Chapter 3: Parallel Trends

Peter Hull

Applied Econometrics II
Brown University
Spring 2024

Outline

Panel Data and Beyond

Often, our data span multiple "fixed" dimensions: e.g. time, geography...

- Regression can partial multiple fixed effects out of a treatment variable that varies across a combination of dimensions
- Intuitively, this isolates variation in treatment "changes" which may be more clean ... but the devil is often in the details

Our agenda for this chapter is threefold:

1. Develop a general "model-based" framework for identification in multiple-FE regressions
2. Link this to familiar (canonical) difference-in-difference strategies
3. Illustrate practical takeaways of the framework

An Aside: Why Are Multi-Way FE Different?

Consider a two-way fixed effect (FE) regression with a binary treatment:

$$Y_{it} = \beta D_{it} + \alpha_i + \tau_t + \varepsilon_{it}$$

Can a selection-on-observables argument justify this specification?

- What would it mean for D_{it} to be as-if-randomly-assigned given (i, t)?
- Can the propensity score be additive in i and t?

A different justification/interpretation seems needed...

Two-Way Fixed Effects as "Model-Based" Identification

Suppose we have standard panel data for a set of units i and time periods t

- Assume again a constant-effects causal model: $Y_{it} = \beta D_{it} + \varepsilon_{it}$
- Let X_{it} be a vector of unit and time indicators

Rather than making an assumption about D_{it} being as-good-as-randomly assigned ("design-based"), let's assume it is a *deterministic function* of X_{it}

- Eg. Treatment $D_{it} \in \{0,1\}$ flips on for certain states i and periods t. Once I know the state and the period, I know the treatment status

For identification, we adopt a *model* for ε_{it}. Namely, $E[\varepsilon_{it} \mid X_{it}] = \alpha_i + \tau_t$

- Implies $E[Y_{it} \mid D_{it}, X_{it}] = \beta D_{it} + \alpha_i + \tau_t$, which can be fit by regression

Can clearly be extended to multiple FE, repeated cross sections, time-varying controls, unit-specific trends, or any other model for $E[\varepsilon \mid X]$

Canonical DiD

A model like $E[\varepsilon_{it} \mid X_{it}] = \alpha_i + \tau_t$ can usually be understood as imposing some kind of *parallel trends* assumption on untreated potential outcomes

- This is most easily seen in the canonical DiD case, with two time periods and a binary D_{it} where $D_{i1} = 0$ for $i = 1, 2$

Useful fact: β in a two-period FE regression of $Y_{it} = \beta D_{it} + \alpha_i + \tau_t + \nu_{it}$ is the same as the β in a first-differenced regression of $\Delta Y_i = \tau + \beta \Delta D_i + \nu_i$

- Here $\Delta Y_i = Y_{i2} - Y_{i1}$, $\Delta D_i = D_{i2} - D_{i1}$, $\tau = \tau_2 - \tau_1$, and $\nu_i - \nu_{i2} - \nu_{i1}$
- Straightforward to prove (see MHE); also true in the sample / for OLS

In the canonical DiD case, $\Delta D_i = D_{i2} \in \{0, 1\}$. So we have a simple reg:

$$\Delta Y_i = \tau + \beta D_{i2} + \nu_i$$

The two-way FE coef is $\beta = E[Y_{i2} - Y_{i1} \mid D_{i2} = 1] - E[Y_{i2} - Y_{i1} \mid D_{i2} = 0]$

Canonical DiD (Cont.)

We can actually study this estimand in a more general (heterogeneous-effects) model: $Y_{it} = Y_{it}(0)(1 - D_{it}) + Y_{it}(1)D_{it}$

The two-way FE coefficient is:

$$\begin{aligned}\beta =& E[Y_{i2} - Y_{i1} \mid D_{i2} = 1] - E[Y_{i2} - Y_{i1} \mid D_{i2} = 0] \\ =& E[Y_{i2}(1) - Y_{i1}(0) \mid D_{i2} = 1] - E[Y_{i2}(0) - Y_{i1}(0) \mid D_{i2} = 0] \\ =& \underbrace{E[Y_{i2}(1) - Y_{i1}(0) \mid D_{i2} = 1] - E[Y_{i2}(0) - Y_{i1}(0) \mid D_{i2} = 1]}_{\equiv \phi} \\ &+ \underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_{i2} = 1] - E[Y_{i2}(0) - Y_{i1}(0) \mid D_{i2} = 0]}_{\equiv \psi}\end{aligned}$$

Where $\phi = E[Y_{i2}(1) - Y_{i2}(0) \mid D_{i2} = 1] = ATT$ and $\psi = 0$ under a parallel trends assumption:

$$E[Y_{i2}(0) - Y_{i1}(0) \mid D_{i2} = 1] = E[Y_{i2}(0) - Y_{i1}(0) \mid D_{i2} = 0]$$

In words: *If not for the treatment, potential outcomes in the $D_{i2} = 1$ group would have followed the same trend as we observe the $D_{i2} = 0$ group*

Illustration: Card and Krueger

On April 1, 1992, New Jersey raised its state minimum wage from \$4.25 to \$5.05. The min wage in nearby Pennsylvania stayed at \$4.25 until 1996

- Card & Krueger '94, working at NJ's Princeton, studied the change in low-wage employment in NJ to the corresponding change in PA

Key assumption: if not for the minimum wage change, low-wage employment in NJ would have trended similarly as it did in PA

- CK surveyed fast food restaurants (innovative!), and famously found a *positive* effect from the DiD
- This was ... uh ... very controversial

"The inverse relationship between quantity demanded and price is the core proposition in economic science, which embodies the presupposition that human choice behavior is sufficiently rational to allow predictions to be made. Just as no physicist would claim that "water runs uphill," no self-respecting economist would claim that increases in the minimum wage increase employment. Such a claim, if seriously advanced, becomes equivalent to a denial that there is even minimal scientific content in economics, and that, in consequence, economists can do nothing but write as advocates for ideological interests. Fortunately, only a handful of economists are willing to throw over the teaching of two centuries; we have not yet become a bevy of camp-following whores."

~**James M. Buchanan**, 1986 Nobel laureate in economics, writing in the Wall Street Journal on April 25, 1996

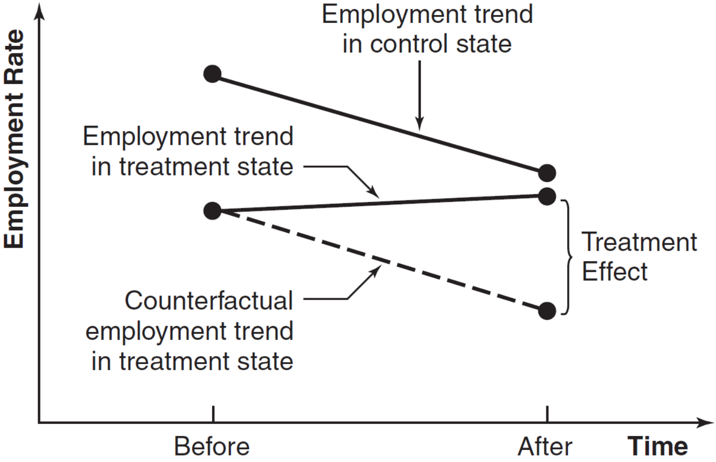
Card and Krueger in a Nutshell

Average employment in fast food restaurants before and after the New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (.51)	–2.89 (1.44)
2. FTE employment after, all available observations	21.17 (.94)	21.03 (.52)	–.14 (1.07)
3. Change in mean FTE employment	–2.16 (1.25)	.59 (.54)	2.76 (1.36)

Notes: Adapted from Card and Krueger (1994), table 3. The table reports average full-time-equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all restaurants with data on employment. Employment at six closed restaurants is set to zero. Employment at four temporarily closed restaurants is treated as missing. Standard errors are reported in parentheses.

Parallel Trends, Illustrated



Interrogating Parallel Trends

The core $E[\varepsilon_{it} \mid X_{it}] = \alpha_i + \tau_t$ assumption can be strong and finicky!

- It is inherently hard to "microfound" with a behavioral model of treatment selection (Ghanem, Sant'Anna, and Wuthrich '22)
- It is inherently sensitive to functional form (Roth and Sant'Anna '22)

People often indirectly validate parallel trends by checking pre-trends (more on this soon) but there are inherent limits to this (Roth '22):

- Pre-trend tests often have low power to reject lack of parallelism
- Conditioning a TWFE regression on a clean pre-trend can cause bias

With more / different FE, it can be hard (but important!) to assess the underlying "parallel trends" assumption

- Usually a good idea to start with "$T = 2$" and a binary treatment

Were CK's Trends Really Parallel?

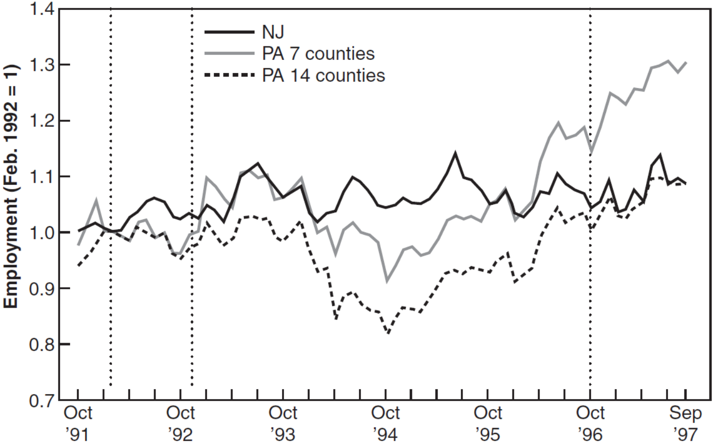


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

Outline

1. Multi-Way Fixed Effects✓
2. DiD and Regression
3. Application: Garthwaite et al. (2014)
4. Bonus Track 1: Relaxing Parallel Trends
5. Bonus Track 2: Synthetic Controls

Automating DiD

As we've seen, the canonical "2×2" DiD can be run by regression:

$$Y_{it} = \beta D_{it} + \alpha_i + \tau_t + \varepsilon_{it}$$

In CK's state-year panel, α_i is a state indicator & $D_{it} = 1$ in NJ post-1992

- If CK observed a panel of fast food workers i, they could have included individual FE (same β if nobody moved / quit)

Of course, running DiD by *reg* has the advantage of automating SEs on β

- With panel data and/or when treatment is "clustered" (e.g. by states) we need to take care with specifying SEs (more on this soon)

The regression toolbox also gives other advantages: generalizing the 2×2

Card (1992): Many States, Variably Treated

Card '92 makes the *federal* min into a DiD using differential exposure:

$$Y_{it} = \beta(E_i \times P_t) + \alpha_i + \tau_t + \varepsilon_{it}$$

where E_i is the fraction of the teen workforce earning $< \$3.80$ in state i and P_t is a post-1990 dummy, when the fed min went from \$3.35 to \$3.80

- Two periods, 1989+1990, and 51 units (all states + DC): $N = 102$

We can again use the first-difference equivalence result:

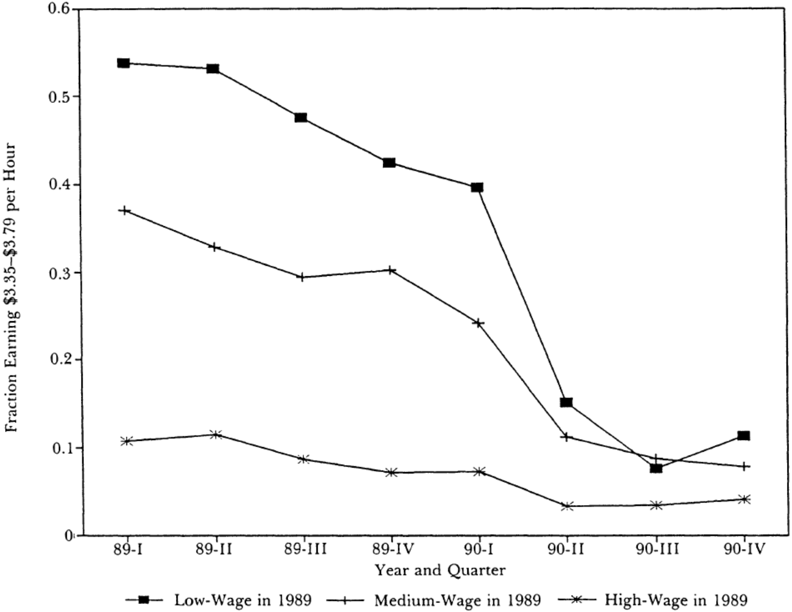
$$\Delta Y_i = \tau + \beta E_i + \varepsilon_i$$

where ΔY_i is the change in teen employment of state i, $1989 - 1990$

- So $\beta = Cov(E_i, \Delta Y_i)/Var(E_i)$, for now the non-binary E_i

See also Finkelstein (2007) – effects of federal Medicare rollout

Card (1992) Identifying Variation



Card (1992) Results

Regression DD estimates of minimum wage effects on teens, 1989 to 1990

	Change in Mean Log Wage		Change in Teen Employment-Population Ratio	
Explanatory Variable	(1)	(2)	(3)	(4)
1. Fraction of affected teens (FA_s)	.15 (.03)	.14 (.04)	.02 (.03)	−.01 (.03)
2. Change in overall emp./pop. ratio	—	.46 (.60)	—	1.24 (.60)
3. R^2	.30	.31	.01	.09

Notes: Adapted from Card (1992). The table reports estimates from a regression of the change in average teen employment by state on the fraction of teens affected by a change in the federal minimum wage in each state. Data are from the 1989 and 1990 CPS. Regressions are weighted by the CPS sample size for each state.

Adding Controls

Adding time-varying controls to (constant-effects) TWFE is easy: if $E[\varepsilon_{it} \mid X_{it}, W_{it}] = \alpha_i + \tau_t + W_{it}'\gamma$, $E[Y_{it} \mid X_{it}, W_{it}] = \beta D_{it} + \alpha_i + \tau_t + W_{it}'\gamma$

- Allow for violations of parallel trends through a common $W_{it}'\gamma$ term
- Abadie (2005) gives a more flexible (inverse-weighting) version of this

We might also want to allow parallel trends to hold conditional on fixed characteristics (e.g. within matched groups, allowing group-specific trends)

- To accomplish this, we just interact the fixed W_i with time FE:

$$Y_{it} = \beta D_{it} + \alpha_i + \tau_t + W_i'\gamma_t + \nu_{it}$$

- E.g. when W_i is a set of group indicators and $T = 2$, we have:

$$\Delta Y_i = \tau + \beta \Delta D_i + W_i'\gamma + \Delta \nu_i$$

The Angrist (1998) result then tells us β estimates a variance-weighted average of within-group DiD regressions, of ΔY_i on ΔD_i

Unit-Specific Trends

When $T \geq 3$, we can take a linear trend out of each unit's outcome:

$$Y_{it} = \beta D_{it} + \alpha_{0i} + \alpha_{1i} t + \tau_t + W_{it}' \gamma + \nu_{it}$$

This can be understood as isolating variation in "trend breaks"

- In a canonical DiD with two pre-periods, the $\alpha_{1i} t$ term differences off any differential pre-trend between treatment and control group
- Seems "weaker"...but is it really? More w/Rambachan-Roth '22 soon

In any event, this can be a useful "robustness check" (when it works)

- Besley and Burgess (2004): labor regulation leads to lower per capita in a conventional TWFE specification
- But adding state-specific trends kills this finding (bad pre-trends!)

Besley-Burgess '04: Regressions of Firm Output Per Capita

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	−.186	−.185	−.104	.0002
	(.064)	(.051)	(.039)	(.020)
Log development expenditure per capita		.240	.184	.241
		(.128)	(.119)	(.106)
Log installed electricity capacity per capita		.089	.082	.023
		(.061)	(.054)	(.033)
Log state population		.720	0.310	−1.419
		(.96)	(1.192)	(2.326)
Congress majority			−.0009	.020
			(.01)	(.010)
Hard left majority			−.050	−.007
			(.017)	(.009)
Janata majority			.008	−.020
			(.026)	(.033)
Regional majority			.006	.026
			(.009)	(.023)
State-specific trends	No	No	No	Yes
Adjusted R^2	.93	.93	.94	.95

Notes: Adapted from Besley and Burgess (2004), table IV. The table reports regression DD estimates of the effects of labor regulation on productivity. The dependent variable is log manufacturing output per capita. All models include state and year effects. Robust standard errors clustered at the state level are reported in parentheses. State amendments to the Industrial Disputes Act are coded 1 = pro-worker, 0 = neutral, −1 = pro-employer and then cumulated over the period to generate the labor regulation measure. Log of installed electrical capacity is measured in kilowatts, and log development expenditure is real per capita state spending on social and economic services. Congress, hard left, Janata, and regional majority are counts of the number of years for which these political groupings held a majority of the seats in the state legislatures. The data are for the sixteen main states for the period 1958–92.

Leads and Lags

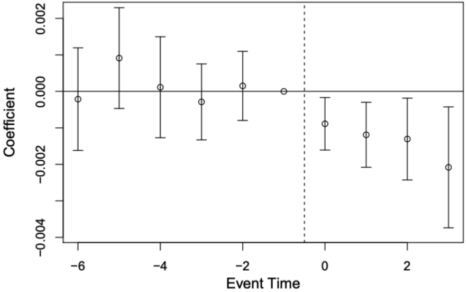
Of course if you have multiple time periods, a more transparent check is a so-called "dynamic" (or "event study") specification. E.g.:

$$Y_{it} = \sum_{k \neq -1} \beta_k (E_i \times P_{tk}) + \alpha_i + \tau_t + \varepsilon_{it}$$

where $D_i \in \{0,1\}$ indicates treatment receipt in periods $k \geq 0$ and $P_{tk} = \mathbf{1}[t = k]$ indicates pre- and post-treatment periods

- Plotting estimates of β_k and 95% gives a nice visual check:

Figure 2: Effect of the ACA Medicaid Expansions on Annual Mortality



Source: Miller et al. (2021)

Higher-Order DiDs

Another extension of the basic DiD setup is the so-called "triple-diff": e.g.

$$Y_{ijt} = \beta(E_i \times P_t \times G_j) + \alpha_{ij} + \tau_{jt} + \gamma_{it} + \nu_{ijt}$$

where $T = 2$, $E_i \in \{0,1\}$ indicates a treatment group (e.g. state), $P_t \in \{0,1\}$ is a post-period indicator, and $G_j \in \{0,1\}$ indicates a different dimension across which treatment exposure varies (e.g. demographics)

- We now control for all two-way interactions, $\alpha_{ij} + \tau_{jt} + \gamma_{it}$

Straightforward to show that the treatment coefficient corresponds to:

$$\begin{aligned}\beta =& E[\Delta Y_{ij} \mid E_i = 1, G_j = 1] - E[\Delta Y_{ij} \mid E_i = 0, G_j = 1] \\ &- E[\Delta Y_{ij} \mid E_i = 1, G_j = 0] - E[\Delta Y_{ij} \mid E_i = 0, G_j = 0]\end{aligned}$$

- If parallel trends holds within groups, β identifies a difference in ATTs
- In fact, PT can fail as long as the failure is the same across groups!

DiD vs. LDV

Since basic DiD can be (and sometimes is!) run in first-differences, sometimes folks think it's nested by a lagged dependent variable spec:

$$Y_{it} = \alpha + \beta D_{it} + \delta Y_{i,t-1} + \nu_{it}$$

where we impose (but can relax) $\delta = 1$... this is not correct

- LDV is motivated by a selection-on-observables assumption: $E[Y_{it}(0) \mid D_{it}, Y_{i,t-1}] = E[Y_{it}(0) \mid Y_{i,t-1}]$; DiD is not
- The LDV spec imposes $Cov(Y_{i,t-1}, \nu_{it}) = 0$; DiD does not

Beware of combining fixed effects and lagged dependent variables!

- The FD version $\Delta Y_{it} = \alpha + \beta \Delta D_{it} + \gamma \Delta Y_{i,t-1} + \Delta \nu_{it}$ likely has problems: $\Delta Y_{i,t-1}$ and $\Delta \nu_{it}$ are likely to be correlated (Nickell '81)
- Motivates dynamic panel IVs (e.g. Arellano-Bond '91)... not great!

Outline

1. Multi-Way Fixed Effects✓
2. DiD and Regression✓
3. Application: Garthwaite et al. (2014)
4. Bonus Track 1: Relaxing Parallel Trends
5. Bonus Track 2: Synthetic Controls

Motivation: Does Public Healthcare Reduce Employment?

Health insurance is tightly linked to employment in the U.S.

- How important is the insurance amenity for labor supply decisions?

In 2005, Tennessee (this guy again!) stopped its expansion of Medicaid (TennCare), causing 170,000 adults to lose coverage over three months

- Garthwaite et al. (2013) look at what happened to employment trends in a DiD and triple-diff framework
- Control groups: other Southern states and adults with children

Big policy motivation: the (then) recent passing of Obamacare (ACA)

- Related lit on private insurance "crowd-out" (Cutler & Gruber '96)

Institutional Setting and Data

TennCare was a bit unusual: "Medicaid for all"?

- Covered all individuals, regardless of income or demographics, deemed "uninsured" or "uninsurable" (i.e. previously denied private coverage)
- In 1995, 40% of enrollees had incomes of $>100\%$ the poverty line

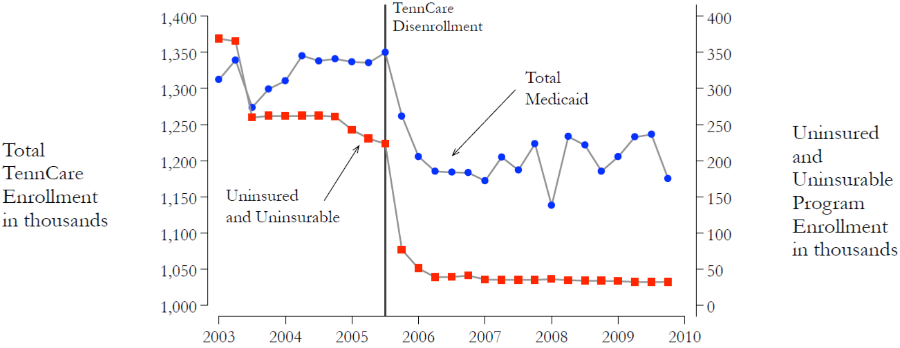
Following some minor eligibility adjustments in 2002-2004, TennCare stopped covering adults who didn't qualify for traditional Medicaid in 2005

- Affected around 4% of non-elderly TN adults for around three months
- Mostly affected childless adults (no longer eligible at any income level)

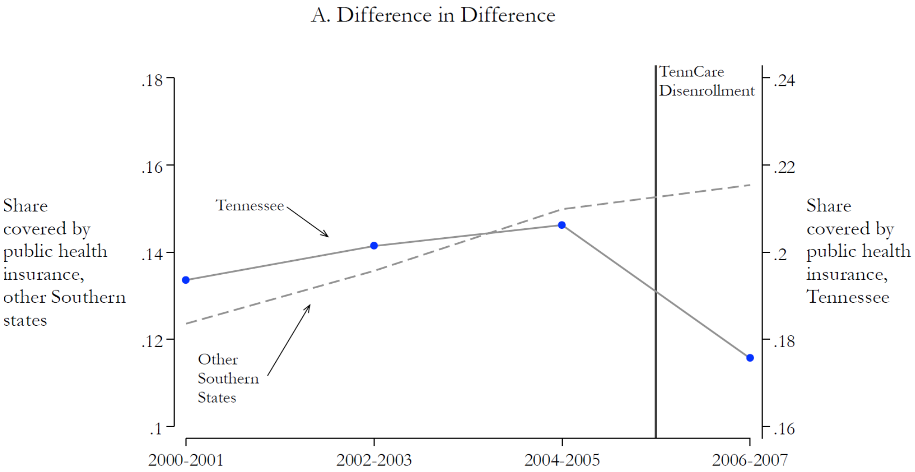
GGN use public data from the CPS and two simple TWFE designs

- Potential issue: measurement error in self-reported coverage status
- More subtle issue: state clustering with 17 Southern states

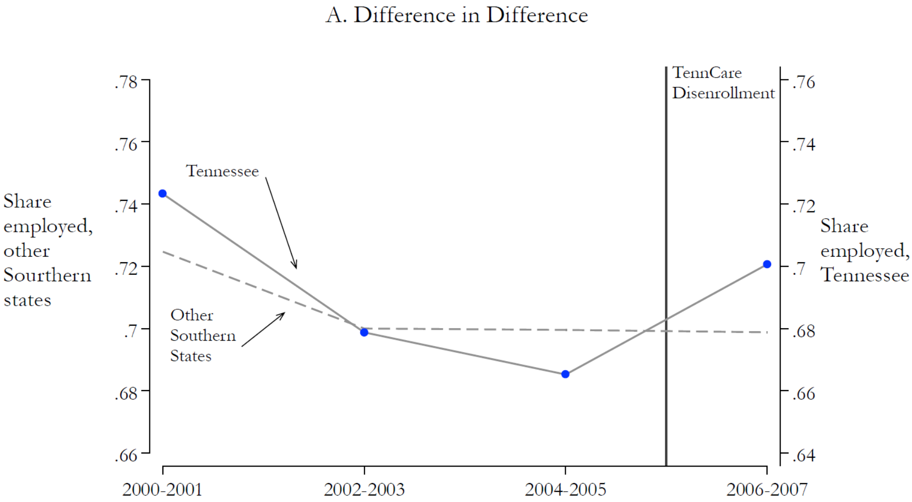
Coverage Trends in Tennessee



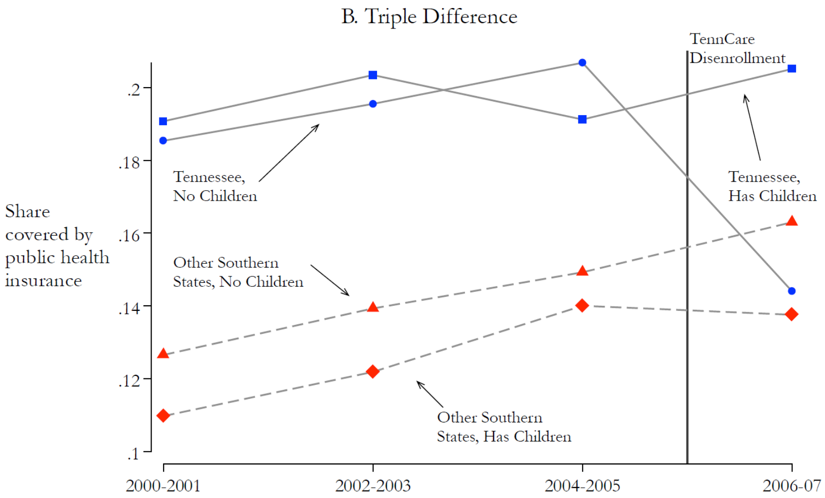
DiD: Insurance Coverage



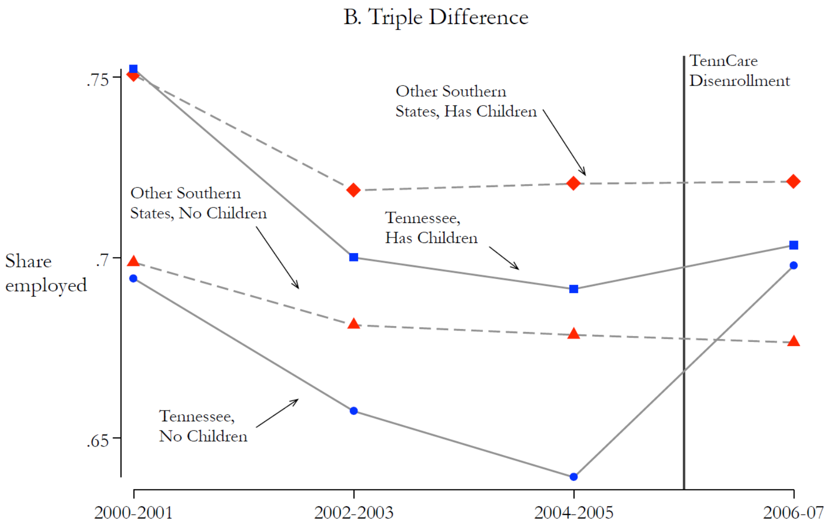
DiD: Employment Rates



Triple-Diff: Insurance Coverage



Triple-Diff: Employment Rates



Main DiD & Triple-Diff Estimates

Table II. The Effect of TennCare Disenrollment on Employment
Dependent Variable: The share of CPS respondents reporting the given outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Has Public Health Insurance	Employed	Employed and Working <20 hours per week	Employed and Working ≥20 hours per week	Employed and Working 20-35 hours per week	Employed and Working ≥35 hours per week	Employed with Private Insurance through Employer
			A. Difference-in-Difference Estimates				
Tennessee × Post 2005	- 0.046	0.025	- 0.001	0.026	0.001	0.025	0.009
	(0.010)	(0.011)	(0.004)	(0.010)	(0.007)	(0.011)	(0.013)
	[0.000]	[0.038]	[0.758]	[0.023]	[0.906]	[0.041]	[0.507]
R^2	0.871	0.867	0.392	0.847	0.418	0.819	0.911
			B. Triple-Difference Estimates				
Tennessee × Post 2005 × No Children	- 0.073	0.046	0.002	0.044	0.018	0.026	0.042
	(0.017)	(0.020)	(0.009)	(0.020)	(0.013)	(0.021)	(0.023)
	[0.001]	[0.032]	[0.843]	[0.042]	[0.195]	[0.236]	[0.084]
R^2	0.952	0.941	0.665	0.931	0.824	0.918	0.942
Mean of dep. variable	0.139	0.705	0.037	0.668	0.097	0.572	0.515

Notes: The sample includes the 17 southern states between 2000 through 2007. For Panel A, N = 136; the sample consists of state-by-year means; state and year fixed effects are included, but not shown. For Panel B, N = 272; the sample consists of means for each state, year, and childless status; state fixed effects, year fixed effects, childless fixed effects, and fixed effects for all possible pairwise interactions are included but not shown. The standard errors in parentheses are modified block bootstrap standard errors that are computed using the following two-stage re-sampling procedure: (1) states are drawn with replacement, and (2) individuals are drawn with replacement within states (resampling independently for state clusters chosen more than once). These standard errors are robust to autocorrelation between observations from the same state and explicitly account for sampling error in the state-by-year means (or state-by-year-by-childless-status means in Panel B). The associated p-values in brackets are based on two-tailed t-test with 16 degrees of freedom.

External Validation

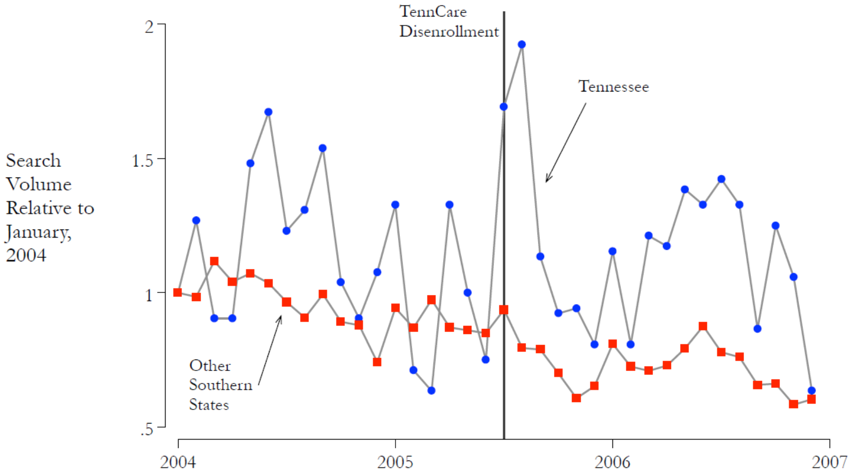


Figure VI
Searches on Google for Phrase "Job Openings"

Heterogeneity

	(1)	(2)	(3)	(4)	(5)
	Has Public Health Insurance	Employed	Employed and Working <20 hours per week	Employed and Working ≥20 hours per week	Employed with Private Insurance through Employer
			B. Heterogeneity by Education		
Triple-difference estimate for high school dropouts	- 0.289	0.125	0.029	0.096	0.087
	(0.057)	(0.054)	(0.024)	(0.056)	(0.049)
	[0.000]	[0.021]	[0.228]	[0.087]	[0.076]
Mean for high school dropouts	0.257	0.533	0.031	0.502	0.246
Triple-difference estimate for those with a high school diploma or more	- 0.034	0.034	- 0.004	0.038	0.036
	(0.017)	(0.023)	(0.009)	(0.023)	(0.025)
	[0.051]	[0.134]	[0.639]	[0.095]	[0.155]
Mean for high school graduates	0.118	0.736	0.038	0.698	0.563
p-value of test for equality across rows	[0.000]	[0.128]	[0.190]	[0.335]	[0.352]
R^2	0.948	0.956	0.584	0.951	0.979
			C. Heterogeneity by Health Status		
Triple-difference estimate for those who report excellent health	- 0.018	0.020	- 0.003	0.024	- 0.014
	(0.023)	(0.037)	(0.021)	(0.041)	(0.050)
	[0.439]	[0.583]	[0.876]	[0.570]	[0.776]
Mean for excellent health	0.065	0.791	0.040	0.750	0.608
Triple-difference estimate for those who report good or poor health	- 0.091	0.053	0.004	0.049	0.061
	(0.021)	(0.024)	(0.009)	(0.025)	(0.027)
	[0.000]	[0.028]	[0.668]	[0.051]	[0.025]
Mean for good or poor health	0.165	0.675	0.036	0.640	0.483
p-value of test for equality across rows	[0.020]	[0.445]	[0.746]	[0.588]	[0.194]
R^2	0.955	0.951	0.603	0.928	0.943

Summary

Two- (or more) way fixed effects designs are a highly tractable way of estimating effects from “big” (e.g. state-level) policy shocks

- Good to start with simple visualizations of the target variation
- Try to “strip down” complex regression specs into simple (i.e. binary) comparisons to figure out the core “parallel trends” assumptions

Check pre-trends when you can, but recognize their limits (more soon...)

- Try as hard as you can to avoid specification search!

Adding more fixed effects (etc) need not bring you closer to identification!

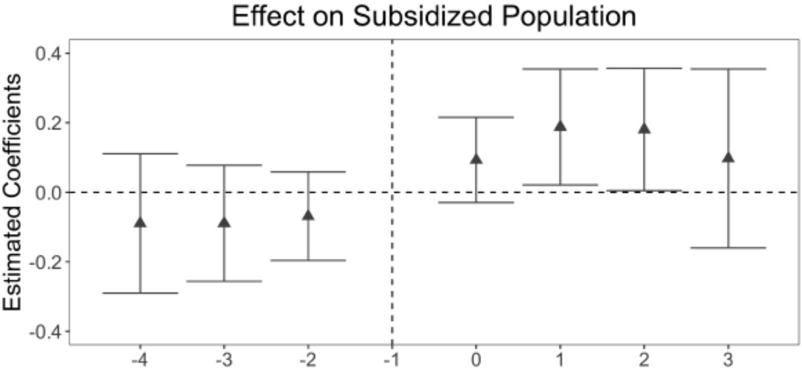
- Are unit-specific trends / triple-diff strategies really more compelling?
- Remember this is all about one (and only one) model for $E[\varepsilon_{it} \mid X_{it}]$

Outline

1. Multi-Way Fixed Effects✓
2. DiD and Regression✓
3. Application: Garthwaite et al. (2014)✓
4. Bonus Track 1: Relaxing Parallel Trends
5. Bonus Track 2: Synthetic Controls

Roth (2022): Pre-Trend Tests Can Have Low Power!

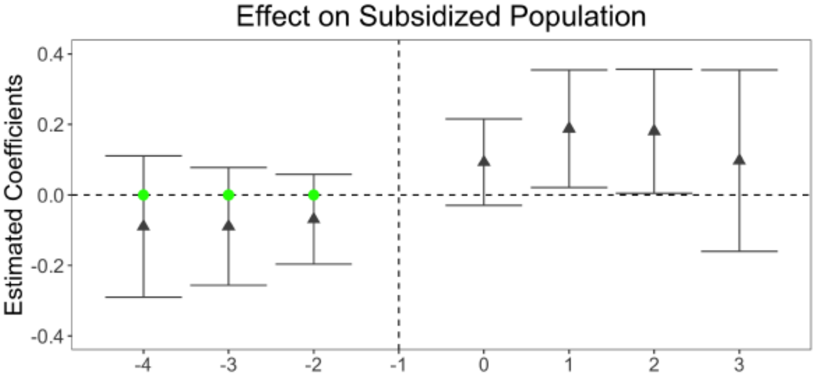
Consider a prototypical "event study" visualization of pre-trends:



(This specific one comes from He and Wang (2017), who study the impacts of placing college grads as village officials in China)

Roth (2022): Pre-Trend Tests Can Have Low Power!

The usual way people want you to interpret such graphs is like this:

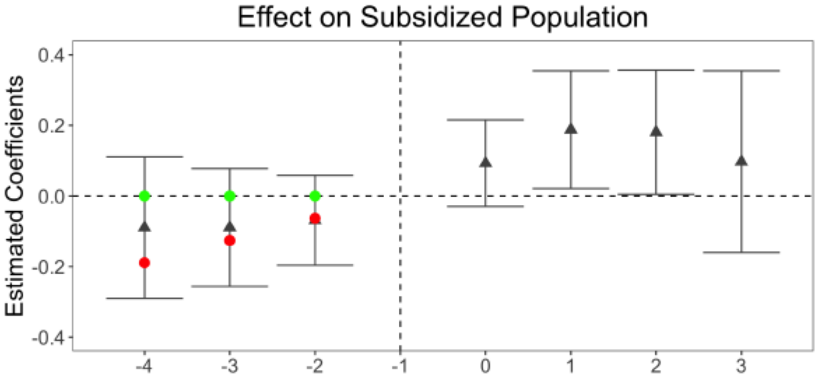


"Zero is in the pre-trend 95% CI's, so we can't reject parallel trends"

- Indeed, here the p-value for H_0: $\beta_{pre} =$ green dots is 0.81

Roth (2022): Pre-Trend Tests Can Have Low Power!

But note that we can also interpret the graph like this:

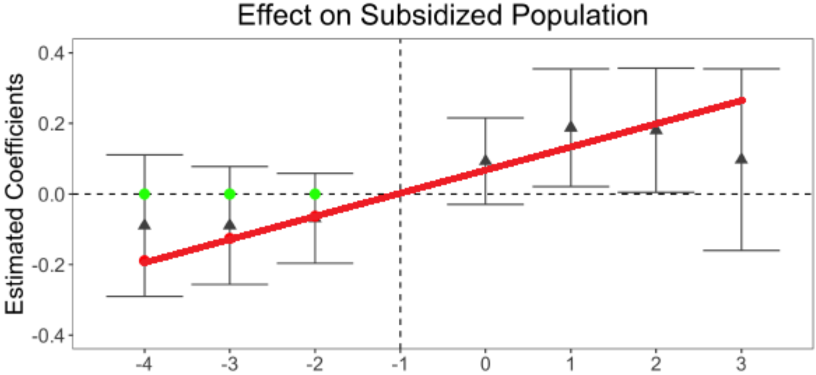


"Uh-oh! A sizable non-zero differential trend is in the pre-trend 95% CI's!"

- For H_0: $\beta_{pre} =$ red dots, we get the same p-value of 0.81

Roth (2022): Pre-Trend Tests Can Have Low Power!

Why uh-oh? If there is such a differential trend, we lose our post-effects

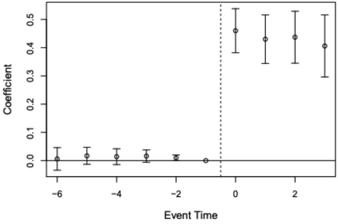


This is a way of visualizing the likely low power of pre-trend tests: there can be important alternative hypotheses that are hard to rule out!

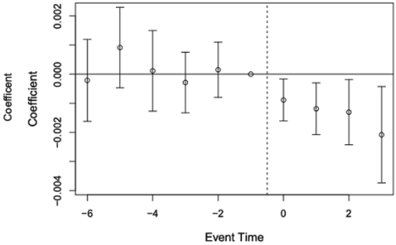
The "Straight Line Test"

One practical upshot from this: try to see if you can draw such lines!

- The more nonlinearity you need to "explain away" effects, the better



Outcome: Medicaid Eligibility



Outcome: Mortality

See Roth (2022) for more, and check out his *pretrends* Stata package

Rambachan and Roth (2022): Relaxing Parallel Trends

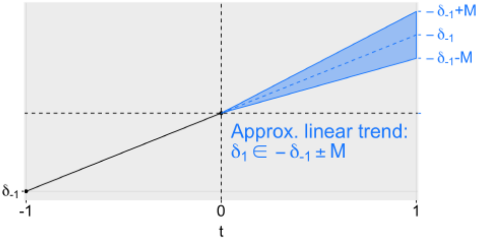
RR '02 go one step further, showing how we can form valid bounds on post-effects under a range of possible differential trends

- See also Manski and Pepper (2018), who apply a similar strategy

Formally, we can restrict possible values of δ_t^+ with estimates of δ_t^-, for:

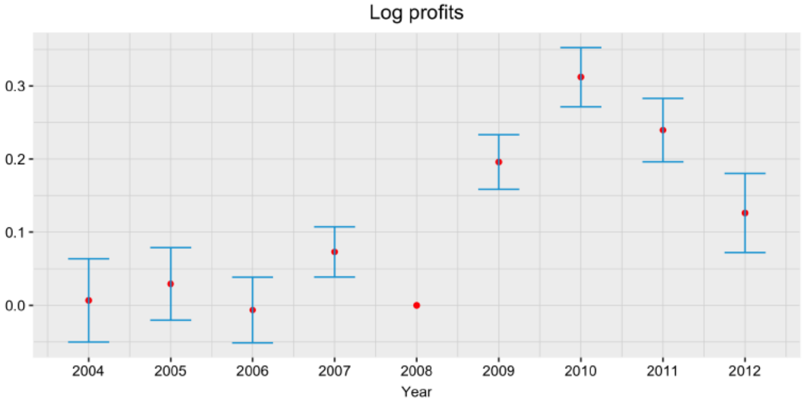
$$\delta_t^+ \equiv E[Y_{i,t}(0) - Y_{i,0}(0) \mid D_i = 1] - E[Y_{i,t}(0) - Y_{i,0}(0) \mid D_i = 0]$$
$$\delta_t^- \equiv E[Y_{i,-t}(0) - Y_{i,0}(0) \mid D_i = 1] - E[Y_{i,-t}(0) - Y_{i,0}(0) \mid D_i = 0]$$

- *Bounds on relative magnitudes*: require $|\delta_t^+| \leq M|\delta_t^-|$ for some $M > 0$
- *Smoothness restrictions*: bound deviations of δ_t^+ from a linear trend



Application: Benzarti and Carloni (2019)

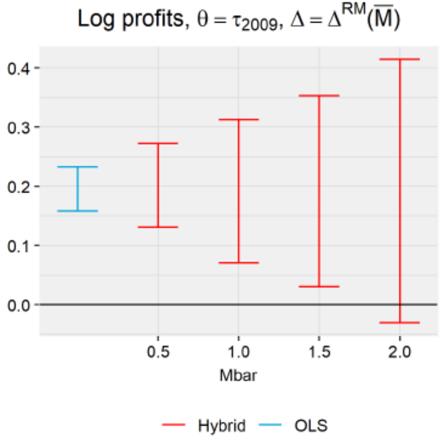
BC '19 study the incidence of a cut in the VAT on restaurants in France: from 19.6% to 5.5% in July 2009



Pretrends pass the "straight line test," but action around 2007 is troubling

Parallel Trends "Breakdown Point"

Using the *honestdid* package, we can find the largest M such that the "honest" 95% CI for ATT_1 excludes zero while requiring $|\delta_1^+| \leq M|\delta_1^-|$



We can rule out a null effect unless we allow for violations of parallel trends that are twice as large than the max in the pre-period!

Outline

1. Multi-Way Fixed Effects✓
2. DiD and Regression✓
3. Application: Garthwaite et al. (2014)✓
4. Bonus Track 1: Relaxing Parallel Trends✓
5. Bonus Track 2: Synthetic Controls

Abadie et al. (2010)

Synthetic controls use a weighted average of untreated unit outcomes to impute a counterfactual outcome for a treated united

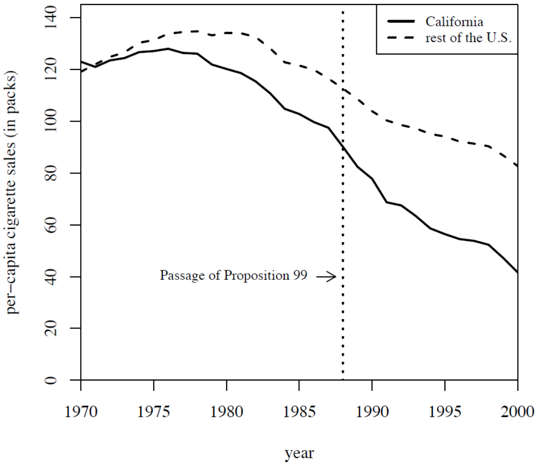
Basic version: observe Y_{it} for $i = 1, \ldots, J+1$ and $t = 1, \ldots, T$ where unit 1 gets treated in period $T_0 \in (1, T]$

- I.e., observe $Y_{1t}(1)$ for $t = T_0, \ldots, T$ and $Y_{jt}(0)$ $\forall$ j, $t = 1, \ldots, T_0 - 1$
- Idea: impute missing $Y_{1t}(0)$, for $t = T_0, \ldots, T$, from $Y_{jt}(0)$
- Estimator: $Y_{1t}(1) - \sum_{j=2}^{J+1} \omega_{jt} Y_{jt}$ for some $\omega_{jt} \geq 0$, $\sum_{j=2}^{J+1} \omega_{jt} = 1$
- Pick weights to minimize distance between Y_{1t} and $\sum_{j=1}^{J+1} \omega_{jt} Y_{jt}$ in the pre-period $(t = 1, \ldots, T_0 - 1)$

Seems fairly intuitive... but does it work?

Illustration: Prop 99

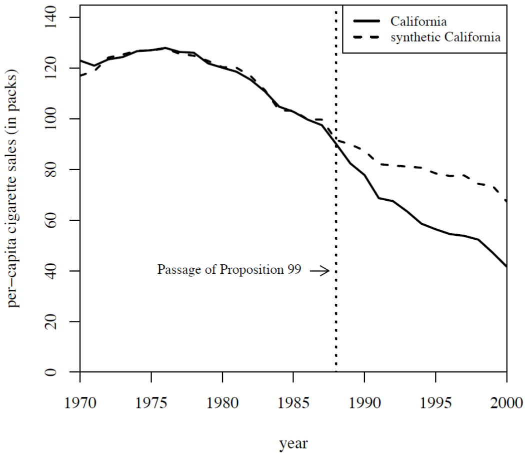
California passed a law in 1989 that raised the cigarette excise tax by 25 cents and implemented a large-scale antitobacco media campaign



Clearly a simple diff-in-diff is not going to cut it here!

Illustration: Synthetic California

Applying standard synthetic control weights renders a pre-trend we could previously only dream about!



What's going on under the hood?

Illustration: Synthetic California

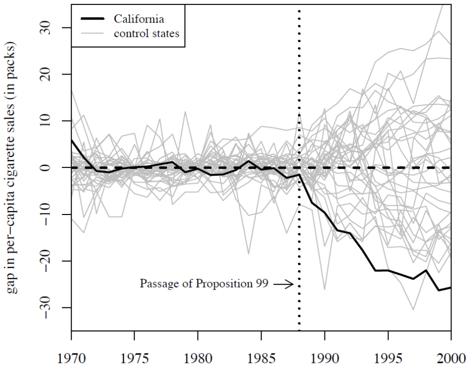
Standard synth weights are "sparse," by design

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	-	Nebraska	0
Arizona	-	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	-
Connecticut	0.069	New Mexico	0
Delaware	0	New York	-
District of Columbia	-	North Carolina	0
Florida	-	North Dakota	0
Georgia	0	Ohio	0
Hawaii	-	Oklahoma	0
Idaho	0	Oregon	-
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	-	Vermont	0
Massachusetts	-	Virginia	0
Michigan	-	Washington	-
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

How Do We Know This is Working?

In some of the original synth papers, Abadie and co. suggest permutation-based inference (note: asymptotics are tricky with one treated unit!)

- Randomly shuffle which state you construct counterfactuals for, and see if the estimated treatment effect is in the tails



But the theoretical justification for this thin (treatment is non-random!)

Current State of the Art: Synthetic Diff-in-Diff

Arkhangelsky et al. (2021), among others, note that conventional synth doesn't allow for unit fixed effects (like DiD does)

- i.e. no within-unit differencing, only across-unit matching

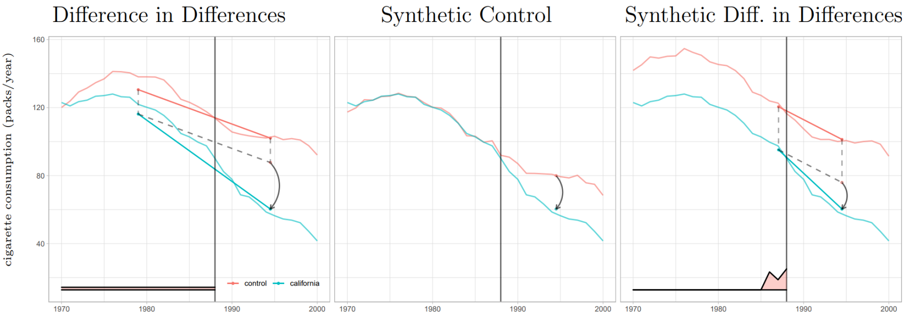
They propose a "hybrid" approach, which works either under parallel trends or when it is possible to reweight to a counterfactual a la synth

- As importantly, they derive valid large-sample asymptotics for this "synthetic diff-in-diff" estimator

I'll leave the technical details to self-study... suffice to say it works well!

- In practice, try the hot-off-the-presses *sdid* Stata package

Illustrating SDiD



Note: SDiD also lets you make the weights balance time-varying covariates

Practical Takeaways

As with the earlier coef. stability stuff, bounding effects to violations of PT can be a useful robustness check...but maybe not a main ID strategy

- Lots of researcher d.o.f. on e.g. how you specify violations

Synthetic controls (and related methods, e.g. "matrix completion") are widely used in industry, but haven't yet really cracked into top-5's

- I have some theories for why this is, but I'm curious to hear yours!
- Is it just that it seems "too easy" ?

Personally, I find any model-based ID strategy (no matter how technically fancy) a bit unsatisfying vs. a compelling design-based story...

- But coming up with compelling design is hard / not always possible!