The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–23

Fuzzy Differences-in-Differences with Stata

Clément de Chaisemartin
University of California at Santa Barbara
Santa Barbara, California
clementdechaisemartin@ucsb.edu

Xavier D'Haultfœuille
CREST
Palaiseau, France
xavier.dhaultfoeuille@ensae.fr

Yannick Guyonvarch
CREST
Palaiseau, France
yannick.guyonvarch@ensae.fr

Abstract. Differences-in-differences (DID) is a method to evaluate the effect of a treatment. In its basic version, a "control group" is untreated at two dates, whereas a "treatment group" becomes fully treated at the second date. However, in many applications of this method, the treatment rate only increases more in the treatment group. In such fuzzy designs, de Chaisemartin and D'Haultfœuille (2018b) propose various estimands that identify local average and quantile treatment effects under different assumptions. They also propose estimands that can be used in applications with a non-binary treatment, multiple periods and groups and covariates. This paper presents the Stata command `fuzzydid`, which computes the various corresponding estimators. We illustrate the use of the command by revisiting Gentzkow et al. (2011).

Keywords: differences-in-differences, fuzzy designs, local average treatment effects, local quantile treatment effects

1 Introduction

Differences-in-differences (DID) is a method to evaluate the effect of a treatment when experimental data are not available. In its basic version, a "control group" is untreated at two dates, whereas a "treatment group" becomes fully treated at the second date. However, in many applications of the DID method the treatment rate increases more in some groups than in others, but there is no group that goes from fully untreated to fully treated, and there is also no group that remains fully untreated. In such fuzzy designs, a popular estimator of treatment effects is the DID of the outcome divided by the DID of the treatment, the so-called Wald-DID.

As shown by de Chaisemartin and D'Haultfœuille (2018b), the Wald-DID identifies a local average treatment effect (LATE) if two assumptions on treatment effects are satisfied. First, the effect of the treatment should not vary over time. Second, when the treatment increases both in the treatment and in the control group, treatment effects should be equal in these two groups. de Chaisemartin and D'Haultfœuille (2018b) also propose two alternative estimands of the same LATE. These estimands do not rely on any assumption on treatment effects, and they can be used when the share of treated units is stable in the control group. The first one, the time-corrected Wald ratio (Wald-

TC), relies on common trends assumptions within subgroups of units sharing the same treatment at the first date. The second one, the changes-in-changes Wald ratio (Wald-CIC), generalizes the changes-in-changes estimand introduced by Athey and Imbens (2006) to fuzzy designs. Finally, under the same assumptions as those used for the Wald-CIC, local quantile treatment effects (LQTE) are also identified.

In this paper, we describe the `fuzzydid` Stata command, which computes the estimators corresponding to these estimands and performs inference on the LATE and LQTE using the bootstrap. In the computation of standard errors and confidence intervals, clustering along one dimension can be allowed for. Equality tests between the Wald-DID, Wald-TC, and Wald-CIC and placebo tests can also be performed. This turns out to be important for choosing between these different estimands, as they identify the LATE under different sets of assumptions.

The identification results mentioned above hold with a control group where the share of treated units does not change over time, a binary treatment, no covariates, and two groups and two periods. Nonetheless, they can be extended in several directions. First, under the same assumptions as those underlying the Wald-TC estimand, the LATE of treatment group switchers can be bounded when the share of treated units changes over time in the control group. Second, non-binary treatments can be easily handled by just modifying the parameter of interest. Third, when the assumptions are more credible conditional on some controls, it is possible to modify the Wald-DID, Wald-TC, and Wald-CIC estimands to incorporate such controls. The `fuzzydid` command handles all these extensions.

Finally, results can be extended to applications with multiple periods and groups. Those are very prevalent in applied work, and researchers then estimate treatment effects through linear regressions including time and group fixed effects. de Chaisemartin and D'Haultfœuille (2018a) show that around 19% of all empirical papers published by the American Economic Review between 2010 and 2012 make use of this research design. This paper also shows that these regressions are extensions of the Wald-DID to multiple periods and groups, and that they identify weighted averages of LATE, with possibly many negative weights.[1] As a result, they do not satisfy the no-sign reversal property: the coefficient of the treatment variable in those regressions may be negative even if the treatment effect is positive for every unit in the population. On the other hand, the Wald-DID, Wald-TC, and Wald-CIC estimands can be extended to applications with multiple groups and periods, and they then identify a LATE under the same assumptions as in the two groups and two periods case. Again, the `fuzzydid` command computes the corresponding estimators.

The remainder of the paper is organized as follows. Section 2 presents the estimands and estimators considerered by de Chaisemartin and D'Haultfœuille (2018b) in the simplest set-up with two groups and periods, a binary treatment and no covariates. Section 3 discusses the various extensions covered by the command. Section 4 presents the `fuzzydid` Stata command. Section 5 illustrates the command by revisiting Gentzkow et al. (2011), who estimate the effect of newspapers on electoral participa-

1. A Stata command computing these weights is available on the authors' webpages.

tion. Section 6 presents the finite sample performances of the various estimators through Monte Carlo simulations. Section 7 concludes.

2 Set-up

2.1 Parameters of interest, assumptions, and estimands

We seek to identify the effect of a treatment D on some outcome. In this section, we assume that D is binary.[2] $Y(1)$ and $Y(0)$ denote the two potential outcomes of the same individual with and without treatment, while $Y = Y(D)$ denotes the observed outcome. We assume the data can be divided into time periods represented by a random variable $T \in \{0, ..., \overline{t}\}$, and into groups represented by a random variable $G \in \{0, ..., \overline{g}\}$. We start by considering the simple case where $\overline{t} = \overline{g} = 1$, thus implying that there are two groups and two periods. In such a case, $G = 1$ (resp. $G = 0$) for units in the treatment (resp. control) group.

We use the following notation hereafter. For any random variable R, $\mathcal{S}(R)$ denotes its support. R_{gt} and R_{dgt} are two other random variables such that $R_{gt} \sim R|\, G = g, T = t$ and $R_{dgt} \sim R|\, D = d, G = g, T = t$, where $\sim$ denotes equality in distribution. For any event or random variable A, F_R and $F_{R|A}$ denote respectively the cumulative distribution function (cdf) of R and its cdf conditional on A. Finally, for any increasing function F on the real line, we let $F^{-1}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$. In particular, F_R^{-1} is the quantile function of R.

We maintain Assumptions 1-3 below in most of the paper.

Assumption 1. *(Fuzzy design)*

$E(D_{11}) > E(D_{10})$, and $E(D_{11}) - E(D_{10}) > E(D_{01}) - E(D_{00})$.

Assumption 2. *(Stable percentage of treated units in the control group)*

For all $d \in \mathcal{S}(D)$, $P(D_{01} = d) = P(D_{00} = d) \in (0, 1)$.

Assumption 3. *(Treatment participation equation)*

There exist $D(0), ..., D(\overline{t})$ such that $D = D(T)$, $D(t) \perp\!\!\!\perp T|G$ ($t \in \{0, ..., \overline{t}\}$) and for all $t \in \{1, ..., \overline{t}\}$,

$$P(D(t) \geq D(t-1)|G) = 1 \;\text{ or } P(D(t) \leq D(t-1)|G) = 1.$$

In standard "sharp" designs, we have $D = G \times T$, meaning that only observations in the treatment group and in period 1 get treated. With Assumption 1, we consider instead "fuzzy" settings where $D \neq G \times T$ in general, but where the treatment group experiences a higher increase of its treatment rate between period 0 and 1. Assumption 2 requires that the treatment rate remain constant in the control group, and be strictly

2. We still define our assumptions and estimands for any scalar treatment, to avoid redefining them when we will extend our results to non-binary treatments.

included between 0 and 1. This assumption is testable. Assumption 3 is equivalent to the latent index model $D = \mathbb{1}\{V \geq v_{GT}\}$ (with $V \perp\!\!\!\perp T|G$) considered in de Chaisemartin and D'Haultfœuille (2018b). In repeated cross sections, $D(t)$ denotes the treatment status of a unit at period t, and only $D = D(T)$ is observed. In single cross sections where cohort of birth plays the role of time, $D(t)$ denotes instead the potential treatment of a unit had she been born at $T = t$. Here again, only $D = D(T)$ is observed.

We consider the subpopulation $S = \{D(0) < D(1), G = 1\}$, called hereafter the treatment group switchers. Our parameters of interest are their Local Average Treatment Effect (LATE) and Local Quantile Treatment Effects (LQTE), which are respectively defined by

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

We now introduce the main estimands considered in de Chaisemartin and D'Haultfœuille (2018b). We start by considering the three estimands of Δ. The first is the Wald-DID defined by

$$W_{DID} = \frac{E(Y_{11}) - E(Y_{10}) - (E(Y_{01}) - E(Y_{00}))}{E(D_{11}) - E(D_{10}) - (E(D_{01}) - E(D_{00}))}.$$

W_{DID} is the coefficient of D in a 2SLS regression of Y on D with G and T as included instruments, and $G \times T$ as the excluded instrument.

The second estimand of Δ is the time-corrected Wald ratio (Wald-TC) defined by

$$W_{TC} = \frac{E(Y_{11}) - E(Y_{10} + \delta_{D_{10}})}{E(D_{11}) - E(D_{10})},$$

where $\delta_d = E(Y_{d01}) - E(Y_{d00})$, for $d \in \mathcal{S}(D)$. Without the $\delta_{D_{10}}$ term, W_{TC} would correspond to the coefficient of D in a 2SLS regression of Y on D using T as the excluded instrument, within the treatment group. δ_0 (resp. δ_1) measures the evolution of the outcome among untreated (resp. treated) units in the control group. Under the assumption that these evolutions are the same in the two groups (see Assumption 4 ' below), the $\delta_{D_{10}}$ term accounts for the effect of time on the outcome in the treatment group.

The third estimand of Δ is the change-in-change Wald ratio (Wald-CIC) defined by

$$W_{CIC} = \frac{E(Y_{11}) - E(Q_{D_{10}}(Y_{10}))}{E(D_{11}) - E(D_{10})},$$

where $Q_d(y) = F^{-1}_{Y_{d01}} \circ F_{Y_{d00}}(y)$ is the quantile-quantile transform of Y from period 0 to 1 in the control group conditional on $D = d$. W_{CIC} is similar to W_{TC}, except that it accounts for the effect of time on the outcome through the quantile-quantile transform instead of the additive term $\delta_{D_{10}}$.

Finally, we consider an estimand of τ_q. Let

$$F_{CIC,d} = \frac{P(D_{11} = d)F_{Y_{d11}} - P(D_{10} = d)F_{Q_d(Y_{d10})}}{P(D_{11} = d) - P(D_{10} = d)}$$

and

$$\tau_{CIC,q} = F^{-1}_{CIC,1}(q) - F^{-1}_{CIC,0}(q).$$

The estimands above identify Δ or τ_q under combinations of the following assumptions.

Assumption 4. *(Common trends)*

For all $t \in \{1, ..., \bar{t}\}$, $E(Y(0)|G, T = t) - E(Y(0)|G, T = t-1)$ does not depend on G.

Assumption 4'. *(Conditional common trends)*

For all $d \in \mathcal{S}(D)$ and all $t \in \{1, ..., \bar{t}\}$, $E(Y(d)|G, T = t, D(t-1) = d) - E(Y(d)|G, T = t-1, D(t-1) = d)$ does not depend on G.

Assumption 5. *(Stable treatment effect over time)*

For all $d \in \mathcal{S}(D)$ and all $t \in \{1, ..., \bar{t}\}$, $E(Y(d) - Y(0)|G, T = t, D(t-1) = d) = E(Y(d) - Y(0)|G, T = t-1, D(t-1) = d)$.

Assumption 6. *(Monotonicity and time invariance of unobservables)*

$Y(d) = h_d(U_d, T)$, with $U_d \in \mathbb{R}$ and $h_d(u, t)$ strictly increasing in u for all $(d, t) \in \mathcal{S}(D) \times \mathcal{S}(T)$. Moreover, $U_d \perp\!\!\!\perp T|G, D(0)$.

Assumption 7. *(Data restrictions)*

1. *$\mathcal{S}(Y_{dgt}) = \mathcal{S}(Y) = [\underline{y}, \overline{y}]$ with $-\infty \leq \underline{y} < \overline{y} \leq +\infty$, for $(d, g, t) \in \mathcal{S}((D, G, T))$.*
2. *$F_{Y_{dgt}}$ is continuous on $\mathbb{R}$ and strictly increasing on $\mathcal{S}(Y)$, for $(d, g, t) \in \mathcal{S}((D, G, T))$.*

Assumption 4 is the usual common trends condition, under which the DID estimand identifies the average treatment effect on the treated in sharp designs where $D = G \times T$. Assumption 4 ' is a conditional version of this common trend condition, which requires that the mean of $Y(0)$ (resp. $Y(1)$) among untreated (resp. treated) units at period 0 follow the same evolution in both groups. Assumption 5 requires that in each group, the average treatment effect among units treated in period 0 remain stable between periods 0 and 1. Assumption 6 requires that potential outcomes be strictly increasing functions of a scalar and stationary unobserved term, as in Athey and Imbens (2006). Assumption 7 is a testable restriction on the distribution of Y that is necessary only for the Wald-CIC and $\tau_{q,CIC}$ estimands.

Theorem 1. *(de Chaisemartin and D'Haultfœuille 2018b) Suppose that Assumptions 1-3 hold.*

1. *If Assumptions 4 and 5 also hold, then $W_{DID} = \Delta$.*
2. *If Assumptions 4 ' also hold, then $W_{TC} = \Delta$.*
3. *If Assumptions 6-7 also hold, then $W_{CIC} = \Delta$ and $\tau_{q,CIC} = \tau_q$.*

Theorem 1 gives several sets of conditions under which we can identify Δ, using one of the three estimands above. It also shows that τ_q can be identified under the same conditions as those under which the Wald-CIC identifies Δ. Compared to the Wald-DID, the Wald-TC and Wald-CIC do not rely on the stable treatment effect assumption, which may be implausible. The choice between the Wald-TC and the Wald-CIC estimands should be based on the suitability of Assumption 4 ' and 6 in the application under consideration. Assumption 4 ' is not invariant to the scaling of the outcome, but it only restricts its mean. Assumption 6 is invariant to the scaling of the outcome, but it restricts its entire distribution. When the treatment and control groups have different outcome distributions conditional on D in the first period, the scaling of the outcome might have a large effect on the Wald-TC. The Wald-CIC is much less sensitive to the scaling of the outcome, so using this estimand might be preferable. On the other hand, when the two groups have similar outcome distributions conditional on D in the first period, using the Wald-TC might be preferable.

To test the assumptions underlying those estimands, one can test whether they are equal. If they are not, at least one of those assumptions must be violated. An alternative approach is to perform placebo tests. For instance, if three time periods are available (T =-1, 0, or 1), and if the treatment rate remains stable in both groups between T =-1 and 0, the numerators of the Wald-DID, Wald-TC, and Wald-CIC estimands for those two periods should be equal to zero.

2.2 Estimators

We now turn to the estimation of Δ and $\tau_{q,CIC}$ using plug-in estimators of the estimands above. Let $(Y_i, D_i, G_i, T_i)_{i=1\dots n}$ denote an i.i.d. sample of (Y, D, G, T) and define $\mathcal{I}_{gt} = \{i : G_i = g, T_i = t\}$ and $\mathcal{I}_{dgt} = \{i : D_i = d, G_i = g, T_i = t\}$. Let n_{gt} and n_{dgt} denote the size of $\mathcal{I}_{gt}$ and $\mathcal{I}_{dgt}$, for all $(d, g, t) \in \mathcal{S}(D) \times \{0, 1\}^2$.

First, let

$$\widehat{W}_{DID} = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}} Y_i - \frac{1}{n_{10}}\sum_{i\in\mathcal{I}_{10}} Y_i - \frac{1}{n_{01}}\sum_{i\in\mathcal{I}_{01}} Y_i + \frac{1}{n_{00}}\sum_{i\in\mathcal{I}_{00}} Y_i}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}} D_i - \frac{1}{n_{10}}\sum_{i\in\mathcal{I}_{10}} D_i - \frac{1}{n_{01}}\sum_{i\in\mathcal{I}_{01}} D_i + \frac{1}{n_{00}}\sum_{i\in\mathcal{I}_{00}} D_i}$$

be the estimator of the Wald-DID. Second, for any $d \in \mathcal{S}(D)$ let $\widehat{\delta}_d = (1/n_{d01})\sum_{i\in\mathcal{I}_{d01}} Y_i - (1/n_{d00})\sum_{i\in\mathcal{I}_{d00}} Y_i$. Then, let

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

be the estimator of the Wald-TC. Third, for all $(d, g, t) \in \mathcal{S}(D) \times \{0, 1\}^2$, let $\widehat{F}_{Y_{dgt}}(y) = \frac{1}{n_{dgt}}\sum_{i\in\mathcal{I}_{dgt}} \mathbb{1}\{Y_i \le y\}$ denote the empirical cdf of Y_{dgt}. Let

$$\widehat{Q}_d(y) = \max\left(\widehat{F}^{-1}_{Y_{d01}} \circ \widehat{F}_{Y_{d00}}(y), \min\{Y_i : i \in \mathcal{I}_{d01}\}\right)$$

be the estimator of the quantile-quantile transform Q_d, and let

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

be the estimator of the Wald-CIC. Finally, let $\widehat{P}(D_{gt} = d) = n_{dgt}/n_{gt}$ and

$$\widehat{F}^{\text{pi}}_{CIC,d} = \frac{\widehat{P}(D_{11} = d)\widehat{F}_{Y_{d11}} - \widehat{P}(D_{10} = d)\widehat{F}_{\widehat{Q}_d(Y_{d10})}}{\widehat{P}(D_{11} = d) - \widehat{P}(D_{10} = d)}.$$

The function $\widehat{F}^{\text{pi}}_{CIC,d}$ is the plug-in estimator of $F_{CIC,d}$ but it has the drawback of not being necessarily a proper cdf. It may not be nondecreasing and may not belong to $[0, 1]$. To avoid these issues, we consider a rearranged version $\widehat{F}^{\text{arr}}_{CIC,d}$ of $\widehat{F}^{\text{pi}}_{CIC,d}$, following Chernozhukov et al. (2010). Moreover, we let

$$\widehat{F}_{CIC,d}(y) = \max\left(\min(\widehat{F}^{\text{arr}}_{CIC,d}(y), 1), 0\right).$$

With this proper cdf at hand, let

$$\widehat{\tau}_q = \widehat{F}^{-1}_{CIC,d}(q) - \widehat{F}^{-1}_{CIC,d}(q)$$

be the estimator of τ_q.

de Chaisemartin and D'Haultfœuille (2018b) show that $\widehat{W}_{DID}$, $\widehat{W}_{TC}$, $\widehat{W}_{CIC}$, and $\widehat{\tau}_q$ are root-n consistent and asymptotically normal under standard regularity conditions.[3] de Chaisemartin and D'Haultfœuille (2018b) also establish the validity of the bootstrap to draw inference on Δ and τ_q based on these estimators. The `fuzzydid` command uses the bootstrap to compute the standard errors of all estimators, and the percentile bootstrap to compute confidence intervals.

3 Extensions

3.1 Special cases

When $P(D_{00} = d) = P(D_{01} = d) = 0$ for $d \in \{0, 1\}$, W_{TC} (resp. W_{CIC}) is not defined because δ_d (resp. Q_d) is not defined. In such cases, we can simply suppose that $\delta_0 = \delta_1$ (resp. $Q_0 = Q_1$). Then, the Wald-TC becomes equal to the Wald-DID, while the Wald-CIC becomes equal to:

$$\widetilde{W}_{CIC} = \frac{E(Y_{11}) - E(Q_{1-d}(Y_{10}))}{E(D_{11}) - E(D_{10})}.$$

3. de Chaisemartin and D'Haultfœuille (2018b) consider an estimator of τ_q based on $\widehat{F}^{\text{pi}}_{CIC,d}$ rather than $\widehat{F}_{CIC,d}$. However, these two estimators are equal on any compact set with probability tending to one whenever $F_{CIC,d}$ is strictly increasing. Thus, the two estimators of τ_q also coincide with probability tending to one, and their result also applies to the estimator considered here.

Similarly, we define

$$\widetilde{F}_{CIC,d}(y) = \frac{P(D_{11}=d)F_{Y_{d11}}(y) - P(D_{10}=d)F_{Q_{1-d}(Y_{d10})}(y)}{P(D_{11}=d) - P(D_{10}=d)},$$
$$\widetilde{\tau}_{CIC,q} = \widetilde{F}^{-1}_{CIC,1}(q) - \widetilde{F}^{-1}_{CIC,0}(q).$$

de Chaisemartin and D'Haultfœuille (2018b) show that $\widetilde{W}_{CIC} = \Delta$ and $\widetilde{\tau}_{CIC,q} = \tau_q$ under the same assumptions as above, and if $h_0(h_0^{-1}(y,1),0) = h_1(h_1^{-1}(y,1),0)$ for every $y \in \mathcal{S}(Y)$. Finally, $\widetilde{W}_{CIC}$ and $\widetilde{\tau}_{CIC,q}$ can be estimated exactly in the same way as W_{CIC} and $\tau_{CIC,q}$.

3.2 No "stable" control group

In some applications (see e.g. Enikolopov et al. 2011), the treatment rate increases in all groups, thus violating Assumption 2. Then, we can still express the Wald-DID as a linear combination of the LATEs of treatment and control group switchers. Specifically, let $S' = \{D(0) \neq D(1), G=0\}$ be the control group switchers, and $\Delta' = E\left(Y(1)-Y(0)|S', T=1\right)$ be their local average treatment effect. Under Assumptions 1, 3, 4 and 5, we have

$$W_{DID} = \alpha\Delta + (1-\alpha)\Delta',$$

where $\alpha = (E(D_{11}) - E(D_{10}))/[E(D_{11}) - E(D_{10}) - (E(D_{01}) - E(D_{00}))]$. Hence, the Wald-DID identifies a weighted sum of Δ and Δ'. Note however that if the treatment rate increases in the control group, $E(D_{01}) > E(D_{00})$ and $\alpha > 1$, so Δ' enters with a negative weight. In such a case, we may have $\Delta > 0$ and $\Delta' > 0$ and yet $W_{DID} < 0$. We will only have $W_{DID} = \Delta$ if $\Delta = \Delta'$.

We can also bound Δ under Assumption 4 ' if Assumption 2 fails. Specifically, suppose that $\mathcal{S}(Y) = [\underline{y}, \overline{y}]$. For any real number x, let $M_0(x) = \max(0,x)$ and let $m_1(x) = \min(1,x)$. Then, let

$$\underline{F}_{d01}(y) = M_0\left(1 - \lambda_{0d}(1 - F_{Y_{d01}}(y))\right) - M_0(1-\lambda_{0d})\mathbb{1}\{y < \overline{y}\},$$
$$\overline{F}_{d01}(y) = m_1\left(\lambda_{0d}F_{Y_{d01}}(y)\right) + (1 - m_1(\lambda_{0d}))\mathbb{1}\{y \geq \underline{y}\},$$
$$\underline{\delta}_d = \int y d\overline{F}_{d01}(y) - E(Y_{d00}),\ \overline{\delta}_d = \int y d\underline{F}_{d01}(y) - E(Y_{d00}),$$
$$\underline{W}_{TC} = \frac{E(Y_{11}) - E(Y_{10} + \overline{\delta}_{D_{10}})}{E(D_{11}) - E(D_{10})},\ \overline{W}_{TC} = \frac{E(Y_{11}) - E(Y_{10} + \underline{\delta}_{D_{10}})}{E(D_{11}) - E(D_{10})}.$$

de Chaisemartin and D'Haultfœuille (2018b) show that under Assumptions 1, 3, and 4 ', $\Delta \in [\underline{W}_{TC}, \overline{W}_{TC}]$.

These bounds can be estimated as follows. Let $\widehat{\lambda}_{0d} = \widehat{P}(D_{01}=d)/\widehat{P}(D_{00}=d)$,

$\widehat{\lambda}_{1d} = \widehat{P}(D_{11} = d)/\widehat{P}(D_{10} = d)$, and

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

Then let

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

We estimate the bounds by

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

Under regularity conditions, these estimators are root-n consistent and asymptotically normal. Moreover, the bootstrap is valid for inference (see de Chaisemartin and D'Haultfœuille 2018c, Section 2.2).

3.3 Non-binary treatment

The Wald-DID, Wald-TC and Wald-CIC still identify a causal parameter if D is not binary but is ordered and takes a finite number of values. Specifically, under Assumptions 1-3 and 4-5 (resp. 4 ' and 6-7), W_{DID} (resp. W_{TC}, and W_{CIC}) point identifies the following average causal response:

$$ACR = \sum_{d=1}^{\overline{d}} w_d E(Y_{11}(d) - Y_{11}(d-1)|D(0) < d \leq D(1)),$$

where $w_d = [P(D_{11} \geq d) - P(D_{10} \geq d)]/[E(D_{11}) - E(D_{10})]$.[4] When Assumption 2 fails to hold, it is still possible to bound ACR. We refer the reader to Section 3.2 in de Chaisemartin and D'Haultfœuille (2018c) for further details.

When the treatment takes a large number of values, its support may differ in the treatment and control groups, and there may be values of D in the treatment group for which δ_d or Q_d are not defined because no unit in the control group has that value of D. This situation includes in particular the special cases discussed in Section 3.1 above. We can then modify slightly W_{TC} and W_{CIC}. Namely, let us consider a recategorized treatment $\widetilde{D} = h(D)$ grouping together some values of D and let

$$\widetilde{\delta}_{\widetilde{d}} = E[Y_{01}|\widetilde{D} = \widetilde{d}] - E[Y_{00}|\widetilde{D} = \widetilde{d}].$$

4. For this result to hold, we have to slightly reinforce Assumption 1 by also assuming that $P(D_{11} \geq d) \geq P(D_{10} \geq d)$ for all $d \in \mathcal{S}(D)$.

We then replace $\delta_{D_{01}}$ by $\widetilde{\delta}_{\widetilde{D}_{01}}$ in the definition of W_{TC}. Then, W_{TC} still identifies Δ provided that $d \mapsto E[Y_{11}(d) - Y_{10}(d)|D(0) = d]$ only depends on $h(d)$. The same applies to W_{CIC}, by using $\widetilde{D}$ instead of D in $Q_d(.)$. Using this recategorized treatment also avoids estimating δ_d and Q_d on a small number of units, thus often lowering the standard errors of the estimators.

Finally, there may also be instances where the treatment has the same support in the treatment and in the control groups, but where bootstrap samples do not satisfy this requirement. For such bootstrap samples, W_{TC} and W_{CIC} cannot be estimated, and the `fuzzydid` command therefore sets them to 10^{15} or -10^{15} with probability 1/2. To avoid distorting inference, those bootstrap samples are not discarded, thus resulting in very large bootstrap standard errors. On the other hand, if such bootstrap samples are not too numerous, their impact on percentile-bootstrap confidence intervals is more limited. This situation is likely to arise when the treatment takes a large number of values. Here again, it may be useful to recategorize the treatment to avoid this issue.

3.4 Including covariates

The basic set-up can be extended to include covariates. Let X denote a vector of covariates, and for any random variable R, let $m^R_{gt}(x) = E(R_{gt}|X = x)$. Let also $\delta_d(x) = E(Y_{d01}|X = x) - E(Y_{d00}|X = x)$ and $\widetilde{\delta}(x) = E\left(\delta_{D_{10}}(X_{10})|X = x\right]$. Then define

$$W^X_{DID} = \frac{E(Y_{11}) - E(m^Y_{10}(X_{11})) - \left(E(m^Y_{01}(X_{11})) - E(m^Y_{00}(X_{11}))\right)}{E(D_{11}) - E(m^D_{10}(X_{11})) - \left(E(m^D_{01}(X_{11})) - E(m^D_{00}(X_{11}))\right)},$$

$$W^X_{TC} = \frac{E(Y_{11}) - E\left[m^Y_{10}(X_{11}) + \widetilde{\delta}(X_{11})\right]}{E(D_{11}) - E(m^D_{10}(X_{11}))}.$$

de Chaisemartin and D'Haultfœuille (2018c) show that W^X_{DID} (resp. W^X_{TC}) identifies Δ under the common support condition $\mathcal{S}(X_{gt}) = \mathcal{S}(X)$ for all (g,t) (resp. $\mathcal{S}(X_{dgt}) = \mathcal{S}(X)$ for all (d,g,t)) and conditional versions of Assumptions 1-3 and 4-5 (resp. 4 ').[5]

Let us turn to estimators of W^X_{DID} and W^X_{TC}. We first consider non-parametric estimators. Let us assume that $X \in \mathbb{R}^r$ is a vector of continuous covariates. Adding discrete covariates is easy by reasoning conditional on each corresponding cell. We take an approach similar to, e.g., Frölich (2007) by estimating in a first step conditional expectations by series estimators. For any positive integer K, let $p^K(x) = (p_{1K}(x), ..., p_{KK}(x))'$ be a vector of basis functions and $P^K_{gt} = (p^K(X_1), ..., p^K(X_n))$. For any random variable R, we estimate $m^R(x) = E(R|X = x)$ by the series estimator

$$\widehat{m}^R(x) = p^{K_n}(x)' \left(P^{K_n} P^{K_n}{}'\right)^- P^{K_n} \left(R_1, ..., R_n\right)',$$

where $(.)^-$ denotes the generalized inverse and K_n is an integer. We then estimate $m^R_{gt}(x) = E(R_{gt}|X = x)$ by the series estimator above on the subsample $\{i : G_i =$

5. de Chaisemartin and D'Haultfœuille (2018c) also propose a Wald-CIC estimand with covariates, but the corresponding estimator is not computed by the `fuzzydid` command.

$g, T_i = t\}$. $m^R_{dgt}(x) = E(R_{dgt}|X = x)$ is estimated similarly. Then our non-parametric estimators of W^X_{DID} and W^X_{TC} are defined as

$$\widehat{W}^X_{DID,NP} = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[Y_i - \widehat{m}^Y_{10}(X_i) - \widehat{m}^Y_{01}(X_i) + \widehat{m}^Y_{00}(X_i)\right]}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[D_i - \widehat{m}^D_{10}(X_i) - \widehat{m}^D_{01}(X_i) + \widehat{m}^D_{00}(X_i)\right]},$$

$$\widehat{W}^X_{TC,NP} = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[Y_i - \widehat{m}^Y_{10}(X_i) - \widehat{m}^D_{10}(X_i)\widehat{\delta}_1(X_i) - (1-\widehat{m}^D_{10}(X_i))\widehat{\delta}_0(X_i)\right]}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[D_i - \widehat{m}^D_{10}(X_i)\right]},$$

where $\widehat{\delta}_d(x) = \widehat{m}^Y_{d01}(x) - \widehat{m}^Y_{d00}(x)$. Under regularity conditions, these estimators are root-n consistent and asymptotically normal (see de Chaisemartin and D'Haultfœuille 2018c, Section 2.3).

Second, we consider semi-parametric estimators of W^X_{DID} and W^X_{TC}. Assume for instance that for $(d, g, t) \in \{0, 1\}^3$, $E(Y_{gt}|X) = X'\beta^Y_{gt}$, $E(Y_{dgt}|X) = X'\beta^Y_{dgt}$, and $E(D_{gt}|X) = X'\beta^D_{gt}$. Under this assumption, we have

$$W^X_{DID} = \frac{E(Y_{11}) - E\left(X'_{11}\beta^Y_{10}\right) - \left(E\left(X'_{11}\beta^Y_{01}\right) - E\left(X'_{11}\beta^Y_{00}\right)\right)}{E(D_{11}) - E\left(X'_{11}\beta^D_{10}\right) - \left(E\left(X'_{11}\beta^D_{01}\right) - E\left(X'_{11}\beta^D_{00}\right)\right)}$$

$$W^X_{TC} = \frac{E(Y_{11}) - E\left[X'_{11}\left(\beta^Y_{10} + X'_{11}\beta^D_{10}\left(\beta^Y_{101} - \beta^Y_{100}\right) + (1 - X'_{11}\beta^D_{10})\left(\beta^Y_{001} - \beta^Y_{000}\right)\right)\right]}{E(D_{11}) - E(X'_{11}\beta^D_{10})}.$$

Then, semi-parametric estimators of W^X_{DID} and W^X_{TC} can be defined as

$$\widehat{W}^X_{DID,OLS} = \frac{\sum_{i\in\mathcal{I}_{11}}\left[Y_i - X'_i\widehat{\beta}^Y_{10} - X'_i\widehat{\beta}^Y_{01} + X'_i\widehat{\beta}^Y_{00}\right]}{\sum_{i\in\mathcal{I}_{11}}\left[D_i - X'_i\widehat{\beta}^D_{10} - X'_i\widehat{\beta}^D_{01} + X'_i\widehat{\beta}^D_{00}\right]},$$

$$\widehat{W}^X_{TC,OLS} = \frac{\sum_{i\in\mathcal{I}_{11}} Y_i - \left[X'_i\widehat{\beta}^Y_{10} + X'_i\left(X'_i\widehat{\beta}^D_{10}(\widehat{\beta}^Y_{101} - \widehat{\beta}^Y_{100}) + (1 - X'_i\widehat{\beta}^D_{10})(\widehat{\beta}^Y_{001} - \widehat{\beta}^Y_{000})\right)\right]}{\sum_{i\in\mathcal{I}_{11}}\left[D_i - X'_i\widehat{\beta}^D_{10}\right]},$$

where for $(d, g, t) \in \{0, 1\}^3$, $\widehat{\beta}^Y_{gt}$ (resp. $\widehat{\beta}^Y_{dgt}$) denotes the coefficient of X in an OLS regression of Y on X in the subsample $\mathcal{I}_{gt}$ (resp. $\mathcal{I}_{dgt}$), and $\widehat{\beta}^D_{gt}$ denotes the coefficient of X in an OLS regression of D on X in the subsample $\mathcal{I}_{gt}$. When either Y or D is binary, one might prefer to posit a probit or a logit model for its conditional expectation functions in the various subsamples. Other semi-parametric estimators can be defined accordingly.

Finally, researchers may sometimes wish to include a large set of controls in their estimation, which may lead to violations of the common support assumptions $\mathcal{S}(X_{gt}) = \mathcal{S}(X)$ and $\mathcal{S}(X_{dgt}) = \mathcal{S}(X)$.[6] For instance, when the researcher wants to estimate the

6. Using a recategorized treatment $\widetilde{D} = h(D)$ may help alleviating this issue, by weakening the support condition to $\mathcal{S}(X_{\widetilde{d}gt}) = \mathcal{S}(X)$ for all $\widetilde{d} \in \mathcal{S}(\widetilde{D})$.

Wald-DID, there might be values of X for which all units belong to the treatment group, thus implying that for those values there are no control units to which the trends experienced by treatment group units can be compared. Let x_0 denote one such problematic value, i.e. $x_0 \in \mathcal{S}(X_{11})$ but $E(Y_{0t}|X = x_0)$ and $E(D_{0t}|X = x_0)$ are not defined for some $t \in \{0, 1\}$. To avoid dropping treatment group units with $X = x_0$, we use all control units to predict their counterfactual trends. Namely, in W^X_{DID} we replace $E(Y_{01}|X = x_0) - E(Y_{00}|X = x_0)$ and $E(D_{01}|X = x_0) - E(D_{00}|X = x_0)$ by $E(Y_{01}) - E(Y_{00})$ and $E(D_{01}) - E(D_{00})$. If instead, the researcher wants to estimate the Wald-TC, the same principle applies.

3.5 Multiple periods and groups

We now extend our initial setting to multiple periods and groups. We first define, at each period $t \in \{1, ..., \bar{t}\}$, the following "supergroup" variable

$$G^*_t = 1\{E(D_{gt}) > E(D_{gt-1})\} - 1\{E(D_{gt}) < E(D_{gt-1})\}.$$

Let $\mathcal{T} = \left\{t \in \{1, ..., \bar{t}\} : P(G^*_t = 0) > 0\right\}$ denote the subset of periods t for which there exists at least one group with stable treatment rate between $t-1$ and t. We let $S = \{D(T) \neq D(T-1), T \in \mathcal{T}\}$ denote the population of units switching between $T-1$ and $T \in \mathcal{T}$ and define Δ in this set-up as $\Delta = E[Y(1) - Y(0)|S]$. For any random variable R and any $(d, g, t) \in \{0, 1\} \times \{-1, 1\} \times \mathcal{T}$, we also define the following quantities:

$$\begin{aligned}
DID^*_R(g,t) =& E(R|G^*_t = g, T = t) - E(R|G^*_t = g, T = t-1) \\
& - (E(R|G^*_t = 0, T = t) - E(R|G^*_t = 0, T = t-1)), \\
\delta^*_{dt} =& E(Y|D = d, G^*_t = 0, T = t) - E(Y|D = d, G^*_t = 0, T = t-1), \\
Q^*_{dt}(y) =& F^{-1}_{Y|D=d,G^*_t=0,T=t} \circ F_{Y|D=d,G^*_t=0,T=t-1}(y), \\
W^*_{DID}(g,t) =& \frac{DID^*_Y(g,t)}{DID^*_D(g,t)}, \\
W^*_{TC}(g,t) =& \frac{E(Y|G^*_t = g, T = t) - E(Y + \delta^*_{Dt}|G^*_t = g, T = t-1)}{E(D|G^*_t = g, T = t) - E(D|G^*_t = g, T = t-1)}, \\
W^*_{CIC}(g,t) =& \frac{E(Y|G^*_t = g, T = t) - E(Q^*_{Dt}(Y)|G^*_t = g, T = t-1)}{E(D|G^*_t = g, T = t) - E(D|G^*_t = g, T = t-1)}.
\end{aligned}$$

When $P(G^*_t = g) = 0$, the three ratios above are not defined. Then, we simply let $W^*_{DID}(g,t) = W^*_{TC}(g,t) = W^*_{CIC}(g,t) = 0$.

Let us then introduce the following weights:

$$\begin{aligned}
w_t &= \frac{DID^*_D(1,t)P(G^*_t = 1, T = t) - DID^*_D(-1,t)P(G^*_t = -1, T = t)}{\sum_{t=1}^{\bar{t}} DID^*_D(1,t)P(G^*_t = 1, T = t) - DID^*_D(-1,t)P(G^*_t = -1, T = t)}, \\
w_{10|t} &= \frac{DID^*_D(1,t)P(G^*_t = 1, T = t)}{DID^*_D(1,t)P(G^*_t = 1, T = t) - DID^*_D(-1,t)P(G^*_t = -1, T = t)},
\end{aligned}$$

where again, we set $DID^*_D(g,t) = 0$ when $P(G^*_t = g) = 0$. The extensions of the Wald-DID, Wald-TC and Wald-CIC to multiple groups and periods are defined as

$$W^*_{DID} = \sum_{t \in \mathcal{T}} w_t \left[w_{10|t} W^*_{DID}(1,t) + (1 - w_{10|t}) W^*_{DID}(-1,t) \right],$$
$$W^*_{TC} = \sum_{t \in \mathcal{T}} w_t \left[w_{10|t} W^*_{TC}(1,t) + (1 - w_{10|t}) W^*_{TC}(-1,t) \right],$$
$$W^*_{CIC} = \sum_{t \in \mathcal{T}} w_t \left[w_{10|t} W^*_{CIC}(1,t) + (1 - w_{10|t}) W^*_{CIC}(-1,t) \right].$$

Finally, we consider the following assumption, which replaces Assumption 2.

Assumption 8. *(Existence of "stable" groups and independence between groups and time)*

$\mathcal{T} \neq \emptyset$, $\mathcal{S}(D|G^*_t \neq 0, T = t-1) \subset \mathcal{S}(D|G^*_t = 0, T = t-1)$ *for all* $t \in \mathcal{T}$, *and* $G \perp\!\!\!\perp T$.

Theorem 2 below shows that under our previous conditions plus Assumption 8, the three estimands point identify Δ. This theorem is proved for the Wald-DID and Wald-TC in de Chaisemartin and D'Haultfœuille (2018a), and can be proved along the same lines for the Wald-CIC.[7]

Theorem 2. *Suppose that Assumptions 3 and 8 hold.*

1. *If Assumptions 4 and 5 are satisfied,* $W^*_{DID} = \Delta$.
2. *If Assumption 4 ' is satisfied,* $W^*_{TC} = \Delta$.
3. *If Assumptions 6 and 7 are satisfied,* $W^*_{CIC} = \Delta$.

To estimate W^*_{DID}, W^*_{TC}, and W^*_{CIC}, we suppose that the $(G^*_t)_{t=1...\bar{t}}$ are known. This is the case in applications where the treatment is constant at the group × period level, as is for instance the case in the example we revisit in Section 5. When the $(G^*_t)_{t=1...\bar{t}}$ are unknown, it is also possible to estimate them consistently, without affecting the asymptotic distribution of the estimators of W^*_{DID}, W^*_{TC} and W^*_{CIC}. We refer to Section 2.1 in de Chaisemartin and D'Haultfœuille (2018c) for details.

Let us focus on the estimator of W^*_{DID}. The estimators of W^*_{TC} and W^*_{CIC} are constructed following exactly the same logic. For any random variable R and any $(g,t) \in \{-1, 0, 1\} \times \mathcal{T}$, let

$$\widehat{DID}^*_R(g,t) = \frac{1}{n^*_{gt,t}} \sum_{i \in \mathcal{I}^*_{gt,t}} R_i - \frac{1}{n^*_{gt,t-1}} \sum_{i \in \mathcal{I}^*_{gt,t-1}} R_i - \left[\frac{1}{n^*_{0t,t}} \sum_{i \in \mathcal{I}^*_{0t,t}} R_i - \frac{1}{n^*_{0t,t-1}} \sum_{i \in \mathcal{I}^*_{0t,t-1}} R_i \right],$$

7. de Chaisemartin and D'Haultfœuille (2018a) obtain the same result on slightly different estimands and without assuming $G \perp\!\!\!\perp T$. Under this additional condition, their estimands are equal to the Wald-DID and Wald-TC considered here. Theorem 2 is also similar to Theorem S1 in de Chaisemartin and D'Haultfœuille (2018c), but they consider slightly different weights and prove the result under stronger conditions.

where $\mathcal{I}^*_{gt,t'} = \{i : G^*_{ti} = g, T_i = t'\}$ and $n^*_{gt,t'}$ is the size of $\mathcal{I}^*_{gt,t'}$. We let, for $g \in \{-1,0,1\}$, $\widehat{P}(G^*_t = g, T = t) = n^*_{gt,t}/n$. We estimate w_t and $w_{10|t}$ by

$$\widehat{w}_t = \frac{\widehat{DID}^*_D(1,t)\widehat{P}(G^*_t = 1, T = t) - \widehat{DID}^*_D(-1,t)\widehat{P}(G^*_t = -1, T = t)}{\sum_{t=1}^{\bar{t}} \widehat{DID}^*_D(1,t)\widehat{P}(G^*_t = 1, T = t) - \widehat{DID}^*_D(-1,t)\widehat{P}(G^*_t = -1, T = t)},$$

$$\widehat{w}_{10|t} = \frac{\widehat{DID}^*_D(1,t)\widehat{P}(G^*_t = 1, T = t)}{\widehat{DID}^*_D(1,t)\widehat{P}(G^*_t = 1, T = t) - \widehat{DID}^*_D(-1,t)\widehat{P}(G^*_t = -1, T = t)}.$$

We then estimate $W^*_{DID}(g,t)$ by $\widehat{W}^*_{DID}(g,t) = \widehat{DID}^*_Y(g,t)/\widehat{DID}^*_D(g,t)$, and we let

$$\widehat{W}^*_{DID} = \sum_{t \in \mathcal{T}} \widehat{w}_t \left[\widehat{w}_{10|t}\widehat{W}^*_{DID}(1,t) + (1 - \widehat{w}_{10|t})\widehat{W}^*_{DID}(-1,t)\right].$$

4 The `fuzzydid` command

The `fuzzydid` command is compatible with Stata 13.1 and later versions. It uses the `moremata` Stata command to compute estimators with covariates. If this command is not already installed, one must type `ssc install moremata` in Stata's command line.

4.1 Syntax

The syntax of `fuzzydid` is as follows:

<u>`fuzzydid`</u> `Y G_list T D` [*if*] [*in*] [`, did tc cic lqte new_categ(`*numlist*`)`
`numerator partial nose cluster(`*varname*`) breps(`*#*`) eqtest`
`continuous(`*varlist*`) qualitative(`*varlist*`) x_param(`*reg1 reg2 reg3*`) sieves`
`sieve_order(`*#*`)`]

4.2 Description

`fuzzydid` estimates Δ or τ_q using one or several of the estimators defined in Sections 2 and 3 above. It also computes their standard errors and confidence intervals.

`Y` is the outcome variable.

`G_list` is the list of group variables. With two periods, it simply corresponds to the variable G described above. With more than two periods, `G_list` should list two variables G_b and G_f, with $G_b = G^*_T$ and $G_f = G^*_{T+1}$. Given the definition of G^*_t, G_b and G_f are respectively missing when $T = 0$ and $T = \bar{t}$. To clarify the definition of these two variables, we consider an example with three groups and three periods. The treatment rate and the values of G_b and G_f for each group and time period are respectively given in the left and right panel of Table 1 below. For instance, group 0 has $G_b = 0$ and $G_f = 1$ in period 1, because its treatment rate is stable between periods 0 and 1, but increases between periods 1 and 2. The support of G, G^*_T and G^*_{T+1} should be included in $\{-1, 0, 1\}$.

Table 1: (G_b, G_f) in an example

	Treatment rates			(G_b, G_f)		
	Group 0	Group 1	Group 2	Group 0	Group 1	Group 2
Period 0	0.3	0.6	0.2	(., 0)	(., -1)	(., 1)
Period 1	0.3	0.4	0.5	(0, 1)	(-1, 0)	(1, 1)
Period 2	0.4	0.4	0.7	(1, .)	(0, .)	(1, .)

Notes: "." stands for missing value.

`T` is the time period variable, with values in $\{0, ..., \bar{t}\}$.

`D` is the treatment variable. It can be any ordered variable.

4.3 Options

General options

`did` computes $\widehat{W}_{DID}$ if no covariates are included in the estimation. If some covariates are included, it computes $\widehat{W}^X_{DID,NP}$, $\widehat{W}^X_{DID,OLS}$, or another estimator with covariates depending on the options specified by the user.

`tc` computes $\widehat{W}_{TC}$ if no covariates are included in the estimation. In the special case where D is binary and $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command actually computes $\widehat{W}_{DID}$, following the discussion in Section 3.1. If some covariates are included, it computes $\widehat{W}^X_{TC,NP}$, $\widehat{W}^X_{TC,OLS}$, or another estimator with covariates depending on the options specified by the user.

`cic` computes $\widehat{W}_{CIC}$. In the special case where D is binary and $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command actually computes $\widetilde{W}_{CIC}$, following the discussion in Section 3.1. This option can only be specified when no covariates are included in the estimation.

`lqte` computes $\widehat{\tau}_q$, for $q \in \{0.05, 0.10, ..., 0.95\}$. This option can only be specified when D, G, and T are binary, and no covariates are included in the estimation. When $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command computes $\widetilde{\tau}_{q,CIC}$, following the discussion in Section 3.1.

At least one of the four options above must be specified. If several of these options are specified, the command computes all the estimators requested by the user.

`new_categ(`*numlist*`)` groups some values of the treatment together when estimating δ_d and Q_d. This option may be useful when the treatment takes a large number of values, as explained in Section 3.3. The user needs to specify the upper bound of each set of values of the treatment she wants to group. For instance, if D takes the values $\{0, 1, 2, 3, 4.5, 7, 8\}$, and she wants to group together units with $D = \{0, 1, 2\}$, $\{3, 4.5\}$, and $\{7, 8\}$ when estimating δ_d and Q_d, she needs to write `new_categ(2 4.5 8)`.

`numerator` computes only the numerators of the $\widehat{W}_{DID}$, $\widehat{W}_{TC}$ and $\widehat{W}_{CIC}$ estimators. As explained in Section 3.3.3 in de Chaisemartin and D'Haultfœuille (2018b), this option is useful to conduct placebo tests of the assumptions underlying each estimator.

`partial` computes the bounds of Δ defined in Section 3.2, $\underline{\widehat{W}}_{TC}$ and $\overline{\widehat{W}}_{TC}$. This option can only be specified when no covariates are included in the estimation.

`nose` computes only the estimators, not their standard errors.

`cluster(`*varname*`)` computes the standard errors of the estimators using a block bootstrap at the *varname* level. Only one clustering variable is allowed.

`breps(`*#*`)` specifies the number of bootstrap replications. The default is 50.

`eqtest` performs an equality test between the estimands, when the user specifies at least two of the `did`, `tc`, and `cic` options.

Options specific to estimators with covariates

`continuous(`*varlist*`)` specifies the names of all the continuous covariates that need to be included in the estimation.

`qualitative(`*varlist*`)` specifies the names of all the qualitative covariates that need to be included in the estimation. For each variable, indicator variables are created for each value except one, and included as controls in the estimation.

`x_param(`*reg1 reg2 reg3*`)` specifies which parametric method should be used to estimate the conditional expectations in W^X_{DID} or W^X_{TC}. *reg1* specifies which method should be used to estimate $E(Y_{gt}|X)$ and $E(Y_{dgt}|X)$. *reg2* specifies which method should be used to estimate $E(D_{gt}|X)$. When D is not binary, *reg3* specifies which method should be used to estimate $\{P(D_{gt} = d|X)\}_{d\in\{1,...,\overline{d}\}}$. The possible methods are: `ols`, `logit`, and `probit`. For instance, if the user writes `x_param(ols logit logit)`, the command estimates $E(Y_{gt}|X)$ and $E(Y_{dgt}|X)$ by OLS, and $E(D_{gt}|X)$ and $\{P(D_{gt} = d|X)\}_{d\in\{1,...,\overline{d}\}}$ by a logistic regession. The `logit` and `probit` options can only be used with binary variables.

`sieves` indicates that the conditional expectations in W^X_{DID} and W^X_{TC} should be estimated nonparametrically (see Section 3.4 above).

When covariates are included in the estimation, and neither `x_param` nor `sieves` is specified, the command estimates by default all conditional expectations by OLS.

`sieve_order(`*#*`)` specifies the order of the sieve basis, when the option `sieves` is used. It must be greater than or equal to 2. For a given order L, the number of basis functions is given by $\binom{p_c+L}{L}$ where p_c is the number of continuous covariates. The command does not allow for more than min $\{4800, n/5\}$ basis functions, where n is the number of observations. If this option is not specified, the choice of the sieve order is done via 5-fold cross-validation with a mean squared error loss function.

4.4 Saved results

The `fuzzydid` command saves the following in `e()`:

1. `e(N)`, a scalar containing the number of observations used in the estimation.

2. If the user specifies at least one of the `did`, `tc`, and `cic` options, `fuzzydid` saves `e(res_table)`, a $k \times 6$ matrix, where k is equal to the number of options specified. The lines of the matrix correspond to each of the requested estimators. The columns of the matrix respectively store the value of each estimator, its bootstrap standard error, its t-statistic, its p-value, and the lower and upper bounds of its 95% confidence interval computed by percentile bootstrap.

3. If the user specifies the `eqtest` option together with at least two of the `did`, `tc`, and `cic` options, `fuzzydid` saves `e(eq_test_table)`, a $\binom{k}{2} \times 6$ matrix, where k is equal to the number of the `did`, `tc`, and `cic` options specified. The lines of the matrix correspond to the difference between each pair of estimators. The columns of the matrix respectively store the value of each difference, its bootstrap standard error, its t-statistic, its p-value, and the lower and upper bounds of its 95% confidence interval computed by percentile bootstrap.

4. If the user specifies the `lqte` option, the command saves `e(LQTE)`, an 19×6 matrix. The lines of the matrix correspond to $\widehat{\tau}_q$ for $q \in \{0.05, 0.10, ..., 0.95\}$. The columns of the matrix respectively store the value of each estimator, its bootstrap standard error, its t-statistic, its p-value, and the lower and upper bounds of its 95% confidence interval computed by percentile bootstrap.

5 Example

To illustrate the use of `fuzzydid`, we use the same dataset as Gentzkow et al. (2011) to study the effect of newspapers on electoral participation.

`turnout_dailies_1868-1928.dta` is a county-level data set. It contains two variables of interest, `pres_turnout` and `numdailies`, that respectively represent the turnout (Y) and the number of newspapers available (D) in each US county and at each presidential election from 1868 and 1928. First, we load the dataset and present summary statistics:

```
. use "turnout_dailies_1868-1928.dta", clear
. sum pres_turnout numdailies
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
pres_turnout |     16,872      .65014    .2210102   .0017981      2.518
  numdailies |     16,872    1.463134    2.210448          0         45
```

The average turnout in the 1868 to 1928 presidential elections across counties is 65.01%. The number of newspapers ranges from 0 to 45, and is on average equal to 1.46.

Second, we use `fuzzydid` to compute $\widehat{W}^*_{DID}$, $\widehat{W}^*_{TC}$, and $\widehat{W}^*_{CIC}$ using the first two time periods in the data set, the 1868 and 1872 elections. Following Section 3.5, we gather counties in three "supergroups". We define for that purpose the `G1872` variable, which is equal to 1 (resp. 0, -1) in counties whose number of newspapers increased (resp. remained stable, decreased) between the 1868 and 1872 elections. `numdailies` takes many values, so there are values taken by counties with `G1872`=1 or -1 that are not taken by any county with `G1872`=0. Therefore, we use `new_categ` to recategorize `numdailies` into four categories: 0, 1, 2, and 3 or more newspapers.[8] Finally, we cluster the bootstrap at the county level, to allow for county-level correlation over time.

```
. gen G1872=(fd_numdailies>0)-(fd_numdailies<0) if (year==1872)&fd_numdailies!=.
> &sample==1

. sort cnty90 year

. replace G1872=G1872[_n+1] if cnty90==cnty90[_n+1]&year==1868

. fuzzydid pres_turnout G1872 year numdailies, did tc cic new_categ(0 1 2 45)
> breps(200) cluster(cnty90)

Estimator(s) of the local average treatment effect with bootstrapped standard
errors.  Cluster variable: cnty90.  Number of observations:  1478 .

             |      LATE    Std_Err          t    p_value   lower_ic   upper_ic
-------------+------------------------------------------------------------------
       W_DID |  .0077888   .0122373     .63648   .5244636  -.0129246   .0344471
        W_TC |  .0172867   .0122715   1.408682   .1589292   -.003978   .0435382
       W_CIC |  .0042715   .0121027   .3529382   .7241348  -.0219849   .0248791
```

The columns of the output table respectively show the value of each estimator, its bootstrap standard error, its t-statistic, its p-value, and the lower and upper bounds of its 95% confidence interval. All point estimates are positive, but none are statistically significant, presumably because this restricted sample with two time periods is too small.

Third, we compute estimators of the LQTEs, using again the 1868 and 1872 elections. We restrict the sample to counties with `G1872`=0 or 1, and we use a binary treatment variable `numdailies_bin` (0 newspaper, 1 or more), because LQTEs can only be estimated with two groups and with a binary treatment.

```
. fuzzydid pres_turnout G1872 year numdailies_bin, lqte breps(200) cluster(cnty90)

Estimators of local quantile treatment effects with bootstrapped standard
errors.  Cluster variable: cnty90.  Number of observations:  1424 .

             |      LQTE    Std_Err          t    p_value   lower_ic   upper_ic
-------------+------------------------------------------------------------------
```

8. Only 17.8% of observations have 3 or more newspapers. Results do not change much if instead we recategorize `numdailies` into five categories: 0, 1, 2, 3, and 4 or more newspapers.

```
        q_20 |       .005    .063113   .0792229   .9368553     -.0825      .1655
        q_40 |      -.052   .0493409  -1.053894   .2919316  -.1244999      .0675
        q_60 |       .011   .0482445   .2280046   .8196427     -.0995        .08
        q_80 |        .02   .0355669   .5623207   .5738975      -.087       .077
```

To preserve space, we only report $\widehat{\tau}_{0.2}$, $\widehat{\tau}_{0.4}$, $\widehat{\tau}_{0.6}$, and $\widehat{\tau}_{0.8}$, but the command computes $\widehat{\tau}_q$ for $q \in \{0.05, 0.10, ..., 0.95\}$. $\widehat{\tau}_{0.4}$ is negative while the other estimates are positive, thus suggesting that numdailies_bin may have heterogeneous effects along the distribution of the outcome. However, none of the point estimates are statistically significant.

Fourth, we compute $\widehat{W}^*_{DID}$, $\widehat{W}^*_{TC}$, and $\widehat{W}^*_{CIC}$ on the full sample. On that purpose, we define the Gb and Gf variables described in Section 4.2. Gb is equal to 1 (resp. 0, -1) for county c $\times$ election-year t observations such that the number of newspapers increased (resp. remained stable, decreased) between election-years $t-1$ and t in that county. Gf is the lead of Gb. We add the eqtest option, to test whether the estimators are significantly different.

```
. xtset cnty90 year

. gen Gb=.

. forvalue i=1872(4)1928 {
  2. replace Gb=(fd_numdailies>0)-(fd_numdailies<0) if (year==`i´)&fd_numdailies
> !=.&sample==1
  3. }

. gen Gf=f4.Gb

. fuzzydid pres_turnout Gb Gf year numdailies, did tc cic new_categ(0 1 2 45)
> breps(200) cluster(cnty90) eqtest
Estimator(s) of the local average treatment effect with bootstrapped standard
errors.  Cluster variable: cnty90.  Number of observations:  16872 .
             |      LATE    Std_Err          t    p_value   lower_ic   upper_ic
-------------+------------------------------------------------------------------
       W_DID |  .0037507   .0012813   2.927357   .0034186   .0009971   .0057828
        W_TC |  .0053305   .0013276   4.015155   .0000594   .0023461   .0075914
       W_CIC |   .004215    .001477   2.853841   .0043194   .0009549   .0067769
Estimators equality test
             |     Delta    Std_Err          t    p_value   lower_ic   upper_ic
-------------+------------------------------------------------------------------
      DID_TC | -.0015798   .0003504  -4.507975   6.54e-06  -.0023752  -.0009441
     DID_CIC | -.0004643   .0007151  -.6492892   .5161515  -.0018629   .0008515
      TC_CIC |  .0011155   .0006505    1.71487    .086369  -.0002291   .0023088
```

The Wald-DID is equal to 0.0038. According to that estimator, increasing the number of newspapers available in a county by one increases voters' turnout in presidential elections by 0.38 percentage points. This estimator is significantly different from 0 at the 5% level. The Wald-TC is larger (0.0053), and significantly different from the Wald-DID (t-stat=-4.51). The Wald-CIC lies in between (0.0042), and this estimator is not significantly different from the other two.

Gentzkow et al. (2011) allow for state-specific trends in their specification, so we compute $\widehat{W}^*_{DID}$ and $\widehat{W}^*_{TC}$ with state indicators as controls, which is equivalent to allowing for state-specific trends.[9]

9. On the other hand, fuzzydid does not compute $\widehat{W}^*_{CIC}$ with controls.

```
. fuzzydid pres_turnout Gb Gf year numdailies, did tc new_categ(0 1 2 45)
> qualitative(st1-st48) breps(200) cluster(cnty90) eqtest
No method was specified to estimate E(Y|X), E(D|X) and P(D=d|X). Method set to
> default: ols.
Estimator(s) of the local average treatment effect with bootstrapped standard
errors.  Cluster variable: cnty90.  Number of observations:  16872 .  Controls
included in the estimation:  st1 st2 st3 st4 st5 st6 st7 st8 st9 st10 st11
st12 st13 st14 st15 st16 st17 st18 st19 st20 st21 st22 st23 st24 st25 st26
st27 st28 st29 st30 st31 st32 st33 st34 st35 st36 st37 st38 st39 st40 st41
st42 st43 st44 st45 st46 st47 st48 .
             |      LATE    Std_Err          t    p_value   lower_ic   upper_ic
-------------+------------------------------------------------------------------
       W_DID |  .0026383   .0012213   2.160195   .0307575   .0002316   .0048236
        W_TC |  .0043428   .0014116   3.076507   .0020944   .0015519   .0066773
Estimators equality test
             |     Delta    Std_Err          t    p_value   lower_ic   upper_ic
-------------+------------------------------------------------------------------
      DID_TC | -.0017046   .0009193   -1.85417   .0637148  -.0034308   .0000123
```

With those controls, $\widehat{W}^*_{DID} = 0.0026$ and $\widehat{W}^*_{TC} = 0.0043$, and the two estimators are significantly different at the 10% level (t-stat=-1.85).

Finally, we compute a placebo Wald-DID (resp. Wald-TC) estimator, to assess if Assumptions 4 and 5 (resp. Assumption 4 ') are plausible in this application. Instead of using the turnout in county g and election-year t as the outcome variable, our placebo estimators use the turnout in the same county in the previous election. Moreover, only counties where the number of newspapers did not change between $t-2$ and $t-1$ are included in the estimation. Therefore, our placebo estimators compare the evolution of turnout from $t-2$ to $t-1$, between counties where the number of newspapers increased or decreased between $t-1$ and t and counties where that number remained stable, restricting the sample to counties where the number of newspapers remained stable from $t-2$ to $t-1$.

```
. gen fd_numdailies_l1=l4.fd_numdailies
. gen pres_turnout_l1=l4.pres_turnout
. gen Gb_placebo=.
. forvalue i=1872(4)1928 {
  2. replace Gb_placebo=(changedailies>0)-(changedailies<0) if (year==`i´)&fd_nu
> mdailies!=.&sample==1&fd_numdailies_l1==0&changeprestout!=.
  3. }
. gen Gf_placebo=f4.Gb_placebo
. fuzzydid pres_turnout_l1 Gb_placebo Gf_placebo year numdailies, did tc
> new_categ(0 1 2 45) qualitative(st1-st48) breps(200) cluster(cnty90)
No method was specified to estimate E(Y|X), E(D|X) and P(D=d|X). Method set to
> default: ols.
Estimator(s) of the local average treatment effect with bootstrapped standard
errors.  Cluster variable: cnty90.  Number of observations:  13221 .  Controls
included in the estimation:  st1 st2 st3 st4 st5 st6 st7 st8 st9 st10 st11
st12 st13 st14 st15 st16 st17 st18 st19 st20 st21 st22 st23 st24 st25 st26
st27 st28 st29 st30 st31 st32 st33 st34 st35 st36 st37 st38 st39 st40 st41
st42 st43 st44 st45 st46 st47 st48 .
             |      LATE    Std_Err          t    p_value   lower_ic   upper_ic
```

W_DID	-.00183	.0016594	-1.102842	.2700959	-.0051247	.0013008
W_TC	-.0008691	.0018412	-.4720226	.6369107	-.0041261	.0025142

The placebo Wald-DID is negative, indicating that the actual Wald-DID may be downward biased due to a violation of Assumptions 4 and 5. However, this placebo estimator is not statistically significant. The placebo Wald-TC is also negative and not statistically significant. It is twice smaller than the placebo Wald-DID, thus indicating that Assumption 4 ' may be more plausible than Assumptions 4 and 5 in this application.

6 Monte Carlo Simulations

This section exhibits the finite sample performance of the estimators of W_{DID}, W_{TC}, W_{CIC} and $\tau_{CIC,q}$. We consider for that purpose the folllowing DGP. Let (G,T) be uniform on $\{0,1\}^2$. Let $(U(0),U(1),V) \sim \mathcal{N}(0,\Sigma)$, with $\Sigma_{ii}=1$ for $i\in\{1,3\}$, $\Sigma_{22}=1.2$, $\Sigma_{12}=0$, $\Sigma_{13}=.5$ and $\Sigma_{23}=-.5$, and with $(U(0),U(1),V) \perp\!\!\!\perp (G,T)$. Then we let

$$Y(d) = d + G + T + U(d),$$
$$D(t) = \mathbb{1}\{V \geq 1 - G\times t\}.$$

In this DGP, all the assumptions in Section 2 hold. Therefore, W_{DID}, W_{TC}, and W_{CIC} all identify Δ, while $\tau_{CIC,q}$ identifies τ_q. We focus on the bias, mean square error, and coverage rate of estimators of Δ and τ_q for $q\in\{.25,.5,.75\}$, and for sample sizes equal to 400, 800, and 1,600. In this DGP, $\Delta \simeq .540$, $\tau_{.25}\simeq .481$, $\tau_{.5}\simeq .536$ and $\tau_{.75}\simeq .595$.

The results are displayed in Table 2. Even with small samples, the Wald-DID and Wald-TC estimators do not exhibit any systematic bias. Their RMSE are also very similar. The Wald-CIC, on the other hand, is more biased and has a RMSE which is 5 to 15% larger. This is probably due to the estimator of the nonlinear transform Q_d. This estimator is likely biased and imprecise in the tails, which may also explain the bias and high RMSE of $\widehat{\tau}_q$ for $n=400$. Note however that the bias of $\widehat{W}_{CIC}$, $\widehat{\tau}_{.25}$, $\widehat{\tau}_{.5}$, and $\widehat{\tau}_{.75}$ decreases quickly with the sample size. For $n=1,600$, the bias of these estimators is already negligible compared to their RMSE. Finally, the percentile bootstrap confidence intervals of all estimators are quite accurate, with all coverage rates lying between .92 and .97 when the nominal level is .95. The levels are slightly more distorted for the Wald-CIC and the $\widehat{\tau}_q$ but again, they become closer to 95% as the sample size increases.

Table 2: Results of the Monte Carlo simulations

		Estimators of Δ			Estimators of τ_q		
n	Statistic	$\widehat{W}_{DID}$	$\widehat{W}_{TC}$	$\widehat{W}_{CIC}$	$\widehat{\tau}_{.25}$	$\widehat{\tau}_{.5}$	$\widehat{\tau}_{.75}$
400	Bias	0.005	-0.002	0.174	0.002	-0.154	-0.497
	RMSE	0.651	0.613	0.682	0.712	0.867	1.223
	Cov. rate	0.948	0.948	0.921	0.971	0.967	0.917
800	Bias	0.015	0.01	0.088	-0.056	-0.029	-0.235
	RMSE	0.422	0.414	0.472	0.539	0.555	0.922
	Cov. rate	0.953	0.951	0.929	0.964	0.961	0.934
1600	Bias	-0.005	-0.005	0.034	-0.054	-0.012	-0.077
	RMSE	0.286	0.284	0.329	0.394	0.382	0.58
	Cov. rate	0.948	0.946	0.943	0.964	0.966	0.955

Notes: “Cov. rate” stands for coverage rates of (percentile bootstrap) confidence intervals, with a nominal level of 95%. The results are based on 1,000 samples and for each, 500 bootstrap samples are drawn to construct the confidence intervals. With our DGP, $\Delta \simeq .540$, $\tau_{.25} \simeq .481$, $\tau_{.5} \simeq .536$ and $\tau_{.75} \simeq .595$.

7 Conclusion

We have discussed how to use `fuzzydid` to estimate local average and quantile treatment effects in fuzzy differences-in-differences designs, following de Chaisemartin and D’Haultfœuille (2018b). In such designs, the popular Wald-DID estimand relies on a stable treatment effect assumption, which may not be plausible. Then, the Wald-TC and Wald-CIC estimands may be valuable alternatives, as they do not hinge upon this assumption. Similarly, when the data bears multiple groups and periods, the Wald-TC and Wald-CIC estimands may be valuable alternatives to commonly used two-way linear regressions. The `fuzzydid` command makes it easy to estimate those estimands.

8 References

Athey, S., and G. W. Imbens. 2006. Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica* 74: 431–497.

de Chaisemartin, C., and X. D’Haultfœuille. 2018a. Two-way fixed effects estimators with heterogeneous treatment effects. Working paper.

———. 2018b. Fuzzy Differences-in-differences. *Review of Economic Studies* 85(2): 999–1028.

———. 2018c. Supplement to “Fuzzy Differences-in-Differences”. *Review of Economic Studies Supplementary Material* 85(2).

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

Enikolopov, R., M. Petrova, and E. Zhuravskaya. 2011. Media and Political Persuasion: Evidence from Russia. *The American Economic Review* 101: 3253.

Frölich, M. 2007. Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics* 139: 35–75.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson. 2011. The Effect of Newspaper Entry and Exit on Electoral Politics. *The American Economic Review* 101: 2980.