Quasi experiments. Difference in differences

Ciprian Domnisoru

September 10, 2021

Educational consultants mini case study

Suppose you run a school district and some consultants offer to sell you educational software to improve math skills for students, since your district ranked last on a regional standardized math skills test. The students can use the software at home, to help them prepare for homework and tests. You accept. To show efficiency, the consultants select a sample of 50 students with the lowest initial test scores and check how they perform next year, arguing they are the ones who are most difficult to teach. The students perform 4 points better on average on the test than the previous year. Are you convinced the educational software had a causal effect on the improvement in student test scores?

Discussion points

- What is the counterfactual: how would students perform in the absence of the intervention. Regression to the mean? Naturally occuring improvements?
- Is a randomized experiment feasible? Ethical?
- How do you expect the treatment effect to differ for the lowest performing students versus the average student?
- Other considerations: are 4 points a quantitatively meaningful effect?
- How do we know the treatment effect is due to the software and not simply due to responses to the intervention itself (Hawthorne effects)?

Natural or quasi experiments

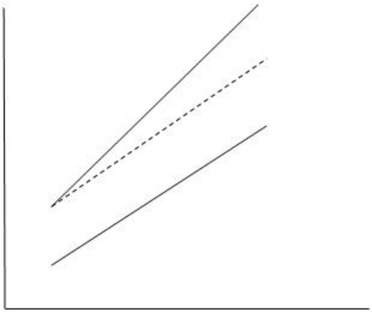
- In the absence of experiments, rely on exogenous changes in policy or circumstances to serve as quasi-experiments . Example: Mariel boatlift, draft lotteries, quarter of birth, test score cutoffs for admission.
- The challenge lies in justifying that the control groups are appropriate (that treatment and control status are as randomly assigned).

Difference-in-differences

- Why not just estimate a before and after difference for policy effects ? $[E(Y_1 \mid T) - E(Y_0 \mid T)]$
- Events occuring simultaneously, maturation.
- Compare outcomes before and after a policy change for a group affected by the change (Treatment Group) to a group not affected by the change (Control Group)
$DD = [E(Y_1 \mid T) - E(Y_0 \mid T)] - [E(Y_1 \mid C) - E(Y_0 \mid C)]$
- Paralell trend assumption: absent the policy change, the average change in Y_1 - Y_0 would have been the same for treatment and controls. The paralell trend assumption suggests the control group may act as a counterfactual.
- The quasi-exogenous nature of the treatment ideally addresses selection bias.

Difference in differences estimator

DD Estimator: $(y_1^T - y_1^C) - (y_0^T - y_0^C)$ or : $(y_1^T - y_0^T) - (y_1^C - y_0^C)$



Card and Krueger (1994)

- Card and Krueger (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania", American Economic Review (AER)
- Rise in minimum wage from $ 4.25 to $ 5.05 in April 1992 in the State of New Jersey.
- Impact on unskilled workers?

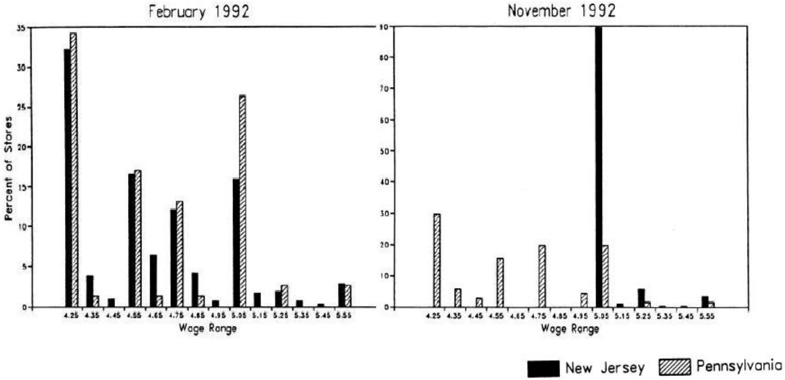
Card and Krueger (1994)

Card and Krueger collected data before and after the minimum wage increase in fast-food restaurants in New Jersey, and in the state of Pennsylvania (survey on wages, employment,...)

Card and Krueger (1994)

Paralell trend assumption: being located along the NJ or PA state line is viewed as randomly assigned- being subject to the minimum wage increase is assumed to be uncorrelated with other determinants of employment changes over the period.

Card and Krueger (1994)



Card and Krueger (1994)

- Causal effect of higher minimum wages? Naive (before and after) approach: $\bar{Y}_{NJ,November93}$ (\$5.05)-$\bar{Y}_{NJ,March93}$ (\$4.25)
- Counterfactual? Unobservable: $\bar{Y}_{NJ,November93}$(\$4.25)
- Use PA as counterfactual: $\bar{Y}_{PA,November93}$(\$4.25)

Card and Krueger(1994) estimation

- Estimatation of DD effect:

$$(y_1^T - y_1^C) - (y_0^T - y_0^C) \text{ or } :(y_1^T - y_0^T) - (y_1^C - y_0^C)$$

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

	Stores by state		
Variable	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)

Card and Krueger (1994)

- The small, albeit insignificant increase in employment in New Jersey makes it hard to accept the hypothesis that employment actually decreased in New Jersey over this time.
- Rise decided at the beginning of 1990. Economic recession in 1992: last-minute political action failed to stop the minimum-wage increase. Sensible to think that the shock was exogenous
- If the treatment and control groups have different time trends, the difference in difference estimator will be biased. One way to help avoid these problems is to get more data on other time periods before and after treatment to see if there are any other pre-existing differences in trends.

DD in regression format

- Where do the standard errors come from in the Card estimatates?:

$$(y_1^T - y_0^T) - (y_1^C - y_0^C)$$

- Regression model for just two periods estimated using data on fast food restaurants:
  $\Delta y_i = \alpha_0 + \beta\, Treatmentgroup_i + u_i$
- Adds some covariates $\Delta y_i = \alpha_0 + \beta\, Treatmentgroup_i + \gamma X_i + u_i$

Difference-in-differences regression

You could rewrite $\Delta y_i = \alpha_0 + \beta\, Treatmentgroup_i + u_i$ as
$Y_{it} = \alpha + \mu POST_i + \gamma Treatmentgroup_i + \beta POST_i \times Treatmentgroup_i + \varepsilon_{it}$

- POST=1 if November; Treatment=1 if fast food restaurant is in New Jersey, DD estimator is β
- OLS estimate is numerically identical to the DD estimate: $DD = [E(Y_1 \mid T) - E(Y_0 \mid T)] - [E(Y_1 \mid C) - E(Y_0 \mid C)]$

Panel data

Entity	*Time*	Y_{it}	$Treat_i$	$Post_t$	$Treat_i \times Post_t$
NJ Restaurant 1	Feb	20	1	0	0
NJ Restaurant 1	Nov	21	1	1	1
......	.....	...	...	....	...
NJ Restaurant N	Feb	22	1	0	0
NJ Restaurant N	Nov	24	1	1	1
PA Restaurant 1	Feb	23	0	0	0
PA Restaurant 1	Nov	21	0	1	0
......	.....	...	...	....	...
PA Restaurant M	Feb	22	0	0	0
PA Restaurant M	Nov	22	0	1	0

Difference in differences regression coefficient

$Y_{it} = \alpha + \mu POST + \gamma Treatmentgroup + \beta POST \times Treatmentgroup + \varepsilon_{it}$

$(y_1^T - y_0^T) - (y_1^C - y_0^C)$
$E(y_i \mid NJ, POST) - E(y_i \mid NJ, PRE) = \mu + \beta$
$E(y_i \mid PA, POST) - E(y_i \mid PA, PRE) = \mu$
DD estimator$= \beta$

You could also write the DD estimator as:
$(y_1^T - y_1^C) - (y_0^T - y_0^C)$
$E(y_i \mid NJ, POST) - E(y_i \mid PA, POST) = \gamma + \beta$
$E(y_i \mid NJ, PRE) - E(y_i \mid PA, PRE) = \gamma$
DD estimator$= \beta$

Difference-in-differences regression with multiple periods

- Why would we need more than two periods?

$Y_{it} = \alpha + \beta * POST_{it} + \gamma Treatmentgroup_{it} + \eta POST_{it} \times Treatmentgroup_{it} + \varepsilon_{it}$

becomes

$Y_{it} = \alpha + \lambda_{TIME} + \gamma Treatmentgroup_{it} + \eta POST_{it} \times Treatmentgroup_{it} + \varepsilon_{it}$

Diff in diff extensions

- Repeated cross sections: random samples from treatment and control groups each period. Same regression model, i now denotes individuals :

$$Y_{it} = \alpha + \beta * POST_{it} + \gamma TREATED_{it} + \eta POST_{it} \times TREATED_{it} + \varepsilon_{it}$$

- Multiple entities in the treatment group :
  $y_{it} = \beta_0 + \alpha_{ENTITY} + \lambda_{TIME} + \beta_{TREATED_{it} * POST_{it}} + u_{it}$
- Many changes and years can be pooled in a single regression: can be informative especially if similar reforms implemented in many different settings (states, regions, schools, etc.)
- Some drawbacks to pooled estimates, especially if already treated states used as controls. Can use only non-treated units as controls.