Diversity and Global Policy: Empirical Tools

Germain Gauthier

Bocconi University

Definitions

Correlation: Two economic variables are correlated if they move together

e.g., height and weight across individuals

Causality: Two economic variables are causally related if the movement of one causes the movement of the other

e.g., good nutrition as an infant increases adult height

Distinction Between Correlation and Causality

There are many examples where causation and correlation can get confused (Cunningham, 2021).

In statistics, this is called the *identification problem*:

Given that two variables are correlated, how do you identify whether one is causing another?

The Identification Problem

Interpreting 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.

Hypothetical Example: Gender Quotas

Among companies that implemented gender quotas, the share of women employees was 10pp lower than among those who hadn't.

$$\text{Share of Women}_i = 45 - 10 \cdot \text{Gender Quotas}_i + \varepsilon_i$$

- Do quotas hurt women? (A causes B)
- Do low shares of women cause companies to implement quotas? (B causes A)
- Did some third factor cause both low shares of women and quotas? (C causes A and B)

Formalizing the Identification Problem

Given observations $i = \{1, ..., N\}$, a treatment indicator D_i, and covariates X_i, researchers typically estimate:

$$Y_i = \alpha + \beta \cdot D_i + \gamma' X_i + \varepsilon_i$$

Goal: interpret β as the causal effect of D_i on Y_i.

Key assumption:

$$\mathbb{E}[\varepsilon_i \mid D_i, X_i] = 0.$$

If violated, β combines the causal effect and the effects of confounders.

Omitted Variable Bias (OVB)

Suppose the true model is

$$Y_i = \alpha + \beta D_i + \delta Z_i + \gamma' X_i + u_i,$$

but Z_i is unobserved.

If we estimate without Z_i:

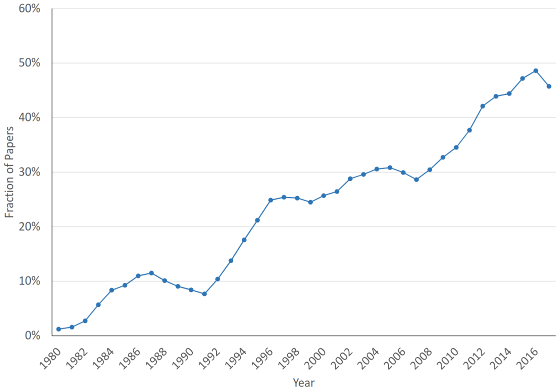
$$Y_i = \alpha + \tilde{\beta} D_i + \gamma' X_i + \varepsilon_i,$$

then

$$\mathbb{E}[\tilde{\beta}] = \beta + \delta \cdot \frac{\mathrm{Cov}(D_i, Z_i \mid X_i)}{\mathrm{Var}(D_i \mid X_i)}.$$

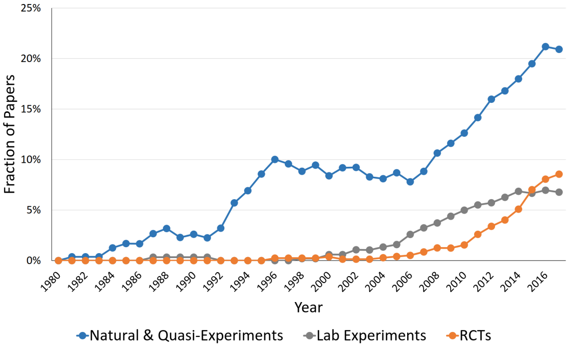
OVB: bias arises when Z_i is correlated with both D_i and Y_i.

The Rise of Identification



Notes: The graph shows the fraction of papers that mention the word “identification” in the context of empirical identification. The sample comprises all NBER working papers 1975–2018 tagged “public economics” (4676 papers). See the original slides here.

The Rise of Experiments



Notes: The graph shows the fraction of papers that refer to each type of experiment. See the original slides here.

Randomized Control Trials

In an RCT, a group of individuals is RANDOMLY divided into two groups.

The treatment group receives the treatment of interest, whereas the control group does not.

Randomization effectively ensures that, in expectation, the treatment is uncorrelated to confounders.

Randomized trials have been used in medicine for many decades and have become very popular in economics in the last 30 years.

Back to our Hypothetical Example: Gender Quotas

Randomize the implementation of gender quotas in a sample of companies.

Then estimate:

$$\text{Share of Women}_i = \alpha + \beta \cdot \text{Gender Quotas}_i + \varepsilon_i$$

β measures the average treatment effect of gender quotas on the share of women employees.

YOU ARE DONE!!!

Methods for Observational Data

In many settings, we cannot design RCTs to answer economic problems, and we thus rely on observational data.

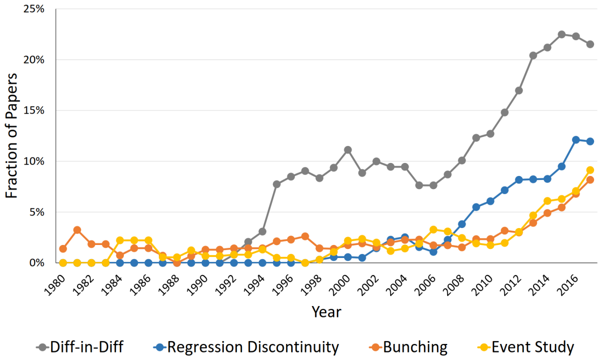
As soon as we do not have proper randomization, the identification problem is very tricky to deal with.

Thankfully, researchers have developed clever identification strategies that aim to "recreate" the random allocation of treatment.

We call those "quasi-experimental methods":

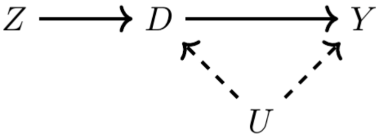
- Instrumental variables (IV)
- Regression Discontinuity Designs (RDD)
- Difference in differences (DID)

The Rise of Quasi-experiments



Notes: The graph shows the fraction of papers that refer to each type of quasi-experiment. See the original slides here.

Instrumental Variables



We care about the relationship between a treatment D and an outcome Y.

There are many confounder variables U that affect D and Y.

An instrumental variable, Z predicts D (relevance) but is uncorrelated with the confounders U (exclusion restriction).

$\Rightarrow$ Allows for the identification of the *causal* effect of D on Y

Two-Stage Least Squares (2SLS)

Step 1: First stage Regress treatment D on the instrument Z (and controls X if any).

$$D = \pi_0 + \pi_1 Z + \pi_2' X + v \tag{1}$$

$\Rightarrow$ Obtain predicted values $\hat{D}$.

Step 2: Second stage Regress outcome Y on $\hat{D}$ (and controls X).

$$Y = \beta_0 + \beta_1 \hat{D} + \beta_2' X + \varepsilon \tag{2}$$

Interpretation: β_1 is the causal effect of D on Y identified using Z.

Example: Fertility and Mothers' Labor Supply

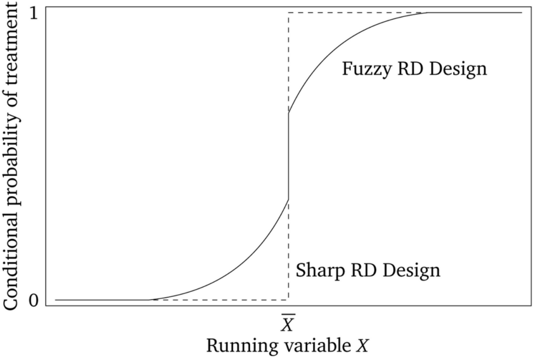
Context: Causal effect of family size on female labor supply (Angrist and Evans, 1998)

- Instrument: Sex composition of the first two children (Z).
- Same-sex first two $\Rightarrow$ higher probability of a third child (D).
- Outcome: Mothers' labor supply / earnings (Y).
- Relevance: Z strongly predicts family size.
- Exclusion: Child sex mix affects labor supply only through family size.

$\rightarrow$ Having more children (due to same-sex first two) reduces mothers' labor supply and earnings, with stronger effects for less-educated women.

Regression Discontinuity Design

RDDs exploit discontinuous jumps in the probability of treatment assignment along some running variable.



Example: Close-election RDD (Casarico et al., 2022)

Women and local public finance☆

Check for updates

Alessandra Casarico [a,b,c], Salvatore Lattanzio [d,e], Paola Profeta [a,b,c,*]

[a] *Bocconi University, Italy*
[b] *CESifo, Germany*
[c] *Dondena, Italy*
[d] *Bank of Italy, Italy*
[e] *University of Cambridge, United Kingdom of Great Britain and Northern Ireland*

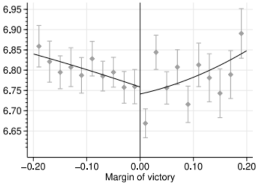
ARTICLE INFO

JEL classification:
E62, J16, H71, H72

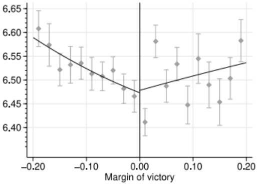
Keywords:
Gender
Municipal government
Local public finance
Regression discontinuity

ABSTRACT

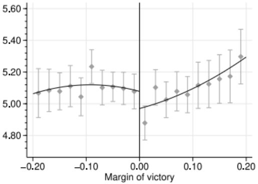
Does the gender of the mayor affect the size and composition of public expenditures and revenues? Using a sharp regression discontinuity design in close mixed gender races for the election of mayors in Italian municipalities in the period 2000–2015, we find no significant differences in policies implemented by male and female mayors. We explore whether the result masks heterogeneity by gender composition of the local government and by electoral rules according to which a mayor is elected. We find some evidence that female mayors devote a larger share of spending to the environment when there are more women in the municipal council, whereas they reduce the amount of resources going to social spending under the run-off relative to the single round system.



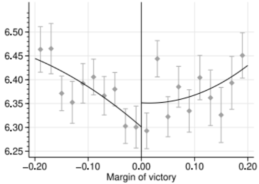
(A) Total expenditures



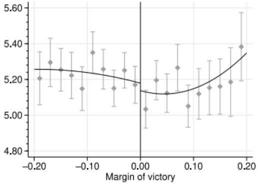
(B) Current expenditures



(C) Capital expenditures



(D) Revenues from taxes and fees



(E) Other revenues

We are interested in estimating:

$$Y_i = \alpha + \beta\, g(X_i) + \delta D_i + \varepsilon_i, \tag{3}$$

where:

- Y_i = the outcome variable
- X_i = the running variable
- g = a flexible function of X_i
- c = the cutoff
- D_i = treatment assignment

δ is the primary quantity of interest (i.e., the "estimand").

It represents the **Local Average Treatment Effect (LATE)** as $X_i \to c$.

Two types of RDDs:

- **Sharp RDD:** The treatment assignment is deterministic at the cutoff.
- **Fuzzy RDD:** The probability of treatment assignment jumps at the cutoff, but it is not a deterministic process.

In the case of the fuzzy RDD, we need to work a bit more.

Let's go ahead and define:

$$Z_i = \begin{cases} 0 & \text{if } X_i \leq c \\ 1 & \text{if } X_i \geq c. \end{cases}$$

We first instrument treatment assignment D_i with Z_i and then estimate Equation 1 in the second stage.

Limitations of RDDs

Core assumption: The only thing that causes the outcome to change abruptly at c is the treatment.

Limitations/threats to identification:

- We cannot guarantee that the control and treatment group are comparable.

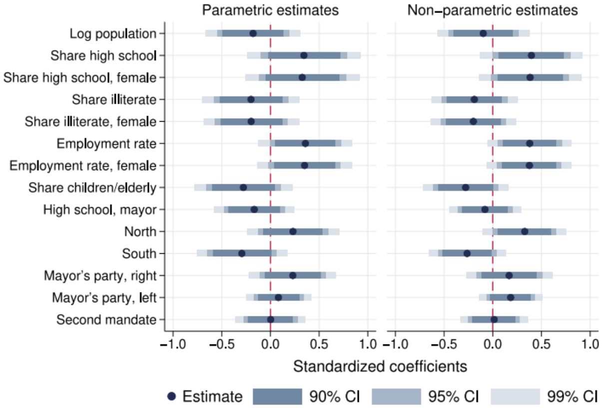
  $\Rightarrow$ We need to do balance tests on observables!

- There is sometimes endogenous sorting at the cutoff.

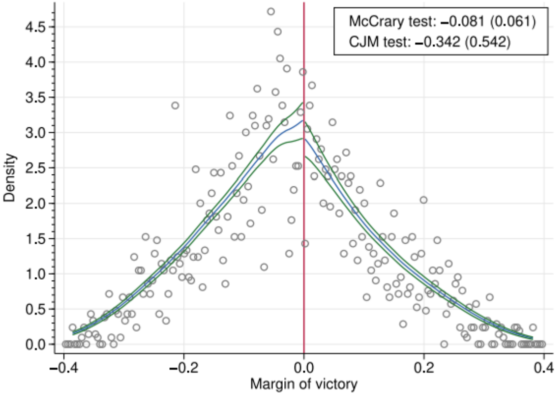
  $\Rightarrow$ McCrary density test assesses bunching at c.

- The LATE is only identified as $X_i \to c$, so our estimates are really based on an extrapolation exercise.

  $\Rightarrow$ The choice of $g(X_i)$ can affect results.



Notes: Balance checks seem OK, comforting the plausible randomness of mixed gender close election results.



Notes: There is no empirical evidence of bunching at the threshold, once again reinforcing our trust in the research design.

Difference-in-differences

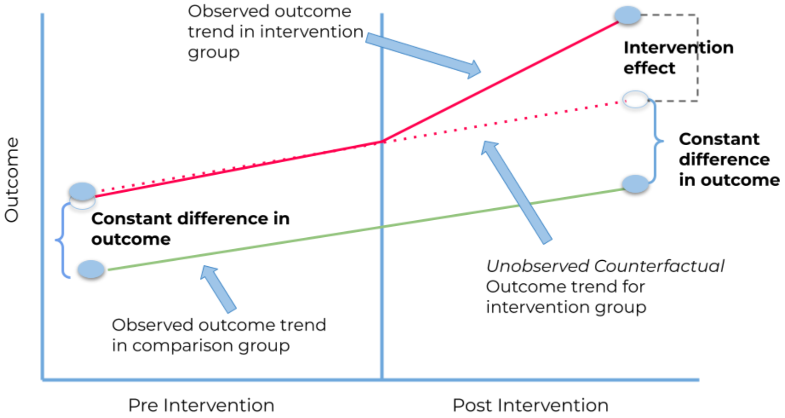
The Difference-in-Difference (DID) estimator:

$$DID = [Y^{T,After} - Y^{C,After}] - [Y^{T,Before} - Y^{C,Before}] \tag{4}$$

This measures whether the difference between treatment and control changes after the policy change.

Parallel trends assumption: DID identifies the causal effect of the treatment if, absent the policy change, the difference between T and C would have stayed the same.

A Graphical Take



Generalizing to Many Groups and Many Periods

In practice, we often observe more than two groups over many periods.

That's fine. A generalized DID estimator can be recovered by running the following regression:

$$Y_{it} = \alpha + \beta D_{it} + \delta_i + \delta_t + \varepsilon_{it},$$

where D_{it} takes value 1 if unit i is treated at period t, and 0 otherwise.

Such specifications are referred to as **"two-way fixed effects" models**.

Can we test parallel trends? **No!**

We will never know if treated and untreated units would have followed the same trends absent the policy intervention.

But we can at least test whether they did display parallel trends in average outcomes *before* the policy intervention:

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} \, D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{lead} D_{i,t}^{k} + \sum_{k=0}^{L} \gamma_k^{lag} D_{i,t}^{k} + \gamma_k^{L+} D_{i,t}^{>L} + \varepsilon_{i,t}$$

,

where $D_{i,t}^{k} = 1\{t - G_i = k\}$ is a variable that takes value 1 if a unit i is k periods away from initial treatment at time t and 0 otherwise.

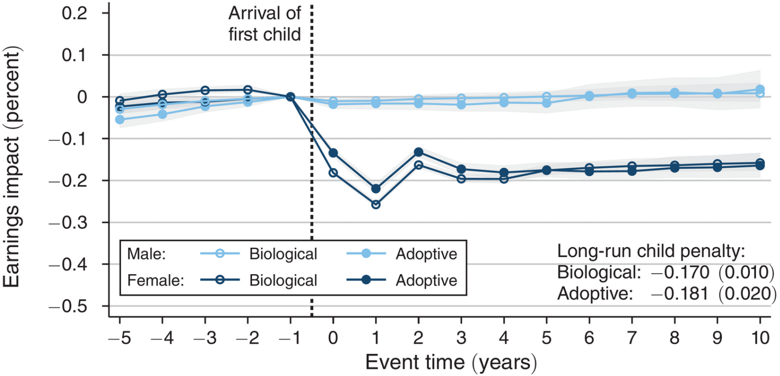
$D_{i,t}^{<-K} = 1\{t - G_i < -K\}$ and $D_{i,t}^{>L} = 1\{t - G_i > L\}$ are defined analogously.

Example: Child Penalty on Earnings

Abstract

This paper investigates whether the impact of children on the labor market outcomes of women relative to men—child penalties—can be explained by the biological links between mother and child. We estimate child penalties in biological and adoptive families using event studies around the arrival of children and almost 40 years of adoption data from Denmark. Short-run child penalties are slightly larger for biological mothers than for adoptive mothers, but their long-run child penalties are virtually identical and precisely estimated. This suggests that biology is not a key driver of child-related gender gaps.

Example: Child Penalty on Earnings



Notes: Estimates from Kleven et al. (2021)

Experiments have their limitations, too.

Even well-designed identification strategies have limitations that you should pay attention to:

- Statistical power
- Selective attrition
- External validity
- Ethics

Statistical Power

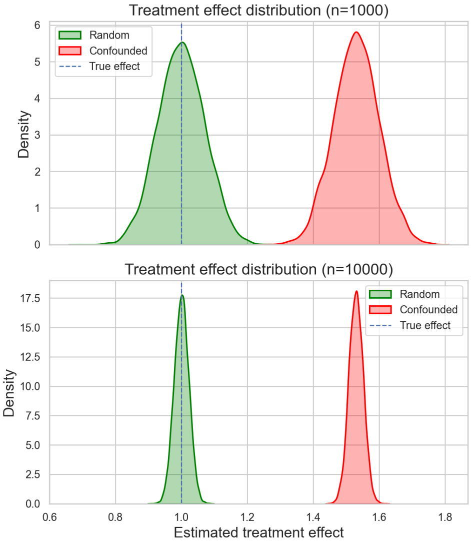
Statistical power is the probability of detecting an effect if it truly exists.

Low power increases the risk of false negatives and produces imprecise estimates.

Small sample sizes or weak interventions often lead to low power.

Researchers should plan sample sizes carefully and report minimum detectable effects.

Simulations on Statistical Power



Selective attrition

Participants may drop out of a study in a non-random way.

If attrition correlates with treatment, estimates become biased.

Example: Job training program where less-motivated treated individuals quit more often.

Solutions: track dropouts carefully and use bounds (Lee bounds).

External Validity

Results from one setting may not generalize.

Treatment effects can vary across populations, locations, and time.

Often, experiments identify Local Average Treatment Effects (LATE) rather than Average Treatment Effects (ATE).

Program scale-up may change effectiveness due to general equilibrium effects (List, 2022).

Always ask: "Would this work elsewhere or at scale?"

Example: Partial vs. General Equilibrium

Partial equilibrium: evaluates direct effect on treated individuals while holding rest of economy fixed.

General equilibrium: accounts for spillovers, price effects, and resource reallocation.

Example: Job training increases wages for participants (partial), but if many are trained, wages in the sector may fall (general).

Ethics

Randomization may raise fairness concerns.

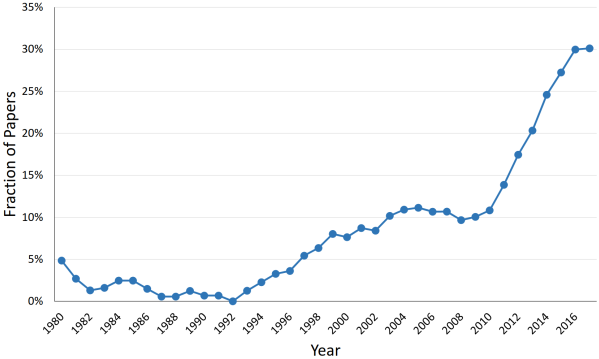
Informed consent and transparency are crucial.

Some interventions may harm participants or deny access to needed services.

Ethical review boards (IRBs) balance scientific value vs. participant rights.

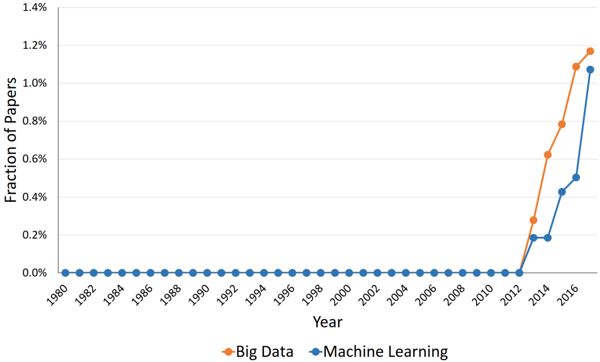
Before concluding, a glimpse of the new trends in empirical economic research...

The Rise of Administrative Data



Notes: The graph shows the fraction of papers that refer to administrative data. See the original slides here.

The Rise of Machine Learning



Notes: The graph shows the fraction of papers that refer to machine learning. See the original slides here.

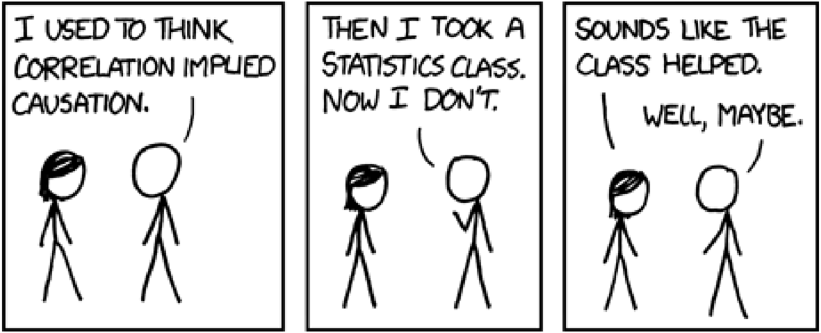
Conclusion

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 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. As a result, we turn to alternative methods, "quasi-experiments".

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.


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3):450–477.

Casarico, A., Lattanzio, S., and Profeta, P. (2022). Women and local public finance. *European Journal of Political Economy*, 72:102096.

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale university press.

Kleven, H., Landais, C., and Søgaard, J. E. (2021). Does biology drive child penalties? evidence from biological and adoptive families. *American Economic Review: Insights*, 3(2):183–198.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Crown Currency.