Diff-in-diff

Advanced Applied Econometrics

Felix Weinhardt

Slides based on Scott Cunningham and Paul Goldsmith-Pinkham
Thanks also to Renke Schmacker

Differences-in-differences

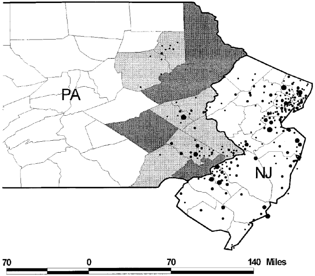
- Goal: estimate effect of a policy
- Naive approach: compare outcomes for treated and untreated groups
- Problem: but policy-adoption is not random $\Rightarrow$ what is the counterfactual?
- Solution: estimate difference-in-difference (DiD)

Differences-in-differences: Card and Krueger (1994)

- Suppose you are interested in the effect of minimum wages on employment (a classic and controversial question in labor economics).
- In a competitive labor market, increases in the minimum wage would move us up a downward sloping labor demand curve $\rightarrow$ employment would fall

DD: Card and Krueger (1994)

- Card and Krueger (1994) analyzed the effect of a minimum wage increase in New Jersey using a differences-in-differences (DD) methodology
- In February 1992, New Jersey increased the state minimum wage from \$4.25 to \$5.05. Pennsylvania's minimum wage stayed at \$4.25.



- They surveyed about 400 fast food stores both in New Jersey and Pennsylvania before and after the minimum wage increase in New Jersey

DD Strategy

- DD is a version of fixed effects estimation. To see this formally:

Y^1_{ist} : employment at restaurant i, state s, time t with a high w^{min}

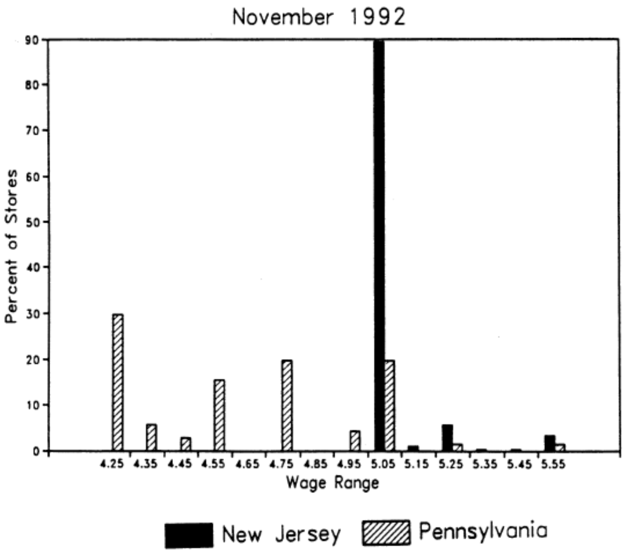
Y^0_{ist} : employment at restaurant i, state s, time t with a low w^{min}

- In practice of course we only see one or the other. We then assume that:

$$E[Y^0_{its}|s,t] = \gamma_s + \lambda_t$$

- In the absence of a minimum wage change, employment is determined by the sum of a time-invariant state effect, γ_s and a year effect, λ_t that is common across states
- Let D_{st} be a dummy for high-minimum wage states and periods
- Assuming $E[Y^1_{its} - Y^0_{its}|s,t] = \delta$ is the treatment effect, observed employment can be written:

$$Y_{its} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{its}$$


November 1992
Percent of Stores
0
10
20
30
40
50
60
70
80
90
4.25
4.35
4.45
4.55
4.65
4.75
4.85
4.95
5.05
5.15
5.25
5.35
5.45
5.55
Wage Range
New Jersey
Pennsylvania

DD Strategy II

- In New Jersey
  - Employment in February is

$$E(Y_{ist}|s = NJ, t = Feb) = \gamma_{NJ} + \lambda_{Feb}$$

  - Employment in November is:

$$E(Y_{ist}|s = NJ, t = Nov) = \gamma_{NJ} + \lambda_{Nov} + \delta$$

  - Difference between November and February

$$E(Y_{ist}|s = NJ, t = Nov) - E(Y_{ist}|s = NJ, t = Feb) = \lambda_{Nov} - \lambda_{Feb} + \delta$$

- In Pennsylvania
  - Employment in February is

$$E(Y_{ist}|s = PA, t = Feb) = \gamma_{PA} + \lambda_{Feb}$$

  - Employment in November is:

$$E(Y_{ist}|s = PA, t = Nov) = \gamma_{PA} + \lambda_{Nov}$$

  - Difference between November and February

$$E(Y_{ist}|s = PA, t = Nov) - E(Y_{ist}|s = PA, t = Feb) = \lambda_{Nov} - \lambda_{Feb}$$

DD Strategy II

- In New Jersey
  - Employment in February is

$$E(Y_{ist}|s = NJ, t = Feb) = \gamma_{NJ} + \lambda_{Feb}$$

  - Employment in November is:

$$E(Y_{ist}|s = NJ, t = Nov) = \gamma_{NJ} + \lambda_{Nov} + \delta$$

  - Difference between November and February

$$E(Y_{ist}|s = NJ, t = Nov) - E(Y_{ist}|s = NJ, t = Feb) = \lambda_{Nov} - \lambda_{Feb} + \delta$$

- In Pennsylvania
  - Employment in February is

$$E(Y_{ist}|s = PA, t = Feb) = \gamma_{PA} + \lambda_{Feb}$$

  - Employment in November is:

$$E(Y_{ist}|s = PA, t = Nov) = \gamma_{PA} + \lambda_{Nov}$$

  - Difference between November and February

$$E(Y_{ist}|s = PA, t = Nov) - E(Y_{ist}|s = PA, t = Feb) = \lambda_{Nov} - \lambda_{Feb}$$

DD Strategy III

- The DD strategy amounts to comparing the change in employment in NJ to the change in employment in PA.
- The population DD are:

$$\Big(E(Y_{ist}|s = NJ, t = Nov) - E(Y_{ist}|s = NJ, t = Feb)\Big)$$
$$-\Big(E(Y_{ist}|s = PA, t = Nov) - E(Y_{ist}|s = PA, t = Feb)\Big) = \delta$$

- This is estimated using the sample analog of the population means

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)

Surprisingly, employment *rose* in NJ relative to PA after the minimum wage change

Regression DD

- We can estimate the DD estimator in a regression framework
- Advantages:
  - It's easy to calculate the standard errors
  - We can control for other variables which can make the parallel trend assumption more credible and may reduce the residual variance (lead to smaller standard errors)
  - It's easy to include multiple periods
  - We can study treatments with different treatment intensity. (e.g., varying increases in the minimum wage for different states)
- The typical regression model we estimate is

$$\text{Outcome}_{it} = \beta_1 + \beta_2 \text{Treat}_i + \beta_3 \text{Post}_t + \beta_4 (\text{Treat} \times \text{Post})_{it} + \varepsilon_{it}$$

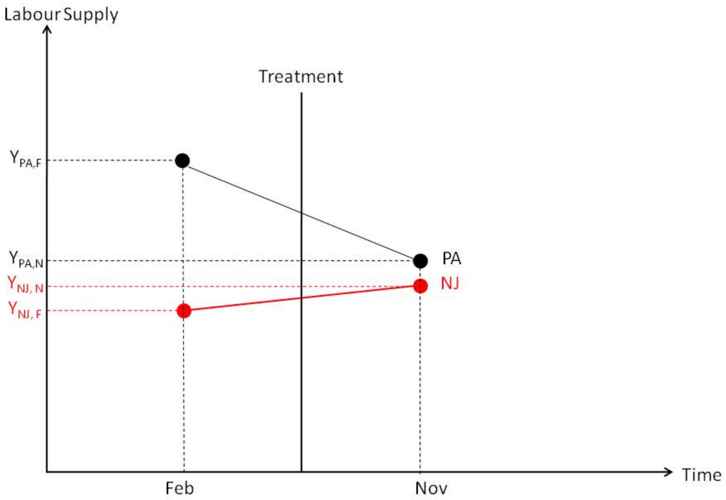
where Treat is a dummy if the observation is in the treatment group and Post is a post treatment dummy

Regression DD - Card and Krueger

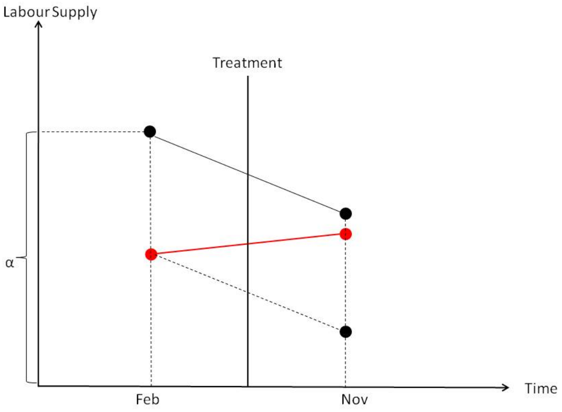
- In the Card and Krueger case, the equivalent regression would be:

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

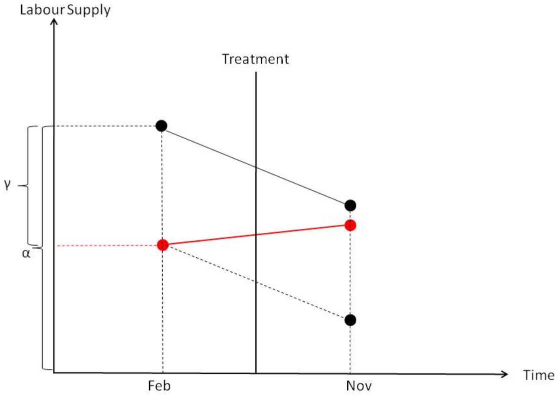
  - NJ is a dummy equal to 1 if the observation is from NJ
  - d is a dummy equal to 1 if the observation is from November (the post period)
- This equation takes the following values
  - PA Pre: α
  - PA Post: $\alpha + \lambda$
  - NJ Pre: $\alpha + \gamma$
  - NJ Post: $\alpha + \gamma + \lambda + \delta$
- DD estimate: (NJ Post - NJ Pre) - (PA Post - PA Pre) $= \delta$
- DD estimates the **ATT**


Labour Supply
Treatment
$Y_{PA,F}$
$Y_{PA,N}$
$Y_{NJ,N}$
$Y_{NJ,F}$
PA
NJ
Time
Feb
Nov

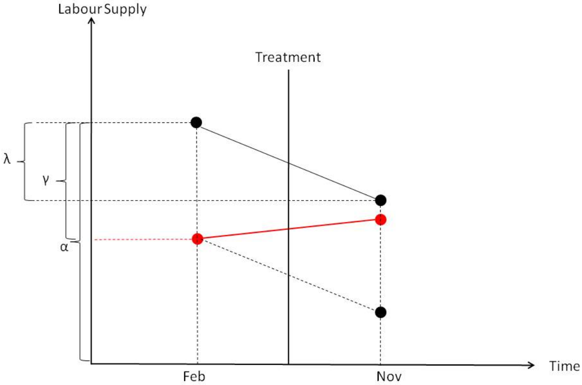
$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$



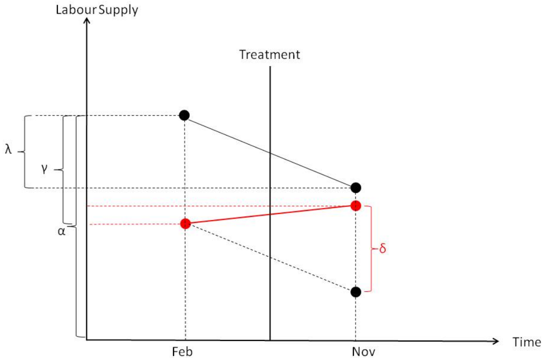
$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$



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



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



Assumptions of DD strategy

- Parallel trends:
  - TWFE estimates:

$$
\begin{aligned}
\beta^{DiD} =& E(Y_{i2}(1) - Y_{i1}(0)|D_{i2} = 1) - E(Y_{i2}(0) - Y_{i1}(0)|D_{i2} = 0) \\
=& E(Y_{i2}(1) - Y_{i1}(0)|D_{i2} = 1) - E(Y_{i2}(0) - Y_{i1}(0)|D_{i2} = 1) \\
& + E(Y_{i2}(0) - Y_{i1}(0)|D_{i2} = 1) - E(Y_{i2}(0) - Y_{i1}(0)|D_{i2} = 0)
\end{aligned}
$$

  - where $E(Y_{i2}(1) - Y_{i2}(0)|D_{i2} = 1) = ATT$ under the parallel trends assumption:

$$
E(Y_{i2}(0) - Y_{i1}(0)|D_{i2} = 1) = E(Y_{i2}(0) - Y_{i1}(0)|D_{i2} = 0)
$$

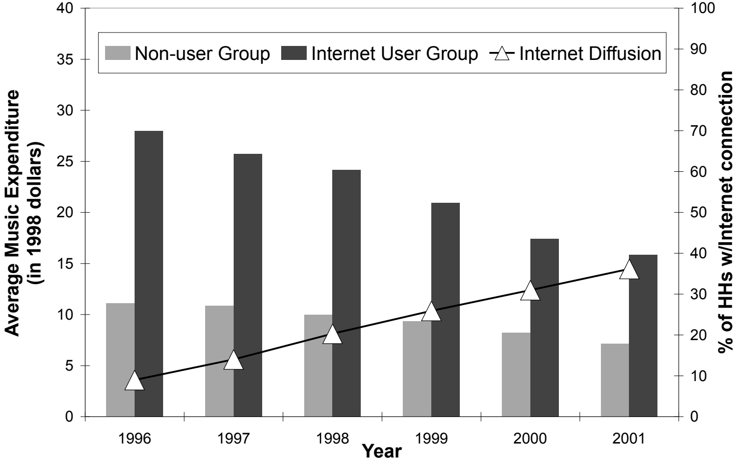
- Stable unit treatment value assumption (SUTVA)
- Fixed sample composition

Threats to validity – compositional differences

- One of the risks of a repeated cross-section (and unbalanced panel) is that the composition of the sample may have changed between the pre and post period
- Hong (2011) uses repeated cross-sectional data from the Consumer Expenditure Survey (CEX) containing music expenditure and internet use for a random sample of households
- Study exploits the emergence of Napter (first file sharing software widely used by Internet users) in June 1999 as a natural experiment
- Study compares internet users and internet non-users before and after emergence of Napster

Compositional differences?

Figure 1: Internet Diffusion and Average Quarterly Music Expenditure in the CEX



Threats to validity – compositional differences

Table 1: Descriptive Statistics for Internet User and Non-user Groups[a]

Year	1997		1998		1999	
	Internet User	Non-user	Internet User	Non-user	Internet User	Non-user
Average Expenditure						
Recorded Music	\$25.73	\$10.90	\$24.18	\$9.97	\$20.92	\$9.37
Entertainment	\$195.03	\$96.71	\$193.38	\$84.92	\$182.42	\$80.19
Zero Expenditure						
Recorded Music	.56	.79	.60	.80	.64	.81
Entertainment	.08	.32	.09	.35	.14	.39
Demographics						
Age	40.2	49.0	42.3	49.0	44.1	49.4
Income	\$52,887	\$30,459	\$51,995	\$28,169	\$49,970	\$26,649
High School Grad.	.18	.31	.17	.32	.21	.32
Some College	.37	.28	.35	.27	.34	.27
College Grad.	.43	.21	.45	.21	.42	.20
Manager	.16	.08	.16	.08	.14	.08

Diffusion of the Internet changes samples (e.g., younger music fans are early adopters)

Threats to validity – SUTVA

- Treatment status of any individual does not affect the potential outcomes of other units (non-interference)
  - Violation: Spillovers and general equilibrium effects
- No change in the composition of the treatment and control group due to the treatment
  - Violation: For example selective migration into area where treatment is introduced

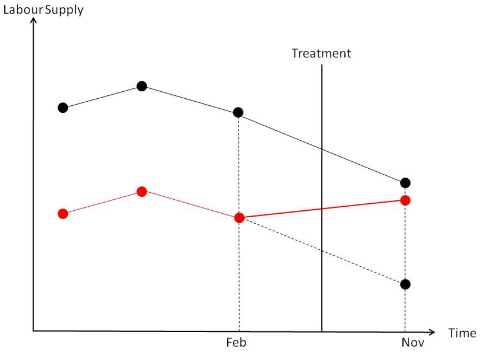
Threats to validity – Non-parallel trends

- 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"
    - Named after Orley Ashenfelter, labor economist at Princeton
    - 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.
- Regional targeting. NGOs may target villages that appear most promising, or worse off, which is a form of selection bias and violates parallel trends

Key assumption of any DD strategy: Parallel trends

- The key assumption for any DD strategy is that the outcome in treatment and control group would follow the same time trend in the absence of the treatment
- This doesn't mean that they have to have the same mean of the outcome
- Parallel trends are difficult to verify because technically one of the parallel trends is an unobserved counterfactual
- But one often will check using pre-treatment data to show that the trends are the same

Key assumption of any DD strategy: Parallel trends



Even if pre-trends are the same one still has to worry about other policies changing at the same time (omitted variable bias)

Regression DD Including Leads and Lags

- Suppose we have multiple periods before and after the treatment
  - Including leads into the DD model is an easy way to analyze pre-treatment trends
  - Lags can be included to analyze whether the treatment effect changes over time after assignment
- The estimated regression would be:

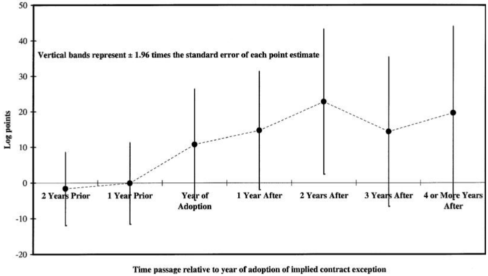
$$Y_{its} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \gamma_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + x_{ist} + \varepsilon_{ist}$$

  - Treatment occurs in year 0
  - Includes q leads or anticipatory effects
  - Includes m lags or post treatment effects
  - One of the coefficients is unidentified (TE relative to this period)

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
- Some states courts have made some exceptions to this employment at will rule and have thus increased employment protection
- The standard thing to do is normalize the adoption year to 0
- Autor then analyzes the effect of these exemptions on the use of temporary help workers.

Results



- The leads are very close to 0. $\rightarrow$ no evidence for anticipatory effects (good news for the parallel trends assumption).
- The lags show the effect increases during the first years of the treatment and then remains relatively constant.

Testing parallel pre-trends

- Diverging pre-trends can falsify the parallel trends assumption
- Showing parallel pre-trends lends support to the validity of the design (worth doing!)
- However, recent literature pointed out issues with testing pre-trends
    - Parallel trends in what?
    - Pre-testing issues