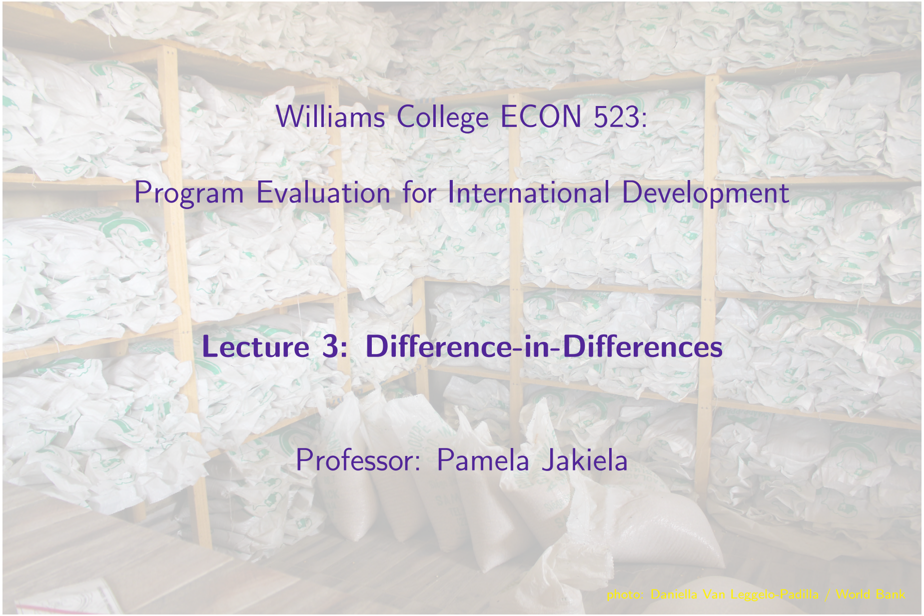


Williams College ECON 523:

Program Evaluation for International Development

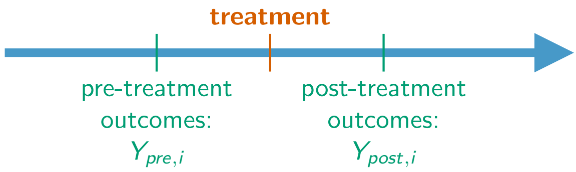
Lecture 3: Difference-in-Differences

Professor: Pamela Jakiela

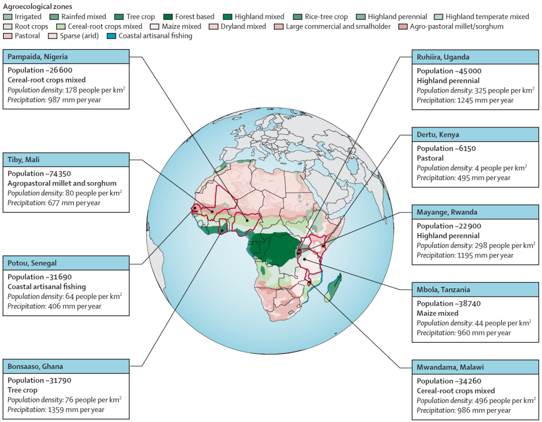
photo: Daniella Van Leggelo-Padilla / World Bank

Intuition

Pre vs. Post Comparisons



The Millennium Villages Project



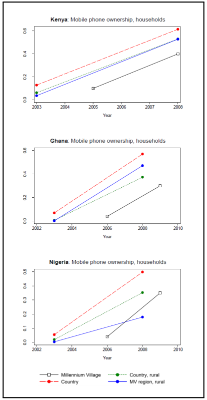
source: Pronyk et al. (2012)

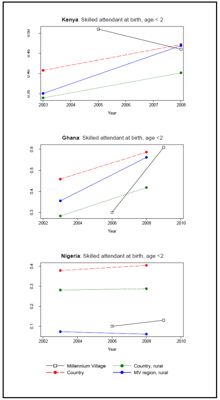
The Impacts of the Millennium Villages Project?

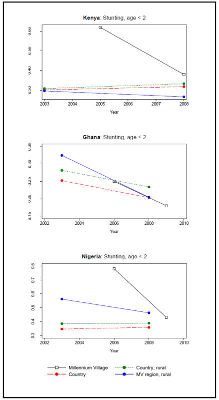
	Observational unit	Millennium Village sites (N=9)				Comparison village sites (N=9)				Millennium Villages vs comparison villages in year 3	
		Year 0 (number)	Year 3 (number)	Absolute change (95% CI)	p value	Year 0 (number)	Year 3 (number)	Absolute change (95% CI)	p value	Absolute difference (95% CI)	p value
Wasting	Children younger than 2 years of age‡	6·4% (271)	5·5% (644)	–0·9% (–4·1 to 2·4)	0·591	..	6·7% (776)	..	..	–1·2% (–6·5 to 4·2)§	0·630
Underweight	Children younger than 2 years of age‡	13·1% (279)	14·3% (660)	1·2% (–4·2 to 6·6)	0·669	..	16·1% (803)	..	..	–1·8% (–8·9 to 5·4)§	0·584
Stunting	Children younger than 2 years of age‡	36·0% (255)	28·2% (709)	–7·9% (–15·6 to –0·2)	0·045	..	35·7% (784)	..	..	–7·5% (–20·0 to 5·0)§	0·205

source: Pronyk et al. (2012)

Critiques of the MVP Evaluation







source: Clemens and Demombynes. (2010)

False Counterfactuals

Pre vs. Post Comparisons:

- **Compares:** same units before vs. after program implementation
- **Drawback:** does not control for time trends (in potential outcomes without treatment)

Participant vs. Non-Participant Comparisons:

- **Compares:** participants to those who choose not to participate in a program
- **Drawback:** potential for selection bias (participants differ from non-participants)

Neither approach provides credible estimates of program impacts

Two Wrongs Sometimes Make a Right

Difference-in-differences combines the two (flawed) false counterfactual approaches

- Observe self-selected treatment, comparison groups before and after treatment (i.e. before and after the treatment group participates in the program)
- May overcome problems of both false counterfactual approaches when:
  - Selection bias relates to fixed characteristics of units
  - Time trends are common to treatment and comparison groups

The difference-in-differences (or diff-in-diff, DD, or DiD) estimator is:

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

Difference-in-Differences Estimation

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

Difference-in-differences estimation is just a comparison of four cell-level means

Difference-in-Differences: A History

Ignaz Semmelweis, Diff-in-Diff Pioneer

In 1840s Vienna, deaths from postpartum infections were higher in one of two maternity wards

- 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 Vienna's Maternity Hospital introduced "anatomical" training of medical students (which involved cadavers)

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

Semmelweis proposed hand-washing with chlorine to remove contamination from cadavers

- 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

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

In the 1860s, Louis Pasteur's research on the germ theory of disease provided a scientific explanation for effect of chlorine hand washing (because chlorine/washing kills germs)

John Snow's Grand Experiment

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

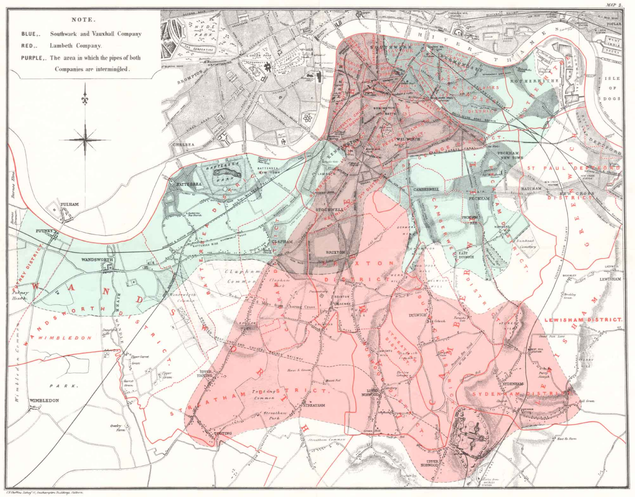
- Two companies supplied water to much of south London
  - The Lambeth Waterworks (LW) and the Southwark and Vauxhall Water Company (SVWC)
  - Both got their water from the Thames, which was dirty
- John Snow believed cholera was spread by contaminated water
  - Most believed cholera transmitted through "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: were LW 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:

- Very few cholera deaths in areas of London that were **only** supplied by LW
- John Snow hired John Whiting to visit the homes of those who died in the cholera outbreak to determine which of the two companies supplied their drinking water
- Using Whiting's data, Snow calculated the death rate:
  - SVWC: 71 cholera deaths/10,000 homes
  - LW: 5 cholera deaths/10,000 homes
- **SVWC responsible for 286 of 334 deaths**
  - 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

- Minimum wage for inexperienced women/girls also increased, but not binding
- Obenauer and Nienburg obtained HR records of 40 firms
- They compared employment of experienced women before and after implementation of new minimum wage law to employment of girls, inexperienced women, and (all) 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

	Men	Girls (16–18)		Women (19+)	
		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 difference-in-differences estimation:
Treatment, comparison outcomes evolving on same trajectory (in the absence of treatment)

- Assumption about treatment group counterfactual
- Referred to as **common trends** assumption (or parallel trends, or equal trends)

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 and period-specific shocks are the same for treatment and control groups

Both necessary conditions for causal inference using difference-in-differences

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:

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

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

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

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

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

$$= \delta$$

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 groups

Does not rely on assumption of homogeneous treatment effects

- DD estimation yields **average treatment effect on the treated** (ATT)

Operationalizing Difference-in-Differences

	treatment	comparison
pre-program		
post-program		

Example:

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

Difference-in-Differences in the Wild

The Labor Market Consequences of School Construction

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 windfall for Indonesia; Suharto uses oil money to fund school construction
- Close to 62,000 schools built by the Indonesian government
  - 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

Diff-in-diff methodology an be used with cross-sectional data to evaluate a nationwide program

The Labor Market Consequences of School Construction

Do children born where more new INPRES schools get more education? Do they earn more?

Treatment status: Children born in communities where many INPRES schools were built (treatment) are compared to children born in areas where fewer schools were built (comparison)

- Duflo operationalizes this by partitioning the sample based on the residuals from a regression of number of primary 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
  - Differences are small, not statistically significant
- Treatment, comparison groups differ in degree of exposure to treatment
  - May understate true effects of the INPRES program (everyone partially treated)
  - When treatment intensity varies continuously, exploiting variation can increase power

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

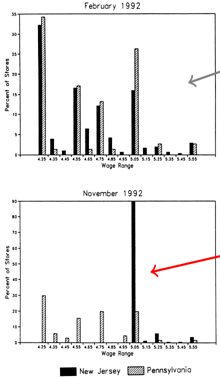


Figure 1. Distribution of Starting Wage Rates

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

Distribution of wages rates similar in NJ, PA

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

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

⇒ 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

2×2 Diff-in-Diff Specifications

Difference-in-Differences Estimation

	treatment	comparison	difference
pre	$\bar{Y}^T_{pre}$	$\bar{Y}^C_{pre}$	$\bar{Y}^T_{pre} - \bar{Y}^C_{pre}$
post	$\bar{Y}^T_{post}$	$\bar{Y}^C_{post}$	$\bar{Y}^T_{post} - \bar{Y}^T_{post}$
difference	$\bar{Y}^T_{post} - \bar{Y}^T_{pre}$	$\bar{Y}^C_{post} - \bar{Y}^C_{pre}$	δ_{DD}

Difference-in-Differences Estimation

To implement diff-in-diff in a regression framework, we estimate:

$$Y_{i,t} = \alpha + \beta D_i + \theta Post_t + \delta \left(D_i * Post_t \right) + \varepsilon_{i,t}$$

Where:

- D_i = treatment dummy
- $Post_i$ = dummy for post-treatment period
- $D_i * Post_i$ = interaction term

Difference-in-Differences Estimation

To implement diff-in-diff in a regression framework, we estimate:

$$Y_{i,t} = \alpha + \beta D_i + \theta Post_t + \delta \left(D_i * Post_t \right) + \varepsilon_{i,t}$$

Where:

- D_i = treatment dummy
- $Post_i$ = dummy for post-treatment period
- $D_i * Post_i$ = interaction term

Panel data: every unit$\times$period data point is an observation

Difference-in-Differences Estimation in Stata

```
. reg y treatment post treatxpost

      Source |       SS           df       MS      Number of obs   =     2,000
-------------+----------------------------------   F(3, 1996)      =     64.75
       Model |  1558.8687         3  519.622901   Prob > F        =    0.0000
    Residual |  16017.7056    1,996  8.02490261   R-squared       =    0.0887
-------------+----------------------------------   Adj R-squared   =    0.0873
       Total |  17576.5743    1,999   8.7926835   Root MSE        =    2.8328

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   treatment |  -.1928937   .1791636    -1.08   0.282     -.544261    .1584737
        post |   .0679519   .1791636     0.38   0.705    -.2834154    .4193193
  treatxpost |   2.110153   .2533757     8.33   0.000     1.613244    2.607061
       _cons |   5.231523   .1266878    41.29   0.000     4.983069    5.479977
------------------------------------------------------------------------------
```

Using ΔY_i as the Outcome Variable

Interacted specification is equivalent* to first differences:

$$Y_{i,t=2} - Y_{i,t=1} = \eta + \gamma D_i + \epsilon_{it}$$

where:

- $Y_{i,t=2} - Y_{i,t=1}$ = change (pre vs. post) in outcome of interest
- γ = coefficient of interest (the treatment effect)
- η = time trend (average change in comparison group)

* Coefficients will be identical, but standard errors may differ

Example: Minimum Wages and Employment in the Fast-Food Industry

Interacted specification is equivalent* to first differences:

$$\Delta \text{FTE}_i = \eta + \gamma \text{NJ}_i + \epsilon_i$$

where:

- ΔFTE_i = change in full-time employment in restaurant i
- γ = difference in mean change in NJ stores (vs. PA stores)
- η = constant (mean change in FTE in PA)

Example: Minimum Wages and Employment in the Fast-Food Industry

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model (i)	Model (ii)
New Jersey dummy	2.33 (1.19)	2.30 (1.20)
Controls for chain and ownership[b]	no	yes
Controls for region[c]	no	no
Standard error of regression	8.79	8.78
Probability value for controls[d]	—	0.34

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).

[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.

source: Card and Krueger (1994)