Fuzzy Differences-in-Differences with Stata*

Clément de Chaisemartin† Xavier D'Haultfœuille‡ Yannick Guyonvarch§

Preliminary

October 22, 2016

Abstract

This paper presents the Stata command `fuzzydid`, which computes the various estimators proposed by de Chaisemartin & D'Haultfœuille (2016*a*) for fuzzy differences-in-differences designs.

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

1 Introduction

Difference-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.

de Chaisemartin & D'Haultfœuille (2016*a*) show that 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. Then,

*The current command is still preliminary. Bug reports are greatly appreciated and should be sent to the authors via email.

†University of California at Santa Barbara, clementdechaisemartin@ucsb.edu

‡CREST, xavier.dhaultfoeuille@ensae.fr

§CREST, yannick.guyonvarch@ensae.fr

they 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 & Imbens (2006) to fuzzy designs. They also show that under the same assumptions as those underlying the Wald-TC and Wald-CIC estimands, the LATE of treatment group switchers can be bounded when the share of treated units changes over time in the control group. They further show that their results extend to applications with non-binary treatments and covariates. Finally, they consider estimators of all the estimands introduced in their paper, and they derive their limiting distributions.

In this paper we describe the `fuzzydid` Stata package. This package computes the estimators proposed in de Chaisemartin & D'Haultfœuille (2016*a*), as well as their standard errors.

The remainder of the paper is organized as follows. Section 2 briefly summarizes the main results in de Chaisemartin & D'Haultfœuille (2016*a*), and gives some background on their estimands and estimators. Section 3 presents the `fuzzydid` Stata package. The appendix gathers the analytical formulas of the asymptotic variances of some of the estimators.

2 Framework

2.1 Parameters of interest and main estimands

We seek to estimate the effect of a treatment D on some outcome. For now we assume D is binary. $Y(1)$ 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 two time periods represented by a random variable $T \in \{0, 1\}$, and into two groups represented by a random variable $G \in \{0, 1\}$. $G = 1$ (resp. $G = 0$) for units in the treatment (resp. control) group. Contrary to the standard "sharp" DID setting where $D = G \times T$, we consider a "fuzzy" setting where $D \neq G \times T$. In this setting, the treatment group is the one that experiences the higher increase of its treatment rate between period 0 and 1 (see Assumption 2 below).

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.

Assumption 1 *(Treatment participation equation)*
$D = 1\{V \geq v_{GT}\}$, *where* $v_{G0} = v_{00}$ *and* $V \perp\!\!\!\perp T|G$.

Assumption 2 *(First stage)*
$E(D_{11}) > E(D_{10})$, *and* $E(D_{11}) - E(D_{10}) > E(D_{01}) - E(D_{00})$.

We then introduce $D(t) = 1\{V \geq v_{Gt}\}$. In repeated cross sections, $D(0)$ and $D(1)$ correspond, under Assumption 1, to the treatment status of a unit at period 0 and 1 respectively. We consider the subpopulation $S = \{D(0) < D(1), G = 1\}$, called hereafter the treatment group switchers. Our parameters of interest are the Local Average Treatment Effect (LATE) and Local Quantile Treatment Effects (LQTE) of treatment group switchers, which are respectively defined by

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

To analyze one of the estimands we consider below, it is also useful to define $\alpha = (E(D_{11}) - E(D_{10}))/[E(D_{11}) - E(D_{10}) - (E(D_{01}) - E(D_{00}))]$, the control group switchers $S' = \{D(0) \neq D(1), G = 0\}$ and their local average treatment effect
$\Delta' = E\left(Y(1) - Y(0)|S', T = 1\right)$.

We now introduce the main estimands considered in de Chaisemartin & D'Haultfœuille (2016*a*). First, let

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

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.

Second, let

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

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 3 below), the $\delta_{D_{10}}$ term accounts for the effect of time on the outcome in the treatment group.

Third, let

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

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, let

$$F_{CIC,d}(y) = \frac{P(D_{11} = d)F_{Y_{d11}}(y) - P(D_{10} = d)F_{Q_d(Y_{d10})}(y)}{P(D_{11} = d) - P(D_{10} = d)} \tag{20}$$

and

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

These estimands identify Δ and τ_q under combinations of the following assumptions.

Assumption 3 *(Common trends)*

$E(Y(0)|G, T = 1) - E(Y(0)|G, T = 0)$ *does not depend on* G.

Assumption 3' *(Conditional common trends)*

For all $d \in \mathcal{S}(D)$, $E(Y(d)|G, T = 1, D(0) = d) - E(Y(d)|G, T = 0, D(0) = d)$ *does not depend on* G.

Assumption 4 *(Homogeneous treatment effect over time)*

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

Assumption 5 *(Homogeneous treatment effect between groups)*

$\Delta = \Delta'$.

Assumption 6 *(Stable percentage of treated units in the control group)*

$0 < E(D_{01}) = E(D_{00}) < 1$.

Assumption 7 *(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,* $U_d \perp\!\!\!\perp T|G, V$.

Assumption 8 *(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* $(d, g, t) \in \mathcal{S}((D, G, T))$.

Theorem 1 *(de Chaisemartin & D'Haultfœuille, 2016*a*) Suppose that Assumptions 1-2 hold.*

1. *If Assumptions 3 and 4 also hold, then* $W_{DID} = \alpha\Delta + (1 - \alpha)\Delta'$. *If Assumption 5 or 6 further holds, then* $W_{DID} = \Delta$.

2. *If Assumptions 3' and 6 also hold, then* $W_{TC} = \Delta$.

3. *If Assumptions 6-8 also hold, then* $W_{CIC} = \Delta$ *and* $\tau_{q,CIC} = \tau_q$.

2.2 Extensions and further estimands

Special cases

When $P(D_{00}=0)=P(D_{01}=0)\in\{0,1\}$, W_{CIC} is not defined because either Q_0 or Q_1 is not defined. In such cases, we can simply suppose that $Q_0=Q_1$. When $P(D_{00}=0)=P(D_{01}=0)=1$, this leads to the following variant of W_{CIC}: $\widetilde{W}_{CIC}=\frac{E(Y_{11})-E(Q_0(Y_{10}))}{E(D_{11})-E(D_{10})}$.[1] Similarly, we define $\widetilde{F}_{CIC,d}(y)=[P(D_{11}=d)F_{Y_{d11}}(y)-P(D_{10}=d)F_{Q_0(Y_{d10})}(y)]/[P(D_{11}=d)-P(D_{10}=d)]$ and $\widetilde{\tau}_{CIC,q}=\widetilde{F}^{-1}_{CIC,1}(q)-\widetilde{F}^{-1}_{CIC,0}(q)$. de Chaisemartin & D'Haultfœuille (2016*a*) 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 $y\in\mathcal{S}(Y)$.

Partial identification

When Assumption 6 does not hold, one can still obtain bounds on Δ using the "TC" approach. 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

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

de Chaisemartin & D'Haultfœuille (2016*a*) show that under Assumptions 1, 2, and 3', $\Delta\in[\underline{W}_{TC},\overline{W}_{TC}]$.

Non-binary treatment

de Chaisemartin & D'Haultfœuille (2016*a*) also show that W_{DID}, W_{TC}, and W_{CIC} still identify a causal parameter if D is not binary but is ordered and takes a finite number of values. Replace the selection equation of Assumption 1 by

$$D=\sum_{d=1}^{\overline{d}}1\{V\geq v^d_{GT}\},$$

with $-\infty=v^0_{GT}<v^1_{GT}...<v^{\overline{d}+1}_{GT}=+\infty$ and $V\perp\!\!\!\perp T|G$. As before, let $D(t)=\sum_{d=1}^{\overline{d}}1\{V\geq v^d_{Gt}\}$. Then, de Chaisemartin & D'Haultfœuille (2016*a*) show that under Assumptions 3, 3', 4, 6, 7,

[1] When $P(D_{00}=0)=P(D_{01}=0)\in\{0,1\}$, W_{TC} is also not defined. However, the corresponding variant of W_{TC} is actually equal to W_{DID}.

and 8, W_{DID}, W_{TC}, and W_{CIC} point identify 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 $w_d = [P(D_{11} \geq d) - P(D_{10} \geq d)]/[E(D_{11}) - E(D_{10})]$. A similar result still holds with a continuous D if $D = h_{GT}(V)$ with $h_{gt}(.)$ strictly increasing. Our estimands then identify the following weighted average derivative:

$$WAD = \int_0^{\overline{d}} E\left(\frac{\partial Y(\delta)}{\partial \delta}\middle| D(0) < \delta \leq D(1)\right) w_\delta d\delta.$$

Finally, when Assumption 6 fails to hold and treatment is ordered and takes a finite number of values, it is still possible to bound ACR. We refer the reader to Subsection 3.1 in de Chaisemartin & D'Haultfœuille (2016*b*) for further details.

Including covariates

If the assumptions in Subsection 2.1 hold conditional on X rather than unconditionally, de Chaisemartin & D'Haultfœuille (2016*a*) show that one can still identify Δ and τ_q using conditional versions of the estimands. Specifically, 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, let

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

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

One can show that W^X_{DID} and W^X_{TC} identify Δ under appropriate generalizations of Assumptions 1-6.[2]

2.3 Estimators

Estimators in the basic set-up

We first consider plug-in estimators of the estimands introduced in Subsection 2.1.[3] Let $\mathcal{I}_{gt} = \{i : G_i = g, T_i = t\}$ (resp. $\mathcal{I}_{dgt} = \{i : D_i = d, G_i = g, T_i = t\}$) and n_{gt} (resp. n_{dgt}) denote the size of $\mathcal{I}_{gt}$ (resp. $\mathcal{I}_{dgt}$) for all $(d, g, t) \in \{0, 1\}^3$.

[2] de Chaisemartin & D'Haultfœuille (2016*a*) also consider a generalization of W_{CIC} identifying Δ when Assumptions 7-8 hold conditional on X rather than unconditionally. This estimand is not presented here since its estimator is not computed by the `fuzzydid` command yet.

[3] Though we present them in the case where treatment is binary, these estimators can be defined similarly when D_i is ordered and takes a finite number of values.

First, let

$$\widehat{\delta}_d = \frac{1}{n_{d01}} \sum_{i \in \mathcal{I}_{d01}} Y_i - \frac{1}{n_{d00}} \sum_{i \in \mathcal{I}_{d00}} Y_i$$

for $d \in \{0, 1\}$, and let

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

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

denote our estimators of W_{DID} and W_{TC}.

Second, let $\widehat{F}_{Y_{dgt}}(y) = \frac{1}{n_{dgt}} \sum_{i \in \mathcal{I}_{dgt}} \mathbb{1}\{Y_i \leq y\}$ denote the empirical cdf of Y_{dgt}, let $\widehat{F}^{-1}_{Y_{dgt}}(q) = \inf\{y : \widehat{F}_{Y_{dgt}}(y) \geq q\}$ denote its empirical quantile of order $q \in (0, 1)$, let $\widehat{Q}_d = \widehat{F}^{-1}_{Y_{d01}} \circ \widehat{F}_{Y_{d00}}$ denote the estimator of the quantile-quantile transform, and let

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

denote our estimator of W_{CIC}.

Finally, let $\widehat{P}(D_{gt} = d)$ denote the proportion of units with $D = d$ in the sample $\mathcal{I}_{gt}$, let $\widehat{H}_d = \widehat{F}_{Y_{d10}} \circ \widehat{F}^{-1}_{Y_{d00}}$, let

$$\widehat{F}_{Y(d)|S,T=1} = \frac{\widehat{P}(D_{10} = d)\widehat{H}_d \circ \widehat{F}_{Y_{d01}} - \widehat{P}(D_{11} = d)\widehat{F}_{Y_{d11}}}{\widehat{P}(D_{10} = d) - \widehat{P}(D_{11} = d)},$$

and let

$$\widehat{\tau}_q = \widehat{F}^{-1}_{Y(1)|S,T=1}(q) - \widehat{F}^{-1}_{Y(0)|S,T=1}(q)$$

denote our estimator of τ_q.

de Chaisemartin & D'Haultfœuille (2016*a*) 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. Their asymptotic variances and the corresponding estimators are displayed in Appendix A. de Chaisemartin & D'Haultfœuille (2016*a*) also establish the validity of the bootstrap to draw inference on Δ and τ_q based on these estimators.

Estimators of bounds under partial identification

We now consider estimators of the bounds $\underline{W}_{TC}$ and $\overline{W}_{TC}$. Let $\widehat{\lambda}_{0d} = \frac{\widehat{P}(D_{01}=d)}{\widehat{P}(D_{00}=d)}$, $\widehat{\lambda}_{1d} = \frac{\widehat{P}(D_{11}=d)}{\widehat{P}(D_{10}=d)}$, and

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

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

Then let

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

Finally, we estimate the bounds by

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

Under regularity conditions, these estimators are root-n consistent and asymptotically normal. Moreover, the bootstrap is valid for inference (see de Chaisemartin & D'Haultfœuille, 2016*b*, Subsection 2.2).

Estimators with covariates

First, we consider non-parametric estimators of W_{DID}^X and W_{TC}^X. 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 $P_{gt}^K = (p^K(X_1), ..., p^K(X_n))$. For any random variable R, we estimate $m^R(x) = E(R|X = x)$ by the series estimator

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

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

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

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

where $\widehat{\delta}_d(x) = \widehat{m}_{d01}^Y(x) - \widehat{m}_{d00}^Y(x)$. Under regularity conditions, these estimators are root-n consistent and asymptotically normal (see de Chaisemartin & D'Haultfœuille, 2016*b*, Subsection 2.3).

The estimators above may not be computable if support conditions on the covariates are not met. For instance, if a discrete covariate can take many values, there may be values for which there are only treatment group units with this specific value of the covariate, thus implying that $m_{0t}^Y(.)$ and $m_{0t}^D(.)$ cannot be estimated for this specific value of the covariate. We explain how the command deals with such cases in Appendix B.

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

$$W_{DID}^X = \frac{E(Y_{11}) - E\left(X_{11}'\beta_{10}^Y\right) - \left(E\left(X_{11}'\beta_{01}^Y\right) - E\left(X_{11}'\beta_{00}^Y\right)\right)}{E(D_{11}) - E\left(X_{11}'\beta_{10}^D\right) - \left(E\left(X_{11}'\beta_{01}^D\right) - E\left(X_{11}'\beta_{00}^D\right)\right)}$$

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

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

$$\widehat{W}_{DID,OLS}^X = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[Y_i - X_i'\widehat{\beta}_{10}^Y - X_i'\widehat{\beta}_{01}^Y + X_i'\widehat{\beta}_{00}^Y\right]}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left[D_i - X_i'\widehat{\beta}_{10}^D - X_i'\widehat{\beta}_{01}^D + X_i'\widehat{\beta}_{00}^D\right]},$$

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

where for $(d,g,t) \in \{0,1\}^3$, $\widehat{\beta}_{gt}^Y$ (resp. $\widehat{\beta}_{dgt}^Y$) denotes the coefficient of X in an OLS regression of Y on X in the subsample $\mathcal{I}_{gt}$ (resp. $\mathcal{I}_{dgt}$), and $\widehat{\beta}_{gt}^D$ 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. Their computation is also automated in the `fuzzydid` command.

3 The `fuzzydid` command

The program uses the distinct Stata command. If not already installed on the user's computer, she must type `ssc install distinct`, in Stata's command line. The program has been thoroughly debugged when used with Stata 13. Small bugs might still appear with other versions of Stata.

3.1 Syntax

The syntax of `fuzzydid` is as follows:

```
fuzzydid Y G T D [if] [in][, did tc cic lqte new_categ(numlist) numerator partial
         nose cluster(varname) breps(#) bwidth_method(method) moreoutput
         nonbinary(varlist) binary(varlist) x_param(reg1 reg2)
         x_no_param(method) sieve_order(#)]
```

3.2 Description

`fuzzydid` estimates Δ or τ_q using one or several of the estimators defined in Subsection 2.3 above. It also computes their standard errors and confidence intervals.

`Y` is the outcome variable.

`G` is a dummy equal to 0 for control group units, and to 1 for treatment group units. We discuss below how one can use the command when the data contains more than two groups.

`T` is a dummy equal to 0 for the first period and to 1 for the second period. We discuss below how one can use the command when the data contains more than two periods.

`D` is the treatment variable. `D` can be binary but also ordered, with finite or infinite support.

3.3 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 special cases where D is binary and $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command actually computes W_{DID} (see Subsection 2.2 above). 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 special cases where D is binary and $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command estimates $\widetilde{W}_{DID}$ defined in Subsection 2.2 above. The `cic` option cannot be used when 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 is binary. When $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$, the command estimates $\widetilde{\tau}_{q,CIC}$ defined in Subsection 2.2 above. It cannot be used when covariates are included in the estimation.

N.B.: 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 in the numerators of W_{TC} and W_{CIC}. When the treatment can take a large number of values, using this option might be necessary to ensure that the support of D is the same in the treatment and in the control group. It also avoids estimating δ_d and Q_d on a small number of units, thus often lowering the variances of $\widehat{W}_{TC}$ and $\widehat{W}_{CIC}$. 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)`. For the last upper bound, any value greater than the highest value in the estimation sample can be used. This might prove convenient if the user wants to estimate `fuzzydid` over various subsamples. In each estimation, she can merely specify a value greater or equal to the highest value of D in the entire sample.

`numerator` computes only the numerators of the estimators. As explained in Subsection 2.5 in de Chaisemartin & D'Haultfœuille (2016*a*), this option is useful to conduct placebos and test the assumptions underlying each estimator.

`partial` estimates $\underline{W}_{TC}$ and $\overline{W}_{TC}$ when D is binary, and the bounds defined in Subsection 3.1 in de Chaisemartin & D'Haultfœuille (2016*b*) when D is ordered and takes a finite number of values.

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

`cluster(`*varname*`)` computes the standard errors of the estimators assuming that observations are clustered at the ***varname*** level instead of assuming that they are independent. Only one clustering variable is allowed.

`breps(`#`)` specifies the number of bootstrap replications to be run. It must be greater than 2.

`bwidth_method(`*method*`)` selects the bandwidth to be used when computing the standard errors of $\widehat{\tau}_q$ analytically. Nonparametric estimators of densities are indeed used in the estimation of these standard errors. ***method*** can be one of the following bandwidths: `bofinger` (the default), `hall_sheather`, or `chamberlain`.

`moreoutput` tests the stability of the distribution of D over time in the control group. When treatment is not binary but ordered and takes a finite number of values, it also estimates the weights w_d in the ACR estimand. If the user has specified more than one of the `did`, `tc`, and `cic` options, it also performs equality tests between the estimators. Finally, when the user has specified the `lqte` option, it plots graphs of the estimated CDFs of $Y(0)$ and $Y(1)$ among switchers.

Options specific to estimation with covariates

`nonbinary(`*varlist*`)` specifies the names of all the non-binary covariates that need to be used in the estimation.

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

`x_param(`*reg1 reg2*`)` specifies which parametric method should be used to estimate the conditional expectations in W_{DID}^X and W_{TC}^X. The available methods are: `ols`, `logit`, and `probit`.

`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 $E(D_{gt}|X)$. For instance, if the user writes `x_param(ols logit)`, the command estimates $E(Y_{gt}|X)$ and $E(Y_{dgt}|X)$ by OLS, and $E(D_{gt}|X)$ by a logistic regession. The `logit` and `probit` options can only be used with binary variables.

`x_no_param(`*method*`)` specifies which non-parametric method should be used to estimate the conditional expectations in W^X_{DID} and W^X_{TC}. The package currently only allows for series estimation, so *method* should be equal to `sieve`. Series estimation consists in approximating the conditional expectations with a linear combination of Legendre polynomials.

N.B.: when covariates are included in the estimation, and neither `x_param` nor `x_no_param` is specified, the package uses by default the `x_param(`*ols ols*`)` option.

`sieve_order(`#`)` specifies the order of the sieve basis, when the option `x_no_param(`*sieve*`)` is chosen. It must be greater than or equal to 2. For a given order L, the number of basis functions is given by $(L+1)^{p_c}$ where p_c is the number of continuous covariates. The package does not allow for more than 450 basis functions. For example, $p_c = 5$ and a sieve order of 4 leads to 625 basis functions, which is too large. In such cases, the order is set to 2. More generally, if order selection fails for some reason, the order is set to 2. If this option is not specified, the choice of the order is done via 5-fold cross-validation with a mean squared error loss function.

3.4 Computation of the standard errors

The standard errors are computed either with analytic formulas or the bootstrap.

When no covariates are included in the estimation, the asymptotic variances of some but not all estimators have been computed. When available, the analytic variances are used by default, so one needs to add the option `breps()` to get bootstrapped standard errors. When analytic variances are not available, the bootstrap is used. The default number of bootstrap replications is 50, but the user can change this number using the `breps()` option. The analytic variances of $\widehat{W}_{DID}$ and $\widehat{W}_{TC}$ have been computed, both for binary and non-binary D, and with or without clustering. The analytic variances of $\widehat{W}_{CIC}$ and $\widehat{\tau}_q$ have also been computed when D is binary and without clustering. However, note that these analytic formulas are written in Mata and use a substantial amount of memory. If an issue with memory allocation occurs, the program switches to the bootstrap. When the `moreoutput` option is specified and an equality test involving $\widehat{W}_{CIC}$ needs to be performed and the analytic formula of its asymptotic variance is not available, all the estimators are bootstrapped. Finally, when the option `numerator` or `partial` is used, bootstrapped standard errors are used.

When covariates are included in the estimation, only bootstraped standard errors are available.

3.5 Using fuzzydid with multiple groups and time periods

When the data contains multiple groups and two periods, one needs to create a variable G_1 equal to 1 for groups whose mean treatment increased between $T = 0$ and $T = 1$, to 0 for groups whose mean treatment remained stable, and missing for groups whose mean treatment decreased. One can then use `fuzzydid` with G_1 as the group variable to estimate, say, the $W^*_{DID}(1, 0)$ estimand defined in de Chaisemartin & D'Haultfœuille (2016*a*). Then, one needs to create a variable G_2 equal to 1 for groups whose mean treatment decreased between $T = 0$ and $T = 1$, to 0 for groups whose mean treatment remained stable, and missing for groups whose mean treatment increased. One can then use `fuzzydid` with G_2 as the group variable to estimate, say, the $W^*_{DID}(-1, 0)$ estimand also defined in de Chaisemartin & D'Haultfœuille (2016*a*). Finally, one can estimate Δ by computing the weighted average of these two estimators proposed in Theorem 3.1 in de Chaisemartin & D'Haultfœuille (2016*a*). The standard error of this weighted average can be obtained by bootstrapping the whole procedure.

When the data contains multiple periods and multiple groups, one can conduct the estimation procedure described in the preceding paragraph for each pair of consecutive periods. This will yield, say, a Wald-DID estimator for each pair of consecutive periods. Then, one can compute the weighted average of these estimators proposed in Theorem S1 in de Chaisemartin & D'Haultfœuille (2016*b*). The standard error of this weighted average can be obtained by bootstrapping the whole procedure.

References

Athey, S. & Imbens, G. W. (2006), 'Identification and inference in nonlinear difference-in-differences models', *Econometrica* **74**(2), 431–497.

Bofinger, E. (1975), 'Estimation of a density function using order statistics', *Australian Journal of Statistics* **17**(1), 1–7.

Chamberlain, G. (1994), Quantile regression, censoring, and the structure of wages, *in* 'Advances in Econometrics: Sixth World Congress', Vol. 2, pp. 171–209.

de Chaisemartin, C. & D'Haultfœuille, X. (2016*a*), Fuzzy differences-in-differences, Technical report.

de Chaisemartin, C. & D'Haultfœuille, X. (2016*b*), Supplement to "fuzzy differences-in-differences", Technical report.

Frölich, M. (2007), 'Nonparametric iv estimation of local average treatment effects with covariates', *Journal of Econometrics* **139**(1), 35–75.

Hall, P. & Sheather, S. J. (1988), 'On the distribution of a studentized quantile', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 381–391.

A Asymptotic variances

de Chaisemartin & D'Haultfœuille (2016*a*) show that under regularity conditions,

$$\begin{aligned}
\sqrt{n}\left(\widehat{W}_{DID} - \Delta\right) &\xrightarrow{d} \mathcal{N}\left(0, V\left(\psi_{DID}\right)\right), \\
\sqrt{n}\left(\widehat{W}_{TC} - \Delta\right) &\xrightarrow{d} \mathcal{N}\left(0, V\left(\psi_{TC}\right)\right), \\
\sqrt{n}\left(\widehat{W}_{CIC} - \Delta\right) &\xrightarrow{d} \mathcal{N}\left(0, V\left(\psi_{CIC}\right)\right), \\
\sqrt{n}\left(\widehat{\tau}_q - \tau_q\right) &\xrightarrow{d} \mathcal{N}\left(0, V\left(\psi_{q,CIC}\right)\right),
\end{aligned}$$

where ψ_{DID}, ψ_{TC}, ψ_{CIC} and $\psi_{q,CIC}$ are random variables defined below.

$$\begin{aligned}
\psi_{DID} = \frac{1}{DID_D}\Bigg[&\frac{GT(\varepsilon - E(\varepsilon_{11}))}{p_{11}} - \frac{G(1-T)(\varepsilon - E(\varepsilon_{10}))}{p_{10}} - \frac{(1-G)T(\varepsilon - E(\varepsilon_{01}))}{p_{01}} \\
&+ \frac{(1-G)(1-T)(\varepsilon - E(\varepsilon_{00}))}{p_{00}}\Bigg]
\end{aligned}$$

where $p_{gt} = P(G = g, T = t)$ and $\varepsilon = Y - \Delta D$.

$$\psi_{TC} = \frac{1}{E(D_{11}) - E(D_{10})} \left\{ \frac{GT(\varepsilon - E(\varepsilon_{11}))}{p_{11}} - \frac{G(1-T)(\varepsilon + (\delta_1 - \delta_0)D - E(\varepsilon_{10} + (\delta_1 - \delta_0)D_{10}))}{p_{10}} \right.$$
$$- E(D_{10})D(1-G)\left[\frac{T(Y - E(Y_{101}))}{p_{101}} - \frac{(1-T)(Y - E(Y_{100}))}{p_{100}}\right]$$
$$\left. - (1 - E(D_{10}))(1-D)(1-G)\left[\frac{T(Y - E(Y_{001}))}{p_{001}} - \frac{(1-T)(Y - E(Y_{000}))}{p_{000}}\right] \right\}.$$

where $p_{dgt} = P(D = d, G = g, T = t)$.

Finally,

$$\psi_{CIC} = \int \Psi_0(y) - \Psi_1(y)dy,$$
$$\psi_{q,CIC} = \left[\frac{\Psi_1}{f_{Y(1)|S,T=1}}\right] \circ F^{-1}_{Y(1)|S,T=1}(q) - \left[\frac{\Psi_0}{f_{Y(0)|S,T=1}}\right] \circ F^{-1}_{Y(0)|S,T=1}(q),$$

where

$$\Psi_d(y) = \frac{1}{p_{d|11} - p_{d|10}} \left\{ \frac{GT}{p_{11}} \left[\mathbb{1}\{D = d\}\mathbb{1}\{Y \leq y\} - p_{d|11}F_{d11}(y) - F_{Y(d)|S,T=1}(y)\left(\mathbb{1}\{D = d\} - p_{d|11}\right)\right] \right.$$
$$+ \frac{G(1-T)}{p_{10}}\left[-\mathbb{1}\{D = d\}\left(\mathbb{1}\{Q_d(Y) \leq y\} - H_d \circ F_{d01}(y)\right) + \left(\mathbb{1}\{D = d\} - p_{d|10}\right)\left(F_{Y(d)|S,T=1}(y) - H_d \circ F_{d01}(y)\right)\right.$$
$$\left. + p_{d|10}(1-G)\mathbb{1}\{D = d\}H'_d \circ F_{d01}(y)\left[\frac{(1-T)(\mathbb{1}\{Q_d(Y) \leq y\} - F_{d01}(y))}{p_{d00}} - \frac{T(\mathbb{1}\{Y \leq y\} - F_{d01}(y))}{p_{d01}}\right]\right\},$$

$H_d = F_{Y_{d10}} \circ F^{-1}_{Y_{d00}}$ and $p_{d|gt} = P(D = d|G = g, T = t)$.

We estimate the asymptotic variances of the estimators using the sample variances of $\widehat{\psi}_{DID}$, $\widehat{\psi}_{TC}$, $\widehat{\psi}_{CIC}$, and $\widehat{\psi}_{q,CIC}$. $\widehat{\psi}_{DID}$ and $\widehat{\psi}_{TC}$ are simple, plug-in estimators of ψ_{DID} and ψ_{TC}, but the estimators of ψ_{CIC} and $\psi_{q,CIC}$ involve nonparametric estimators. Specifically, we estimate $H'_d = f_{Y_{d10}} \circ F^{-1}_{Y_{d00}} / f_{Y_{d00}} \circ F^{-1}_{Y_{d00}}$ by

$$\widehat{H'_d}(y) = \frac{\widehat{f}_{Y_{d10}} \circ \widehat{F}^{-1}_{Y_{d00}}(y)}{\widehat{f}_{Y_{d00}} \circ \widehat{F}^{-1}_{Y_{d00}}(y)},$$

where $\widehat{f}_{Y_{dgt}}$ denotes the kernel density estimator of Y on the subsample $\mathcal{I}_{dgt}$. An estimator of $f_{Y(d)|S,T=1}$ is also required to estimate $\psi_{q,CIC}$. As $F_{Y(d)|S,T=1} = F_{CIC,d}$,

$$f_{Y(d)|S,T=1}(y) = \frac{P(D_{11} = d)f_{Y_{d11}}(y) - P(D_{10} = d)f_{Q_d(Y_{d10})}(y)}{P(D_{11} = d) - P(D_{10} = d)}.$$

To form an estimator of $f_{Y(d)|S,T=1}(y)$, we estimate $f_{Y_{d11}}$ and $f_{Q_d(Y_{d10})}$ as above. The bandwidth used in the estimation is computed using either the Bofinger, Hall and Sheather, or Chamberlain method, as in the `qreg` command. We refer the reader to Bofinger (1975), Hall & Sheather (1988), and Chamberlain (1994) for more details.

B Support issues with control variables.

Researchers may sometimes wish to include a potentially large set of mutually exclusive binary variables as controls in their estimation. This might however lead to violations of the common support assumption $\mathcal{S}(X_{dgt}) = \mathcal{S}(X)$ underlying the estimands with covariates in Subsection 1.3 of de Chaisemartin & D'Haultfœuille (2016*b*).

To describe the problem and the way the command handles it, let us consider an example where the units of observation are, say, US counties, and where the researcher wishes to allow for, say, state-specific trends when estimating the Wald-DID. To do so, in principle the researcher merely needs to include state dummies as controls using the `binary` option. However, there might be a state where all counties belong to the treatment group, thus implying that within this state there are no control counties to which the trends experienced by treatment group counties can be compared. Let X be a random variable denoting the state a county belongs to, and let x_0 denote that problematic state. To avoid dropping all observations from state x_0, when this happens the command uses all control counties to predict the counterfactual trends of treatment group counties in state x_0. Specifically, when estimating $W_{DID}(x_0)$ (see Subsection 1.3 in de Chaisemartin & D'Haultfœuille, 2016*b*), the command estimates $E(Y_{01}) - E(Y_{00})$ and $E(D_{01}) - E(D_{00})$ instead of $E(Y_{01}|X = x_0) - E(Y_{00}|X = x_0)$ and $E(D_{01}|X = x_0) - E(D_{00}|X = x_0)$, as these latter quantities are not defined. This implies that $W_{DID}(x_0)$ relies on common trends between the treatment group counties in state x_0 and all the control counties. For the other states where both treatment and control group counties are present, the command estimates $W_{DID}(x)$ following the definition in Subsection 1.3 in de Chaisemartin & D'Haultfœuille (2016*b*). If the researcher instead wants to estimate the Wald-TC, the same principle applies. If there is a state × treatment status cell (x_0, d_0) where there are only treatment group counties at $T = 0$, the command estimates $E(Y_{d_001}) - E(Y_{d_000})$ instead of $E(Y_{d_001}|X = x_0) - E(Y_{d_000}|X = x_0)$ when estimating $W_{TC}(x_0)$ (see Subsection 1.3 in de Chaisemartin & D'Haultfœuille, 2016*b*). Putting it differently, the command uses all control counties with treatment d_0 to predict the counterfactual trends of treatment group counties in state x_0 and with treatment d_0.