Public Finance

Davide Cipullo

Università Cattolica del Sacro Cuore

a.a. 2025/2026

Review of empirical tools

Definitions

Definition

Empirical public finance: The use of data and statistical methods to measure the impact of government policy on individuals and markets

- **Example**: estimate how labor supply varies due to a tax reform

Definition

Correlation: Two economic variables are correlated if they move together

- **Example**: height and weight across individuals

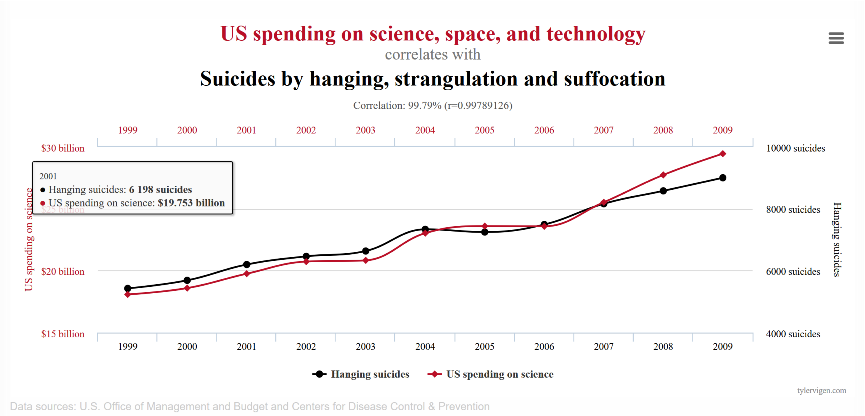
Definition

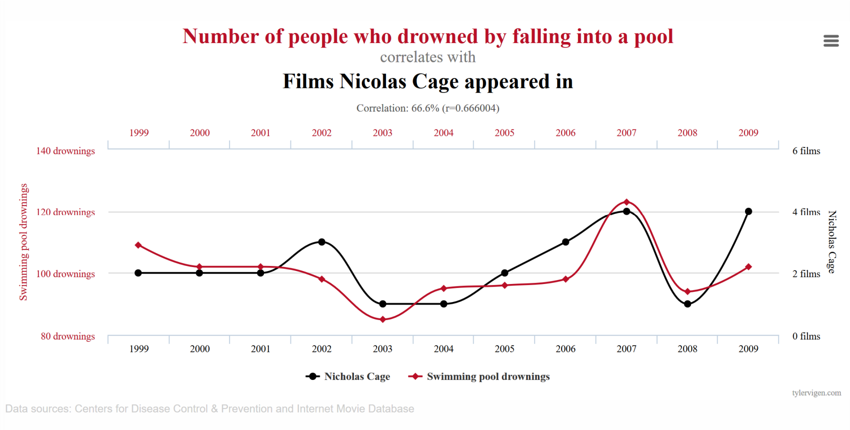
Causality: One economic variable causes another if the movement of the former variable causes a movement of the other variable

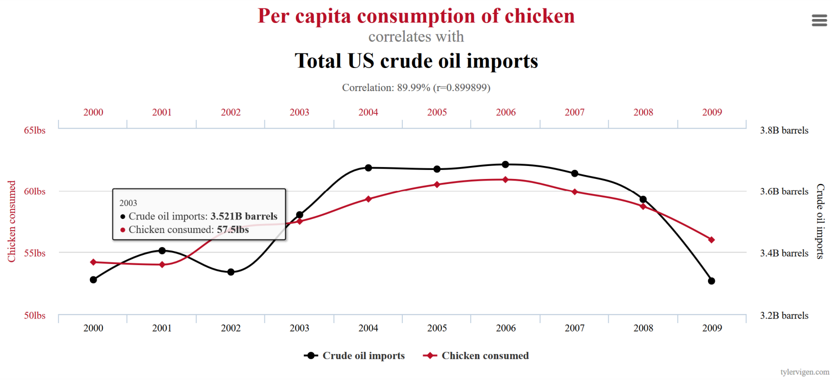
- **Example**: good nutrition as an infant increases adult height

Correlation $\neq$ causation

- There are many examples where causation and correlation can get confused
- In statistics, this is called the *identification problem*: given that two series are correlated, how do you identify whether one series is causing another?


US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation
Correlation: 99.79% (r=0.99789126)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
$30 billion
$25 billion
$20 billion
$15 billion
US spending on science
10000 suicides
8000 suicides
6000 suicides
4000 suicides
Hanging suicides
2001
Hanging suicides: 6 198 suicides
US spending on science: $19.753 billion
Hanging suicides
US spending on science
tylervigen.com
Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
Correlation: 66.6% (r=0.666004)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drownings
120 drownings
100 drownings
80 drownings
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
Nicholas Cage
Swimming pool drownings
tylervigen.com
Data sources: Centers for Disease Control & Prevention and Internet Movie Database



Data sources: U.S. Department of Agriculture and Dept. of Energy

The identification problem

- The attempt to interpret a correlation as a causal relationship without sufficient thought to the underlying process generating the data is a common problem.
- For any correlation between two variables A and B, there are three possible explanations, one or more of which could result in the correlation:
    1. A is causing B
    2. B is causing A
    3. Some third factor is causing both
- The general problem that empirical economists face in trying to use existing data to assess the causal influence of one factor on another is that one cannot immediately go from correlation to causation.

Observational data

- Data that are observed as they are produced in the real world (e.g., surveys, administrative records)
  1. **Cross-sectional data**:
     - ⋆ Different units are observed: workers, households, cities, firms, etc....
     - ⋆ No time dimension, or it is neglected–¿ Order of data does not matter
     - ⋆ Analysis of the relationship between two or more variables exhibited by many individuals at one point in time.
  2. **Time series data**:
     - ⋆ Data for a single unit collected at multiple points in time
     - ⋆ Observations are time-dependent–¿ Order of data matters
     - ⋆ The concept of population correspond to that of Data Generating Process (DGP)
     - ⋆ Analysis of the co-movement of two series over time
  3. **Panel (or longitudinal) data**:
     - ⋆ Different units are observed at multiple points in time
     - ⋆ Combination of cross-sectional and time series data
     - ⋆ They have several advantages with respect to both cross-sectional and time series data

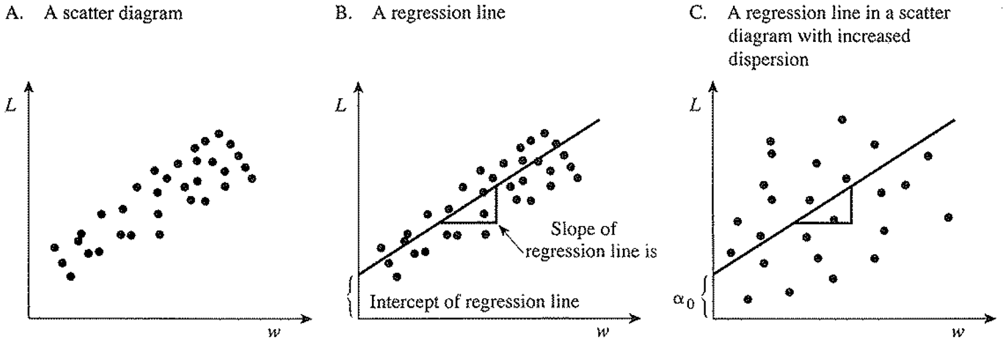
Regression analysis

- Suppose a relationship of the form

$$Y_i = \beta X_i + u_i$$

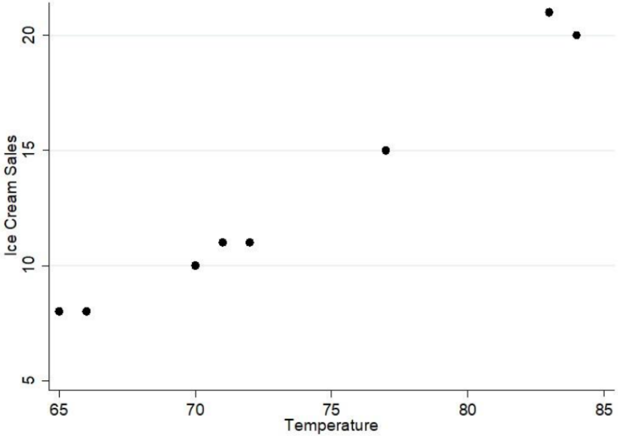
  for many individuals $i = \{1, ..., N\}$
- X_i is the independent variable (e.g., hospitalization)
- Y_i is the dependent variable (e.g., health outcome)
- β is the coefficient that measures the causal effect of X_i on Y_i
- ε is a random error term (captures variations in Y_i not related to X_i)
- The simplest way to estimate β is linear estimation (i.e., assuming the relation between Y_i and X_i to be linear) through Ordinary Least Square regression (OLS).

$$\hat{y}_i = \hat{\beta} x_i + \hat{u}_i$$


A. A scatter diagram
L
w
B. A regression line
L
Slope of
regression line is
Intercept of regression line
w
C. A regression line in a scatter
diagram with increased
dispersion
L
α_0
w

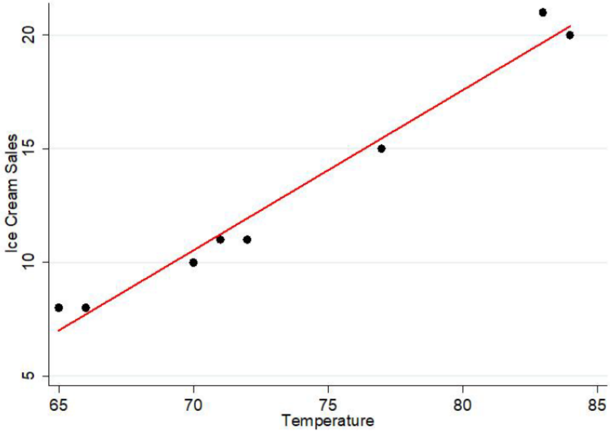
Intuition behind linear regression

RQ: What is the (causal) effect of temperature on ice cream sales?



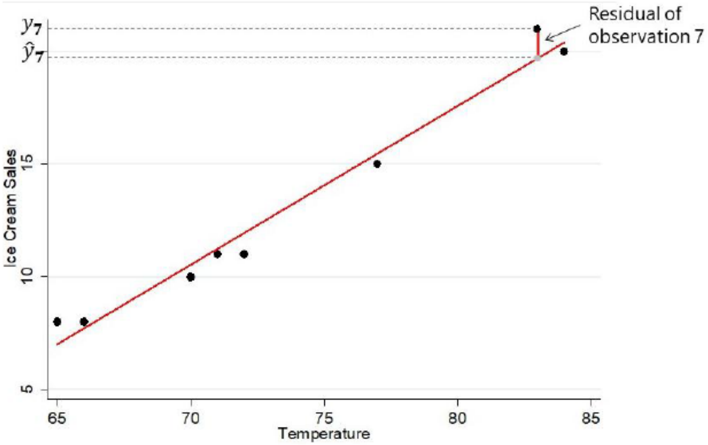
Intuition behind linear regression

RQ: What is the (causal) effect of temperature on ice cream sales?



Intuition behind linear regression

RQ: What is the (causal) effect of temperature on ice cream sales?



Interpreting OLS regression results

$$Y_i = \beta X_i + u_i$$

- The estimated coefficient $\hat{\beta}$ is reported with standard errors in parentheses

Example

$\hat{\beta} = 0.5 \ \ (0.1)$ should be understood as β is in confidence interval $(0.5 - 1.96 \cdot 0.1, \ 0.5 + 1.96 \cdot 0.1) \approx (0.3, \ 0.7)$ with probability 95%.

- We have standard errors because we do not know the exact value of β since it is a random variable, whose $\hat{\beta}$ is only a specific realization related to a specific sample
- When the confidence interval does not overlap with 0, we can conclude that it is significantly different from zero (i.e., we reject with 95% probability the null hypothesis that the true β coefficient is equal to 0).

Interpreting OLS regression results

TABLE 2
FRAGMENTATION AND STABILITY

	(1) Mayor uns.	(2) Mayor uns.	(3) Mayor uns.	(4) Mayor uns.
N. Parties	0.053**	0.049**	0.052**	0.052**
	(0.026)	(0.024)	(0.024)	(0.025)
Mean of dep.var.	0.036	0.036	0.036	0.036
Obs.	11293	11109	11293	11109
Fixed Effects	N	N	Y	Y
Controls	N	Y	N	Y

Source: Carozzi et al. (2022), p. 39

- The entry in municipality council of one additional party increases the probability of replacing the amyor via a vote of no confidence by 5 p.p.
- The coefficient is statistically significant at the 5% level
- The results are stable across four different specifications
- To understand how large the effect is, compare with mean of dep. var. (3.6 %).

Bias of OLS estimator

The OLS estimator $\hat{\beta}$ is **biased** (and inconsistent), which means that it does not identifies the true population parameter β, in four main cases:

1. **Functional mis-specification**: The actual relatioship between Y and X is not linear
2. **Omitted variables**: The actual relationship is spouriuos and there is a third variable affecting both Y and X
3. **Measurement error in the *X***: the variable is mismeasured and the error in not random
4. **Reverse causality**: Y causes X and at the same time X causes Y

- In those cases, we generally refer to as **endogeneity**. Formally, $\mathbb{E}(u_i|x_i) = 0$

Correlation $\neq$ causation every time one of these issues are present

Endogeneity

- Endogeneity arises every time the error term is correlated with the independent variable of interest.
- In the example above, hospitalized individuals also had worse health than not-hospitalized ones before entering the hospital $\rightarrow$ past health is likely an omitted variable
- There might be other sources of endogeneity at work: reverse causality and measurement error.
  - It's not only the case that hospitalization improves health, but also that poor existing health increases the probability of being hospitalized
  - Health outcomes might be more precisely measured for hospitalized individuals than for not hospitalized individuals

Formalizing the identification problem

- We make use of the **potential outcome framework**.
- Hospitalization is a binary random variable $D_i = \{0, 1\}$
- Health status Y_i is also a random variable
- $Y_{1,i} - Y_{0,i}$ is the causal effect of hospitaliation for a generic individual i.
- The average effect of hospitalization on health (Average Treatment Effect) is

$$ATE = E[Y_{1,i}] - E[Y_{0,i}]$$

- The average effect of hospitalization on health on those who got hospitalized (Average Treatment Effect on the Treated) is

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

- When treatment effects are homogeneous across individuals, $ATE = ATT$.

Formalizing the identification problem

- Unfortunately, both $Y_{1,i}$ and $Y_{0,i}$ are not contemporaneusly observable for the same individual
  - **Fundamental problem of causal inference**
- It is not possible to observe **what would have happened** to an individual who was treated in the hospital **had she not been treated**
- The observable outcome Y_i of each individual i can be written as linear combination of potential outcomes

$$Y_i = Y_{0,i} + (Y_{1,i} - Y_{0,i})D_i$$

- At most, we can estimate the average effect of hospitalization by comparing the average health of those who were hospitalized and those who were not, which instead are measurable.

$$\overbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}^{\text{Observed differences in average health}}$$

Formalizing the identification problem

- Data from the US National Health Interview Survey (NHIS) 2005
- Health status measured from 1 (poor health) to 5 (excellent health)
- Hospitalization in the last 12 months (1/0)

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No hospital	90,049	3.93	0.003

- Hospitalized individuals have on averge worse health
  - Hospitals are bad?
  - Hospitalized had worse health also before entereing the hospital?

Formalizing the identification problem

- Let us decompose the average difference in observed health in its components:

$$E[Y_i|D_i = 1] = E[Y_{0,i} + (Y_{1,i} - Y0, i)D_i|D_i = 1] = E[Y_{1,i}|D_i = 1]$$

$$E[Y_i|D_i = 0] = E[Y_{0,i} + (Y_{1,i} - Y0, i)D_i|D_i = 0] = E[Y_{0,i}|D_i = 0]$$

- In turn,

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

- Which can be re-written as

$$\overbrace{E[Y_{1,i}|D_i = 1] - E[Y_{0,i}|D_i = 1]}^{ATT} + \overbrace{E[Y_{0,i}|D_i = 1] - E[Y_{0,i}|D_i = 0]}^{Bias}$$

Formalizing the identification problem

- $E[Y_{0,i}|D_i = 1] - E[Y_{0,i}|D_i = 0]$ is, in general, not equal to 0
- There are no reasons to believe that, absent hospitalization, health of hospitalized and not hospitalized individuals would have been the same
- To summarize all possible sources of bias, we often refer to **selection bias**.
  - Individuals self-select themselves into the treatment (treatment is an individual choice) for reasons that unobservable from a researcher's point of view.

- In what circumstances we can be confident that $E[Y_{0,i}|D_i = 1] - E[Y_{0,i}|D_i = 0] = 0$?

Randomized Controlled Trials (RCTs) as a solution

Definition

Randomized controlled trial: The ideal type of experiment designed to test causality. A group of individuals is randomly divided into a treatment group, which receives the treatment of interest, and a control group, which does not.

Definition

Treatment group: The set of individuals who are subject to an intervention being studied.

Definition

Control group: The set of individuals comparable to the treatment group who are not subject to the intervention being studied.

- Randomized trials have been used in medicine for many decades and have become very popular in economics in the last 15 years (2019 Nobel Prize)

Randomized Controlled Trials (RCTs) as a solution

- Random assignment of D_i solves the selection problem making the treatment independent of Y_i
- In fact, if $D_i \perp Y_i$, $E[Y_{0,i}|D_i = 1] = E[Y_{0,i}|D_i = 0]$, thus making the selection bias term to vanish
- Moreover, if $D_i \perp Y_i$ we can consistently estimate the ATE even if we only observe a treatment effect on treated individuals

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

- The effect of randomly assigned hospitalization on hospitalized individuals is (on average) the same as the effect of hospitalization on a randomly choosen patient

Example of RCT: Tennessee STAR Project (Krueger, 1999)

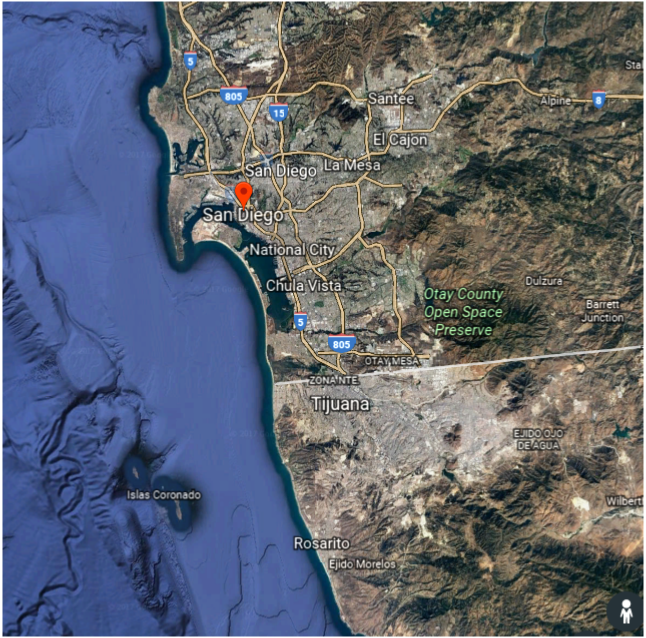
- Does class size matter for student achievements?
  - Class size is a very expensive tool for improving learning $\rightarrow$ smaller classes, more teachers
  - This makes the study of such relationship worth for policy makers
  - Non-experimental studies suggest that there is little or no link $\rightarrow$ Schools may save money by reducing teachers!
- **Tennesse Student/Teacher Achievement Ratio (STAR)** run in 1980s:
  - 11,600 students and their teachers were randomly assigned to classes of different size
  - After the assignment, student had to remain in the class for 4 years
  - Randomization occurred within schools
- Sizeable and persistent negative effect of class size on students achievement

Beyond RCTs

- Even if RCTs are the *gold standard*, they still have some potential problems
  1. **External validity:** The results are only valid for the sample of individuals who volunteer to be either treatments or controls, and this sample may be different from the population at large
  2. **Attrition**: Individuals may leave the experiment before it is complete. Reduction in the size of samples over time, which, if not random, can lead to bias estimates.
  3. **Ethics**: Not all experiments are "ethically viable"
  4. **Costs**: Experiments are expensive to implement, and difficult to replicate
  5. **Time**: Experiments require a lot of time to be run, while policy making may need quicker responses
- When a randomized experiment is not feasible, bias is a pervasive problem that is not easily remedied. There are, however, methods available that can allow us to approach the gold standard of randomized trials.

With a little help from nature: Quasi-experiments

- Also known as natural experiments
- Nature may sometimes mimimic a randomized experiment by changing the variable of interest while other factors are kept balanced.
- Changes in the economic environment that create nearly identical treatment and control groups for studying the effect of that change, allowing public finance economists to take advantage of quasi-randomization created by external forces
- Convincing natural experiment are difficult to find!


5
805
15
Santee
Alpine
8
El Cajon
San Diego
La Mesa
San Diego
National City
Chula Vista
Dulzura
Otay County
Open Space
Preserve
Barrett
Junction
5
805
OTAY MESA
ZONA NTE
Tijuana
EJIDO OJO
DE AGUA
Islas Coronado
Rosarito
Ejido Morelos

Example of natural experiment: the Maimonides' Rule (Angrist and Lavy, 1999)

- In Israel, class size is capped at 40 students by law
- Cohorts of 41 students end up in two classes, one of 20 and one of 21 students (i.e. Maimonides' rule).
- Students in cohort size of 40 and 41 are likely to be similar on other dimensions (e.g. ability, family background).
- We can think of the difference between 40 and 41 students enrolled "as good as randomly assigned".
- By comparing students in grades with enrollement below and above the cutoff they can estimate the effect of a sharp change in class size without the benefit of a real experiment.
- Very consistent with Krueger (1999) $\rightarrow$ Strong negative effect of class size on achievements

Natural experiments: estimation

- We consider a Treatment group ($D_i = 1$) and a Control group ($D_i = 0$) and outcome Y_i
- **Simple difference**: $\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$ is the difference in average outcomes between treatment and control after the (natural) treatment
- In randomized experiment, simple difference is sufficient because treatment is independent on Y_i.
- In quasi-experiment, treatment and control group might still vary in other dimensions besides the treatment.

Natural experiments: estimation

- Suppose we can rely on panel data. We observe both the treatment and the control groups before and after treatment occurs.
- With panel data, we observe Y_i^{Before} and Y_i^{After}.
- Can compute $\mathbb{E}[Y_i^{Before}|D_i = 1] - \mathbb{E}[Y_i^{Before}|D_i = 0]$
- If $\mathbb{E}[Y_i^{Before}|D_i = 1] - \mathbb{E}[Y_i^{Before}|D_i = 0] = 0$, can be fairly confident that $\mathbb{E}[Y_i^{After}|D_i = 1] - \mathbb{E}[Y_i^{After}|D_i = 0]$ identifies the ATE

This approach is known as the **difference estimator**.

Difference-in-Differences (DiD)

- If simple difference $\mathbb{E}[Y_i^{Before}|D_i = 1] - \mathbb{E}[Y_i^{Before}|D_i = 0] \neq 0$, we can identify the ATT using the **Difference-in-Differences (DiD)** technique

$$DiD = \overbrace{\{\mathbb{E}[Y_i^{After}|D_i = 1] - \mathbb{E}[Y_i^{After}|D_i = 0]\}}^{\text{Difference after treatment}} - \overbrace{\{\mathbb{E}[Y_i^{Before}|D_i = 1] - \mathbb{E}[Y_i^{Before}|D_i = 0]\}}^{\text{Difference before treatment}}$$

- Measures whether the difference between treatment and control changes after the policy change
- DiD identifies the ATT if, in the absence the policy change, the difference between T and C would have stayed the same over time (this is called the **parallel trends assumption**)

DiD in regression framework

- We can estimate DiD using regression analysis

$$Y_{i,t} = \beta_0 + \beta_1 Treated_i + \beta_2 Post_t + \beta_3 Treat_i \times Post_t + u_{i,t}$$

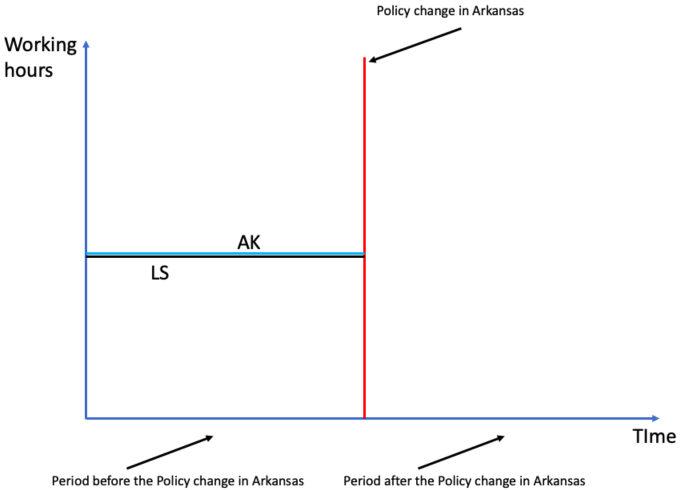
- where $Treated_i = 1$ if i in treated group (0 otherwise)
- where $Post_t = 1$ if t is after treatment (0 otherwise)

Difference-in-Differences: Example

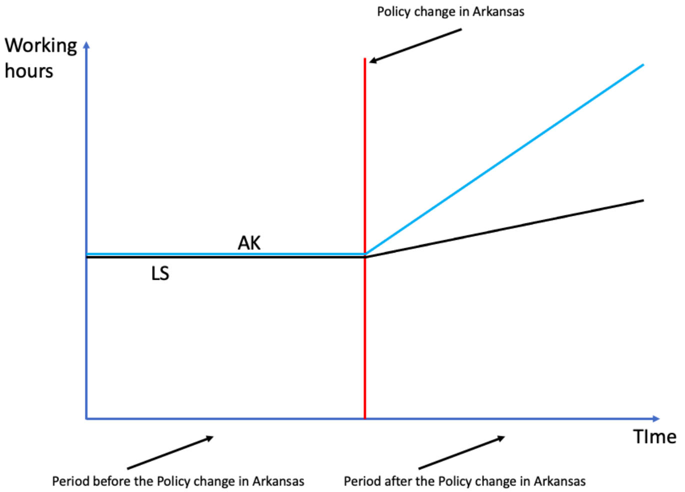
- **RQ**: What is the effect of welfare benefit for single mothers on their supply of labour?
- **Natural Experiment**: One state (Arkansas) decreased generosity of welfare benefits in 1997 while another comparable state (Louisiana) did not
- **Setting** Single mothers in Arkansas are the Treatment (T) group, Single mothers in Louisiana are the control (C) group.

Arkansas			
	1996	1998	Difference
Benefit guarantee ($)	5,000	4,000	−1,000
Hours worked	1,000	1,200	200
Louisiana			
	1996	1998	Difference
Benefit guarantee ($)	5,000	5,000	0
Hours worked	1,050	1,100	50

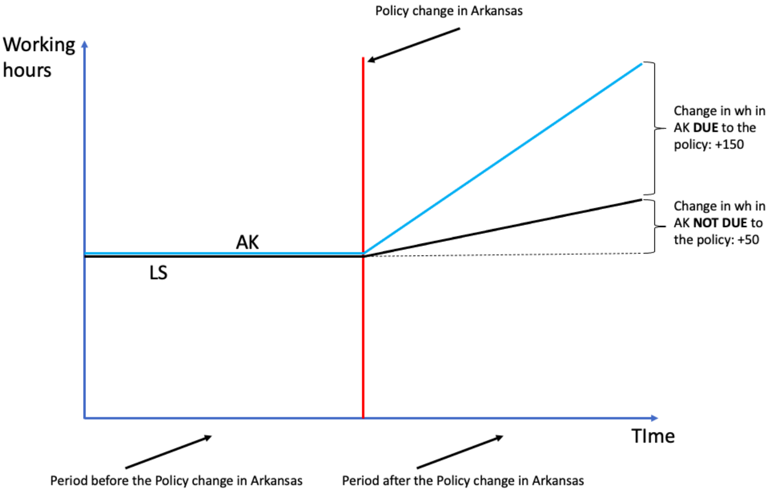
Difference-in-Differences: Example



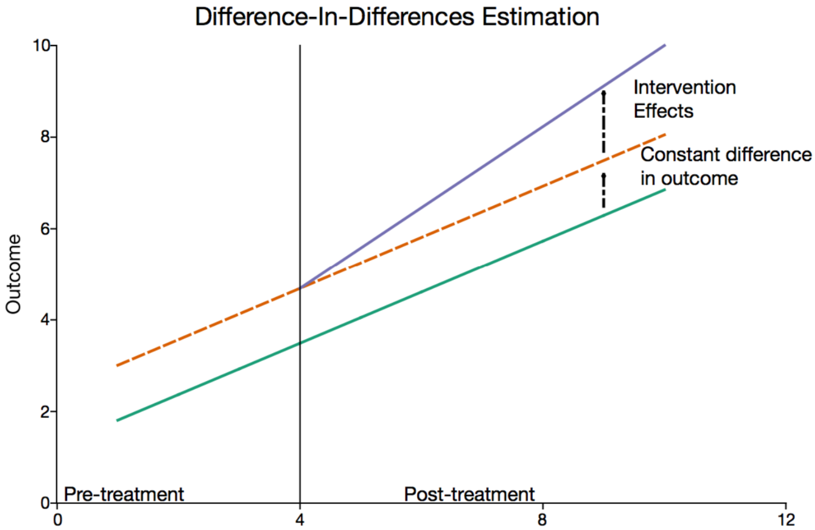
Difference-in-Differences: Example



Difference-in-Differences: Example



Difference-in-Differences: Example



Limitations of natural experiments

- With quasi-experimental studies, we can never be completely certain that we have purged all bias from the treatment–control comparison
- Quasi-experimental studies present various robustness checks to try to make the argument that they have obtained a causal estimate
- Examples: find alternative control groups, do a placebo comparing treatment and control DiD when no policy change took place, etc.
- Best way to check validity of DiD estimator is to plot times series and assess whether a clear break between the two groups happens at the time of the reform

Two graphical examples

Example

Effects of lottery winnings on labor supply from Imbens, Rubin, Sacerdote (2001)

Ideal quasi-experiment to measure income effects as lottery generates random assignment conditional on playing $\Rightarrow$ Very compelling graph, DD is convincing

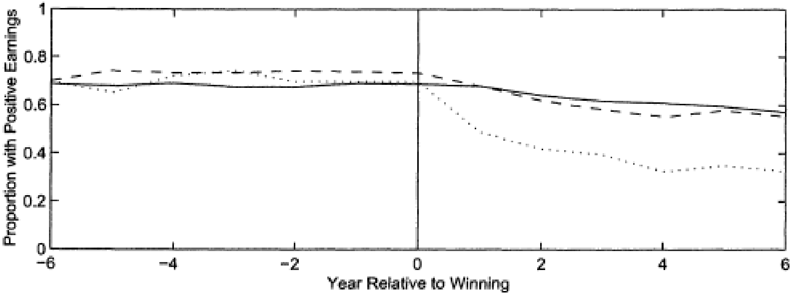


FIGURE 2. PROPORTION WITH POSITIVE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

Note: Solid line = nonwinners; dashed line = winners; dotted line = big winners.

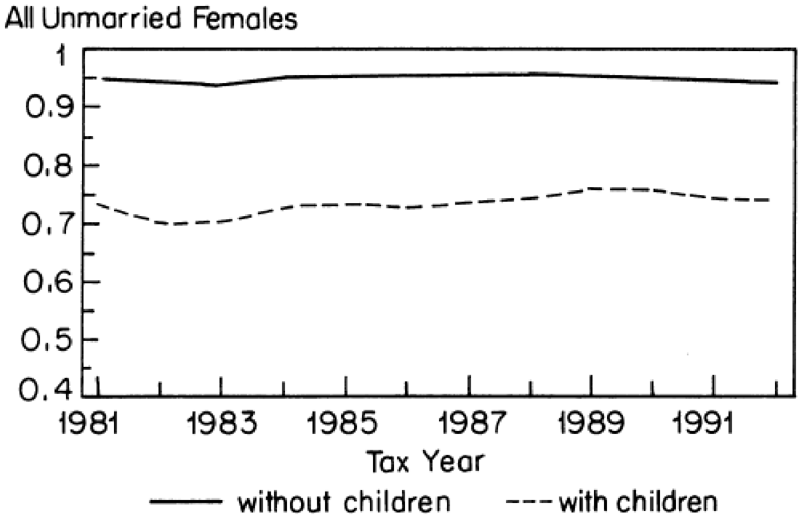
Source: Imbens et al (2001), p. 784

Two graphical examples

Example

Effects of the 1987 EITC expansion (tax credit for low income workers with kids) on labor supply from Eissa and Liebman (1996)

Compares single mothers (Treatment) to single females with no kids (Control) $\Rightarrow$ No compelling break in graph around 1987, DD is not convincing



Source: Eissa and Liebman (1996), p. 624

Structural modelling

- **Structural estimates**: Builds a theoretical model of individual behavior and then estimates the parameters of the model.
  - Estimates of the features that drive individual decisions, such as income and substitution effects or parameters of the utility function.
- **Reduced form estimates**: Measures of the total impact of an independent variable on a dependent variable, without decomposing the source of that behavior response in terms of underlying parameters of the utility functions.
  - Reduced form estimates are more transparent and convincing but structural estimates are more directly useful to make predictions for alternative policies.

Conclusion

- The central issue for any policy question is establishing a causal relationship between the policy in question and the outcome of interest.
- We discussed several approaches to distinguish causality from correlation. The gold standard for doing so is the randomized controlled trial, which removes bias through randomly assigning treatment and control groups.
- Unfortunately, however, such trials are not available for every question we wish to address in empirical public finance. As a result, we turn to alternative methods such as time series analysis, cross-sectional regression analysis, and quasi-experimental analysis.
- Each of these alternatives has weaknesses, but careful consideration of the problem at hand can often lead to a sensible solution to the bias problem that plagues empirical analysis.

References

- Rosen, Harvey, and Ted Gayer. Public finance, 2014, McGraw Hill Education, Chapter 2
- Gruber, Jonathan, Public Finance and Public Policy, 2016 Worth Publishers, Chapter 3
- Carozzi, Felipe, Davide Cipullo, and Luca Repetto (2022). Political Fragmentation and Government Stability: Evidence from Local Governments in Spain. American Economic Journal: Applied Economics, 14(2), 23-50.
- Alan B. Krueger (1999). Experimental Estimates of Education Production Functions, The Quarterly Journal of Economics, 114(2), 497–532.
- Joshua D. Angrist, Victor Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement, The Quarterly Journal of Economics, 114(2) 533–575.
- Imbens, Guido W., Donald B. Rubin, and Bruce I. Sacerdote. "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players." American Economic Review 91(4), 778-794.
- Eissa, Nada, and Jeffrey B. Liebman. "Labor supply response to the earned income tax credit." The Quarterly Journal of Economics 111(2), 605-637.