corinna.birner@stud-mail.uni-wuerzburg.de
max.mueller@stud-mail.uni-wuerzburg.de

Understanding Metrics based on Mastering Metrics

Corinna Birner & Max Müller

University of Würzburg

July 15, 2020

Chapter 5: Differences in Differences

Overview

Today we will continue our journey on the path from cause to effect. Therefore, we will discuss Differences in Differences designs as one possible way of finding causal relations in our data. Our topics will be the following:

Introduction

- most of the times, conducting a pure regression will not deliver a credible causal relationship
- but randomized experiments are expensive and sometimes not feasible
- it is also hard to find good instruments
- clear discontinuities are not always available
- we need another tool in our metrics tool kit!

Differences-in-Differences (DiD) method

- method recognizes that in absence of random assignment, treatment and control groups are likely to differ in a lot of ways
- sometimes the treatment and control outcomes still move in a parallel way over time in the absence of treatment
- the DiD takes the time dimension into account
- in that case, the difference of a post-treatment path from the trend (established by a comparison group) may signal the treatment effect

A Mississippi Experiment

- banking is built on confidence
- maturity mismatch: banks don't hold enough money to pay back all depositors at once
- normally, this is not a problem because depositors won't withdraw all their money at the same time
- during a crisis, if you are the last one that tries to get the money from the bank, you might lose all your savings - if you see people running to get their money from the bank, you should worry!

the Great Depression

- before the Great Depression in the early 1920s, the US economy was prospering
- banks were prospering and wealthy
- after the stock market crash in October 1929 Caldwell and Company collapsed
- this event let to a confidence crisis throughout the American South
- in Mississippi people wanted to withdraw their savings in Dec 1930: bank run
- in 1933 the depression-era experienced its peak
- during this time the Federal Reserve used a tight monetary policy - was that a good idea?

A Mississippi Experiment

- Should the central bank intervene?
- easy credit might have helped with the urgent withdrawal demands and prevented a panic
- but if banks are really in trouble, temporary liquidity wont hold for a long term
- moral hazard as another issue
- providing aditional liquidity vs. speed up survival of the fittest bank: which scenario is more likely to end an economic downturn?

A Mississippi Experiment

- the US Federal Reserve Bank is organized into 12 districts, where each is governed by a regional Federal Reserve Bank
- Atlanta Fed (sixth district): providing aditional liquidity to banks through loans
- St.Louis Fed (eigth district): restricted credit during recession; Real Bills Doctrine
- border between the districts runs through the state of mississippi
- this provides a within-state natural experiment!

A DiD Mississippi Experiment

- was the lending policy of the Atlanta Fed successfull?

A Mississippi Experiment

- compare outcome Y: number of banks in each district on July 1,1931:
  - sixth district: 132
  - eight district: 121
- but the number of banks was diffent before July 1, 1930:
  - sixth district: 165
  - eight district: 135
- Set Y_{dt} for the number of banks open in the district d in the year t and δ_{DD} the DD estimate of the causal effect of easy money in the sixth district:

$$\delta_{DD} = (Y_{6,1931} - Y_{6,1930}) - (Y_{8,1931} - Y_{8,1930})$$

$$= (121 - 135) - (132 - 165) = 19$$

A DiD Mississippi Experiment

- instead of comparing outcomes, DD compares the change in the outcome between the districts
- comparing changes and not levels will help us adjust for differences in the pre-treatment period
- alternative form:

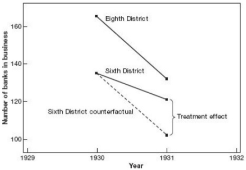
$$\delta_{DD} = (Y_{6,1931} - Y_{8,1931}) - (Y_{6,1930} - Y_{8,1930})$$

$$= (121 - 132) - (135 - 165) = 19$$

Logic of a DiD

- Central Assumption: common trends
- that means that in absence of treatment, the trend of treatment and control groups would have followed a parallel trend
- in our example: the trend in the eight districts is what we should have seen in the sixth district.
- you should test this assumption!

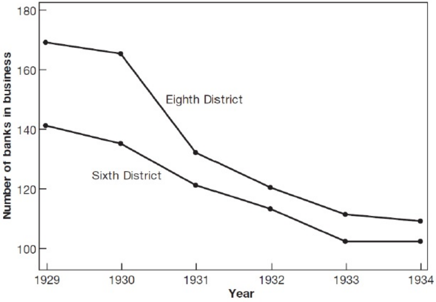
Common Trend Assumption



Notes: This figure shows the number of banks in operation in Mississippi in the Sixth and Eighth Federal Reserve Districts in 1930 and 1931. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district in this period as did in the Eighth.

- dotted line: counterfactual outcome
- shows what would have happened, had they evolved the same way

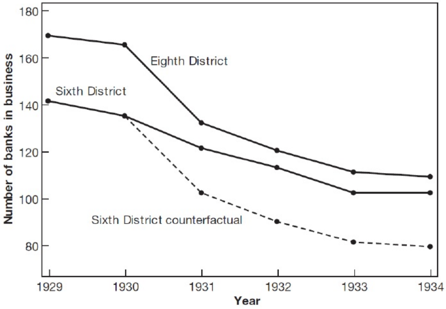
Is there a common trend?



Note: This figure shows the number of banks in operation in Mississippi in the Sixth and Eighth Federal Reserve Districts between 1929 and 1934.

- long time series on bank activity
- before 1930, similar policy
- bank failure moved almost in parallel in the two districts
- banks declining slightly in both districts
- common trends assumption seems valid

Adding the Sixth District counterfactual



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

- counterfactual implied by extrapolating Eight District trends ti the Sixth
- in 1931, St.Louis Fed started lending money and the two districts had a similar policy again

A Depression Regression

- we will use regression models fit to samples of more than four data points and more than two cross-sectional units
- our regession DD recipe contains
  - a dummy for the treatment $TREAT_d$ (equals 1 for data points from Sixths Districts)
  - a dummy for post-treatment periods $POST_t$ (observations from 2931 onward)
  - the interaction term $TREAT_d \times POST_t$ (observations in the Sixth District in periods when Fed's response mattered for number of active banks)
- what would our regression equation look like?

A Depression Regression

- putting our three ingredients together, we obtain:

$$Y_{dt} = \alpha + \beta TREAT_d + \gamma POST_t + \delta_{rDD}(TREAT_d \ x \ POST_t) + e_{dt}$$

- this sample is constructed by tacking observations from both districts and six available years for each district
- δ_{rDD} is our causal effect of interest

A Depression Regression

- fitting this equation to our 12 observations from Figure 5.2. we get:

$$Y_{dt} = 167 - 29 TREAT_d - 49 POST_t$$

$$+20.5_{rDD}(TREAT_d \ x \ POST_t) + e_{dt}$$

- standard errors β: 8.8 , γ: 7.6, δ: 10.7
- this suggests that roughly 21 banks were kept alive by Sixth District lending
- looking at the standard error for δ, this is a marginally significant result

Did the policy support real economic activity?

	1929	1933	Difference (1933–1929)
Panel A. Number of wholesale firms			
Sixth Federal Reserve District (Atlanta)	783	641	–142
Eighth Federal Reserve District (St. Louis)	930	607	–323
Difference (Sixth–Eighth)	–147	34	181
Panel B. Net wholesale sales ($ million)			
Sixth District Federal Reserve (Atlanta)	141	60	–81
Eighth District Federal Reserve (St. Louis)	245	83	–162
Difference (Sixth–Eighth)	–104	–23	81

Notes: This table presents a DD analysis of Federal Reserve liquidity effects on the number of wholesale firms and the dollar value of their sales, paralleling the DD analysis of liquidity effects on bank activity in Figure 5.1.

- DD analysis of Fed Reserve liquidity effects on the number of active wholesalers and their sales
- the reduction in bank credit in the Eigth District brought wholesale business activity down unlike in the Sixth
- as only four-number DD, the evidence for a liquidity treatment effect is weaker than that produced by our larger sample for bank activity

DiD on the MLDA

- Alabama lowered its MLDA to 19 in 1975, geographically near Arkansas has had an MLDA of 21.
- Did Alabama's indulgence of its youthful drinkers cost some of them their lives?
- We want to find out by fitting a regression DD model to data on the death rates of 18–20-year-olds from 1970 to 1983.
- The dependent variable is denoted Y_{st}, for death rates in state s and year t.

$$Y_{st} = \alpha + \beta \, TREAT_s + \gamma \, POST_t + \sigma_{rDD}(TREAT_s \times POST_t) + \epsilon_{st}$$

Explaining the regression

- $TREAT_s$ is a dummy variable indicating Alabama.
- $POST_t$ is a dummy indicating years from 1975 onward.
- Interaction term $TREAT_s \times POST_t$ indicates Alabama observations from low-drinking-age years.
- The coefficient σ_{rDD} captures the effect of an age-19 MLDA on death rates.

Larger Sample = More Success

- Remember: Larger sample produces more precise results.
- So: We should take all districts with different MLDA laws into account.
- We swap the single $POST_t$ dummy for a set of dummies indicating each year in the sample, with one omitted as a reference group.
- The coefficients on these dummies, known as time effects, capture temporal changes in death rates that are common to all states.

Multistate setup

- Multistate setup controls for the differing death rates in each of many states.
- This is accomplished by introducing state effects, a set of dummies for every state in the sample, except for one, which is omitted as a reference group.
- A regression DD analysis of data from Alabama, Arkansas, and Tennessee, for example, includes two state effects.
- State effects replace the single $TREAT_s$ dummy included in a two-state (or twogroup) analysis.

Explaining the regression

- A final complication in this scenario is the absence of a common treatment variable that discretely switches off and on.
- We replace $TREAT_d POST_t$ with a variable well call $LEGAL_{st}$.
- This variable measures the proportion of 18–20-year-olds allowed to drink in state s and year t.
- In some states, no one under 21 is allowed to drink, while in states with an age-19 MLDA, roughly two-thirds of 18–20-year-olds can drink, and in states with an age-18 MLDA, all 18–20-year-olds can drink

Explaining the regression

- Multistate regression DD model looks like:

$$Y_{st} = \alpha + \sigma_{rDD} LEGAL_{st} + \sum_{k=Alaska}^{Wyoming} \beta_k \; STATE_{ks} + \sum_{j=1971}^{1983} \gamma_j \; YEAR_{jt} + \epsilon_{st}$$

- Here every state but one (the reference state) gets its own dummy variable, indexed by the subscript k for state k.
- Observations from California, for example, have $STATE_{CA}$, s switched on, and all other state dummies switched off.

Explaining the regression

- β_k, are the coefficients on the state dummies.
- For example, the California state effect, β_{CA} is the coefficient on $STATE_{CA,s}$
- Every state except the reference state, the one omitted when constructing state dummies, has a state effect in this equation.
- Same goes for year effects, e.g. The 1975 year effect, γ_{1975}, is the coefficient on $YEAR1975, t$.

Explaining the regression

- Our multistate MLDA analysis uses a data set with 14 years and 51 states for a total of 714 observations.
- This data structure is called a state-year panel.
- The state effects control for fixed differences between states (for example, fatal car accidents are more frequent, on average, in rural states with high average travel speeds).
- The time (year) effects in this equation control for trends in death rates that are common to all states (due, for example, to national trends in drinking or vehicle safety).
- This equation attributes changes in mortality within states to changes in $LEGAL_{st}$.

Effect of different MLDA on death

TABLE 5.2
Regression DD estimates of MLDA effects on death rates

Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.80	8.47	12.41	9.65
	(4.59)	(5.10)	(4.60)	(4.64)
Motor vehicle accidents	7.59	6.64	7.50	6.46
	(2.50)	(2.66)	(2.27)	(2.24)
Suicide	.59	.47	1.49	1.26
	(.59)	(.79)	(.88)	(.89)
All internal causes	1.33	.08	1.89	1.28
	(1.59)	(1.93)	(1.78)	(1.45)
State trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds. The table shows coefficients on the proportion of legal drinkers by state and year from models controlling for state and year effects. The models used to construct the estimates in columns (2) and (4) include state-specific linear time trends. Columns (3) and (4) show weighted least squares estimates, weighting by state population. The sample size is 714. Standard errors are reported in parentheses.

- Legal alcohol access caused about 11 additional deaths per 100,000 18–20-year-olds, of which seven or eight deaths were the result of motor vehicle accidents.
- Generates little evidence of an effect of legal drinking on deaths from internal causes.

Always common trends?

- Samples that include many states and years allow us to relax the common trends assumption, that is, to introduce a degree of nonparallel evolution in outcomes between states in the absence of a treatment effect.
- Such a regression looks like:

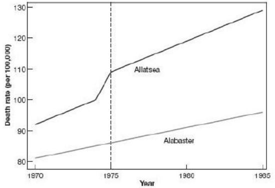
$$Y_{st} = \alpha + \sigma_{rDD} LEGAL_{st} + \sum_{k=Alaska}^{Wyoming} \beta_k \; STATE_{ks}$$

$$+ \sum_{j=1971}^{1983} \gamma_j \; YEAR_{jt} + \sum_{k=Alaska}^{Wyoming} \theta_k (STATE_{ks} \times t) + \epsilon_{st}$$

- In the absence of a treatment effect, death rates in state k deviate from common year effects by following the linear trend captured by the coefficient θ_k.

Uncommon Trends

FIGURE 5.6
A real MLDA effect, visible even though trends are not parallel



- Regression DD captures treatment effects with of uncommon trends.
- Death rates in Allatsea increase more steeply than in Alabaster.
- But the Allatsea increase is especially steep from 1974 to 1975, when Allatsea lowered its MLDA.
- The coefficient on $LEGAL_{st}$ picks this up, while the model allows for the fact that death rates in different states were on different trajectories from the get-go.

Controls

- But: We should control for other factors too.
- Important consideration in research on alcohol is the price of a drink.
- Taxes are the most powerful tool the government uses to affect the price.
- States might raise tax rates at the same time that they increase their MLDA.

With Controls

TABLE 5.3
Regression DD estimates of MLDA effects controlling for beer taxes

	Without trends		With trends	
Dependent variable	Fraction legal (1)	Beer tax (2)	Fraction legal (3)	Beer tax (4)
All deaths	10.98 (4.69)	1.51 (9.07)	10.03 (4.92)	−5.52 (32.24)
Motor vehicle accidents	7.59 (2.56)	3.82 (5.40)	6.89 (2.66)	26.88 (20.12)
Suicide	.45 (.60)	−3.05 (1.63)	.38 (.77)	−12.13 (8.82)
Internal causes	1.46 (1.61)	−1.36 (3.07)	.88 (1.81)	−10.31 (11.64)

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds, controlling for state beer taxes. The table shows coefficients on the proportion of legal drinkers by state and year and the beer tax by state and year, from models controlling for state and year effects. The fraction legal and beer tax variables are included in a single regression model, estimated without trends to produce the estimates in columns (1) and (2) and estimated with state-specific linear trends to produce the estimates in columns (3) and (4). The sample size is 700. Standard errors are reported in parentheses.

- Generate MLDA estimates similar to those without such controls.
- The beer tax estimates from models that include state trends are especially noisy.

Weighted least squares (WLS)

- States are not created equal: some, like Texas and California, are bigger than most countries, while others, like Vermont and Wyoming, have populations smaller than those of many American cities.
- Solution: Use weighted least squares (WLS)
- Standard OLS estimator fits a line by minimizing the sample average of squared residuals, with each squared residual getting equal weight in the sum.
- WLS weights each term in the residual sum of squares by population size or some other researcher-chosen weight.
- Capture a weighted average of effects for the groups or cells represented in our data

In a nutshell

- Treatment and control groups may differ in the absence of treatment, yet move in parallel.
- This pattern opens the door to DD estimation of causal effects.
- Comparing changes instead of levels, we eliminate fixed differences between groups that might otherwise generate omitted variables bias.
- It has power and flexibility, in a state-year panel we need only control for state and year effects.
- Important: Parallel trends, the claim that in the absence of treatment, treatment and control group outcomes would move in parallel (Although we can control for nonparallel trends).

Serial correlations

- Regression DD is a special case of estimation with panel data.
- A state-year panel consists of repeated observations on states over time.
- The repetitive structure of such data sets raises special statistical problems.
- Economic data of this sort typically exhibit a property called serial correlation.
- Serially correlated data are persistent, meaning the values of variables for nearby periods are likely to be similar.
- When the dependent variable in a regression is serially correlated, the residuals from any regression model explaining this variable are often serially correlated as well.

Serial correlations

- Serial correlation is a deviation from randomness, with the important consequence that each new observation in a serially correlated time series contains less information than would be the case if the sample were random.
- The formula used in this case is the clustered standard errors formula
- Clustering allows for correlated data within researcher defined clusters.

Clustered standard errors

- In contrast with the assumption that all data are randomly sampled, the formula for clustered standard errors requires only that clusters be sampled randomly, with no random sampling
- Assumption invoked for what's inside them.
- In principle, clustering solves any sort of dependence problem in your data
- Especially useful for panel data
- But: Large standard errors as a result

Thank you very much for listening