Lecture 11: Difference-in-Differences and Extensions

Introduction to Econometrics,Fall 2023

Zhaopeng Qu

Nanjing University Business School

May 31 2023

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.

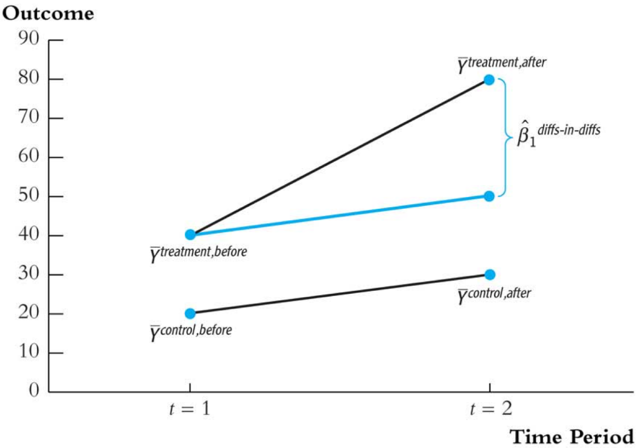
RCT and Difference in Differences

- 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

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



Card and Krueger(1994): Minimum Wage on Employment

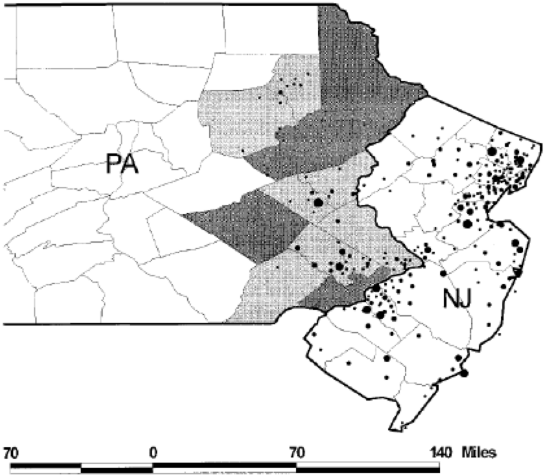
Introduction

- 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

- 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.

Card & Krueger(1994): Geographic Background



Card & Krueger(1994): Model Graph

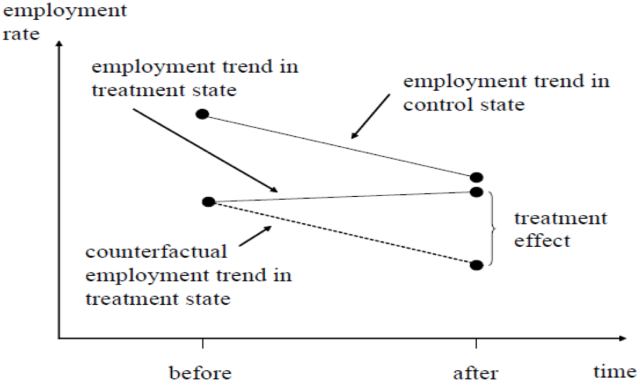


Figure 5.2.1: Causal effects in the differences-in-differences model

Card & Krueger(1994):Result

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, all available observations	23.33 (1.35)	20.44 (0.51)	-2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	-0.14 (1.07)
3. Change in mean FTE employment	-2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

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

Regression DD - Card and Krueger

- DID model:

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

  - NJ is a dummy equal to 1 if the observation is from NJ,
  - otherwise equal to 0(from Penny)
  - d is a dummy equal to 1 if the observation is from November (the post period),
  - otherwise equal to 0(Feb. the pre period)
- Which estimate coefficient does present 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

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

  - $Treat_s$ is a dummy variable indicate whether or not is **treated**.
  - $Post_t$ 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**: *Outcomes are often measured at the individual level i,while treatment takes place at the group levels.*(The S.E. has to be adjusted in “cluster” way).

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.
- 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.

TWFE: 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

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

  - α_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**.

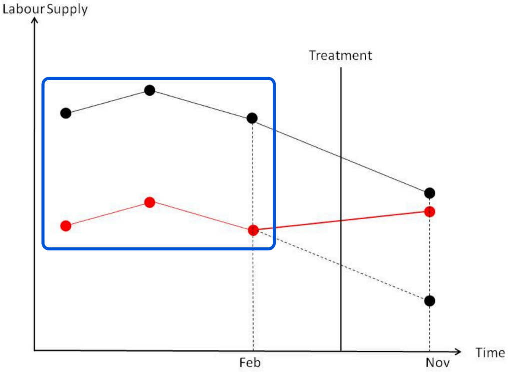
Key Assumption For DID

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

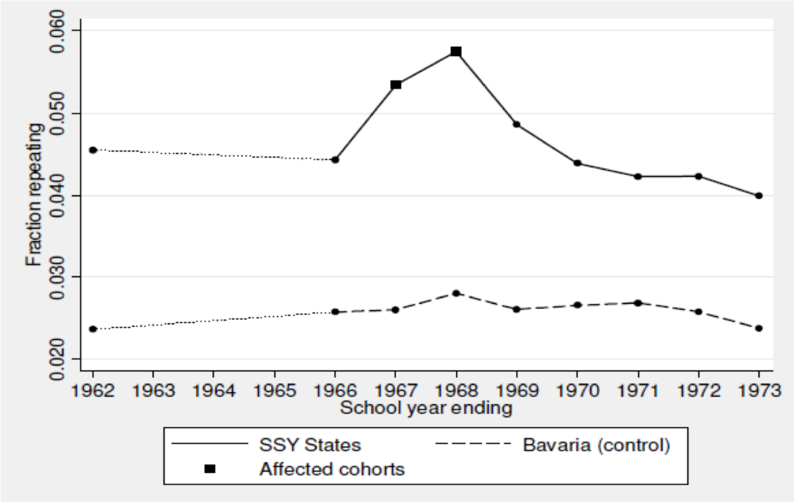
- **Common Trend**: It is difficult to verify but one often uses pre-treatment data to show that the trends are the same.
    - If you only have two-period data, you can do **nothing**.
    - If you luckly have multiple-period data, then you can show something graphically.



An Encouraging Example: Pischeke(2007)

- 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**.

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.

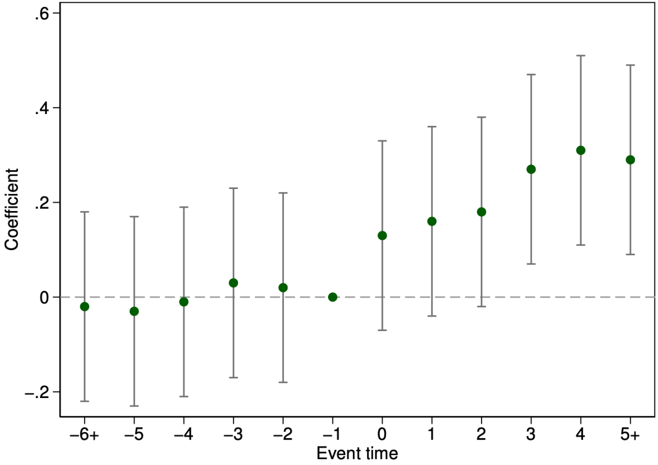
The Event Study Design: Including Leads and Lags

- 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

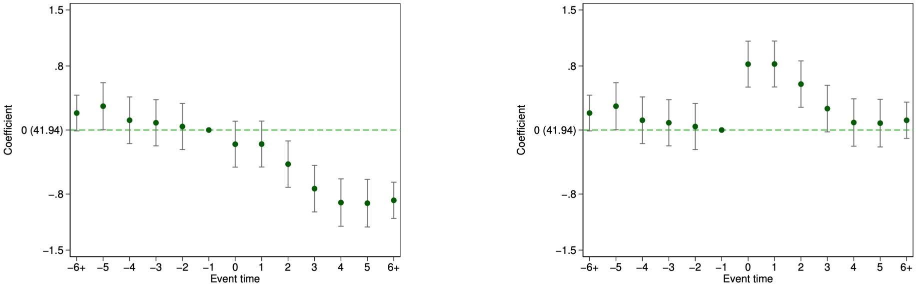
$$Y_{its} = \alpha_s + \delta_t + \sum_{\tau=-q}^{-1} \theta_\tau D_{st} + \sum_{\tau=0}^{p} \delta_\tau D_{st} + X_{ist} + u_{its}$$

- Treatment occurs in year 0
- Includes q leads or anticipatory effects
- Includes p leads or post treatment effects

The Event Study Design: Including Leads and Lags



The Event Study Design: Including Leads and Lags



(a) "Smooth" event-time trend

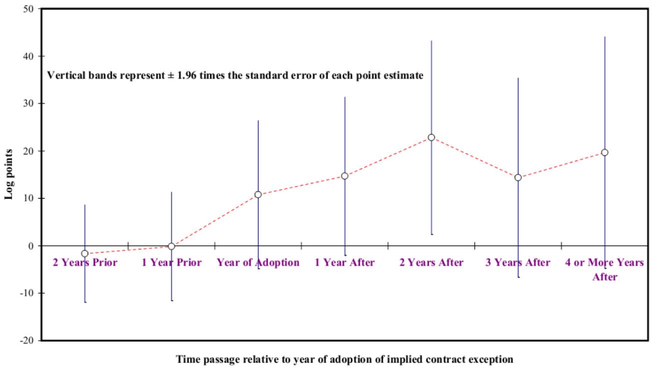
(b) "Jump" at the time of the event

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.

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.
- 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.

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 when the pretreatment data establish a clear trend that can be extrapolated into the posttreatment period.

Add group-specific time trends

- Besley and Burgess (2004),"Can Labor Regulation Hinder Economic Performance? Evidence from India", *The Quarterly Journal of Economics*.
    - 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)

- Jonathan Gruber (1994), “The Incidence of Mandated Maternity Benefits”, *American Economic Review*
  - 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 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:	−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:	−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)**		

More Extensions

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 “80s,” “90s,” or “00s.”
- 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

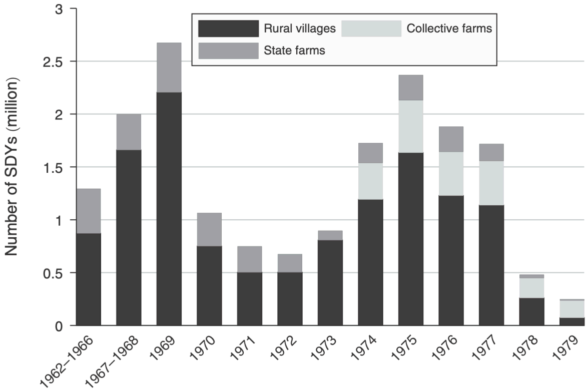


FIGURE 1. NUMBER OF SDYS BY RESETTLEMENT, 1962–1979

Source: Gu (2009)

Sent-Down Movement and Rural Education in China

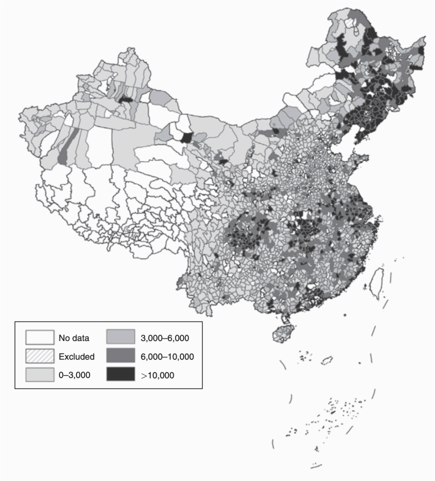


FIGURE 2. NUMBER OF RECEIVED SDYS IN EACH COUNTY

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.
- $\mathbf{X}_{i,g,c,p}$ 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

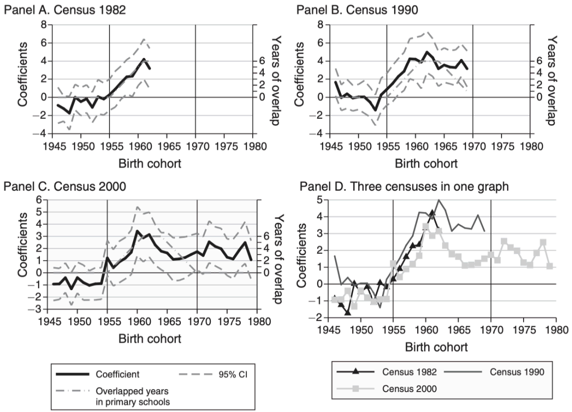


FIGURE 3. EFFECT OF SDYS ON THE EDUCATIONAL ATTAINMENT OF DIFFERENT COHORTS

Note: The left *y*-axis represents the coefficients from equation (1), which captures the effect of SDYs densities on different cohorts.

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	0.151	0.441	−0.0658	0.767	−0.0517		
	(0.701)	(0.517)	(0.0873)	(0.0611)	(0.121)	(0.103)		
Local density of received SDYs × affected cohorts (placebo)							−0.817	−0.432
							(0.576)	(0.319)
Male	1.874	0.668	0.201	0.0319	0.203	0.0546	2.286	0.665
	(0.0284)	(0.0256)	(0.00361)	(0.00227)	(0.00285)	(0.00316)	(0.0300)	(0.0150)
Han ethnic	0.150	3.34e-05	0.0213	0.00962	0.00657	0.0177	0.0802	0.477
	(0.0565)	(0.0811)	(0.00769)	(0.00540)	(0.00679)	(0.00875)	(0.0554)	(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 for different treatment intensity

Card(1992): Minimum Wage on Employment

- Study treatments with different treatment intensity. (e.g., varying increases in the minimum wage for different states)'
- Card(1992) exploits regional variation in the impact of the federal minimum wage.
- **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.

Card(1992): DID for different treatment intensity

- It can also use the first-differences equivalence

$$\Delta Y_s = \gamma + \beta(Intense_s) + u_s$$

- Where ΔY_s is a measure of the change in teen employment of state i, from 1989 to 1990.
- Then base on simple OLS formula

$$\beta = \frac{Cov(E_i, \Delta Y_i)}{Var(E_i)}$$

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 aslo 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.
- Assume that we have following two-way fixed effects regression

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

  - $Y_i t$ is the outcome.
  - $S_i t$ is the endongenous variable.
  - α_i and τ_t are entity-fixed effect and time-fixed effect respectively.
- **Potential bias** of $\hat{\beta}$?
  - some unobservable and varing factors could be omitted into ϵ_{it}

Introduction

- 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.

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.
- Although a *negative* relationship has been widely observed, the cross-sectional association cannot be interpreted as the causal effect of quantity on quality.
    - OVB
    - simultanouse bias

One-Child Policy as a natural experiment

- Plausibly exogenous variation in family size
  - the natural occurrence: twin births or the sibling sex composition,
  - fertility policy: one-child policy
- One-child policy(OCP) as a exogenous policy to the nubmer of children.
  - The OCP formally implemented in 1980 has varied significantly between rural and urban areas, over time, and across provinces, ethnicities, and even entities.
- *"Does population control lead to better child quality? Evidence from China's one-child policy enforcement", Journal of Comparative Economics 45 (2017), by Bingjing Li and Hongliang Zhang*
- 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.

DID: One-Child Policy on Family Size

- DID regression model on the treatment is

$$\text{FamilySize}_{ijt} = (EFR_j \times T_i)\,\alpha_1 + \boldsymbol{X_i}\gamma_1 + (\boldsymbol{C_j} \times T_i)\,\delta_1 + \phi_j + \lambda_t + u_{ijt},$$

- where $FamilySize_{ijt}$ is the family size of firstborn child i from prefecture j in census year t;
- T_i is a dummy that equals 1 if the child belongs to the 1990 sample;
- 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, mother's education level, father's education level, mother's employment sector, and father's employment sector;
- C_j is a vector of prefecture-specific control variables that account for pre-existing fertility preferences and socio-economic characteristics;
- ϕ_j and λ_t are the prefecture and census year fixed effects, respectively. and an interaction term $\boldsymbol{C_j} \times T_i$ to net out regional EFR differences attributable to their differences in pre-existing fertility preferences and socio-economic characteristics.

DID: One-Child Policy on Education

- The DID regression model on the outcome

$$Y_{ijt} = \left(EFR_j \times T_i\right)\alpha_2 + \boldsymbol{X_i}\gamma_2 + \left(\boldsymbol{C_j} \times T_i\right)\delta_2 + \phi_j + \lambda_t + u_{ijt}$$

- 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**.

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	4.045***	−0.482**	−0.539***	5.660***	−0.397†	−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

$$y_{ijt} = \text{FamilySize}_{ijt}\beta + \boldsymbol{X_i}\pi + \left(\boldsymbol{C_j} \times T_i\right)\eta + \boldsymbol{\phi}_j + \lambda_t + \varepsilon_{ijt}$$

- A DDIV-2SLS model

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

  - 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) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 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×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× year level are reported in parentheses. [5] *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Paralled Trends

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

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

  - 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

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

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

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

Paralled Trend(I)

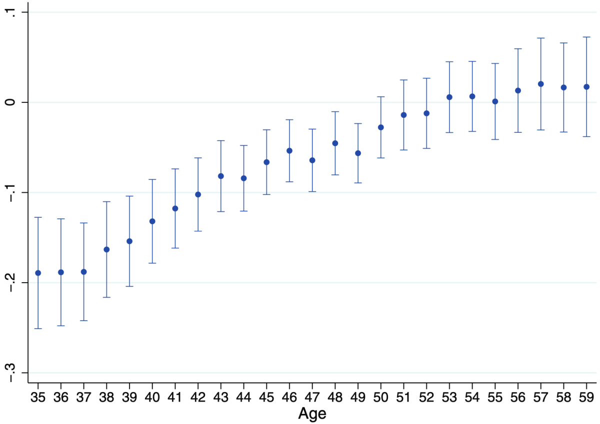


Fig. 1. EFR and intercensus change in fertility by women's age.
Notes: The figure displays the estimated coefficients and 95% confidence intervals of θ_l of Eq. (2) using all Han females aged 33–57 in the 1982 and 1990 Censuses.

Paralled Trend(II)

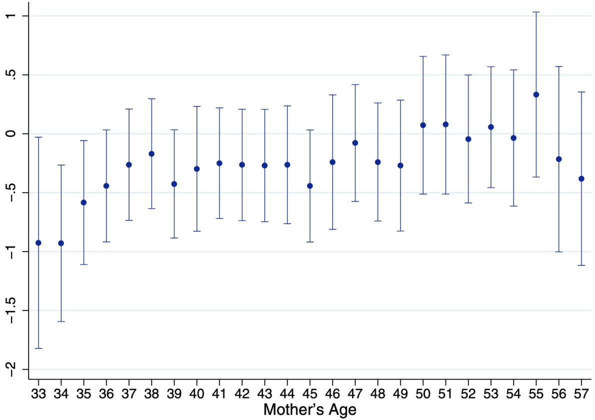


Fig. 2. EFR and intercensus change in children's education level by mother's age.
Notes: The figure displays the estimated coefficients and 95% confidence intervals of δ_l of Eq. (4) using 14- to 17-year-old Han children with mothers aged between 33 and 57 in the 1982 and 1990 Censuses.

DID with RDD

Terrorism on Individual Wellbeing

- **Title**: The Impact of Terrorism on Individual Well-Being: Evidence from the Boston Marathon Bombing,The Economic Journal 2020, by Clark,Doyle and Stancanelli.
- **Topic**: Terrorism on Individual Wellbeing.
- **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
  - Negative Affect

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.

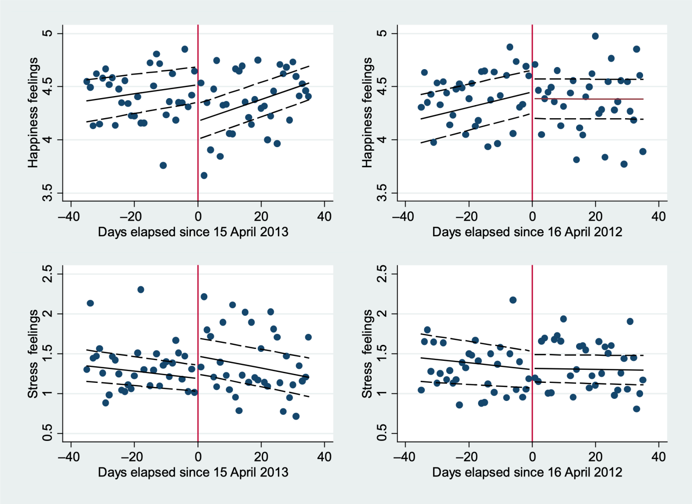
Terrorism on Individual Wellbeing: RDD-DID

- The combination of the RDD with the DID

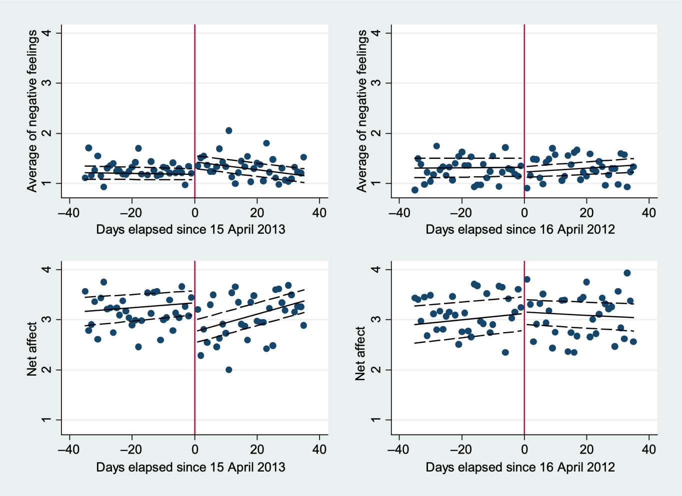
$$
\begin{aligned}
W_{it} = & \xi T_i \times \text{ Year }_t + \delta f\left(D_t\right) \times T_i \times \text{ Year }_i + \rho f\left(D_t\right) \times \left(1 - T_i\right) \times \text{Year}_t + \alpha f\left(D_i\right) \times T_i \\
& + \eta f\left(D_i\right) \times \left(1 - T_i\right) + \omega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

- 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**
Bandwidth 35 days, 2012 and 2013 data, standard errors are robust but not clustered	(0.245)	(0.253)	(0.172)	(0.345)
Observations	4,366	4,396	4,341	4,316
R^2	0.062	0.068	0.069	0.063

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.

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

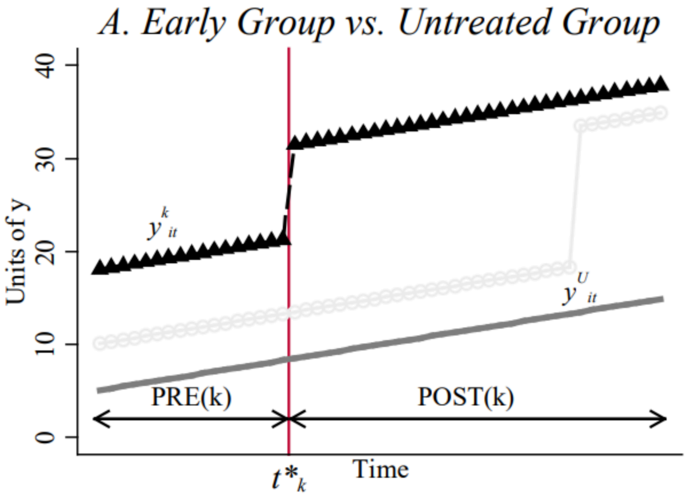
DID with multiple treatment times

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

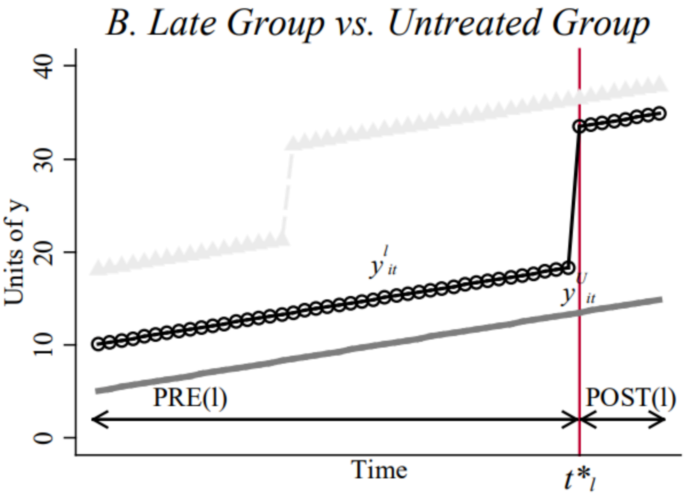
$$Y_{ist} = \alpha + \beta D_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

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

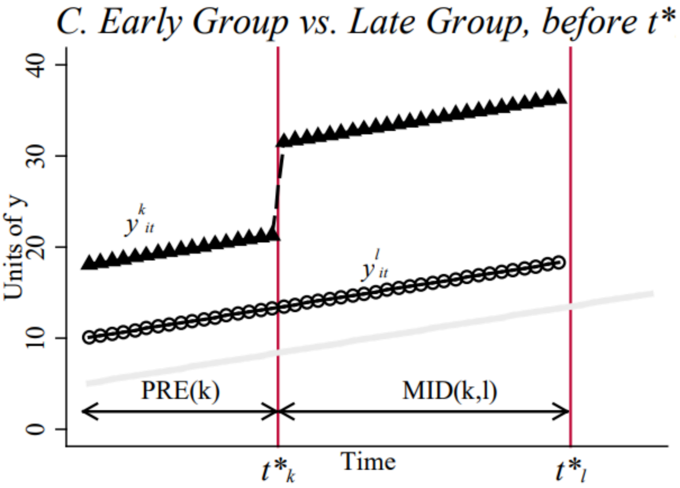
DID with multiple treatment times



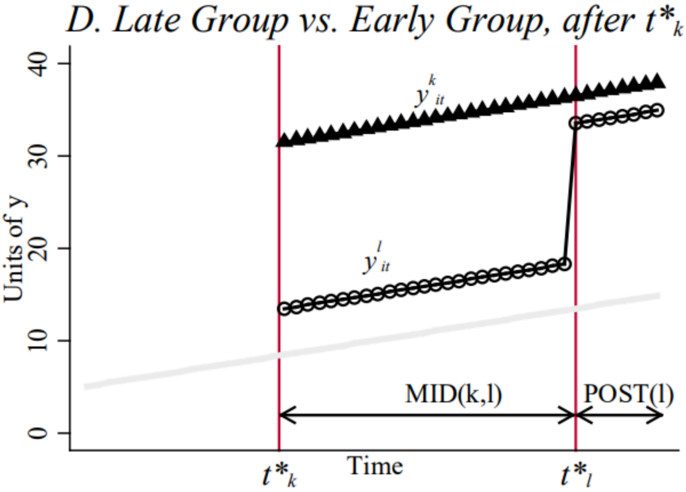
DID with multiple treatment times



DID with multiple treatment times



DID with multiple treatment times



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!
- Some weighted balance test to make sure this isn't a problem.
    - eg. Bacon decomposition by Goodman-Bacon (2021) and many others

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.