Applied Quantitative Methods II

Lecture 5: Difference in differences

Klára Kalíšková

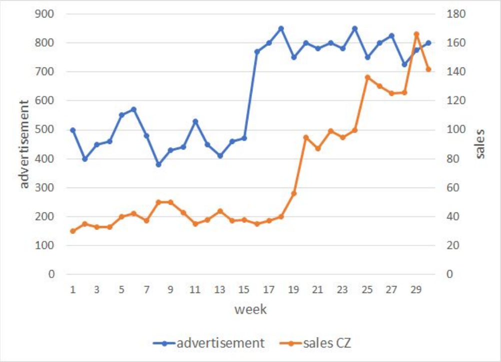
Outline

Motivation - example

- You work for a company selling a mobile app for job seekers offering a list of personalised job offers and a simple way to send job applications to companies
- The company have advertised the app on Google in the past year
- Now, they consider launching a bigger advertisement campaign to increase sales of the app and but would first like to know:
  - what is the impact of advertisement on sales
  - whether it is profitable to advertise
- How do you estimate the impact of adds on sales?

Motivation - example

- Take advantage of the variation in advertisement over time to see what was the impact on sales!



- Is this necessarily the impact of advertisement?

Motivation - example

- You found out the your company also sells in SK, but they do much less advertising there
  - They spend exactly 1/2 of the money on advertisement in SK than in CZ
- Take advantage of the variation in advertisement across two countries to see what was the impact on sales
- Sales in CZ are 1,7 times higher than in SK
  - Does it mean that twice as much advertisement causes 1,7 increase in sales?

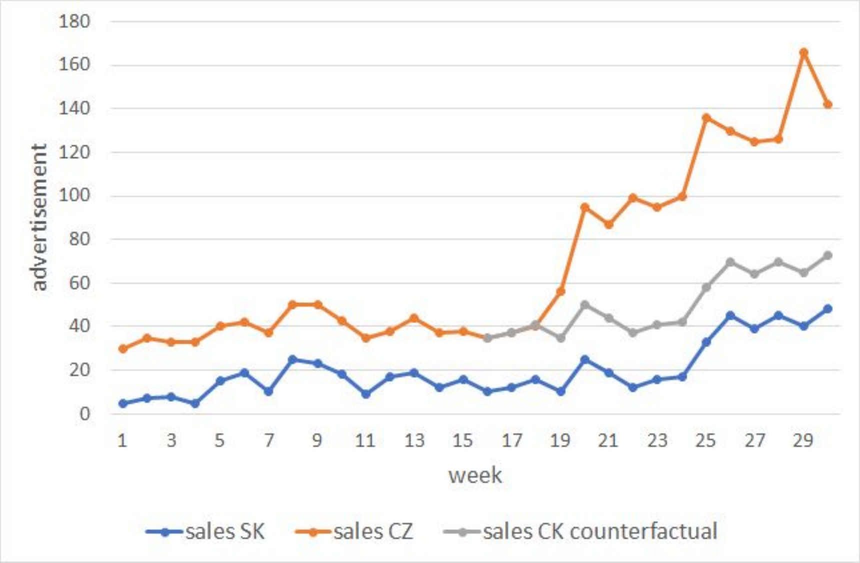
Diff-in-diff – Basic idea

- We already established that:
  - Comparison of before and after is problematic (aggregate trends, other factors)
  - Comparison of treated (T) and control (C) often suffers from selection bias (unobservable differences)
- What if we **combine the before-after and treated-control comparisons**?
  - Thus, **compare the change in outcome of the T group with the change in outcome of the C group**
  - Counterfactual: Use the change in outcome for the control group as a proxy for what would be the change in outcome of the treatment group if there was no treatment
- What you need for diff-in-diff estimation:
  - a treatment group that was affected and a control group that was not
  - at least 1 data point for before the policy and 1 for after the policy for both groups (better to have several observations before and after)

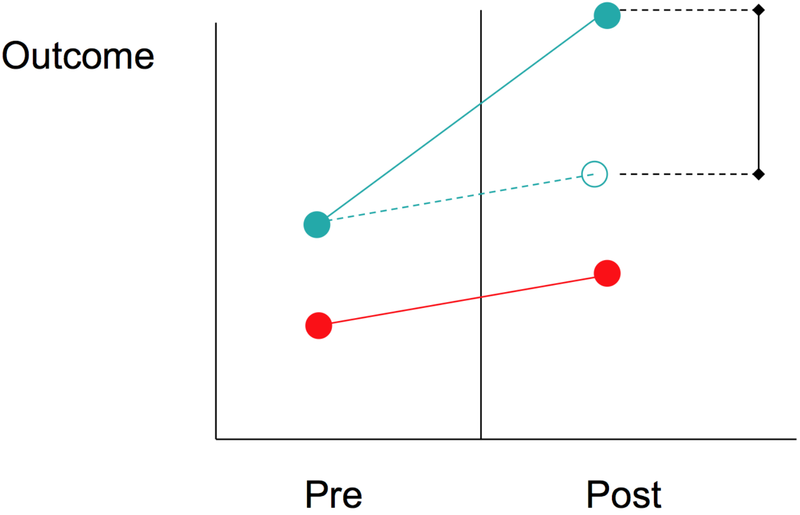
Diff-in-diff – Application to our example

- Your company increased addvertisement in CZ in week 16 to twice as much, but made no change to SK advertisement amount
- Take advatange of this variation in one country
- Use the change in sales in SK as a proxy (counterfactual) for what would be the change in sales in CZ had your company never increased advertisement there
- What is the main assumption here?

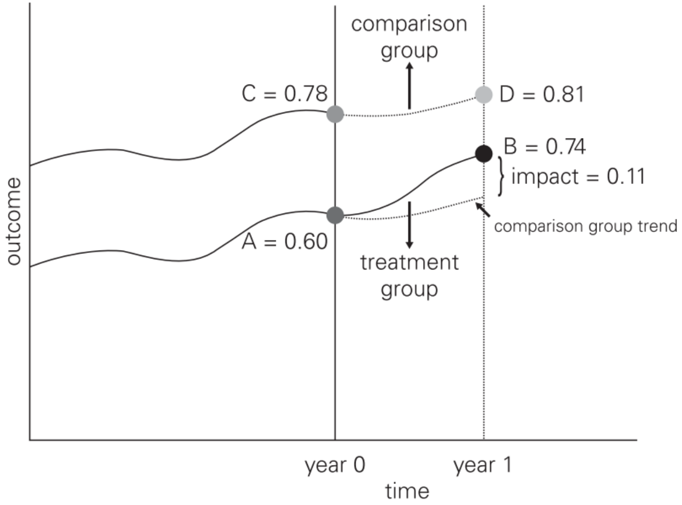
Diff-in-diff – Application to our example



Diff-in-diff – basic idea



Diff-in-diff – basic idea



Diff-in-diff – basic idea

- Write a table:

	After	Before
Treatment/enrolled	*B*	*A*
Comparison/ nonenrolled	*D*	*C*

Diff-in-diff – basic idea

- Comparison Treatment-control

	After	**Before**
Treatment/enrolled	B	A
Comparison/ nonenrolled	D	C
Difference	$B - D$	$A - C$

Diff-in-diff – basic idea

- Comparison before-after

	After	Before	Difference
Treatment/enrolled	B	A	$B - A$
Comparison/ nonenrolled	D	C	$D - C$

- influence of time-invariant chars cancel out
  - observed and **unobserved**!

Diff-in-diff – basic idea

- Final effect calculation: Difference in differences

	After	Before	Difference
Treatment/enrolled	B	A	$B - A$
Comparison/ nonenrolled	D	C	$D - C$
Difference	$B - D$	$A - C$	$DD = (B - A) - (D - C)$

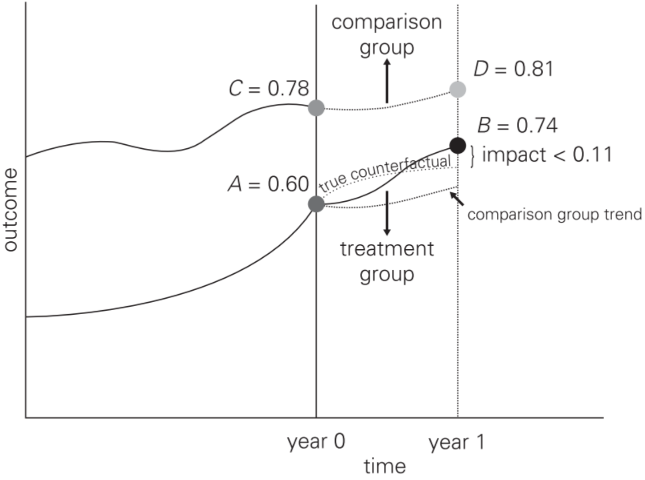
- Assume equal trends
  - no time-variying differences exist
  - no way to prove this

Diff-in-diff – basic idea

- Example with numbers:

	After	Before	Difference
Treatment enrolled	0.74	0.60	*0.14*
Comparison/ nonenrolled	0.81	0.78	*0.03*
Difference	*–0.07*	*–0.18*	*DD = 0.14 – 0.03 = 0.11*

Diff-in-diff – basic idea



Diff-in-diff – Theory

- Dif-in-dif: compare how outcome variable of the treatment group changed compared to the outcome variable of the control group:
  - basically, from previous slides: $DD = (B - A) - (D - C)$
- Formally: difference between population conditional means:

$$
\begin{aligned}
\tau &= \mathbb{E}\left[(Y_i(1) - Y_i(0)) \mid X\right] \\
&= \{\mathbb{E}[Y_{i,t=2} \mid G = T, X] - \mathbb{E}[Y_{i,t=2} \mid G = C, X]\} \\
&- \{\mathbb{E}[Y_{i,t=1} \mid G = T, X] - \mathbb{E}[Y_{i,t=1} \mid G = C, X]\}
\end{aligned}
$$

Diff-in-diff – Practice

- Regression:

$$Y_i = \alpha_0 + \alpha_1 G_i + \alpha_2 T_i + \tau(G_i \times T_i)_i + X_i\beta + \varepsilon_i$$

- G_i is a dummy for being in the Treatment group
- T_i is a dummy for post-treatment period
- $G_i \times T_i$ - interaction term: Treated-group X post-intervention

Diff-in-diff – Assumptions

- Have baseline data for both groups
  - 3 points: 2 before, 1 after
- Ind. in treatment group affected, in control unaffected by intervention
  - No/small cross-contamination
- No time-variying differences exist = proper control group
  - no way to prove this
  - check 1: if more than 1 period before treatment, check for tandem movement
  - check 2: placebo test - choose a "fake" treatment or control group
- Problem: Control and treatment experiencing different (macro) trends
  - Treated and untreated are in different markets
  - Cohort specific characteristics
    - e.g. Unemployment of youth compared to adults is more volatile

Macro trends

- Different macro trends affecting treatment and control group
- Example: generation specific characteristics
  - Cohort specific shocks (e.g. born before/after 1989)
  - Different trends for unemployment of older/younger people
- We need to check if they respond similarly in past

Diff-in-diff – Problems

- Calculate correct standard errors
  - When using more than two time periods
  - When the unit of observation is more detailed than the level of variation (e.g. individual person vs. the region)
- Ashenfelter's Dip
  - Selection on idiosyncratic temporary shocks
  - e.g. individuals experience a shock (unemployment) and enter program when things are especially bad
- Anticipation of policy step
- When treatment is a **choice** of participants

Diff-in-diff: Do we need panel data?

- Having cross-section is enough, if:
  - we can separate treatment and control group before and after the policy change
  - Sample representative with respect to the population
  - Composition of treated and untreated the same before and after the policy change
- We need slightly different specification for panel and cross-section data

Outline

Example: Health-insurance policy program (HISP)

- Setting:
  - 2 rounds of data for 2 groups of households
  - one group enrolled to HISP
    - Why? We don't know
  - thus comparison T vs C impossible

	After (follow-up)	**Before (baseline)**
Enrolled	7.8	14.4
Nonenrolled	21.8	20.6

- How to implement Dif-in-dif estimation?
  - table
  - regression
- What are the basic assumptions?

Example: Health-insurance policy program (HISP)

- Results

	After (follow-up)	Before (baseline)	Difference
Enrolled	7.8	14.4	−6.6
Nonenrolled	21.8	20.6	1.2
Difference			DD = −6.6 − 1.2 = −7.8

- Regression: $Y_i = \alpha_0 + \alpha_1 G_i + \alpha_2 T_i + \tau(G_i \times T_i)_i + X_{ij} + \varepsilon_i$; $\tau = -7.8$

	Linear regression	Multivariate linear regression
Estimated impact on household health expenditures	−7.8** (0.33)	−7.8** (0.33)

Outline

Diff-in-diff – policy evaluations

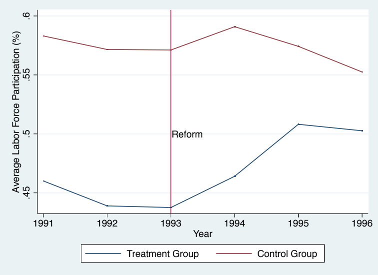
- Despite drawbacks, DiD one of most frequently used methods
- Easy to use method of causal-effect estimation
- *When timing of policy implementation differ for different states, countries, regions, we can have a lot of treatment and control groups*
    - Abortion law
    - Divorce law
- *When eligibility of individuals based on observables*
    - Tax law
    - Maternity leave

Diff-in-diff – Examples - policy evaluations

- Introduction of a tax credit that is applicable only to single mothers with children – Eissa&Liebman (1996)
  - women don't suddenly become single mothers to claim tax credits (at least not initially)
  - Control group: single women without children
- Increase in the minimum wage in one US state but not another – Card and Krueger (1993)
- Effect of tax changes in the UK on labor supply – Blundell et al. (1998)
- effect of flat tax reform in Russia on the size of the shadow economy (Gorodnichenko&al, 2009)
  - Control group: People who had the same marginal tax rate before and after the reform
- effect of no-fault divorce laws on divorce rates in European countries

Before we use diff-in-diff...

- Understanding the policy change in detail
  - Was it possible to anticipate it?
- What could be the correct control and treatment group?
  - What are differences between them?
  - Do they share same pre-treatment trend (common trend assumption)



  - Do they respond similarly to potential and plausible shocks?

Outline

Example 1 – effect of divorce laws

González and Viitanen, 2007 (EER)

- Pronounced rise in divorce rates in Europe since the 1960s
- Reform of divorce legislation –
  - allow divorce or “no-fault” or unilaterally
- **Question:** To what extend do these reforms making divorce easier contributed to increase in divorce rates?
- Panel of 18 countries for 1950-2003 (Eurostat)
  - Different timing of legislation introduction (Netherlands vs. Italy / Ireland)
  - Data on # of divorces, total population and married population + control variables
- Dependent variable = annual divorces / 1000 people, alternative is annual divorces / 1000 married people
  - Effect of law on quality and quantity of matches

Example 1 – effect of divorce laws

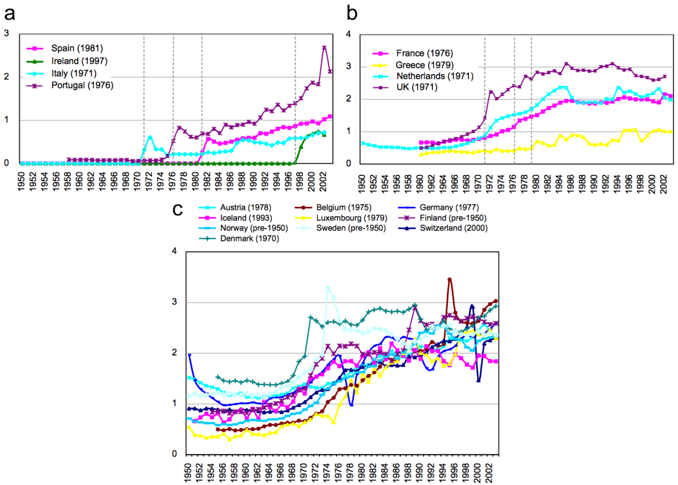


Fig. 2. Divorce rates in 18 European countries, 1950–2003. (a) Countries that legalized divorce during the period. (b) Countries that introduced no-fault during the period (excluding those in (a)). *Note:* The dotted lines indicate the years when the reforms took place. (c) Countries that introduced no-fault pre-1950, unilateral during the period.

Example 1 – effect of divorce laws

Methodology

- Comparison of reform and control countries
  - Are legislative reforms across countries exogenous?
  - Probably a difference in observable/unobservable traits

$$divorce_{it} = \beta law_{it} + \sum_i c_i + \sum_t T_t + \sum_i c_i \times time_t + X_{it}\gamma + \varepsilon_{it}$$

  - law_{it} – dummy if the reform is effective in country i in year t
    - β- average rise in divorce rates
  - c_i – country fixed effects - preexisting differences
  - T_t – year fixed effects -
  - $c_i \times time_t$ – country specific time trend

Example 1 – effect of divorce laws
Methodology II

- The preceding methodology captures only a discrete series break

$$div_{it} = \sum_{k>1} \beta_k law\ eff\ for\ k\ periods_{it} + \sum_i c_i + \sum_t T_t + \sum_i c_i \times time_t + X_{it}\gamma +$$

- this captures dynamic response of divorce while country spec. effect identify pre-existing trends

Example 1 – effect of divorce laws

Results I - legalizing divorce

Table 2
Static and dynamic effects of legalizing divorce; dependent variable: annual divorces per 1000 people

	Static 1	Dynamic 1	Static 2	Dynamic 2	Static 3	Dynamic 3
Legal	0.343*** (0.039)		0.354*** (0.027)		0.323*** (0.130)	
Legal yrs 1–2		0.419*** (0.050)		0.420*** (0.031)		0.388*** (0.039)
Legal yrs 3–4		0.394*** (0.043)		0.392*** (0.022)		0.361*** (0.054)
Legal yrs 5–6		0.331** (0.098)		0.378*** (0.022)		0.339*** (0.045)
Legal yrs 7–8		0.274* (0.093)		0.356*** (0.042)		0.312*** (0.044)
Legal yrs 9–10		0.350** (0.094)		0.424*** (0.069)		0.378*** (0.059)
Legal yrs 11–12		0.423** (0.113)		0.482*** (0.037)		0.426*** (0.065)
Legal yrs 13–14		0.362* (0.140)		0.447*** (0.014)		0.388*** (0.087)
Legal yrs 15+		0.453* (0.182)		0.537*** (0.049)		0.468*** (0.086)
Country trends	No	No	Yes ($F = 2819$)	Yes ($F = 1038$)	Yes ($F = 2200$)	Yes ($F = 86$)
Quadratic trends	No	No	No	No	Yes ($F = 40285$)	Yes ($F = 663$)
Adjusted R^2	0.927	0.931	0.963	0.965	0.973	0.975

Sample: 1950–2003, $n = 206$ (unbalanced panel). Estimated using country population weights. All specifications include dummies for year and country as well as country-specific controls for total fertility rate, unemployment rate and female labor force participation rate and dummies if they are missing for any year. Standard errors are clustered by country and shown in parentheses.
*** Statistical significance at 1%.
** Statistical significance at 5%.
* Statistical significance at 10% level.

- Legalizing divorce increases divorce rate by 0.32-0.35 divorces / 1000 people
- Dynamic specification suggests U-shaped effect

Table 3
Static and dynamic effects of divorce law change (no fault); dependent variable: annual divorces per 1000 people

	Static 1	Dynamic 1	Static 2	Dynamic 2	Static 3	Dynamic 3
No fault	0.175 (0.163)		0.549*** (0.115)		0.262*** (0.055)	
No fault yrs 1–2		0.417*** (0.138)		0.629*** (0.099)		0.407*** (0.088)
No fault yrs 3–4		0.586*** (0.113)		0.860*** (0.087)		0.543*** (0.066)
No fault yrs 5–6		0.905*** (0.148)		1.063*** (0.123)		0.675*** (0.103)
No fault yrs 7–8		1.129*** (0.237)		1.397*** (0.158)		0.851*** (0.121)
No fault yrs 9–10		1.447*** (0.252)		1.613*** (0.132)		0.982*** (0.123)
No fault yrs 11–12		1.729*** (0.339)		1.646*** (0.093)		0.980*** (0.042)
No fault yrs 13–14		1.909*** (0.373)		1.670*** (0.083)		0.960*** (0.066)
No fault yrs 15+		2.218*** (0.324)		1.818*** (0.060)		1.01*** (0.086)
Country trends	No	No	Yes ($F = 203$)	Yes ($F = 78$)	Yes ($F = 221$)	Yes ($F = 158$)
Quadratic trends	No	No	No	No	Yes ($F = 93$)	Yes ($F = 133$)
Adjusted R^2	0.922	0.948	0.962	0.980	0.982	0.986

Sample: 1950–2003, $n = 185$ (unbalanced panel). Estimated using country population weights. All specifications include dummies for year and country as well as country-specific controls for total fertility rate, unemployment rate and female labor force participation rate and dummies if they are missing for any year. Standard errors are clustered by country and shown in parentheses.
*** Statistical significance at 1%.

Example 1 – effect of divorce laws

Results III - unilateral

Table 4
Static and dynamic effects of divorce law change (unilateral); dependent variable: annual divorces per 1000 people

	Static 1	Dynamic 1	Static 2	Dynamic 2	Static 3	Dynamic 3
Unilateral	0.083 (0.142)		0.400*** (0.112)		0.243** (0.076)	
Unilateral yrs 1–2		0.052 (0.077)		0.254** (0.083)		0.118 (0.073)
Unilateral yrs 3–4		0.148 (0.102)		0.386*** (0.111)		0.181* (0.095)
Unilateral yrs 5–6		0.246 (0.166)		0.563*** (0.153)		0.286* (0.147)
Unilateral yrs 7–8		0.345* (0.173)		0.712*** (0.183)		0.351* (0.160)
Unilateral yrs 9–10		0.174 (0.204)		0.613** (0.215)		0.201 (0.201)
Unilateral yrs 11–12		0.066 (0.241)		0.582* (0.294)		0.113 (0.249)
Unilateral yrs 13–14		−0.150 (0.215)		0.414 (0.264)		−0.081 (0.246)
Unilateral yrs 15+		0.003 (0.206)		0.644* (0.301)		−0.114 (0.242)
Country trends	No	No	Yes	Yes	Yes	Yes
Quadratic trends			($F = 2.8$e+06)	($F = 2.9$e+05)	($F = 2.3$e+07)	($F = 4.9$e+05)
Adjusted R^2	No	No	No	No	Yes	Yes

Sample: 1950–2003, $n = 525$ (unbalanced panel). Estimated using country population weights. All specifications include dummies for year and country as well as country-specific controls for total fertility rate, unemployment rate and female labor force participation rate and dummies if they are missing for any year. Standard errors are clustered by country and shown in parentheses.

*** Statistical significance at 1%.

** Statistical significance at 5%.

* Statistical significance at 10%.

Outline

Example 2: Malaria eradication campaign

Malaria Eradication in the Americas (Bleakley, 2010)

- Question: How much childhood exposure to malaria depresses labor productivity?
- Data: Malaria Eradication campaign in Southern United States (1920's)
  - + Brazil, Colombia, Mexico (1950's)
- Diff-in-Diff:
  1. birth cohorts - old vs. young people at the time of campaign (only younger were exposed to campaign)
  2. regions with high vs. low incidence of malaria (only former should experience an effect)

Example 2: Malaria eradication campaign
Methodology

- Areas with high pre-treatment malaria will benefit more from malaria eradication
- Treatment group: Young people living in high pre-treatment malaria areas $=>$ should experience the full effect of campaign
- Comparison group: young and older people living in low pre-treatment malaria areas $=>$ control for natural evolution of income over cohorts (without malaria)

Example 2: Malaria eradication campaign
Methodology

$$Y_{jkt} = \beta_k M_j + \delta_k + X_j \Gamma_k + \nu_{jkt}$$

where:

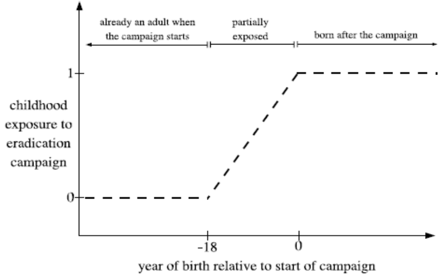
- Y_{jkt}– average outcome (income) in state j for cohort k at time t
- M_j– pre-campaign malaria intensity in state j
- β_k– cohort-specific coefficient on malaria
- X_j – state controls (health and education related)

They have run this separately for each cohort k and obtained β_k

Example 2: Malaria eradication campaign

Hypothesis about β_k (exposure to malaria in younger age has effect):

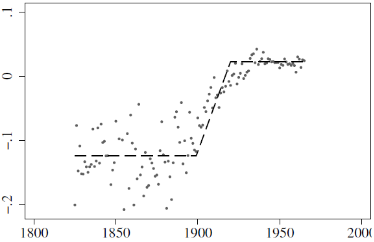
- For older cohorts (before 1900) – negative relationship between malaria intensity and outcomes
- For younger cohorts (after 1920) – relationship was purged by the effect of campaign
- In-between – decreasing strengh of the relationship (more and more exposure to campaign in the childhood)



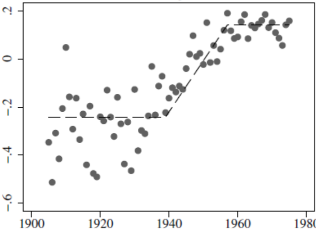
Example 2: Malaria eradication campaign

Results

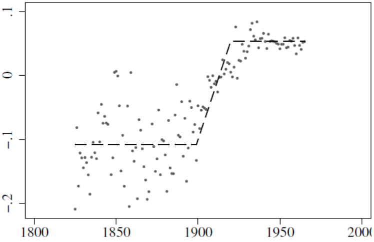
Basic Specification, Occupational Income Score



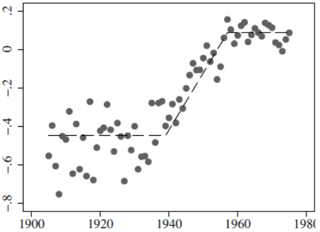
Brazil, Basic Specification



Additional controls, Occupational Income Score



Brazil, Additional Controls



Outline

Example 3: D-in-D-in-D

- Implementation of (imaginary) health care policy, aiming at people of age 65 and older in country A
- Looking at effect on health outcomes (y)
- DD approach:
  - 2 periods (before x after);
  - control group = people of age 55-65 ?
- What problems do you see?

Example 3: D-in-D-in-D

- Let's use elderly patients from the country B, where the health reform wasn't introduced at all
- Assuming that age effects are the same across countries 3 dummies:
    - Age eligibility: $d_i = 1$ if age of person i >65
    - Time eligibility: $T_t = 1$ if time period t is AFTER
    - Country identificator: $A_i = 1$ if person i from country A

$$y_{it} = \beta_0 + \beta_1 d_i + \beta_2 A_i + \beta_3 d_i A_i + \\ + \delta_0 T_t + \delta_1 T_t A_i + \delta_2 d_i T_t + \delta_3 d_i T_t A_i + \varepsilon_{it}$$

Example 3: D-in-D-in-D

$$\begin{aligned}\hat{\delta}_3 =&(\hat{y}_{A,d=1,t=2} - \hat{y}_{A,d=1,t=1})\\ &-(\hat{y}_{B,d=1,t=2} - \hat{y}_{B,d=1,t=1})\\ &-(\hat{y}_{A,d=0,t=2} - \hat{y}_{A,d=0,t=1})\end{aligned}$$

- By including different control groups, we hope to control for different confounding factors
  - Cohort (age) specific – comparing with same cohort from B
  - State specific – comparing with younger cohort from A

Outline

Reality check – Bertrand et al. (2004)

- How much should we trust diff-in-diff estimates?
- General specification of D-in-D model:

$$Y_{ist} = A_s + B_t + cX_{ist} + \beta I_{st} + \varepsilon_{ist}$$

  - A_s– state (group) fixed effect
  - B_t– time fixed effect effect
  - X_{ist}– individual controls
  - I_{st}– indication whether policy has effect on state s at time t
- Usually cluster by year & state (group) Are standard errors OK?

Reality check – Bertrand et al. (2004)

- How does DD perform on placebo laws?
- Take typical data used in DD estimations CPS, women 25-50 with positive earnings, 50 years
- Assign randomly treated states and years of introduction
- "If hundreds of researchers analyzed the effects of various laws in the CPS, what fraction would find a significant effect even when laws have no effect?"
- Significant effect at 5% level should be found in ?? % of cases?

Reality check – Bertrand et al. (2004)

- Result: Bertrand et al. has found significant effect in 45% of cases!! (even after clustering)
  - Reason = serial (time) correlation problem
    - Use of fairly long time-series (avg.16.5 periods)
    - Dependent variables (e.g. income) are typically highly positively serially correlated (And not only AR(1))
  - Treatment variable has small variation over time; usually 0 before and 1 after – think malaria

Reality check – Bertrand et al. (2004)

- Solution: Block-bootstrapping: OK if large number of groups
- Aggregate data to 2 periods – before and after, for each group (small # of groups)
- Allow for unrestricted covariance over time within states – cluster on states!!! (EASY, but also for larger number of groups)

Outline

Summary

- We covered diff-in-diff method of estimation of a causal effect
  - it works when selection to treatment group is external to our model
  - i.e. once we control for observable characteristics, the treatment dummy is exogenous
- Some drawbacks – assumptions
- Still, one of very popular methods

Bibliography

- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates ? Quarterly Journal Of Economics, (February).
- Bleakley, H. (2010). Malaria eradication in the americas: A retrospective analysis of childhood exposure. American Economic Journal: Applied Economics, 2(2), 1–45. doi:10.1257/app.2.2.1
- González, L., & Viitanen, T. K. (2009). The effect of divorce laws on divorce rates in Europe. European Economic Review, 53(2), 127–138. doi:10.1016/j.euroecorev.2008.05.005
- Barrera-Osorio, F., & Linden, L. L. (2009). The use and misuse of computers in education: Evidence from a randomized experiment in Colombia. Policy Research Working Paper, (February), 43. Retrieved from http://go.worldbank.org/BZZT7KNLG0

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY