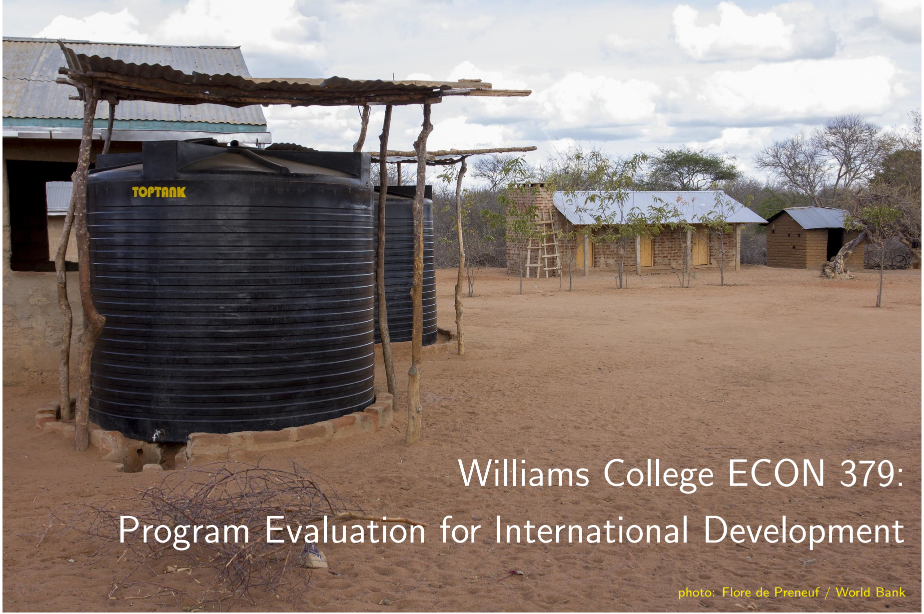


Williams College ECON 379: Program Evaluation for International Development

photo: Flore de Preneuf / World Bank

Williams College ECON 379:

Program Evaluation for International Development

Module 4: Difference-in-Differences

Professor: Pamela Jakiela

photo: Daniella Van Leggelo-Padilla / World Bank

Intuition

False Counterfactuals

Before vs. After Comparisons:

- **Compares:** same units before/after program
- **Drawback:** does not control for time trends

Participant vs. Non-Participant Comparisons:

- **Compares:** participants to those not in the program
- **Drawback:** potential for selection bias
  Are participants different in absence of treatment?

Two Wrongs Sometimes Make a Right

Difference-in-differences estimation (or "diff-in-diff" or "DD") combines the two (flawed) false counterfactual approaches

- Observe self-selected treatment group, comparison groups before and after treatment group participates in treatment
- May overcome problems of both when [1] selection bias is on fixed traits of units and [2] time trends common to both groups

The diff-in-diff estimator is:

$$DD = \bar{Y}^{treatment}_{post} - \bar{Y}^{treatment}_{pre} - \left(\bar{Y}^{comparison}_{post} - \bar{Y}^{comparison}_{pre}\right)$$

Difference-in-Differences Estimation

	treatment	comparison
pre-program	$\bar{Y}^{treatment}_{pre}$	$\bar{Y}^{comparison}_{pre}$
post-program	$\bar{Y}^{treatment}_{post}$	$\bar{Y}^{comparison}_{post}$

Diff-in-diff estimation is just a comparison of 4 cell-level means

Difference-in-Differences: A History

Ignaz Semmelweis, Diff-in-Diff Pioneer

In the 1840s, observers of Vienna's maternity wards noted that death rates from postpartum infections were higher in one wing

- Division 1 patients attended by doctors and trainee doctors
- Division 2 patients attended by midwives and trainee midwives

Ignaz Semmelweis noted that the difference emerged in 1841, when hospital moved to "anatomical" training involving cadavers

- Doctors received new training, but midwives didn't
- Did transference of "cadaveric particles" explain death rate?

Semmelweis proposed hand-washing with chlorine

- Policy implemented in May of 1847

Ignaz Semmelweis, Diff-in-Diff Pioneer

	Physicians' Wing			**Midwives' Wing**		
		Deaths			Deaths	
Year	Births	No.	%	Births	No.	%
1841	3036	237	7.7	2442	86	3.5
1842	3287	518	15.8	2659	202	7.5
1843	3060	274	8.9	2739	169	6.2
1844	3157	260	8.2	2956	68	2.3
1845	3492	241	6.8	3241	66	2.03
1846	4010	459	11.4	3754	105	2.7
Intervention introduced in May of 1847						
1847	3,975	176	4.4	3306	32	0.9
1848	3356	45	1.27	3219	43	1.33
1849	3,858	103	2.7	3,371	87	2.6

Ignaz Semmelweis: Epilogue

Ignaz Semmelweis was fired (for political reasons) in 1849

- Doctors in Vienna continued washing their hands
- Semmelweis' theory of "cadaveric particles" not widely accepted by European medical community at the time

In the 1860s, Louis Pasteur's research on the germ theory of disease provided a scientific explanation for effect of chlorine hand washing

John Snow's Grand Experiment

1849: London's worst cholera epidemic claims 14,137 lives

- Two companies supplied water to much of south London: Lambeth Waterworks and Southwark and Vauxhall Water Co.
  - Both got their water from the Thames, which was dirty
- John Snow believed cholera was spread by contaminated water; most believed cholera transmitted through "miasma" in the air

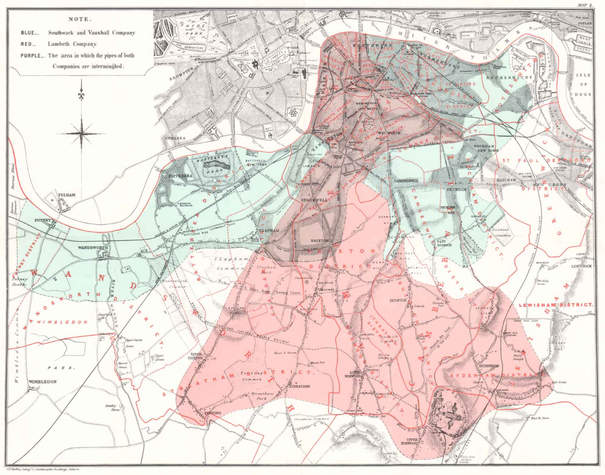
1852: Lambeth Waterworks moved their intake upriver

- Everyone knew the Thames was dirty below central London

1853: London has another cholera outbreak

- Are Lambeth Waterworks customers less likely to get sick?

John Snow's Grand Experiment



Source: John Snow Archive and Research Companion

John Snow's Grand Experiment

John Snow's Grand Experiment:

- Mortality data showed very few cholera deaths in areas of London that were **only** supplied by Lambeth Waterworks
- Snow hired John Whiting to visit the homes of the deceased to determine which company (if any) supplied their drinking water
- Using Whiting's data, Snow calculated the death rate:
  - Southwark and Vauxhall: 71 cholera deaths/10,000 homes
  - Lambeth: 5 cholera deaths/10,000 homes
- **Southwark and Vauxhall responsible for 286 of 334 deaths**
  - Southwark and Vauxhall moved their intake upriver in 1855

John Snow: Epilogue

source: wikimedia commons

Broad Street cholera outbreak killed 616 people in 1854

$\Rightarrow$ Snow convinced many pump was source

Diff-in-Diff Estimation by Economists

BULLETIN OF THE
U. S. BUREAU OF LABOR STATISTICS.

WHOLE NO. 176. WASHINGTON. JULY, 1915.

EFFECT OF MINIMUM-WAGE DETERMINATIONS IN OREGON.[1]

BY MARIE L. OBENAUER AND BERTHA VON DER NIENBURG.

Source: Obenauer and Nienburg (1915)

Diff-in-Diff Estimation by Economists

In 1913, Oregon increased minimum wage for experienced women to \$9.25 per week, with a maximum of 50 hours of work/week

- Minimum wage for inexperienced women (and girls) also increased, but new minimum (\$6/week) not binding constraint
- Obenauer and Nienburg obtain HR records of 40 firms
- Compared employment of experienced women before/after new law to employment of girls, inexperienced women, men

Diff-in-Diff Estimation by Economists

TABLE 1.—ESTABLISHMENTS COVERED IN THE INVESTIGATION AND WOMEN AND MEN EMPLOYED DURING PERIOD STUDIED IN 1914.

[This table does not include extra male or female help whose identity from week to week could not be traced, such female help being equivalent to 3 women working full time; nor does it include 20 saleswomen whose regular employment began with the opening of a new department on the last day of the period covered in the investigation.]

Type of store.	Number of establishments covered.	Number of persons employed during period studied in 1914.	
		Women and girls.	Men.
PORTLAND.			
Department, dry-goods, and 5 and 10 cent stores	6	1,345	802
Specialty stores	11	181	49
Neighborhood stores	16	20	17
Total	33	1,546	868
SALEM.			
Dry-goods, specialty, and 5 and 10 cent stores	7	96	34
Grand total	40	1,642	902

[1] See note [1], p. 57.
[2] One firm, Olds, Wortman & King, a Portland department store, refused the Federal agents access to their records. They offered to furnish a summary statement, but the Bureau did not regard this as comparable with material obtained direct from other firms' books.

Source: Obenauer and Nienburg (1915)

Diff-in-Diff Estimation by Economists

		Girls (16–18)		Women (19+)	
	Men	No.	G/M	No.	W/M
1913 (before)	940	138	0.146	1,543	1.641
1914 (after)	868	160	0.184	1,327	1.529
Change	−72	22	0.038	−216	−0.113

Data collected for March and April of each year. G/M indicates the ratio of girls (aged 16 to 18) employed to men employed. W/M indicates the ratio of women (aged 19 and above) employed to men employed.

Source: Kennan (1995)

Identifying Assumptions

Common Trends

Identifying assumption underlying diff-in-diff: treatment, comparison outcomes evolving on same trajectory (in the absence of treatment)

- Assumption about treatment group counterfactual
- Referred to as **common trends** assumption

There are two (implicit) parts to this assumption:

- Selection bias relates to fixed characteristics of individuals
  - Magnitude of the selection bias term isn't changing over time
- Time trend/shocks same for treatment and control groups

Both necessary conditions for causal inference using diff-in-diff

An Example of a Data-Generating Process

In absence of program, unit i's expected outcome at time τ is:

$$E[Y_{0i}|D_i = 0, t = \tau] = \gamma_i + \lambda_\tau$$

An Example of a Data-Generating Process

In absence of program, unit i's expected outcome at time τ is:

$$E[Y_{0i}|D_i = 0, t = \tau] = \gamma_i + \lambda_\tau$$

Outcomes in the comparison group:

$$E[\bar{Y}_{pre}^{comparison}] = E[Y_{0i}|D_i = 0, t = 1] = E[\gamma_i|D_i = 0] + \lambda_1$$

$$E[\bar{Y}_{post}^{comparison}] = E[Y_{0i}|D_i = 0, t = 2] = E[\gamma_i|D_i = 0] + \lambda_2$$

An Example of a Data-Generating Process

The comparison group allows us to estimate the **time trend**:

$$E[\bar{Y}_{post}^{comparison}] - E[\bar{Y}_{pre}^{comparison}]$$

$$= E[\gamma_i | D_i = 0] + \lambda_2 - (E[\gamma_i | D_i = 0] + \lambda_1)$$

$$= \lambda_2 - \lambda_1$$

An Example of a Data-Generating Process

Let δ denote the true impact of the program:

$$\delta = E[Y_{1i}|D_i = 1, t = \tau] - E[Y_{0i}|D_i = 1, t = \tau]$$

which does not depend on time period or i's characteristics

Outcomes in the treatment group:

$$E[\bar{Y}_{pre}^{treatment}] = E[Y_{0i}|D_i = 1, t = 1] = E[\gamma_i|D_i = 1] + \lambda_1$$

$$E[\bar{Y}_{post}^{treatment}] = E[Y_{1i}|D_i = 1, t = 2] = E[\gamma_i|D_i = 1] + \delta + \lambda_2$$

Differences in outcomes pre-treatment vs. post treatment cannot be attributed to program; treatment effect is conflated with time trend

An Example of a Data-Generating Process

If we were to calculate a pre vs. post estimator, we'd have:

$$E[\bar{Y}_{post}^{treatment}] - E[\bar{Y}_{pre}^{treatment}]$$

$$= E[\gamma_i|D_i = 1] + \delta + \lambda_2 - (E[\gamma_i|D_i = 1] + \lambda_1)$$

$$= {\color{red}\delta} + \underbrace{\lambda_2 - \lambda_1}_{\text{time trend}}$$

If we calculated a treatment vs. comparison estimator, we'd have:

$$E[\bar{Y}_{post}^{treatment}] - E[\bar{Y}_{post}^{comparison}]$$

$$= E[\gamma_i|D_i = 1] + \delta + \lambda_2 - (E[\gamma_i|D_i = 0] + \lambda_2)$$

$$= {\color{red}\delta} + \underbrace{E[\gamma_i|D_i = 1] - E[\gamma_i|D_i = 0]}_{\text{selection bias}}$$

An Example of a Data-Generating Process

Substituting in the terms from our model:

$$DD = \bar{Y}^{treatment}_{post} - \bar{Y}^{treatment}_{pre} - \left(\bar{Y}^{comparison}_{post} - \bar{Y}^{comparison}_{pre}\right)$$

$$= E[Y_{1i}|D_i = 1, t = 2] - E[Y_{0i}|D_i = 1, t = 1]$$

$$- \left(E[Y_{0i}|D_i = 0, t = 2] - E[Y_{0i}|D_i = 0, t = 1]\right)$$

$$= E[\gamma_i|D_i = 1] + \delta + \lambda_2 - (E[\gamma_i|D_i = 1] + \lambda_1)$$

$$- \left[E[\gamma_i|D_i = 0] + \lambda_2 - \left(E[\gamma_i|D_i = 0] + \lambda_1\right)\right]$$

$$= \delta$$

When Does Diff-in-Diff Work?

Diff-in-diff recovers true impact of program on participants (as long as common trends assumption isn't violated)

- Magnitude of selection bias cannot change over time
  - In model: $E[\gamma_i | D_i = 1] - E[\gamma_i | D_i = 0]$ is constant
- Time trends, shocks not correlated with treatment
  - In model: $\lambda_2 - \lambda_1$ same for treatment, comparison

Does not rely on assumption of homogeneous treatment effects

- When treatment effects are heterogeneous, DD estimation yields **average treatment effect on the treated** (ATT)

When Does Diff-in-Diff Work? An Example

	treatment	comparison
pre-program		
post-program		

Example:

Government introduces program for 8^{th} graders in public schools

Diff-in-Diff in the Wild

Minimum Wages and Employment

Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

By DAVID CARD AND ALAN B. KRUEGER*

On April 1, 1992, New Jersey's minimum wage rose from \$4.25 to \$5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above \$5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (*JEL* J30, J23)

source: Card and Krueger (*AER*, 1994)

Minimum Wages and Employment: Impacts on Wages

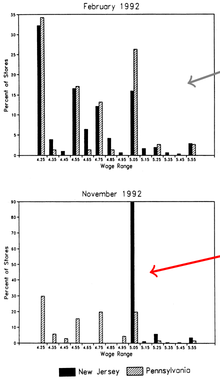


Figure 1. Distribution of Starting Wage Rates

Distribution of wages rates similar in NJ, PA

Minimum wage law shifts wage distribution in NJ: 90 percent at new legal minumum

source: Card and Krueger (*AER*, 1994)

Minimum Wages and Employment: Impacts on Employment

Variable	Stores by state: 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)

source: Card and Krueger (*AER*, 1994)

Outcome: employment (store-level)

Treatment group: New Jersey

$\Rightarrow$ Only one cell is treated

	NJ	PA	
pre	20.44	23.33	−2.89
post	21.03	21.17	−0.14
	0.59	−2.16	2.76

Minimum Wages and Employment: Impacts on Employment

Variable	Stores by state: 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)	$\boxed{2.76\ (1.36)}$

source: Card and Krueger (*AER*, 1994)

Standard errors – from whence?

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

when X and Y are independent

Minimum Wages and Employment: Impacts on Employment

Variable	Stores by state: 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)	$\boxed{2.76 \ (1.36)}$

source: Card and Krueger (*AER*, 1994)

$\rightarrow$ Standard errors – from whence?

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

when X and Y are independent

$$\boxed{\Rightarrow SE_{DD} = \sqrt{SE^2_{\Delta \bar{Y}^T} + SE^2_{\Delta \bar{Y}^C}}}$$

The Labor Market Consequences of School Construction

In a paper in *American Economic Review*, Esther Duflo examines the impacts of a large school construction program in Indonesia

Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment

By Esther Duflo*

Between 1973 and 1978, the Indonesian government engaged in one of the largest school construction programs on record. Combining differences across regions in the number of schools constructed with differences across cohorts induced by the timing of the program suggests that each primary school constructed per 1,000 children led to an average increase of 0.12 to 0.19 years of education, as well as a 1.5 to 2.7 percent increase in wages. This implies estimates of economic returns to education ranging from 6.8 to 10.6 percent. (JEL I2, J31, O15, O22)

source: Duflo (*AER*, 2001)

The Labor Market Consequences of School Construction

The Sekolar Dasar INPRES program (1973–1979):

- Oil crisis creates large windfall for Indonesia
- Suharto uses oil money to fund school construction
- Close to 62,000 schools built by national gov't
    - Approximately 1 school built per 500 school-age children
- More schools built in areas which started with fewer schools
- Schools intended to promote equality, national identity

The Labor Market Consequences of School Construction

Do children who were born into areas with more newly built INPRES primary schools get more education? Do they earn more as adults?

Treatment status: Children born in communities where many schools were built (treatment) vs. fewer schools (comparison)

- Partition sample based on residuals from a regression of number of schools built on number of school-aged children

Timing: Data on children born before and after program

- Children aged 12 and up in 1974 did not benefit from program
- Children aged 6 and under were young enough to be treated

The Labor Market Consequences of School Construction

Dep. Var.: Years of Education

	more schools	fewer schools	difference
over 11 in 1974	8.02	9.40	
under 7 in 1974	8.49	9.76	
difference	0.47	0.36	0.12

The Labor Market Consequences of School Construction

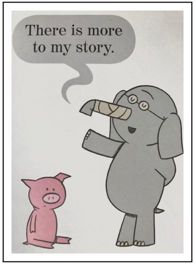
Dep. Var.: Log Wages

	more schools	fewer schools	difference
over 11 in 1974	6.87	7.02	-0.15
under 7 in 1974	6.61	6.73	-0.12
difference	-0.26	-0.29	0.026

The Labor Market Consequences of School Construction

Educational attainment, wages grew faster in “treatment” areas

- Difference are small, not statistically significant



source: Willems (2011)

Treatment, comparison groups differ in degree of exposure to treatment, not whether exposed

- May understate impact of treatment

When treatment is continuous in $[0, 1]$, a continuous treatment measure increases power

The End