Lecture 8: Difference-in-Differences and Extensions

Applied MicroEonometrics,Fall 2024

Zhaopeng Qu

Nanjing University Business School

November 13 2024

Difference in Differences

Introduction

Difference in Differences: Introduction

- DD(or DID) is a special case for “twoway fixed effects” under certain assumption, which is one of most popular research designs in applied microeconomics.
- It was introduced into economics via Orley Ashenfelter in the late 1970s and then popularized through his student David Card (with Alan Krueger) in the 1990s.

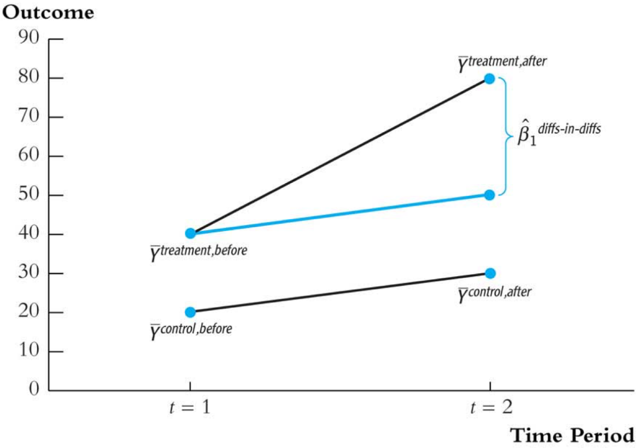
RCT and Difference in Differences

- A typical RCT design requires a causal studies to do as follow
  1. Randomly assignment of treatment to divide the population into a "treatment" group and a "control" group.
  2. Collecting the data at the time of post-treatment then comparing them.
- It works because *treatment* and *control* are randomized.
- What if we have the treatment group and the control group, but they are not fully randomized?
- If we have observations across two times at least with one before treatment and the other after treatment, then an easy way to make causal inference is **Difference in Differences(DID)** method.

DID estimator

- The DID estimator is

$$\hat{\beta}_{DID} = (\bar{Y}_{treat,post} - \bar{Y}_{treat,pre}) - (\bar{Y}_{control,post} - \bar{Y}_{control,pre})$$



Card and Krueger(1994): Minimum Wage on Employment

Introduction

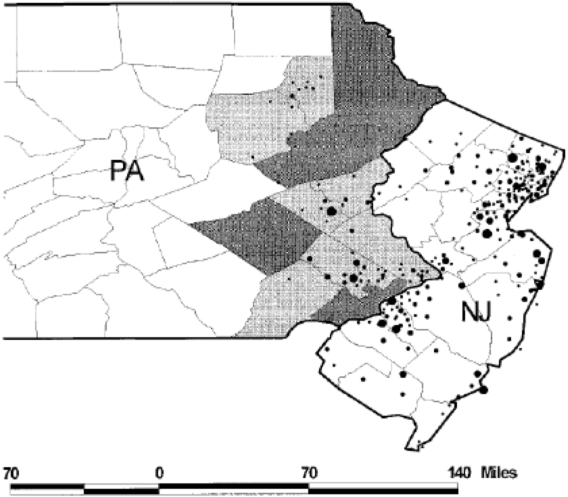
- Theoretically,in competitive labor market, increasing binding minimum wage decreases employment.But what about the reality?
- Ideal experiment: randomly assign labor markets to a control group (minimum wage kept constant) and treatment group (minimum wage increased), compare outcomes.
- Policy changes affecting some areas and not others create natural experiments.
  - Unlike ideal experiment, control and treatment groups here are not randomly assigned.

Card and Krueger(1994): Backgroud [1]

- Policy Change: in April 1992
  - Minimum wage in New Jersey from $4.25 to $5.05
  - Minimum wage in Pennsylvania constant at $4.25
- Research Design:
  - Collecting the data on employment at 400 fast food restaurants in NJ(treatment group) in Feb.1992 (before treatment)and again November 1992(after treatment).
  - Also collecting the data from the same type of restaurants in eastern Pennsylvania(PA) as control group where the minimum wage stayed at $4.25 throughout this period.

[1] Card, D., and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. The American Economic Review, 84(4)

Card & Krueger(1994): Geographic Background



Card & Krueger(1994): Model Graph[2]

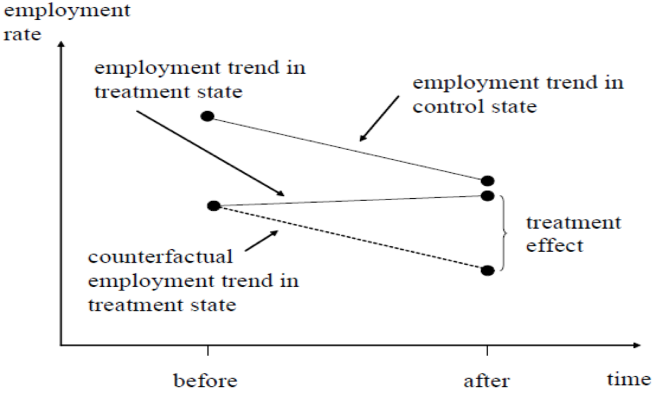


Figure 5.2.1: Causal effects in the differences-in-differences model

[2]Source:Angrist and Pischke(2009)

Card & Krueger(1994):Result[3]

Table 5.2.1: Average employment per store before and after the New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1. FTE employment before,	23.33	20.44	-2.89
all available observations	(1.35)	(0.51)	(1.44)
2. FTE employment after,	21.17	21.03	-0.14
all available observations	(0.94)	(0.52)	(1.07)
3. Change in mean FTE	-2.16	0.59	2.76
employment	(1.25)	(0.54)	(1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The

[3]Source:Angrist and Pischke(2009)

Regression DD - Card and Krueger

- DID model:

$$Y_{st} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + u_{st}$$

  - NJ is a dummy equal to 1 if the observation is from NJ(treat).Otherwise equal to 0 from Penny(control).
  - d is a dummy equal to 1 if the observation is from November (the **post** period),otherwise equal to 0(Feb. the **pre** period)
  - $(NJ \times d)$ is the **interaction term** of NJ and d.
  - u_{st} is the error term.
- Which estimated coefficient represents the DID estimator?

Regression DD - Card and Krueger

- A 2×2 matrix table

		treat or control	
		NJ=0(control)	NJ=1(treat)
pre or post	d=0(pre)	α	$\alpha + \gamma$
	d=1(post)	$\alpha + \lambda$	$\alpha + \gamma + \lambda + \delta$

- Then DID estimator

$$\begin{aligned}
\hat{\beta}_{DID} &= (\bar{Y}_{treat,post} - \bar{Y}_{treat,pre}) - (\bar{Y}_{control,post} - \bar{Y}_{control,pre}) \\
&= (NJ_{post} - NJ_{pre}) - (PA_{post} - PA_{pre}) \\
&= [(\alpha + \gamma + \lambda + \delta) - (\alpha + \gamma)] - [(\alpha + \lambda) - \alpha] \\
&= \delta
\end{aligned}$$

Specifications of DID

A Simple(2×2)DID Regression

- The simple DID regression on the individual level can be written as

$$Y_{ist} = \alpha + \beta(Treat \times Post)_{st} + \gamma Treat_s + \delta Post_t + u_{ist}$$

  - $Treat_s$(or D)is a dummy variable indicate whether or not is **treated**.
  - $Post_t$(or T) is a dummy variable indicate whether or not is **post-treatment** period.
  - γ captures the outcome gap between treatment and control group that *are constant over time*.
  - δ captures the outcome gap across post and pre period that *are common to both two groups*.
  - β is the coefficient of interest which is the **difference-in-differences** estimator
- **Note**: The outcomes are often measured at the individual level i,while treatment takes place at the group level.(The S.E. has to be adjusted).

A Simple(2×2)DID Regression with Covariates

- Add more covariates as **control variables** which may reduce the residual variance (lead to smaller standard errors)

$$Y_{ist} = \alpha + \beta(Treat \times Post)_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

- X_{ist} is a vector of control variables, which can include **individual level characteristics** and **time-varying measured at the group level**. Γ is the corresponding estimate coefficient vector.
- Those *time-invariant Xs* may not helpful because they are part of fixed effect which will be differential(absorted in α and γ).
- *Time-varying Xs* may be problematic if they are the outcomes of the treatment which are **bad controls**.
- So *Pre-treatment covariates* which could include *Xs* on both group and individual level are more favorable.

A Simple 2 × 2 DID Regression with Many Periods

- We can slightly change the notations and generalize it into

$$Y_{ist} = \alpha + \beta D_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

  - Where D_{st} means $(Treat \times Post)_{st}$
- Using **Fixed Effect** models further to transform it into

$$Y_{ist} = \beta D_{st} + \alpha_s + \delta_t + \Gamma X'_{ist} + u_{ist}$$

  - $\alpha_s = \alpha + \gamma_s$ is a set of groups fixed effects, which captures $Treat_s$.
  - δ_t is a set of time fixed effects, which captures $Post_t$.
- Note:
  - Samples enter the treatment and control groups **at the same time**.
  - The frame work can also apply to **Repeated(Pooled) Cross-Section Data**.

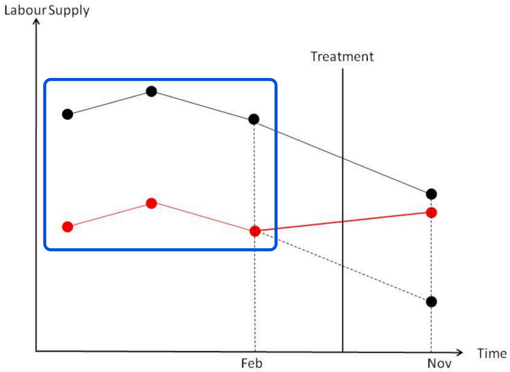
Key Assumption For DID

Paralled Trend

- A key identifying assumption for DID is: **Common trends** or **Parallel trends**
  - Treatment would be the same "trend" in both groups in the absence of treatment.
- This doesn't mean that they have to have the same mean of the outcome.
- There may be some unobservable factors affected on outcomes of both group. But as long as the effects have the same trends on both groups, then DID will eliminate the factors.
- It is difficult to verify because technically one of the parallel trends can be an unobserved counterfactual.

Assessing Graphically

- **Common Trend**: It is difficult to verify but one often uses pre-treatment data to show that the trends are the same.
    - If you only have two-period data, you can do **nothing**.
    - If you luckly have multiple-period data, then you can show something graphically.

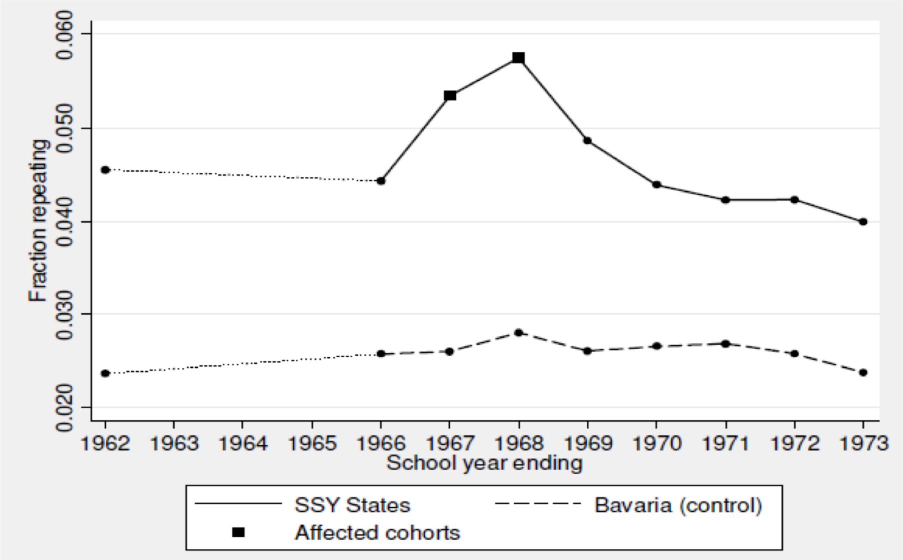


An Encouraging Example: Pischke(2007)[4]

- Topic: the length of school year on student performance
- Background:
  - Until the 1960s, children in all German states except Bavaria started school in the Spring. In 1966-1967 school year, the Spring moved to Fall.
  - It make two shorter school years for affected cohort, 24 weeks long instead of 37.
- Research Design:
  - Dependent Variable: Retreating rate
  - Independent Variable: spending time on school
  - Treatment group: Students in the German **States except Bavaria**.
  - Control group: Students in **Bavaria**.

[4] Pischke, J. (2007). The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years. The Economic Journal, 117(5)

An Encouraging Example: Pischeke(2007)



An Encouraging Example: Pischeke(2007)

- This graph provides strong visual evidence of treatment and control states with a common underlying trend.
- A treatment effect that induces a sharp but transitory deviation from this trend.
- It seems to be clear that a short school years have increased repetition rates for affected cohorts.

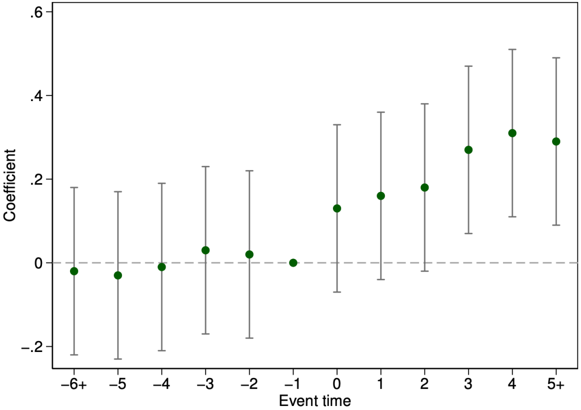
The Event Study Design: Including Leads and Lags

- If you have a multiple years panel data, then including leads into the DD model is an easy way to analyze pre-treatment trends.
- Lags can be also included to analyze whether the treatment effect changes over time after assignment.
- The estimated regression would be

$$Y_{ist} = \alpha_s + \delta_t + \sum_{\tau=-q}^{-1} \theta_\tau D_{st} + \sum_{\tau=0}^{p} \delta_\tau D_{st} + X_{ist} + u_{ist}$$

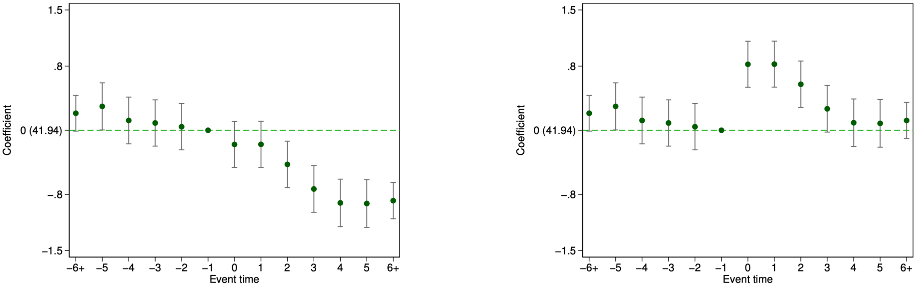
- Treatment occurs in year 0
- Includes q leads or anticipatory effects,thus θ_τ should be no different from 0.
- Includes p leads or post treatment effects, thus δ_τ had better be different from 0 significantly,at least for some periods.

The Event Study Design: Including Leads and Lags[5]



[5] Source:Freyaldenhoven, S., Hansen, C., Pérez, J. P., and Shapiro, J. M. (2021). Visualization, Identification, and Estimation in the Linear Panel Event-Study Design. SSRN Electronic Journal

The Event Study Design: Including Leads and Lags[6]



(a) "Smooth" event-time trend

(b) "Jump" at the time of the event

Figure 3: Label for normalized coefficient. Exemplary event-study plot for two possible datasets. Relative to Figure 2, a parenthetical label for the average value of the outcome corresponding to the normalized coefficient has been added, in accordance with Suggestion 2.

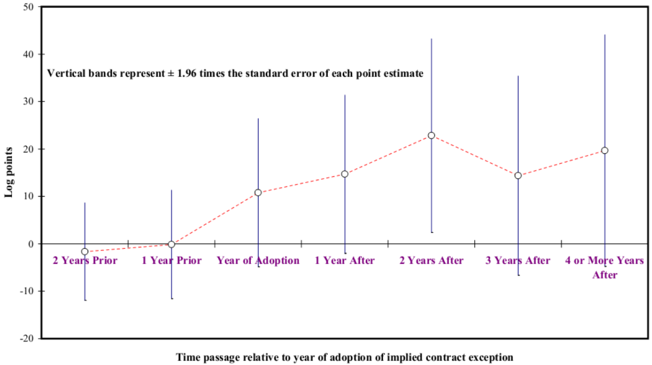
[6] Source:Freyaldenhoven, S., Hansen, C., Pérez, J. P., and Shapiro, J. M. (2021). Visualization, Identification, and Estimation in the Linear Panel Event-Study Design. SSRN Electronic Journal

Study including leads and lags: Autor (2003)

- Autor (2003) includes both leads and lags in a DD model analyzing the effect of increased employment protection on the firm's use of temporary help workers.[7]
- In the US employers can usually hire and fire workers at will.
- U.S labor law allows **employment-at-will** but in some state courts have allowed a number of exceptions to the doctrine, leading to lawsuits for *unjust dismissal*.
- The employment of temporary workers in a state to dummy variables indicating state court rulings that allow exceptions to the *employment-at-will* doctrine.
- The standard thing to do is normalize the adoption year to 0
- Autor(2003) then analyzes the effect of these exemptions on the use of temporary help workers.

[7] Autor, D. H. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. Journal of Labor Economics, 21(1), 1–42.

Study including leads and lags: Autor (2003)



- The leads are very close to 0: Common trends assumption may hold.
- The lags show that the effect increases during the first years of the treatment and then remains relatively constant.

Loose Common Trend Assumption

Add group-specific time trends

- This setting can eliminate the effect of group-specific time trend in outcome on our DID estimates

$$Y_{ist} = \beta D_{st} + \alpha_s + \delta_t + \tau_{st} + \Gamma X'_{ist} + u_{ist}$$

- τ_{st} is group-specific dummies multiplying the time trend variable t, which can be *quadratic* to capture some nonlinear trend.
- The **group specific time trend** in outcome means that treatment and control groups can follow different trends.
- It make DID estimate more robust and convincing.
- **Strong Assumption**: the pre-treatment data establish a clear trend that can be **extrapolated** into the post-treatment period.

Add group-specific time trends

- Besley, T., & Burgess, R. (2004). Can Labor Regulation Hinder Economic Performance? Evidence from India. *The Quarterly Journal of Economics, 119(1)*
  - Topic: labor regulation on businesses in Indian states
  - Method: Difference-in-Differences
  - Data: States in India
  - Dependent Variable: log manufacturing output per capita on states levels
  - Independent Variable: Labor regulation(lagged) coded $1 = pro - worker; 0 = neutral; -1 = pro - employer$ and then accumulated over the period to generate the labor regulation measure.

TABLE 5.2.3
Estimated effects of labor regulation on the performance of firms in Indian states

	(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

- Controlling the group specific time trend- thus the long-term propensity of pro-labor of the states- makes the estimate to **zero**.

Within control group – DDD(Triple D)

- More convincing analysis sometime comes from higher-order contrasts: **DDD** or **Triple D** design.
  - Build the third dimension of contrast to eliminate the potential bias.
- e.g: Minimum Wage
  - Treatment group: Low-wage-workers in NJ.
  - Control group 1: High-wage-workers in NJ.
  - Assumption 1: the low wage group would have the same trends as high wage group if there were not the new law.
  - Control group 2: Low-wage workers in PA.
  - Assumption 2: the low wage group in NJ would have the same trends as those in PA if there were not the new law.
- It can loose the simple *common trend* assumption in simple DID.

Within control group – DDD(Triple D)

- Gruber, J. (1994). The incidence of mandated maternity benefits. *The American Economic Review, 84(3), 622–641*.
  - Topic: how the *mandated maternity* benefits affects female's wage and employment.
  - Several state government passed the law that mandated childbirth be covered comprehensively in health insurance plans.
  - Dependent Variable: log hourly wage
  - Independent Variable: mandated maternity benefits law
- Econometric Method: **Triple D**
  1. DID estimates for treatment group (women of childbearing age) in treatment state v.s. control state before and after law change.
  2. DID estimates for control group (women not in childbearing age) in treatment state v.s. control state before and after law change.
  3. **DDD** estimate of the effect of mandated maternity benefits on wage is (1) − (2)

Within control group – DDD(Triple D)

- DDD in Regression

$$Y_{isct} = \beta D_{sct} + \alpha_s + \gamma_c + \delta_t + \lambda_{1st} + \lambda_{2sc} + \lambda_{3ct} + \Gamma X'_{icst} + u_{isct}$$

- α_s:a set of dummies indicating whether or not treatment state
- δ_t: a set of dummies indicating whether or not law change
- γ_c: a set of dummies indicating whether or not women of childbearing age

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

Table 3—DDD Estimates of the Impact of State Mandates on Hourly Wages

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	$\boxed{\begin{matrix}-0.062\\(0.022)\end{matrix}}$		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	$\boxed{\begin{matrix}-0.008:\\(0.014)\end{matrix}}$		
DDD:	**−0.054** **(0.026)**		

Table 3—DDD Estimates of the Impact of State Mandates on Hourly Wages

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	$\boxed{-0.062 \; (0.022)}$		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	$\boxed{-0.008: \; (0.014)}$		
DDD:	**−0.054** **(0.026)**		

More Extensions

DID for different treatment intensity

Card(1992): Minimum Wage on Employment[8]

- Study treatments with different treatment intensity, e.g., varying increases in the minimum wage for different states.
- **Background**: the federal minimum increased from \$3.35 to \$3.80. It means that *there is NO control group*, because all states have to follow without exemption.
- The DID regression can be

$$Y_{ist} = \beta(Intense_s \times D_t) + \gamma_s + \delta_t + u_{ist}$$

  - Where the variable $Intense_s$ is a measure of the fraction of teenagers likely to be affected by a minimum wage increase in each state and D_t is a dummy for observations after 1990,
  - β means that how much does wage increase when increasing the one fraction of affected teenagers by an increase of the federal minimum wage.

[8] Card, D. (1992). Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage. Industrial and Labor Relations Review, 46(1), 22–37.

Card(1992): DID for different treatment intensity

- In the t period, the DID regression model can be

$$Y_{ist} = \beta(Intense_s \times D_t) + \gamma_s + \delta_t + u_{ist}$$

- In the t-1 period, the DID regression model can be

$$Y_{is,t-1} = \beta(Intense_s \times D_{t-1}) + \gamma_s + \delta_{t-1} + u_{is,t-1}$$

- The **first-difference** between pre and post treatment equivalence is

$$\Delta \overline{Y}_s = \bar{\gamma} + \beta(Intense_s) + \Delta \bar{u}_s$$

- Where $\Delta \overline{Y}_s = \frac{1}{n_s}\Sigma_i(Y_{ist} - Y_{ist-1})$ is a measure of the change in teen employment and average wage of state s, from 1989 to 1990.
- And $D_t - D_{t-1} = 1$ and $D_{t-1} = 0$.

Card(1992): DID for different treatment intensity[9]

Table 5.2.2: Regression-DD estimates of minimum wage effects on teens, 1989 to 1992

Explanatory Variable	Equations for Change in Mean Log Wage: (1)	(2)	Equations for change in Teen Employment-Population Ratio: (3)	(4)
1. Fraction of Affected Teens	0.15 (0.03)	.14 (0.04)	0.02 (0.03)	-.01 (0.03)
2. Change in Overall Emp./Pop. Ratio	–	0.46 (0.60)	–	1.24 (0.60)
3. R-squared	0.30	0.31	0.01	0.09

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 1992 CPS. Regressions are weighted by the CPS sample size by state and year.

[9] Source:Angrist and Pischke(2009)

DID in Cross-Sectional Data(Cohort DID)

Introduction

- When using the Difference-in-Differences (DID) method, having **at least two time periods of panel data** is generally required.
- However, there are situations where we can still construct a valid DID design using **cross-sectional data** alone if *the shock is related to time or other dimensions*.
- This is especially useful for researchers who may not have access to panel data, or for those who are working with data that is hard to come by.
  - **Cohort-DID**
- **Cohort** here refers on *groups of people who share the same birth year or a period with a birth year*, such as the "1980s," "1990s," or "2000s" etc.
- In a DID design, when an unexpected shock or institutional change occurs that is **related to age**, some cohorts may be exposed to it while others may not.
- This creates a treated group and a control group in the DID design, which can help us better understand the effects of the shock or change.

Introduction

- A simple(2×2) Cohort-DID regression model can be

$$Y_{isg} = \alpha + \beta(TArea \times TCohort)_{sg} + \gamma TArea_s + \delta TCohort_g + u_{isg}$$

  - $TArea_s$ is a dummy variable indicate that the living areas of respondents whether or not are **treated**.
  - $TCohort_g$ is a dummy variable indicate that the cohorts of respondents whether or not are **treated**.
- A Standard Cohort-DID regression model

$$Y_{isg} = \beta D_{sg} + \alpha_g + \delta_s + \Gamma X'_{isg} + u_{isg}$$

  - δ_s controls area fixed effects.
  - α_g controls cohort fixed effecs.
  - X_{isg} is a vector of control variables, which can include individual level characteristics and time-varying measured at the group level.

Sent-Down Movement and Rural Education in China

- *Arrival of Young Talent: The Send-Down Movement and Rural Education in China*,American Economic Review 2020, 110(11): 3393–3430.By Yi Chen, Ziying Fan, Xiaomin Gu, and Li-An Zhou.
- **Topic**: The long-term consequence of Sent-Down Movement("上山下乡" 运动)
- **Background**:
  - The origins of the send-down movement can be traced back to the 1950s.
  - Before the Cultural Revolution, the program operated on a relatively small scale, and participation was largely voluntary.
  - After the outbreak of the **Cultural Revolution**, the send-down movement made a decisive turnaround and **mandated** about 16 million urban youths to go to the countryside.

Sent-Down Movement and Rural Education in China

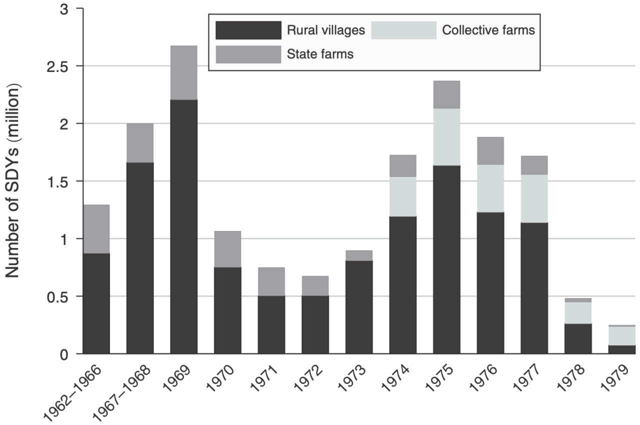
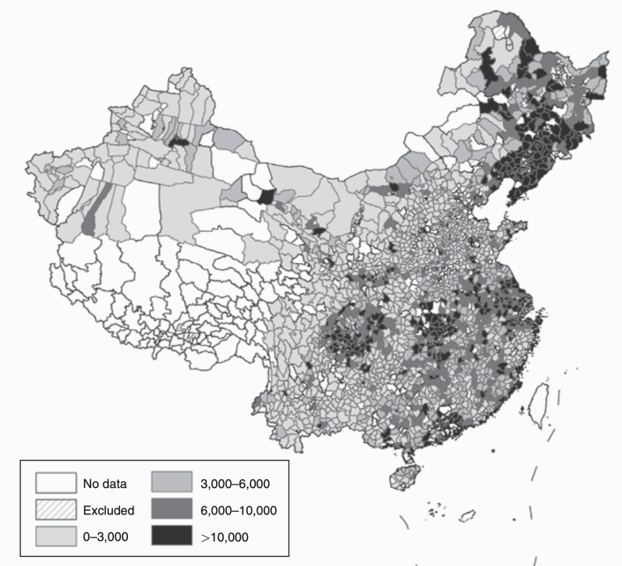


FIGURE 1. NUMBER OF SDYS BY RESETTLEMENT, 1962–1979

Source: Gu (2009)

Sent-Down Movement and Rural Education in China



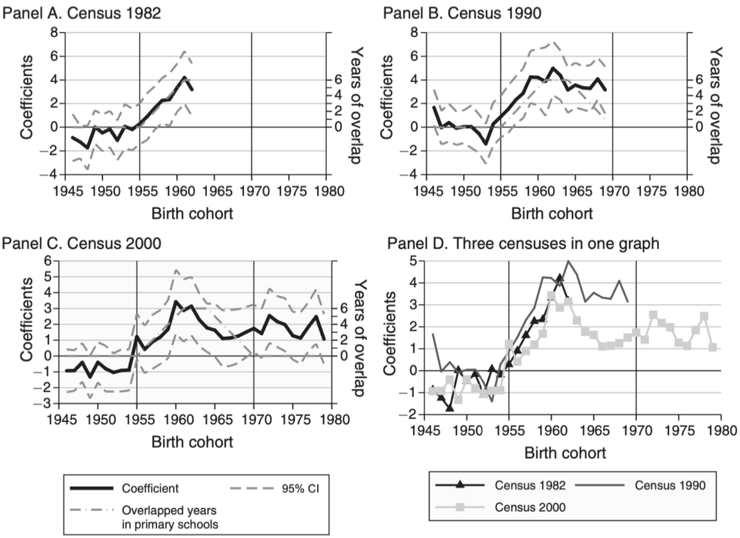
Sent-Down Movement and Rural Education in China

- A cohort DID regression model as following

$$Y\_Edu_{i,g,c,p} = \beta_0 + \beta_1 \%SDY_{c,p} \times I(1956 \leq g \leq 1969) + \beta_2 \mathbf{X}_{i,g,c,p} + \lambda_c + \mu_{g,p} + \Lambda_c \times \mu_g + \varepsilon_{i,g,c,p}$$

- $Y\_Edu_{i,g,c,p}$ refers to the years of education of individual i of cohort g in county c of province p.
- $\%SDY_{c,p}$ is the **density** of received SDYs in county c during the movement.
- $I(1956 \leq g \leq 1969)$ is an indicator function that equals 1 if the individual belongs to the cohort of 1956-1969, which is the exposure cohort.
- $\mathbf{X}_{i,g,c,p}$ is a vector of individual-level controls, including gender and ethnicity.
- λ_c is **county fixed effects**, which absorb all time-invariant county-level characteristics.
- $\mu_{g,p}$ is **province-cohort fixed effects** and an interaction terms between **county base education with cohort dummies**$(\Lambda_c \times \mu_g)$

Sent-Down Movement and Rural Education in China



Sent-Down Movement and Rural Education in China

TABLE 3—THE EFFECT OF SDYS ON THE EDUCATIONAL ATTAINMENT OF RURAL CHILDREN (1990 CENSUS)

Dependent variables:	Years of education		Complete primary		Complete junior high		Placebo I (1990) (1946–1950) versus (1951–1955)	Placebo II (2000) (1970–1974) versus (1975–1979)
Sample:	Rural	Urban	Rural	Urban	Rural	Urban		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Local density of received SDYs × affected cohorts (1956–1969)	3.237 (0.701)	0.151 (0.517)	0.441 (0.0873)	−0.0658 (0.0611)	0.767 (0.121)	−0.0517 (0.103)		
Local density of received SDYs × affected cohorts (placebo)							−0.817 (0.576)	−0.432 (0.319)
Male	1.874 (0.0284)	0.668 (0.0256)	0.201 (0.00361)	0.0319 (0.00227)	0.203 (0.00285)	0.0546 (0.00316)	2.286 (0.0300)	0.665 (0.0150)
Han ethnic	0.150 (0.0565)	3.34e-05 (0.0811)	0.0213 (0.00769)	0.00962 (0.00540)	0.00657 (0.00679)	0.0177 (0.00875)	0.0802 (0.0554)	0.477 (0.0401)
Observations	2,775,858	417,883	2,775,858	417,883	2,775,858	417,883	960,123	947,025
R^2	0.293	0.225	0.258	0.106	0.212	0.198	0.267	0.216
$\bar{Y}$ of control group	5.372	8.882	0.616	0.911	0.205	0.670		
County FE	✓	✓	✓	✓	✓	✓	✓	✓
Province-cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Base education × cohort FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors are clustered at the county level. Local density of received SDYs is computed by dividing the number of received SDYs by the county population in 1964. Base education is calculated as the primary and junior high graduation rates of the control group.

DID as an Instrument(DID+IV)

Introduction

- Recall IV: Instrument Exogeneity
  - Hard to test the assumption statistically that the instruments are exogenous.Instead, "telling good story"
- DID can also be treated as an IV, which apply the DID effect on the treatment variable instead of outcomes.
  - Advantage over simple IV: The exogeneity of the instrument depends on whether the DID strategy works or not which can be tested formally in DID frameworks.

Recall: Endogeneity in Two-Way Fixed Effects Model

- Assume that we have following **two-way fixed effects** model

$$Y_{it} = \alpha_i + \tau_t + \beta S_{it} + \epsilon_{it}$$

  - Y_{it} is the outcome.
  - α_i and τ_t are entity-fixed effect and time-fixed effect respectively.
  - S_{it} is our interest variable.
- **Potential bias** of $\hat{\beta}$?
  - some unobservable and time-varing factors could be omitted into ϵ_{it}, which leads to an **OVB**
  - And S_{it} is **endogenous**,which is correlated with ϵ_{it}.
- **Solution**: Find a valid instrument for S_{it}.

DID as an Instrument(DID+IV)

- Assume that we have a simple 2×2 DID policy change: Z_i and a time term $T \in \{0, 1\}$

$$Z_i = \begin{cases} 0, & \text{not exposed to the policy} \\ 1, & \text{not exposed to the policy} \end{cases} \text{and}, T_t = \begin{cases} 0, & \text{before the policy carring out} \\ 1, & \text{after the policy carring out} \end{cases}$$

- Then, the **first stage** of the **DDIV** is

$$S_{it} = \gamma_i + \delta_t + \pi Z_i T_t + \eta_{it}$$

- In other words, the interaction term $Z_i \times T_t$ is the instrument, which is essentially a DID design.

Case: Quantity and Quality Trade-off

- **Q**-**Q** model implication:
  - the reduction in the number of children increases parental investment per child and therefore improves child quality.
- A simple *Q*-*Q* model is

$$Y_i = \alpha + \beta Q_i + \gamma X_i + \epsilon_i$$

  - Y_i is the child quality, Q_i is the quantity of children, and X_i is a vector of control variables.
- Although a *negative* relationship has been widely observed, the cross-sectional association cannot be interpreted as the causal effect of quantity on quality.
  - OVB
  - Simultaneous bias

One-Child Policy: Li and Zhang(2017)[10]

- Plausibly exogenous variation in family size is due to
  - the **natural occurrence**: twin births or the sibling sex composition,
- **One-child policy(OCP)** as a exogenous policy to the number of children.
  - The OCP formally implemented in 1980s has varied significantly between rural and urban areas, over time, and across provinces, ethnicity, and even entities.
- They construct a quantitative indicator of the extent of local violation of the OCP using the percentage of current Han mothers of primary childbearing age who gave a higher order birth in 1981.
- Thus the **“excess fertility rate”(EFR)** as the measurement of local one-child policy intensity.

[10] Bingjing Li and Hongliang Zhang(2017),Does population control lead to better child quality? Evidence from China’s one-child policy enforcement, Journal of Comparative Economics 45 (2017)

Q-Q Trade-off: Family Size on Education

- A fixed effects model is as following

$$Y_{ijt} = \text{FamilySize}_{ijt}\beta + \boldsymbol{X_i}\pi + (\boldsymbol{C_j} \times T_i)\,\eta + \phi_j + \lambda_t + \varepsilon_{ijt}$$

- $FamilySize_{ijt}$ is the family size of firstborn child i from prefecture j in census year t;
- X_i contains a set of individual controls, including mother's age at first birth, mother's age at first birth squared, and dummy indicators for child's age, parents' education, and their employment sectors;
- ϕ_j and λ_t are the prefecture and census year fixed effects, respectively.
- C_j is a vector of prefecture-specific control variables that account for pre-existing fertility preferences and socio-economic characteristics;
- $\boldsymbol{C_j} \times T_i$ to net out regional EFR differences attributable to their differences in pre-existing fertility preferences and socioeconomic characteristics.
- Still suffer some **bias**?
  - time-varying unobservable factors could be omitted into ϵ_{ijt}, which leads to an **OVB**

DDIV: One-Child Policy on Family Size

- Instrument on family size: the **excess fertility rate**(EFR) as the measurement of local one-child policy intensity.

$$\text{FamilySize}_{ijt} = \beta\left(EFR_j \times T_i\right) + \boldsymbol{X_i}\gamma_1 + \left(\boldsymbol{C_j} \times T_i\right)\delta_1 + \phi_j + \lambda_t + u_{ijt},$$

- **the key variable of interest**, $EFR_j \times T_i$, is the interaction of the EFR in prefecture j and the post policy period dummy T_i.
- Other variables are the same as defined for previous equation.
- An *intensity DID design as an IV*, which apply the DID effect on the treatment variable instead of outcomes.
- It is the **first stage** of the **DDIV**.

DDIV: One-Child Policy on Education

- The DID regression model on the outcome

$$\mathrm{Y}_{ijt} = (EFR_j \times T_i)\,\alpha_2 + \boldsymbol{X_i}\gamma_2 + (\boldsymbol{C_j} \times T_i)\,\delta_2 + \phi_j + \lambda_t + u_{ijt}$$

- where Y_{ijt} denotes the **educational outcome** of firstborn child i from prefecture j in census year t;
- Other variables are the same as defined for previous equation.
- This is the **reduced-form** of DDIV.

DID: First stages and Reduced Forms

Table 3
Effect of policy enforcement intensity on family size and firstborn children's education.

	Boys			Girls		
	Family size	Education level	Junior secondary school attendance	Family size	Education level	Junior secondary school attendance
	(1)	(2)	(3)	(4)	(5)	(6)
EFR×Year1990	$\boxed{4.045^{***}}$	-0.482^{**}	-0.539^{***}	$\boxed{5.660^{***}}$	$-0.397^{\dagger}$	-0.626^{***}
	(0.367)	(0.189)	(0.146)	(0.390)	(0.245)	(0.160)
Control variables:						
Individual controls	Y	Y	Y	Y	Y	Y
Prefecture initial controls×Year1990	Y	Y	Y	Y	Y	Y
N	120,273	120,273	120,273	115,637	115,637	115,637

Notes: [1] All regressions include prefecture fixed effects and census fixed effects. [2] Individual controls include mother's age at first birth, mother's age at first birth squared, mother's education level, father's education level, mother's employment sector, father's employment sector, and child age fixed effects. [3] Prefecture-specific initial control variables include the average total number of births of females aged 45–54; the shares of females aged 25–44 with 1, 2, 3, and 4+ births, respectively; the shares of females aged 25–29, 30–34, 35–39, and 40–44, respectively; the agricultural sector's employment share among adults aged 25–49 by gender; and the shares of each education level category among adults aged 25–49 by gender. [4] Robust standard errors clustered at prefecture× year level are reported in parentheses. [5] *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$, † $p< 0.15$.

IV: Family Size on Education

- An OLS fixed-effects model

$$y_{ijt} = \text{FamilySize}_{ijt}\beta + \boldsymbol{X_i}\pi + (\boldsymbol{C_j} \times T_i)\,\eta + \phi_j + \lambda_t + \varepsilon_{ijt}$$

- A DDIV-2SLS model

$$y_{ijt} = \widehat{\text{FamilySize}}_{ijt}\beta + \boldsymbol{X_i}\pi + (\boldsymbol{C_j} \times T_i)\,\eta + \phi_j + \lambda_t + \varepsilon_{ijt}$$

  - where $\widehat{\text{FamilySize}}_{ijt}$ is the predicted value from **first stage** regression.

IV: Family Size on Education

Table 4
Effect of family size on firstborn children's education.

	Boys				Girls			
	Education level		Junior secondary school attendance		Education level		Junior secondary school attendance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Family size	-0.033^{***}	-0.119^{**}	-0.020^{***}	-0.133^{***}	-0.078^{***}	-0.070^{*}	-0.047^{***}	-0.111^{***}
	(0.003)	(0.048)	(0.002)	(0.039)	(0.003)	(0.042)	(0.002)	(0.027)
Control variables:								
Individual controls	Y	Y	Y	Y	Y	Y	Y	Y
Prefecture initial controls$\times$Year1990	Y	Y	Y	Y	Y	Y	Y	Y
Kleibergen and Paap rk statistic		121.26		121.26		210.64		210.64
Stock-Yogo critical value 10% maximal IV size		16.38		16.38		16.38		16.38
N	120,273	120,273	120,273	120,273	115,637	115,637	115,637	115,637

Notes: [1] All regressions include prefecture fixed effects and census fixed effects. [2] Individual controls include mother's age at first birth, mother's age at first birth squared, mother's education level, father's education level, mother's employment sector, father's employment sector, and child age fixed effects. [3] Prefecture-specific initial control variables include the average total number of births of females aged 45–54; the shares of females aged 25–44 with 1, 2, 3, and 4+ births, respectively; the shares of females aged 25–29, 30–34, 35–39, and 40–44, respectively; the agricultural sector's employment share among adults aged 25–49 by gender; and the shares of each education level category among adults aged 25–49 by gender. [4] Robust standard errors clustered at prefecture$\times$ year level are reported in parentheses. [5] *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Parallel Trends Assumption

- **Parallel Trend(I)**: EFR on Fertility by women's age
- To examine whether the EFR-fertility link indeed differs by age

$$\text{TotalBirths}_{ijt} = \sum_{l=33}^{57} \left(EFR_j \times T_i \times d_{il}\right) \theta_l + \boldsymbol{W}_{1i}\zeta + \sum_{l=33}^{57} \left(\boldsymbol{C_j} \times T_i \times d_{il}\right) \kappa_l + \phi_j + \lambda_t + v_{ijt}$$

  - where TotalBirths_{ijt} is the total number of births of female i from prefecture j in census year t. And d_{il} is a dummy that equals 1 if she is aged l

- **Parallel Trend(II)**: EFR on children's education by women's age

$$\text{EduLevel}_{ijt} = \sum_{l=33}^{57} \left(EFR_j \times T_i \times d_{il}\right) \delta_l + \boldsymbol{W}_{2i}\psi + \sum_{l=33}^{57} \left(\boldsymbol{c_j} \times T_i \times d_{il}\right) \tau_l + \phi_j + \lambda_t + v_{ijt}$$

  - where EduLevel_{ijt} is the education level of child i from prefecture j in census year t.

Paralled Trend(I): First Stage

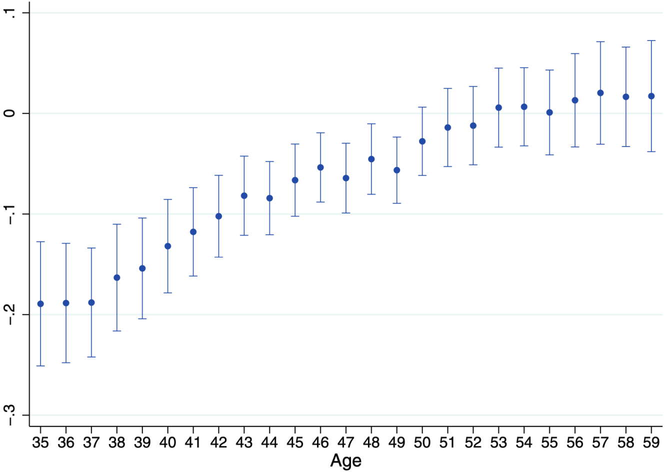


Fig. 1. EFR and intercensus change in fertility by women's age.
Notes: The figure displays the estimated coefficients and 95% confidence intervals of θ_l of Eq. (2) using all Han females aged 33–57 in the 1982 and 1990 Censuses.

Paralled Trend(II): Reduced form

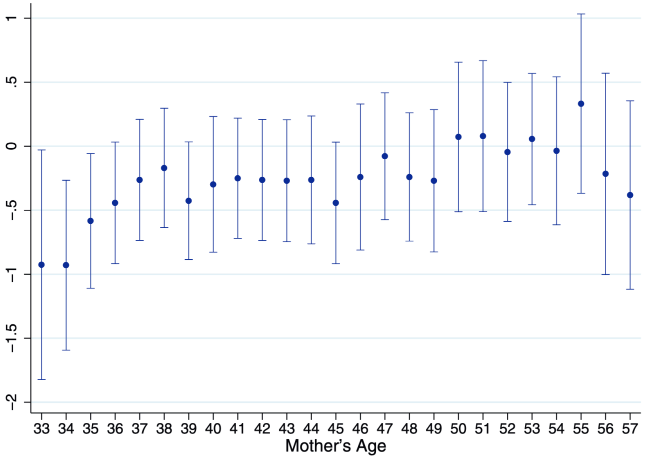


Fig. 2. EFR and intercensus change in children's education level by mother's age.
Notes: The figure displays the estimated coefficients and 95% confidence intervals of δ_l of Eq. (4) using 14- to 17-year-old Han children with mothers aged between 33 and 57 in the 1982 and 1990 Censuses.

DID with RDD

Terrorism on Individual Wellbeing

- Clark, A. E., Doyle, O., & Stancanelli, E. (2020). *The Impact of Terrorism on Individual Well-Being: Evidence from the Boston Marathon Bombing*. The Economic Journal, 130(631), 2065–2104.
- **Topic**: Terrorism on Individual Well-being.
- **Background**: The Boston marathon bombing took place on Monday 15 April 2013, when two bombs were detonated near the finish line, causing the death of three spectators and a policeman, and injuring 264 spectators.
- **Data**: The data come from the 2012 and 2013 ATUS and WB, which gather information on respondents' emotional well-being and a diary recording the activities over the past 24 hours.
- **Methods**: DID, RD and RDD-DID
- **Outcomes**:
  - Happy
  - Stress

Terrorism on Individual Wellbeing: RDD Model

- A RDD on Time regression model is

$$W_{it} = \gamma T_i + \beta f(D_t)\, T_i + \lambda f(D_t)\,(1 - T_i) + \mathbf{V}_t + u_{it}$$

  - T_i is the individual i whether expose to the treatment T.
  - D_t is the running variable which is the distance to D-Day. And $f(D_t)$ is a polynomial function of the running variable interacted with the treatment dummy T, to allow for different effects on either side of the cut-off.
  - $\mathbf{V}_t$: day(Monday to Sunday) fixed effects to control variations on weekdays versus weekends.
- Any potential bias?
  - The Boston marathon is itself an important sporting event in the United States and the runners come from all over the country to participate in it or watch it.
  - Emotional responses may therefore respond to the marathon itself, **independently** of any major terrorist attack.

Terrorism on Individual Wellbeing: DID Model

- A DID regression model is

$$W_{it} = \beta T_i \times \text{Year}_t + \tau T_i + \gamma \mathbf{Z}_i + \boldsymbol{v}_{st} + u_{it},$$

  - *Year* denotes the survey in 2012 or 2013.
  - Z_i is a matrix of individual characteristics, including demographic characteristics (age, age-squared, race and gender), education, economic status, and household characteristics.
  - V_{st} are state, year and day (Monday to Sunday) fixed effects.

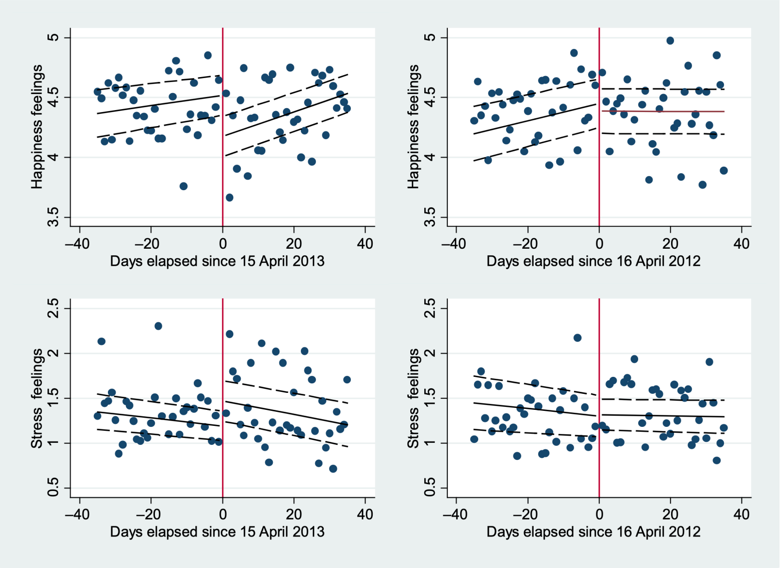
Terrorism on Individual Wellbeing: RDD-DID

- The combination of the RDD with the DID

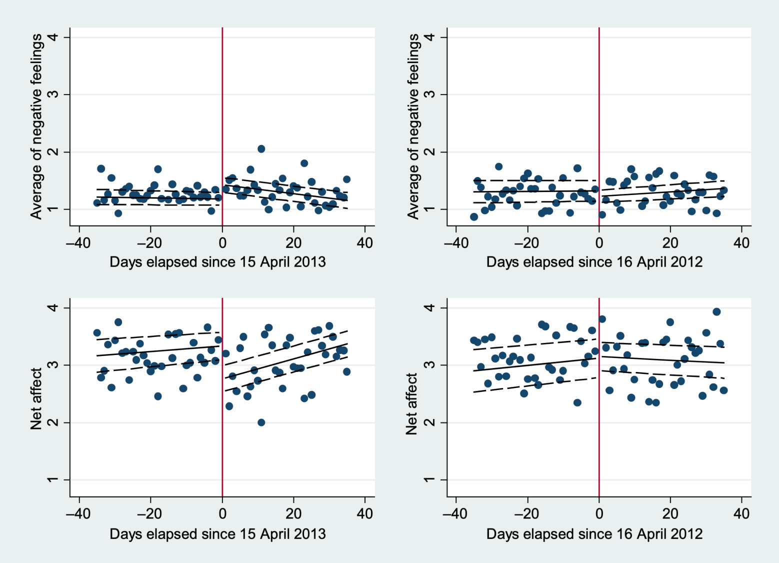
$$
\begin{aligned}
W_{it} = & \xi T_i \times \text{ Year }_t + \delta f(D_t) \times T_i \times \text{ Year }_i + \rho f(D_t) \times (1 - T_i) \times \text{Year}_t + \alpha f(D_i) \times T_i \\
& + \eta f(D_i) \times (1 - T_i) + \omega T_i + \psi Z_i + V_{st} + \theta_{it}.
\end{aligned}
$$

- Identification: use responses around the day of the 2012 Boston marathon, when there was no bombing, as a control group and combine this with the RDD model above.

Results: RD and DID(I)



Results: RD and DID(II)



Results: RD and DID(III)

Table 2. *The Effect of the Boston Marathon Bombing on Individual Well-being.*

	Happy	Stress	Negative affect	Net affect
Mean month before (SD)	4.44 (1.23)	1.24 (1.42)	1.20 (1.01)	3.25 (1.86)
1a) RDD (2)	−0.351**	0.351**	0.327***	−0.651***
Bandwidth 35 days, 2013 data	(0.136)	(0.172)	(0.117)	(0.196)
Observations	2,124	2,142	2,110	2,095
R^2	0.097	0.105	0.102	0.098
1b) RDD (non-parametric estimates)	−0.383**	0.298	0.277***	−0.618***
Optimal bandwidth, 2013 data	(0.171)	(0.191)	(00968)	(0.199)
2) Diff-in-Diff (3)	0.00973	−0.000760	0.00230	0.00333
Pooled 2012 and 2013 data	(0.0609)	(0.0678)	(0.0493)	(0.0937)
Observations	20,902	21,075	20,879	20,712
R^2	0.028	0.047	0.052	0.034

Results: RD and DID(IV)

3) RDD* Diff-in-Diff (4)	−0.379*	0.272	0.355**	−0.720**
Bandwidth 35 days, 2012 and 2013 data	(0.216)	(0.266)	(0.154)	(0.307)
Observations	4,366	4,396	4,341	4,316
R^2	0.062	0.068	0.069	0.063
4) RDD* Diff-in-Diff (4)	−0.514**	0.436	0.519**	−0.992***
Bandwidth 21 days, 2012 and 2013 data	(0.248)	(0.305)	(0.193)	(0.373)
Observations	2,708	2,729	2,693	2,675
R^2	0.075	0.083	0.070	0.072
5) RDD* Diff-in-Diff (4)	−0.626**	0.572	0.560**	−1.082**
Bandwidth 14 days, 2012 and 2013 data	(0.296)	(0.423)	(0.253)	(0.437)
Observations	1,877	1,891	1,870	1,856
R^2	0.098	0.091	0.099	0.096
6) RDD* Diff-in-Diff (4)	−0.435**	0.231	0.336**	−0.771***
Bandwidth 56 days, 2012 and 2013 data	(0.185)	(0.202)	(0.144)	(0.278)
Observations	6,571	6,616	6,543	6,502
R^2	0.046	0.058	0.065	0.047
7) RDD* Diff-in-Diff (4)	−0.456*	0.314	0.364*	−0.789**
Bandwidth 56 days, 2012 and 2013 data, and quadratic functional form	(0.271)	(0.328)	(0.190)	(0.388)
Observations	6,571	6,616	6,543	6,502
R^2	0.048	0.059	0.066	0.049
8) RDD* Diff-in-Diff (4)	−0.0725	0.152	0.326**	−0.385
Bandwidth 35 days, 2012 and 2013 data, including observations on the day of the marathon	(0.306)	(0.258)	(0.145)	(0.385)
Observations	4,420	4,451	4,395	4,369
R^2	0.078	0.075	0.071	0.074
9) RDD* Diff-in-Diff (4)	−0.379	0.272	0.355**	−0.720**

Standard errors and Other Threats

Standard errors in DD strategies

- Many paper using DD strategies use data from many years: not just 1 pre and 1 post period.
- The variables of interest in many of these setups only vary at a group level (say a state level) and outcome variables are often serially correlated.
- In the Card and Krueger study, it is very likely that employment in each state is not only correlated within the state but also serially correlated.
- As Bertrand, Duflo and Mullainathan (2004) point out, conventional standard errors often severely *understate* the standard deviation of the estimators – standard errors are biased downward.[11]

[11] Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics

Standard errors in Practice

- Simple solution:
  - **Clustering standard errors at the group level**,but the number of groups does matter($c \geq 50$).
  - It may also cluster at both the group level and time level.
- Other solutions: Bootstrapping

Other Threats to Validity

- Non-parallel trends
- Functional form dependence
- Multiple treatment times(Staggered DID)
- Other simultaneous shocks

Non-parallel trends

- Often policymakers will select the treatment and controls based on pre-existing differences in outcomes: practically guaranteeing the parallel trends assumption will be violated.
- "Ashenfelter dip"
  - Participants in job trainings program often experience a "dip" in earnings just prior to entering the program.
  - Since wages have a natural tendency to mean reversion,comparing wages of participants and non-participants using DD leads to an upward biased estimate of the program effect.

Function Forms

- So far our specifications of DID regression equation is linear, but what if it is wrong?
- Several nonparametric or semi-parametric methods can be used
  - Matching DID: Propensity Score Matching and Kernel Density Matching DID
  - Semiparametric DID

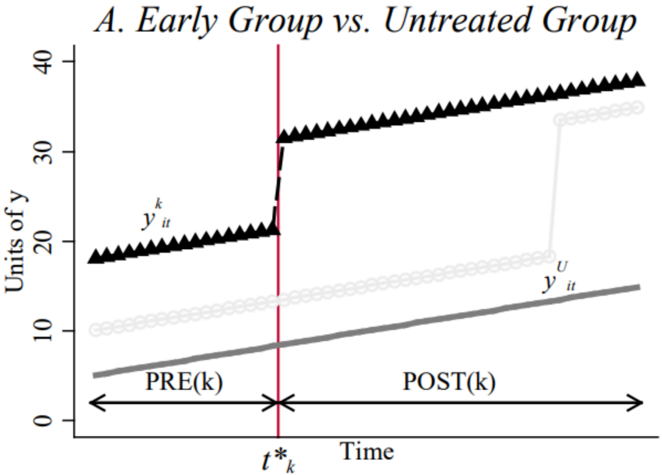
DID with multiple treatment times

- What happens if we have treated units who **get treated at different times**?
  - Staggered DID(**交错或渐进**)
- The simple DID model

$$Y_{ist} = \alpha + \beta D_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

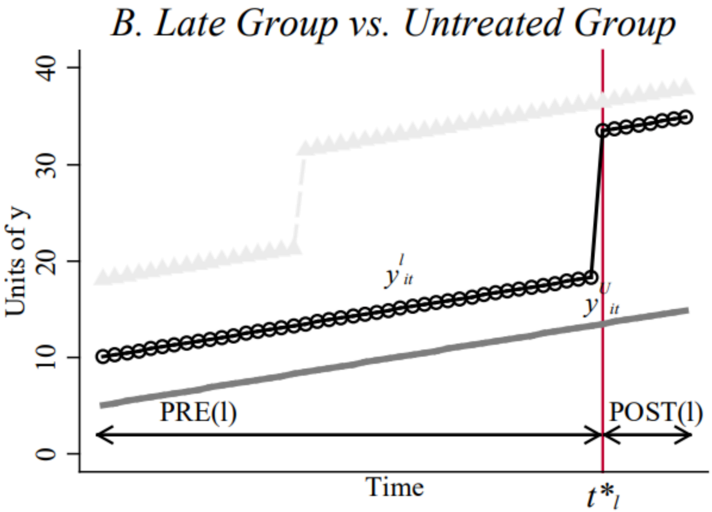
- But now D_{st} can turn from 0 to 1 at different times for different units.
  - eg. China's High-speed rail

Goodman-Bacon(2021)[12]

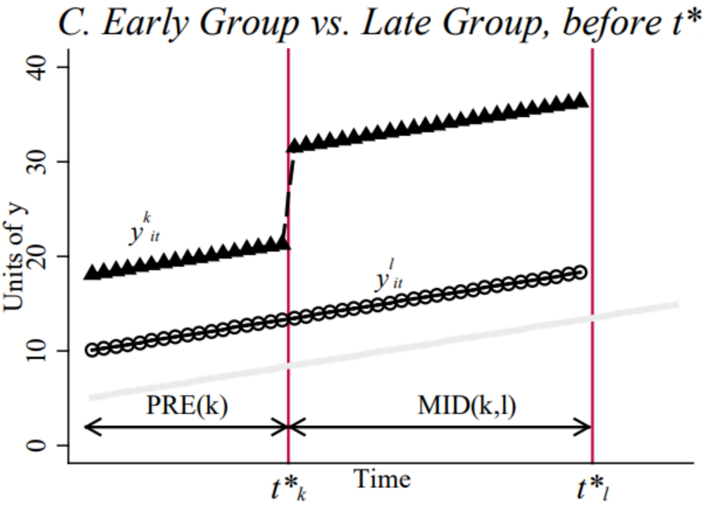


[12] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254–277.

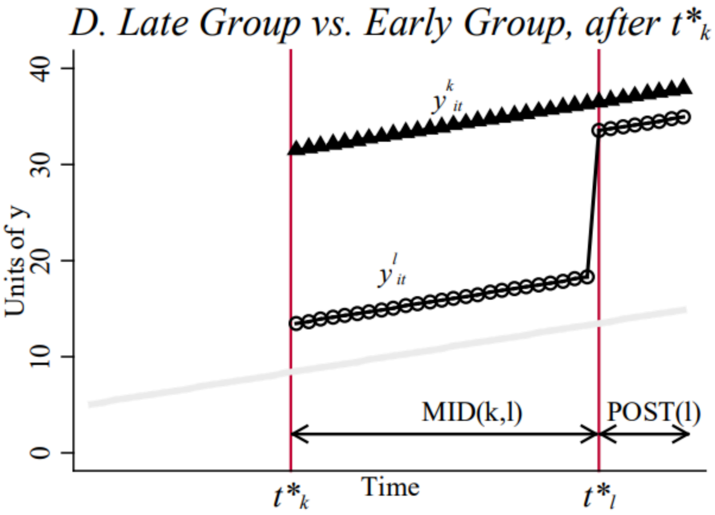
Goodman-Bacon(2021)



Goodman-Bacon(2021)



Goodman-Bacon(2021)



DID with multiple treatment times

- **Caution**: the TWFE specification gets you a **weighted average** of several comparisons.This may not be exactly what you want with heterogeneous treatment effects.
- New diagnostic approaches such as
  - **Bacon decomposition** by Goodman-Bacon (2021)
  - de Chaisemartin and D'Haultfoeuille (2020) [13]
- Alternative estimators
  - Callaway and Sant'Anna (2021) [14]
  - Borusyak, Jaravel and Spiess (2021) [15]
  - Others...

[13] Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,American Economic Review, 2020, 110 (9), 2964–2996.

[14] Difference-in-Differences with multiple time periods, Journal of Econometrics, 2021, 225 (2), 200–230.

[15] Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. arXiv. https://doi.org/10.48550/arxiv.2108.12419

Checks for DD Design

- Very common for readers and others to request a variety of “robustness checks” from a DID design.
- Think of these as along the same lines as the leads and lags
  - Falsification test using data for prior periods
  - Falsification test using data for alternative control group(kind of triple DDD)
  - Falsification test using alternative “placebo” outcome that should not be affected by the treatment

Summary

Wrap up

- Difference-in-differences is a special case of fixed effect model with much more powers in our toolbox to make causal inference.
- The key assumption is common trend which is not easy to testify using data.
- DID can be mixed with other methods such as IV and RD to obtain a more reliable causal inference.
- Noting that using the right way to inference the standard error.

Extensions of DID(II): Synthetic Control Method(SCM)

Basic Idea

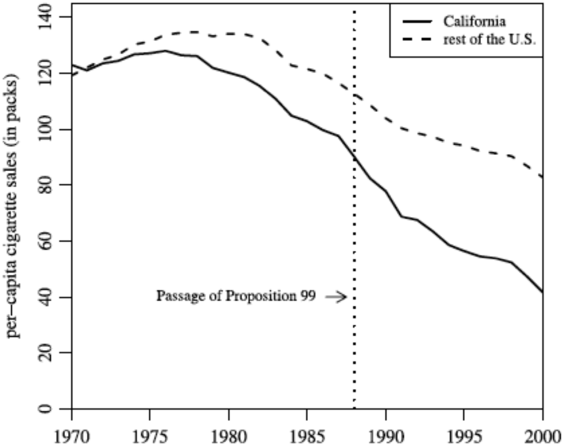
- The **synthetic control method**(SCM) were originally proposed in Abadie and Gardeazabal (2003) and Abadie et al. (2010) with the aim to estimate the effects of aggregate interventions,
- Interventions that are implemented at an aggregate level affecting a small number of large units (such as a cities, regions, or countries), on some aggregate outcome of interest.
- The basic idea behind synthetic controls is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone.
    - a data-driven procedure to use a small number of non-treated units to build the suitable counterfactuals.
- It is useful for case studies, which is nice because that is often all we have.
- Continues to also be methodologically a frontier for applied econometrics and is widely used in many field, even outside academia.

Extensions of DID: Synthetic Controls Method

- The basic idea is use (long) longitudinal data to build the **weighted average of non-treated** units that best reproduces characteristics of the treated unit over time in pre-treatment period.
- The weighted average of non-treated units is the **synthetic cohort**.
- Causal effect of treatment can be quantified by a simple difference after treatment:
    - **treated vs synthetic cohort**.

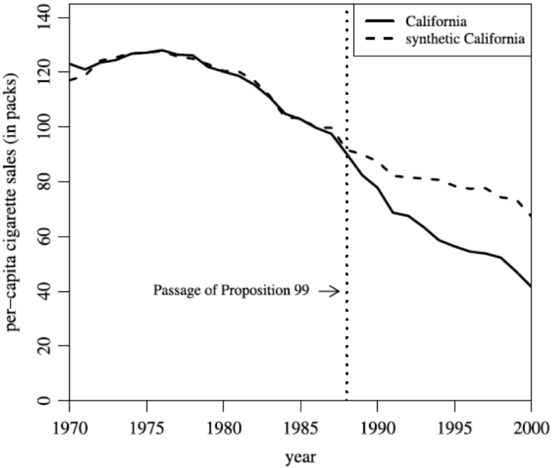
Abadie et.al(2010): Tax on Cig-Consumption

- In 1988, California passed comprehensive tobacco control legislation: Increased cigarette taxes by $0.25 per pack.



Abadie et.al(2010): Tax on Cig-Consumption

- Using 38 states that had never passed such programs as controls: **Synthetic CA**



Predictor Means: Actual vs Synthetic California

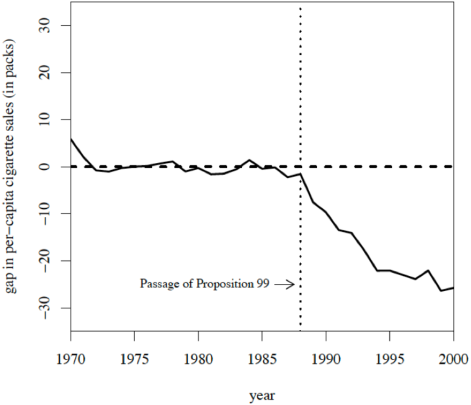
- Most observables are similar between Actual and Synthetic

	California		Average of
Variables	Real	Synthetic	38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).

The Application: Actual vs Synthetic California

- The treatment effect is measured by the gap in ciga-sales between Actual and Synthetic



Formalization

Formalization: The Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$
  - Assume that the first unit ($j = 1$) is the **treated unit**, that is, the unit affected by the policy intervention of interest.
  - Then the set of potential comparisons,$j = 2, ..., J+1$ is a collection of **untreated units**, not affected by the intervention.
- Assume also that our data span T periods and that the first T_0 periods are **before the intervention**.
- Let Y_{jt} and Y_{jt}^C be the real and counterfactual outcomes of interest for unit j of $J+1$ aggregate units at time t with and without intervention.
- Then the effect of the intervention of interest for the affected unit in period $t(t > T_0)$(ATT)

$$\tau_{1t} = Y_{1t} - Y_{1t}^C$$

Formalization: The Setting

- How to reproduce Y_{1t}^C which is totally unobservable?
  - Use unaffected units in control groups to predict it like matching in cross-sectional data.
- More specifically, a weighted average of the units in the comparison group use to construct the potential outcome of treated units, which define as **synthetic control**.Thus,

$$\hat{Y}_{1t}^C = \Sigma_{j=2}^{J+1} w_j Y_{jt}$$

- Then the question is how to determine these values of the weights, w_j or $W = (w_2, w_3, ...w_{J+1})$

Formalization: Weights

- Let more specifically, $W = (w_2, ..., w_{J+1})'$ have to satisfy two restriction conditions
  - $w_j \geq 0$ for $j = 2, ..., J+1$
  - $\Sigma_{j=2}^{J+1} w_j = 1$
- **Key Question**: how to determine these values of the weights, w_j or how to construct a proper control group?
  - eg. assigning equal weights, thus
$$w_j = \frac{1}{J}$$
  - or a fraction of the total population in the comparison group(at the time of the intervention),thus
$$w_j = \frac{N_j}{\Sigma_{j=2}^{J+1} N_j}$$

Formalization: Weights of Xs

- For each unit, j, we also observe a set of characteristics or covariates which can be use to predict the outcome Y_{jt}, denoted as $X_{1j}, ...X_{kj}$
- Let X_1 is a $k \times 1$ vector of pre-intervention characteristics for the treated unit. Similarly, let X_0 be a $(k \times J)$ matrix which contains the same variables for the unaffected units.
- Recall: how to measure the closeness or similarity between two vectors?

Formalization: Weight by Matching

- The rule to choose the optimal weight vector $W^* = (w_2, ..., w_{J+1})'$ will be

$$argmin_W \parallel (X_1 - X_0 W) \parallel$$

  - Thus,the optimal vector of weight W should **minimize the "distance"** between treated unit and untreated group before the treatment,subject to two weight constraints.

- More specifically, *Abadie, et al(2010)* consider

$$\parallel (X_1 - X_0 W) \parallel_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$$

where V can be some $(k \times k)$ symmetric and positive semidefinite matrix.

Formalization: More on the V matrix

- Typically, V is diagonal with main diagonal $v_1, ..., v_k$. Then the synthetic control weights minimize

$$\sum_{m=1}^{k} v_m \big(X_{1m} - \sum_{j=2}^{J+1} w_j^* X_{jm}\big)^2$$

  - Where v_m is a **weight** that reflects the *relative importance* that we assign to the m^{th} variable when we measure the discrepancy between the treated unit and the synthetic controls.
  - And v_m is critical because it weights directly shape w_j,which help reproducing the counterfactual outcome for the treated unit in the absence of the treatment.

Formalization: Estimating the V matrix

- Various ways to choose V
  - In practice,most people choose V that minimizes *the mean squared prediction error(MSPE)*.Thus,

$$\sum_{t=1}^{T_0}\left(Y_{1t}-\sum_{j=2}^{J+1} w_j^*(V)Y_{jt}\right)^2$$

- If the number of pre-intervention periods in the data is “large”, then matching on pre-intervention outcomes can allow us to control for the heterogeneous responses to multiple unobserved factors.
- The intuition here is that only units that are alike on unobservables and unobservables would follow a similar trajectory pre-treatment.

A Machine Learning Procedure

1. Divide the pre-intervention periods(T_0) into a initial **training** period($t = 1, \ldots t_0$) and a subsequent **validation** period($t = t_0 + 1, \ldots T_0$).
2. Select a value V^* make the MSPE is small

$$\sum_{t=t_0+1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt}\right)^2$$

3. Use the resulting V^* and data on the predictors for the last t_0 before in the intervention, $t = t_0 + 1, t_0 + 2, \ldots, T_0$ to calculate $w^* = w(V^*)$

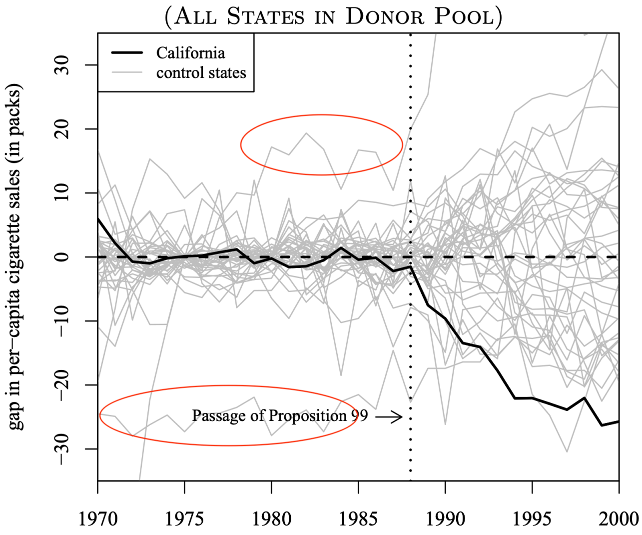
Inference

- Permutation Strategy: whether the effect estimated by the synthetic control for the unit affected by the intervention is **large** relative to the effect estimated for a unit chosen at random.
- Implementation: "randomization" of the treatment to each unit, re-estimating the model, and calculating a set of root mean squared prediction error (RMSPE) values for the pre- and post-treatment period.
- For $0 \leq t_1 \leq t_2 \leq T$ and $j = 1, 2, ..., J+1$,let

$$R_j(t_1, t_2) = \left(\frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} (Y_{jt} - \hat{Y}_{jt}^N)^2\right)^{\frac{1}{2}}$$

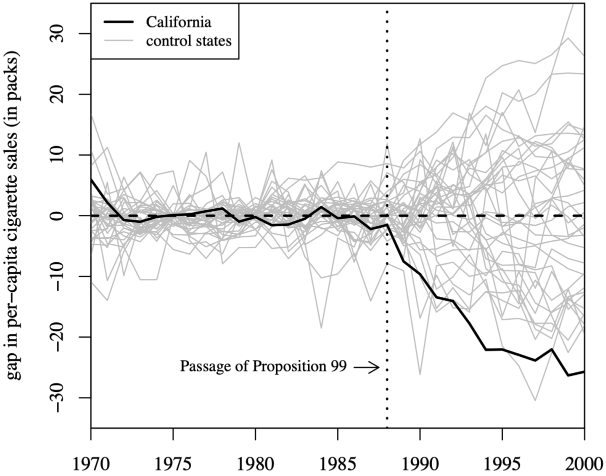
- Some states whose pre-treatment RMSPE is considerably different than California's can be dropped.

Inference: Dropping Sample



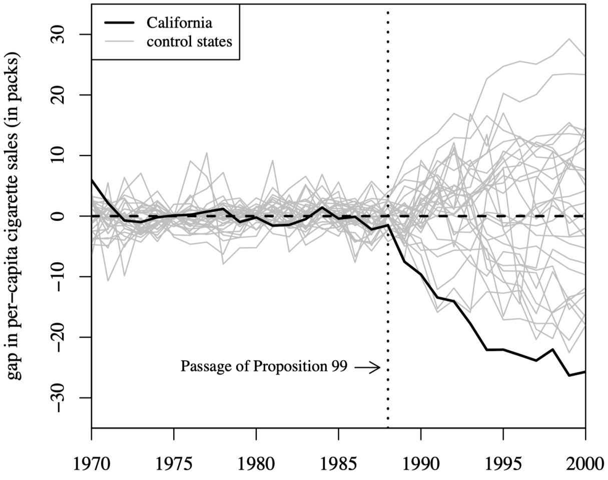
Inference: Dropping Sample

(Pre-Prop. 99 MSPE $\leq$ 20 Times Pre-Prop. 99 MSPE for CA)



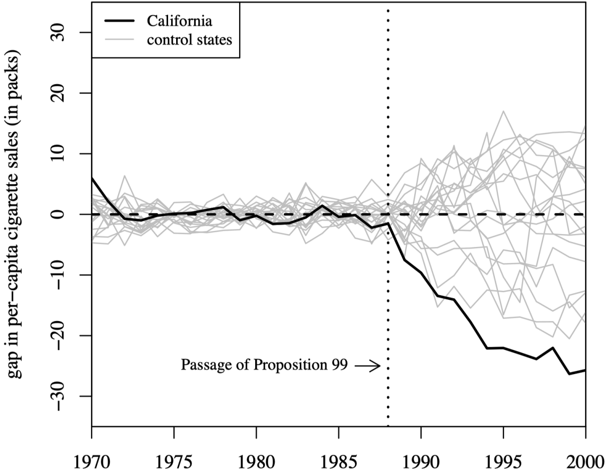
Inference: Dropping Sample

(Pre-Prop. 99 MSPE $\leq$ 5 Times Pre-Prop. 99 MSPE for CA)



Inference: Dropping Sample

(PRE-PROP. 99 MSPE $\leq$ 2 TIMES PRE-PROP. 99 MSPE FOR CA)



Inference: Procedure

1. Iteratively apply the synthetic method to each state in the unaffected group and obtain a distribution of placebo effects.
2. Calculate the RMSPE(root mean squared prediction error) for *each placebo* for the pre-treatment and post-treatment.

- Post-treatment $R_{j,post} = RMSPE_j(T_0 + 1, T)$
- Pre-treatment $R_{j,pre} = RMSPE_j(1, T_0)$

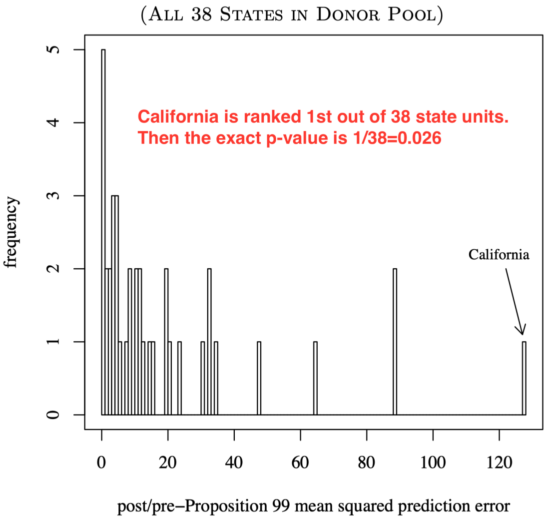
3. Compute the ratio of the post-to-pre-treatment and sort it in descending order from greatest to highest. Thus

$$r_j = \frac{R_{j,post}}{R_{j,pre}}$$

4. The exact p-value is defined as

$$p - value = \frac{rank_{th}}{J + 1}$$

Inference: P-Value

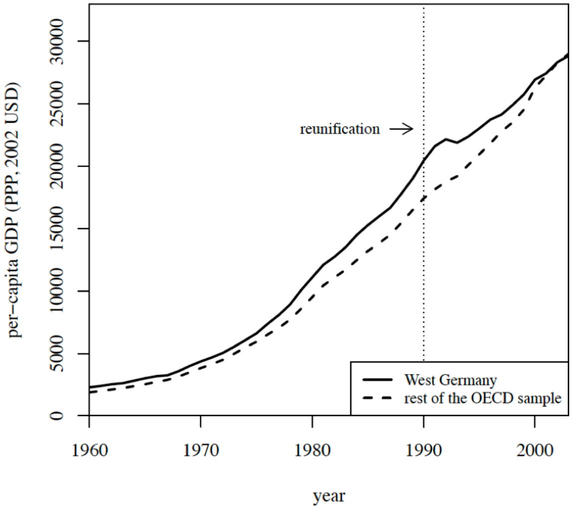


An Application: The 1990 German Reunification

The Economic Effect of the German Reunification

- Cross-country regressions are often criticized because they put side-by-side countries of very different characteristics.
  - "What do Thailand, the Dominican Republic, Zimbabwe,Greece and Bolivia have in common that merits their being put in the same regression analysis? Answer: For most purposes, nothing at all." (Harberger 1987)
- Application: The economic effect of "Berlin Wall" Falling,thus the 1990 German reunification,on West Germany.
- Control group is compositional restricted to 16 OECD countries

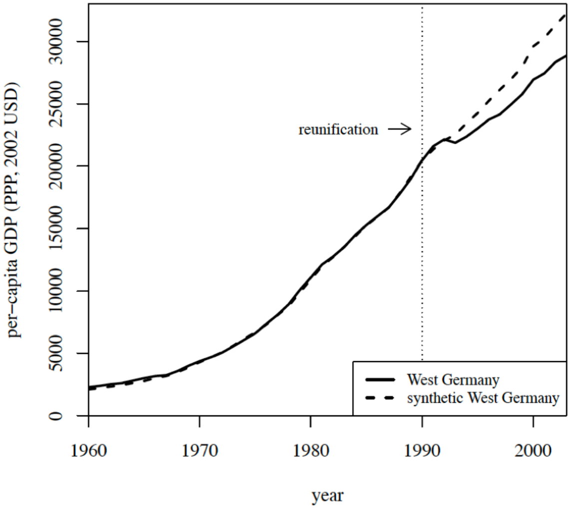
West Germany v.s. OECD countries



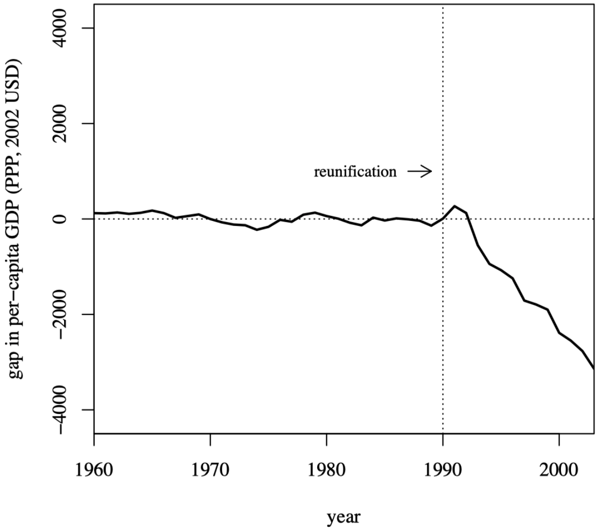
Economic Growth Predictors Means across groups

	West Germany	Synthetic West Germany	OECD Sample
GDP per-capita	15808.9	15800.9	8021.1
Trade openness	56.8	56.9	31.9
Inflation rate	2.6	3.5	7.4
Industry share	34.5	34.4	34.2
Schooling	55.5	55.2	44.1
Investment rate	27.0	27.0	25.9

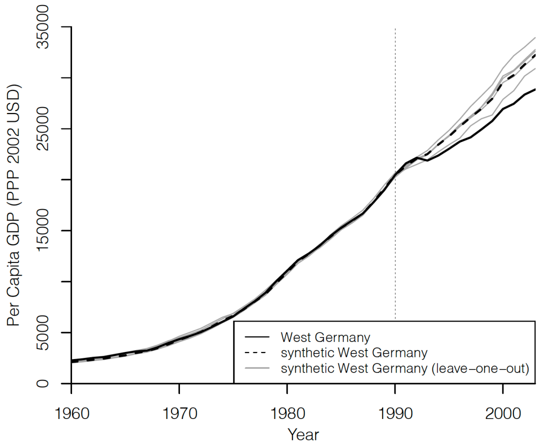
West Germany v.s Sythetic West Germany



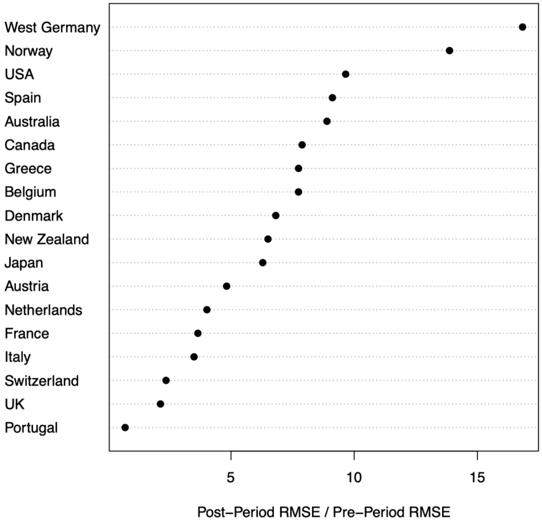
GDP Gap: West Germany and synthetic West Germany



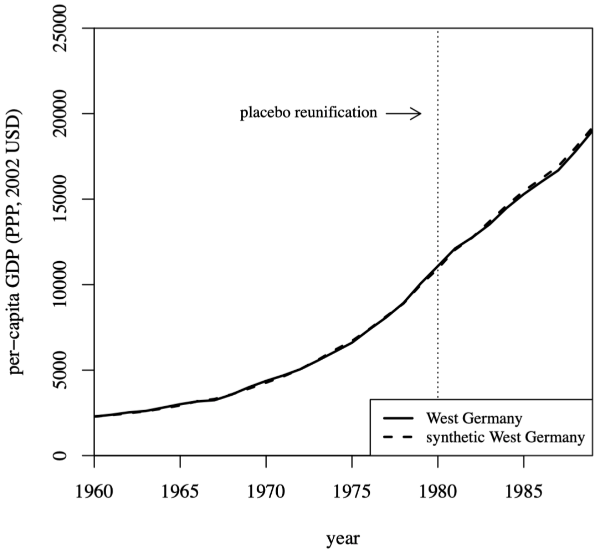
The 1990 German Reunification: Leave-one-out estimates



RMSE Test


West Germany
Norway
USA
Spain
Australia
Canada
Greece
Belgium
Denmark
New Zealand
Japan
Austria
Netherlands
France
Italy
Switzerland
UK
Portugal
5
10
15
Post−Period RMSE / Pre−Period RMSE

Placebo Test: What if '1980' German Reunification



Wrap Up

- **Synthetic control method** provide many practical advantages for causal inference.
- The credibility of the results depends on
  - the level of diligence exerted in the application
  - whether contextual and data requirements are met

A Summary of Causal Inference Method

The goal of causal inference

- Build a reasonable counterfactual world by naturally occurring data to find or construct a proper control group is the core of econometrical methods.
- Common Idea: match similar units, and produce a proper comparison
  - OLS: gives conditional mean comparison
  - Matching: a weighted conditional mean comparison
  - IV: compares difference between instrumented and non-instrumented groups.
  - RD: compares means around the cutoff.
  - DID: compares the changes of the difference across locations.
  - SCM: compares the gaps between treated and sythetic control groups.
- All are about a a **believable** and **reliable** comparison.

Final Thoughts(Angrist and Pischeke,2008)

- A good research design is one you are excited to tell people about
  - that's basically what characterizes all research designs, whether **instrumental variable**,**regression discontinuity designs** or **difference-in-differences**,**synthetic control method** among others(**Seven Magic Weapons**).
- Causality is *easy and hard*. Don't get confused which is the hard part and which is the easy part.
- Always understand *what assumptions you must make*, be clear which parameters you are and are not identifying.
- Last but not least, Remember: **Good question is always the first priority**. Along with good research design is in the second place.
- What is a good research question?
  - interesting(people cares) and/or relevent(does matter something)
  - should not simply duplicate existing research, but instead should aim to be innovative and unique.

Though still a long way to go but now we could take a break and enjoy the landscape.