Lecture 14: Repeated Observations II

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

March 23, 2017

- Standard difference in differences (DID) (MHE).
- "Conditional" DID (Abadie, 2005).
- "Changes-in-changes," generalizing DID (Athey & Imbens, 2006).
- Synthetic control (Abadie & Gardeazabal, 2003).
- Recent generalizations (Doudchenko & Imbens, 2017; Athey et al. 2017).

Differences-in-Differences

DID is a special case of FE:

Differences-in-Differences

DID is a special case of FE:

- Suppose *i*s are grouped into 2 sets indexed by $g = 0, 1$ for "treated" and "control" and two time periods labeled $t = 0, 1$.
- $D_{g[i]0} = 0$ for all g, while $D_{11} = 1$ and $D_{01} = 0$

Differences-in-Differences

DID is a special case of FE:

- Suppose is are grouped into 2 sets indexed by $g = 0, 1$ for "treated" and "control" and two time periods labeled $t = 0, 1$.
- $D_{g[i]0} = 0$ for all g, while $D_{11} = 1$ and $D_{01} = 0$
- An FE model with time and group effects is:

$$\begin{aligned} Y_{it} &= \mu + \alpha_{g[i]} + \lambda_t + \delta D_{g[i]t} + \varepsilon_{it} \\ &= \beta_0 + \beta_1 \cdot 1(g[i] = 1) + \beta_2 \cdot 1(t = 1) + \delta \cdot 1(g[i] = 1) \cdot 1(t = 1) + \varepsilon_{it} \end{aligned}$$

where latter is how DID is often estimated.

Differences-in-Differences

DID is a special case of FE:

- Suppose *i*s are grouped into 2 sets indexed by $g = 0, 1$ for "treated" and "control" and two time periods labeled $t = 0, 1$.
- $D_{g[i]0} = 0$ for all g, while $D_{11} = 1$ and $D_{01} = 0$
- An FE model with time and group effects is:

$$\begin{aligned} Y_{it} &= \mu + \alpha_{g[i]} + \lambda_t + \delta D_{g[i]t} + \varepsilon_{it} \\ &= \beta_0 + \beta_1 \cdot 1(g[i] = 1) + \beta_2 \cdot 1(t = 1) + \delta \cdot 1(g[i] = 1) \cdot 1(t = 1) + \varepsilon_{it} \end{aligned}$$

  where latter is how DID is often estimated.
- With a little algebra,

$$\delta = \mathrm{E}[Y_{i1} - Y_{i0} | g[i] = 1] - \mathrm{E}[Y_{i1} - Y_{i0} | g[i] = 0]$$

  which shows how δ is indeed a "difference in differences."
  (NB: 1 and 0 subscripts are time periods here, not potential outcomes.)

Differences-in-Differences

DID is a special case of FE:

- To avoid confusion, let's call potential outcomes under control Y_{it}^C and under treatment Y_{it}^T.
- We observe $Y_{i0}^C = Y_{i0}$ for everyone.
- We observe $Y_{i1}^C = Y_{i1}$ for $g[i] = 0$, and $Y_{i1}^T = Y_{i1}$ for $g[i] = 1$.
- We would like the ATT: $\mathrm{E}\left[Y_{i1}^T - Y_{i1}^C | g[i] = 1\right]$.
- Assumption 1: Suppose a form of mean independence:

$$\mathrm{E}\left[Y_{i1}^C - Y_{i0}^C | g[i] = 0\right] = \mathrm{E}\left[Y_{i1}^C - Y_{i0}^C | g[i] = 1\right].$$

Trend in control is equal to what trend *would have been* among treated *had treatment never been applied*.

Differences-in-Differences

- Then, δ is identified for the ATT:

$$\delta = \underbrace{\mathrm{E}\left[Y_{i1}^{T} - Y_{i0}^{C} | g[i] = 1\right] - \mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | g[i] = 0\right]}_{\text{observed}}$$

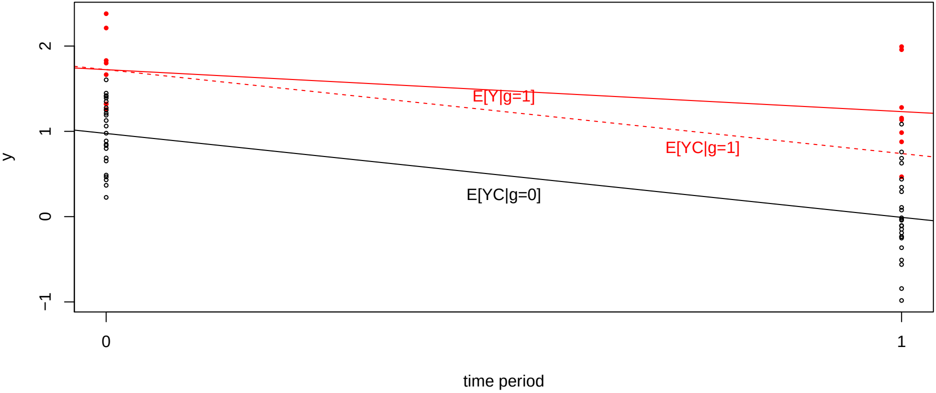
$$= \mathrm{E}\left[Y_{i1}^{T} - Y_{i0}^{C} | g[i] = 1\right] - \underbrace{\mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | g[i] = 1\right]}_{\text{counterfactual}}$$

$$= \mathrm{E}\left[Y_{i1}^{T} - Y_{i1}^{C} | g[i] = 1\right].$$

Differences-in-Differences

- Assumption 1 is the standard DID "parallel trends" assumption.
- Implies control group trend is parallel to what *would have happened* to treatment group members were there no treatment.
- I.e., control group trend is parallel to counterfactual trend for treatment group.
- You can exploit Assumption 1 with repeated cross sections (you don't need panel data).

Differences-in-Differences



Differences-in-Differences

Some considerations for identification:

- Trend assumptions are sensitive to transformations! (Linear trend in natural scale implies non-linear trend in log scale.)
- Trend assumptions may not be plausible on levels, though perhaps on differences or other higher order differences. Identification is still possible (Mora & Reggio, 2013).
- Trend assumptions may be plausible only for classes of similar units and not for treated and control groups as a whole.

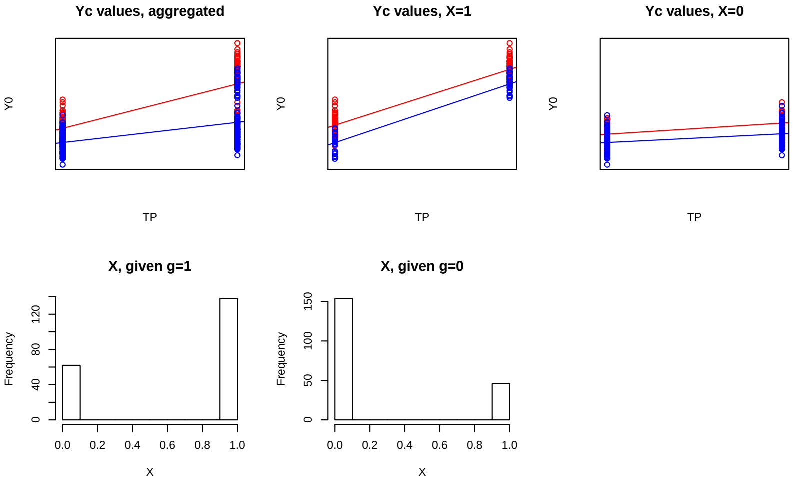
Conditional Differences-in-Differences

- Abadie (2005) considers conditional mean independence:
- Assumption 2:

$$\mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | g[i] = 0, X_i\right] = \mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | g[i] = 1, X_i\right],$$

and $\Pr[g[i] = 1 | X_i] < 1$ for all X_i.

Conditional Differences-in-Differences



Conditional Differences-in-Differences

- Then,

$$\begin{aligned}\delta_x &\equiv \mathrm{E}[Y_{i1} - Y_{i0}|g[i] = 1, X_i] - \mathrm{E}[Y_{i1} - Y_{i0}|g[i] = 0, X_i] \\ &= \mathrm{E}[Y_{i1}^T - Y_{i0}^C|g[i] = 1, X_i] - \mathrm{E}[Y_{i1}^C - Y_{i0}^C|g[i] = 0, X_i] \\ &= \mathrm{E}[Y_{i1}^T - Y_{i0}^C|g[i] = 1, X_i] - \mathrm{E}[Y_{i1}^C - Y_{i0}^C|g[i] = 1, X_i] \\ &= \mathrm{E}[Y_{i1}^T - Y_{i1}^C|g[i] = 1, X_i],\end{aligned}$$

and so the ATT is identified, since

$$\begin{aligned}\int_x \mathrm{E}[Y_{i1}^T - Y_{i1}^C|g[i] = 1, X_i = x] f(x|g[i] = 1)dx \\ = \mathrm{E}[Y_{i1}^T - Y_{i1}^C|g[i] = 1].\end{aligned}$$

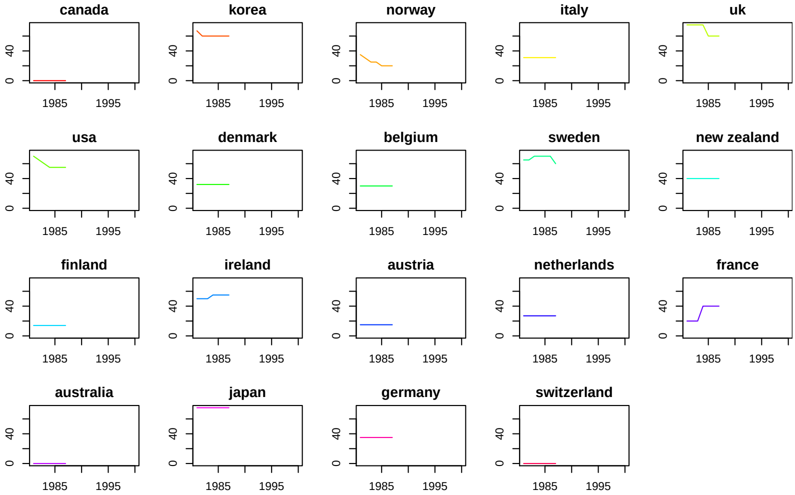
Conditional Differences-in-Differences

- Three different ways to exploit Assumption 2:
  1. Regression model that incorporate X_i.
     - Consider interactions with time period and group dummies, higher order X_i terms, etc.
     - Key is to trace out outcome trajectories under control.
     - Risk of specification or aggregation biases.
  2. Inverse-propensity score weighting using $e(X_i)$.
  3. Matching on X_i.

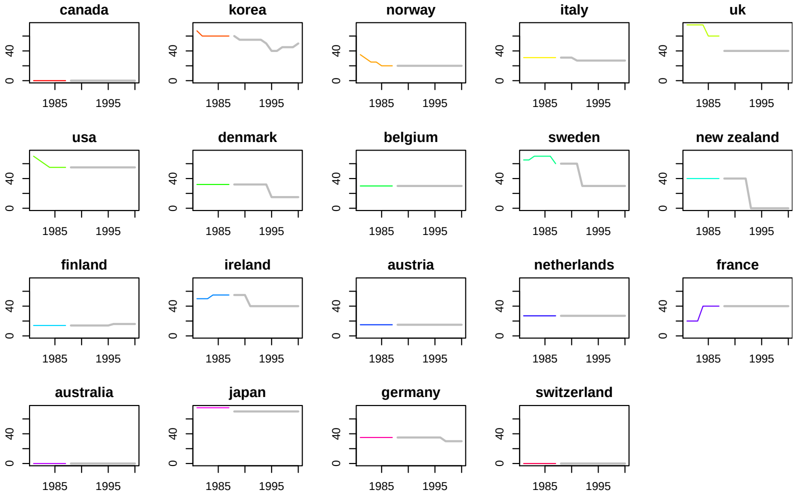
Conditional Differences-in-Differences

- If you only have a repeated cross section, conditional DID methods are limited by the need to have *retrospective* covariate data so that you can adjust/weight/match with pre-treatment covariates.
- Often, X_i contains pre-treatment time series on key variables (e.g., pre-treatment values of the outcome of interest). Matching or weighting to achieve balance on these is a compelling way to make the "parallel trends" assumption believable.

Conditional Differences-in-Differences



Conditional Differences-in-Differences



Extended DID

"Triple differences":

Extended DID

"Triple differences":

- Policy in state (S) targets poor (P) but leaves rich (R) unaffected.

Extended DID

"Triple differences":

- Policy in state (S) targets poor (P) but leaves rich (R) unaffected.
- Under parallel trends for P and R, could estimate effect as,

$$\delta = \mathrm{E}\left[Y_{i1}^{T} - Y_{i0}^{C} | P\right] - \mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | R\right].$$

Extended DID

"Triple differences":

- Policy in state (S) targets poor (P) but leaves rich (R) unaffected.
- Under parallel trends for P and R, could estimate effect as,

$$\delta = \mathrm{E}\left[Y_{i1}^T - Y_{i0}^C | P\right] - \mathrm{E}\left[Y_{i1}^C - Y_{i0}^C | R\right].$$

- Controls for any "state" effects.
- But poor may have different trends than rich.

Extended DID

"Triple differences":

- Policy in state (S) targets poor (P) but leaves rich (R) unaffected.
- Under parallel trends for P and R, could estimate effect as,

$$\delta = \mathrm{E}\left[Y_{i1}^{T} - Y_{i0}^{C} | P\right] - \mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | R\right].$$

- Controls for any "state" effects.
- But poor may have different trends than rich.
- We can incorporate poor from another state (S'):

$$\delta_3 = \mathrm{E}\left[Y_{i1}^{T} - Y_{i0}^{C} | P, S\right] - \Big(\underbrace{\mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | R, S\right]}_{\text{state effect}} + \underbrace{\mathrm{E}\left[Y_{i1}^{C} - Y_{i0}^{C} | P, S'\right]}_{\text{poor effect}}\Big)$$

- With P, S, and T as poor, state S, and $t = 1$ indicators, estimate

$$Y_{it} = \beta_0 + \beta_1 P + \beta_2 S + \beta_3 PS + \delta_0 T + \delta_1 PT + \delta_2 ST + \delta_3 PST + \varepsilon_{it}$$

- Incorporate covariates as above.

DID inference

- Traditionally, inference proceeds by just accounting for sampling variability in estimating the time- and group-specific means
  - Just "robust" or "cluster robust" within g.
- If there is outcome clustering by g, then there is a problem.
  - Traditional approach is anti-conservative (Bertrand et al., 2004).
  - By definition there is treatment clustering by g.
  - So, with outcome clustering by g, we need lots of groups.
  - Similar for outcomes clustering by t.
  - See recent "Ask Guido" post:

`http://blogs.worldbank.org/impactevaluations/introducing-ask-guido`

- Recent contributions on DID inference with few groups: MacKinnon & Webb (2016), Ferman & Pinto (2015).
- A different, and I think especially promising angle, is Doudchenko & Imbens (2017)–more later.

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:

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.
- And so,

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

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.

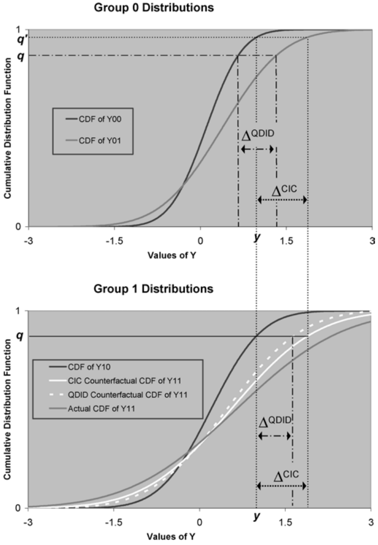


FIGURE 1.—Illustration of transformations.

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).
- 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.

Synthetic Control

- Abadie & Gardeazabal (2003) develop a DID method for "quantitative case studies."
- Their application is to estimate the effects of terrorism in the Basque region on the prosperity of the region.
- They do so by creating a "synthetic" Basque region out of the rest of Spain, and then estimating effects via DID.

Synthetic Control

TABLE 3—PRE-TERRORISM CHARACTERISTICS, 1960'S

	Basque Country (1)	Spain (2)	"Synthetic" Basque Country (3)
Real per capita GDP[a]	5,285.46	3,633.25	5,270.80
Investment ratio (percentage)[b]	24.65	21.79	21.58
Population density[c]	246.89	66.34	196.28
Sectoral shares (percentage)[d]			
Agriculture, forestry, and fishing	6.84	16.34	6.18
Energy and water	4.11	4.32	2.76
Industry	45.08	26.60	37.64
Construction and engineering	6.15	7.25	6.96
Marketable services	33.75	38.53	41.10
Nonmarketable services	4.07	6.97	5.37
Human capital (percentage)[e]			
Illiterates	3.32	11.66	7.65
Primary or without studies	85.97	80.15	82.33
High school	7.46	5.49	6.92
More than high school	3.26	2.70	3.10

Sources: Authors' computations from Matilde Mas et al. (1998) and Fundación BBV (1999).

[a] 1986 USD, average for 1960–1969.

[b] Gross Total Investment/GDP, average for 1964–1969.

[c] Persons per square kilometer, 1969.

[d] Percentages over total production, 1961–1969.

[e] Percentages over working-age population, 1964–1969.

Synthetic Control

- Suppose 1 treated and J potential control regions.
- Let $W = (w_1, ..., w_J)'$ be a vector of non-negative weights for each control region. Let $\sum_k w_j = 1$.
- Control region outcomes are combined using W to create a synthetic counterfactual for the treated region.
- We want to choose W^* to create the best possible synthetic counterfactual.

Synthetic Control

- Let X_1 be covariates for the treated region and $\mathbf{X}_0$ a matrix of covariates for the control region.
- Choose W^* to minimize weighted L_2 ("squared") distance,

$$(X_1 - \mathbf{X}_0 W)' \mathbf{V} (X_1 - \mathbf{X}_0 W),$$

  where $\mathbf{V}$ weights the different covariate discrepancies on the basis of their relative "importance."
- You are free to choose $\mathbf{V}$ as you see fit.
- Abadie & Gardeazabal choose $\mathbf{V}$ to give priority to minimizing distance between the pre-conflict GDP trend.
- Then, synthetic control outcomes for the treated region are computed as $\hat{Y^C}_{it} = Y'_{jt} W^*$.

Synthetic Control

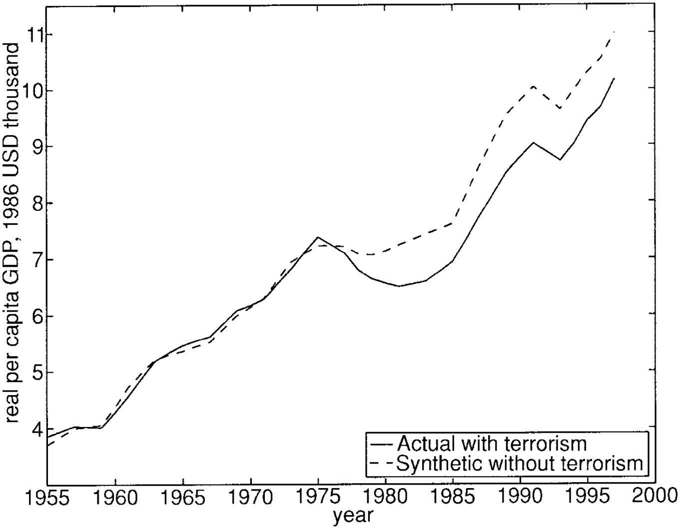


Figure 1. Per capita GDP for the Basque Country

Synthetic Control

- They use time series techniques (essentially ADL models) to estimate how effects transmit over time.
- Cf. DeBoef & Keele (2008) or Wooldridge (undergrad or grad textbook) for more on time series techniques.

Synthetic Control

- Inference in Abadie & Gardeazabal (2003) conditioned on the weighting solution.
- Abadie et al. (2008) propose a permutation based method to account for uncertainty about the weighting solution.
- See Xu (2017) for generalization to many treated units.

Recent Generalizations

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

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

Recent Generalizations

- Following Douchenko 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 nonnegative, although can vary.
  - DID: $\mu \neq 0$, but ω_i's add to 1, nonnegative, and constant.
- Douchenko and Imbens (2017) propose to drop these 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.

Recent Generalizations

- Douchenko and Imbens (2017) develop useful ideas for inference in this setting.
- The crucial thing here is to account for uncertainty in the counterfactual estimate:

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

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

Remarks

- DID is “cleaner” than the kinds of FE-panel or -TSCS applications that we considered last time. In those cases, the treatment could switch “on” and “off” at any time.
- This makes interpretation of the estimates fraught (Imai & Kim 2012; Sobel 2012).
- With DID, pre-treatment and post-treatment distinction is clear.
- This makes the identification and inference challenges and the identifying assumptions clearer and more plausible.