

This can be achieved by constructing $\widetilde{V}$ using an appropriate Bernoulli random variable to split untreated observations in the treatment group in period 0 between some for which $\widetilde{V} \in [v_{11}, v_{00})$, and some for which $\widetilde{V} < v_{11}$, exactly as we did for $\lambda_{00} > 1$. Using the definition of $\widetilde{h}_0^k(.,1)$ and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k, 1) = Y$ for all observations such that $G = 0, T = 1, \widetilde{V} < v_{11}$, one can show after a few computations that

$$F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_{00}} = F_{010} \circ F_{000}^{-1}.$$

Plugging the last two equations into (40) finally yields

$$\begin{aligned} F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}\in[v_{11},v_{00})} &= \frac{p_{0|10}F_{010} \circ F_{000}^{-1} - p_{0|11}F_{011} \circ G_0(\underline{T}_0^k)^{-1}}{p_{0|10} - p_{0|11}} \\ &= C_0(\underline{T}_0^k) \circ G_0(\underline{T}_0^k)^{-1}. \end{aligned}$$

Therefore, the distributions of $\widetilde{U}_0^k|G = 1, T = t, \widetilde{V} < v_{11}$ and $\widetilde{U}_0^k|G = 1, T = t, \widetilde{V} \in [v_{11}, v_{00})$ satisfy (iii). This completes the proof when $\lambda_{00} < 1$.

**Sharpness of the bounds for $\Delta$ and $\tau_q$**

We prove that the bounds on $\Delta$ and $\tau_q$ are sharp under Assumption 9. We only focus on the lower bound, the result being similar for the upper bound. The model and data impose no condition on the joint distribution of $(U_0, U_1)$. Hence, by the previous sharpness proof we can rationalize the fact that $(F_{Y_{11}(0)|S_1}, F_{Y_{11}(1)|S_1}) = (\underline{F}_{CIC,0}, \overline{F}_{CIC,1})$ when $\lambda_{00} > 1$. Sharpness of $\Delta$ and $\tau_q$ follows directly. When $\lambda_{00} < 1$, on the other hand, we can only rationalize the fact that $(F_{Y_{11}(0)|S_1}, F_{Y_{11}(1)|S_1}) = (C_{0k}, \overline{F}_{CIC,1})$, where $C_{0k}$ converges pointwise to $\underline{F}_{CIC,0}$. To show the sharpness of the LATE and LQTE, we thus have to prove that $\lim_{k\to\infty} \int y dC_{0k}(y) = \int y d\underline{F}_{CIC,0}(y)$ and $\lim_{k\to\infty} C_{0k}^{-1}(q) = \underline{F}_{CIC,0}^{-1}(q)$.

As for the LATE, we have, by integration by parts for Lebesgue-Stieljes integrals,

$$\int y dC_{0k}(y) = \overline{y} - \int_{\underline{y}}^{\overline{y}} C_{0k} dy = -\int_{\underline{y}}^{0} C_{0k}(y)dy + \int_{0}^{\overline{y}} [1 - C_{0k}(y)]\, dy. \quad (41)$$

We now prove the convergence of each integral in the right-hand side. As shown by Lemma S2, $C_{0k}$ can be defined as $C_{0k} = C_0(\underline{T}_0^k)$ with $\underline{T}_0^k \leq T_0$, $T_0$ denoting $F_{Y_{11}(0)|S_0}$. Because $C_0(T_0) = F_{Y_{11}(0)|S_1}$ and $C_0(.)$ is increasing when $\lambda_{00} < 1$, $C_{0k} \leq F_{Y_{11}(0)|S_1}$. $E(|Y_{11}(0)|\ |S_1) < +\infty$ implies that $\int_{\underline{y}}^0 F_{Y_{11}(0)|S_1}(y)dy < +\infty$. Thus, by the dominated convergence theorem,

$$\lim_{k\to\infty} \int_{\underline{y}}^{0} C_{0k} dy = \int_{\underline{y}}^{0} \underline{F}_{CIC,0}(y)dy < +\infty.$$

Now consider the second integral in (41). If $\overline{y} < +\infty$, we can also apply the dominated convergence theorem: $1 - C_{0k} \leq 1$ implies that $\int_0^{\overline{y}} [1 - C_{0k}(y)]\, dy \to \int_0^{\overline{y}} \left[1 - \underline{F}_{CIC,0}(y)\right] dy$. If $\overline{y} = +\infty$, $\lim_{y\to+\infty} \underline{F}_{CIC,0}(y) = \ell < 1$ so that

$$\int_0^{\overline{y}} \left[1 - \underline{F}_{CIC,0}(y)\right] dy = +\infty.$$

By Fatou's lemma,

$$\liminf \int_0^{\overline{y}} [1 - C_{0k}(y)]\, dy \geq \int_0^{\overline{y}} \left[1 - \underline{F}_{CIC,0}(y)\right] dy = +\infty.$$

Thus, in this case as well the second integral in (41) converges to $\int_0^{\overline{y}} \left[1 - \underline{F}_{CIC,0}(y)\right] dy$. Finally, because $\int_{\underline{y}}^0 C_{0k}(y)dy$ converges to a finite limit, $\int ydC_{0k}(y)$ converges to $\int yd\underline{F}_{CIC,0}(y)$. Hence, the lower bound of $\Delta$ is sharp.

Now, let us turn to $\tau_q$. Following Lemma S2 , we can let $C_{0k} = C_0(\underline{T}_0^k)$, where $\underline{T}_0^k$ and $C_0(\underline{T}_0^k)$ satisfy the three following requirements:

1. $\underline{T}_0^k \geq \underline{T}_0$
2. for all $y_* \in \overset{\circ}{\mathcal{S}}(Y)$, there is a $k \in \mathbb{N}$ such that for every $k' \geq k$, $\underline{T}_0^{k'}(y) = \underline{T}_0(y)$ for all $y \leq y_*$.
3. $C_0(\underline{T}_0^k)$ is increasing.

Suppose first that $y_q \equiv \underline{F}_{CIC,0}^{-1}(q) \in \overset{\circ}{\mathcal{S}}(Y)$. Then point 2 above implies that for all $k$ large enough, $C_{0k}(y) = \underline{F}_{CIC,0}(y)$ for every $y \leq y_q$. This implies that $C_{0k}^{-1}(q) = y_q$. Hence, $C_{0k}^{-1}(q)$ converges to $y_q$. Now suppose that $y_q \notin \overset{\circ}{\mathcal{S}}(Y)$. Given that $\mathcal{S}(Y) = [\underline{y}, \overline{y}]$, $y_q \in \{\underline{y}, \overline{y}\}$. If $y_q = \underline{y}$, $\underline{y} \leq C_{0k}^{-1}(q) \leq \underline{F}_{CIC,0}^{-1}(q)$, where the second inequality follows from the fact that point 1 above implies that $C_{0k} \geq \underline{F}_{CIC,0}$. Therefore, $C_{0k}^{-1}(q) = y_q$. Finally, if $y_q = \overline{y}$, the proof of Lemma S2 shows that there exists a sequence $(y_k)_{k\in\mathbb{N}}$ converging towards $\overline{y}$ such that, for every $k \geq 1$, $C_{0k}(y_k - 1/k) = \underline{F}_{CIC,0}(y_k - 1/k)$. Moreover, by definition, $\underline{F}_{CIC,0}(y_k - 1/k) < q$. Thus, $C_{0k}(y_k - 1/k) < q$, and $\overline{y} \geq C_{0k}^{-1}(q) \geq y_k - 1/k$, where the second inequality holds by point 3 above. Hence, in this case as well, $C_{0k}^{-1}(q)$ converges to $\overline{y}$. This proves that the lower bound of $\tau_q$ is sharp, which completes the proof □

**Theorem S1**

**Proof of 1**

As the first stage regression is fully saturated in $(T, G)$, the predicted value of $D$ from this regression is $E(D|T, G)$. As a result, the second stage is a regression of $Y$ on a constant, the time and group dummies, and $E(D|T, G)$. Then, $\beta = \frac{cov(Y,\widetilde{Z})}{V(\widetilde{Z})}$, where $\widetilde{Z}$ is the residual from an OLS regression of $E(D|T, G)$ on the constant and the group and time dummies. Let $\alpha$, $\alpha_g$, and $\alpha_t$ respectively denote the coefficients of the constant and of the group and time dummies in that regression. We have $V(\widetilde{Z}) = cov(E(D|T, G), \widetilde{Z}) = cov(D, \widetilde{Z})$. The first equality follows from the fact that

$$E(D|T, G) = \alpha + \sum_{g=1}^{\overline{g}} \alpha_g 1\{G = g\} + \sum_{t=1}^{\overline{t}} \alpha_t 1\{T = t\} + \widetilde{Z}$$

and $\widetilde{Z}$ is by construction uncorrelated with the time and group dummies. The second equality follows from the law of iterated expectations. Therefore, $\beta = \frac{cov(Y,\widetilde{Z})}{cov(D,\widetilde{Z})}$.

As $G \perp\!\!\!\perp T$, it follows from the Frisch-Waugh theorem that the $\alpha_t$ are equal to the coefficient of the time dummies in a regression of $E(D|G,T)$ on the constant and the time dummies alone. Therefore,

$$\alpha_t = E(E(D|G,T)|T=t) - E(E(D|G,T)|T=0) = E(D|T=t) - E(D|T=0).$$

Similarly, as $G \perp\!\!\!\perp T$, it follows from the Frisch-Waugh theorem that the $\alpha_g$ are equal to the coefficient of the group dummies in a regression of $E(D|G,T)$ on the constant and the group dummies alone. Therefore,

$$\alpha_g = E(E(D|G,T)|G=g) - E(E(D|G,T)|G=0) = E(D|G=g) - E(D|G=0).$$

Then, we have

$$\widetilde{Z} = E(D|G,T) - \alpha - (E(D|G) - E(D|G=0)) - (E(D|T) - E(D|T=0)).$$

Let us first consider the numerator of $\beta$.

$$\begin{aligned}
cov(Y,\widetilde{Z}) &= cov(Y, E(D|G,T) - E(D|G) - E(D|T)) \\
&= E((E(D|G,T) - E(D|G) - E(D|T) + E(D))E(Y|G,T)) \\
&= \sum_{t=0}^{\bar{t}}\sum_{g=0}^{\bar{g}} P(G=g)P(T=t)(E(D_{gt}) - E(D|G=g) - E(D|T=t) + E(D))E(Y_{gt}) \\
&= \sum_{t=1}^{\bar{t}}\sum_{g=0}^{\bar{g}} P(G=g)P(T=t)(E(D_{gt}) - E(D|G=g) - E(D|T=t) + E(D))\left(E(Y_{gt}) - E(Y_{g0}) - (E(Y_{0t}) - E(Y_{00}))\right) \\
&= \sum_{t=1}^{\bar{t}}\sum_{g=1}^{\bar{g}} P(G=g)P(T=t)(E(D_{gt}) - E(D|G=g) - E(D|T=t) + E(D))\sum_{t'=1}^{t}\sum_{g'=1}^{g} W_{DID}(g',g'-1,t')DID_D(g',g'-1,t') \\
&= \sum_{t=1}^{\bar{t}}\sum_{g=1}^{\bar{g}} W_{DID}(g,g-1,t)DID_D(g,g-1,t)\sum_{t'=t}^{\bar{t}}\sum_{g'=g}^{\bar{g}} P(G=g')P(T=t')(E(D_{g't'}) - E(D|G=g') - E(D|T=t') + E(D)) \\
&= \sum_{t=1}^{\bar{t}}\sum_{g=1}^{\bar{g}} W_{DID}(g,g-1,t)DID_D(g,g-1,t)P(G\geq g)P(T\geq t)\left(E\left(D|G\geq g, T\geq t\right) - E\left(D|G\geq g\right) - E\left(D|T\geq t\right) + E(D)\right).
\end{aligned}$$

Similarly, one can show that

$$cov(D,\widetilde{Z}) = \sum_{t=1}^{\bar{t}}\sum_{g=1}^{\bar{g}} DID_D(g,g-1,t)P(G\geq g)P(T\geq t)\left(E\left(D|G\geq g, T\geq t\right) - E\left(D|G\geq g\right) - E\left(D|T\geq t\right) + E(D)\right).$$

**Proof of 2**

The result directly follows after combining the first point with Theorem 3.1 and rearranging the summations. □

**Theorem S2**

**Proof of 1**

We have $\beta = \frac{cov(Y,\widetilde{Z})}{cov(D,\widetilde{Z})}$, where $\widetilde{Z}$ is the residual from an OLS regression of $T^{**} \times f(G)$ on the constant and the group and time dummies. Let $\alpha$, $\alpha_g$, and $\alpha_t$ respectively denote the coefficients of the constant and of the group and time dummies in that regression. As $G \perp\!\!\!\perp T$, it follows from the Frisch-Waugh theorem that the $\alpha_t$ are equal to the coefficient of the time dummies in a regression of $T^{**} \times f(G)$ on the constant and the time dummies alone. Therefore,

$$\alpha_t = E(T^{**} \times f(G)|T = t) - E(T^{**} \times f(G)|T = 0) = 1\{t \geq t_0\}E(f(G)).$$

Similarly, as $G \perp\!\!\!\perp T$, it follows from the Frisch-Waugh theorem that the $\alpha_g$ are equal to the coefficient of the group dummies in a regression of $T^{**} \times f(G)$ on the constant and the group dummies alone. Therefore,

$$\alpha_g = E(T^{**} \times f(G)|G = g) - E(T^{**} \times f(G)|G = 0) = (f(g) - f(0))E(T^{**}).$$

Then, we have

$$\widetilde{Z} = T^{**}f(G) - \alpha - T^{**}E(f(G)) - (f(G) - f(0))E(T^{**}).$$

Therefore, a few computations yield

$$\beta = \frac{cov(Y, f(G)|T^{**} = 1) - cov(Y, f(G)|T^{**} = 0)}{cov(D, f(G)|T^{**} = 1) - cov(D, f(G)|T^{**} = 0)}.$$

Now, let us consider the numerator.

$$\begin{aligned}
& cov(Y, f(G)|T^{**} = 1) - cov(Y, f(G)|T^{**} = 0) \\
= \; & E((f(G) - E(f(G)))Y|T^{**} = 1) - E((f(G) - E(f(G)))Y|T^{**} = 0) \\
= \; & E((f(G) - E(f(G)))E(Y|G, T^{**} = 1)|T^{**} = 1) - E((f(G) - E(f(G)))E(Y|G, T^{**} = 0)|T^{**} = 0) \\
= \; & \sum_{g'=1}^{\bar{g}} P(G = g'|T^{**} = 1)(f(g') - E(f(G)))E(Y|G = g', T^{**} = 1) \\
- \; & \sum_{g'=1}^{\bar{g}} P(G = g'|T^{**} = 0)(f(g') - E(f(G)))E(Y|G = g', T^{**} = 0) \\
= \; & \sum_{g'=1}^{\bar{g}} P(G = g')(f(g') - E(f(G)))(E(Y|G = g', T^{**} = 1) - E(Y|G = g', T^{**} = 0)) \\
= \; & \sum_{g'=2}^{\bar{g}} P(G = g')(f(g') - E(f(G)))DID_Y^{**}(g', 0) \\
= \; & \sum_{g'=2}^{\bar{g}} P(G = g')(f(g') - E(f(G))) \sum_{g=2}^{g'} DID_Y^{**}(g, g - 1)
\end{aligned}$$

Hence,

$$
\begin{aligned}
& cov(Y, f(G)|T^{**} = 1) - cov(Y, f(G)|T^{**} = 0) \\
= & \sum_{g'=2}^{\bar{g}} P(G = g')(f(g') - E(f(G))) \sum_{g=2}^{g} W_{DID}^{**}(g, g-1) DID_D^{**}(g, g-1) \\
= & \sum_{g=2}^{\bar{g}} W_{DID}^{**}(g, g-1) DID_D^{**}(g, g-1) \sum_{g'=g}^{\bar{g}} P(G = g')(f(g') - E(f(G))) \\
= & \sum_{g=2}^{\bar{g}} W_{DID}^{**}(g, g-1) DID_D^{**}(g, g-1) P(G \geq g)(E(f(G)|G \geq g) - E(f(G))).
\end{aligned}
$$

Similarly, one can show that

$$
\begin{aligned}
& cov(D, f(G)|T^{**} = 1) - cov(D, f(G)|T^{**} = 0) \\
= & \sum_{g=2}^{\bar{g}} DID_D^{**}(g, g-1) P(G \geq g)(E(f(G)|G \geq g) - E(f(G))).
\end{aligned}
$$

**Proof of 2**

The result directly follows after combining the first point with Theorem 3.1 and rearranging the summations. □

**Theorem S3**

We let hereafter $\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_{00}, \lambda_{10}, \lambda_{01}, \lambda_{11})$.

**Proof of 1**

We already showed in the proof of Theorem 5.1 that each term of the bounds, except $\int y d\widehat{\overline{F}}_{d10}(y)$ and $\int y d\widehat{\underline{F}}_{d10}(y)$, could be linearized. Therefore, it suffices to prove that these integrals can be linearized as well. Let us focus on $\int y d\widehat{\overline{F}}_{d10}(y)$, as the reasoning is similar for the other. An integration by part yields

$$
\begin{aligned}
& \int y d\widehat{\overline{F}}_{d10}(y) - \int y d\overline{F}_{d10}(y) \\
= & -\int_{\underline{y}}^{\overline{y}} \left[\widehat{\overline{F}}_{d10}(y) - \overline{F}_{d10}(y)\right] dy \\
= & -\int_{\underline{y}}^{\overline{y}} \left[m_1\left(\widehat{\lambda}_{0d}\widehat{F}_{Y_{d01}}(y)\right) - m_1\left(\lambda_{0d}F_{Y_{d01}}(y)\right)\right] dy + (\overline{y} - \underline{y})\left[m_1\left(\widehat{\lambda}_{0d}\right) - m_1\left(\lambda_{0d}\right)\right].
\end{aligned}
$$

By assumption, the equation $\lambda_{0d}F_{Y_{d01}}(y) = 1$ admits at most one solution. Hence, by Point 2 of Lemma 6 and the chain rule, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} m_1\left[\lambda_{0d}F_{Y_{d01}}(y)\right] dy + (\overline{y} - \underline{y}) m_1\left(\lambda_{0d}\right)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. The result then follows from Lemma 4, the functional delta method, and the functional delta method for the bootstrap.

**Proof of 2**

Let $\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_{00}, \lambda_{10}, \lambda_{01}, \lambda_{11})$. By Lemma 6, for $d \in \{0,1\}$ and $q \in \mathcal{Q}$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \underline{F}_{CIC,d}(y)dy$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \overline{F}_{CIC,d}(y)dy$, $\theta \mapsto \overline{F}_{CIC,d}^{-1}(q)$, and $\theta \mapsto \underline{F}_{CIC,d}^{-1}(q)$ are Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. Because $\underline{\Delta} = \int_{\mathcal{S}(Y)} \underline{F}_{CIC,0}(y) - \overline{F}_{CIC,1}(y)dy$, $\underline{\Delta}$ is also a Hadamard differentiable function of $\theta$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. The same reasoning applies for $\overline{\Delta}$, and for $\underline{\tau_q}$ and $\overline{\tau_q}$ for every $q \in \mathcal{Q}$. The result follows as previously $\square$

# 5 Technical lemmas

## 5.1 Lemmas related to identification

**Lemma S1** *Assume Assumptions 7 and 9 hold, and $\lambda_{0d} > 1$. Then:*

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

**Proof:** we only prove the result for $d = 0$, the reasoning being similar otherwise. One can show that when $\lambda_{00} > 1$,

$$G_0(\underline{T}_0) = \min\left(\lambda_{00}F_{001}, \max\left(\lambda_{00}F_{001} + (1 - \lambda_{00}), H_0^{-1} \circ (\lambda_{10}F_{011})\right)\right). \tag{42}$$

By Assumption 7, $\lambda_{10}F_{011}$ is strictly increasing. Moreover, $\mathcal{S}(Y_{10}|D = 0) = \mathcal{S}(Y_{00}|D = 0)$ implies that $H_0^{-1}$ is strictly increasing on $[0,1]$. Consequently, $H_0^{-1} \circ (\lambda_{10}F_{011})$ is strictly increasing on $\mathcal{S}(Y)$ since $\lambda_{10} < 1$. Therefore, $G_0(\underline{T}_0)$ is strictly increasing on $\mathcal{S}(Y)$ as a composition of the max and min of strictly increasing functions, which in turn implies that $G_0(\underline{T}_0) \circ F_{001}^{-1}$ is strictly increasing on $[0,1]$. Moreover, it is easy to see that since $\mathcal{S}(Y_{1t}|D = 0) = \mathcal{S}(Y_{0t}|D = 0)$,

$$\begin{aligned}
\lim_{y \to \underline{y}} H_0^{-1} \circ (\lambda_{10}F_{011}) \circ F_{001}^{-1}(y) &= 0, \\
\lim_{y \to \overline{y}} H_0^{-1} \circ (\lambda_{10}F_{011}) \circ F_{001}^{-1}(y) &\leq 1.
\end{aligned}$$

Hence, by Equation (42),

$$\lim_{y \to \underline{y}} G_0(\underline{T}_0)(y) = 0, \ \lim_{y \to \overline{y}} G_0(\underline{T}_0)(y) = 1. \tag{43}$$

Finally, $G_0(\underline{T}_0) \circ F_{001}^{-1}$ is also continuous by Assumption 7, as a composition of continuous functions. Point 1 then follows, by the intermediate value theorem.

Now, we have

$$C_0(\underline{T}_0) = \frac{p_{0|10}F_{010} \circ F_{001}^{-1} \circ G_0(\underline{T}_0) - p_{0|11}F_{011}}{p_{0|10} - p_{0|11}}.$$

(43) implies that $G_0(\underline{T}_0)$ is a cdf. Hence, by Assumption 7,

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

Moreover, $C_0(\underline{T}_0)$ is increasing by Assumption 9. Combining this with Assumption 7 yields Point 2, since $C_0(\underline{T}_0)$ is continuous by Assumption 7 once more □

**Lemma S2** *Suppose Assumptions 7 and 9 hold, $p_{0|g0} > 0$ for $g \in \{0;1\}$ and $\lambda_{00} < 1$. Then there exists a sequence of cdf $\underline{T}_0^k$ such that*

1. *$\underline{T}_0^k(y) \to \underline{T}_0(y)$ for all $y \in \overset{\circ}{\mathcal{S}(Y)}$;*
2. *$G_0(\underline{T}_0^k)$ is an increasing bijection from $\mathcal{S}(Y)$ to $[0,1]$;*
3. *$C_0(\underline{T}_0^k)$ is increasing and onto $[0,1]$.*

*The same holds for the upper bound.*

**Proof:** we consider a sequence $(y_k)_{k\in\mathbb{N}}$ converging to $\overline{y}$ and such that $y_k < \overline{y}$. Since $y_k < \overline{y}$, we can also define a strictly positive sequence $(\eta_k)_{k\in\mathbb{N}}$ such that $y_k + \eta_k < \overline{y}$. By Assumption 9, $H_0$ is continuously differentiable. Moreover,

$$H_0' = \frac{F_{010}' \circ F_{000}^{-1}}{F_{000}' \circ F_{000}^{-1}}$$

is strictly positive on $\mathcal{S}(Y)$ under Assumption 9. $F_{011}'$ is also strictly positive on $\mathcal{S}(Y)$ under Assumption 9. Therefore, using a Taylor expansion of $H_0$ and $F_{011}$, it is easy to show that there exists constants $A_{1k} > 0$ and $A_{2k} > 0$ such that for all $y < y' \in [y_k, y_k + \eta_k]^2$,

$$H_0(y') - H_0(y) \ \geq \ A_{1k}(y' - y), \tag{44}$$
$$F_{011}(y') - F_{011}(y) \ \leq \ A_{2k}(y' - y). \tag{45}$$

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

$$\varepsilon_k = \min\left(\eta_k, \frac{A_{1k}(1-\lambda_{00})\left(T_0(y_k) - \underline{T}_0(y_k)\right)}{\lambda_{10}A_{2k}}\right). \tag{46}$$

Note that as shown in (26), since $\lambda_{00} < 1$, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ implies that we must have

$$\underline{T}_0 \leq T_0,$$

which implies in turn that $\varepsilon_k \geq 0$. Consequently, since $0 \leq \varepsilon_k \leq \eta_k$, inequalities (44) and (45) also hold for $y < y' \in [y_k, y_k + \varepsilon_k]^2$.

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

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

For every $k$ such that $\varepsilon_k = 0$, let

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

Then, we verify that $\underline{T}_0^k$ defines a sequence of cdf which satisfy Points 1, 2 and 3. Under Assumption 9, $\underline{T}_0(y)$ is increasing, which implies that $\underline{T}_0^k(y)$ is increasing on $(\underline{y}, y_k)$. Since $T_0(y)$ is a cdf, $\underline{T}_0^k(y)$ is also increasing on $(y_k + \varepsilon_k, \overline{y})$. Finally, it is easy to check that when $\varepsilon_k > 0$, $\underline{T}_0^k(y)$ is increasing on $[y_k, y_k + \varepsilon_k]$. $\underline{T}_0^k$ is continuous on $(\underline{y}, y_k)$ and $(y_k + \varepsilon_k, \overline{y})$ under Assumption 7. It is also continuous at $y_k$ and $y_k + \varepsilon_k$ by construction. This proves that $\underline{T}_0^k(y)$ is increasing on $\mathcal{S}(Y)$. Moreover,

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

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

$G_0(\underline{T}_0^k) = \lambda_{00} F_{001} + (1 - \lambda_{00}) \underline{T}_0^k$ is strictly increasing and continuous as a sum of the strictly increasing and continuous function $\lambda_{00} F_{001}$ and an increasing and continuous function. Moreover, $G_0(\underline{T}_0^k)$ tends to 0 (resp. 1) when $y$ tends to $\underline{y}$ (resp. to $\overline{y}$). Point 2 follows by the intermediate value theorem.

Finally, let us show Point 3. Because $G_0(\underline{T}_0^k)$ is a continuous cdf, $C_0(\underline{T}_0^k)$ is also continuous and converges to 0 (resp. 1) when $y$ tends to $\underline{y}$ (resp. to $\overline{y}$). Thus, the proof will be completed if we show that $C_0(\underline{T}_0^k)$ is increasing. By Assumption 9, $C_0(\underline{T}_0^k)$ is increasing on $(\underline{y}, y_k)$. Moreover, since $F_{Y_{11}(0)|S_1} = C_0(T_0)$, $C_0(\underline{T}_0^k)$ is also increasing on $(y_k + \varepsilon_k, \overline{y})$. Finally, when $\varepsilon_k > 0$, we have that for all $y < y' \in [y_k, y_k + \varepsilon_k]^2$,

$$\begin{aligned} & H_0(\lambda_{00} F_{001}(y') + (1 - \lambda_{00}) \underline{T}_0^k(y')) - H_0(\lambda_{00} F_{001}(y) + (1 - \lambda_{00}) \underline{T}_0^k(y)) \\ \geq \; & A_{1k}(1 - \lambda_{00}) \left( \underline{T}_0^k(y') - \underline{T}_0^k(y) \right) \\ \geq \; & \frac{A_{1k}(1 - \lambda_{00}) \left( T_0(y_k) - \underline{T}_0(y_k) \right)}{\varepsilon_k} (y' - y) \\ \geq \; & \lambda_{10} A_{2k} (y' - y) \\ \geq \; & \lambda_{10} \left( F_{011}(y') - F_{011}(y) \right), \end{aligned}$$

where the first inequality follows by (44) and $F_{001}(y') \geq F_{001}(y)$, the second by the definition of $\underline{T}_0^k$ and $T_0(y_k+\varepsilon_k) \geq T_0(y_k)$, the third by (46) and the fourth by (45). This implies that $C_0(\underline{T}_0^k)$ is increasing on $[y_k, y_k+\varepsilon_k]$, since

$$C_0(\underline{T}_0^k) = \frac{H_0(\lambda_{00}F_{001} + (1-\lambda_{00})\underline{T}_0^k) - \lambda_{10}F_{011}}{1-\lambda_{10}}.$$

It is easy to check that under Assumption 7 $C_0(\underline{T}_0^k)$ is continuous on $\mathcal{S}(Y)$. This completes the proof □

## 5.2 Lemmas related to inference

In the following lemmas, we let, for any functional $R$, $dR_F$ denote the Hadamard differential of $R$ taken at $F$. Whenever it exists, this differential is the continuous linear form satisfying

$$dR_F(h) = \lim_{t\to 0} \frac{R(F+th_t) - R(F)}{t}, \text{ for any } h_t \text{ s.t. } ||h_t - h||_\infty \to 0.$$

In absence of ambiguity, we let the point at which the differential is taken implicit and simply denote it by $dR$. In addition to the sets $\mathcal{C}^0(\Theta)$ and $\mathcal{C}^1(\Theta)$, we also denote by $\mathcal{D}(\Theta)$ (resp. $\mathcal{D}_c(\Theta)$) the set of bounded càdlàg (resp. cdfs) functions on $\Theta$. Once more, $\Theta$ is left implicit when it is $\mathcal{S}(Y)$.

Also, for any $(r,k) \in \mathbb{N}^*$, $u = (u_1, ..., u_r) \in \mathbb{R}^r$ and any function $h = (h_1, ...h_k)'$ from $\mathbb{R}^r$ to $\mathbb{R}^k$, let $\|u\|_1 = \sum_{j=1}^r |u_j|$ and $\|h\|_\infty = \max_{j=1,...,k} \sup_{x\in\mathbb{R}^r} |h_j(x)|$ denote the usual $L^1$ norm of $u$ and the supremum norm of $h$, respectively. The following inequality on ratios is used repeatedly in the proofs of Theorems 5.1 and 5.2. It is probably well-known but we prove it for the sake of completeness.

**Lemma S3** *Let $(x_1, y_1)$ and $(x_2, y_2)$ be such that $y_2 \geq C > 0$ and $\max(|x_1-x_2|, |y_1-y_2|) \leq C/2$. Then*

$$\left| \frac{x_1}{y_1} - \frac{x_2}{y_2} - \frac{1}{y_2}\left(x_1 - x_2 - \frac{x_2}{y_2}(y_1-y_2)\right)\right| \leq \frac{2(1+|x_2/y_2|)}{C^2}\max(|x_1-x_2|, |y_1-y_2|)^2.$$

**Proof:**

First, some algebra shows that

$$\frac{x_1}{y_1} - \frac{x_2}{y_2} - \frac{1}{y_2}\left(x_1 - x_2 - \frac{x_2}{y_2}(y_1-y_2)\right) = \frac{y_1-y_2}{y_2^2}\left[(x_2-x_1) + \frac{x_1}{y_1}(y_1-y_2)\right].$$

As a result,

$$\left| \frac{x_1}{y_1} - \frac{x_2}{y_2} - \frac{1}{y_2}\left(x_1 - x_2 - \frac{x_2}{y_2}(y_1-y_2)\right)\right| \leq \frac{1+|x_1/y_1|}{C^2}\max(|x_1-x_2|, |y_1-y_2|)^2.$$

Besides, $y_1 \geq y_2 - |y_1 - y_2| \geq C/2$. Thus,

$$\left|\frac{x_1}{y_1} - \frac{x_2}{y_2}\right| \leq \frac{|x_2||y_2 - y_1|}{y_1 y_2} + \frac{|x_1 - x_2|}{y_1} \leq \frac{C}{2y_1}\left(\frac{|x_2|}{y_2} + 1\right) \leq 1 + |x_2/y_2|.$$

The triangular inequality then yields

$$\left|\frac{x_1}{y_1} - \frac{x_2}{y_2} - \frac{1}{y_2}\left(x_1 - x_2 - \frac{x_2}{y_2}(y_1 - y_2)\right)\right| \leq \frac{2(1 + |x_2/y_2|)}{C^2}\max(|x_1 - x_2|, |y_1 - y_2|)^2_\square$$

The following lemma is used to establish the asymptotic normality of the Wald-CIC estimator in the proof of Theorem 5.1.

**Lemma S4** *Suppose that* $p_{d|g0} > 0$ *for* $(d, g) \in \{0, 1\}^2$ *and let*

$$\theta = (F_{000}, F_{001}, ..., F_{111}, \lambda_{00}, \lambda_{10}, \lambda_{01}, \lambda_{11})$$

*and*

$$\widehat{\theta} = (\widehat{F}_{000}, \widehat{F}_{001}, ..., \widehat{F}_{111}, \widehat{\lambda}_{00}, \widehat{\lambda}_{10}, \widehat{\lambda}_{01}, \widehat{\lambda}_{11}).$$

*Then*

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

*where* $\mathbb{G}$ *denotes a gaussian process defined on* $\mathcal{S}(Y)^8 \times \{0\}^4$*. Moreover,* $\mathbb{G}$ *is continuous in its* $k$*-th component (*$k \in \{1, ..., 8\}$ *if the corresponding* $F_{dgt}$ *is continuous.*[10] *Finally, the bootstrap is consistent for* $\widehat{\theta}$*.*

**Proof:** let $\mathbb{G}_n$ denote the standard empirical process. We prove the result for

$$\eta = (F_{000}, F_{001}, ..., F_{111}, p_{1|00}, p_{1|01}, p_{1|10}, p_{1|11})$$

instead of $\theta$. The result on $\theta$ then follows as $(\lambda_{00}, \lambda_{10}, \lambda_{01}, \lambda_{11})$ is a smooth function of $(p_{1|00}, p_{1|01}, p_{1|10}, p_{1|11})$. For any $(y, d, g, t) \in (\mathcal{S}(Y) \cup \{+\infty\}) \times \{0, 1\}^3$, let

$$f_{dgty}(Y, D, G, T) = \frac{\mathbb{1}\{D = d\}\mathbb{1}\{G = g\}\mathbb{1}\{T = t\}\left[\mathbb{1}\{Y \leq y\} - F_{dgt}(y)\right]}{p_{dgt}},$$
$$f_{gt}(Y, D, G, T) = \mathbb{1}\{G = g\}\mathbb{1}\{T = t\}\left[\mathbb{1}\{D = 1\} - p_{1|gt}\right]/p_{gt}.$$

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

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

[10] Formally, the link between $(d, g, t)$ and $k$ is $k = 1 + t + 2g + 4t$.

Similarly, $\sqrt{n}\left(\widehat{p}_{1|gt} - p_{1|gt}\right) = \mathbb{G}_n f_{g,t}\left(1 + o_P(1)\right)$. Hence, letting

$$f_y = (f_{000y}, ..., f_{111y}, f_{00}, f_{01}, f_{10}, f_{11})',$$

we obtain $\sqrt{n}\left(\widehat{\eta} - \eta\right) = \mathbb{G}_n f_y\left(1 + o_P(1)\right)$. Weak convergence of the left-hand side to a gaussian process follows because each class $\{f_{dgty} : y \in \mathcal{S}(Y)\}$ is Donsker. Moreover, remark that $\sqrt{n_{dgt}}\left(\widehat{F}_{dgt}(y) - F_{dgt}(y)\right)$ is the standard empirical process on the sample $\mathcal{I}_{dgt}$ of random size $n_{dgt}$. Therefore (see, e.g. Theorem 3.5.1 in van der Vaart & Wellner, 1996), it converges in distribution to a process $B \circ F_{dgt}$, where $B$ is a Brownian bridge. Hence, continuity follows as long as $F_{dgt}$ is continuous.

Now let us turn to the bootstrap. Observe that

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

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

The next two lemmas allows us to use the functional delta method for the CIC estimators of average and quantile treatment effects, both in the point and partially identified case, with and without covariates.

**Lemma S5** *1. Let $R_1(F_1, F_2, F_3, F_4, \lambda, \mu) = \frac{\mu F_4 - F_1 \circ F_2^{-1} \circ q_1(F_3,\lambda)}{\mu - 1}$ and $R_2(F_1, F_2, F_3, F_4, \lambda, \mu) = \frac{\mu F_4 - F_1 \circ F_2^{-1} \circ q_2(F_3,\lambda)}{\mu - 1}$, with $q_1(F_3, \lambda) = \lambda F_3$ and $q_2(F_3, \lambda) = \lambda F_3 + 1 - \lambda$. $R_1$ and $R_2$ are Hadamard differentiable at any $(F_{10}, F_{20}, F_{30}, F_{40}, \lambda_{00}, \lambda_{10}) \in (\mathcal{C}^1)^4 \times [0, \infty) \times ([0, \infty)\backslash\{1\})$, tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. Moreover, $dR_1\left((\mathcal{C}^0)^4 \times \mathbb{R}^2\right)$ and $dR_2\left((\mathcal{C}^0)^4 \times \mathbb{R}^2\right)$ are included in $\mathcal{C}^0$.*

*2. Let $R_3(F_1) = \int_{\underline{y}}^{\overline{y}} m_1(F_1)(y)dy$ and $R_4(F_1, F_2) = \int_{\underline{y}}^{\overline{y}} F_2(m_1(F_1))(y)dy$. Tangentially to $\mathcal{C}^0$, $R_3$ is Hadamard differentiable at any $F_{10} \in \mathcal{D}_c$ and the equation $F_{10}(y) = 1$ admits at most one solution on $\overset{\circ}{\mathcal{S}}(Y)$. Tangentially to $(\mathcal{C}^0)^2$, $R_4$ is Hadamard differentiable at any $(F_{10}, F_{20})$ such that $F_{10}$ satisfies the same conditions as for $R_3$ and $F_{20}$ is continuously differentiable on $[0, 1]$. The same holds if we replace $m_1$ (and the equation $F_{10}(y) = 1$) by $M_0$ (and $F_{10}(y) = 0$).*

*3. Let $R_4(F, Q_{1|X}, Q_{2|X}, Q_{3|X}) = \int m^D_{10}(x)\int_0^1 Q_{1|X}\{Q^{-1}_{2|X}[Q_{3|X}(u|x)|x]|x\}dudF(x)$, where $m^D_{10}(x) = E(D_{10}|X = x)$. Then, tangentially to $\mathcal{C}^0(\mathcal{S}(X)) \times \mathcal{C}^0((0, 1) \times \mathcal{S}(X))^3$, $R_4$ is Hadamard differentiable at any $(F_0, Q_{10|X}, Q_{20|X}, Q_{30|X})$ such that $F_0 \in \mathcal{D}_c(\mathcal{S}(X))$, $(Q_{1|X}(.|x), Q_{2|X}(.|x), Q_{3|X}(.|x)) \in$*

$(\mathcal{C}^1(0,1))^3$ *for all* $x \in \mathcal{S}(X)$ *and* $G(x) = m_{10}^D(x)\int_0^1 Q_{10|X}\{Q_{20|X}^{-1}[Q_{30|X}(u|x)|x]|x\}du$ *is of bounded variation. Moreover, for all* $h_1$ *such that* $h_1(\inf \mathcal{S}(X)) = h_1(\sup \mathcal{S}(X)) = 0$,

$$\begin{aligned} dR_4(h_1,h_2,h_3,h_4) = & \int m_{10}^D(x)\int_0^1 \left\{h_2\left[Q_{20|X}^{-1}[Q_{30|X}(u|x)],x\right] + \partial_u\left[Q_{10|X}\circ Q_{20|X}^{-1}\right][(Q_{30|X}(u|x)|x)|x]\right. \\ & \left.\times\left[-h_3\left[Q_{20|X}^{-1}[Q_{30|X}(u|x)],x\right] + h_4(u,x)\right]\right\}dudF_0(x) - \int h_1(x)dG(x). \end{aligned}$$

**Proof of 1.** We first prove that $\phi_1(F_1,F_2,F_3) = F_1\circ F_2^{-1}\circ F_3$ is Hadamard differentiable at $(F_{10},F_{20},F_{30}) \in (\mathcal{C}^1)^3$. Because $(F_{10},F_{20}) \in (\mathcal{C}^1)^2$, the function $\phi_2 : (F_1,F_2,F_3) \mapsto (F_1\circ F_2^{-1}, F_3)$ is Hadamard differentiable at $(F_{10},F_{20},F_{30})$ tangentially to $\mathcal{D}\times\mathcal{C}^0\times\mathcal{D}$ (see, e.g., van der Vaart & Wellner, 1996, Problem 3.9.4), and therefore tangentially to $(\mathcal{C}^0)^3$. Moreover computations show that its derivative at $(F_{10},F_{20},F_{30})$ satisfies

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

This shows that $d\phi_2\left((\mathcal{C}^0)^3\right) \subseteq (\mathcal{C}^0)^2$.

Then, the composition function $\phi_3 : (U,V) \mapsto U\circ V$ is Hadamard differentiable at any $(U_0,V_0) \in (\mathcal{C}^1)^2$, tangentially to $\mathcal{C}^0\times\mathcal{D}$ (see, e.g., van der Vaart & Wellner, 1996, Lemma 3.9.27), and therefore tangentially to $(\mathcal{C}^0)^2$. It is thus Hadamard differentiable at $(F_{10}\circ F_{20}^{-1}, F_{30})$, and one can show that $d\phi_3\left((\mathcal{C}^0)^2\right) \subseteq \mathcal{C}^0$. Thus, by the chain rule (see van der Vaart & Wellner, 1996, Lemma 3.9.3), $\phi_1 = \phi_3\circ\phi_2$ is also Hadamard differentiable at $(F_{10},F_{20},F_{30})$ tangentially to $(\mathcal{C}^0)^3$, and $d\phi_1\left((\mathcal{C}^0)^3\right) \subseteq \mathcal{C}^0$.

Finally, because $q_1(F_3,\lambda)$ is a smooth function of $F_3$ and $\lambda$, and $R_1$ is a smooth function of $(\phi_1(F_1,F_2,q_1(F_3,\lambda)),F_4,\mu)$, it is also Hadamard differentiable at $(F_{10},F_{20},F_{30},F_{40},\lambda_{00},\lambda_{10})$ tangentially to $(\mathcal{C}^0)^4\times\mathbb{R}^2$, and $dR_1\left((\mathcal{C}^0)^4\times\mathbb{R}^2\right) \subseteq \mathcal{C}^0$.

**Proof of 2.** We only prove the result for $R_4$ and $m_1$, the reasoning being similar (and simpler) for $R_3$ and $M_0$. For any collections of functions $(h_{t1})$ and $(h_{t2})$ in $\mathcal{C}^0$, respectively converging uniformly towards $h_1$ and $h_2$ in $\mathcal{C}^0$, we have

$$\begin{aligned} \frac{R_4(F_{10}+th_{t1},F_{20}+th_{t2}) - R_4(F_{10},F_{20})}{t} = & \int_{\underline{y}}^{\overline{y}} h_{t2}\circ m_1(F_{10}+th_{t1})(y)dy \\ & + \int_{\underline{y}}^{\overline{y}} \frac{F_{20}\circ m_1(F_{10}+th_{t1}) - F_{20}\circ m_1(F_{10})}{t}(y)dy. \end{aligned}$$

Consider the first integral $I_1$.

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

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

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

Moreover, for $t$ small enough,

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

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

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

$$\frac{F_{20} \circ m_1(F_{10}(y)+th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t} \to F'_{20}(F_{10}(y))h_1(y)\mathbb{1}\{y < \underline{y}_1\}. \qquad (47)$$

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

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

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

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

so that (47) holds for $y < \underline{y}_1$. Now, if $\overline{y} > y > \underline{y}_1$, $F_{10}(y) > 1$ because $F_{10}$ is increasing. Thus, for $t$ small enough, $F_{10}(y)+th_{t1}(y) > 1$. Therefore, for $t$ small enough,

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

so that (47) holds as well. Thus, (47) holds almost everywhere.

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

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

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

$$\int_{\underline{y}}^{\overline{y}} \frac{F_{20} \circ m_1(F_{10} + th_{t1}) - F_{20} \circ m_1(F_{10})}{t}(y)dy \to \int_{\underline{y}}^{y_1} F'_{20}(F_{10}(y))h_1(y)dy.$$

The right-hand side is linear with respect to $h_1$. It is also continuous since the integral is taken over a bounded interval. The second point follows.

**Proof of 3.** Combining the same reasoning as in part 1 with a dominated convergence argument, we obtain that $R_5(Q_{1|X}, Q_{2|X}, Q_{3|X}) = \int_0^1 Q_{1|X}\{Q_{2|X}^{-1}[Q_{3|X}(u|x)|x]|x\}du$ is Hadamard differentiable at $(Q_{10|X}, Q_{20|X}, Q_{30|X})$, with

$$\begin{aligned} dR_5(h_1, h_2, h_3) = \int_0^1 & \left\{ h_1 \left[ Q_{20|X}^{-1}[Q_{30|X}(u|x)], x \right] + \partial_u \left[ Q_{10|X} \circ Q_{20|X}^{-1} \right] [(Q_{30|X}(u|x)|x), x] \right. \\ & \left. \times \left[ -h_2 \left[ Q_{20|X}^{-1}[Q_{30|X}(u|x)], x \right] + h_3(u, x) \right] \right\} du. \end{aligned}$$

Besides, by the same reasoning as in the proof of Lemma 20.10 of van der Vaart (2000), $R_6(F_X, G) = \int m_{10}^D(x)G(x)dF_X(x)$ is Hadamard differentiable at any $(F_X, G)$ such that $F_X$ is a cdf and $G$ is of bounded variation. Moreover,

$$dR_6(h_1, h_2) = -\int h_1 d[m_{10}^D \times G] + \int \left[ m_{10}^D \times h_2 \right] dF_X.$$

The result follows by the chain rule □

**Lemma S6** *Assume Assumptions 2, 7, 9, 12 and S2 hold. Let*

$$\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_{00}, \lambda_{10}, \lambda_{01}, \lambda_{11}).$$

*For $d \in \{0,1\}$ and $q \in \mathcal{Q}$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \underline{F}_{CIC,d}(y)dy$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \overline{F}_{CIC,d}(y)dy$, $\theta \mapsto \overline{F}_{CIC,d}^{-1}(q)$ and $\theta \mapsto \underline{F}_{CIC,d}^{-1}(q)$ are Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.*

**Proof:** the proof is complicated by the fact that even if the primitive cdf are smooth, the bounds $\underline{F}_{CIC,d}$ and $\overline{F}_{CIC,d}$ may admit kinks, so that Hadamard differentiability is not trivial to derive. The proof is also lengthy as $\underline{F}_{CIC,d}$ and $\overline{F}_{CIC,d}$ take different forms depending on $d \in \{0,1\}$ and whether $\lambda_{00} < 1$ or $\lambda_{00} > 1$. Before considering all possible cases, note that by Assumption 9, $\underline{F}_{CIC,d} = C_d(\underline{T}_d)$.

**1. Lower bound $\underline{F}_{CIC,d}$**

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

$$\begin{aligned} \underline{T}_d &= M_0\left(m_1\left(U_d\right)\right), \\ C_d(\underline{T}_d) &= \frac{\lambda_{1d}F_{d11} - H_d\left(\lambda_{0d}F_{d01} + (1-\lambda_{0d})\underline{T}_d\right)}{\lambda_{1d} - 1}. \end{aligned}$$

Also, let

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

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

*Case 1: $\lambda_{00} < 1$ and $d = 0$.*

In this case, $U_0 = \frac{H_0^{-1}(\lambda_{10}F_{011}) - \lambda_{00}F_{001}}{1-\lambda_{00}}$. We first prove by contradiction that $y_{00}^u = +\infty$. First, because $\lim_{y\to+\infty} U_0(y) < 1$, we have

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

Thus, by Assumption 9, $U_0(y) < 1$ for all $y$, otherwise $\underline{T}_0(y)$ would be decreasing. Hence, $y_{10}^u = +\infty$.

Therefore, when $y_{00}^u < +\infty$, there exists $y$ such that $0 < U_0(y) < 1$. Assume that there exists $y' \geq y$ such that $U_0(y') < 0$. By continuity and the intermediate value theorem, this would imply that there exists $y'' \in (y, y')$ such that $U_0(y'') = 0$. But since both $U_0(y)$ and $U_0(y'')$ are included in $[0, 1]$, this would imply that $\underline{T}_0$ is strictly decreasing between $y$ and $y''$, which is not possible under Assumption 9. This proves that when $y_{00}^u < +\infty$, there exists $y$ such that for every $y' \geq y$, $0 \leq U_0(y') < 1$.

Consequently, $\underline{T}_0 = U_0$ for every $y' \geq y$. This in turn implies that $C_0(\underline{T}_0) = 0$ for every $y' \geq y$. Moreover, $C_0(\underline{T}_0)$ is increasing under Assumption 9, which implies that $C_0(\underline{T}_0) = 0$ for every $y$. This proves that when $y_{00}^u < +\infty$, $C_0(\underline{T}_0) = 0$. This implies that $\underline{\mathcal{S}}_0$ is empty, which violates Assumption S2. Therefore, under Assumption 9, we cannot have $y_{00}^u < +\infty$ when $\lambda_{00} < 1$. Because $y_{00}^u = +\infty$, $\underline{T}_0 = 0$. Therefore,

$$C_0(\underline{T}_0)(y) \quad = \quad \frac{\lambda_{10}F_{011}(y) - H_0\left(\lambda_{00}F_{001}(y)\right)}{\lambda_{10} - 1}.$$

The map $F \mapsto \int_{\mathcal{S}(Y)} F(y)dy$ is linear and continuous with respect to the supremum norm at any continuous $F$ because $\mathcal{S}(Y)$ is bounded. It is thus Hadamard differentiable, tangentially to $\mathcal{C}^0$. Therefore, by Assumption S2, the first point of Lemma S5, and the chain rule,

$$\theta \mapsto \int_{\mathcal{S}(Y)} \underline{F}_{CIC,0}(y)dy$$

is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.

Then, the map $F \mapsto F^{-1}$ is Hadamard differentiable at any $F$ with strictly positive derivative, tangentially to $\mathcal{C}^0$ (see, e.g., van der Vaart, 2000, Lemma 21.4). Moreover, by Assumption S2, $C_0(\underline{T}_0)$ is increasing and differentiable with strictly positive derivative on $\underline{\mathcal{S}}_0$, which is equal to

$\mathcal{S}(Y)$ in this case. Thus, by the first point of Lemma 5 and the chain rule, $\theta \mapsto \underline{F}_{CIC,0}^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$ for any $q \in \mathcal{Q}$.

*Case 2:* $\lambda_{00} > 1$ *and* $d = 0$.

In this case,

$$U_0 = \frac{\lambda_{00}F_{001} - H_0^{-1}(\lambda_{10}F_{011})}{\lambda_{00} - 1}.$$

Therefore, $\lim_{y \to \underline{y}} U_0(y) = 0$, and $\lim_{y \to \overline{y}} U_0(y) > 1$. As a result, $-\infty < y_{10}^u < +\infty$, and $\underline{T}_0(y_{10}^u) = U_0(y_{10}^u) = 1$. This in turn implies $C_0(\underline{T}_0)(y_{10}^u) = 0$. Combining this with Assumption 9 implies that $C_0(\underline{T}_0)(y) = 0$ for every $y \leq y_{10}^u$. Moreover, Assumption 9 also implies that $\underline{T}_d(y) = 1$ for every $y \geq y_{10}^u$. Therefore,

$$C_0(\underline{T}_0)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y_{10}^u, \\ \frac{\lambda_{10}F_{011}(y) - H_0(\lambda_{00}F_{001}(y) + (1-\lambda_{00}))}{\lambda_{10}-1} & \text{if } y > y_{10}^u. \end{array} \right.$$

Thus, $C_0(\underline{T}_0)(y) = M_0(R_2(F_{011}, F_{010}, F_{000}, F_{001}, \lambda_{00}, \lambda_{10}))$, where $R_2$ is defined as in Lemma S5. Hadamard differentiability of $\int_{\underline{y}}^{\overline{y}} C_0(\underline{T}_0)(y)dy$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$ thus follows by Points 1 and 2 of Lemma S5, the chain rule and the fact that by Assumption 13, $(F_{011}, F_{010}, F_{000}, F_{001}, \lambda_{00}, \lambda_{10}) \in (\mathcal{C}^1)^4 \times [0, \infty) \times ([0, \infty) \backslash \{1\})$. As for the LQTE, note that by Point 1 of Lemma S5, $\theta \mapsto C_0(\underline{T}_0)$ is Hadamard differentiable as a function on $(y_{10}^u, \overline{y})$, tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. By Assumption S2, $C_0(\underline{T}_0)$ is also strictly increasing and differentiable with positive derivative on $\underline{\mathcal{S}}_0 = (y_{10}^u, \overline{y})$. Thus, by point 1 of Lemma S5, Hadamard differentiability of $F \mapsto F^{-1}(q)$ at $(C_0(\underline{T}_0), q)$ for $q \in \mathcal{Q}$ tangentially to $\mathcal{C}^0$, and the chain rule, $\theta \mapsto \underline{F}_{CIC,0}^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.

*Case 3:* $\lambda_{00} < 1$ *and* $d = 1$.

In this case,

$$U_1 = \frac{\lambda_{01}F_{100} - H_1^{-1}(\lambda_{11}F_{111})}{\lambda_{01} - 1}.$$

$\lambda_{11} > 1$ implies that $\frac{1}{\lambda_{11}} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{1}{\lambda_{11}})$ is in $\mathring{\mathcal{S}}(Y)$ under Assumption 7.

*Case 3.a:* $\lambda_{00} < 1$, $d = 1$ *and* $y_{01}^u < y^*$.

We have $U_1(y^*) = \frac{\lambda_{01}F_{100}(y^*) - 1}{\lambda_{01} - 1} < 1$. Assume that $U_1(y^*) < 0$. Since $y_{01}^u < y^*$, this implies that there exists $y < y^*$ such that $0 < U_1(y)$. Since $U_1$ is continuous, there also exists $y' < y^*$ such that $0 < U_1(y') < 1$. By continuity and the intermediate value theorem, this finally implies that there exists $y''$ such that $y' < y''$ and $U_1(y'') = 0$. This contradicts Assumption 9 since this would imply that $\underline{T}_1$ is decreasing between $y'$ and $y''$. This proves that

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

Therefore, $\underline{T}_1(y^*) = U_1(y^*)$, which in turn implies that $C_1(\underline{T}_1)(y^*) = 0$. By Assumption 9, this implies that for every $y \leq y^*$, $C_1(\underline{T}_1)(y) = 0$.

For every $y$ greater than $y^*$,

$$U_1(y) \quad = \quad \frac{\lambda_{01}F_{100}(y) - 1}{\lambda_{01} - 1}.$$

$U_1(y) < 1$. Since $U_1(y^*) \geq 0$ and $y \mapsto \frac{\lambda_{01}F_{100}(y)-1}{\lambda_{01}-1}$ is increasing, $U_1(y) \geq 0$. Consequently, for $y \geq y^*$, $\underline{T}_1(y) = U_1(y)$.

Finally, we obtain

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y^*, \\ \frac{\lambda_{11}F_{111}(y)-1}{\lambda_{11}-1} & \text{if } y > y^*. \end{array} \right.$$

The result follows as in Case 2 above.

*Case 3.b:* $\lambda_{00} < 1$, $d = 1$ *and* $y^u_{01} \geq y^*$.

For all $y \geq y^*$, $U_1(y) = \frac{\lambda_{01}F_{100}(y)-1}{\lambda_{01}-1}$. This implies that $y^u_{01} = F^{-1}_{100}(1/\lambda_{01}) < +\infty$ and $U_1(y^u_{01}) = 0$. Because $y \mapsto \frac{\lambda_{01}F_{100}(y)-1}{\lambda_{01}-1}$ is increasing, $U_1(y) \geq 0$ for every $y \geq y^u_{01}$. Moreover, $U_1(y) \leq 1$. Therefore, $\underline{T}_1(y) = U_1(y)$ for every $y \geq y^u_{01}$. Beside, for every $y$ lower than $y^u_{01}$, $\underline{T}_1(y) = 0$. As a result,

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} \frac{\lambda_{11}F_{111}(y) - H_1(\lambda_{01}F_{101}(y))}{\lambda_{11}-1} & \text{if } y \leq y^u_{01}, \\ \frac{\lambda_{11}F_{111}(y)-1}{\lambda_{11}-1} & \text{if } y > y^u_{01}. \end{array} \right.$$

This implies that

$$\int_{\underline{y}}^{\overline{y}} C_1(\underline{T}_1)(y)dy = \frac{1}{\lambda_{11} - 1}\left[\lambda_{11}\int_{\underline{y}}^{\overline{y}} F_{111}(y)dy - R_4(\lambda_{01}F_{101}, H_1)\right],$$

where $R_4$ is defined in Lemma S5. $\theta \mapsto \int_{\underline{y}}^{\overline{y}} F_{111}(y)dy$ is Hadamard differentiable at $F_{111}$, tangentially to $\mathcal{C}^0$. As shown in the proof of Lemma S5, $H_1 = F_{110} \circ F^{-1}_{100}$ is a Hadamard differentiable function of $(F_{110}, F_{100})$, tangentially to $(\mathcal{C}^0)^2$. Thus, by Lemma S5 and the chain rule, $R_4(\lambda_{01}F_{101}, H_1)$ is a Hadamard differentiable function of $(F_{101}, F_{110}, F_{100})$, tangentially to $(\mathcal{C}^0)^3$. The result follows for $\int_{\underline{y}}^{\overline{y}} C_1(\underline{T}_1)(y)dy$.

The previous display also shows that $C_1(\underline{T}_1)$ is Hadamard differentiable as a function of

$$(F_{100}, F_{101}, F_{110}, F_{111}, \lambda_{01}, \lambda_{11})$$

when considering the restriction of these functions to $(\underline{y}, y^u_{01})$ only. By Assumption S2, $C_1(\underline{T}_1)$ is also a differentiable function with positive derivative on $(\underline{y}, y^u_{01})$. Therefore, using once again the first point of Lemma S5 and the chain rule, $\theta \mapsto C_1(\underline{T}_1)^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$, for $q \in (C_1(\underline{T}_1)(\underline{y}), C_1(\underline{T}_1)(y^u_{01})) = (0, q_1)$. The same holds when considering the interval $(y^u_{01}, \overline{y})$ instead of $(\underline{y}, y^u_{01})$. Hence, $\theta \mapsto \underline{F}^{-1}_{CIC,1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$, for $q \in (0, 1)\backslash\{q_1\} = \mathcal{Q}$.

*Case 4: $\lambda_{00} > 1$ and $d = 1$.*

In this case,

$$U_1 = \frac{H_1^{-1}(\lambda_{11}F_{111}) - \lambda_{01}F_{100}}{1-\lambda_{01}}.$$

Therefore, $\lim_{y\to\underline{y}} U_1(y) = 0$, which implies that $y_{11}^u > -\infty$. As above, $\lambda_{11} > 1$ implies that $y^*$ is in $\overset{\circ}{\mathcal{S}}(Y)$ under Assumption 7. $U_1(y^*) = \frac{1-\lambda_{01}F_{100}(y^*)}{1-\lambda_{01}} > 1$, which implies that $y_{11}^u < +\infty$. Therefore, reasoning as for Case 2, we obtain

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y_{11}^u, \\ \frac{\lambda_{11}F_{111}(y) - H_1(\lambda_{01}F_{100}(y)+(1-\lambda_{01}))}{\lambda_{11}-1} & \text{if } y > y_{11}^u. \end{array} \right.$$

The result follows as in Case 2 above.

**2. Upper bound $\overline{F}_{CIC,d}$.**

Let $V_d = \frac{\lambda_{0d}F_{d01} - H_d^{-1}(M_0(\lambda_{1d}F_{d11}+(1-\lambda_{1d})))}{\lambda_{0d}-1}$, so that

$$\begin{aligned} \overline{T}_d &= M_0\left(m_1\left(V_d\right)\right), \\ C_d(\overline{T}_d) &= \frac{\lambda_{1d}F_{d11} - H_d\left(\lambda_{0d}F_{d01} + (1-\lambda_{0d})\overline{T}_d\right)}{\lambda_{1d}-1}. \end{aligned}$$

Also, let

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

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

*Case 1: $\lambda_{00} < 1$ and $d = 0$.*

In this case,

$$V_0 = \frac{H_0^{-1}(\lambda_{10}F_{011} + (1-\lambda_{10})) - \lambda_{00}F_{001}}{1-\lambda_{00}}.$$

Since $\lambda_{10} < 1$, $\lim_{y\to\underline{y}} V_0(y) > 0$ and can even be greater than 1.

First, let us prove by contradiction that $y_{10}^v = -\infty$. $V_0(y) \leq 1$ for every $y \leq y_{10}^v$. Using the fact that $\lim_{y\to\underline{y}} V_0(y) > 0$ and that $\overline{T}_0$ must be increasing under Assumption 9, one can also show that $0 \leq V_0(y)$ for every $y \leq y_{10}^v$. This implies that $\overline{T}_0(y) = V_0(y)$ which in turn implies that $C_0(\overline{T}_0)(y) = 1$ for every $y \leq y_{10}^v$. Since $C_0(\overline{T}_0)$ must be increasing under Assumption 9, this implies that for every $y \in \mathcal{S}(Y)$,

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

This implies that $\overline{\mathcal{S}}_0$ is empty, which violates Assumption S2. Therefore, $y_{10}^v = -\infty$.

$y_{10}^v = -\infty$ implies that $\lim_{y\to\underline{y}} \overline{T}_0(y) = 1$. This combined with Assumption 9 implies that $\overline{T}_0(y) = 1$ for every $y \in \mathcal{S}(Y)$. Therefore,

$$C_0(\overline{T}_0)(y) = \frac{\lambda_{10}F_{011}(y) - H_0\left(\lambda_{00}F_{001}(y) + (1-\lambda_{00})\right)}{\lambda_{10}-1}.$$

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

*Case 2: $\lambda_{00} > 1$ and $d = 0$.*

In this case,

$$V_0 = \frac{\lambda_{00}F_{001} - H_0^{-1}(\lambda_{10}F_{011} + (1-\lambda_{10}))}{\lambda_{00}-1}.$$

Since $\lambda_{10} < 1$, $\lim_{y\to\underline{y}} V_0(y) < 0$. Therefore, $y_{00}^v > -\infty$.

*Case 2.a): $\lambda_{00} > 1$, $d = 0$ and $y_{00}^v < +\infty$.*

If $y_{00}^v \in \mathbb{R}$, $\overline{T}_0(y_{00}^v) = V_0(y_{00}^v)$ which in turn implies that $C_0(\overline{T}_0)(y_{00}^v) = 1$. By Assumption 9, this implies that for every $y \geq y_{00}^v$, $C_0(\overline{T}_0)(y) = 1$. For every $y \leq y_{00}^v$, $\overline{T}_0(y) = 0$, so that

$$C_0(\overline{T}_0) = \frac{\lambda_{10}F_{011} - H_0\left(\lambda_{00}F_{001}\right)}{\lambda_{10}-1}.$$

As a result,

$$C_0(\overline{T}_0)(y) = \left| \begin{array}{ll} \frac{\lambda_{10}F_{011}(y) - H_0(\lambda_{00}F_{001}(y))}{\lambda_{10}-1} & \text{if } y \leq y_{00}^v, \\ 1 & \text{if } y > y_{00}^v. \end{array} \right.$$

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

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

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

$$C_0(\overline{T}_0)(y) = \frac{\lambda_{10}F_{011}(y) - H_0\left(\lambda_{00}F_{001}(y)\right)}{\lambda_{10}-1}.$$

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

*Case 3: $\lambda_{00} < 1$ and $d = 1$.*

In this case,

$$V_1 = \frac{\lambda_{01}F_{101} - H_1^{-1}(\lambda_{11}F_{111} - (\lambda_{11}-1))}{\lambda_{01}-1}.$$

Therefore, $\lim_{y\to\underline{y}} V_1(y) = 0$, which implies that $y_{11}^v > -\infty$. $\lambda_{11} > 1$ implies that $\frac{\lambda_{11}-1}{\lambda_{11}} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{\lambda_{11}-1}{\lambda_{11}})$ is in $\overset{\circ}{\mathcal{S}}(Y)$ under Assumption 7.

*Case 3.a): $\lambda_{00} < 1$, $d = 1$ and $y_{11}^v > y^*$.*

We have $V_1(y^*) = \lambda_{01}F_{101}(y^*)/(\lambda_{01}-1) > 0$. If $y^* < y_{11}^v$, $V_1(y^*) < 1$. Therefore, $0 < \overline{T}_1(y^*) = V_1(y^*) < 1$. This implies that $C_1(\overline{T}_1)(y^*) = 1$ which in turn implies that $C_1(\overline{T}_1)(y) = 1$ for every $y \geq y^*$ under Assumption 9.

For every $y$ lower than $y^*$,

$$V_1(y) = \frac{\lambda_{01}F_{101}(y)}{\lambda_{01}-1}.$$

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

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\lambda_{11}F_{111}(y)}{\lambda_{11}-1} & \text{if } y \leq y^*, \\ 1 & \text{if } y > y^*. \end{array} \right.$$

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

*Case 3.b): $\lambda_{00} < 1$, $d = 1$, and $y_{11}^v \leq y^*$.*

First, $V_1(y_{11}^v) = 1$, implying $\overline{T}_1(y_{11}^v) = 1$. By Assumption 9, $\overline{T}_1(y) = 1$ for all $y \geq y_{11}^v$. Second, if $y \leq y_{11}^v \leq y^*$, $V_1(y) = \frac{\lambda_{01}F_{101}(y)}{\lambda_{01}-1}$. Thus $V_1$ is increasing on $(-\infty, y_{11}^v)$. Moreover $V_1(y_{11}^v) = 1$. Hence, $V_1(y) \leq 1$ for every $y \leq y_{11}^v$. Because we also have $V_1(y) \geq 0$, $\overline{T}_1(y) = V_1(y)$ for every $y \leq y_{11}^v$.

As a result,

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\lambda_{11}F_{111}(y)}{\lambda_{11}-1} & \text{if } y \leq y_{11}^v, \\ \frac{\lambda_{11}F_{111}(y) - H_1(\lambda_{01}F_{101}(y)+1-\lambda_{01})}{\lambda_{11}-1} & \text{if } y > y_{11}^v. \end{array} \right.$$

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

*Case 4: $\lambda_{00} > 1$ and $d = 1$.*

In this case,

$$V_1 = \frac{H_1^{-1}(\lambda_{11}F_{111} - (\lambda_{11}-1)) - \lambda_{01}F_{101}}{1-\lambda_{01}}.$$

$\lim_{y \to \overline{y}} V_1(y) = 1$, which implies that $y_{01}^v < +\infty$. As above, $\lambda_{11} > 1$ implies that $\frac{\lambda_{11}-1}{\lambda_{11}} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{\lambda_{11}-1}{\lambda_{11}})$ is in $\mathring{\mathcal{S}}(Y)$ under Assumption 7. $V_1(y^*) = \frac{-\lambda_{01}F_{101}(y^*)}{1-\lambda_{01}} < 0$. Since $\overline{T}_1$ is increasing under Assumption 9, one can show that this implies that $y_{01}^v > y^*$. Therefore, reasoning as for Case 2, we obtain that

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\lambda_{11}F_{111}(y) - H_1(\lambda_{01}F_{101}(y))}{\lambda_{11}-1} & \text{if } y \leq y_{01}^v, \\ 1 & \text{if } y > y_{01}^v. \end{array} \right.$$

The result follows as in Case 2 of the lower bound □

The following lemma on kernel estimators is classical but stated and proved for completeness. We let hereafter $\widehat{m}^U$ denote the Nadaraya-Watson estimator of $m^U(x) = E(U|X = x)$ with bandwidth $h_n$.

**Lemma S7** *Suppose that Assumptions 12 and 14-15-hold. Then* $\left\|\widehat{m}^U - m^U\right\|_\infty = o_P(n^{-1/4})$ *for* $U \in \{Y, D\}$.

**Proof:** Let $N^U(x) = E(U|X = x)f_X(x)$ and

$$\widehat{N}^U(x) = \frac{1}{nh_n}\sum_{i=1}^n U_i K\left(\frac{x - X_i}{h_n}\right),$$

so that $m^U = N^U/f_X$ and $\widehat{m}^U = \widehat{N}^U/\widehat{f}_X$. Assumptions 14 and 15 imply that the conditions of Lemma 8.10 of Newey & McFadden (1994) are met for $U = D$ and $U = Y$. As a result,

$$\max\left(\left\|\widehat{N}^U - N^U\right\|_\infty, \left\|\widehat{f}_X - f_X\right\|_\infty\right) = O_P\left[(\ln n)^{1/2}(nh_n^r)^{-1/2} + h_n^m\right].$$

Moreover, by Assumption 14, the right-hand side is an $o_P(n^{-1/4})$. Hence, with probability approaching one, the left-hand side is smaller than $c/2$, where $c = \inf_{x\in\mathcal{S}(X)} f_X(x) > 0$. Then, by Lemma S3 and the triangular inequality,

$$\begin{aligned}\left\|\widehat{m}^U - m^U\right\|_\infty \leq & \frac{1}{c}\left[\left\|\widehat{N}^U - N^U\right\|_\infty + \left\|m^U\right\|_\infty\left\|\widehat{f}_X - f_X\right\|_\infty\right] \\ & + \frac{2(1 + \|m^U\|_\infty)}{c^2}\max\left(\left\|\widehat{N}^U - N^U\right\|_\infty, \left\|\widehat{f}_X - f_X\right\|_\infty\right)^2.\end{aligned}$$

The result follows □

The proof of Theorem 5.2 uses repeatedly Lemma S8 below, which establishes a linear representation result on two-steps estimators involving a nonparametric first step. Let us consider two dummy variable $I$ and $J$ and $U$ and $V$ be other two random variables. In the proof of Theorem 5.2, $I$ and $J$ are functions of $D$, $G$ and $T$, $U$ is $D$ or $Y$ and $V$ is a function of $X$. Let also $\beta_0 = E[VE[U|X, J = 1]|I = 1]$ and

$$\widehat{\beta} = \frac{\sum_{i=1}^n I_i V_i \widehat{m}^U_{J=1}(X_i)}{\sum_{i=1}^n I_i},$$

where $\widehat{m}^U_{J=1}$ is the Nadaraya-Watson estimator of $m^U_{J=1}(x) = E(U|X, J = 1)$ a regression function on another subgroup:

$$\widehat{m}^U_{J=1}(x) = \frac{\sum_{i=1}^n J_i K_{h_n}(x - X_i)U_i}{\sum_{i=1}^n J_i K_{h_n}(x - X_i)}.$$

We have let $K_h(x) = K(x/h)/h$, with $K$ a kernel function, and $h_n$ a bandwidth parameter. The following lemma shows that under suitable conditions, $\widehat{\beta}$ admits a linear representation.

**Lemma S8** *Suppose that* $(I_i, J_i, U_i, V_i, X_i)_{i=1,...,n}$ *are i.i.d. and Assumption 14 and the first point of Assumption 15 hold. Suppose also that* $x \mapsto E(JU|X = x)$ *and* $x \mapsto E(J|X = x)$ *are* $m$ *times differentiable,* $E(V^4) < \infty$, $x \mapsto E(V^4|X = x)$ *is Lipschitz,* $x \mapsto E(IV|X = x)$ *and* $x \mapsto P(J = 1|X = x)$ *are continuous,* $P(J = 1|X) \geq \underline{p} > 0$ *almost surely and* $P(I = 1) > 0$. *Then*

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{I_i(V_i m^U_{J=1}(X_i) - \beta_0) + \lambda(X_i)J_i(U_i - m^U_{J=1}(X_i))}{P(I = 1)} + o_P(1), \tag{48}$$

*where* $\lambda(x) = E(IV|X = x)/P(J = 1|X = x)$.

**Proof:** let $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} I_i V_i \widehat{m}^U_{J=1}(X_i)$ and $\theta_0 = E(IVm^U_{J=1}(X))$. We first prove that $\widehat{\theta}$ is root-n consistent and can be linearized. For that purpose, we check that the conditions (i)-(iv) of Theorem 8.11 of Newey & McFadden (1994) apply here. We adopt the same notations as Newey & McFadden (1994). Let $\gamma_0 = (\gamma_{01}, \gamma_{02})'$, with $\gamma_{01}(x) = E[JU|X = x]f_X(x)$ and $\gamma_{02}(x) = E[J|X = x]f_X(x)$. Let also $\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2)'$, with $\widehat{\gamma}_1(x) = \frac{1}{n}\sum_{i=1}^{n} J_i K_{h_n}(x - X_i)U_i$ and $\widehat{\gamma}_2(x) = \frac{1}{n}\sum_{i=1}^{n} J_i K_{h_n}(x-X_i)$. Let $Z_i = (I_i, J_i, X_i, U_i, V_i)$ and let $g(Z, \gamma) = IV\gamma_1(X)/\gamma_2(X)-\theta_0$, so that $\widehat{\theta} - \theta_0 = \frac{1}{n}\sum_{i=1}^{n} g(Z_i, \widehat{\gamma})$. Finally, let $G(Z, \gamma) = \frac{IV}{\gamma_{02}(X)}\left[\gamma_1(X) - m^U_{J=1}(X) \times \gamma_2(X)\right]$. Let $C = \underline{p}c$, so that $\|\gamma_{02}\|_\infty \geq C$. By Lemma S3 applied to $x_1 = IV\gamma_1(X)$, $y_1 = \gamma_2(X)$, $x_2 = IV\gamma_{01}(X)$ and $y_2 = \gamma_{02}(X)$, with $\gamma$ satisfying $\|\gamma - \gamma_0\|_\infty < C/2$,

$$\begin{aligned}|g(Z, \gamma) - g(Z, \gamma_0) - G(Z, \gamma - \gamma_0)| &\leq \frac{2(1 + |Vm^U_{J=1}(X)|)}{C^2} \max\left(|V|\,\|\gamma_1 - \gamma_{01}\|_\infty, \|\gamma_2 - \gamma_{02}\|_\infty\right)^2 \\ &\leq \frac{2(1 + |Vm^U_{J=1}(X)|)}{C^2} (1 + |V|)^2 \|\gamma - \gamma_0\|^2_\infty .\end{aligned}$$

Moreover, $E[|V|^3] < \infty$ and there exists $K_0$ such that $|m^U_{J=1}(x)| \leq K_0$ on $\mathcal{S}(X)$. Thus,

$$E\left[(1 + |Vm^U_{J=1}(X)|)|V|^2\right] < \infty$$

and (i) of Theorem 8.11 of Newey & McFadden (1994) holds.

Still using $\|\gamma_{02}\|_\infty \geq C$, we get

$$|G(z, \gamma)| \leq \frac{|v|}{2C}(1 + |m^U_{J=1}(x)|)\,\|\gamma\|_\infty ,$$

with $E[|V|(1 + |m^U_{J=1}(X)|)^2] < \infty$. This proves the condition (ii) in Theorem 8.11 of Newey & McFadden (1994).

Now remark that for all function $\gamma$ satisfying $\|\gamma\|_\infty < \infty$,

$$E\left[G(Z, \gamma)\right] = \int v(x)'\gamma(x)dx,$$

with $v(x) = E(IV|X = x)(1, -m^U_{J=1}(x))' \mathbb{1}_{\mathcal{S}(X)}(x)/P(J = 1|X = x)$. Hence, condition (iii) in Theorem 8.11 of Newey & McFadden (1994) holds.

Finally, to prove their (iv), note first that $v$ is continuous on $\mathcal{S}(X)$ by assumption. Moreover, Jensen's inequality, the triangular inequality and Assumption 15 yield

$$\begin{aligned}\int \|v(x)\|_1 \, dx &\leq \frac{1+K_0}{\underline{p}} \int \frac{f_X(x)}{c} E(|V||X=x)dx \\ &\leq \frac{(1+K_0)E(|V|}{c\underline{p}} < \infty.\end{aligned}$$

Besides, for all $t$ such that $\|t\|_1 \leq 1$, we have, by Jensen's inequality and the fact that $x \mapsto E(V^4|X = x)$ is Lipschitz (with a constant $K_1$, say),

$$\begin{aligned}\|v(x+t)\|_1^4 &= \frac{E(IV|X=x+t)^4}{P(J=1|X=x)^4} \left[1 + |m(x+t)|\right]^4 \mathbb{1}_{\mathcal{S}(X)}(x+t) \\ &\leq \frac{E(V^4|X=x+t)}{\underline{p}^4} \left[1+K_0\right]^4 \\ &\leq \frac{(K_1 + E(V^4|X=x))}{\underline{p}^4} \left[1+K_0\right]^4.\end{aligned}$$

This and $E(V^4) < \infty$ imply that $E\left[\sup_{t \in \mathbb{R}^p : \|t\|_1 \leq 1} \|v(X+t)\|_1^4\right] < \infty$. Hence, Condition (iv) of Theorem 8.11 of Newey & McFadden (1994) holds, and this theorem applies. The proof of Theorem 8.11 then implies that the conditions of Theorem 8.1 of Newey & McFadden (1994) are satisfied, with

$$\delta(Z) = v(X)'(JU, J)' - E[v(X)'(JU, J)'] = \frac{E(IV|X)J}{P(J=1|X)}(U - m^U_{J=1}(X)).$$

The proof of Theorem 8.1 of Newey & McFadden (1994) finally implies that

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(Z_i, \gamma_0) + \delta(Z_i) + o_P(1).$$

Now, applying Lemma S3 with $x_1 = \widehat{\theta}$, $y_1 = \widehat{P}(I_i = 1)$, $x_2 = \theta_0$ and $y_2 = P(I = 1)$, we obtain, with a large probability,

$$\begin{aligned}&\left|\widehat{\beta} - \beta_0 - \frac{1}{P(I=1)}\left[\widehat{\theta} - \theta_0 - \beta_0\left(\widehat{P}(I=1) - P(I=1)\right)\right]\right| \\ &\leq \frac{2(1+|\beta_0|)}{P(I=1)^2} \max(|\widehat{\theta} - \theta_0|, |\widehat{P}(I=1) - P(I=1)|)^2.\end{aligned}$$

Moreover, the right-hand side is an $o_P(1/\sqrt{n})$. By rearranging the left-hand side, we finally obtain the linear decomposition (48) □

Finally, the asymptotic normality of the CIC-type estimator with covariates, established in Part 3 of Theorem 5.2, uses the following Lemma S9, together with Part 3 of Lemma S5 above.

**Lemma S9** *1. Under Assumption 7X and 16X, we have*

$$\sqrt{n}\left[\widehat{F}^{-1}_{dgt|x}(\tau) - F^{-1}_{dgt|x}(\tau)\right] = \frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_{dgt}} \frac{x'J_\tau X_i}{p_{dgt}}\left(\tau - \mathbb{1}\{Y_i - X_i'\beta(\tau)\le 0\}\right) + o_P(1),$$

*where* $J_\tau = E\left[f_{Y|X}(X'\beta(\tau))XX'\right]^{-1}$ *and the* $o_P(1)$ *is uniform over* $(\tau, x)\in(0,1)\times\mathcal{S}(X)$.
*2. For any* $(x,\tau)\in\mathcal{S}(X)\times(0,1)$, *let* $\widehat{G}(\tau,x) = (\widehat{F}_{X_{11}}(x), \widehat{F}^{-1}_{101|x}(\tau), \widehat{F}^{-1}_{100|x}(\tau), \widehat{F}^{-1}_{110|x}(\tau))$. *Then*

$$\sqrt{n}\left[\widehat{G} - G\right] \Longrightarrow \mathbb{G},$$

*where the convergence is in the space of continuous process on* $(0,1)\times\mathcal{S}(X)$ *and* $\mathbb{G}$ *denotes a continuous gaussian process defined on that space.*

**Proof: Part 1.** We prove that uniformly over $(\tau, x)$,

$$\sqrt{n_{dgt}}\left[\widehat{F}^{-1}_{dgt|x}(\tau) - F^{-1}_{dgt|x}(\tau)\right] = \frac{1}{\sqrt{n_{dgt}}}\sum_{i\in\mathcal{I}_{dgt}} x'J_\tau X_i\left(\tau - \mathbb{1}\{Y_i - X_i'\beta_{dgt}(\tau)\le 0\}\right) + o_P(1). \tag{49}$$

The result then follows directly from $n_{dgt}/[np_{dgt}] \xrightarrow{\mathbb{P}} 1$, as in the proof of Lemma S4. To alleviate the notational burden, we let the dependency in $(d,g,t)$ implicit hereafter. For instance, we let $\mathcal{I}$ denote $\mathcal{I}_{dgt}$, $n$ denote $n_{dgt}$, etc.. We denote by $P_n$ the empirical distribution of $(X,Y)$ on $\mathcal{I}$, $P$ denote its true distribution and $\mathbb{G}_n = \sqrt{n}(P_n - P)$. We write, e.g., $Ph$ as a shortcut for $\int hdP$. We also let $\rho_{\tau,\beta}(x,y) = (\tau - \mathbb{1}\{y - x'\beta\le 0\})(y - x'\beta)$, $h_{\tau,\beta}(x,y) = x(\tau - \mathbb{1}\{y\le x'\beta\})$, $\mathcal{R} = \{\rho_{\tau,\beta}, (\tau,\beta)\in[0,1]\times B\}$ and $\mathcal{H} = \{h_{\tau,\beta}, (\tau,\beta)\in[0,1]\times B\}$. To establish our proof of (49), we first show that $\widehat{\beta}(\tau)$ is uniformly consistent in $\tau$. Then we prove a uniform Bahadur representation on $\widehat{\beta}(\tau)$.

*a. Uniform consistency*
Let $M_\tau(\beta) = -P\rho_{\tau,\beta}$ and $M_{n\tau}(\beta) = -P_n\rho_{\tau,\beta}$. First, $\mathcal{R}$ is Glivenko-Cantelli because it satisfies the conditions of pointwise compact classes considered in Example 19.8 in van der Vaart (2000). As a result,

$$\sup_{\beta,\tau}|M_{n\tau}(\beta) - M_\tau(\beta)| \xrightarrow{\mathbb{P}} 0.$$

Following the proof of Theorem 5.7 of van der Vaart (2000), this implies

$$0 \le \sup_{\tau\in(0,1)} M_\tau(\beta(\tau)) - M_\tau(\widehat{\beta}(\tau)) \xrightarrow{\mathbb{P}} 0. \tag{50}$$

Second, using Equation (4.3) of Koenker (2005), we obtain, for any $\beta$,

$$\begin{aligned}M_\tau(\beta(\tau)) - M_\tau(\beta) &= E[\rho_\tau(Y - X'\beta)] - E[\rho_\tau(Y - X'\beta(\tau))]\\ &= E\left[\int_0^{X'(\beta-\beta(\tau))} F_{Y|X}(s + X'\beta(\tau)) - F_{Y|X}(X'\beta(\tau))ds\right].\end{aligned}$$

Because $\inf_{(y,x)} f_{Y|X}(y|x) = c > 0$ and $X$ is assumed to have bounded support, this yields

$$M_\tau(\beta(\tau)) - M_\tau(\beta) \geq K \left\|\beta(\tau) - \beta\right\|^2, \tag{51}$$

for some constant $K > 0$ independent of $\tau$. Fix $\varepsilon > 0$. If $\sup_{\tau\in(0,1)} \left\|\widehat{\beta}(\tau) - \beta(\tau)\right\| > \varepsilon$, then there exists $\tau_0$ such that $\left\|\widehat{\beta}(\tau_0) - \beta(\tau_0)\right\| > \varepsilon/2$. Then (51) implies that

$$\sup_{\tau\in(0,1)} M_\tau(\beta(\tau)) - M_\tau(\widehat{\beta}(\tau)) \geq K\varepsilon^2/4,$$

which happens with proability approaching 0 in view of (50). The result follows.

*b. Uniform Bahadur representation*

Let $\mathbf{X}$ (resp. $\mathbf{Y}$) denote the matrix (resp. the vector) stacking all $X_i$ (resp. $Y_i$), for $i \in \mathcal{I}$. For all $\tau \in (0,1)$, there exists a subset $h \subset \mathcal{I}$ of $r$ elements such that the corresponding submatrix (resp. subvector) $X(h)$ (resp. $Y(h)$) of $\mathbf{X}$ (resp. of $\mathbf{Y}$) satisfies $\widehat{\beta}(\tau) = X(h)^{-1}Y(h)$ (see Koenker, 2005, p.34). Note also that by Assumption 16, $\mathbf{Y}$ and $\mathbf{X}$ are in general position with probability one (see Koenker, 2005, p.35). Then

$$\sum_{i\in h} X_i(\tau - \mathbb{1}\{Y_i \leq X_i'\widehat{\beta}(\tau)\} = (\tau - 1)X(h)'\iota_r,$$

where $\iota_r$ is a vector of one of size $r$. Moreover, by Theorem 2.1 of Koenker (2005), there exists $\lambda = (\lambda_1, ..., \lambda_r)'$ with $|\lambda_j| \leq 1$ such that

$$\sum_{i\in \overline{h}} X_i(\tau - \mathbb{1}\{Y_i \leq X_i'\widehat{\beta}(\tau)\} = X(h)'\lambda,$$

where $\overline{h}$ denotes the complement of $h$ in $\mathcal{I}$. By Assumption 16, $\|X_i\|_1 \leq C$ for some $C > 0$. Hence, we obtain,

$$\left\|\sum_{i\in\mathcal{I}} X_i(\tau - \mathbb{1}\{Y_i \leq X_i'\widehat{\beta}(\tau)\})\right\|_1 \leq 2\sum_{i\in h} \|X_i\|_1 \leq 2Cr,$$

which holds uniformly over $(d, g, t, \tau)$. Thus,

$$\sup_{\tau\in(0,1)} \left\|\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}} X_i(\tau - \mathbb{1}\{Y_i \leq X_i'\widehat{\beta}(\tau)\})\right\|_1 \xrightarrow{\mathbb{P}} 0.$$

Now, using $Ph_{\tau,\beta(\tau)} = 0$, we obtain

$$-\sqrt{n}P\left[h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right] = \mathbb{G}_n\left[h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right] + \mathbb{G}_n h_{\tau,\beta(\tau)} + o_P(1),$$

uniformly over $\tau$. Moreover, by the intermediate value theorem,

$$\sqrt{n}P\left[h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right] = E\left[f_{Y|X}(X'(t_\tau\widehat{\beta}(\tau) + (1-t_\tau)\beta(\tau))|X)XX'\right]\sqrt{n}\left(\widehat{\beta}(\tau) - \beta(\tau)\right).$$

for some random $t_\tau \in [0,1]$. Now, by uniform consistency of $\widehat{\beta}(\tau)$ and continuity of $f_{Y|X}(.|x)$,

$$\sup_{\tau\in(0,1)} \left| f_{Y|X}(X'(t_\tau\widehat{\beta}(\tau) + (1-t_\tau)\beta(\tau))|X) - f_{Y|X}(X'(t_\tau\widehat{\beta}(\tau) + (1-t_\tau)\beta(\tau))|X)\right| \xrightarrow{\mathbb{P}} 0.$$

Because $f_{Y|X}(.|x)$ is bounded and $\mathcal{S}(X)$ is compact, Theorem 2.20 in van der Vaart (2000) implies that

$$\sqrt{n}P\left[h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right] = \left(J_\tau^{-1} + o_P(1)\right)\sqrt{n}\left(\widehat{\beta}(\tau) - \beta(\tau)\right),$$

where the $o_P(1)$ is uniform over $\tau$.

Next, remark that $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$, with $\mathcal{H}_1 = \{(x,y) \mapsto x\tau, \tau \in [0,1]\}$ and $\mathcal{H}_2 = \{(x,y) \mapsto -x\mathbb{1}\{y - x'\beta \leq 0\}, \beta \in B\}$. The sets $\mathcal{H}_1$ and $\{(x,y) \mapsto y - x'\beta\}, \beta \in B\}$ are Donsker as subsets of vector spaces (see van der Vaart, 2000, Example 19.17). Still by Example 19.17 in van der Vaart, 2000, this imlies that $\mathcal{H}_2$, and then also $\mathcal{H}$, is Donsker. Besides,

$$\begin{aligned}
P\left\|h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right\|_1^2 &= E\left[\|X\|_1^2 \left|\mathbb{1}\{Y \leq X'\widehat{\beta}(\tau)\} - \mathbb{1}\{Y \leq X'\widehat{\beta}(\tau)\}\right|^2\right] \\
&\leq C^2 E\left[\left|F_{Y|X}(X'\widehat{\beta}(\tau)) - F_{Y|X}(X'\beta(\tau))\right|\right] \\
&\leq K' \sup_{(y,x)} f_{Y|X}(y|x)\left\|\widehat{\beta}(\tau) - \beta(\tau)\right\|_1.
\end{aligned}$$

Hence, $\sup_{\tau\in(0,1)} P\left\|h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right\|_1^2 \xrightarrow{\mathbb{P}} 0$. Then, following the proof of Theorem 19.26 of van der Vaart (2000), we get, uniformly over $\tau$,

$$\mathbb{G}_n\left[h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right] \xrightarrow{\mathbb{P}} 0.$$

For all $\tau \in (0,1)$, the smallest eigenvalue of $J_\tau^{-1}$ is greater than the one of $cE[XX']$. It is thus bounded away from 0, uniformly over $\tau$. This, combined with the boundedness of $\mathcal{S}(X)$ and what precedes, yields

$$\sup_{x,\tau}\left|x'J_\tau\mathbb{G}_n\left[h_{\tau,\widehat{\beta}(\tau)} - h_{\tau,\beta(\tau)}\right]\right| \xrightarrow{\mathbb{P}} 0.$$

Equation (49) follows.

**2.** We prove the result for $\widehat{F}^{-1}$ only. By the Cramer-Wold device, a similar reasoning applies for $\widehat{G}$. By the stability properties of Donsker classes (see, e.g.,van der Vaart, 2000, Example 19.18), it is easy to see that the set of functions

$$\{(d,g,t,x,y) \mapsto \mathbb{1}\{d = \widetilde{d}, g = \widetilde{g}, t = \widetilde{t}\}\widetilde{x}'J_\tau(y - \mathbb{1}\{y - x'\beta \leq 0\}), (\widetilde{x},\tau,\beta) \in \mathcal{S}(X) \times (0,1) \times B\}$$

is Donsker, for any $(\widetilde{d},\widetilde{g},\widetilde{t}) \in \{0,1\}^3$. Hence,

$$\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}} x'J_\tau X_i\left(\tau - \mathbb{1}\{Y_i - X_i'\beta(\tau) \leq 0\}\right) \Longrightarrow \mathbb{G},$$

where the convergence is in the space of continuous process on $(0,1) \times \mathcal{S}(X)$ and $\mathbb{G}$ denotes a continuous gaussian process. Part 1 and, e.g. Theorem 18.10-(iv) of van der Vaart (2000) then imply the result $\square$

## References


Andrews, D. W. K. & Barwick, P. J. (2012), 'Inference for parameters defined by moment inequalities: A recommended moment selection procedure', *Econometrica* **80**, 2805–2826.

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

Chernozhukov, V., Lee, S. & Rosen, A. M. (2013), 'Intersection bounds: Estimation and inference', *Econometrica* **81**(2), 667–737.

de Chaisemartin, C. & D'Haultfœuille, X. (2015), Fuzzy differences-in-differences, Technical report.

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

Enikolopov, R., Petrova, M. & Zhuravskaya, E. (2011), 'Media and political persuasion: Evidence from russia', *The American Economic Review* **101**(7), 3253.

Field, E. (2005), 'Property rights and investment in urban slums', *Journal of the European Economic Association* **3**(2-3), 279–290.

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

Gentzkow, M., Shapiro, J. M. & Sinkinson, M. (2011), 'The effect of newspaper entry and exit on electoral politics', *The American Economic Review* **101**(7), 2980.

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

Koenker, R. (2005), *Quantile regression*, Cambridge university press.

Newey, W. K. & McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Romano, J. P., Shaikh, A. M. & Wolf, M. (2014), 'A practical two-step method for testing moment inequalities', *Econometrica* **82**, 1979–2002.

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

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

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

White, T. H. (2003), 'United states early radio history'.