Lecture 12: Repeated Observations II

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

March 28, 2022

Recap

(i indexes units, t time periods, and k periods under treatment)

- One-way FE: $Y_{it} = \rho D_{it} + \alpha_i + \varepsilon_{it}$

Recap

(i indexes units, t time periods, and k periods under treatment)

- One-way FE: $Y_{it} = \rho D_{it} + \alpha_i + \varepsilon_{it}$
  - Equiv. of dummy var, within, and unit-mean centering (sweep).
  - Yields conditional variance weighted estimate of ρ under strong parametric restrictions (confounders are linearly separable and time-invariant, effects don't accumulate over time).
  - Centered-interaction specification can target, e.g., ATE.

Recap

(i indexes units, t time periods, and k periods under treatment)

- One-way FE: $Y_{it} = \rho D_{it} + \alpha_i + \varepsilon_{it}$
  - Equiv. of dummy var, within, and unit-mean centering (sweep).
  - Yields conditional variance weighted estimate of ρ under strong parametric restrictions (confounders are linearly separable and time-invariant, effects don't accumulate over time).
  - Centered-interaction specification can target, e.g., ATE.
- Two-way FE: $Y_{it} = \sum_k \rho^k D_{it}^k + \alpha_i + \lambda_t + \varepsilon_{it}$

Recap

(i indexes units, t time periods, and k periods under treatment)

- One-way FE: $Y_{it} = \rho D_{it} + \alpha_i + \varepsilon_{it}$
  - Equiv. of dummy var, within, and unit-mean centering (sweep).
  - Yields conditional variance weighted estimate of ρ under strong parametric restrictions (confounders are linearly separable and time-invariant, effects don't accumulate over time).
  - Centered-interaction specification can target, e.g., ATE.
- Two-way FE: $Y_{it} = \sum_k \rho^k D_{it}^k + \alpha_i + \lambda_t + \varepsilon_{it}$
  - Single-post-period DID or multi-post-period event study.
  - Identifies ATTs under parallel trends. (i.e., for two-period DID, constructs counterfactual trend in treatment group: $E[Y_{i1}^C - Y_{i0}^C \mid g[i] = 1] = E[Y_{i1} - Y_{i0} \mid g[i] = 0]$.)
  - Extend to triple differences (debiasing argument).

Recap

(i indexes units, t time periods, and k periods under treatment)

- One-way FE: $Y_{it} = \rho D_{it} + \alpha_i + \varepsilon_{it}$
  - Equiv. of dummy var, within, and unit-mean centering (sweep).
  - Yields conditional variance weighted estimate of ρ under strong parametric restrictions (confounders are linearly separable and time-invariant, effects don't accumulate over time).
  - Centered-interaction specification can target, e.g., ATE.
- Two-way FE: $Y_{it} = \sum_k \rho^k D_{it}^k + \alpha_i + \lambda_t + \varepsilon_{it}$
  - Single-post-period DID or multi-post-period event study.
  - Identifies ATTs under parallel trends. (i.e., for two-period DID, constructs counterfactual trend in treatment group: $E[Y_{i1}^C - Y_{i0}^C \mid g[i] = 1] = E[Y_{i1} - Y_{i0} \mid g[i] = 0]$.)
  - Extend to triple differences (debiasing argument).
- Synthetic control: construct counterfactual post-treatment trend using weighted average of controls: $Y_{it}^C = \sum_j w_j Y_{jt}^C$.

Today

- Multiple post-treatment periods and variable treatment timing.
- Changes in changes.
- Generalized synthetic control and interactive FE.
- Regularized high-dimensional estimators.

Vanilla TWFE

- Vanilla TWFE involves using OLS with multiperiod data to fit specifications like,

$$Y_{it} = \rho D_{it} + \alpha_i + \lambda_t + \varepsilon_{it}$$

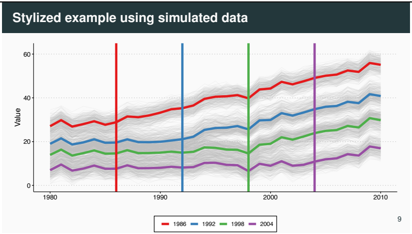
or (with multiple pre or post periods)

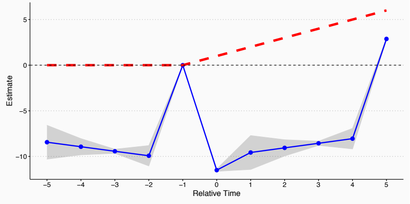
$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-K}^{-1} \gamma_t^{lead} D_{it}^k + \sum_{k=1}^{L} \gamma_t^{lag} D_{it}^k + \varepsilon_{it}$$

- Problems arise when there are multiple post-treatment periods, variable treatment timing, and effect heterogeneity over time.

Problems with vanilla TWFE

DGP: | Result:





Specification: $Y_{it} = \alpha_i + \lambda_t + \sum_{k=-K}^{-1} \gamma_t^{lead} D_{it}^k + \sum_{k=1}^{L} \gamma_t^{lag} D_{it}^k + \varepsilon_{it}$

(Callaway & Sant'Anna, 2020 lecture slides)

Problems with vanilla TWFE

	Time		
Unit	1	2	3
1	0	1	1
2	0	1	1
3	0	$\boxed{0}$	1
4	0	$\boxed{0}$	1
5	0	$\boxed{0}$	0

Matched controls

	Time		
Unit	1	2	3
1	0	1	1
2	0	1	1
3	$\boxed{0}$	0	1
4	$\boxed{0}$	0	1
5	$\boxed{0}$	0	0

Valid second differences

	Time		
Unit	1	2	3
1	0	1	1
2	0	1	1
3	0	0	$\boxed{1}$
4	0	0	$\boxed{1}$
5	0	0	0

Invalid second differences

(Strezhnev, 2018)

To construct counterfactual Y_{12}^{C}, TWFE involves the following:

- *Match* within period 2: take mean for units 3,4,5 (panel 1). Biased, bc doesn't account for cross-sectional heterogeneity.
- *De-bias* with second differences across periods:
  - First, subtract differences between unit 1 and 3,4,5 when all *under control* (panel 2). Fine, what traditional two-period DID does.
  - Second, subtract differences btwn 1 and 3,4,5 when all are *under treatment*. Problem: 3,4,5 under treatment for one period.

Problems with TWFE

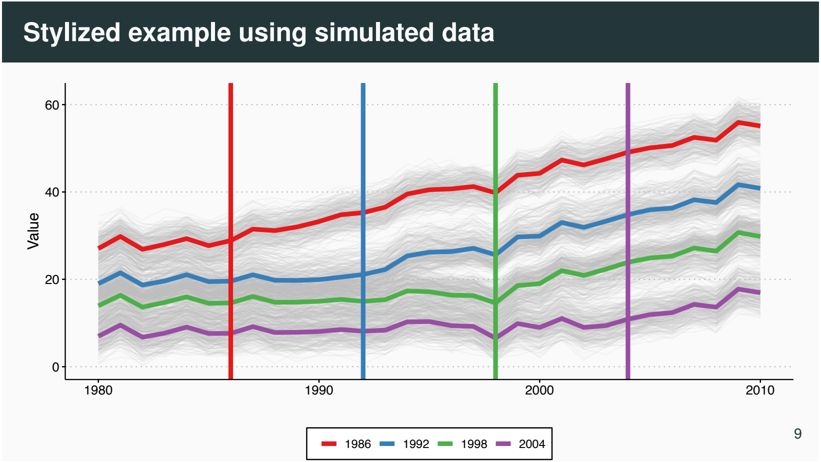
- Result “caution[s] against summarizing time-varying effects with a single-coefficient” (Goodman-Bacon, 2019, 3).
- A few versions of this result: Borusyak & Jaravel (2017), de Chaisemartin & D’Haultfoeuille (2020), Goodman-Bacon (2019), Imai & Kim (2020), Strezhnev (2018).
- de Chaisemartin & D’Haulftfoeuille (2020) and Goodman-Bacon (2019) propose diagnostic measures to examine the extent of the problem.
- Abraham & Sun (2018) derive a similar result for bias when using leads to estimate pre-trend differences in DID or event study designs.

In place of TWFE

Abraham & Sun (2018), de Chaisemartin & D'Haultfoeuille (2020), and Goodman-Bacon (2019):

- Separate DID estimates for each cohort entering into treatment.
- Take an average of these separate DIDs.
- Under other parallel trends assumption for *ATC*, can work with groups that *drop out* of treatment too (see de Chaisemartin & D'Haultfoeuille, 2020).

In place of TWFE



(Callaway & Sant'Anna, 2020 lecture slides)

In place of TWFE

- Key take-away: make sure you are always using *clean controls* (don't use treated-treated differences, only control-control differences to do debiasing).
- Callaway & Sant'Anna (2018) and Strezhnev (2018) note that parallel trends between *same control group* and different treatment cohorts may be implausible.
- Implies *conditional DID* important in this setting.
- C & S'A and Strezhnev propose IPW estimators and augmented-IPW (doubly-robust) estimators that use covariates to line up treated cohorts with appropriately reweighted/modeled untreated trajectories.
- C & S'A develop methods for uniform inference for effect trajectories.
- Method extends to event studies as well.

Changes-in-Changes

- Athey & Imbens (2006) drop linearity assumptions.
- Develop a more agnostic "changes in changes" approach.
- Characterize not just conditional mean effects, but entire distributional effects.
- Can be used to estimate, e.g., median and other quantile effects.

Changes-in-Changes

Consider an unconditional DID setting (could incorporate X via matching or IPW). Now suppose:

Changes-in-Changes

Consider an unconditional DID setting (could incorporate X via matching or IPW). Now suppose:

1. $Y_{it}^C = h(U_i, t)$, where U_i unobserved variables whose distribution may vary over groups (g).

Changes-in-Changes

Consider an unconditional DID setting (could incorporate X via matching or IPW). Now suppose:

1. $Y_{it}^C = h(U_i, t)$, where U_i unobserved variables whose distribution may vary over groups (g).
2. $h(u,t)$ is strictly increasing in u (thus, u a latent score). Fixing u, this preserves rank ordering of units from $h(u,0)$ to $h(u,1)$.

Changes-in-Changes

Consider an unconditional DID setting (could incorporate X via matching or IPW). Now suppose:

1. $Y_{it}^C = h(U_i, t)$, where U_i unobserved variables whose distribution may vary over groups (g).
2. $h(u,t)$ is strictly increasing in u (thus, u a latent score). Fixing u, this preserves rank ordering of units from $h(u,0)$ to $h(u,1)$.
3. Support of U_i for $g = 0$ fully overlaps support of U_i for $g = 1$.

Changes-in-Changes

- Then for the counterfactual of interest,

$$\mathrm{E}\left[Y_{i1}^{C}|g[i]=1\right]=\mathrm{E}\left[F_{Y,01}^{-1}\left(F_{Y,00}(Y_{i0})\right)|g[i]=1\right],$$

where $F_{Y,00}(.)$ and $F_{Y,01}^{-1}(.)$ are the CDF and inverse CDF of outcomes for $g=0,t=0$ and $g=0,t=1$, respectively.

Changes-in-Changes

- Then for the counterfactual of interest,

$$\mathrm{E}\left[Y_{i1}^{C}|g[i]=1\right]=\mathrm{E}\left[F_{Y,01}^{-1}\left(F_{Y,00}(Y_{i0})\right)|g[i]=1\right],$$

  where $F_{Y,00}(.)$ and $F_{Y,01}^{-1}(.)$ are the CDF and inverse CDF of outcomes for $g=0,t=0$ and $g=0,t=1$, respectively.
- And so,

$$\begin{aligned}&\mathrm{E}\left[Y_{i1}^{T}-Y_{i1}^{C}|g[i]=1\right]\\&\quad=\mathrm{E}\left[Y_{i1}^{T}|g[i]=1\right]-\mathrm{E}\left[F_{Y,01}^{-1}\left(F_{Y,00}(Y_{i0})\right)|g[i]=1\right].\end{aligned}$$

Changes-in-Changes

Constructing $\mathrm{E}\left[F_{Y,01}^{-1}\left(F_{Y,00}(Y_{i0})\right) | g[i] = 1\right]$ requires three steps:

1. Take y from quantile q of pretreatment distribution (Y_{10}).
2. Feed it into pre-treatment control group CDF (Y_{00}), match quantile (q') with post-treatment control group CDF (Y_{01}).
3. Then cast back onto the outcome to form quantile q of post-treatment *counterfactual* Y_{11} CDF.

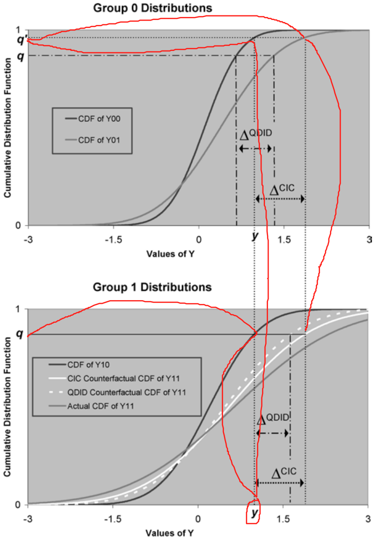


FIGURE 1.—Illustration of transformations.

Changes-in-Changes

- $y + \Delta CIC$ is the counterfactual value for $g = 1$ units with $Y_{i0} = y$.
- We can do this over the support of the outcomes for $g = 1$ and $t = 0$, completing the distribution of counterfactual Y_{i1}^{C}'s for $g = 1$.
- We can then use these to compute ATT, or any difference-in-distribution effect (e.g., quantile effects).

Changes-in-Changes

- $y + \Delta CIC$ is the counterfactual value for $g = 1$ units with $Y_{i0} = y$.
- We can do this over the support of the outcomes for $g = 1$ and $t = 0$, completing the distribution of counterfactual Y_{i1}^C's for $g = 1$.
- We can then use these to compute ATT, or any difference-in-distribution effect (e.g., quantile effects).
- Athey and Imbens show quantile effect estimator is asymptotically normal, so could use bootstrap inference. They also provide analytical standard errors.
- If the standard DID assumptions hold, this ATT estimator converges to the standard DID ATT estimator.
- For discrete outcomes, Athey and Imbens provide bounds results and additional point identification results.

Generalized Synthetic Control

- Synthetic control different than DID:
- DID: You think the untreated trends subject only to uniform time shocks. Allows for selection on level differences pre-treatment such that you don't even need overlap between treated and control trends.
- Synthetic control: construct a counterfactual trend as a convex combination of untreated trends. Requires overlap between treated trend and control trends.

Generalized Synthetic Control

- Following Xu (2017) (cf. Bai (2009) "interactive fixed effects")
- N_{tr} (not just one) and N_{co} control, T periods.
- For those treated, T_0 pretreatment periods ($D_{it} = 1$ as of $T_0 + 1$).

Generalized Synthetic Control

- Following Xu (2017) (cf. Bai (2009) "interactive fixed effects")
- N_{tr} (not just one) and N_{co} control, T periods.
- For those treated, T_0 pretreatment periods ($D_{it} = 1$ as of $T_0 + 1$).
- DGP is a "factor model": $Y_{it} = \delta_{it} D_{it} + X'_{it}\beta + \lambda'_i f_t + \varepsilon_{it}$
  - ε_{it} independent mean zero error,
  - f_t vector of period-specific factors, normalized and orthogonal,
  - λ_i corresponding vector of unit-specific factor loadings.

Generalized Synthetic Control

- Following Xu (2017) (cf. Bai (2009) "interactive fixed effects")
- N_{tr} (not just one) and N_{co} control, T periods.
- For those treated, T_0 pretreatment periods ($D_{it} = 1$ as of $T_0 + 1$).
- DGP is a "factor model": $Y_{it} = \delta_{it} D_{it} + X'_{it}\beta + \lambda'_i f_t + \varepsilon_{it}$
  - ε_{it} independent mean zero error,
  - f_t vector of period-specific factors, normalized and orthogonal,
  - λ_i corresponding vector of unit-specific factor loadings.
- Estimating ATT:
  - Estimate $\hat{f}_t$ on control,
  - Use these $\hat{f}_t$s to estimate $\hat{\lambda}_i$ and $\hat{\beta}$ in treated,
  - Construct counterfactual trend for treated with $X'_{it}\hat{\beta} + \hat{\lambda}'_i \hat{f}_t$
  - Use observed treated trend and counterfactual trend to estimate ATT.

Regularized high-dimensional estimators

- Following Doudchenko and Imbens (2017), the inference problem here is one of missing counterfactual data. We see

$$\mathbf{Y}^{obs} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix} \Rightarrow \mathbf{Y}(0) = \begin{pmatrix} ? & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}.$$

Regularized high-dimensional estimators

- Following Doudchenko and Imbens (2017), the inference problem here is one of missing counterfactual data. We see

$$\mathbf{Y}^{obs} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix} \Rightarrow \mathbf{Y}(0) = \begin{pmatrix} ? & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}.$$

- Most counterfactual estimators have the form,

$$\hat{Y}_{0,T}(0) = \mu + \sum_{i=1}^{N} \omega_i Y_{i,T}^{obs}.$$

  - Synth: $\mu = 0$, ω_i's add to 1 and ≥ 0, although can vary.
  - DID: $\mu \neq 0$, but ω_i's add to 1, nonnegative, and constant.

Regularized high-dimensional estimators

- Following Doudchenko and Imbens (2017), the inference problem here is one of missing counterfactual data. We see

$$\mathbf{Y}^{obs} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix} \Rightarrow \mathbf{Y}(0) = \begin{pmatrix} ? & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}.$$

- Most counterfactual estimators have the form,

$$\hat{Y}_{0,T}(0) = \mu + \sum_{i=1}^{N} \omega_i Y_{i,T}^{obs}.$$

  - Synth: $\mu = 0$, ω_i's add to 1 and ≥ 0, although can vary.
  - DID: $\mu \neq 0$, but ω_i's add to 1, nonnegative, and constant.
- Doudchenko and Imbens (2017) drop the restrictions. This introduces an identification problem—MMSE solution is non-unique. Their solution: regularization.

Regularized high-dimensional estimators

- Following Doudchenko and Imbens (2017), the inference problem here is one of missing counterfactual data. We see

$$\mathbf{Y}^{obs} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix} \Rightarrow \mathbf{Y}(0) = \begin{pmatrix} ? & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}.$$

- Most counterfactual estimators have the form,

$$\hat{Y}_{0,T}(0) = \mu + \sum_{i=1}^{N} \omega_i Y_{i,T}^{obs}.$$

  - Synth: $\mu = 0$, ω_i's add to 1 and ≥ 0, although can vary.
  - DID: $\mu \neq 0$, but ω_i's add to 1, nonnegative, and constant.
- Doudchenko and Imbens (2017) drop the restrictions. This introduces an identification problem—MMSE solution is non-unique. Their solution: regularization.
- Athey et al. (2017) use another approach: "complete" $\mathbf{Y}(0)$ based on a best-fitting factorized decomposition of the matrix, under matrix regularization constraints (approx. rank minimization).

Regularized high-dimensional estimators

- Doudchenko and Imbens (2017) develop useful ideas for inference.

Regularized high-dimensional estimators

- Doudchenko and Imbens (2017) develop useful ideas for inference.
- We want to account for uncertainty in the counterfactual estimate:

$$\hat{\tau} = Y_{11} - \hat{Y}_{11}^{C} \Rightarrow \mathrm{E}\left(\hat{\tau} - \tau\right)^{2} = \mathrm{E}\left(\hat{Y}_{11}^{C} - Y_{11}^{C}\right)^{2}.$$

Regularized high-dimensional estimators

- Doudchenko and Imbens (2017) develop useful ideas for inference.
- We want to account for uncertainty in the counterfactual estimate:

$$\hat{\tau} = Y_{11} - \hat{Y}_{11}^C \Rightarrow \mathrm{E}\left(\hat{\tau} - \tau\right)^2 = \mathrm{E}\left(\hat{Y}_{11}^C - Y_{11}^C\right)^2.$$

- One way to approximate this is to assume that Y_{11}^C is exchangeable wrt the Y_{01}^C values that we observe.
- Then, you can estimate $\mathrm{E}\left(\hat{Y}_{11}^C - Y_{11}^C\right)^2$ with the residual distribution from *placebo estimates* of the Y_{01}^C values.

What to use?

Estimator	Situation
Classic DID	▶ Single treatment cohort ▶ Control trends don't overlap treated trend ▶ Use CIC for distributional effects, issues with linearity
Classic Synth	▶ Single treatment unit ▶ Control trends overlap treated trend ▶ Long pretrend data
New DID (dCM & dH; C & S'A)	▶ Multiple cohorts receiving treatment at different times ▶ Control trends don't overlap treated trend
Gen. Synth	▶ Multiple treated units, possibly at different times ▶ Control trends overlap treated trends ▶ Long pretrend data
High-dim. est.	▶ Multiple treated units, possibly at different times ▶ Large N (needed to "train" these models) ▶ At least some control trends overlap treated trends

Remarks

- Liu, Wang & Xu (2019) offer tests for different identifying assumptions (R package: `fect`).
- Recently proposed, improved methods address problems with classic TWFE and event studies in different ways:
- Some proposed improvements to TWFE are "going back to basics"— run many simple DIDs and aggregate. Idea is to restore design-based foundation.
- Others, like generalized synthetic control and the regularized high-dimensional estimators, work with more flexible and complete DGPs, and then use modest restrictions or regularization to address underidentification.