Political Methodology II
Section: Difference in Differences

Apoorva Lal

February 12, 2022

Stanford University

Roadmap

SOO Loose-ends

Difference-in-differences

General Exam Advice

Generalising to multiple time periods: Fixed Effects

Synthetic Control Methods

Roadmap

SOO Loose-ends

Difference-in-differences

General Exam Advice

Generalising to multiple time periods: Fixed Effects

Synthetic Control Methods

Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T \beta + \delta_R D_i$$

where X_i is a vector of covariates and D_i is the treatment indicator.

Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where X_i is a vector of covariates and D_i is the treatment indicator. Using our partialling out formula, we can write δ_R as

$$\delta_R = \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)}$$

Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where X_i is a vector of covariates and D_i is the treatment indicator. Using our partialling out formula, we can write δ_R as

$$\delta_R = \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]}$$

Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where X_i is a vector of covariates and D_i is the treatment indicator. Using our partialling out formula, we can write δ_R as

$$\delta_R = \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]]}$$

Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where X_i is a vector of covariates and D_i is the treatment indicator. Using our partialling out formula, we can write δ_R as

$$\begin{aligned}\delta_R &= \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]]} \\ &= \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[Var[D_i|X]]}\end{aligned}$$

Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where X_i is a vector of covariates and D_i is the treatment indicator. Using our partialling out formula, we can write δ_R as

$$\begin{aligned}\delta_R &= \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]]} \\ &= \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[Var[D_i|X]]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]}\end{aligned}$$

Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates X_i.

Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates X_i. Compare this to the unweighted propensity score estimand:

$$\delta_{ATE} = \mathbb{E}\Big[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big]$$

When will these two coincide?

Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates X_i. Compare this to the unweighted propensity score estimand:

$$\delta_{ATE} = \mathbb{E}\Big[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big]$$

When will these two coincide? Constant treatment effects across strata of X_i.

Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates X_i. Compare this to the unweighted propensity score estimand:

$$\delta_{ATE} = \mathbb{E}\Big[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big]$$

When will these two coincide? Constant treatment effects across strata of X_i. Otherwise, OLS does not estimate ATE/ATT.

OLS with heterogeneous treatment effects

In this case your estimat**or** is picking your estim**and**.

Imputation estimators

$$\widehat{\tau}_{\text{Reg}} = \frac{1}{N}\sum_{i=1}^{n}(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))$$

$$\widehat{\tau}_{\text{IPW}} = \frac{1}{N}\sum_{i=1}^{n}\left(\frac{D_i Y_i}{\widehat{\pi}(X_i)}\frac{(1-D_i)Y_i}{1-\widehat{\pi}(X_i)}\right)$$

$$\widehat{\tau}_{\text{AIPW}} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left[\overbrace{\hat{\mu}_1(\boldsymbol{X}_i)}^{\text{Regression}} + \overbrace{\frac{D_i(Y_i - \hat{\mu}_1(\boldsymbol{X}_i))}{\hat{\pi}(\boldsymbol{X}_i)}}^{\text{IPW}}\right]}_{\text{estimator for } \mathbb{E}[Y_i(1)]} - \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left[\hat{\mu}_0(\boldsymbol{X}_i) + \frac{(1-D_i)(Y_i - \hat{\mu}_0(\boldsymbol{X}_i))}{1-\hat{\pi}(\boldsymbol{X}_i)}\right]}_{\text{estimator for } \mathbb{E}[Y_i(0)]}$$

Fit $\widehat{\mu}, \widehat{\pi}$ using learner of choice.

Roadmap

SOO Loose-ends

Difference-in-differences

General Exam Advice

Generalising to multiple time periods: Fixed Effects

Synthetic Control Methods

DiD: Two Groups and Two Periods

Denote potential outcomes $Y_{(d)}(t)$ for $d \in \{0,1\}$, $t \in \{0,1\}$

Estimand (ATT in the 2nd period)

$\tau_{ATT} = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D=1	$E[Y_1(1)\|D = 1]$	$E[Y_0(0)\|D = 1]$
Control D=0	$E[Y_0(1)\|D = 0]$	$E[Y_0(0)\|D = 0]$

Problem

Missing potential outcome: $E[Y_0(1)|D = 1]$*, ie. what is the average post-period outcome for the treated in the absence of the treatment?*

Identification with Difference-in-Differences

Identification Assumption (parallel trends)

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$$

Identification Result

Given parallel trends the ATT is identified as:

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D = 1] &= \Big\{E[Y(1)|D = 1] - E[Y(1)|D = 0]\Big\} \\ &- \Big\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\Big\} \end{aligned}$$

Implicit functional form assumption: Parallel trends in levels $\neq$ Parallel trends in logs (growth rates). (cf Jensen's Inequality)

Non-parametric Identification with Difference-in-Differences

Start with the estimand we want:

$$\tau_{ATT} = E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1]$$

What is the missing data here?

Non-parametric Identification with Difference-in-Differences

Start with the estimand we want:

$$\tau_{ATT} = E[Y_1(1)|D=1] - E[Y_0(1)|D=1]$$

What is the missing data here? The missing data is $E[Y_0(1)|D=1]$ (the control counterfactual in the post-period for the treated group). To obtain an estimate of this, we assume parallel trends:

$$E[Y_0(1) - Y_0(0)|D=1] = E[Y_0(1) - Y_0(0)|D=0]$$

$$\implies E[Y_0(1)|D=1] = \underbrace{E[Y_0(0)|D=1]}_{\text{Level at } t=0} + \underbrace{E[Y_0(1) - Y_0(0)|D=0]}_{\text{Trend for control group}}$$

Non-parametric Identification with Difference-in-Differences

Start with the estimand we want:

$$\tau_{ATT} = E[Y_1(1)|D=1] - E[Y_0(1)|D=1]$$

What is the missing data here? The missing data is $E[Y_0(1)|D=1]$ (the control counterfactual in the post-period for the treated group). To obtain an estimate of this, we assume parallel trends:

$$\begin{aligned} E[Y_0(1) - Y_0(0)|D=1] &= E[Y_0(1) - Y_0(0)|D=0] \\ \implies E[Y_0(1)|D=1] &= \underbrace{E[Y_0(0)|D=1]}_{\text{Level at } t=0} + \underbrace{E[Y_0(1) - Y_0(0)|D=0]}_{\text{Trend for control group}} \end{aligned}$$

Substitute the assumption in for the missing data and rearrange:

$$\begin{aligned} \tau_{ATT} &= E[Y_1(1)|D=1] - E[Y_0(1)|D=1] \\ &= E[Y_1(1)|D=1] - E[Y_0(0)|D=1] + E[Y_0(1)|D=0] - E[Y_0(0)|D=0] \\ &= \underbrace{(E[Y(1)|D=1] - E[Y(0)|D=1])}_{\text{Before after for treated}} - \underbrace{(E[Y(1)|D=0] - E[Y(0)|D=0])}_{\text{Before after for control}} \end{aligned}$$

This is just the difference of pre-post differences between the treated and control groups.

Difference-in-Differences

Estimand (Sample Means: Panel)

$$\left\{ \frac{1}{N_1} \sum_{D_i=1} Y_i(1) - \frac{1}{N_0} \sum_{D_i=0} Y_i(1) \right\} - \left\{ \frac{1}{N_1} \sum_{D_i=1} Y_i(0) - \frac{1}{N_0} \sum_{D_i=0} Y_i(0) \right\}$$

$$= \left\{ \frac{1}{N_1} \sum_{D_i=1} \{Y_i(1) - Y_i(0)\} - \frac{1}{N_0} \sum_{D_i=0} \{Y_i(1) - Y_i(0)\} \right\},$$

where N_1 and N_0 are the number of treated and control units respectively.

Regression Estimator

This implies a fully saturated regression model with a two-way interaction:

$$E[Y_{igt}|g,t] = \beta_0 + \beta_1 \text{Treated}_g + \beta_2 \text{Post}_t + \beta_3 (\text{Treated}_g \times \text{Post}_t)$$

The subscripts are conventions to indicate the level of variation (i = Individual, g = Group, t = Time).

Difference-in-Differences estimator

This implies a fully saturated regression model with a two-way interaction:

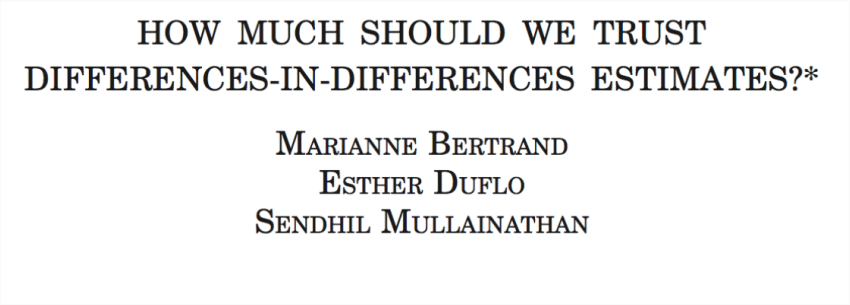
$$E[Y_{igt}|g,t] = \beta_0 + \beta_1 \text{Treated}_g + \beta_2 \text{Post}_t + \beta_3 (\text{Treated}_g \times \text{Post}_t)$$

The subscripts are conventions to indicate the level of variation (g = Group, t = Time). This translates to our two-by-two table as follows:

	Post-Period (T=1)	**Pre-Period (T=0)**	**Pre/Post Diff.**
Treated D=1	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$	$\beta_2 + \beta_3$
Control D=0	$\beta_0 + \beta_2$	β_0	β_2
Treated/Control Diff.	$\beta_1 + \beta_3$	β_1	$\boldsymbol{\beta_3}$

β_3 is the diff-in-diff estimate.

Inference



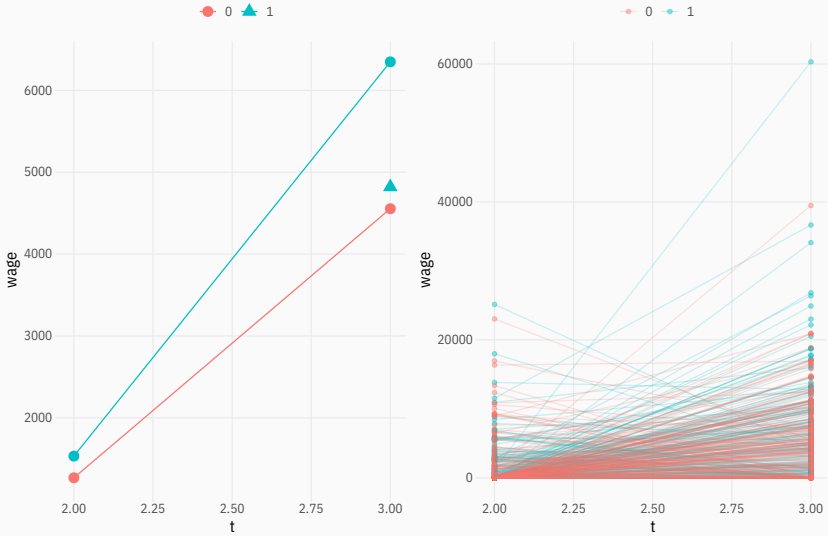
HOW MUCH SHOULD WE TRUST DIFFERENCES-IN-DIFFERENCES ESTIMATES?*

MARIANNE BERTRAND
ESTHER DUFLO
SENDHIL MULLAINATHAN

Standard advice: Cluster at least the unit level. More on this later.

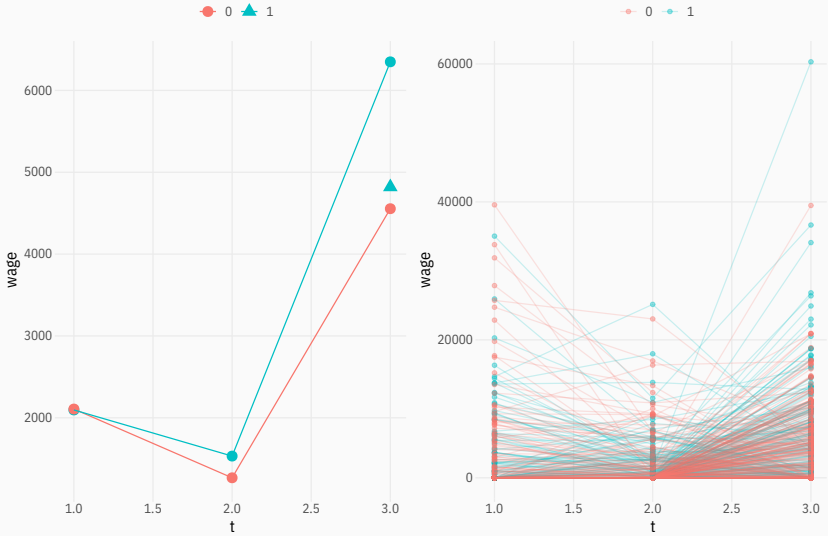
Lalonde Experimental Sample: 2 time periods

Lalonde: Experimental



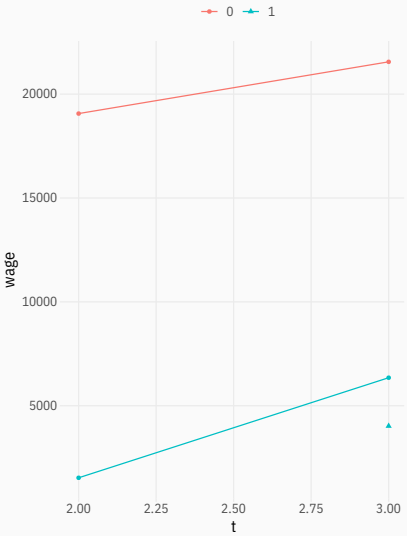
Lalonde Experimental Sample: 3 time periods - Ashenfelter Dip

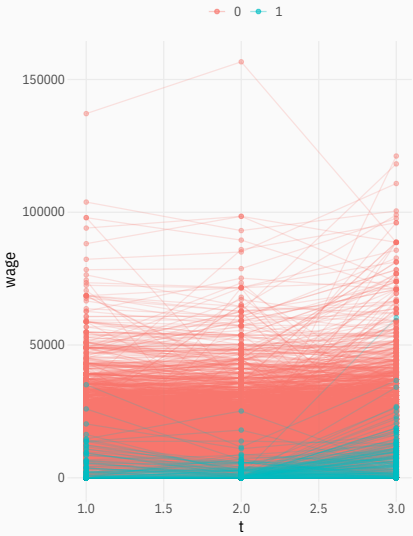
Lalonde: Experimental



Lalonde PSID: 2 time periods

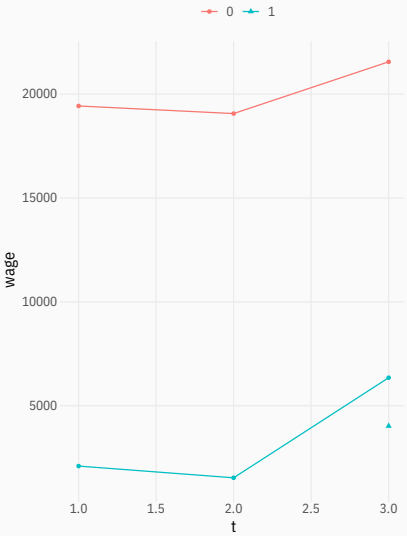
Lalonde: Observational

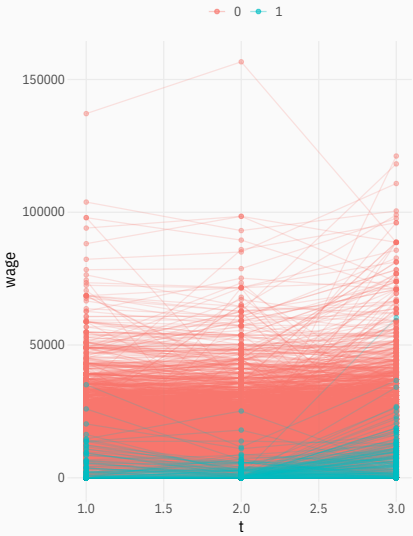




Lalonde PSID: 3 time periods

Lalonde: Observational





Roadmap

SOO Loose-ends

Difference-in-differences

General Exam Advice

Generalising to multiple time periods: Fixed Effects

Synthetic Control Methods

- Identify the relevant counterfactual
- What assumptions can we use to impute that counterfactual? Are those assumptions plausible?
- When writing identification results:
  - Start with observables (i.e., $E[Y_i \mid D_i = 1]$, not $E[Y_{1i} \mid D_i = 1]$)
  - Manipulate the expressions to get to causal estimands
  - Explicitly note whenever you invoke an assumption. E.g., don't simply write $E[Y_{1i} \mid X_i, D_i = 1] = E[Y_{1i} \mid X_i]$ without explaining what assumption justifies it.

- Identify the relevant counterfactual
- What assumptions can we use to impute that counterfactual? Are those assumptions plausible?
- When writing identification results:
  - Start with observables (i.e., $E[Y_i \mid D_i = 1]$, not $E[Y_{1i} \mid D_i = 1]$)
  - Manipulate the expressions to get to causal estimands
  - Explicitly note whenever you invoke an assumption. E.g., don't simply write $E[Y_{1i} \mid X_i, D_i = 1] = E[Y_{1i} \mid X_i]$ without explaining what assumption justifies it.
- When showing properties of estimators:
  - First write down an expression for the estimator
  - Then apply expectation, variance, etc.
  - Get to some population quantity, then your identification results can kick in

Review Topics

- Power: intuition and calculation
- SOO assumptions
- OLS estimator
- Sensitivity analysis

Roadmap

SOO Loose-ends

Difference-in-differences

General Exam Advice

Generalising to multiple time periods: Fixed Effects

Synthetic Control Methods

Fixed Effects Regressions

- We often have access to panel data, wherein each individual $i \in \{1, \ldots, N\}$ is observed for $T \geq 2$ time periods
- Stipulate following potential outcomes $Y_{it}^{(d)}$
  - $\mathbb{E}\left[Y_{it}^0 | \alpha_i, t, D_{it}\right] = \alpha_i + \lambda_t$
    - α_i is a unit fixed-effect: each individual has an intercept α_i - absorbs time-invariant unit-specific confounders
    - λ_t is a time fixed-effect: each time period has an intercept λ_t - absorbs unit-invariant time-specific confounders
  - Suppose D_{it} is as-good as randomly assigned conditional on α_i
  - Stipulate constant, additive effect of treatment. Then, $\mathbb{E}\left[Y_{it}^1\right] = \mathbb{E}\left[Y_{it}^0\right] + \tau$
- This motivates the popular **two-way fixed-effects** regression

$$Y_{it} = \tau D_i + \alpha_i + \gamma_t + \varepsilon_{it}$$

Within Estimator

- With large datasets, estimating individual α_is can involve inverting a very large matrix
  - With short panels, the estimates of α_is are inconsistent anyway := incidental parameters problem (Neyman-Scott)
- Instead, we can use Frisch-Waugh-Lovell (again!) and partial out FEs
- Calculate individual averages of the 2wFE equation

$$\overline{Y}_i = \alpha_i + \overline{\lambda} + \tau \overline{D}_i + \overline{\varepsilon}$$

Within Estimator

- With large datasets, estimating individual α_is can involve inverting a very large matrix
  - With short panels, the estimates of α_is are inconsistent anyway := incidental parameters problem (Neyman-Scott)
- Instead, we can use Frisch-Waugh-Lovell (again!) and partial out FEs
- Calculate individual averages of the 2wFE equation

$$\overline{Y}_i = \alpha_i + \overline{\lambda} + \tau \overline{D}_i + \overline{\varepsilon}$$

- Subtract this from the FE equation

$$Y_{it} - \overline{Y}_i = \lambda_t - \overline{\lambda} + \tau(D_{it} - \overline{D}_i) + (\varepsilon_{it} - \overline{\varepsilon})$$

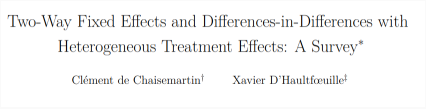
Staggered Adoption, Treatment Reversals, and other complications

- Recall that we stipulated a **constant, additive treatment effect** and **treatment timing as-good-as-random (conditional on FEs)**
- Last 5 years of methods literature on panel data studies what happens when we relax these parametric assumptions
- Weird weights redux: 2WFE no longer consistent for ATT


What's Trending in Difference-in-Differences?
A Synthesis of the Recent Econometrics Literature

Jonathan Roth* Pedro H. C. Sant'Anna† Alyssa Bilinski‡ John Poe§

January 3, 2022


Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey*

Clément de Chaisemartin† Xavier D'Haultfœuille‡

- review paper 1
- review paper 2

Roadmap

SOO Loose-ends

Difference-in-differences

General Exam Advice

Generalising to multiple time periods: Fixed Effects

Synthetic Control Methods

Synthetic Control: Motivation

- We're sometimes interested in estimating a treatment effect where only a single unit is treated.
- In these cases, it's really important to estimate a good counterfactual for that particular unit – not just on average as in the matching or traditional diff-in-diff cases.
- The synthetic control method introduced by Abadie, Diamond, and Hainmueller (2010) is useful when there are many pre-treatment outcome observations, and perhaps relatively few untreated units.
- The intuition is to create a "synthetic control" that is a weighted average of control units. We pick the weights so that the pre-treatment outcome of the synthetic control looks similar to the pre-treatment outcomes of the treated unit.

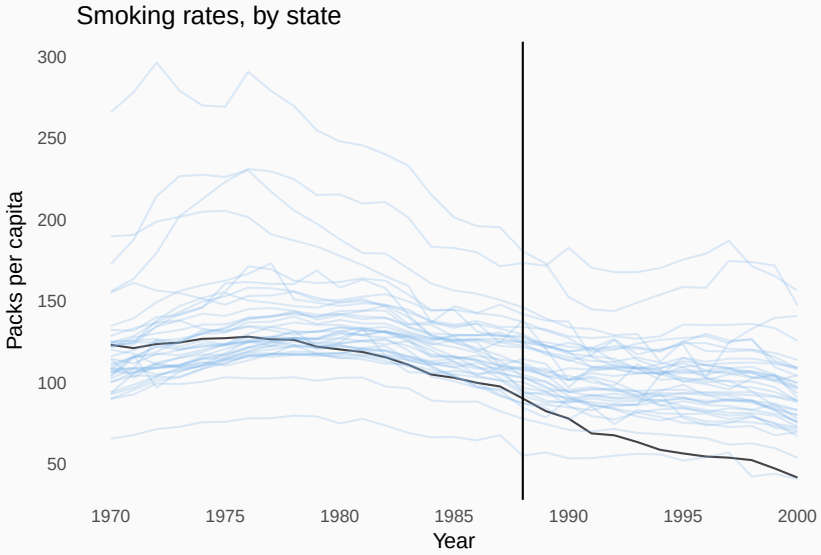
Smoking Data

ADH (2010) study the effect on smoking rates of an increase in the tobacco tax in California in 1988.

```
load("synth.rdata")
head(synth.long[c(1:3, 1206:1209), c("statename", "year", "smoking")])
```

```
##          statename year smoking
## 1          Alabama 1970    89.8
## 2         Arkansas 1970   100.3
## 3         Colorado 1970   124.8
## 1206 West Virginia 2000   107.9
## 1207     Wisconsin 2000    80.1
## 1208       Wyoming 2000    90.5
```

Smoking Data



Potential Strategies

The problem is to impute what California's level of smoking would have been in the post-treatment period if Proposition 99 hadn't been passed.

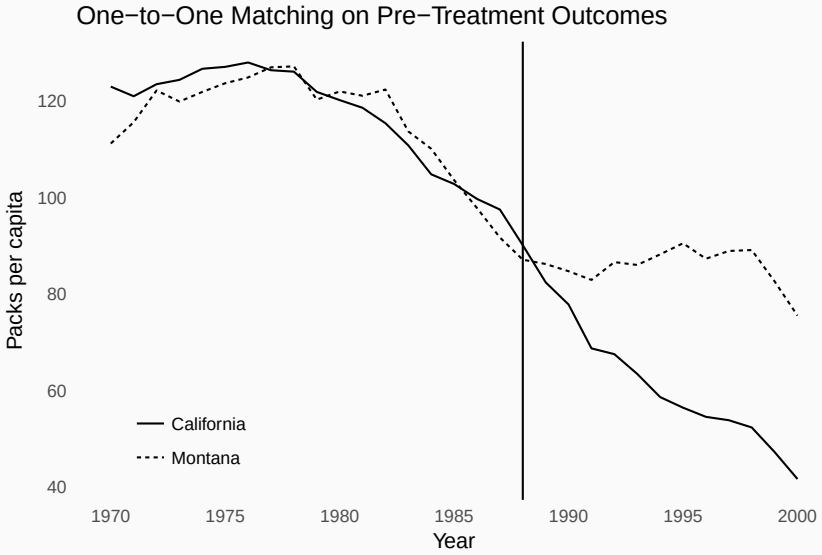
Potential strategies:

- Matching on pre-treatment covariates (possibly including lagged outcomes).
- Difference-in-differences.
- Regression.
- Synthetic control.

Matching on Pre-Treatment Outcomes

```
genmatch = GenHatch(synth$statename == "California",
                    synth[, paste0("smoking_", 1970:1988)])
matched.unit = genmatch$matches[,2]
match.statename = synth[matched.unit, statename]
out = Match(Y = synth$smoking_1996, Tr = synth$statename == "California",
            X =  synth[, paste0("smoking_", 1970:1988), with=F],
            Weight.matrix = genmatch, estimand = "ATT")
```

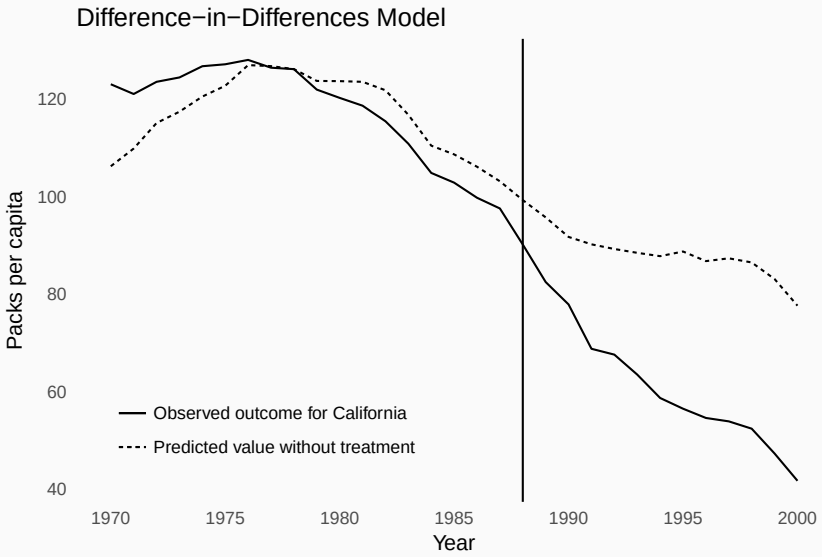
Matching on Pre-Treatment Outcomes



Difference-in-Differences

```
synth.long = synth.long %>%
  mutate(treatment = statename == 'California' & year ≥ 1989)
did = lm(smoking ~ treatment + statename + factor(year), synth.long)
calif = synth.long[synth.long$statename == 'California', ]
calif$treatment = F
counterfactuals = predict(did, newdata = calif)
```

Difference-in-Differences



Synthetic Control Method

Synthetic control setup: suppose unit $i = 0$ is treated after time T_0, and units $i = 1, \ldots, n$ are never treated. We observe outcomes for time periods $t = 1 < T_0 \leq T$. In potential outcomes notation, we observe:

$$
\begin{aligned}
&Y_{0,t}(0) \quad \text{for } t = 1, \ldots, T_0 - 1 \\
&Y_{0,t}(1) \quad \text{for } t = T_0, \ldots, T \\
&Y_{i,t}(0) \quad \text{for } t = 1, \ldots, T \text{ and } i = 1, \ldots, n
\end{aligned}
$$

We want to estimate $Y_{0,t}(1) - Y_{0,t}(0)$ for time periods $t \geq T_0$, but we can't observe $Y_{0,t}(0)$ after T_0.

What We Observe

The matrix we observe is:

$$\mathbf{Y}^{obs} = \begin{pmatrix} \mathbf{Y}^{obs}_{treat,pre} & \mathbf{Y}^{obs}_{control,pre} \\ \mathbf{Y}^{obs}_{treat,post} & \mathbf{Y}^{obs}_{control,post} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{treat,pre}(0) & \mathbf{Y}_{control,pre}(0) \\ \mathbf{Y}_{treat,post}(1) & \mathbf{Y}_{control,post}(0) \end{pmatrix}$$

What We Observe

The matrix we observe is:

$$\mathbf{Y}^{obs} = \begin{pmatrix} \mathbf{Y}^{obs}_{treat,pre} & \mathbf{Y}^{obs}_{control,pre} \\ \mathbf{Y}^{obs}_{treat,post} & \mathbf{Y}^{obs}_{control,post} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{treat,pre}(0) & \mathbf{Y}_{control,pre}(0) \\ \mathbf{Y}_{treat,post}(1) & \mathbf{Y}_{control,post}(0) \end{pmatrix}$$

To estimate the ATT we need $\mathbf{Y}_{treat,post}(0)$. We only observe:

$$\mathbf{Y}(0) = \begin{pmatrix} \mathbf{Y}_{treat,pre}(0) & \mathbf{Y}_{control,pre}(0) \\ ? & \mathbf{Y}_{control,post}(0) \end{pmatrix}$$

Synthetic Control Method

The synthetic control estimator imputes $Y_{0,t}(0)$ as a weighted average of the observed outcomes for the control units (plus possibly an intercept shift):

$$\hat{Y}_{0,t}(0) = \mu + \sum_{i=1}^{n} \omega_i Y_{i,t}$$

for some weight vector $\omega = (\omega_1, \ldots, \omega_n)$ and intercept μ. The parameters (μ, ω) define the "synthetic control" for the treated unit.

Synthetic Control Method

The synthetic control estimator imputes $Y_{0,t}(0)$ as a weighted average of the observed outcomes for the control units (plus possibly an intercept shift):

$$\hat{Y}_{0,t}(0) = \mu + \sum_{i=1}^{n} \omega_i Y_{i,t}$$

for some weight vector $\omega = (\omega_1, \ldots, \omega_n)$ and intercept μ. The parameters (μ, ω) define the "synthetic control" for the treated unit.

The parameters are typically picked by minimizing the squared distance between the synthetic control's pre-treatment outcomes and the treatment unit's pret-treatment outcomes:

$$\omega^* = \underset{(\mu,\omega_1,\ldots,\omega_n)}{\arg\min} \sum_{t=1}^{T_0-1} \left(Y_{0,t} - \mu - \sum_{i=1}^{n} \omega_i Y_{i,t} \right)^2$$

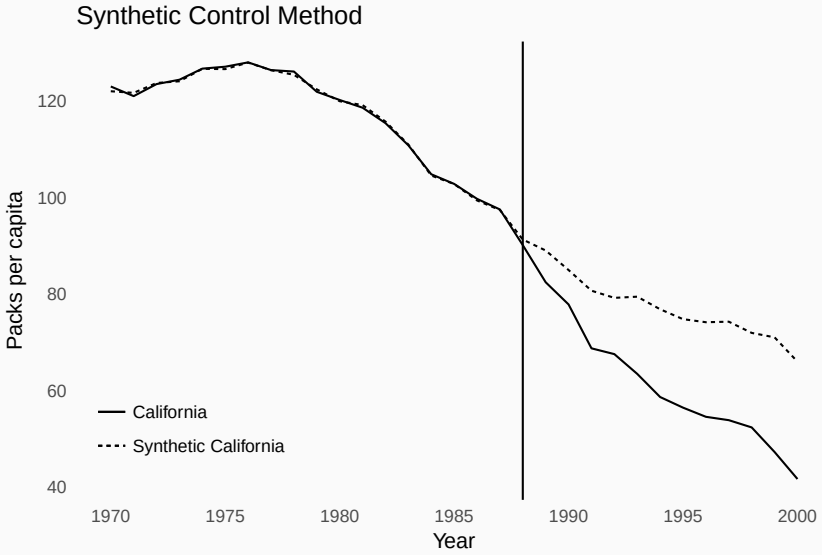
The natural method would be to estimate (μ, ω) using OLS, but if $n > T_0 - 1$ it requires additional constraints (e.g., weights summing to 1 or regularization).

Synthetic Control Method: Estimation Using glmnet

```r
library(glmnet)
control.pre = synth[synth$state≠3, paste0("smoking_",1970:1988)] %>%
  as.matrix %>% t
control.full = synth[synth$state≠3, paste0("smoking_",1970:2000)] %>%
  as.matrix %>% t
treat.pre = synth[synth$state==3, paste0("smoking_",1970:1988)] %>%
  as.matrix %>% t

# estimate weights and intercept using LASSO
weightsout = cv.glmnet(x = control.pre, y = treat.pre)
predictions = predict(weightsout, newx = control.full, s = "lambda.min")
calif$scm.pred = as.numeric(predictions)
```

Synthetic Control Method: Results



Synthetic Control Method: Uncertainty

How to assess uncertainty with synthetic control methods? There's only one treated unit so asymptotics are not helpful. Instead, ADH suggest a procedure like Bertrand, Duflo, and Mullainathan's placebo laws to assess what the null distribution of the SCM estimator looks like.