Difference-in-Differences II: Extensions

Lingguo Cheng

Nanjing University

December 15, 2025

Outline

1. Extension 1: Incorporate Covariates, X_{ist}
2. Extension 2: Non-binary Treatment
3. Extension 3: Not just restricted to calendar year, Cohort DD
4. Extension 4: From 2×2 to N×T
5. Extension 5: Heterogeneous treatment effect and event study
6. Extension 6: Difference-in-difference-in-differences(DDD) strategy
7. Extension 7: Placebo test

Extension 1: Incorporate Covariates, X_{ist}

- It's easy to add additional covariates in this framework. In other words, we can model counterfactual employment in the absence of a change in the minimum wage as (what is the philosophy behind adding X):

$$E\left(Y_{0ist} \mid s, t, X_{ist}\right) = \alpha_s + \lambda_t + X_{ist}'\beta$$

$$E\left(Y_{1ist} \mid D = 1; s, t, X_{st}\right) = \alpha_s + \lambda_t + \delta + X_{ist}'\beta$$

$$Y_{ist} = \alpha_s + \lambda_t + \delta D_{st} + X_{ist}'\beta + \varepsilon_{ist}$$

- Variables in X_{ist} could be individual-level variables or time-varying variables at the state level.

Addition of state-specific time trend

- Adds state-specific time trends to the regressors in X_{ist}.

$$Y_{ist} = \beta D_{s,t} + X_{ist}\delta + \alpha_s + \gamma_s t + \lambda_t + \varepsilon_{ist}$$

  Where α_s is a state-specific intercept, and γ_s is a state-specific trend coefficient multiplying the time trend variable, t. This allows treatment and control states to follow different trends in a limited but potentially revealing way.

- **Province-specific more general time trend**

$$Y_{ispt} = \beta D_{s,t} + X_{ispt}\delta + \alpha_s + \lambda_t + \gamma_p t + \varepsilon_{ispt}$$

  Where p indexes province, and s, say, prefecture; as an alternative to state-prefecture trend, imperfect but sometimes acceptable (s is affiliated to p, then number of p is smaller than that of s).

Technical Note

- Technically, how to incorporate a linear time trend in panel data analyses using Stata?
  - Method 1: Using a baseline year
    gen t = year - 2000 // 2000 is the baseline year
  - Method 2: Creating a sequence within each panel unit sort id year
    bysort id: gen t = _n
  - Method 3: Using the calendar year itself (simply use the 'year' variable as the trend)

Or, more generally

- **State-specific linear trend (Typical)**

$$Y_{ist} = \beta D_{s,t} + X_{ist}\delta + \alpha_s + \lambda_t + \gamma_s t + \varepsilon_{ist} \quad \text{(Version 1)}$$

- **State-time fixed effects**

$$Y_{ist} = \beta D_{s,t} + X_{ist}\delta + \gamma_s + \lambda_t + \gamma_s * \lambda_t + \varepsilon_{ispt}$$

or

$$Y_{ist} = \beta D_{s,t} + X_{ist}\delta + \gamma_s + \lambda_t + \phi_{st} + \varepsilon_{ispt}$$

More strict than version 1; ideal in theory, but often infeasible.

Or, more complex

控制其它项目

- Controlling the confounding program, P_{st}

$$Y_{ist} = \beta D_{s,t} + \alpha P_{st} + X_{ist}\delta + \gamma_s + \lambda_t + \varepsilon_{ist}$$

宽带中国项目　电子下乡项目等…

  - See Duflo (2003) for a good example.

- To conserve the degrees of freedom, Z_s can be a time-invariant continuous covariate at s level

节约自由度

$$Y_{ist} = \beta D_{s,t} + X_{ist}\delta + \gamma_s + \lambda_t + Z_s t + \varepsilon_{ist}$$

- An ultimate control form: combining together but avoid duplicate or redundant controls

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

Notes: Adapted from Besley and Burgess (2004), table IV. The table reports regression DD estimates of the effects of labor regulation on productivity. The

- Besley and Burgess (2004) study the effects of labor regulation on businesses in Indian states
- The dependent variable is log manufacturing output per capita. All models include state and year effects. Robust SE clustered at the state level.
- Apparently, labor regulation in India increased in states where output was declining anyway. Control for this trend therefore drives the estimated regulation effect to zero.

Outline

1. Extension 1: Incorporate Covariates, X_{ist}
2. Extension 2: Non-binary Treatment
3. Extension 3: Not just restricted to calendar year, Cohort DD
4. Extension 4: From 2×2 to N×T
5. Extension 5: Heterogeneous treatment effect and event study
6. Extension 6: Difference-in-difference-in-differences(DDD) strategy
7. Extension 7: Placebo test

Extension 2: Non-binary Treatment

- **Example: Card (1992)**

  Look at all state minimum wages in the United States.

- Some of these are a little higher than the federal minimum (which covers everyone regardless of where they live), some are a lot higher, and some are the same. The minimum wage is therefore a variable with differing "treatment intensity" across states and over time.

- Moreover, in addition to statutory variation in state minima, the local importance of a minimum wage varies with average state wage levels. For example, the early-1990s federal minimum of $4.25 was probably irrelevant in Connecticut—with high average wages—but a big deal in Mississippi.

Card (1992) - Minimum Wage Study

- **Card (1992)** exploits regional variation in the impact of the federal minimum wage. His approach is motivated by an equation like:

$$Y_{ist} = \gamma_s + \lambda_t + \beta(FA_s \cdot d_t) + \varepsilon_{ist}$$

- The variable FA_s is a measure of the fraction of teenagers likely to be affected by a minimum wage increase in each state, more specifically, the fraction of workers who are paid less than \$3.80 just before the increase of the minimum wages (pre-increase).
- d_t is a dummy for observations after April 1990, when the federal minimum increased from \$3.35 to \$3.80.
- Since there are still only two time periods in the Card (1992) setup, the equation can be differenced over time to obtain:

$$\Delta Y_{st} = \lambda^* + \beta FA_s + \Delta\varepsilon_{st}$$

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

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

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

Outline

1. Extension 1: Incorporate Covariates, X_{ist}
2. Extension 2: Non-binary Treatment
3. Extension 3: Not just restricted to calendar year, Cohort DD
4. Extension 4: From 2×2 to N×T
5. Extension 5: Heterogeneous treatment effect and event study
6. Extension 6: Difference-in-difference-in-differences(DDD) strategy
7. Extension 7: Placebo test

Extension 3: Not just restricted to calendar year, Cohort DD

- In the typical DD setting, T is by default, set to "Calendar year"
- But many times, other dimensions can be possible. One common example is the cohort. 出生年份
- When the cohort takes the place of the time dimension, DD can be adopted even with a cross-sectional dataset

long term effect of early-life exposure

Example 2: Qian (2008)

- **Empirical Strategy**

$$\text{sex}_{ic} = (tea_i \times post_c)\beta + (orchard_i \times post_c)\delta + (cashcrop_i \times post_c)\rho + Han_i\zeta + \alpha + \varphi_i + \gamma_c + \varepsilon_{ic} + \gamma_{ic}.$$

$post_c$ is a dummy variable that indicates if an individual is born after 1979, that is

$$post_c = \begin{cases} 1 & \text{if individuals are born after 1979} \\ 0 & \text{otherwise} \end{cases}$$

  - sex_{ic}, the fraction of male in county i, cohort c is a function of: the interactions between $post_c$ and tea_i, the amount of tea planted for each county i; $orchard_i$, the amount of orchard planted for each county i; $cashcrop_i$, the amount of cash crops planted for each county i, respectively; Han_{ic}, the fraction that is ethnically Han; φ_i, county fixed effect; γ_c, cohort fixed effect
  - The reference group is composed of individuals born during 1970-1979. It and all its interaction terms are dropped.

TABLE III

OLS AND 2SLS ESTIMATES OF THE EFFECT OF PLANTING TEA AND ORCHARDS ON SEX RATIOS CONTROLLING FOR COUNTY LEVEL LINEAR COHORT TRENDS

	Dependent variables					
	Fraction of males			Tea × post	Fraction of males	
	(1) OLS	(2) OLS	(3) OLS	(4) 1st	(5) IV	(6) IV
Tea × post	−0.012	−0.013	−0.012		−0.072	−0.011
	(0.007)	(0.006)	(0.005)		(0.031)	(0.007)
Orchard × post	0.005					
	(0.002)					
Slope × post	−0.002			0.26		
	(0.002)			(0.057)		
Linear trend	No	No	Yes	Yes	No	Yes
Observations	28,349	37,756	37,756	37,756	37,756	37,756

Notes. Coefficients of the interactions between dummies indicating whether a cohort was born post-reform and the amount of tea planted in the county of birth. All regressions include county and birth year fixed effects and controls for Han, and cashcrop × post. All standard errors are clustered at the county level. In column (1), the sample includes all individuals born during 1970–1986. In columns (2)–(6), the sample includes all individuals born during 1962–1990. Post = 1 if birthyear > 1979. Data for land area sown are from the 1997 China Agricultural Census.

- When treatment is suspicious to be exogeneous??
- More generally, the exogeneity of the treatment assignment or treatment intensity should not be taken for granted, it should be justified instead.

Example 3: Duflo, Esther (2001)

- The questions of whether investments in infrastructure can cause an increase in educational attainment, and whether an increase in educational attainment causes an increase in earnings are basic concerns for development economists.
- This paper exploits a dramatic change in policy to evaluate the effect building schools has on education and earnings in Indonesia.
- Indonesian children normally attend primary school between the ages of 7 and 12. All children born in 1962 or before were 12 or older in 1974, when the first INPRES schools were constructed. Thus, they did not benefit from the program, since they should have left primary school before the first INPRES schools were opened.
- For younger children, the exposure is an increasing function of their date of birth. Hence, the effect of the program should be close to 0 for children 12 or older in 1974 and increasing for younger children.

Identification strategy

- The program intensity was related to enrollment rates in 1972, which differed widely across regions. Region of birth is a second dimension of variation in the intensity of the program.
- The regression function is as follows

$$S_{ijk} = c + \alpha_j + \beta_k + P_j T_i \gamma + (C_j T_i)\delta + \varepsilon_{ijk}$$

- where S_{ijk} is the education of individual i born in region j in year k, c is a constant, α_j is a district of birth fixed effect, β_k is a cohort of birth fixed effect, T_i is a dummy indicating whether the individual belongs to the "young" cohort in the subsample, P_j denotes the intensity of the program in the region of birth, and C_j is a vector of region-specific variables.

Table 3—Means of Education and Log(Wage) by Cohort and Level of Program Cells

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

TABLE 4—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN COHORT DUMMIES AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN THE REGION OF BIRTH

		Dependent variable					
		Years of education			Log(hourly wage)		
	Observations	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Experiment of Interest: Individuals Aged 2 to 6 or 12 to 17 in 1974							
(Youngest cohort: Individuals ages 2 to 6 in 1974)							
Whole sample	78,470	0.124	0.15	0.188			
		(0.0250)	(0.0260)	(0.0289)			
Sample of wage earners	31,061	0.196	0.199	0.259	0.0147	0.0172	0.0270
		(0.0424)	(0.0429)	(0.0499)	(0.00729)	(0.00737)	(0.00850)
Panel B: Control Experiment: Individuals Aged 12 to 24 in 1974							
(Youngest cohort: Individuals ages 12 to 17 in 1974)							
Whole sample	78,488	0.0093	0.0176	0.0075			
		(0.0260)	(0.0271)	(0.0297)			
Sample of wage earners	30,225	0.012	0.024	0.079	0.0031	0.00399	0.0144
		(0.0474)	(0.0481)	(0.0555)	(0.00798)	(0.00809)	(0.00915)
Control variables:							
Year of birth*enrollment rate in 1971		No	Yes	Yes	No	Yes	Yes
Year of birth*water and sanitation program		No	No	Yes	No	No	Yes

Notes: All specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummies and the number of children in the region of birth (in 1971). The number of observations listed applies to the specification in columns (1) and (4). Standard errors are in parentheses.

Outline

Extension 4: From 2×2 to N×T

- Transformation:

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s^* d_t) + \varepsilon_{ist}$$
$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

where $E(\varepsilon_{ist} \mid s, t) = 0$; D_{st}, is a policy dummy that is defined to be unity for groups and time periods subject to the policy.

$$D_{st} = \begin{cases} 1, & \text{if } t = 2 for NJ \\ 0, & \text{otherwise} \end{cases}$$

ind	year	event_year	year_till_event	Yit	a_sh	bt	as_bt	Xit
BJ	2020				0	0	0	
BJ	2022				0	1	0	
SH	2020	2021	-1		1	0	0	
SH	2022	2021	1		1	1	1	

ind	year	event_year	ar_till_eve	Yit	a_sh	bt	as_bt	dst	Xit
BJ	2020				0	0	0	0	
BJ	2022				0	1	0	0	
SH	2020	2021	-1		1	0	0	0	
SH	2022	2021	1		1	1	1	1	

From 2×2 to N×T

- Transformation:

$$y_{ist} = \alpha + \gamma_s + \lambda d_t + \delta(\gamma_s \times d_t) + \varepsilon_{ist}$$
$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

where $E(\varepsilon_{ist} \mid s,t) = 0, N > 2,$ and $T > 2$; D_{st}, is a policy dummy that is defined to be unity for groups and time periods subject to the policy.

$$D_{st} = \begin{cases} 1, & \text{if } t \geq \tau \text{ for state } s, \tau \text{ is the time that policy implements} \\ 0, & \text{otherwise} \end{cases}$$

- A seeming irrelevant but essential transformation: a great leap from 2×2 to N×T
- What's the loss? This restricts that the policy has the same effect in every year (Jeffrey M. Wooldridge, 2010).

ind	year	event_year	year_till_event	Yit	a_sh	a_gz	b_2022	b_2023	a_sh*b_2022	a_sh*b_2023	a_gz*b_2022	a_gz*b_2023	Dst
BJ	2020				0	0	0	0	0	0	0	0	0
BJ	2022				0	0	1	0	0	0	0	0	0
BJ	2023				0	0	0	1	0	0	0	0	0
SH	2020	2021	-1		1	0	0	0	0	0	0	0	0
SH	2022	2021	1		1	0	1	0	1	0	0	0	1
SH	2023	2021	2		1	0	0	1	0	1	0	0	1
GZ	2020	2021	-1		0	1	0	0	0	0	0	0	0
GZ	2022	2021	1		0	1	1	0	0	0	1	0	1
GZ	2022	2021	2		0	1	0	1	0	0	0	1	1

With more periods, we can fly

- As a start, let's consider a policy that occurs all at t_0 (e.g. single timing rolled out to treated units)
- More time periods helps in several ways:
  - If we have multiple periods before the policy implementation, we can partially test the underlying assumptions. Sometimes referred to as "pre-trends"
  - If we have multiple periods after the policy implementation, we can examine the timing of the effects
    - Is it an immediate effect? Does it die off? Is it persistent?
- If you pool all time periods together into one "post" variable, this estimates the average effect. Or alternatively, you implicitly assume a constant treatment effect over time
- If the sample is not balanced, can have unintended effects!

Outline

1. Extension 1: Incorporate Covariates, X_{ist}
2. Extension 2: Non-binary Treatment
3. Extension 3: Not just restricted to calendar year, Cohort DD
4. Extension 4: From 2×2 to N×T
5. Extension 5: Heterogeneous treatment effect and event study
6. Extension 6: Difference-in-difference-in-differences(DDD) strategy
7. Extension 7: Placebo test

Extension 5: Heterogeneous treatment effect and event study

- Heterogeneous treatment effect: q lags (or q after) - posttreatment effect

post拆开为q期。 time的异质性

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{j=0}^{q} \delta_j D_{st}(s, t = \tau + j) + X'_{st}\beta + \varepsilon_{ist}$$

- Moreover, the δ_j, $\forall j > 0$ may not be identical. For example, the effect of the treatment could accumulate over time, so that increases in j; or fades away.
- Heterogeneous treatment effect: m leads (or m before) - anticipatory effect

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{j=-m}^{-1} \delta_j D_{st}(s, t = \tau + j) + X'_{st}\beta + \varepsilon_{ist}$$

- A test of the differences assumption is $\delta_j = 0 \ \forall j < 0$, i.e. the coefficients on all leads of the treatment should be zero. This is thought of as a direct test of the parallel trend hypothesis.

Granger causality and event study

- When the sample includes many years, the regression-DD model lends itself to a test for causality in the spirit of Granger (1969).
- The Granger idea is to see whether causes happen before consequences, and not vice versa.
- Suppose the policy variable of interest, D_{st}, changes at different times in different states. In this context, Granger causality testing means a check on whether, conditional on state and year effects, past D_{st} predicts Y_{ist} while future D_{st} does not.

Granger causality and event study

- If it holds true, then dummies for future policy changes should not matter in an equation like:

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau=0}^{m} \delta_{-\tau} D_{s,t-\tau} + \sum_{\tau=1}^{q} \delta_{+\tau} D_{s,t+\tau} + X'_{ist}\beta + \varepsilon_{ist}$$

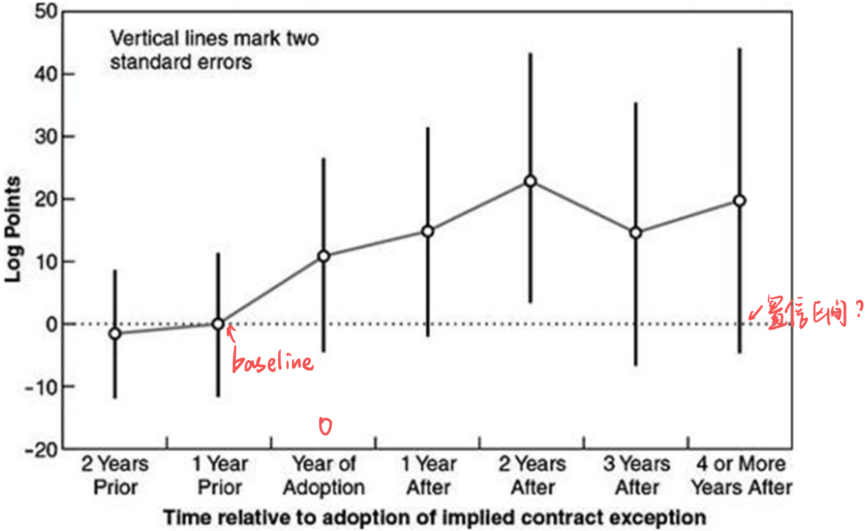
where the sums on the right-hand side allow for m lags (δ_{-1}, δ_{-2}, ..., δ_{-m}) or posttreatment effects and q leads (δ_{+1}, δ_{+2}, ..., δ_{+q}) or anticipatory effects. The pattern of lagged effects is usually of substantive interest as well. We might, for example, believe that causal effects should grow or fade as time passes.

Autor (2003)

- **The effect of employment protection on firms' use of temporary help.**
- In the US, Employment protection is a type of labor law (Particularly true in China) - promulgated by state legislatures or, more typically, through common law as made by state courts - that makes it harder to fire workers.
- As a rule, US labor law allows employment at will, which means that workers can be fired for just cause or no cause, at the employer's whim. But some state courts have allowed a number of exceptions to the employment-at-will doctrine, leading to lawsuits for unjust dismissal.

Autor (2003)

- Autor is interested in whether fear of employee lawsuits makes firms more likely to use temporary workers for tasks for which they would otherwise have increased their workforce.
- Temporary workers are employed by someone else besides the firm for which they are executing tasks. As a result, firms using them cannot be sued for unjust dismissal when they let temporary workers go.



Vertical lines mark two standard errors
Log Points
50
40
30
20
10
0
-10
-20
baseline
置信区间？
0
2 Years Prior
1 Year Prior
Year of Adoption
1 Year After
2 Years After
3 Years After
4 or More Years After
Time relative to adoption of implied contract exception

Cunningham, S. and Cornwell, C. (2013)

- **Cunningham, S. and Cornwell, C. (2013).** "The Long-Run Effect of Abortion on Sexually Transmitted Infections." *American Law and Economics Review*, 15(1):381-407.
- The design exploited the early repeal of abortion in five states in 1970 and compared those states to the states that were legalized under Roe vs. Wade in 1973.

Cunningham, S. and Cornwell, C. (2013)

- To do this, I needed cohort-specific data on gonorrhea incidence by state and year, but as those data are not collected by the CDC, I had to settle for second best.
- That second best was the CDC's gonorrhea data broken into five-year age categories (e.g., age 15-19, age 20-24). But this might still be useful because even with aggregate data, it might be possible to test the model I had in mind.
- **Literature context:**
  - **Gruber et al.(1999):** the characteristics of the marginal child aborted had that child reached their teen years. —any far-reaching effects?
  - **Donohue and Levitt (2001), Levitt (2004);** abortion legalization $\rightarrow$ decrease in crime rates

Empirical Strategy

- Our estimating equation is as follows:

$$Y_{st} = \beta_1 Repeal_s + \beta_2 DT_t + \beta_{3t} Repeals \times DT_t + X_{st}\psi + \alpha_{1s} DS_s$$
$$+ \gamma_1 t + \gamma_{2s} DS_s \cdot t + \varepsilon_{st}$$

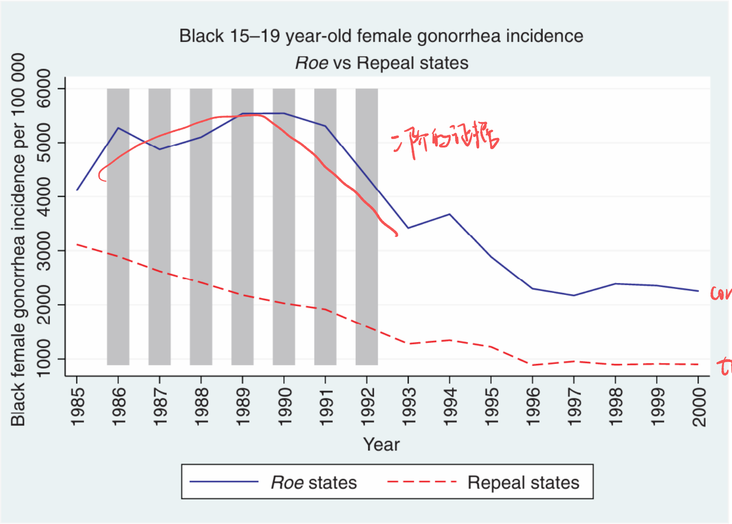
  - Y_{st} is the log number of new gonorrhea cases for 15- to 19-year-olds (per 100,000 of the population)
  - $Repeal_s$ equals 1 if the state legalized abortion prior to Roe; DT_t is a year dummy; DS_s is a state dummy; t is a time trend; X_{st} is a matrix of covariates
  - ε_{st} is a structural error term assumed to be conditionally independent of the regressors.
  - All standard errors were clustered at the state level allowing for arbitrary serial correlation.
- Is there anything wrong with the regression function?

CDC Surveillance Data in Calendar Year

Age in calendar year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
15	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
16	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
17	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
18	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82
19	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
20	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
21	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
22	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
23	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77
24	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
25	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
26	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
27	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
28	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
29	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
Number of cohorts (age 15-19) exposed, reforms in 71, 74																
Repeal (1)	0	1	2	3	4	5	5	5	5	5	5	5	5	5	5	5
No Repeal (2)	0	0	0	0	1	2	3	4	5	5	5	5	5	5	5	5
Difference (3)	0	1	2	3	3	3	2	1	0	0	0	0	0	0	0	0

Figure 1. Period–cohort diagram.

- Theoretical predictions of abortion legalization on age profiles of gonorrhea incidence.

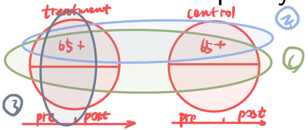


- Differences in gonorrhea incidence among black females between repeal and Roe cohorts expressed as coefficient plots.

Outline

Extension 6: Difference-in-difference-in-differences(DDD) strategy (2→N)

- In some cases, a more convincing analysis of a policy change is available by further refining the definition of treatment and control groups. For example, suppose a state implements a change in health care policy aimed at the elderly, say people 65 and older, and the response variable, y, is a health outcome.
- **DD strategy 1**: using data only on people in the state with the policy change
  - **Treatment group**: 65 and older in the state
  - **Control group**: people under 65 in the state



- **DD strategy 2**: Use another state as the control and use the elderly from the non-policy state as the control group
- **Difference-in-difference-in-differences(DDD) strategy**

- We again label the two time periods as 1 and 2, let B represent the state implementing the policy (vs. A, the states not implementing the policy), and let E denote the group of elderly (vs. N, nonelderly). Then

$$\begin{aligned} y = & \beta_0 + \beta_1 dB + \beta_2 dE + \beta_3 dB \cdot dE \\ & + \delta_0 d2 + \delta_1 d2 \cdot dB + \delta_2 d2 \cdot dE + \delta_3 d2 \cdot dB \cdot dE + u \end{aligned}$$

- The coefficient of interest is now δ_3. The OLS estimate $\hat{\delta}_3$ can be expressed as follows:

$$\begin{aligned} \hat{\delta}_3 &= [(\bar{y}_{B,E,2} - \bar{y}_{B,E,1}) - (\bar{y}_{A,E,2} - \bar{y}_{A,E,1})] - [(\bar{y}_{B,N,2} - \bar{y}_{B,N,1}) - (\bar{y}_{A,N,2} - \bar{y}_{A,N,1})] \\ &= [(\bar{y}_{B,E,2} - \bar{y}_{B,N,2}) - (\bar{y}_{B,E,1} - \bar{y}_{B,N,1})] - [(\bar{y}_{A,E,2} - \bar{y}_{A,N,2}) - (\bar{y}_{A,E,1} - \bar{y}_{A,N,1})] \end{aligned}$$

- The hope is that this controls for two kinds of potentially confounding trends: (1) changes in health status of elderly across states which would have nothing to do with the policy; (2) changes in health status of all people living in the policy-change state (possibly due to other state policies that affect everyone's health).

From DD to DDD

JS实施 P　　ZJ不实施 P.

JS (Treated =1)				ZJ (Treated =0)				
	t_0	t_1			t_0	t_1	DD	DDD
Elder	$r_{JS,E}$	$r_{JS,E}+T+T_{JS}+T_E+ATE$		Elder	$r_{ZJ,E}$	$r_{ZJ,E}+T+T_{ZJ}+T_E$		
	D	$T+T_{JS}+T_E+ATE$			D	$T+T_{ZJ}+T_E$	$(T_{JS} - T_{ZJ})+ATE$	
Youth	$r_{JS,Y}$	$r_{JS,Y}+T+T_{JS}+T_Y$		Youth	$r_{ZJ,Y}$	$r_{ZJ,Y}+T+T_{ZJ}+T_Y$		ATE
	D	$T+T_{JS}+T_Y$			D	$T+T_{ZJ}+T_Y$	$(T_{JS}-T_{ZJ})$	
	DD1	$(T_E - T_Y)+ATE$			DD2	(T_E-T_Y)		
		DD1-DD2	DDD	ATE				

T: $t_0 \to t_1$ effect

$T_{JS/ZJ}$: $t_0 \to t_1$ JS/ZJ effect

$T_{E/Y}$ $t_0 \to t_1$ elder/young effect

Table 74. Triple differences design.

States	Group	Period	Outcomes	D_1	D_2	D_3
NJ	Low-wage workers	After	$NJ_l + T + NJ_t + l_t + D$	$T + NJ_t + l_t + D$	$(l_t - h_t) + D$	D
		Before	NJ_l			
	High-wage workers	After	$NJ_h + T + NJ_t + h_t$	$T + NJ_t + h_t$		
		Before	NJ_h			
PA	Low-wage workers	After	$PA_l + T + PA_t + l_t$	$T + PA_t + l_t$	$l_t - h_t$	
		Before	PA_l			
	High-wage workers	After	$PA_h + T + PA_t + h_t$	$T + PA_t + h_t$		
		Before	PA_h			

Gruber (1994)

- Study state-level policies providing maternity benefits.
- He uses as his treatment group married women of childbearing age in treatment and control states, but he also uses a set of placebo units (older women and single men 20-40) as within-state controls.
- He then goes through the differences in means to get the difference-in-differences for each set of groups, after which he calculates the DDD as the difference between these two difference-in-differences.

Table 75. DDD Estimates of the Impact of State Mandates on Hourly Wages.

Location/year	Pre-law	Post-law	Difference
A. Treatment: Married women, 20−40yo			
Experimental states	1.547	1.513	−0.034
	(0.012)	(0.012)	(0.017)
Control states	1.369	1.397	0.028
	(0.010)	(0.010)	(0.014)
Difference	0.178	0.116	
	(0.016)	(0.015)	
Difference-in-difference	−0.062		
	(0.022)		
B. Control: Over 40 and Single Males 20−40			
Experimental states	1.759	1.748	−0.011
	(0.007)	(0.007)	(0.010)
Control states	1.630	1.627	−0.003
	(0.007)	(0.007)	(0.010)
Difference	1.09	1.21	
	(0.010)	(0.010)	
Difference-in-difference	−0.008		
	(0.014)		
DDD	−0.054		
	(0.026)		

Note: Standard errors in parentheses.

Aaron S. Yelowitz (1995)

- The regression function:

$$y_{iast} = \gamma_{st} + \lambda_{at} + \theta_{as} + \delta D_{ast} + X'_{iast}\beta + \varepsilon_{iast}$$

- Where s indexes states, t indexes time, and a is the age of the youngest child in a family. This model provides full nonparametric control for state-specific time effects that are common across age groups (γ_{st}), time-varying age effects (λ_{at}), and state-specific age effects (θ_{as}).
- The regressor of interest, D_{ast}, indicates families with children in affected age groups in states and periods where Medicaid coverage is provided. This triple-differences model may generate a more convincing set of results than a traditional DD analysis that exploits differences by state and time alone.

Outline

Extension 7: Placebo test and DDD

- Observable pre-treatment dynamics
- Event study
- First-stage evidence
- Placebo test:
  - In most cases, the start point of analysis of DD design is the ITT (Intention-to-treat) estimate (Treatment group vs. nontreated group).
  - That is, the treated group contains treated individuals (the target subject of the policy) and untreated individuals (nontarget subjects). ITT is the average treatment effect on the treated group. Further disentangling the effects on these two subgroups is interesting and important.

- Suppose ITT exists, a natural expectation is:
  - The treatment effect on the treated individuals should be much stronger—as an alternative, first-stage effect should be provided.
  - The treatment effect on the untreated individuals should be zero.
- Placebo test is increasingly emphasized in modern empirical research in economics; it is even viewed as being necessary by many researchers; neat placebo test is extremely crucial for the credibility of one research.

- Subgroup analysis
- DDD analysis: many times not a literally DDD form of DDD (see Hoynes.2016.AER) for a typical example
- Picking controls and picking different life meanings
- Economic research is partly scientific and partly artistic, an art of persuasion.
  - Pre-trends
  - Event-study
  - Placebo test

$Y_{ist} = \alpha + \gamma D_{st} + X_{ist}\beta + \varepsilon_i$

① 分组: male vs female

subgroup analysis: if

male	female	
0.08***	0.001	✓
0.08***	0.065**	?

结果不一致

② interaction:

$+ \rho \cdot D_{st} \times male$

$\rho = 0.03^{**}$ ✓

强假设: male 只在 D_{st} 上有异质性.

放松假设

③ $+ \rho \cdot D_{st} \times male + \delta \cdot X_{ist} \times male$

Inference in DD

- Anytime, keep in mind to report robust standard error clustered at group level

std error

- Sometimes, it gets tricky about this points, such as Card and Kruger (1994).
- When the number of groups is too limited, clustered SE at the group-year level may be considered
- Bootstrap standard errors often serve as the last resort. See the wild cluster bootstrap-t procedure (Cameron et al., 2008)
- Actually, in the research of China's economy, when the policy occurs in the province level (31 provincial units), it is becoming the new norm to provide bootstrapped standard errors.
- Reference: Cameron, A. C., J. B. Gelbach and D. L. Miller (2008). 'Bootstrap-based improvements for inference with clustered errors', The Review of Economics and Statistics, vol. 90(3), pp. 414–427.

Thanks for your Attention!

Any questions or comments please write to:

chenglingguo@nju.edu.cn