

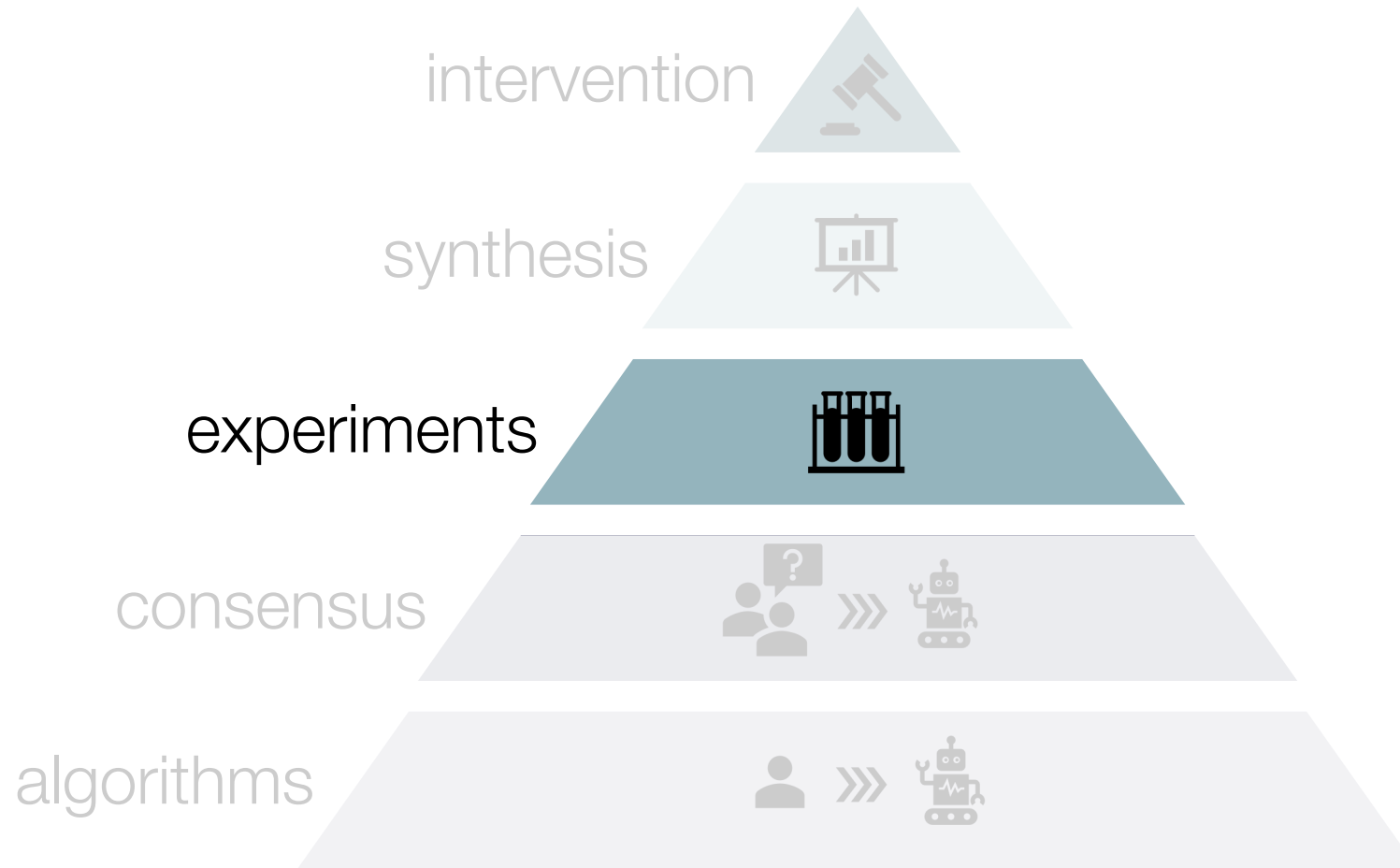
Evidence-based Decision Making

Counterfactuals: Alternatives to Experiments

Loreen Tisdall, FS 2025

Version: April 10, 2025

Climbing the pyramid of evidence



Goals for today

- Understand the nature of causal inference as the comparison of treatment to some counterfactual
- Familiarize yourself with different methods of causal inference - randomization, regression, regression discontinuity, instrumental variables - and associated limitations
- Consider the untapped potential of natural experiments

Recap: Obstacles for RCTs

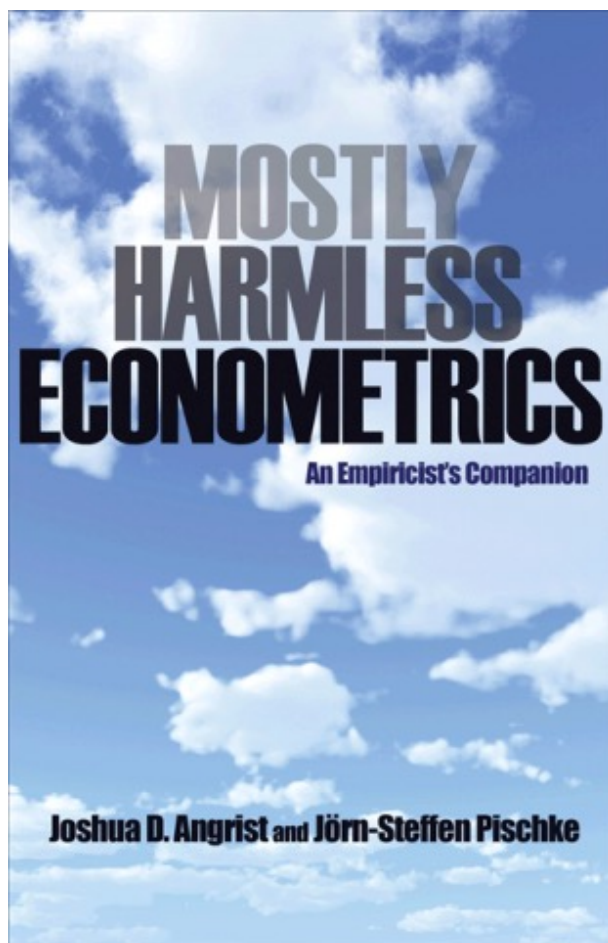
Experimental equipoise

Aspect	What It Means	Example (Mental Health Context)
Definition	Researchers honestly do not know which treatment is better.	Psychologists are unsure if CBT or a new mindfulness therapy works better for anxiety.
Why It Matters Ethically	It would be unfair to give someone a treatment that is already known to be worse.	You would not want to give a client a new therapy if you knew CBT definitely worked better.
Why It Matters Scientifically	A fair test needs two treatments that are both reasonable options.	The new therapy must have some evidence that it could work—otherwise, why test it?
When It Exists	There is no clear winner based on current research.	CBT has lots of support, but early studies on mindfulness therapy show promise—so there is debate.
When It Doesn't	One treatment is clearly better based on strong evidence.	If CBT is clearly more effective than a new, untested therapy, it is not ethical to compare them.
Purpose of the Study	To fairly figure out which approach is more helpful.	To learn if the new therapy actually helps more, or if CBT remains the better choice.

A colorful bouquet of creating counterfactuals

“The stronger the demonstrated consistency of an association under conditions that rule out alternative hypotheses and the stronger the evidence regarding a mechanism that can explain the observed association, the more likely we are to accept the causal hypothesis. Usually the evidence required to confirm a causal hypothesis is **cumulated across multiple studies, many of which are, of necessity, observational.** Although a wide variety of research designs and analytic techniques are available to assist in gathering evidence to support a causal inference, they are helpful only to the extent that their use is guided and constrained by appropriate subject-matter considerations. **No method or set of methods defines causality.”**

“Furious Five” statistical methods for causal inference



- Randomization
- Regression
- Difference in differences
- Regression discontinuity
- Instrumental variables

Your turn!



Image created with AI (Bing), February 13, 2024

Take 3 minutes and write down what you think is the key idea behind each of the “Furious Five”.

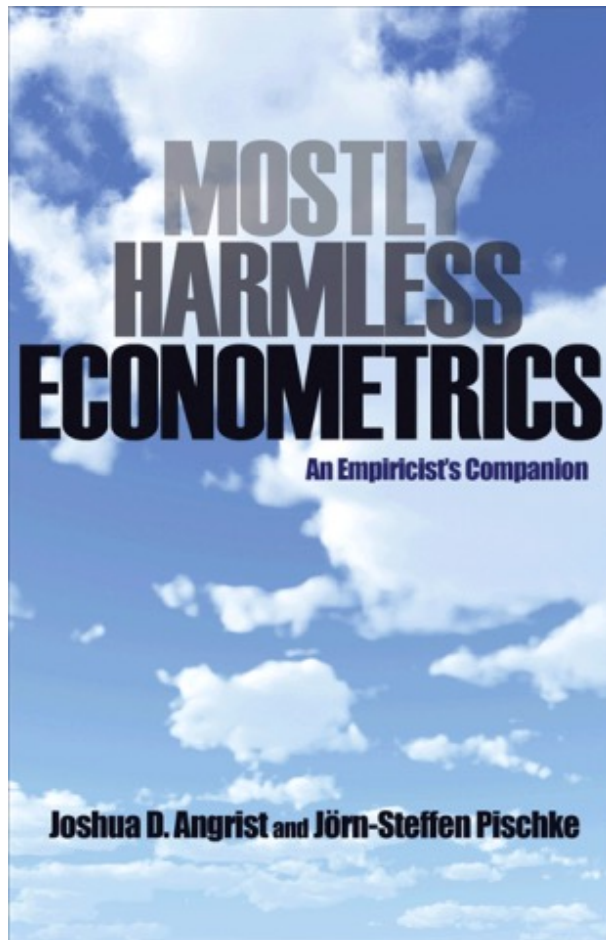
- 1) Randomization,
- 2) Regression,
- 3) Difference in differences,
- 4) Regression discontinuity,
- 5) Instrumental variables

The Fundamentally Different “Furious Five”?

Furious Five	What is it (category)?
Random assignment	Design
Regression	Estimation (statistical) method
Difference-in-Differences***	Design
Regression Discontinuity***	Design
Instrumental Variables***	Design / Both

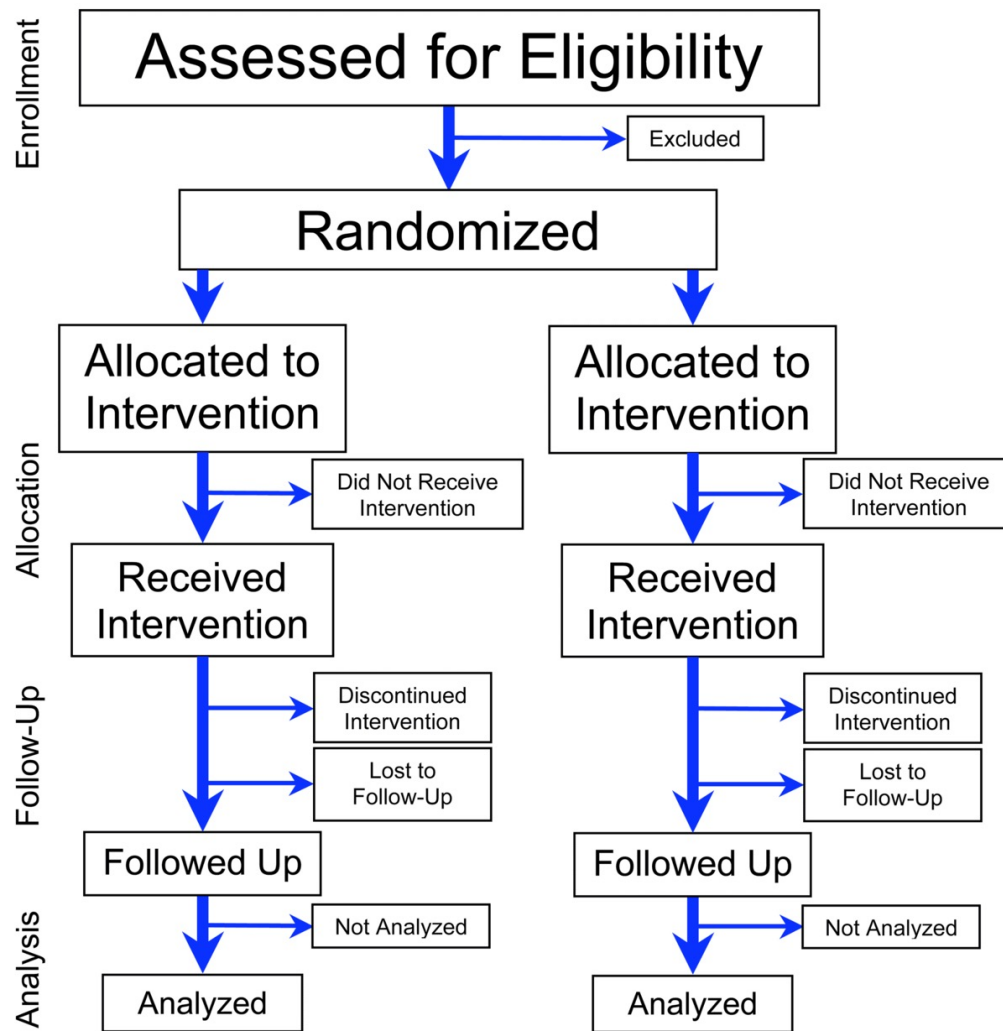
*** often made possible as a result of a **Natural Experiment**

“Furious Five” statistical methods for causal inference



- **Randomization**
- Regression
- Difference in differences
- Regression discontinuity
- (Instrumental variables)

Randomisation



→ Calculated as

Difference between group means:

$$\mu_{\text{treatment}} - \mu_{\text{control}}$$

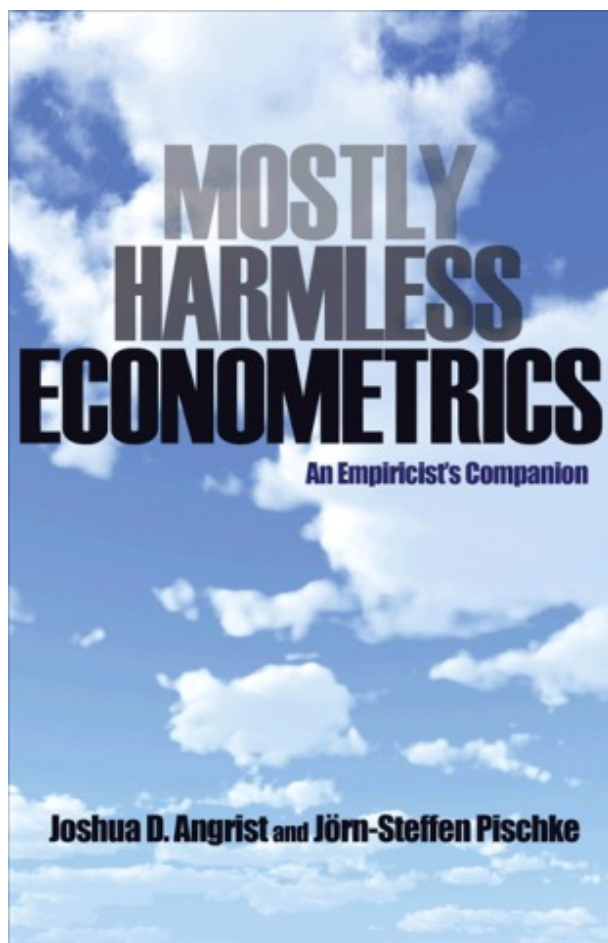
→ or as

Effect of treatment on outcome:

$$Y = b_0 + b_1\text{group} + e$$

Regression is not inherently causal, but it becomes a powerful causal inference tool when paired with a good design—like randomization (or natural experiments 😊)!

“Furious Five” statistical methods for causal inference



- Randomization
- **Regression**
- Difference in differences
- Regression discontinuity
- (Instrumental variables)

Regression as a statistical tool

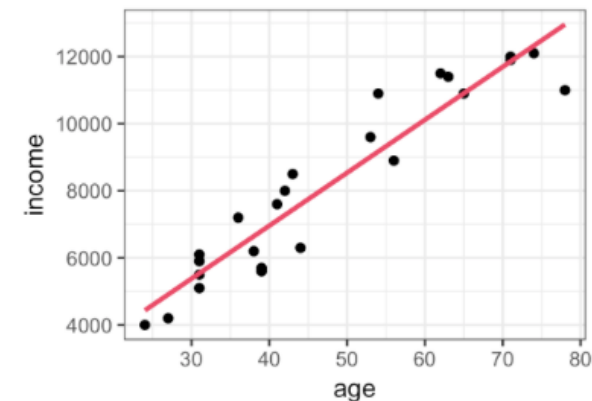
Regression analysis is a set of statistical processes for **estimating the relationships among variables**. It includes many techniques for modeling and analyzing several variables, when the focus is on the **relationship between a dependent variable (criterion) and one or more independent variables (predictors)**. More specifically, regression analysis helps one understand how the typical **value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are fixed**.

Regression

Simple Linear Regression

Definition: Simple linear regression is a linear model with one predictor x , and where the error term ϵ is Normally distributed.

$$y = \beta_0 + \beta_1 x + \epsilon$$



Parameter	Description	In words
β_0	Intercept	When $x = 0$, what is the predicted value for y ?
β_1	Coefficient for x	For every increase of 1 in x , how does y change?

Formula

$$income = 885 + 149.3 \times age + \epsilon$$

Coefficients

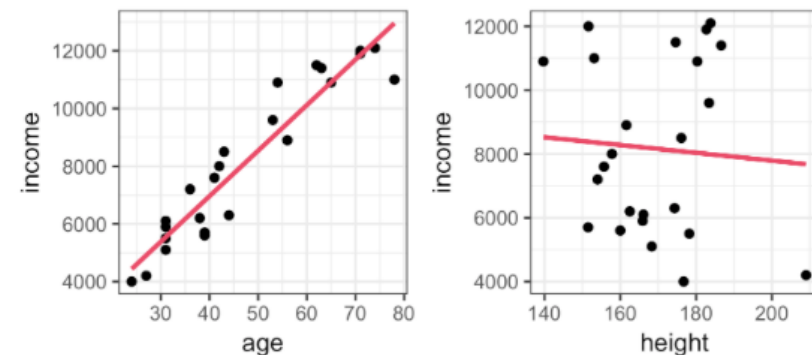
$$\beta_0 = 885, \beta_{age} = 149.3$$

Regression

Multiple Linear Regression

Definition: Multiple linear regression is a linear model with many predictors x_1, x_2, \dots, x_n , and where the error term ϵ is Normally distributed.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$



Parameter	Description	In words
β_0	Intercept	When all x values are 0, what is the predicted value for y?
β_1, β_2, \dots	Coefficient for x_1, x_2, \dots	For every increase of 1 in coefficient for x_1, x_2, \dots how does y change?

Formula

$$income = 1628 + 147 \times age - 4.1 \times height + \epsilon$$

Coefficients

$$\beta_0 = 1628, \beta_{age} = 147, \beta_{height} = -4.1$$

Full randomisation is seldom available in practice...

The “ideal” data, from the viewpoint of the analyst, would be data from an incompetent advertiser who allocated expenditures randomly across cities. If ad expenditure is truly random, then we do not have to worry about confounding variables because the predictors will automatically be orthogonal to the error term. However, statisticians are seldom lucky enough to have a totally incompetent client.

Instead ... natural randomizers



Natural experiments in economics and psychology

Natural experiments: “offer unique opportunities to combine features of randomized experiments and observational studies. A natural experiment is a **“naturally” occurring event or condition (i.e., an event or condition not created by researchers) that affects some but not all units of a population.** [...] Natural experiments differ from (non-natural) randomized experiments in that **participants are not randomly assigned to treatment and control groups by researchers**, and researchers do not control the experimental manipulations and conditions.”

Examples

Pros

Cons

-
- | | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none">• A longitudinal survey spans the occurrence of a relevant event (e.g., natural disaster) that affects parts of the population• Educational reforms lead to the raising of minimum school-leaving age from one year to the next• A lottery decides who gets drafted into military service during war times• Lottery players who won large versus small sums of money | <ul style="list-style-type: none">• An option when randomized experiments are unethical or unfeasible• Limited demand effects• More “realistic” effect sizes (cf. efficacy versus effectiveness)• No self-selection bias (who volunteers for vaccine trials?) and lower participation burden | <ul style="list-style-type: none">• Outcomes / dependent variables often not assessed immediately after the events/treatment occurred• Hypotheses usually come after the event/treatment |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Natural experiments in economics and psychology

Table 1. Review of a Random Sample of Economics and Psychology Articles

	Economics	Psychology
Number of articles reviewed	108	108
Number of articles containing . . .		
An empirical study	88	96
A randomized experiment	17	42
A natural experiment	36 ^a	0
A standard natural experiment with true randomization	1	0
A standard natural experiment with as-if randomization	19	0
An instrumental-variable design using a natural experiment with true randomization	1	0
An instrumental-variable design using a natural experiment with as-if randomization	18	0
A sharp regression-discontinuity design	3	0
A fuzzy regression-discontinuity design	2	0

“Reasons for the more frequent use of natural experiments in economics than psychology might be that randomized experiments are hardly feasible in macro-economics because researchers cannot experiment with countries’ economies, rendering natural experiments such as public policies an attractive alternative to randomized experiments. Moreover, economists often use administrative observational data to exploit a natural experiment.”

A suitable natural experiment?

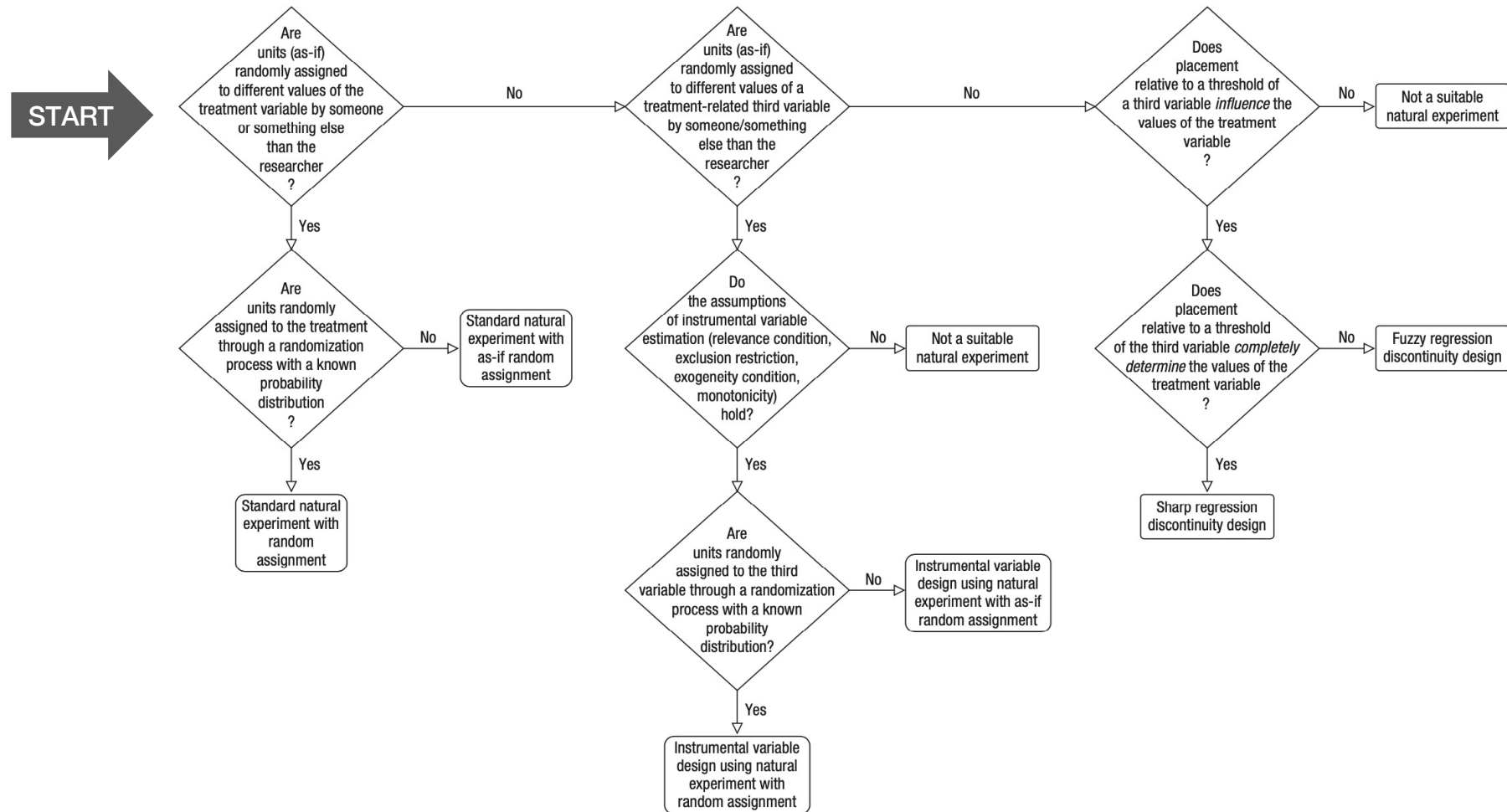
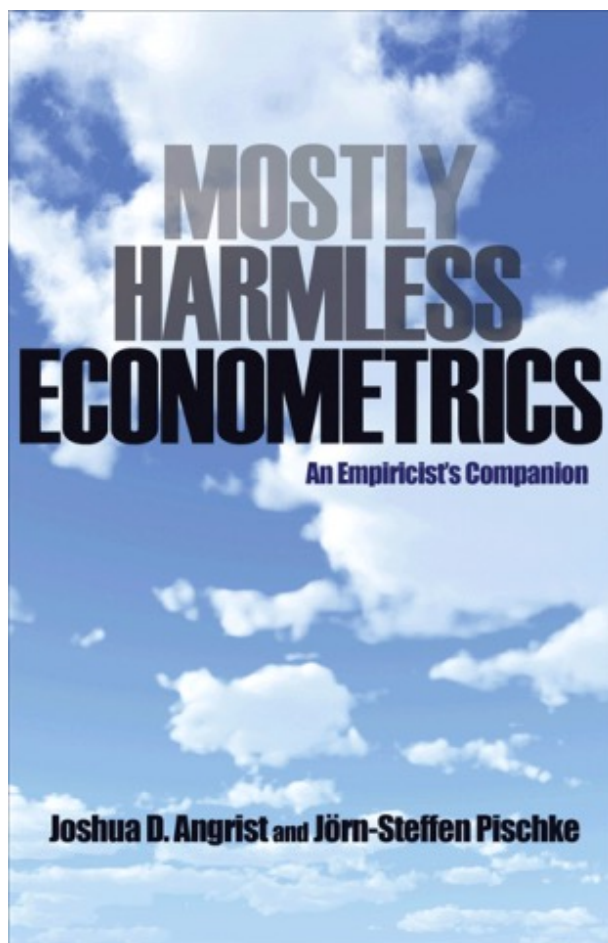


Fig. 2. Decision tree for identifying (types of) natural experiments.

“Furious Five” statistical methods for causal inference



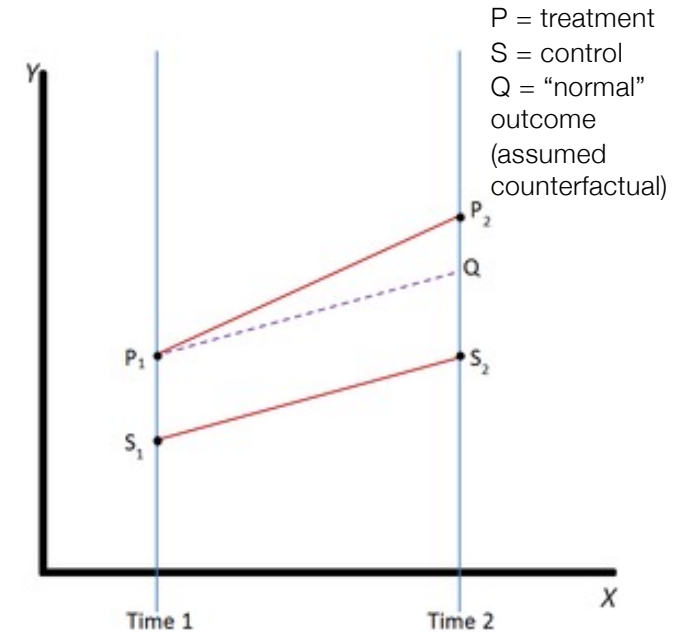
- Randomization
- Regression
- **Difference in differences**
- Regression discontinuity
- (Instrumental variables)

Difference in differences

Example: Effect of minimum wage on employment (Card & Krueger, 1994)

Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to **mimic an experimental research design using observational study data**, by studying the **differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment**.

It calculates the effect of a treatment on an outcome by **comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group**. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be **subject to certain biases** (e.g., omitted variable bias).



treatment effect = the difference between the observed value of P2 and what the value of P2 would have been with parallel trends, had there been no treatment (Q)

→ **Calculated as:**

$$\text{treatment effect} = P_2 - Q = (P_2 - P_1) - (S_2 - S_1)$$

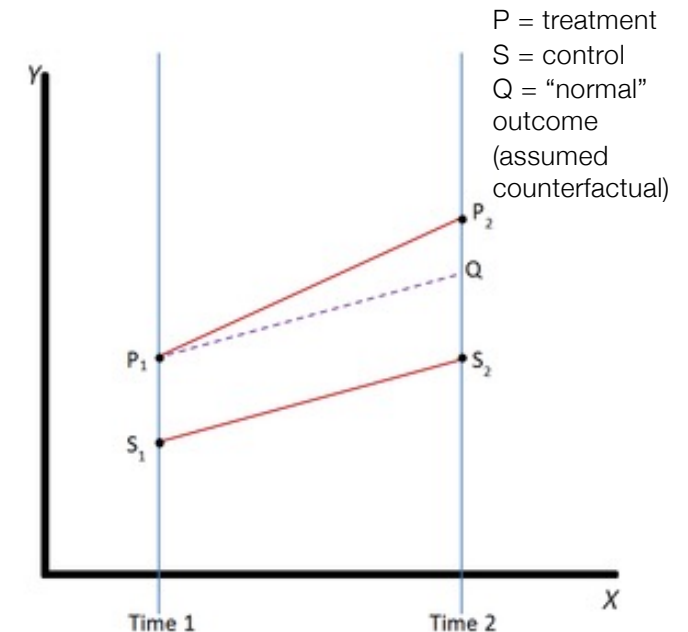
→ **or as**

$$Y = b_0 + b_1 \text{Group} + b_2 \text{Time} + b_3 \text{Group} * \text{Time} + e$$

Difference in differences

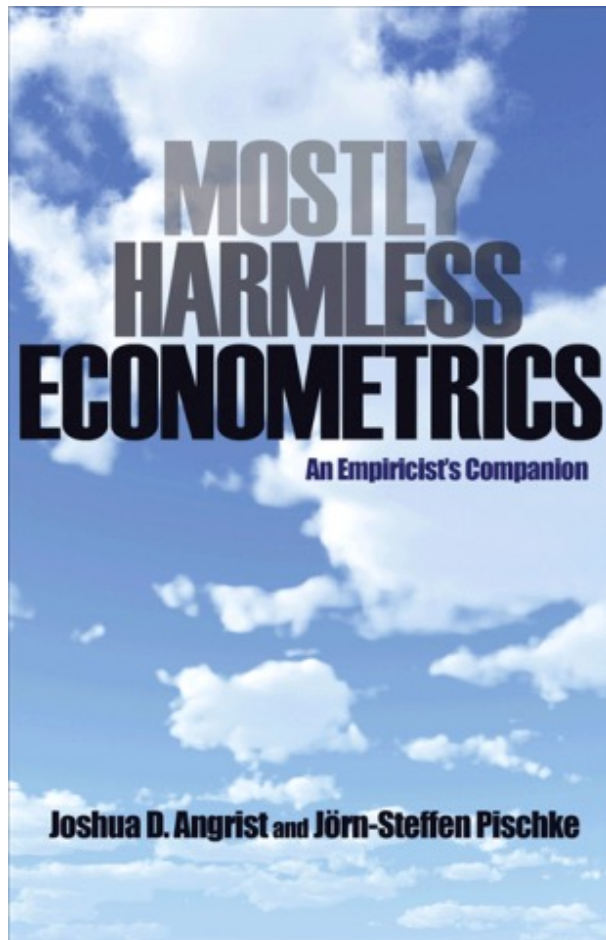
Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to **mimic an experimental research design using observational study data**, by studying the **differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment**.

It calculates the effect of a treatment on an outcome by **comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group**. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be **subject to certain biases** (e.g., omitted variable bias).



Problem: Assumption that the change in outcomes from pre- to post-intervention in the control group (S) is a good proxy for the (counterfactual) change in untreated potential outcomes in the treated group (P) may not be warranted; choice of treatment/control groups is crucial (an additional trick may be *matching* on observables)...

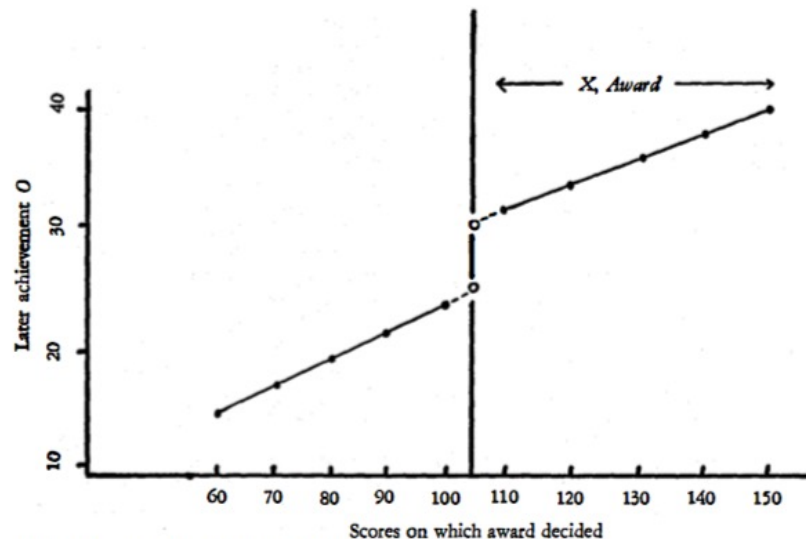
“Furious Five” statistical methods for causal inference



- Randomization
- Regression
- Difference in differences
- **Regression discontinuity**
- (Instrumental variables)

Regression discontinuity

A regression discontinuity design (RDD) is a **quasi-experimental pretest-posttest design** that elicits the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned. By **comparing observations lying closely on either side of the threshold**, it is possible to estimate the average treatment effect in environments in which randomization is unfeasible. RDD was first applied by Donald Thistlethwaite and Donald Campbell to the evaluation of scholarship programs.



→ Calculated as

Difference between award groups for achievement: $\mu_{\text{above}} - \mu_{\text{below}}$

→ or as

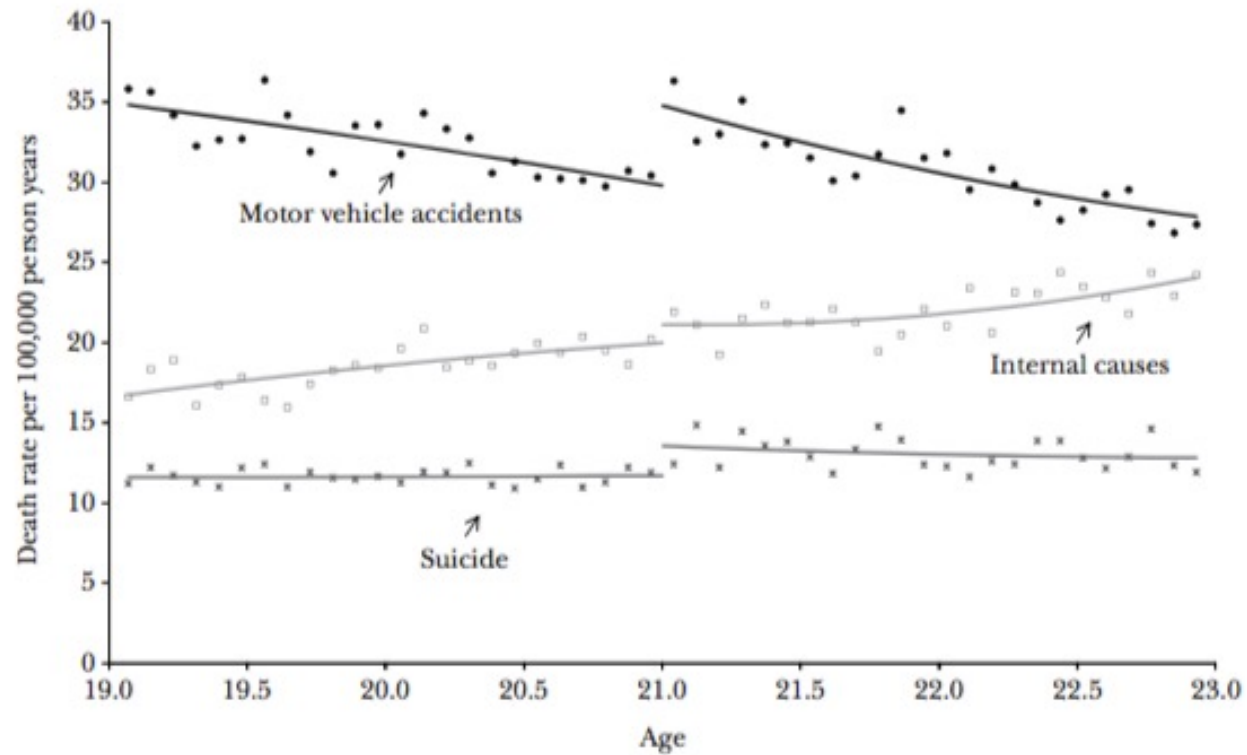
Effect of award on achievement:

$$Y = b_0 + b_1 \text{Score}_{\text{centered}} + b_2 \text{Award} + e$$

Problem: Assumption that the individuals just below the cutoff are not systematically different from those just above can be wrong (e.g., individuals just above the threshold could try harder); the estimation may not generalize to observations away from the cutoff (e.g., awards could have different results at different levels of ability).

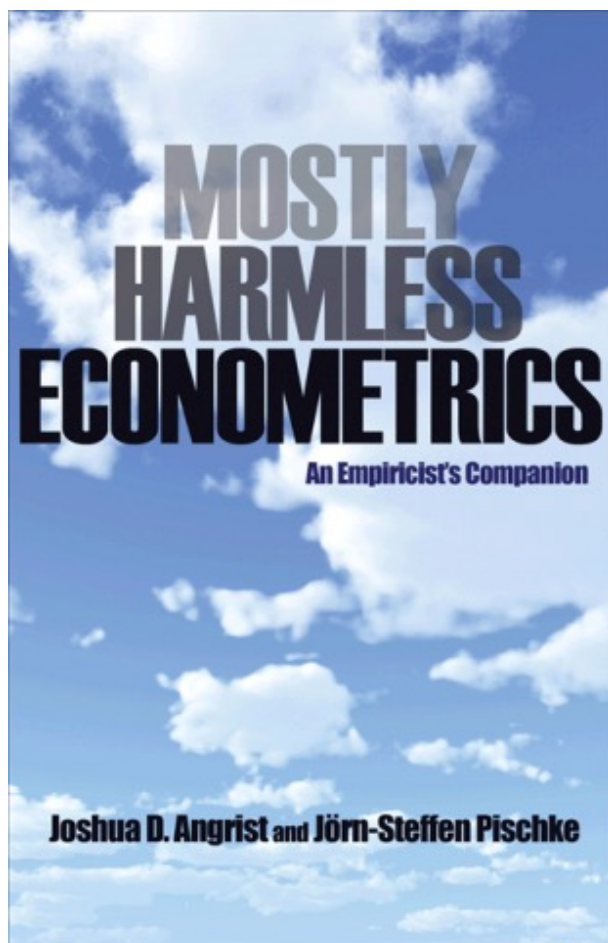
Regression discontinuity

Figure 2
Age Profiles for Death Rates in the United States



Notes: The death rates are estimated by combining the National Vital Statistics records with population estimates from the U.S. Census.

“Furious Five” statistical methods for causal inference



- Randomization
- Regression
- Difference in differences
- Regression discontinuity
- **(Instrumental variables)**

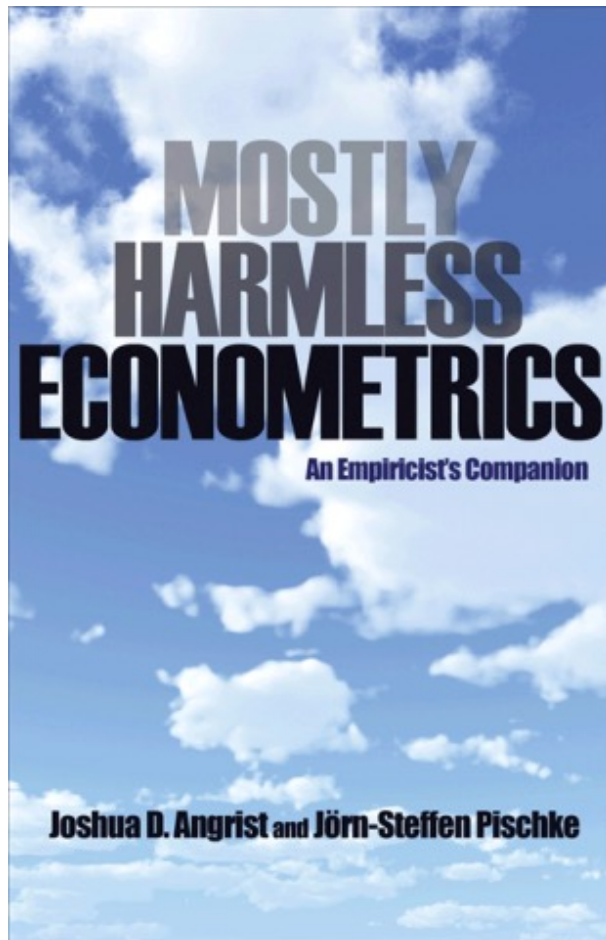
Instrumental variables

Table 1
Examples of Studies That Use Instrumental Variables to Analyze Data From Natural and Randomized Experiments

<i>Outcome Variable</i>	<i>Endogenous Variable</i>	<i>Source of Instrumental Variable(s)</i>	<i>Reference</i>
<i>1. Natural Experiments</i>			
Labor supply	Disability insurance replacement rates	Region and time variation in benefit rules	Gruber (2000)
Labor supply	Fertility	Sibling-Sex composition	Angrist and Evans (1998)
Education, Labor supply	Out-of-wedlock fertility	Occurrence of twin births	Bronars and Grogger (1994)
Wages	Unemployment insurance tax rate	State laws	Anderson and Meyer (2000)
Earnings	Years of schooling	Region and time variation in school construction	Duflo (2001)
Earnings	Years of schooling	Proximity to college	Card (1995)
Earnings	Years of schooling	Quarter of birth	Angrist and Krueger (1991)
Earnings	Veteran status	Cohort dummies	Imbens and van der Klaauw (1995)
Earnings	Veteran status	Draft lottery number	Angrist (1990)
Achievement test scores	Class size	Discontinuities in class size due to maximum class-size rule	Angrist and Lavy (1999)
College enrollment	Financial aid	Discontinuities in financial aid formula	van der Klaauw (1996)
Health	Heart attack surgery	Proximity to cardiac care centers	McClellan, McNeil and Newhouse (1994)
Crime	Police	Electoral cycles	Levitt (1997)
Employment and Earnings	Length of prison sentence	Randomly assigned federal judges	Kling (1999)
Birth weight	Maternal smoking	State cigarette taxes	Evans and Ringel (1999)

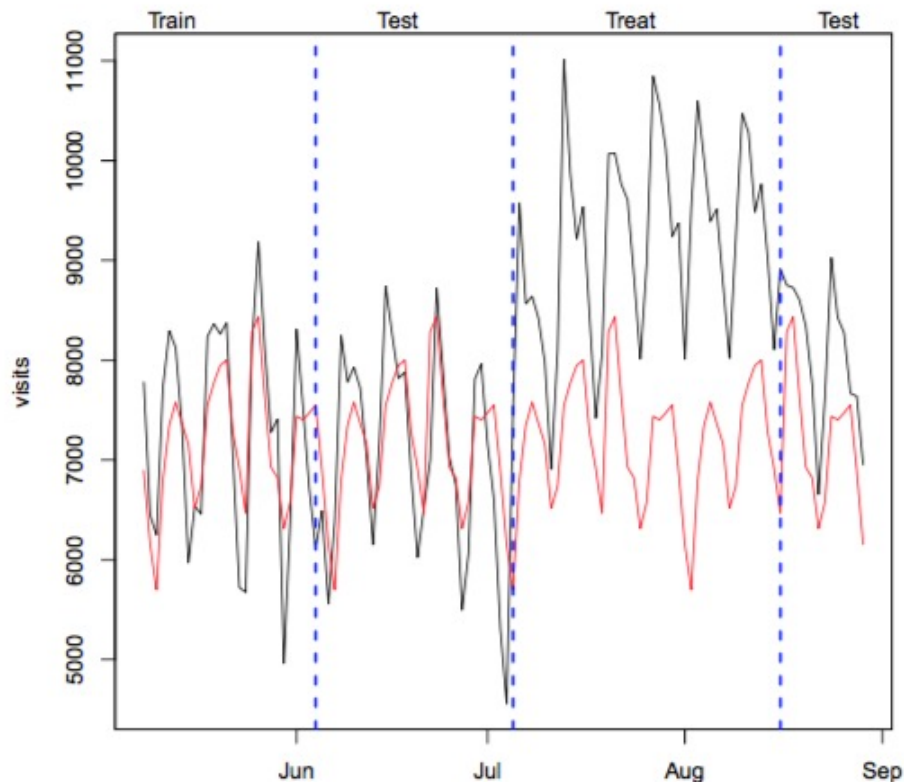
Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4), 69–85.

“Furious Six” statistical methods for causal inference?



- Randomization
- Regression
- Difference in differences
- Regression discontinuity
- (Instrumental variables)
- **Models as counterfactuals**

Using models as the counterfactual (Train-test-treat-compare)



An online advertiser might ask “if I increase my ad expenditure by some amount, how many extra sales do I generate?”

A predictive statistical model (based on number of “searches” about topics related to the subject matter of the website) is estimated during the training period and its predictive performance is assessed during the test period. The extrapolation of the model during the treat period (red line) serves as a counterfactual. This counterfactual is compared with the actual outcome (black line), and the difference is the estimated treatment effect. When the treatment is ended, the outcome returns to something close to the original level.

Summary

- **Counterfactuals:** “The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment.”
- **Causal Inference in Psychology:** There are many types of causal inference analyses that can be (and are) used in the behavioural sciences - in psychology, experiments and multiple regression from observational data are the most commonly used inference methods.
- **Furious Five / Six:** It is helpful to be aware of other methods (e.g., instrumental variables, regression discontinuity, difference in differences) and, more importantly, “the possibility of creatively utilizing the idiosyncratic features of any research situation in designing tests of causal hypotheses”.
- **Natural Experiments:** Frequently exploited in economic research, natural experiments have untapped potential for psychological science (but also limitations!).

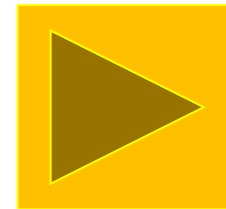
Appendix (not mandatory)

Instrumental variables

Instrumental variables



Introduction to Instrumental Variables (IV)

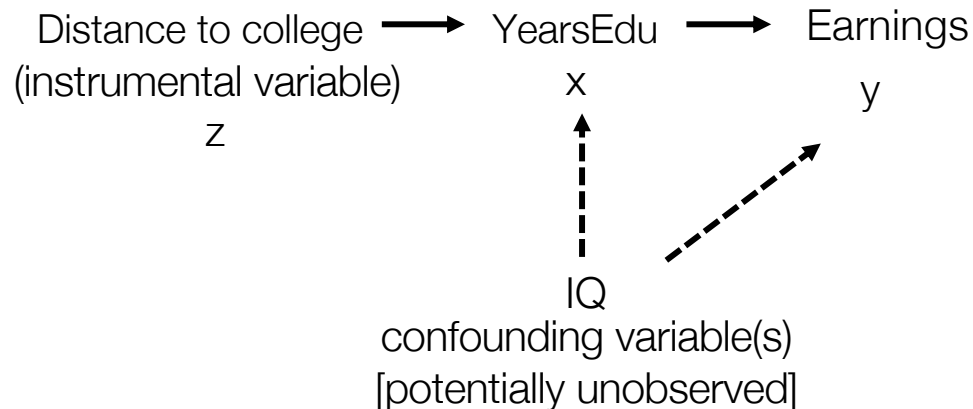


Econometrics - Instrumental Variables



Instrumental variables

The method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment. Intuitively, the method is used when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares gives biased results. **A valid instrument (instrumental variable z) induces changes in the explanatory variable (x) but has no independent effect on the dependent variable (y),** allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.



Estimation though two-stage least squares:

Stage 1: generate predictions of YearsEdu:

$$\text{YearsEdu}_{\text{pred}} = B_0 + B_1 \text{DisttoCollege} + \text{Error}$$

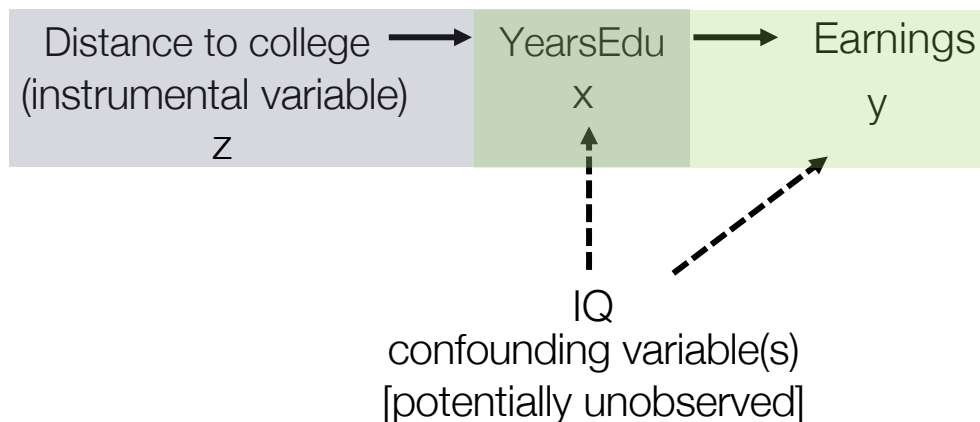
Stage 2: test whether YearsEdu_pred is

significantly associated with earnings:

$$\text{Earnings} = B_0 + B_1 \text{YearsEdu}_{\text{pred}} + \text{Error}$$

Instrumental variables

The method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment. Intuitively, the method is used when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares gives biased results. **A valid instrument (instrumental variable z) induces changes in the explanatory variable (x) but has no independent effect on the dependent variable (y),** allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.



Estimation through two-stage least squares:

Stage 1: generate predictions of YearsEdu:
$$\text{YearsEdu}_{\text{pred}} = B_0 + B_1 \text{Dist2College} + \text{Error}$$

Stage 2: test whether YearsEdu_pred is significantly associated with earnings:
$$\text{Earnings} = B_0 + B_1 \text{YearsEdu}_{\text{pred}} + \text{Error}$$

Problem: Good instrumental variables (i.e., that are correlated with x but not any confounding variables) are hard to find, and you are only estimating a local average treatment effect (not an average treatment effect)...