Fuzzy Difference-in-Differences with Grouped Data

EEA-ESEM Congress - Bocconi University, Milan

Clément de Chaisemartin — Sciences Po

Xavier D'Haultfœuille — CREST

Félix Pasquier — CREST

August 24^{nd}, 2022

Difference-in-Differences (DiD)

- One of the most popular research designs to estimate causal effects of a binary treatment
- In practice, implemented using a **Two-Way Fixed Effect (TWFE) regression**

$$Y_{g,t} = \alpha_g + \gamma_t + \beta_{fe} D_{g,t} + \varepsilon_{g,t}$$

TWFE and Heterogeneous Treatment Effects

- TWFE regressions can provide very misleading estimates of treatment effects

  e.g. de Chaisemartin & D'Haultœuille (2020), Borusyak et al. (2021), Sun & Abraham (2021), Goodman-Bacon (2021), Callaway & Sant'Anna (2020), Imai & Kim (2018)

- Alternative estimators only apply to **sharp designs**, i.e. designs in which all units belonging to the same (g, t) cell have the same treatment status

Example of a Non-Sharp Design : Adena et al. (2015)

- Investigate the impact of biased political radio programs in Germany over the period 1928-1933 on votes for the Nazi Party
- TWFE regression at the electoral district $\times$ election date level
  - **Outcome** $Y_{g,t}$: vote share for the Nazi party
  - **Treatment** $D_{g,t}$: share of households having access to the radio
- Heterogeneous increase of radio subscriptions over districts
- "Treated" and "untreated" voters in all (g, t) cells

Example of a Non-Sharp Design : Adena et al. (2015)

- Investigate the impact of biased political radio programs in Germany over the period 1928-1933 on votes for the Nazi Party
- TWFE regression at the electoral district × election date level
    - **Outcome** $Y_{g,t}$: vote share for the Nazi party
    - **Treatment** $D_{g,t}$: share of households having access to the radio
- Heterogeneous increase of radio subscriptions over districts
- "Treated" and "untreated" voters in all (g, t) cells

de Chaisemartin & D'Haultfœuille (2020)'s review of all AER papers between 2010-2012 : **27% of TWFE specifications are fuzzy DiD designs**

This paper

- Develop an alternative estimation strategy to TWFE
  - robust to heterogeneous treatment effects
  - applicable to *fuzzy* designs
  - allowing for some endogenous selection into treatment
  - only requires aggregated data at the group-level
- Rationalize Chamberlain (1992)'s Correlated Random Coefficient (CRC) model
- Exploit results from Chamberlain (1992), Arellano & Bonhomme (2012) and Graham & Powell (2012) to identify treatment effects
- Revisit Adena et al. (2015)

Related Paper

- **de Chaisemartin & D'Haultfœuille (2018)**
  - specifically considers *fuzzy* designs
  - provide robust estimators to heterogeneous treatment effects
- Main differences with our estimators
  - cannot be computed with only average variables at the group level
  - require "stable" groups for each time period

Overview

1. Introduction
2. Model
3. Identification
4. Application (in Progress) : Impact of Biased Political Radio Programs on the Rise of Nazism

Overview

Set-up

- Panel of G groups over T periods
- Number of units in the "cell" (g, t) : $N_{g,t}$
- For any variable $A_{i,g,t}$ defined at the individual level, let

$$A_{g,t} = \sum_{i=1}^{N_{g,t}} A_{i,g,t}/N_{g,t}$$
$$\mathbf{A}_g^a = (A_{g,1}, \ldots, A_{g,T})'$$

- Binary (unit-level) treatment : $D_{i,g,t} \in \{0,1\}$
- Potential outcomes : $Y_{i,g,t}(0)$, $Y_{i,g,t}(1)$
- Treatment effect : $\Delta_{i,g,t} := Y_{i,g,t}(1) - Y_{i,g,t}(0)$
- Realised outcome : $Y_{i,g,t} := Y_{i,g,t}(D_{i,g,t})$

Model

For all $(i, g, t) \in \{1, \dots, N_{g,t}\} \times \{1, \dots, G\} \times \{1, \dots, T\}$,

$$\begin{cases} Y_{i,g,t}(0) & = \alpha_g + \beta_t + \varepsilon_{i,g,t} \\ \Delta_{i,g,t} & = \gamma_{i,g(t)} + \delta_t + \eta_{i,g,t} \end{cases} \tag{1}$$

Without loss of generality, we suppose that $\beta_1 = \delta_1 = 0$

Model

For all $(i, g, t) \in \{1, \ldots, N_{g,t}\} \times \{1, \ldots, G\} \times \{1, \ldots, T\}$,

$$\begin{cases} Y_{i,g,t}(0) & = \alpha_g + \beta_t + \varepsilon_{i,g,t} \\ \Delta_{i,g,t} & = \gamma_{i,g(t)} + \delta_t + \eta_{i,g,t} \end{cases} \tag{1}$$

Without loss of generality, we suppose that $\beta_1 = \delta_1 = 0$

Assumption 1 (Balanced Panel of Groups)

For all $(g, t) \in \{1, \ldots, G\} \times \{1, \ldots, T\}$, $N_{g,t} > 0$

Assumptions

Let $U_{i,g,t} = (D_{i,g,t}, \xi_{i,g,t}, \zeta_{i,g,t})$ and $\mathbf{U}_g = (U_{i,g,t})_{1\leq t\leq T, 1\leq i\leq N_{g,t}}$

Assumption 2 (Independent Groups)

The G random vectors $(\mathbf{U}_g)_{g=1,\dots,G}$ are independent

Assumption 3 (Strong Exogeneity)

For all $(i,g,t) \in \{1,\dots,N_{g,t}\} \times \{1,\dots,G\} \times \{1,\dots,T\}$,

1. $\mathbb{E}[\varepsilon_{g,t}|\mathbf{D}_g^a] = 0$
2. $\mathbb{E}[\eta_{i,g,t}|\mathbf{D}_g^a, D_{i,g,t}] = 0$

Assumptions

Assumption 4 (Treatments Independent of Cell's Treatment Rates at Other Periods)

For all $(g,t) \in \{1,\dots,G\} \times \{1,\dots,T\}$,

$$(D_{i,g,t})_{1 \leq i \leq N_{g,t}} \perp\!\!\!\perp (D_{g,1},\dots,D_{g,t-1},D_{g,t+1},\dots,D_{g,T}) | \; D_{g,t}$$

Assumption 5 (Irrelevance of Identities, Conditional on Cells' Treatment Effects)

For all g*, there is a function* $\Gamma_g : d \mapsto \Gamma_g(d)$ *that does not depend on* t *and such that for all* t*,*

$$E\left[\frac{1}{D_{g,t}N_{g,t}} \sum_{i:D_{i,g,t}=1} \gamma_{i,g(t)} \middle| D_{g,t}\right] = \Gamma_g(D_{g,t})$$

Common Trends Roy Model

Assumptions

Assumption 6 (Parametrization of $\Gamma_g(.)$)

There are known $K \leq T - 2$ *and functions* $r_0, \ldots, r_k$ *such that for all* g *and all* $d \in (0,1]$,

$$\Gamma_g(d) = \sum_{k=0}^{K} \gamma_{k,g} r_k(d)$$

for some real numbers $(\gamma_{0,k}, \ldots, \gamma_{K,g})$.

Overview

Rationalizing Chamberlain (1992) CRC Model

Lemma 1

Suppose Model (1) and Assumptions 1-6 hold. Then, $\forall g \in \{1,\ldots,G\}$, $\forall t \times \{1,\ldots,T\}$,

$$\mathbb{E}[Y_{g,t}|\mathbf{D}_g^a] = \beta_t + D_{g,t}\delta_t + \alpha_g + \sum_{k=0}^{K} \gamma_{k,g} D_{g,t} r_k(D_{g,t})$$

Steps

$$\mathbb{E}\left[\begin{pmatrix} Y_{g,1} \\ \vdots \\ Y_{g,T} \end{pmatrix} \middle| \mathbf{D}_g^a\right] = \underbrace{\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & D_{g,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & D_{g,T} \end{pmatrix}}_{W_g} \underbrace{\begin{pmatrix} \beta_2 \\ \delta_2 \\ \vdots \\ \beta_T \\ \delta_T \end{pmatrix}}_{\lambda_0} + \underbrace{\begin{pmatrix} 1 & D_{g,1}r_0(D_{g,1}) & \ldots & D_{g,1}r_K(D_{g,1}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & D_{g,T}r_0(D_{g,T}) & \ldots & D_{g,T}r_K(D_{g,T}) \end{pmatrix}}_{X_g} \underbrace{\begin{pmatrix} \alpha_g \\ \gamma_{0,g} \\ \vdots \\ \gamma_{K,g} \end{pmatrix}}_{\mu_{0,g}}$$

$$\Longrightarrow \mathbb{E}\left[Y_g|\mathbf{D}_g^a\right] = W_g\lambda_0 + X_g\mu_{0,g} \qquad (2)$$

(2) is a particular case of Chamberlain (1992)'s CRC Model

Identification

Let $\Pi(X_g) = I_T - X_g X_g^+$ be the orthogonal projector on the kernel of X_g

Assumption 7 (Design Restriction)

$E\left[\frac{1}{G}\sum_{g=1}^{G} W_g'\Pi(X_g)W_g\right]$ *is non-singular*

Conditions

Theorem 1

Suppose Model (1) and Assumptions 1-7 hold. Then,

$$\lambda_0 = \mathbb{E}\left[\frac{1}{G}\sum_{g=1}^{G} W_g'\Pi(X_g)W_g\right]^{-1}\mathbb{E}\left[\frac{1}{G}\sum_{g=1}^{G} W_g'\Pi(X_g)Y_g\right], \quad (3)$$

Moreover, for all g such that $P(\det(X_g'X_g) \neq 0) > 0$,

$$\mu_{0,g} = \mathbb{E}\left[(X_g'X_g)^{-1}X_g'(Y_g - W_g\lambda_0)\middle|\det(X_g'X_g) \neq 0\right], \quad (4)$$

Identification

Let $\mathcal{M} = \{g : \det(X_g' X_g) \neq 0\}$, $N_t^m = \sum_{g \in \mathcal{M}} D_{g,t} N_{g,t}$ and

$$\Delta_t^m = E\left[\frac{1}{N_t^m} \sum_{g \in \mathcal{M}} \sum_{i: D_{i,g,t}=1} \Delta_{i,g,t}\right]$$

Corollary 1

Suppose that Model (1) and Assumptions 1-7 hold. Then, Δ_t^m is identified and

$$\Delta_t^m = \delta_t + E\left[\sum_{g \in \mathcal{M}} \frac{N_{g,t} D_{g,t}}{N_t^m} \sum_{k=0}^{K} \gamma_{k,g} r_k(D_{g,t})\right]$$

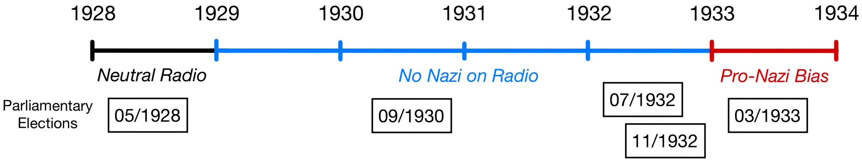
Moreover, if $r_k(d) = d^k$ for all $k = 0, \ldots, K$,

$$\mathcal{M} = \{g : \textit{card}(\{D_{g,1}, \ldots, D_{g,T}\}) \leq K + 2\}$$

Overview

Adena et al. (2015)

- TWFE regression of votes for the Nazi party on radio subscription rates at the electoral district × election date level



- Adena et al. (2015)'s treatment value for a voter i in group g with an access to the radio:

$$D_{i,g,t} = \begin{cases} -1 & \text{if election } t \text{ is held between 1929 and 1932} \\ 0 & \text{if election } t \text{ is held in 1928} \\ 1 & \text{if election } t \text{ is held in 1933} \end{cases}$$

Effects of Radio on Voting for the Nazis - Our Estimates

Parameter	1928	1930	07/1932	11/1932	1933
$\widehat{\Delta}_t$	0.171	0.177	0.237	0.225	0.364
	(0.344)	(0.355)	(0.360)	(0.358)	(0.367)
$\widehat{\mu}_t$	0	0.006	0.067	0.054	0.193^{**}
	–	(0.038)	(0.066)	(0.070)	(0.084)

Notes: $G = 850$. $K = 0$. We use our estimator with the covariates selected by Belloni (2014)'s double selection procedure. Standard errors, under parentheses, are clustered at the electoral region level.

* p < .1, ** p < .05, *** p < .01.

Conclusion

- Rationalize Chamberlain (1992)'s CRC model in the context of a *fuzzy* DiD design
- Provide an estimation strategy
  1. Robust to heterogeneous treatment effects
  2. Allowing for some endogenous selection into treatment
  3. Only requiring aggregated data at the group level
- Revisit Adena et al. (2015) and estimate the effect of political programs on the radio on voting for the Nazi party

Common Trends Assumption

- Assumption 3-(1) implies

$$E[Y_{g,t}(0) - Y_{i,g,t-1}(0)] = \beta_t - \beta_{t-1} \quad (5)$$

- Assumptions 3, 4 and 5 imply that

$$E[Y_{g,t}(1) - Y_{i,g,t-1}(1)] = \delta_t - \delta_{t-1} + \beta_t - \beta_{t-1} \quad (6)$$

- Similar conditions are imposed in de Chaisemartin & d'Haultfœuille (2018, 2020)

Back

Generalized Roy Model

Proposition 1

Suppose a balanced panel of individuals, Model (1) *holds, and for all* $(i, g, t) \in \{1, ..., N_{g,t}\} \times \{1, ..., G\} \times \{1, ..., T\}$, $E[\eta_{i,g,t}] = 0$ *and*

$$D_{i,g,t} = \mathbf{1}\{E[\Delta_{i,g,t}] \geq c_{g,t}\} \quad (7)$$

for some real numbers $c_{g,t}$, *then Assumptions 3 - (2), 4 and 5 are satisfied*

Back

Steps to Lemma 1

- Model (1) $\implies$

$$E[Y_{g,t}|\mathbf{D}_g^a] = \beta_t + \alpha_g + E\left[\frac{1}{N_{g,t}}\sum_{i=1}^{N_{g,t}}\varepsilon_{i,g,t}\middle|\mathbf{D}_g^a\right] + E\left[\frac{1}{N_{g,t}}\sum_{i=1}^{N_{g,t}}D_{i,g,t}(\gamma_{i,g(t)}+\delta_t+\eta_{i,g,t})\middle|\mathbf{D}_g^a\right]$$

- Assumption 3-part(1) $\implies$

$$E[Y_{g,t}|\mathbf{D}_g^a] = \beta_t + \alpha_g + E\left[\frac{1}{N_{g,t}}\sum_{i=1}^{N_{g,t}}D_{i,g,t}\gamma_{i,g(t)}\middle|\mathbf{D}_g^a\right] + D_{g,t}\delta_t + E\left[\frac{1}{N_{g,t}}\sum_{i=1}^{N_{g,t}}D_{i,g,t}\eta_{i,g,t}\middle|\mathbf{D}_g^a\right]$$

$$E[Y_{g,t}|\mathbf{D}_g^a] = \beta_t + \alpha_g + D_{g,t}E\left[\frac{1}{N_{g,t}D_{g,t}}\sum_{i=1}^{N_{g,t}}\gamma_{i,g(t)}\middle|\mathbf{D}_g^a\right] + D_{g,t}\delta_t + D_{g,t}E\left[\frac{1}{N_{g,t}D_{g,t}}\sum_{i=1}^{N_{g,t}}\eta_{i,g,t}\middle|\mathbf{D}_g^a\right]$$

Steps to Lemma 1

- Assumptions 3-part(2),4 and 5 $\Longrightarrow$

$$E[Y_{g,t} \Big| \mathbf{D}_g^a] = \beta_t + \alpha_g + D_{g,t}\Gamma_g(D_{g,t}) + D_{g,t}\delta_t$$

- Assumptions 6 $\Longrightarrow$

$$E[Y_{g,t} \Big| \mathbf{D}_g^a] = \beta_t + D_{g,t}\delta_t + \alpha_g + D_{g,t}\sum_{k=0}^{K}\gamma_{k,g}r_k(D_{g,t})$$

Back

Necessary and Sufficient Conditions for Assumption 7

Let $S_g = \mathbf{1}\{D_{g,1} = \cdots = D_{g,T}\}$

Proposition 2

Assumption 7 holds if:

1. *Either* $P(S_g = 1) > 0$ *and* $V(D_{g,1}|S_g = 1) > 0$
2. *or* $P(S_g = 0) > 0$, $T \geq 3$, $K = 0$, $r_0(d) = 1$ *and* $(1, D_1, D_t)$ *are not collinear conditional on* $S_g = 0$, *for all* $t = 2, \ldots, T$

Conversely, if Assumption 7 holds and:

1. $T = 2$, *then* $P(S_g = 1) > 0$ *and* $V(D_{g,1}|S_g = 1) > 0$
2. $r_k(d) = d^k$ *for all* k *and* $K = T - 2$, *then* $P(card(\{D_{g,1}, \ldots, D_{g,T}\}) = T) < 1$

Back