Difference in Differences

April 10, 2025

Motivation

- We now move on to a setting where the assignment of treatment is not necessarily uncorrelated with unit unobservable.
- In other words, when the unconfoundedness assumption were violated, can we still obtain an estimate of ATT?
- Panel data + some assumptions can potentially help!

- We start with a simple example with two periods $T = 2$.
- Expand our notation for the observed outcomes
  - $Y_{it}(0,0)$ outcome in period $t = 1,2$ if untreated both periods
  - $Y_{it}(0,1)$ outcome in period $t = 1,2$ if untreated in first period but treated in second
- Since there are only two periods and neither groups were treated in the first period, we can simplify notations slightly to $Y_{it}(0)$ and $Y_{it}(1)$.
- We can then define the average treatment effect of the treated (ATT)

$$\tau^{ATT} = E(Y_{i,2}(1) - Y_{i,2}(0)|w_i = 1)$$

- The challenge is that we do not observe the counterfactual outcome $E(Y_{i,2}(0)|w_i = 1)$ (yet the treatment is not random assigned so we can not use $E(Y_{i,2}(0)|w_i = 0)$ as its estimand.)

Two Key Assumptions

- Parallel Trends: we assume that

$$E(Y_{i,2}(0) - Y_{i,1}(0)|w_i = 1) = E(Y_{i,2}(0) - Y_{i,1}(0)|w_i = 0)$$

  which implies that the average outcome of the treated units would have evolved in parallel to the untreated units, absent treatment.

  - Consider a linear formulation $Y_{it}(0) = \gamma_t + \alpha_i + \epsilon_{it}$ where w_i can be correlated with α_i.
  - The within-unit difference is useful to eliminate α_i.

- No Anticipation: in other words we assume that $Y_{i,1}(0) = Y_{i,1}(1)$.
  - A natural implication is that $E(Y_{i,1}(0)|w_i = 1) = E(Y_{i,1}(1)|w_i = 1)$.

- Rewrite parallel trend assumption as

$$E(Y_{i,2}(0)|w_i = 1] = E(Y_{i,1}(0)|w_i = 1) + E(Y_{i,2}(0) - Y_{i,1}(0)|w_i = 0)$$

- Use the no anticipation, we have

$$E(Y_{i,2}(0)|w_i = 1] = E(Y_{i,1}(1)|w_i = 1) + E(Y_{i,2}(0) - Y_{i,1}(0)|w_i = 0)$$

- The RHS all have an empirical analog, so we have constructed an estimand for the unobserved $E(Y_{i,2}(0)|w_i = 1)$.
- We can now have

$$\tau^{ATT} \equiv E(Y_{i,2}(1) - Y_{i,2}(0)|w_i = 1) =$$
$$E(Y_{i,2}(1) - Y_{i,1}(1)|w_i = 1) - E(Y_{i,2}(0) - Y_{i,1}(0)|w_i = 0)$$

  This gives the so-called "Difference-in-Difference" estimator.

Linear Regression

- In practice, we often set up a regression to identify the τ^{ATT}

$$Y_{it} = \alpha_i + \gamma_t + w_{it}\beta + \epsilon_{it}$$

where w_{it} is the treatment status of unit i at time t.

- This is the Two-way Fixed Effects estimator (TWFE).
- It is equivalent to estimate the effect in a constructive way by

$$\hat{\tau} = N^{-1} \sum_i w_i(Y_{i,2} - Y_{i,1}) - (1 - w_i)(Y_{i,2} - Y_{i,1})$$

- Can be generalized to a multiple periods setting (although recent literature has cautioned against the "staggered treatment" problems). Let us consider the simple situation that the policy occurs to all units at time t_0, we can then use

$$Y_{it} = \alpha_i + \gamma_t + \sum_{t=1, t \neq t_0}^{T} \beta_t w_{it} + \epsilon_{it}$$

where all the coefficients measure the policy effect relative to period t_0

Card and Krueger (1994): Minimum Wage

Background
How to evaluate the effect of the minimum wage increase on employment in New Jersey?

- Look into a specific industry that is **affected/treated** the most by the minimum wage increase – fast-food stores in NJ.
- Find good **comparison/control groups** – fast-food stores in eastern Pennsylvania.
- Compare the **Difference** in the **Differences** of average employment rates before and after the law implementation in the two states.

Card and Krueger (1994): Minimum Wage

TABLE 1—SAMPLE DESIGN AND RESPONSE RATES

	All	Stores in: NJ	PA
Wave 1, February 15–March 4, 1992:			
Number of stores in sample frame:[a]	473	364	109
Number of refusals:	63	33	30
Number interviewed:	410	331	79
Response rate (percentage):	86.7	90.9	72.5
Wave 2, November 5–December 31, 1992:			
Number of stores in sample frame:	410	331	79
Number closed:	6	5	1
Number under rennovation:	2	2	0
Number temporarily closed:[b]	2	2	0
Number of refusals:	1	1	0
Number interviewed:[c]	399	321	78

[a]Stores with working phone numbers only; 29 stores in original sample frame had disconnected phone numbers.
[b]Includes one store closed because of highway construction and one store closed because of a fire.
[c]Includes 371 phone interviews and 28 personal interviews of stores that refused an initial request for a phone interview.

Card and Krueger (1994): Minimum Wage

TABLE 2—MEANS OF KEY VARIABLES

Variable	Stores in: NJ	Stores in: PA	t^a
1. *Distribution of Store Types (percentages):*			
a. Burger King	41.1	44.3	−0.5
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	−1.1
e. Company-owned	34.1	35.4	−0.2
2. *Means in Wave 1:*			
a. FTE employment	20.4	23.3	−2.0
	(0.51)	(1.35)	
b. Percentage full-time employees	32.8	35.0	−0.7
	(1.3)	(2.7)	
c. Starting wage	4.61	4.63	−0.4
	(0.02)	(0.04)	
d. Wage = $4.25 (percentage)	30.5	32.9	−0.4
	(2.5)	(5.3)	
e. Price of full meal	3.35	3.04	4.0
	(0.04)	(0.07)	
f. Hours open (weekday)	14.4	14.5	−0.3
	(0.2)	(0.3)	
g. Recruiting bonus	23.6	29.1	−1.0
	(2.3)	(5.1)	
3. *Means in Wave 2:*			
a. FTE employment	21.0	21.2	−0.2
	(0.52)	(0.94)	
b. Percentage full-time employees	35.9	30.4	1.8
	(1.4)	(2.8)	
c. Starting wage	5.08	4.62	10.8
	(0.01)	(0.04)	
d. Wage = $4.25 (percentage)	0.0	25.3	—
		(4.9)	
e. Wage = $5.05 (percentage)	85.2	1.3	36.1
	(2.0)	(1.3)	
f. Price of full meal	3.41	3.03	5.0
	(0.04)	(0.07)	
g. Hours open (weekday)	14.4	14.7	−0.8
	(0.2)	(0.3)	
h. Recruiting bonus	20.3	23.4	−0.6
	(2.3)	(4.9)	

- The two states are quite comparable before the policy assignment.

Card and Krueger (1994): Minimum Wage

Identification

We can identify the treatment effect using the following specifications:

$$\Delta E_i = a + b\mathbf{X}_i + cNJ_i + \varepsilon_i$$
$$\Delta E_i = a' + b'\mathbf{X}_i + c'GAP_i + \varepsilon_i'$$

where:

$$GAP_i = \begin{cases} 0 & \text{for stores in Pennsylvania} \\ 0 & \text{for stores in New Jersey with } W_{1i} \geq \$5.05 \\ \dfrac{5.05 - W_{1i}}{W_{1i}} & \text{for other stores in New Jersey.} \end{cases}$$

Card and Krueger (1994): Minimum Wage

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model (i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap[a]	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls[d]	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are −0.237 and 8.825, respectively. All models include an unrestricted constant (not reported).

[a] Proportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

[b] Three dummy variables for chain type and whether or not the store is company-owned are included.

[c] Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

[d] Probability value of joint F test for exclusion of all control variables.

Meyer et al. 1990: Workers' Compensation

Background

- State-run insurance program for workers' compensation in Kentucky and Michigan.
- Compensate workers for medical expenses and lost work due to on-the-job accident.
- The premiums are paid by firms as a function of **previous claims** and **wages paid** with a cap.

$$Premium = \min(pY, C)$$

where P: percent of replacement. Y: earnings. C: payment cap.
e.g. 65% of earnings up to \$ 400/month

Meyer et al. 1990: Workers' Compensation

How workers' compensation influences the duration of time employees take off work due to injury?

- Concern: Moral hazard. Benefits will discourage return to work.
- Previous studies: Regress duration Y on replaced wages X
  - Given progressive nature of benefits, replaced wages reveal a lot about the workers.
  - Replacement rates higher in higher wage states.

Meyer et al. 1990: Workers' Compensation

- Previous studies: Regress duration Y on replaced wages X

$$Y_i = X_i\beta + \alpha R_i + \varepsilon_i$$

- Y: Duration. R: Replacement rate.
- Expect: $\alpha > 0$
- Expect $Cov(R_i, \varepsilon_i)$
  - Higher wage workers have lower R and higher duration (understate).
  - Higher wage states have longer duration and longer R (overstate).

Meyer et al. 1990: Workers' Compensation

Identification

Quasi experiment in KY and MI: Increased earnings cap.

- Increased benefit for higher-wage workers (Treated)
- Did nothing to those already below original cap. (Comparison/Control)

Compare the change in duration of spell before/after change for these groups.

Meyer et al. 1990: Workers' Compensation

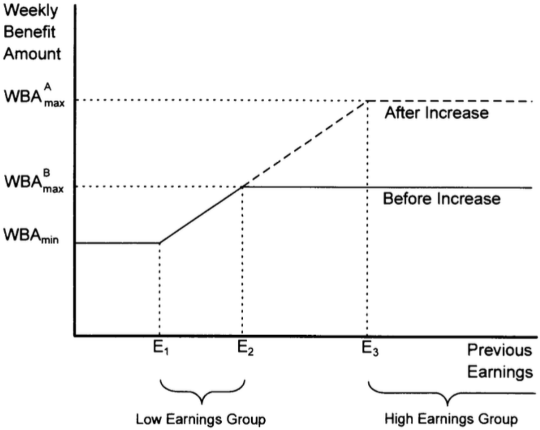

FIGURE 1. TEMPORARY TOTAL BENEFIT SCHEDULE BEFORE AND AFTER AN INCREASE IN THE MAXIMUM WEEKLY BENEFIT

Meyer et al. 1990: Workers' Compensation

TABLE 4—KENTUCKY AND MICHIGAN: DURATION AND MEDICAL COSTS OF TEMPORARY TOTAL DISABILITIES DURING THE YEARS BEFORE AND AFTER BENEFIT INCREASES

Variable	High earnings: Before increase (1)	High earnings: After increase (2)	Low earnings: Before increase (3)	Low earnings: After increase (4)	Differences: [(2)−(1)] (5)	Differences: [(4)−(3)] (6)	Difference in differences: [(5)−(6)] (7)
Mean duration (weeks)							
Kentucky	11.16	12.89	6.25	7.01	1.72	0.76	0.96
	(0.83)	(0.83)	(0.30)	(0.41)	(1.17)	(0.51)	(1.28)
Michigan	14.76	19.42	10.94	13.64	4.66	2.70	1.96
	(2.25)	(2.67)	(1.09)	(1.56)	(3.49)	(1.90)	(3.97)
Median duration (weeks)							
Kentucky	4.00	5.00	3.00	3.00	1.00	0.00	1.00
	(0.14)	(0.20)	(0.11)	(0.12)	(0.25)	(0.16)	(0.29)
Michigan	5.00	7.00	4.00	4.00	2.00	0.00	2.00
	(0.45)	(0.67)	(0.22)	(0.28)	(0.81)	(0.35)	(0.89)

Meyer et al. 1990: Workers' Compensation

Model

- Meyer et al. estimated the following model:

$$Y_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 A_{it} + \beta_3 A_{it} H_{it} + \beta_4 X'_{it} + \varepsilon_{it}$$

  where Y_{it}: Duration of spell on WC. A_{it}: Period after benefits hike.H_{it}: High earnings group(Income $> E_3$)

- Diff-in-Diff estimate: $\beta_{\mathbf{3}}$

Meyer et al. 1990: Workers' Compensation

TABLE 6—REGRESSION EQUATIONS FOR NATURAL LOGARITHM OF DURATION, HIGH- AND LOW-EARNINGS GROUPS POOLED, AND HIGH-EARNINGS GROUP SEPARATELY

	Specification							
	High- and low-earnings groups pooled				High-earnings group only			
	Kentucky		Michigan		Kentucky		Michigan	
Explanatory variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
After-increase indicator variable	0.016 (0.045)	−0.004 (0.038)	0.082 (0.084)	0.003 (0.073)	0.228 (0.054)	0.149 (0.044)	0.244 (0.136)	0.260 (0.113)
High-earnings-group indicator variable	−1.522 (1.099)	−0.594 (0.930)	5.577 (4.811)	3.607 (4.162)				
After-increase × high-earnings-group indicator variable	0.215 (0.069)	0.162 (0.059)	0.157 (0.153)	0.203 (0.132)				
ln(Previous earnings)	0.258 (0.104)	0.207 (0.088)	0.901 (0.648)	0.139 (0.562)	0.492 (0.163)	0.229 (0.133)	0.067 (0.496)	−0.335 (0.414)
ln(Previous earnings) × high-earnings group	0.232 (0.187)	0.065 (0.158)	−0.973 (0.803)	−0.587 (0.695)				

Gruber and Poterba(1994): Tax incentive and health insurance

Background

- Tax deduction of employer-provided health insurance
- Governments target: High and rising medical expenditures v.s. lowering the number of uninsured person
- Key point: Estimate the price elasticity of demand of health insurance

Identification

- Split the differences in demand as well as the cost of providing the care
- the Tax Reform Act of 1986 (TRA86) - a tax deduction equal to 25 percent of health insurance costs
- Self-employed individuals are eligible for higher tax deduction

Gruber and Poterba(1994): Tax incentive and health insurance

TABLE I

AVERAGE AFTER-TAX PRICE OF HEALTH INSURANCE

	Before TRA86	After TRA86
Self-employed	1 410 (0 074)	1 334 (0 055)
Employed	0 922 (0 045)	0 920 (0 045)
High-income self-employed	1 455 (0 065)	1 307 (0 041)
Low-income self-employed	1 389 (0 078)	1 355 (0 068)
High-income employed	0 900 (0 038)	0 902 (0 029)
Low-income employed	0 950 (0 046)	0 953 (0 042)

Notes Figures are tax price of health insurance based on authors calculations as described in Appendix A Values in parens are standard deviations of estimated prices for individuals in each category

Gruber and Poterba(1994): Tax incentive and health insurance

Data

- Pre-TRA samples: 1985 and 1986 Current Population Survey (CPS)
- Post-TRA samples: 1988 and 1989 CPS

Classfication

- Self-employed: report to be self-employed based on their main job last year and report at least $2000 in self-employment income
- Employed: report to be employed, report no self-employment income, and report at least $2000 in wage and salaries
- Uninsured: employed more than the self-employed in terms of demographic characteristics, but in terms of occupational and industrial distribution they are more similar to the self-employed

Gruber and Poterba(1994): Tax incentive and health insurance

TABLE II
CHARACTERISTICS OF THE SAMPLE

	Employed pre-TRA86	Employed post-TRA86	Self-emp pre-TRA86	Self-emp post-TRA86	Working unins pre-TRA86
Experience	17 9	18 1	20 1	20 3	18 2
Education	13 1	13 1	13 3	13 4	11 7
Female (%)	49 1	49 2	25 8	30 1	41 2
Married (%)	69 2	68 0	78 8	76 2	49 4
Nonwhite (%)	13 5	13 4	6 1	6 3	18 0
Working (%)	100	100	100	100	100
<35 hours (%)	14 4	13 6	15 5	16 5	20 5
<26 weeks (%)	11 0	9 6	5 1	4 9	19 2
Family income ($1985)	35,839	36,856	37,852	39,837	27,641
Management, technician (%)	29 5	30 7	31 8	31 1	14 1
Sales, services (%)	39 8	39 1	30 0	31 1	41 0
Manual (%)	30 7	30 2	38 6	37 8	44 8
Ag, mining, construct (%)	8 7	8 4	30 3	30 4	22 0
Manufacturing (%)	21 5	20 8	3 7	3 8	12 9
Trade & services (%)	69 8	70 8	66 1	65 8	65 1
Sample Size	85560	87515	6786	7306	14902

Notes Figures are tabulated from the 1986 1990 CPS for 25–54 year olds as described in text

Gruber and Poterba(1994): Tax incentive and health insurance

TABLE III

INSURANCE COVERAGE, SELF-EMPLOYMENT STATUS, AND INCOME

Pre-TRA86						
	All		Single		Married	
Income	SE	Empl	SE	Empl	SE	Empl
0–5K	30 0	36 0	23 7	35 4	39 7	38 9
5–10K	38 3	51 7	38 8	53 5	38 0	46 3
10–20K	55 4	80 0	48 0	83 7	59 0	74 8
20–30K	71 9	92 4	57 4	94 3	75 0	91 4
30–50K	81 3	96 5	68 7	95 6	82 9	96 6
50K+	88 1	97 9	73 5	96 2	89 4	98 0
Overall	69 4	87 9	50 1	80 2	74 7	91 5
Sample	85560	6786	1447	26487	5339	59073
Post-TRA86						
	All		Single		Married	
Income	SE	Empl	SE	Empl	SE	Empl
0–5K	33 5	30 9	29 9	30 0	42 9	35 3
5–10K	39 1	45 5	38 0	47 7	40 8	38 1
10–20K	57 2	74 3	50 7	79 8	61 0	66 0
20–30K	73 6	88 9	59 2	92 0	77 7	87 2
30–50K	84 3	95 2	68 8	94 5	86 5	95 3
50K+	91 6	97 2	87 0	93 9	92 1	97 4
Overall	73 3	84 9	54 0	77 0	79 4	89 1
Sample	87515	7306	1744	28013	5562	59502

- Single and married: Control for the changes in the self-employed's other family members
- Higher income, higher coverage; especially in the employed group
- Post-TRA: Increasing coverage in the self-employed group and slightly decreasing in the employed group

Gruber and Poterba(1994): Tax incentive and health insurance

DID Identification

- No health insurance's expenditure in the CPS, using the Probit model
- baseline

$$I_i^* = X_i\beta + Y_i\alpha + \mathrm{SELF}_i\,\delta_1 + \mathrm{POST}\,86_i\delta_2 + \mathrm{SELF}_i \times \mathrm{POST}\,86_i\delta_3 + \epsilon_1$$

where I_i^*: underlying demand for health insurance

SELF: a dummy variable for self-Employed

POST86: a dummy variable for the post-TRA period

X_i: socio-demographic characteristics

Y_i: income

$$\mathrm{prob}\,(I_i = 1) = \mathrm{prob}\,(I_i^* > 0)$$

Gruber and Poterba(1994): Tax incentive and health insurance

	All	Single	Married
Diff-in-Diff Self-employed versus employed (δ_3 in equation (5))			
Probit coefficient	0.236 (0.026)	0.232 (0.050)	0.261 (0.031)
Marginal probability	0.037	0.049	0.033
Implied elasticity	−0.500	−0.620	−0.440

Gruber and Poterba(1994): Tax incentive and health insurance

Income difference

- The post-1986 deduction should be more valuable to higher income self-employed individuals than to lower income self-employed individuals, since the marginal tax rate for the former group is higher.

$$I_i^* = X_i\beta + Y_i\alpha + \text{HIINC}_i\,\delta_1 + \text{POST}\,86_i^*\delta_2 + \text{HIINC}_i \times \text{POST}\,86_i\delta_3 + \epsilon_i$$

where HIINC: 1 for individuals over $50000 in real family income, and 0 for individuals below $20000

Diff-in-Diff High-income versus low-income self-employed (δ_3 in equation (6))			
Probit coefficient	0 148 (0 075)	0 538 (0 192)	0 098 (0 085)
Marginal probability	0 042	0 186	0 025
Implied elasticity	−0 368	−2 188	−0 216

Gruber and Poterba(1994): Tax incentive and health insurance

DDD setting

Whether the difference in the rate of coverage growth for high- and low-income self-employed individuals is the same as the employed individuals.

$$
\begin{aligned}
I_i^* = X_i\beta + Y_i\alpha + \mathrm{HIINC}_i\delta_1 + \mathrm{POST86}_i\delta_2 \\
+\mathrm{HIINC}_i \times \mathrm{POST86}_i\delta_3 + \mathrm{SELF}_i\delta_4 \\
+\mathrm{HIINC}_i \times \mathrm{SELF}_i\delta_5 + \mathrm{SELF}_i \times \ \mathrm{POST\ 86}_i\delta_6 \\
+\mathrm{HIINC}_i \times \mathrm{POST86}_i \times \mathrm{SELF}_i\delta_7 + \epsilon_i
\end{aligned}
$$

Diff-in-Diff-in-Diff High-income versus low-income self-employed versus high-income versus low-income employed (δ_7 in equation (7))			
Probit coefficient	0 135 (0 082)	0 730 (0 227)	−0 008 (0 094)
Marginal probability	0 028	0 170	−0 001
Implied elasticity	-0 248	−1 270	0 008