Fuzzy differences-in-differences with Stata

Clément de Chaisemartin
University of California at Santa Barbara
Santa Barbara, CA
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 evaluates 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 increases more only in the treatment group. In such fuzzy designs, de Chaisemartin and D'Haultfœuille (2018b, *Review of Economic Studies* 85: 999–1028) 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 nonbinary treatment, multiple periods, and groups and covariates. In this article, we present the command `fuzzydid`, which computes the various corresponding estimators. We illustrate the use of the command by revisiting Gentzkow, Shapiro, and Sinkinson (2011, *American Economic Review* 101: 2980–3018).

Keywords: st0560, fuzzydid, 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 Wald DID, which is the DID of the outcome divided by the DID of the treatment.

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

In this article, we describe the `fuzzydid` command, which computes the estimators corresponding to these estimands and performs inference on the LATE and LQTE using the bootstrap. When one computes standard errors and confidence intervals, clustering along one dimension can be allowed for. One can also perform equality tests between the Wald DID, Wald TC, Wald CIC, and placebo tests. This is important for choosing between these different estimands because 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, 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, nonbinary treatments can be easily handled by modifying the parameter of interest. Third, when the assumptions are more credible conditional on some controls, one can modify the Wald DID, Wald TC, and Wald CIC estimands to incorporate such controls. The `fuzzydid` command handles these extensions.

Finally, results can be extended to applications with multiple periods and groups that are prevalent in applied work. Researchers then estimate treatment effects via linear regressions, including time and group fixed effects. de Chaisemartin and D'Haultfœuille (2018a) show that around 19% of all empirical articles published by the *American Economic Review* between 2010 and 2012 use this research design. They also show that these regressions are extensions of the Wald DID to multiple periods and groups and that they identify weighted averages of LATEs with possibly many negative weights.[1] Thus, 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 article is organized as follows. Section 2 presents the estimands and estimators considered by de Chaisemartin and D'Haultfœuille (2018b) in the simplest setup with two groups and periods, a binary treatment, and no covariates. Section 3 discusses the various extensions covered by the `fuzzydid` command. Section 4 presents `fuzzydid`. Section 5 illustrates `fuzzydid` by revisiting Gentzkow, Shapiro, and Sinkinson (2011), who estimate the effect of newspapers on electoral participation. Section 6 presents the finite sample performances of the various estimators through Monte Carlo simulations. Section 7 concludes.

---

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

2   Setup

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 $T \in \{0, \ldots, \bar{t}\}$ and into groups represented by a random variable $G \in \{0, \ldots, \bar{g}\}$. We start by considering the simple case where $\bar{t} = \bar{g} = 1$, thus implying that there are two groups and two periods. In this case, $G = 1$ for units in the treatment group, and $G = 0$ for units in the control group, respectively.

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 the cumulative distribution function (CDF) of R and its CDF conditional on A, respectively. 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 this article.

Assumption 1. *Fuzzy design*

$$E(D_{11}) > E(D_{10}) \quad \text{and} \quad E(D_{11}) - E(D_{10}) > E(D_{01}) - E(D_{00})$$

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 $D(0), \ldots, D(\bar{t})$ such that $D = D(T)$, $D(t) \perp\!\!\!\perp T | G$ ($t \in \{0, \ldots, \bar{t}\}$) and for all $t \in \{1, \ldots, \bar{t}\}$,

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

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 periods 0 and 1. Assumption 2 requires that the treatment rate remain constant in the control group and be strictly 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

---

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

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 he or she been born at $T = t$. Here again, only $D = D(T)$ is observed.

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

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

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

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

W_{DID} is the coefficient of D in a two-stage least-squares 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 Wald TC ratio defined by

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

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 two-stage least-squares regression of Y on D using T as the excluded instrument within the treatment group. δ_0 and δ_1 measure the evolution of the outcome among untreated and treated units in the control group, respectively. 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 Wald CIC defined by

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

where $Q_d(y) = F_{Y_{d01}}^{-1} \circ F_{Y_{d00}}(y)$ 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

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

and

$$\tau_{\text{CIC},q} = F_{\text{CIC},1}^{-1}(q) - F_{\text{CIC},0}^{-1}(q)$$

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

Assumption 4. *Common trends*

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

Assumption 4'. *Conditional common trends*

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

Assumption 5. *Stable treatment effect over time*

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

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 means of $Y(0)$ and $Y(1)$ among untreated and treated units at period 0 follow the same evolution in both groups, respectively. Assumption 5 requires that in each group, the average treatment effect among units treated in period 0 remains 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 $\tau_{q,\text{CIC}}$ estimands.

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

1. If assumptions 4 and 5 also hold, then $W_{\mathrm{DID}} = \Delta$.

2. If assumptions 4' also hold, then $W_{\mathrm{TC}} = \Delta$.

3. If assumptions 6–7 also hold, then $W_{\mathrm{CIC}} = \Delta$ and $\tau_{q,\mathrm{CIC}} = \tau_q$.

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 with 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 assumptions 4' and 6 in the application under consideration. Assumption 4' is not invariant to the scaling of the outcome, but it restricts only 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 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, then the numerators of the Wald DID, Wald TC, and Wald CIC estimands for those two periods should be equal to 0.

2.2 Estimators

We now turn to the estimation of Δ and $\tau_{q,\mathrm{CIC}}$ using plugin estimators of the estimands above. Let $(Y_i, D_i, G_i, T_i)_{i=1\ldots n}$ denote an independent and identically distributed 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

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

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

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

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

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

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

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

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

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

The function $\widehat{F}_{\mathrm{CIC},d}^{\mathrm{pi}}$ is the plugin estimator of $F_{\mathrm{CIC},d}$, 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 $\widehat{F}_{\mathrm{CIC},d}^{\mathrm{arr}}$ of $\widehat{F}_{\mathrm{CIC},d}^{\mathrm{pi}}$, following Chernozhukov, Fernández-Val, and Galichon (2010). Moreover, we let

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

With this proper CDF at hand, let

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

be the estimator of τ_q.

de Chaisemartin and D'Haultfœuille (2018b) show that $\widehat{W}_{\mathrm{DID}}$, $\widehat{W}_{\mathrm{TC}}$, $\widehat{W}_{\mathrm{CIC}}$, 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. de Chaisemartin and D'Haultfœuille (2018b) consider an estimator of τ_q based on $\widehat{F}_{\mathrm{CIC},d}^{\mathrm{pi}}$ rather than $\widehat{F}_{\mathrm{CIC},d}$. However, these two estimators are equal on any compact set with probability tending to one whenever $F_{\mathrm{CIC},d}$ 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.

3 Extensions

3.1 Including covariates

The basic setup can be extended to include covariates. Let $\mathbf{X}$ denote a vector of covariates, and for any random variable R, let $m_{gt}^R(\mathbf{x}) = E(R_{gt}|\mathbf{X} = \mathbf{x})$. Also, let $\delta_d(\mathbf{x}) = E(Y_{d01}|\mathbf{X} = \mathbf{x}) - E(Y_{d00}|\mathbf{X} = \mathbf{x})$ and $\widetilde{\delta}(\mathbf{x}) = E\{\delta_{D_{10}}(\mathbf{X}_{10})|\mathbf{X} = \mathbf{x}\}$. Then, define

$$W_{\mathrm{DID}}^X = \frac{E(Y_{11}) - E\{m_{10}^Y(\mathbf{X}_{11})\} - \left[E\{m_{01}^Y(\mathbf{X}_{11})\} - E\{m_{00}^Y(\mathbf{X}_{11})\}\right]}{E(D_{11}) - E\{m_{10}^D(\mathbf{X}_{11})\} - \left[E\{m_{01}^D(\mathbf{X}_{11})\} - E\{m_{00}^D(\mathbf{X}_{11})\}\right]}$$

$$W_{\mathrm{TC}}^X = \frac{E(Y_{11}) - E\left\{m_{10}^Y(\mathbf{X}_{11}) + \widetilde{\delta}(\mathbf{X}_{11})\right\}}{E(D_{11}) - E\{m_{10}^D(\mathbf{X}_{11})\}}$$

In their article's supplement, de Chaisemartin and D'Haultfœuille (2018b) show that W_{DID}^X or W_{TC}^X identifies Δ under the common support condition $\mathcal{S}(\mathbf{X}_{gt}) = \mathcal{S}(\mathbf{X})$ for all (g, t) and $\mathcal{S}(\mathbf{X}_{dgt}) = \mathcal{S}(\mathbf{X})$ for all (d, g, t) and conditional versions of assumptions 1–3 and 4–5 or 4', respectively.[4]

Let us turn to estimators of W_{DID}^X and W_{TC}^X. We first consider nonparametric estimators. Let us assume that $\mathbf{X} \in \mathbb{R}^r$ is a vector of continuous covariates. Adding discrete covariates is easy by reasoning conditional on each corresponding cell. Like for instance Frölich (2007), we first estimate conditional expectations by series estimators. For any positive integer K, let $p^K(\mathbf{x}) = \{p_{1K}(\mathbf{x}), \ldots, p_{KK}(\mathbf{x})\}'$ be a vector of basis functions and $\mathbf{P}_{gt}^K = \{p^K(\mathbf{X}_1), \ldots, p^K(\mathbf{X}_n)\}$. For any random variable R, we estimate $m^R(\mathbf{x}) = E(R|\mathbf{X} = \mathbf{x})$ by the series estimator

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

where $(.)^-$ denotes the generalized inverse and K_n is an integer. We then estimate $m_{gt}^R(\mathbf{x}) = E(R_{gt}|\mathbf{X} = \mathbf{x})$ by the series estimator above on the subsample $\{i: G_i = g, T_i = t\}$. $m_{dgt}^R(\mathbf{x}) = E(R_{dgt}|\mathbf{X} = \mathbf{x})$ is estimated similarly. Then, our nonparametric estimators of W_{DID}^X and W_{TC}^X are defined as

$$\widehat{W}_{\mathrm{DID,NP}}^X = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left\{Y_i - \widehat{m}_{10}^Y(\mathbf{X}_i) - \widehat{m}_{01}^Y(\mathbf{X}_i) + \widehat{m}_{00}^Y(\mathbf{X}_i)\right\}}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left\{D_i - \widehat{m}_{10}^D(\mathbf{X}_i) - \widehat{m}_{01}^D(\mathbf{X}_i) + \widehat{m}_{00}^D(\mathbf{X}_i)\right\}}$$

$$\widehat{W}_{\mathrm{TC,NP}}^X = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[Y_i - \widehat{m}_{10}^Y(\mathbf{X}_i) - \widehat{m}_{10}^D(\mathbf{X}_i)\widehat{\delta}_1(\mathbf{X}_i) - \{1 - \widehat{m}_{10}^D(\mathbf{X}_i)\}\widehat{\delta}_0(\mathbf{X}_i)\right]}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left\{D_i - \widehat{m}_{10}^D(\mathbf{X}_i)\right\}}$$

where $\widehat{\delta}_d(\mathbf{x}) = \widehat{m}_{d01}^Y(\mathbf{x}) - \widehat{m}_{d00}^Y(\mathbf{x})$. Under regularity conditions, these estimators are root-n consistent and asymptotically normal (see the supplement to de Chaisemartin and D'Haultfœuille [2018b, sec. 2.3]).

---

4. In their supplement, de Chaisemartin and D'Haultfœuille (2018b) also propose a Wald CIC estimand with covariates, but the corresponding estimator is not computed by the `fuzzydid` command.

Second, we consider semiparametric estimators of W_{DID}^X and W_{TC}^X. For instance, assume that for $(d, g, t) \in \{0, 1\}^3$, $E(Y_{gt}|\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}_{gt}^Y$, $E(Y_{dgt}|\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}_{dgt}^Y$, and $E(D_{gt}|\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}_{gt}^D$. Under this assumption, we have

$$
W_{\text{DID}}^X = \frac{E(Y_{11}) - E\left(\mathbf{X}_{11}'\boldsymbol{\beta}_{10}^Y\right) - \left(E\left(\mathbf{X}_{11}'\boldsymbol{\beta}_{01}^Y\right) - E\left(\mathbf{X}_{11}'\boldsymbol{\beta}_{00}^Y\right)\right)}{E(D_{11}) - E\left(\mathbf{X}_{11}'\boldsymbol{\beta}_{10}^D\right) - \left\{E\left(\mathbf{X}_{11}'\boldsymbol{\beta}_{01}^D\right) - E\left(\mathbf{X}_{11}'\boldsymbol{\beta}_{00}^D\right)\right\}}
$$

$$
W_{\text{TC}}^X =
$$
$$
\frac{E(Y_{11}) - E\left[\mathbf{X}_{11}'\left\{\boldsymbol{\beta}_{10}^Y + \mathbf{X}_{11}'\boldsymbol{\beta}_{10}^D\left(\boldsymbol{\beta}_{101}^Y - \boldsymbol{\beta}_{100}^Y\right) + (1 - \mathbf{X}_{11}'\boldsymbol{\beta}_{10}^D)\left(\boldsymbol{\beta}_{001}^Y - \boldsymbol{\beta}_{000}^Y\right)\right\}\right]}{E(D_{11}) - E(\mathbf{X}_{11}'\boldsymbol{\beta}_{10}^D)}
$$

Then, semiparametric estimators of W_{DID}^X and W_{TC}^X can be defined as

$$
\widehat{W}_{\text{DID,OLS}}^X = \frac{\sum_{i \in \mathcal{I}_{11}} \left(Y_i - \mathbf{X}_i'\widehat{\boldsymbol{\beta}}_{10}^Y - \mathbf{X}_i'\widehat{\boldsymbol{\beta}}_{01}^Y + \mathbf{X}_i'\widehat{\boldsymbol{\beta}}_{00}^Y\right)}{\sum_{i \in \mathcal{I}_{11}} \left(D_i - \mathbf{X}_i'\widehat{\boldsymbol{\beta}}_{10}^D - \mathbf{X}_i'\widehat{\boldsymbol{\beta}}_{01}^D + \mathbf{X}_i'\widehat{\boldsymbol{\beta}}_{00}^D\right)}
$$

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

where for $(d, g, t) \in \{0, 1\}^3$, $\widehat{\boldsymbol{\beta}}_{gt}^Y$ and $\widehat{\boldsymbol{\beta}}_{dgt}^Y$ denote the coefficient of $\mathbf{X}$ in an ordinary least-squares (OLS) regression of Y on $\mathbf{X}$ in the subsamples $\mathcal{I}_{gt}$ and $\mathcal{I}_{dgt}$, respectively, and $\widehat{\boldsymbol{\beta}}_{gt}^D$ denotes the coefficient of $\mathbf{X}$ in an OLS regression of D on $\mathbf{X}$ 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 semiparametric 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 $\mathcal{S}(\mathbf{X}_{gt}) = \mathcal{S}(\mathbf{X})$ and $\mathcal{S}(\mathbf{X}_{dgt}) = \mathcal{S}(\mathbf{X})$.[5] For instance, when the researcher wants to estimate the Wald DID, there might be values of $\mathbf{X}$ for which all units belong to the treatment group, thus implying that for those values, there are no control units with which the trends experienced by treatment group units can be compared. Let $\mathbf{x}_0$ denote one such problematic value; that is, $\mathbf{x}_0 \in \mathcal{S}(\mathbf{X}_{11})$ but $E(Y_{0t}|\mathbf{X} = \mathbf{x}_0)$ and $E(D_{0t}|\mathbf{X} = \mathbf{x}_0)$ are not defined for some $t \in \{0, 1\}$. To avoid dropping treatment group units with $\mathbf{X} = \mathbf{x}_0$, we use all control units to predict their counterfactual trends. Namely, in W_{DID}^X, we replace $E(Y_{01}|\mathbf{X} = \mathbf{x}_0) - E(Y_{00}|\mathbf{X} = \mathbf{x}_0)$ and $E(D_{01}|\mathbf{X} = \mathbf{x}_0) - E(D_{00}|\mathbf{X} = \mathbf{x}_0)$ 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.

---

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

3.2 Multiple periods and groups

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

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

Let $\mathcal{T} = \{t \in \{1, \ldots, \bar{t}\} : P(G_t^* = 0) > 0\}$ 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 setup 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:

$$\mathrm{DID}_R^*(g, t) = E(R|G_t^* = g, T = t) - E(R|G_t^* = g, T = t - 1)$$
$$- \{E(R|G_t^* = 0, T = t) - E(R|G_t^* = 0, T = t - 1)\}$$
$$\delta_{dt}^* = E(Y|D = d, G_t^* = 0, T = t) - E(Y|D = d, G_t^* = 0, T = t - 1)$$
$$Q_{dt}^*(y) = F_{Y|D=d,G_t^*=0,T=t}^{-1} \circ F_{Y|D=d,G_t^*=0,T=t-1}(y)$$
$$W_{\mathrm{DID}}^*(g, t) = \frac{\mathrm{DID}_Y^*(g, t)}{\mathrm{DID}_D^*(g, t)}$$
$$W_{\mathrm{TC}}^*(g, t) = \frac{E(Y|G_t^* = g, T = t) - E(Y + \delta_{Dt}^*|G_t^* = g, T = t - 1)}{E(D|G_t^* = g, T = t) - E(D|G_t^* = g, T = t - 1)}$$
$$W_{\mathrm{CIC}}^*(g, t) = \frac{E(Y|G_t^* = g, T = t) - E\{Q_{Dt}^*(Y)|G_t^* = g, T = t - 1\}}{E(D|G_t^* = g, T = t) - E(D|G_t^* = g, T = t - 1)}$$

When $P(G_t^* = g) = 0$, the three ratios above are not defined. Then, we simply let $W_{\mathrm{DID}}^*(g, t) = W_{\mathrm{TC}}^*(g, t) = W_{\mathrm{CIC}}^*(g, t) = 0$.

Let us then introduce the following weights,

$$w_t = \frac{\mathrm{DID}_D^*(1, t)P(G_t^* = 1, T = t) - \mathrm{DID}_D^*(-1, t)P(G_t^* = -1, T = t)}{\sum_{t=1}^{\bar{t}} \mathrm{DID}_D^*(1, t)P(G_t^* = 1, T = t) - \mathrm{DID}_D^*(-1, t)P(G_t^* = -1, T = t)}$$
$$w_{10|t} = \frac{\mathrm{DID}_D^*(1, t)P(G_t^* = 1, T = t)}{\mathrm{DID}_D^*(1, t)P(G_t^* = 1, T = t) - \mathrm{DID}_D^*(-1, t)P(G_t^* = -1, T = t)}$$

where again we set $\mathrm{DID}_D^*(g, t) = 0$ 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

$$W_{\mathrm{DID}}^* = \sum_{t \in \mathcal{T}} w_t \left\{ w_{10|t} W_{\mathrm{DID}}^*(1, t) + (1 - w_{10|t}) W_{\mathrm{DID}}^*(-1, t) \right\}$$
$$W_{\mathrm{TC}}^* = \sum_{t \in \mathcal{T}} w_t \left\{ w_{10|t} W_{\mathrm{TC}}^*(1, t) + (1 - w_{10|t}) W_{\mathrm{TC}}^*(-1, t) \right\}$$
$$W_{\mathrm{CIC}}^* = \sum_{t \in \mathcal{T}} w_t \left\{ w_{10|t} W_{\mathrm{CIC}}^*(1, t) + (1 - w_{10|t}) W_{\mathrm{CIC}}^*(-1, t) \right\}$$

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

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

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

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.[6]

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

1. If assumptions 4 and 5 are satisfied, $W_{\text{DID}}^* = \Delta$.

2. If assumption 4' is satisfied, $W_{\text{TC}}^* = \Delta$.

3. If assumptions 6 and 7 are satisfied, $W_{\text{CIC}}^* = \Delta$.

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 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's (2018b) supplement 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

$$\widehat{\text{DID}}_R^*(g, t) = \frac{1}{n_{gt,t}^*} \sum_{i \in \mathcal{I}_{gt,t}^*} R_i - \frac{1}{n_{gt,t-1}^*} \sum_{i \in \mathcal{I}_{gt,t-1}^*} R_i$$

$$- \left(\frac{1}{n_{0t,t}^*} \sum_{i \in \mathcal{I}_{0t,t}^*} R_i - \frac{1}{n_{0t,t-1}^*} \sum_{i \in \mathcal{I}_{0t,t-1}^*} R_i \right)$$

---

6. 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's (2018b) supplement, 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

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

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

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

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

3.3 Other extensions

We now briefly review some other extensions, for which more details can be found in de Chaisemartin and D'Haultfœuille (2018b) and its supplement.

Special cases

When $P(D_{00} = d) = P(D_{01} = d) = 0$ for $d \in \{0, 1\}$, W_{TC}, W_{CIC}, and $\tau_{\mathrm{CIC},q}$ are not defined, because δ_d and Q_d are not defined, respectively. In such cases, we can simply suppose that $\delta_0 = \delta_1$ and $Q_0 = Q_1$, respectively, and modify the estimators accordingly. Then the Wald TC becomes equal to the Wald DID, while the modified CIC estimands identify Δ and τ_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 $y \in \mathcal{S}(Y)$.

No "stable" control group

In some applications (see, for example, Enikolopov, Petrova, and Zhuravskaya [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 $\Delta' = E\{Y(1) - Y(0)|S', T = 1\}$ be their LATE. Under assumptions 1, 3, 4, and 5, we have

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

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 this case, we may have $\Delta > 0$ and $\Delta' > 0$ and yet $W_{\mathrm{DID}} < 0$. We will have only $W_{\mathrm{DID}} = \Delta$ if $\Delta = \Delta'$.

We can also bound Δ under assumption 4' if assumption 2 fails. We refer to de Chaisemartin and D'Haultfœuille (2018b) for such bounds and to de Chaisemartin and D'Haultfœuille's (2018b) supplement for their corresponding estimators.

Nonbinary 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, as shown in de Chaisemartin and D'Haultfœuille (2018b). When the treatment takes many 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 particularly includes the special cases discussed above. We can then slightly modify W_{TC} and W_{CIC}. Namely, let us consider a recategorized treatment $\widetilde{D} = h(D)$ grouping together some values of D, and let

$$\widetilde{\delta}_{\widetilde{d}} = E\left(Y_{01}|\widetilde{D} = \widetilde{d}\right) - E\left(Y_{00}|\widetilde{D} = \widetilde{d}\right)$$

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 $d \mapsto E\{Y_{11}(d) - Y_{10}(d)|D(0) = d\}$ depends only 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, these bootstrap samples are not discarded in the computation of the percentile-bootstrap confidence intervals, thus enlarging these intervals.[7] This situation is likely to arise when the treatment takes many values. Here again, it may be useful to recategorize the treatment to avoid this issue.

4   The fuzzydid command

The `fuzzydid` command is compatible with Stata 13.1 and later versions. It uses the `moremata` command (Jann 2005) to compute estimators with covariates. If this command is not already installed, one must type `ssc install moremata` in Stata's command line.

---

7. However, they are discarded when computing the bootstrap standard errors.

4.1 Syntax

The syntax of `fuzzydid` is as follows:

fuzzydid Y G T D $\big[\textit{if}\,\big]$ $\big[\textit{in}\,\big]$ $\big[$, did tc cic lqte <u>newc</u>ateg(*numlist*)
<u>num</u>erator <u>partial</u> nose <u>cl</u>uster(*varname*) breps(*#*) eqtest <u>tagobs</u>
<u>cont</u>inuous(*varlist*) <u>qual</u>itative(*varlist*) modelx(*reg1 reg2 reg3*) sieves
<u>sieveo</u>rder(*#*) $\big]$

Y is the outcome variable.

G is the group variable or variables. When the data bear only two groups and two
   periods, G merely corresponds to the variable G defined in section 2, an indicator
   for units in the treatment group. Outside of this special case, G should list the
   variables G_T^* and G_{T+1}^* defined in section 3.2. Below are some lines of code that
   users can follow to create these two variables:

   ```
   sort G T
   by G T: egen mean_D = mean(D)
   by G: generate lag_mean_D = mean_D[_n-1] if G==G[_n-1]&T-1==T[_n-1]
   generate G_T = sign(mean_D - lag_mean_D)
   generate G_Tplus1 = G_T[_n+1] if G==G[_n+1]&T+1==T[_n+1]
   ```

   Sometimes, there may not be groups where the treatment is perfectly stable between
   consecutive periods, thus implying that the Wald DID, Wald TC, and Wald CIC
   estimators cannot be computed with the `G_T` and `G_Tplus1` variables defined above.
   The user may then replace the fourth line of code above with

   generate G_T = (mean_D - lag_mean_D$> \varepsilon$)-(mean_D - lag_mean_D$< -\varepsilon$)

   where ε is a positive number small enough to consider that the mean treatment
   did not really change in groups where it changed by less than ε. See section 2.1 in
   de Chaisemartin and D'Haultfœuille's (2018b) supplement for one possible method
   to choose ε.

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

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

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.

4.3   Options

General options

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

tc computes $\widehat{W}_{\text{TC}}$ if no covariates are included in the estimation. If D is binary and $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command actually computes $\widehat{W}_{\text{DID}}$, following the discussion in section 3.3. If some covariates are included, it computes $\widehat{W}^X_{\text{TC,NP}}$, $\widehat{W}^X_{\text{TC,OLS}}$, or another estimator with covariates depending on the options specified by the user.

cic computes $\widehat{W}_{\text{CIC}}$. If D is binary and $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command actually computes $\widetilde{W}_{\text{CIC}}$, following the discussion in section 3.3. This option can be specified only when no covariates are included in the estimation.

lqte computes $\widehat{\tau}_q$, for $q \in \{0.05, 0.10, \dots, 0.95\}$. This option can be specified only 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 $\widetilde{\tau}_{q,\text{CIC}}$, following the discussion in section 3.3.

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

newcateg(*numlist*) groups some values of the treatment together when estimating δ_d and Q_d. This option may be useful when the treatment takes many values, as explained in section 3.3. One must specify the upper bound of each set of values of the treatment one wants to group. For instance, if D takes the values $\{0, 1, 2, 3, 4.5, 7, 8\}$, and one wants to group together units with $D = \{0, 1, 2\}$, $\{3, 4.5\}$, and $\{7, 8\}$ when estimating δ_d and Q_d, one must write newcateg(2 4.5 8).

numerator computes only the numerators of the $\widehat{W}_{\text{DID}}$, $\widehat{W}_{\text{TC}}$, and $\widehat{W}_{\text{CIC}}$ 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 3.3, $\underline{\widehat{W}}_{\text{TC}}$ and $\overline{\widehat{W}}_{\text{TC}}$. This option can be specified only 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 breps(50).

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

tagobs creates a new variable named tagobs, which identifies the observations used by
  fuzzydid.

Options specific to estimators with covariates

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

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

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

sieves indicates that the conditional expectations in W_{DID}^X and W_{TC}^X should be esti-
  mated nonparametrically (see section 3.1 above).

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

sieveorder(#) 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. By default, the choice of the sieve order is done via
  fivefold cross-validation with a mean squared error loss function.

4.4   Stored results

fuzzydid stores 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(b_LATE), a $k \times 1$ matrix, where k is equal to the number of options specified.
   The lines of the matrix correspond to each of the requested estimators. If nose
   is not specified, fuzzydid also saves e(se_LATE) and e(ci_LATE), which are a
   $k \times 1$ and a $k \times 2$ matrix, respectively. The lines of e(se_LATE) correspond to
   the bootstrap standard error associated with each of the requested estimators.

The columns of e(ci_LATE) store the lower and upper bounds, respectively, of the 95% confidence interval computed by percentile bootstrap for each requested estimator.

3. If the user specifies the eqtest option and at least two of the did, tc, and cic options, fuzzydid saves three matrices e(b_LATE_eqtest), e(se_LATE_eqtest), and e(ci_LATE_eqtest). The first two matrices have dimension $\binom{k}{2} \times 1$ while the third has dimension $\binom{k}{2} \times 2$, where k is equal to the number of the did, tc, and cic options specified. The matrices e(b_LATE_eqtest) and e(se_LATE_eqtest) store the value of the difference between each pair of estimators and the associated bootstrap standard error, respectively. The columns of e(ci_LATE_eqtest) store the lower and upper bounds, respectively, of the 95% confidence interval computed by percentile bootstrap associated with each difference.

4. If the user specifies the lqte option, the command saves e(b_LQTE), a 19×1 matrix. The lines of the matrix store the value of $\widehat{\tau}_q$ for $q \in \{0.05, 0.10, \ldots, 0.95\}$. If nose is not specified, fuzzydid also saves e(se_LQTE) and e(ci_LQTE), a 19×1 and a 19×2 matrix, respectively. The lines of e(se_LQTE) correspond to the bootstrap standard error associated with $\widehat{\tau}_q$ for $q \in \{0.05, 0.10, \ldots, 0.95\}$. The columns of e(ci_LQTE) store the lower and upper bounds, respectively, of the 95% confidence interval computed by percentile bootstrap for each of the 19 LQTE estimators.

5   Example

To illustrate the use of fuzzydid, we rely on the same dataset as Gentzkow, Shapiro, and Sinkinson (2011) to study the effect of newspapers on electoral participation.

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

```
. use turnout_dailies_1868-1928

. summarize 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 $\widehat{W}_{\mathrm{DID}}^*$, $\widehat{W}_{\mathrm{TC}}^*$, and $\widehat{W}_{\mathrm{CIC}}^*$ using the first two time periods in the dataset, the 1868 and 1872 elections. We then define the G1872 variable, which is equal to 1 or 0 in counties whose number of newspapers increased or remained stable, respectively, between the 1868 and 1872 elections. For now, counties

where that number decreased are excluded from the analysis. `numdailies` takes many values, so there are values taken by counties with $G1872 = 1$ that are not taken by any county with $G1872 = 0$. Therefore, we use `newcateg()` to recategorize `numdailies` into four categories: zero, one, two, and three or more newspapers.[8] Finally, we cluster the bootstrap at the county level to allow for county-level correlation over time.

```
. generate G1872=(fd_numdailies>0) if (year==1872)&fd_numdailies!=.
> &fd_numdailies>=0&sample==1
(16,160 missing values generated)
. sort cnty90 year
. replace G1872=G1872[_n+1] if cnty90==cnty90[_n+1]&year==1868
(712 real changes made)
. fuzzydid pres_turnout G1872 year numdailies, did tc cic newcateg(0 1 2 45)
> breps(200) cluster(cnty90)
  (output omitted )
Estimator(s) of the local average treatment effect with bootstrapped standard
errors.  Cluster variable: cnty90.  Number of observations:  1424 .
```

	LATE	Std_Err	t	p_value	lower_ic	upper_ic
W_DID	.0047699	.0160903	.2964428	.766892	−.0230387	.0377381
W_TC	.0266618	.0164816	1.617671	.1057335	−.0021458	.0586236
W_CIC	.0133223	.0132744	1.003613	.3155653	−.0116416	.0348834

The columns of the output table show, respectively, 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. In this simple example with two periods and no controls, the computation of the estimators and of 200 bootstrap replications takes only about 3 seconds on a Dell Optiplex 9020 with an Intel Core i7-4790 CPU 3.60 GHz processor and 16 GB of RAM, using Stata/MP with 4 cores.

---

8. Only 17.8% of observations have 3 or more newspapers. Results do not change much if instead we recategorize `numdailies` into five categories: zero, one, two, three, and four or more newspapers.

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

```
. generate numdailies_bin=(numdailies>0)
. fuzzydid pres_turnout G1872 year numdailies_bin, lqte breps(200)
> cluster(cnty90)
  (output omitted)
```

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
(output omitted)						
q_20	.005	.063113	.0792229	.9368553	−.0825	.1655
(output omitted)						
q_40	−.052	.0493409	−1.053894	.2919316	−.1244999	.0675
(output omitted)						
q_60	.011	.0482445	.2280046	.8196427	−.0995	.08
(output omitted)						
q_80	.02	.0355669	.5623207	.5738975	−.087	.077
(output omitted)						

To preserve space, we report only $\widehat{\tau}_{0.2}$, $\widehat{\tau}_{0.4}$, $\widehat{\tau}_{0.6}$, and $\widehat{\tau}_{0.8}$, but the command computes $\widehat{\tau}_q$ for $q \in \{0.05, 0.10, \ldots, 0.95\}$. $\widehat{\tau}_{0.4}$ 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 $\widehat{W}^*_{\mathrm{DID}}$, $\widehat{W}^*_{\mathrm{TC}}$, and $\widehat{W}^*_{\mathrm{CIC}}$ on the full sample. For that purpose, we define the `G_T` and `G_Tplus1` variables described in section 4.2. `G_T` is equal to 1, 0, or -1 for county $c \times$ election-year t observations such that the number of newspapers increased, remained stable, or decreased, respectively, between election-years $t-1$ and t in that county. `G_Tplus1` is the lead of `G_T`. We add the `eqtest` option to test whether the estimators are significantly different.

```
. sort cnty90 year
. by cnty90 year: egen mean_D = mean(numdailies)
. by cnty90: generate lag_mean_D = mean_D[_n-1] if cnty90==cnty90[_n-1]&year-4==
> year[_n-1]
(1,243 missing values generated)
. generate G_T = sign(mean_D - lag_mean_D) if sample==1
(1,245 missing values generated)
. generate G_Tplus1 = G_T[_n+1] if cnty90==cnty90[_n+1]&year+4==year[_n+1]
(1,245 missing values generated)
```

```
. fuzzydid pres_turnout G_T G_Tplus1 year numdailies, did tc cic
> newcateg(0 1 2 45) breps(200) cluster(cnty90) eqtest
  (output omitted)
```

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 statistic $= -4.51$). The Wald CIC lies in between (0.0042), and this estimator is
not significantly different from the other two. In this more complicated example with
16 periods and almost 17,000 observations, computing the estimators and 200 bootstrap
replications still takes only around two minutes.

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

```
. fuzzydid pres_turnout G_T G_Tplus1 year numdailies, did tc newcateg(0 1 2 45)
> qualitative(st1-st48) breps(200) cluster(cnty90) eqtest
  (output omitted)
```

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	2.160196	.0307575	.0002316	.0048236
W_TC	.0043428	.0014116	3.076507	.0020944	.0015519	.0066772

Estimators equality test

	Delta	Std_Err	t	p_value	lower_ic	upper_ic
DID_TC	-.0017046	.0009193	-1.85417	.0637149	-.0034308	.0000123

---

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

With those controls, $\widehat{W}_{\mathrm{DID}}^* = 0.0026$ and $\widehat{W}_{\mathrm{TC}}^* = 0.0043$, and the two estimators are significantly different at the 10% level (t statistic $= -1.85$). Adding the control variables substantially increases the computation time to 79 minutes.

Finally, we compute a placebo Wald DID or Wald TC estimator to assess if assumptions 4 and 5 or assumption 4', respectively, is 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 $t-1$.

```
. xtset cnty90 year
  (output omitted)
. generate fd_numdailies_l1=l4.fd_numdailies
(2,486 missing values generated)
. generate pres_turnout_l1=l4.pres_turnout
(1,243 missing values generated)
. sort cnty90 year
. generate G_T_placebo = sign(mean_D - lag_mean_D) if
> sample==1&fd_numdailies_l1==0
(6,761 missing values generated)
. generate G_Tplus1_placebo = G_T_placebo[_n+1] if
> cnty90==cnty90[_n+1]&year+4==year[_n+1]
(6,761 missing values generated)
. fuzzydid pres_turnout_l1 G_T_placebo G_Tplus1_placebo year numdailies, did tc
> newcateg(0 1 2 45) qualitative(st1-st48) breps(200) cluster(cnty90)
(running estim_wrapper on estimation sample)
  (output omitted)
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	.270096	−.0051247	.0013008
W_TC	−.0008691	.0018412	−.4720226	.6369106	−.0041261	.0025142

The placebo Wald DID is negative, indicating that the actual Wald DID may be downward biased because of 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 $\tau_{\mathrm{CIC},q}$. For that purpose, we consider the following data-generating process (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 $i \in \{1,3\}$, $\Sigma_{22} = 1.2$, $\Sigma_{12} = 0$, $\Sigma_{13} = 0.5$, and $\Sigma_{23} = -0.5$ 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 $\tau_{\mathrm{CIC},q}$ identifies τ_q. We focus on the bias, mean square error, and coverage rate of estimators of Δ and τ_q for $q \in \{0.25, 0.5, 0.75\}$ and for sample sizes equal to 400, 800, and 1,600. In this DGP, $\Delta \simeq 0.540$, $\tau_{0.25} \simeq 0.481$, $\tau_{0.5} \simeq 0.536$, and $\tau_{0.75} \simeq 0.595$.

The results are displayed in table 1. Even with small samples, the Wald DID and Wald TC estimators do not exhibit any systematic bias. Their root mean squared errors (RMSE) are also similar. The Wald CIC, conversely, is more biased and has an RMSE that 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 $\widehat{W}_{\mathrm{CIC}}$, $\widehat{\tau}_{0.25}$, $\widehat{\tau}_{0.5}$, and $\widehat{\tau}_{0.75}$ decreases quickly with the sample size. For $n = 1600$, the bias of these estimators is already negligible compared with their RMSE. Finally, the percentile bootstrap confidence intervals of all estimators are quite accurate, with all coverage rates lying between 0.92 and 0.97 when the nominal level is 0.95. 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 1. Results of the Monte Carlo simulations

n	Statistic	Estimators of Δ			Estimators of τ_q		
		$\widehat{W}_{\mathrm{DID}}$	$\widehat{W}_{\mathrm{TC}}$	$\widehat{W}_{\mathrm{CIC}}$	$\widehat{\tau}_{0.25}$	$\widehat{\tau}_{0.5}$	$\widehat{\tau}_{0.75}$
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	−0.013	−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, $\Delta \simeq 0.540$, $\tau_{0.25} \simeq 0.481$, $\tau_{0.5} \simeq 0.536$, and $\tau_{0.75} \simeq 0.595$.

7   Conclusion

We have discussed how to use `fuzzydid` to estimate LATE and LQTE in fuzzy DID 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 because they do not hinge upon this assumption. Similarly, when the data bear 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.

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

de Chaisemartin, C., and X. D'Haultfœuille. 2018a. Two-way fixed effects estimators with heterogeneous treatment effects. ArXiv Working Paper No. arXiv:1510.01757. https://arxiv.org/abs/1803.08807.

———. 2018b. Fuzzy differences-in-difference. *Review of Economic Studies* 85: 999–1028.

Enikolopov, R., M. Petrova, and E. Zhuravskaya. 2011. Media and political persuasion: Evidence from Russia. *American Economic Review* 101: 3253–3285.

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. *American Economic Review* 101: 2980–3018.

Jann, B. 2005. moremata: Stata module (Mata) to provide various functions. Statistical Software Components S455001, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s455001.html.

About the authors

Clément de Chaisemartin is an assistant professor of economics at the University of California, Santa Barbara.

Xavier D'Haultfœuille is a professor of economics at CREST.

Yannick Guyonvarch is a PhD candidate in economics at CREST.