# Empirical Strategies in Economic History

Jesús Fernández-Villaverde[1]

September 27, 2021

[1]University of Pennsylvania

# Our task

## Questions

"For the historian, constructing the object means beginning with a question and not an archive." André Burguière, *The Annales School, An Intellectual History*, p. 23.

- Economic history is not the accumulation of one fact after another.
- Obviously, facts will help us to frame our arguments.
- But facts will not be our primary focus. Too many facts even for small periods of time and space.
- Why did things happen in the way they did?

## Searching for causes

"Life is a perpetual instruction in cause and effect." Ralph Waldo Emerson.

- How can we learn from the data?
- In particular, how can we assert relations of causality?
- Why do we care about causality?
- Difference between forecasting (conditional and unconditional) and statements of causality.
- Difference with views in history departments (Bing, 2012).

# The fragility of induction

- Problems of induction and generalization have preoccupied thinkers for centuries.
- David Hume's (1711-1776): *A Treatise of Human Nature* and *An Enquiry Concerning Human Understanding*.
- There are always hidden conditionals to any causal statement.
- Equivalently, we cannot map precisely the *prediction or reference class* of a causal statement.
- The problem of induction is that we can only be sure about a *reference class* of size 1 (particular observation).

# David Hume's explanation

**An Enquiry Concerning Human Understanding, Section 4**

"These two propositions are far from being the same, *I have found that such an object has always been attended with such an effect*, and *I foresee, that other objects, which are, in appearance, similar, will be attended with similar effects*. I shall allow, if you please, that the one proposition may justly be inferred from the other: I know in fact, that it always is inferred. But if you insist, that the inference is made by a chain of reasoning, I desire you to produce that reasoning. The connexion between these propositions is not intuitive."

# Causality in social sciences

- While the problem of inference is serious in natural sciences, it becomes considerably acuter in social sciences:

  1. Measurement issues. For example, how do we measure a "social norm"?

  2. Expectations matter. An electron does not change its behavior depending on what you are planning to do next in a lab. Humans *do* change their behavior depending on what their expectations of future policy are.

  3. Behavior is endogenous. Milton Friedman's thermostat metaphor.

  4. Performing controlled lab experiments is much harder (although not totally impossible) and limited in scope.

  5. Social phenomena live in a world of high causal density (Jim Manzi).

  6. Ideological positions biased our reading of the evidence (although this problem also appears sometimes in natural sciences: evolution, climate change, ...).

# Beliefs

- The main consequence of the previous concerns is that, instead of a high degree of certainty, most "empirical findings" in social sciences are only under a "degree of belief."
- Some findings have very high degrees of belief, some findings have lower degrees of belief.
- What determines the "degree of belief"?
    1. Credibility of empirical strategy (for example, the strength of assumptions).
    2. Temporality and size of effect.
    3. Repeated findings in the literature by different scholars.
    4. Agreement with what we believe we know in other contexts.
- Judgement by researcher and community.

## Beliefs and decisions

- Even if we only have moderate "degree of belief," we still need to make decisions in real life.
- Think about many choices in economic policy.
- Different decisions require different "degrees of belief": preponderance of the evidence in civil adjudications vs. evidence beyond reasonable doubt in criminal procedures.
- One needs to think about decision making with an objective function (and its possible asymmetries).
- Unlikely to reach good decisions/research conclusions unless we have a candid assessment of the existing uncertainty.
- Scylla of ignoring evidence and Charybdis of inordinate reliance on empirical results.

# A sound balance

**Chris Sims, Journal of Economic Perspectives (2010)**

"Because economics is not an experimental science, economists face difficult problems of inference. The same data generally are subject to multiple interpretations. It is not that we learn nothing from data, but that we have at best the ability to use data to narrow the range of substantive disagreement. We are always combining the objective information in the data with judgment, opinion and/or prejudice to reach conclusions. Doing this well can require technically complex modeling. Doing it in a scientific spirit requires recognizing and taking account of the range of opinions about the subject matter that may exist in one's audience. That is, it requires balancing the need to use restrictive assumptions on which there may be substantial agreement against the need to leave lightly restricted those aspects of the model on which the data might help resolve disagreement.".

## An example

- Let's look at a concrete example:

  The Colonial Origins of Comparative Development: An Empirical Investigation by Daron Acemoglu, Simon Johnson, and James Robinson; AJR.

- Many economists have argued that secure property rights are key for economic growth.
- Intuition: incentives for investment, technological development, etc.
- If true, this hypothesis has important consequences for economic history.
- How would you document that secure property rights matter?

(potential) settler mortality $\Rightarrow$ settlements

$\Rightarrow$ early institutions $\Rightarrow$ current institutions

$\Rightarrow$ current performance.

# Three main approaches

- Three approaches for understanding the data:

  1. Analytic narratives (i.e., historical narratives disciplined by formal reasoning).
  2. Statistical models (i.e., models based on flexible statistical representations of the data).
  3. Structural models (i.e., models based on economic theory).

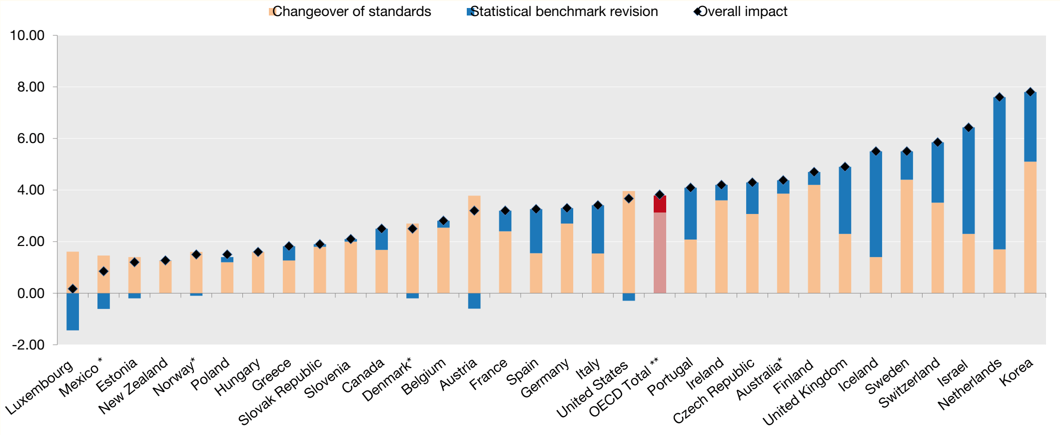
- In practice, best economic history work combines all three approaches.

- In these slides, we will focus on statistical models.

## Step I: measurement

- First, one needs to gather data.
- Data are "construed." Thus, "let the data speak by themselves" is an oxymoron.
- Most obvious example: gross domestic product.
- Financial service, R&D expenditure, etc.
- But even measures such as life expectancy, child mortality, or population suffer from this problem.
- Other problems:
  1. Faulty data collection.
  2. Imputation mistakes.
  3. Outright lies (i.e., most official data from communist countries during the 20th century).

## The overall impact of the benchmark revision on GDP-levels, in year 2010 (changeover to SNA 2008 and statistical benchmark revision)



# Measurement in the era of big data

- Furthermore, there is a large number of possible sources of data we can look at.
- For example: geospatial data (GIS), internet searches, video, library records.
- Much of the best recent work in economic history has come from original data sources.
- Big data techniques do not eliminate this problem, it just transforms it in subtle ways.

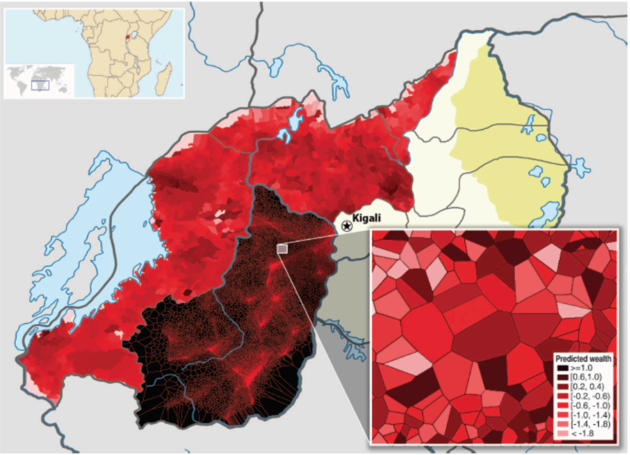



**Fig. 2. Construction of high-resolution maps of poverty and wealth from call records.** Information derived from the call records of 1.5 million subscribers is overlaid on a map of Rwanda. The northern and western provinces are divided into cells (the smallest administrative unit of the country), and the cell is shaded according to the average (predicted) wealth of all mobile subscribers in that cell. The southern province is overlaid with a Voronoi division that uses geographic identifiers in the call data to segment the region into several hundred thousand small partitions. (**Bottom right inset**) Enlargement of a 1-km$^2$ region near Kiyonza, with Voronoi cells shaded by the predicted wealth of small groups (5 to 15 subscribers) who live in each region.

Blumenstock et. al. (2015)

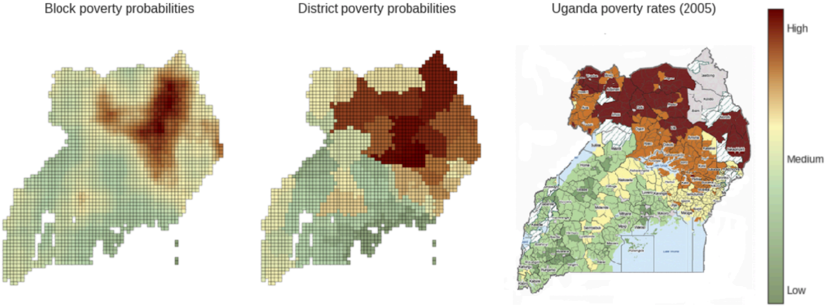
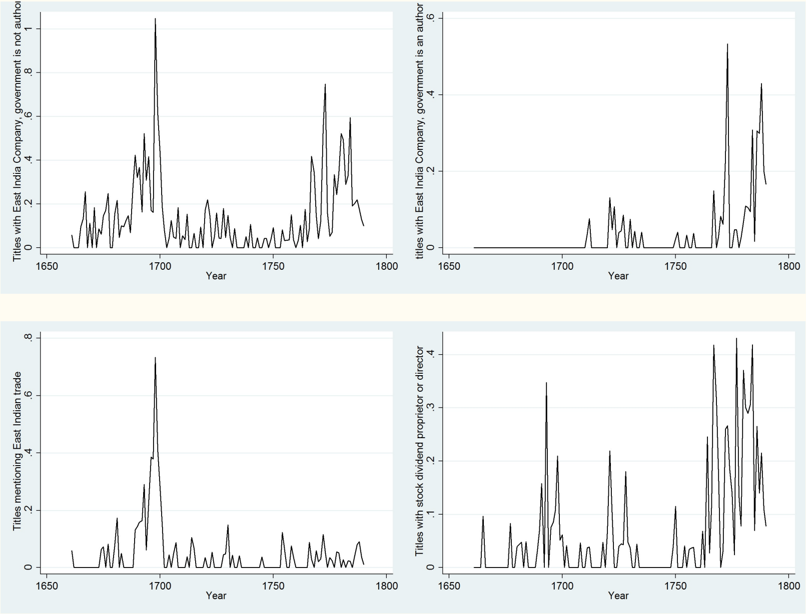



Figure 3: **Left:** Predicted poverty probabilities at a fine-grained 10km × 10km block level. **Middle:** Predicted poverty probabilities aggregated at the district-level. **Right:** 2005 survey results for comparison (World Resources Institute 2009).

Xie et. al. (2016)

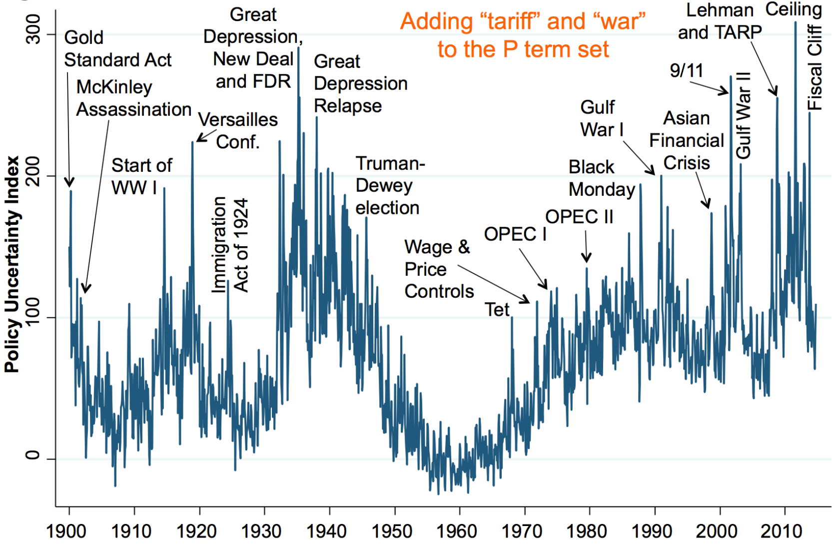
Figure 5: East India published titles as a percentage of all English titles



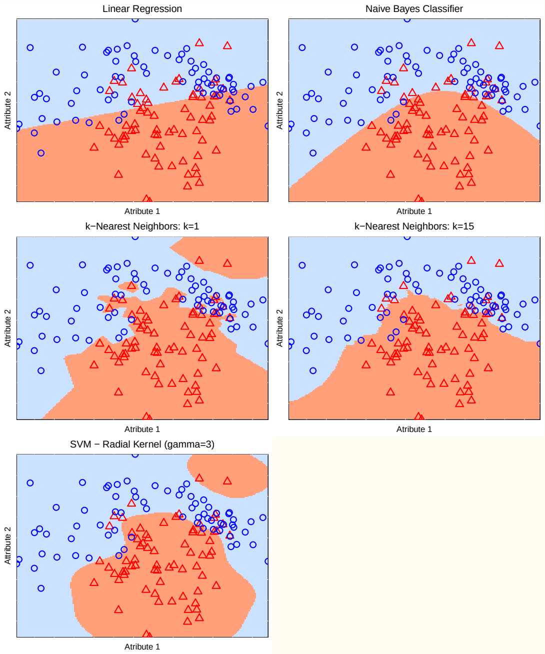
## Machine learning and measurement

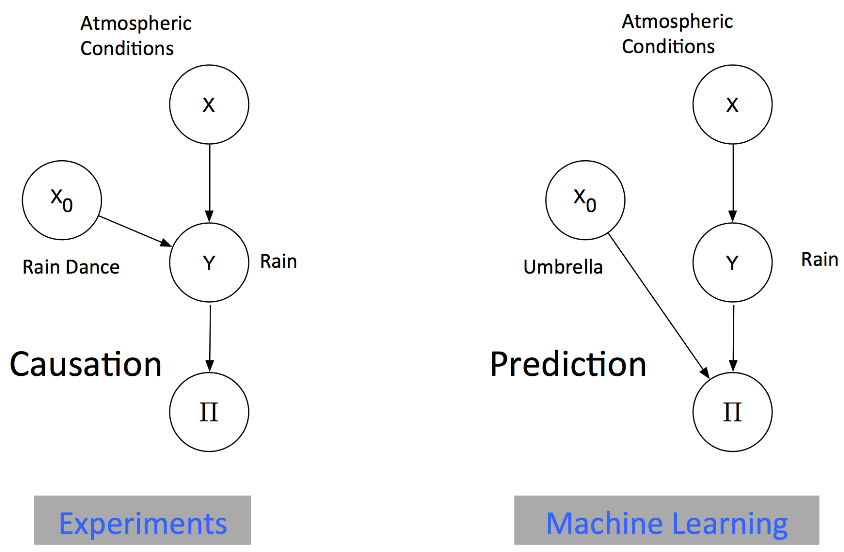
- Machine learning is having a growing impact on economic history.
- Deep learning is potentially promising.
- Role of machine learning:
    1. Building data.
    2. Reading data.
    3. Supervised vs. unsupervised learning.

## Figure 2: Historical U.S. EPU Index, Jan. 1900 to Dec. 2012




**Notes:** Index reflects scaled monthly counts of articles in 6 major newspapers (Washington Post, Boston Globe, LA Times, NY Times, Wall Street Journal, and Chicago Tribune) that contain the same triple as in Figure 1, except the E term set includes "business", "commerce" and "industry" and the P term set includes "tariffs" and "war". Data normalized to 100 from 1900-2011.


Linear Regression
Naive Bayes Classifier
k−Nearest Neighbors: k=1
k−Nearest Neighbors: k=15
SVM − Radial Kernel (gamma=3)
Attribute 2
Atribute 1


Atmospheric Conditions
X
$X_0$
Y
Rain Dance
Rain
Causation
Π
Experiments
Atmospheric Conditions
X
$X_0$
Y
Umbrella
Rain
Prediction
Π
Machine Learning

## Some of the variables in AJR

- Log of GDP per capita, 1975 and 1995, and output per worker, 1988.
- Average protection against expropriation risk, 1985-1995.
- Constraint on executive in 1900, 1970, 1990, and in the first year of independence.
- Democracy in 1900 and first year of independence.
- Ethnolinguistic fragmentation.
- Religion variables.
- Log European settler mortality.
- Yellow fever.
- Distance from the coast.

## Step II: descriptive statistics

- Once we have compiled the data, we can analyze it.
- Simple descriptive statistics (means, median, s.d., quantiles, ...) and plotting the data.
- Also, hypothesis testing.
- Often, descriptive statistics can be surprisingly effective.
- No amount of formal treatment can substitute the "reality-check" of assessing the raw data and basic statistics.

| | Whole world | Base sample | By quartiles of mortality (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|---|
| Log GDP per capita (PPP) in 1995 | 8.3 (1.1) | 8.05 (1.1) | 8.9 | 8.4 | 7.73 | 7.2 |
| Log output per worker in 1988 (with level of United States normalized to 1) | −1.70 (1.1) | −1.93 (1.0) | −1.03 | −1.46 | −2.20 | −3.03 |
| Average protection against expropriation risk, 1985–1995 | 7 (1.8) | 6.5 (1.5) | 7.9 | 6.5 | 6 | 5.9 |
| Constraint on executive in 1990 | 3.6 (2.3) | 4 (2.3) | 5.3 | 5.1 | 3.3 | 2.3 |
| Constraint on executive in 1900 | 1.9 (1.8) | 2.3 (2.1) | 3.7 | 3.4 | 1.1 | 1 |
| Constraint on executive in first year of independence | 3.6 (2.4) | 3.3 (2.4) | 4.8 | 2.4 | 3.1 | 3.4 |
| Democracy in 1900 | 1.1 (2.6) | 1.6 (3.0) | 3.9 | 2.8 | 0.19 | 0 |
| European settlements in 1900 | 0.31 (0.4) | 0.16 (0.3) | 0.32 | 0.26 | 0.08 | 0.005 |
| Log European settler mortality | n.a. | 4.7 (1.1) | 3.0 | 4.3 | 4.9 | 6.3 |
| Number of observations | 163 | 64 | 14 | 18 | 17 | 15 |

## Step III: reduced-form analysis

- Slightly more involved analysis: reduced-form.
- Why "reduced-form"?
- Observed statistical behavior might be the consequence of complicated non-linear interactions in the "structural-form."
- However, what is "structural-form" is context-dependent and it is a function of the class of policy interventions we are interested in evaluating (Hurwicz, 1962).

# Benchmark linear regression

- Imagine that we have observations $\{y_i, x_i\}_{i=1}^{N}$. $i$ is the level of the observation (individuals, regions, countries,...) and $N$ is the # of observations.
- Linear regression:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

  where:

  1. $y_i$ is the dependent variable (also known as the regresand or the left-hand variable).
  2. $y_i$ is the independent variable (also known as the regressor or the right-hand variable).
  3. $\varepsilon_i$ is the error term.
  4. $\beta_0$ is the intercept.
  5. $\beta_1$ is the slope.

# Conditional expectation function

- For two random variables $\{Y, X\}$, $\mathbb{E}[Y|X]$ is the expectation of $Y$ conditional on $X$.
- $\mathbb{E}[Y|X]$ is known as the conditional expectation function (CEF).
- Given realizations $y_i$ and $x_i$ of $\{Y, X\}$, we can always write:

$$y_i = \mathbb{E}[Y|X = x_i] + \varepsilon_i$$

  where:

  1. $\mathbb{E}[Y|X = x_i]$ is the expectation of $Y$ conditional on $X = x_i$.
  2. $\mathbb{E}[\varepsilon_i|X = x_i] = 0$.
- The CEF is the MMSE predictor of Y conditional on X.

# Why linear regression?

- The CEF is an unknown function.
- We can approximate an unknown function by using a basis of monomials: $1, x, x^2, \dots$ multiplied a vector of coefficients $\beta_0, \beta_1, \beta_2, \dots$
- The Stone-Weierstrass theorem ensures us that this approximation converges in the "right" sense.
- Then,

$$\mathbb{E}\left[Y|X = x_i\right] = \beta_0 + \beta_1 x_i + \beta_2 x^2 + \dots$$

- In practice, we want to truncate the approximation at a low degree of the polynomial. For example, linear:

$$\mathbb{E}\left[Y|X = x_i\right] \simeq \beta_0 + \beta_1 x_i$$

- And we get:

$$\begin{aligned} y_i &= \mathbb{E}\left[Y|X = x_i\right] + \varepsilon_i \\ &\simeq \beta_0 + \beta_1 x_i + \varepsilon_i \end{aligned}$$

# Why OLS?

- We still need to determine $\beta = \{\beta_0, \beta_1\}$.
- We use some criterira that minimizes the distance between $\mathbb{E}[Y|X = x_i]$ and $\beta_0 + \beta_1 x_i$.
- Since we are dealing with the difference between two functions, we need a function metric.
- A standard choice: minimize the square of the error terms in the sample:

$$\widetilde{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} [y_i - (\beta_0 + \beta_1 x_i)]^2$$

- This is called "ordinary least squares" (or OLS for short).
- By construction, it is the MMSE linear estimator of the CFE.

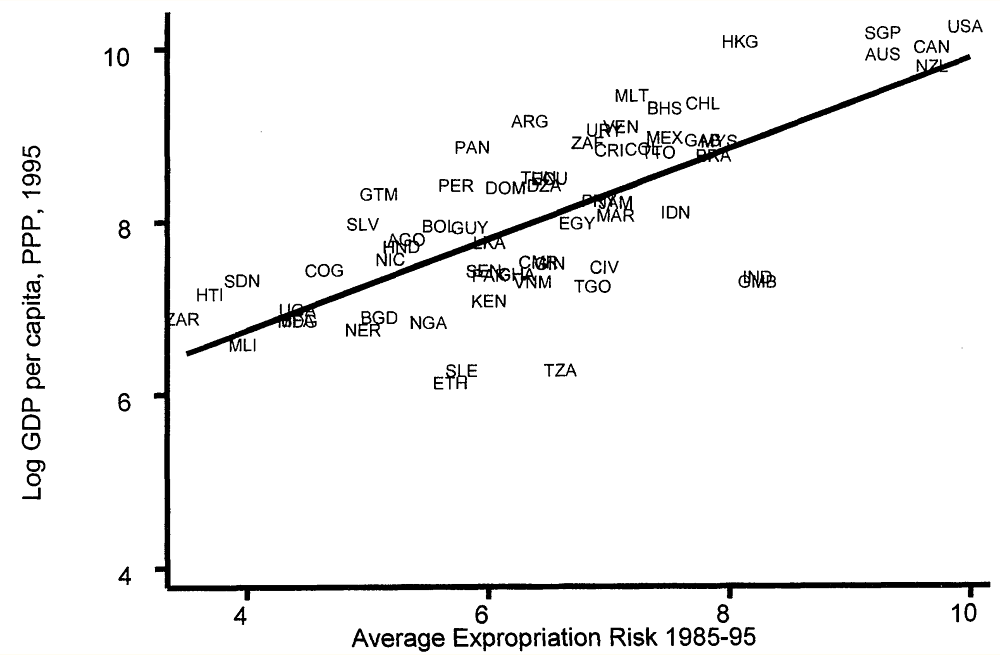

Log GDP per capita, PPP, 1995
Average Expropriation Risk 1985-95
10
8
6
4
4
6
8
10
HKG
SGP
USA
AUS
CAN
NZL
MLT
BHS
CHL
ARG
VEN
MEX
PAN
TTO
BRA
GTM
PER
DOM
JAM
IDN
MAR
SLV
EGY
GUY
CIV
COG
NIC
SDN
TGO
HTI
KEN
ZAR
BGD
NGA
NER
MLI
TZA
SLE
ETH
VNM

TABLE 2—OLS REGRESSIONS

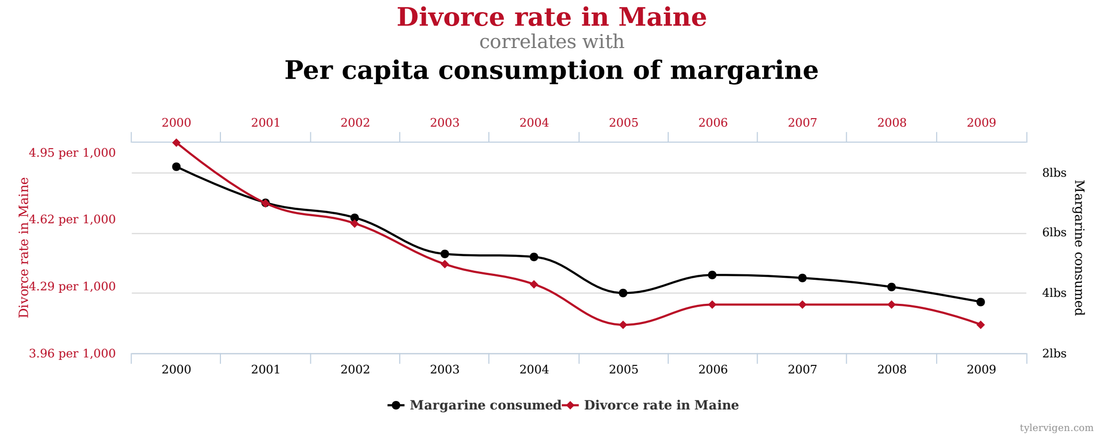
| | Whole world (1) | Base sample (2) | Whole world (3) | Whole world (4) | Base sample (5) | Base sample (6) | Whole world (7) | Base sample (8) |
|---|---|---|---|---|---|---|---|---|
| | Dependent variable is log GDP per capita in 1995 | | | | | | Dependent variable is log output per worker in 1988 | |
| Average protection against expropriation risk, 1985–1995 | 0.54 | 0.52 | 0.47 | 0.43 | 0.47 | 0.41 | 0.45 | 0.46 |
| | (0.04) | (0.06) | (0.06) | (0.05) | (0.06) | (0.06) | (0.04) | (0.06) |
| Latitude | | | 0.89 | 0.37 | 1.60 | 0.92 | | |
| | | | (0.49) | (0.51) | (0.70) | (0.63) | | |
| Asia dummy | | | | −0.62 | | −0.60 | | |
| | | | | (0.19) | | (0.23) | | |
| Africa dummy | | | | −1.00 | | −0.90 | | |
| | | | | (0.15) | | (0.17) | | |
| "Other" continent dummy | | | | −0.25 | | −0.04 | | |
| | | | | (0.20) | | (0.32) | | |
| $R^2$ | 0.62 | 0.54 | 0.63 | 0.73 | 0.56 | 0.69 | 0.55 | 0.49 |
| Number of observations | 110 | 64 | 110 | 110 | 64 | 64 | 108 | 61 |

# OLS: strengths

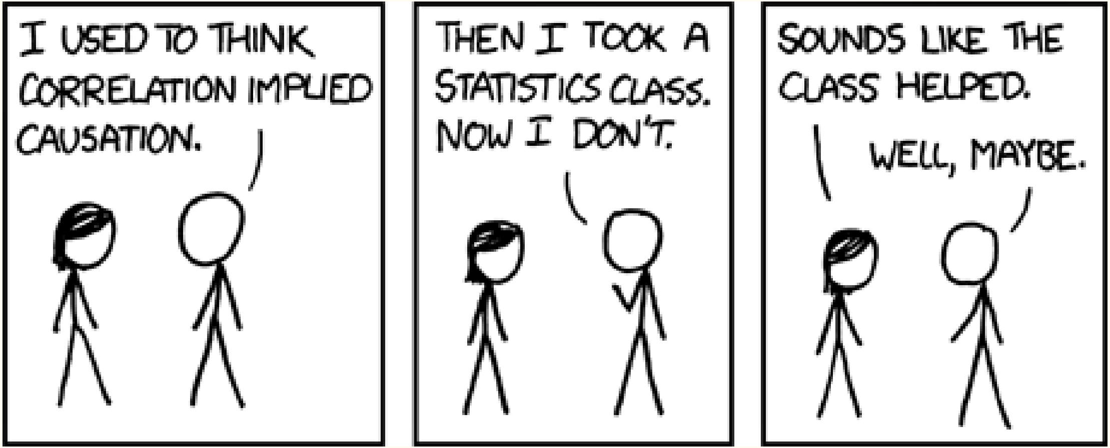
- We have not assumed anything except that the CEF can be well approximated by a linear function.
- This is enough to gives us a powerful way to look at the data: $\beta_1 = \frac{\sigma_{XY}}{\sigma_X{}^2}$.
  1. Document "stylized facts."
  2. Forecasting.
  3. Assess performance of a formal model.
- Alternative interpretations of OLS: best linear predictor, linear projection.

## OLS: weaknesses

- We are not making additional assumptions (normality of innovations, etc.) required to prove properties such as unbiasedness or consistency.
- We often care about these properties.
- However, these additional assumptions limit the scope of interpretability.
- No causal interpretation.
- Also, it maximizes bias in the bias-variance tradeoff (BLUE).
- Simple alternative: regularization (Lasso).



Divorce rate in Maine
correlates with
Per capita consumption of margarine
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
4.95 per 1,000
4.62 per 1,000
4.29 per 1,000
3.96 per 1,000
Divorce rate in Maine
8lbs
6lbs
4lbs
2lbs
Margarine consumed
Margarine consumed
Divorce rate in Maine
tylervigen.com


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

# Potential outcomes

- Neyman-Rubin counterfactual framework.

- *Potential outcomes*: the outcome of interest for the researcher that agent $i$ (a country, a region, a firm, a family, an individual) would have if not treated ($D_i = 0$) or treated ($D_i = 1$).

- Notation:

$$
\begin{aligned}
Y_i &= \begin{cases} Y_{0i} \text{ if } D_i = 0 \\ Y_{1i} \text{ if } D_i = 1 \end{cases} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i
\end{aligned}
$$

- Easy to generalize to continuous and/or multivariate treatments.

# Causal effect

- *Causal effect of treatment*: the difference in the outcome of interest due to the treatment.
- Notation:

$$Y_{1i} - Y_{0i}$$

- Example: causal effect of imposing property rights protections on economic growth of a country $i$.
- Mill (1848), Marshall (1890), and Haavelmo (1943).

# Challenge

- Counterfactuals: we do not observe both potential outcomes, only one of them.
- Furthermore, $Y_{1i} - Y_{0i}$, $Y_{1i}$, and $Y_{0i}$ are heterogenous across $i$, even after controlling for observables.
- Conceptually: missing data problem.
- Economists deal with ways to get around these two challenges:
    1. Definitions of counterfactuals.
    2. Identification of causal models from population distributions.
    3. Identification of causal models from actual data.

## A first approach

- We can compare the observed averages of agents $i$ depending on the treatment:

$$\frac{1}{n_1}\sum_{i}^{n_1} Y_{1i} - \frac{1}{n_2}\sum_{i}^{n_2} Y_{0i}$$

where

$$\underbrace{n}_{\text{Total \# agents}} = \underbrace{n_1}_{\text{Total \# treated agents}} + \underbrace{n_1}_{\text{Total \# non-treated agents}}$$

- Problem: *selection bias*.

- Intuition.

# A formal explanation I

- We start from the definition of the *average treatment effect* or ATE given controls $X$:

$$\underbrace{\mathbb{E}\left[Y_i|X, D_i = 1\right] - \mathbb{E}\left[Y_i|X, D_i = 0\right]}_{\text{ATE(X)}}$$

- Decomposition:

$$\begin{aligned}
ATE(X) &= \mathbb{E}\left[Y_{0i} + \left(Y_{1i} - Y_{0i}\right) D_i|X, D_i = 1\right] \\
&\quad -\mathbb{E}\left[Y_{0i} + \left(Y_{1i} - Y_{0i}\right) D_i|X, D_i = 0\right] \\
&= \mathbb{E}\left[Y_{1i}|X, D_i = 1\right] - \mathbb{E}\left[Y_{0i}|X, D_i = 0\right] \\
&= \underbrace{\mathbb{E}\left[Y_{1i}|X, D_i = 1\right] - \mathbb{E}\left[Y_{0i}|X, D_i = 1\right]}_{\text{Average treatment effect on the treated}} \\
&\quad +\underbrace{\mathbb{E}\left[Y_{0i}|X, D_i = 1\right] - \mathbb{E}\left[Y_{0i}|X, D_i = 0\right]}_{\text{Selection bias}}
\end{aligned}$$

# A formal explanation II

- *Average treatment effect on the treated* or TT is the effect of the treatment on those actually treated.

- *Selection bias* is the difference in outcome between the untreated group had they been treated and what happens to them.

- *Selection bias* is sometimes attributed to omitted (observable) variable bias, but it actually depends on unobservable heterogeneity.

# An additional decomposition

- *Average treatment effect on the untreated* or UT is the effect of the treatment on those not treated:.

$$UT(X) = \mathbb{E}\left[Y_{1i}|X, D_i = 0\right] - \mathbb{E}\left[Y_{0i}|X, D_i = 0\right]$$

- Then, using standard probability arguments:

$$ATE(X) = Pr\left(D_i = 1|X\right) TT(X) + Pr\left(D_i = 0|X\right) UT(X)$$

# Inference problem

- We can directly estimate, with observable data, $\mathbb{E}[Y_{1i}|X, D_i = 1]$ and $\mathbb{E}[Y_{0i}|X, D_i = 0]$.
- However, we cannot direcly estimate, with observable data, $\mathbb{E}[Y_{0i}|X, D_i = 1]$ and $\mathbb{E}[Y_{1i}|X, D_i = 0]$.
- Thus, we cannot directly estimate TT and UT, which are the real objects of interest.
- How do we address this estimation problem?

# Randomization I

- Imagine that an agent $i$ is assigned to a treatment randomly.
- Recall that:

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

- Then:

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

where the third and fifth lines come from the independence of $Y_{0i}$ and $D_i$.

# Randomization II

- Intuition: groups being treated are equivalent and, hence, we control for material hidden conditionals that go beyond $X$.
- Note that random assignment is is different from random sampling.
- Achieving true randomization can be harder in practice than it theory.
- Ethical problems (related to the irrelevance of stopping rule principle).

# Randomized field trials (RFTs)

- Also called randomized controlled trials (RCTs).

- Two components:

  1. Treated vs. control groups.

  2. Randomized assigment (if feasible, double-blind).

- Two technical conditions:

  1. Checking for balance (although the test is less informative than it might seem). Stratification by simple observables.

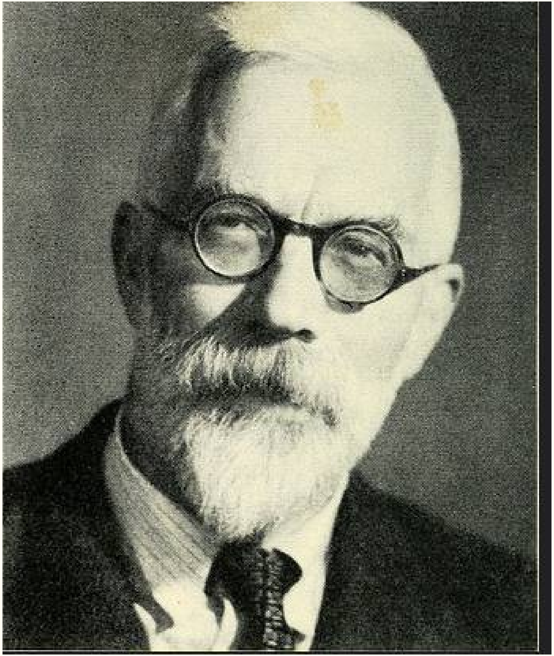
  2. Sufficient power ("signal-to-noise" ratio: the size of the causal effect to be measured in comparison with underlying variation in the data).

## Treated vs. control groups

- Precursors: Book of Daniel, al-Razi, Avicenna, Ben Cao Tu Jing.
- First well-documented case: James Lind in 1742 dealing scurvy in HMS *Salisbury*.
- Louis Pasteur in 1882 anthrax vaccine experiment.

# Randomized assignment

- Historical precedent: Van Helmont in 17th century.
- C.S. Pierce: 1884 in weight-feeling experiments.
- Social sciences: 1928 at Purdue University (effect of exemptions in exams).
- Joseph Bell in 1938: pertussis vaccine trial in Norfolk (Virgina) $\Rightarrow$ "intent-to-treat" principle (compliance effect vs. causal effect).
- Theory developed by R.A. Fisher in 1925 and, particularly, in his 1935 classic The Design of Experiments. Also, Jerzy Neyman.
- Recent technology: only became generally applied in the 1950s and 1960s. Until the late 1970s suffered from opposition.



# Randomized field trials in practice

- Randomization can be applied in:

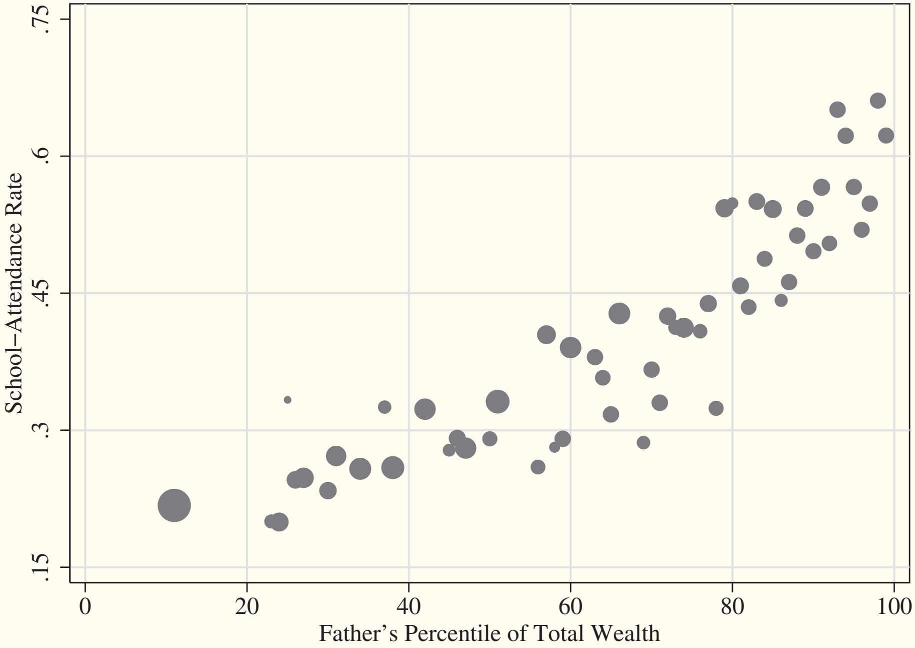
  1. Assessing therapeutic efficacy in phase III of clinical trials.
  2. Natural sciences.
  3. Marketing.
  4. Micro programs (school choice, class size, labor, ..).

- Randomization is difficult to apply to:

  1. Aggregate agents.
  2. Historical events.

# A question: education and wealth

- Is investment in education of children limited by the wealth of parents?
- Strong correlation between father's wealth and children educational attaintment.
- Clark and Cummins (2014) and Clark (2014): persistence lasts 8 centuries.
- Public policies could improve outcomes for children from low wealth parents.
- However, characteristics of parents may be passed to children (genes, cultural norms, position in society, ...).


School–Attendance Rate
.15
.3
.45
.6
.75
0
20
40
60
80
100
Father's Percentile of Total Wealth

# Randomization as an answer

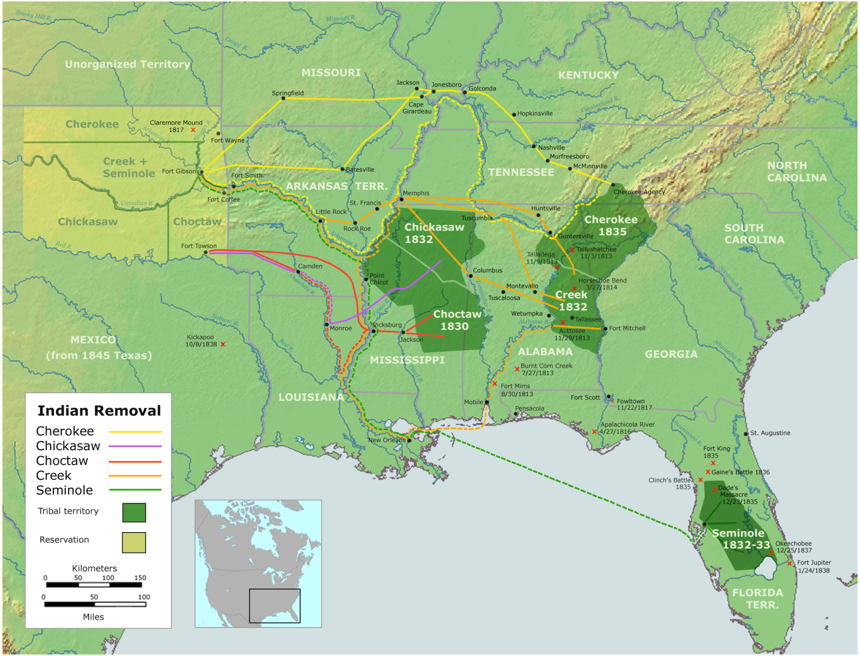
- Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations, by Bleakley and Ferrie (2016).

- Five Civilized Tribes (Cherokee, Chickasaw, Choctaw, Muscogee, and Seminole) evicted by the *Indian Removal Act of 1830*.

- Cherokee are expelled from Northwest Georgia in 1832 (although the Trail of Tears deportation happened in 1838 after a legal battle).

- State of Georgia divides the Cherokee former lands in around 18,309 160-acre parcels and allocates them in a land lottery (a tradition in the state).

- Good and productive land. Some of it with gold deposits.


Indian Removal
Cherokee
Chickasaw
Choctaw
Creek
Seminole
Tribal territory
Reservation
Kilometers
Miles
Unorganized Territory
MISSOURI
KENTUCKY
TENNESSEE
NORTH CAROLINA
SOUTH CAROLINA
ARKANSAS TERR.
MISSISSIPPI
ALABAMA
GEORGIA
LOUISIANA
FLORIDA TERR.
MEXICO (from 1845 Texas)
Cherokee
Creek + Seminole
Chickasaw
Choctaw
Chickasaw 1832
Cherokee 1835
Creek 1832
Choctaw 1830
Seminole 1832-33
Claremore Mound 1817
Fort Wayne
Fort Gibson
Fort Smith
Fort Coffee
Fort Towson
Springfield
Jackson
Jonesboro
Golconda
Cape Girardeau
Hopkinsville
Nashville
Murfreesboro
McMinnville
Cherokee Agency
Batesville
Memphis
St. Francis
Little Rock
Rock Roe
Camden
Point Chicot
Monroe
Vicksburg
Jackson
Tuscumbia
Huntsville
Guntersville
Tallushatchee 11/3/1813
Talladega 11/9/1813
Horseshoe Bend 3/27/1814
Columbus
Montevallo
Tuscaloosa
Wetumpka
Tallassee
Autossee 11/29/1813
Fort Mitchell
Burnt Corn Creek 7/27/1813
Fort Mims 8/30/1813
Mobile
Pensacola
Fort Scott
Fowltown 11/22/1817
Apalachicola River 4/27/1816
New Orleans
St. Augustine
Fort King 1835
Gaine's Battle 1836
Clinch's Battle 1835
Dade's Massacre 12/23/1835
Okeechobee 12/25/1837
Fort Jupiter 1/24/1838
Kickapoo 10/8/1838

## The land lottery

- Rule: every white male resident in Georgia for at least three years, 18 and over, enters once. If, in addition, he has a wife or children under 18, he enters twice. Some exceptions (prisoners out, widows in).

- Over 98 percent of eligible men registered for the lottery: value of the parcel was around five years of an unskilled worker wage ($\approx$ median wealth in Georgia at the time).

- Roughly 85,000 slips with names in one drum and an identical number of slips (18,309 with locations of parcels, more than 66,000 blank slips) in a second drum. Thus, both winning and which parcel was won was random.

- Winners could sell their claims right away: no homesteading requirement.

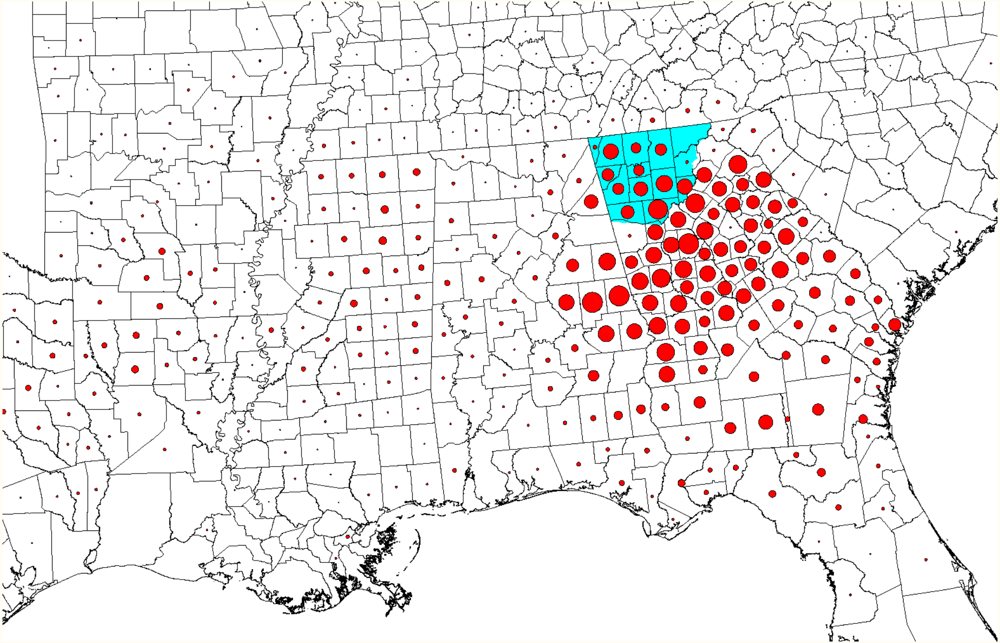
- Winners: treated group; losers: control group.

## The data

- James Smith published in 1838 the list of winners.
- Bleakley and Ferrie find all men residing in Georgia in the 1830 U.S. Census and locate them in the 1850 U.S. Census.
- Also, slave data from the 1840 U.S. Census and education and wealth data of children and grandchildren up to the 1880 U.S. Census.
- Those who do not appear in the list of winners were the losers.
- More general point: importance of micro data and computer power.

# The sample

- Bleakley and Ferrie select all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period (14,306 individuals).
- 1,758 linked to a winner in Smith's list and 1,177 to more than one individual.
- 14,963 male children and 40,658 grandchildren observed in 1850 and 1880.
- 12,235 male children observed in 1870 (wealth data).
- Matching names and surnames requires some care with spelling.



## Checks

- Balancing tests.
- Favorably checked with Columbia and Oglethorpe Counties, where there are actual lists of both lottery participants and lottery winners.
- Placebo analysis with residents from South Carolina and from using all Georgia's census.
- Also, winners did not move to statistically different counties.

| | (1) Whole Sample | (2) Lottery "Losers" | (3) Lottery "Winners" | (4) $p$-Value, Mean Difference [$N$] |
|---|---|---|---|---|
| Panel A: Lottery winner or loser | | | | |
| Dummy for unique match to Smith (1838) list | 0.124 (0.329) | 0 | 1 | — |
| Dummy for match to Smith (1838), deflated to $\frac{1}{n}$ in case of ties | 0.155 (0.335) | 0.037 (0.121) | 0.995 (0.053) | 0.000 [14,375] |
| Panel B: Predetermined outcomes | | | | |
| Age, in years | 51.2 (8.5) | 51.3 (8.5) | 50.9 (8.6) | 0.122 [14,375] |
| Born in Georgia | 0.497 (0.500) | 0.497 (0.500) | 0.498 (0.500) | 0.889 [14,375] |
| Born in South Carolina | 0.212 (0.408) | 0.210 (0.407) | 0.222 (0.416) | 0.263 [14,375] |
| Born in North Carolina | 0.180 (0.384) | 0.180 (0.384) | 0.178 (0.383) | 0.804 [14,375] |
| Number of Georgia-born children in the three years prior to the lottery | 1.333 (0.542) | 1.333 (0.541) | 1.332 (0.542) | 0.910 [14,375] |
| Cannot read and write | 0.147 (0.354) | 0.147 (0.354) | 0.142 (0.350) | 0.593 [14,340] |
| Number of letters in surname | 6.19 (1.61) | 6.20 (1.62) | 6.13 (1.51) | 0.072 [14,375] |
| Frequency with which surname appears in sample | 36.2 (46.3) | 36.3 (46.9) | 35.3 (41.9) | 0.380 [14,375] |
| Surname begins with "M" or "O" | 0.101 (0.302) | 0.101 (0.301) | 0.104 (0.305) | 0.740 [14,375] |
| Mean wealth of families in the South with same surname | 1,203.4 (445.4) | 1,204.5 (455.1) | 1,195.7 (370.5) | 0.373 [14,093] |
| Median wealth of families in the South with same surname | 184.1 (162.3) | 184.6 (167.3) | 181.0 (121.8) | 0.276 [14,093] |
| Mean illiteracy of adults in the South with same surname | 0.175 (0.043) | 0.175 (0.044) | 0.176 (0.039) | 0.124 [14,093] |
| Mean school attendance of children in the South with same surname | 0.323 (0.052) | 0.323 (0.052) | 0.323 (0.049) | 0.998 [13,975] |
| Panel C: Fertility and school attendance | | | | |
| Number of children in household born after the 1832 lottery | 3.955 (2.546) | 3.930 (2.539) | 4.135 (2.586) | 0.002 [14,375] |
| School attendance among children aged 5–17, inclusive | 0.342 (0.474) | 0.342 (0.475) | 0.341 (0.474) | 0.799 [47,749] |

| | (1) Whole Sample | (2) Lottery "Losers" | (3) Lottery "Winners" | (4) $p$-Value, Mean Difference [$N$] |
|---|---|---|---|---|
| Panel D: Other outcomes | | | | |
| Spouse cannot read and write | 0.235 | 0.236 | 0.231 | 0.676 |
| | (0.424) | (0.424) | (0.421) | [11,563] |
| Resides in Georgia | 0.723 | 0.722 | 0.729 | 0.548 |
| | (0.447) | (0.448) | (0.445) | [14,375] |
| Resides in Alabama | 0.144 | 0.144 | 0.145 | 0.935 |
| | (0.351) | (0.351) | (0.352) | [14,375] |
| Resides in Old Cherokee | 0.113 | 0.111 | 0.126 | 0.074 |
| County | (0.317) | (0.314) | (0.332) | [14,375] |
| Panel E: Measures of wealth in 1850 (18 years after the lottery) | | | | |
| Real estate wealth | 1,999.0 | 1,970.8 | 2,198.2 | 0.068 |
| | (4,694.2) | (4,422.0) | (6,290.1) | [13,094] |
| Slave wealth | 1,339.1 | 1,297.3 | 1,635.3 | 0.021 |
| | (5,761.0) | (5,329.7) | (8,189.0) | [14,375] |
| Total wealth (sum of wealth in | 3,323.7 | 3,245.5 | 3,876.5 | 0.006 |
| real estate and slaves) | (8,691.0) | (7,952.9) | (12,734.4) | [13,094] |
| | {100,800, 3,000} | {100,800, 3,000} | {100,1,000, 3,550} | |
| Panel F: Select variables for those with below $300 in 1850 total wealth | | | | |
| Number of children in | 3.905 | 3.878 | 4.098 | 0.063 |
| household born after | (2.471) | (2.453) | (2.591) | [4,506] |
| the 1832 lottery | | | | |
| Number of slaves in 1840 | 1.4 | 1.3 | 2.3 | 0.074 |
| | (6.7) | (6.6) | (7.4) | [1,761] |
| Has at least one slave in 1840 | 0.190 | 0.179 | 0.255 | 0.012 |
| | (0.392) | (0.384) | (0.437) | [1,761] |

## The regression

- Regression:

$$y_{ij} = \gamma T_j + \delta_{ai} + \beta \mathbf{x}_{ij} + \varepsilon_{ij}$$

- $i$: individual.
- $j$: lottery-elegible person.
- $y_{ij}$: outcome. If measured outcome is for elegible person, $i = j$.
- $T_j$: treatment dummy.
- $\delta_{ai}$: age dummies.
- $\mathbf{x}_{ij}$: controls.
- Some specifications with fixed-effects on surnames.

Effects of Lottery Winning on Fertility and School Attendance, 1850 Census

| | | | Additional Fixed Effects or Alternative Estimators: | | | | |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Specification: | Match to List of Winners: | None | Given Name | State of Residence | State and County of Residence | Urban Residence | Poisson (A) and Logit (B) |
| Panel A: Post-1832 fertility of lottery-eligible men [$N$ = 14,306] | | | | | | | |
| Basic | Binary | 0.132 | 0.146 | 0.124 | 0.102 | 0.126 | 0.033 |
| | | (0.058)** | (0.061)** | (0.058)** | (0.059)* | (0.058)** | (0.014)** |
| | $\frac{1}{n}$ | 0.137 | 0.156 | 0.128 | 0.104 | 0.130 | 0.034 |
| | | (0.056)** | (0.060)*** | (0.056)** | (0.058)* | (0.056)** | (0.014)** |
| Surname | Binary | 0.184 | 0.137 | 0.106 | 0.090 | 0.089 | 0.028 |
| | | (0.073)** | (0.069)** | (0.065) | (0.067) | (0.066) | (0.016)* |
| | $\frac{1}{n}$ | 0.175 | 0.131 | 0.095 | 0.075 | 0.074 | 0.026 |
| | | (0.072)** | (0.068)* | (0.064) | (0.065) | (0.065) | (0.016) |
| Panel B: School attendance of children aged 5–17 [$N$ = 47,749] | | | | | | | |
| Basic | Binary | −0.005 | −0.002 | −0.005 | 0.001 | −0.004 | −0.021 |
| | | (0.011) | (0.011) | (0.011) | (0.011) | (0.011) | (0.051) |
| | $\frac{1}{n}$ | −0.004 | 0.000 | −0.004 | 0.004 | −0.003 | −0.017 |
| | | (0.011) | (0.011) | (0.011) | (0.010) | (0.011) | (0.050) |
| Surname | Binary | −0.005 | 0.000 | −0.005 | 0.002 | −0.004 | −0.017 |
| | | (0.011) | (0.012) | (0.011) | (0.011) | (0.011) | (0.033) |
| | $\frac{1}{n}$ | −0.006 | −0.002 | −0.006 | 0.004 | −0.005 | −0.023 |
| | | (0.011) | (0.011) | (0.011) | (0.011) | (0.011) | (0.033) |

## Outcomes of Grandchildren of Lottery-Eligible Men in 1880

| Match to List of Winners: | Grandchildren: Unable to Read and Write (1) | Enrolled in School (2) | Number Children under 10 (3) | Number Children under 18 (4) |
|---|---|---|---|---|
| 1. Estimates of the effect of grandfather winning the lottery | | | | |
| Panel A: Basic specification | | | | |
| Binary | −0.004 | −0.021 | −0.055 | −0.097 |
| | (0.014) | (0.012)* | (0.041) | (0.060) |
| $\frac{1}{n}$ | 0.003 | −0.012 | −0.059 | −0.089 |
| | (0.013) | (0.012) | (0.041) | (0.059) |
| Panel B: Control for surname fixed effects | | | | |
| Binary | −0.006 | −0.026 | −0.044 | −0.086 |
| | (0.014) | (0.013)** | (0.046) | (0.066) |
| $\frac{1}{n}$ | 0.001 | −0.020 | −0.051 | −0.076 |
| | (0.014) | (0.013) | (0.046) | (0.066) |
| Panel C: Control for surname effects and length of given name | | | | |
| Binary | −0.005 | −0.032 | −0.049 | −0.120 |
| | (0.016) | (0.014)** | (0.056) | (0.079) |
| $\frac{1}{n}$ | 0.007 | −0.024 | −0.055 | −0.099 |
| | (0.016) | (0.014)* | (0.055) | (0.079) |
| 2. Estimation sample | | | | |
| | Children in 1880, ages 10–19 [$N$ = 23,544] | Children in 1880, ages 5–19 [$N$ = 40,658] | 1850 children as adults in 1880 [$N$ = 14,963] | 1850 children as adults in 1880 [$N$ = 14,963] |

## Limitations of randomized field trials

- Often, RTFs are called the “gold standard” for testing theories.
- While, RTFs are extremely useful, calling them the “gold standard” is too optimistic (and unfair to other empirical methods, both experimental and non-experimental).
- Difficult to integrate with previous knowledge.
- Daniel T. Campbell (1916-1996): internal and external validity.

# Internal validity

- Heterogeneity of effects among different agents. Role of outliers.
- Decompose effects of different elements of a treatment.
- Low take-up rates.
- Attrition, compliance, and contamination.
- Hawthorne effect: randomization changes how a treatment works.
- John Henry effect: members of the control group change their behavior as consequence of the perceived disadvantage of being in the control group. Also related to substitution toward other treatments.
- Pioneer effects.
- Long-term effects (Price and Song, 2016).

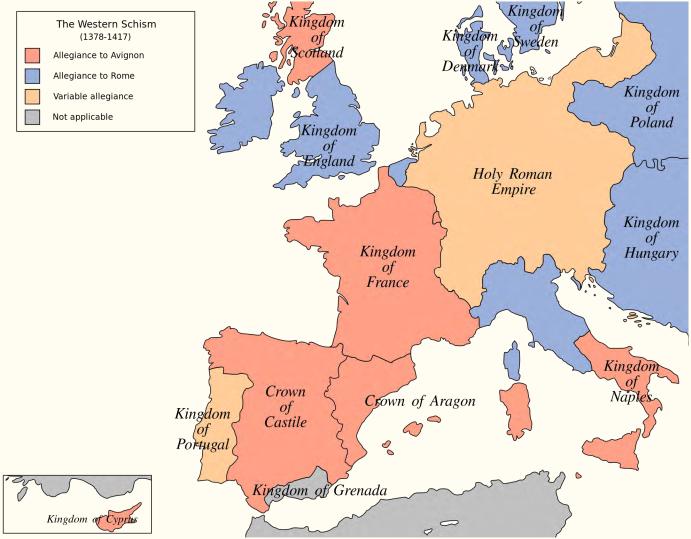
# External validity

- Replicability with other populations (meta-analysis).
- Ideal vs. realistic conditions.
- Randomization bias (getting into an RTF is already evidence of something. Ashenfelter, 1981 ).
- Local spillovers.
- General equilibrium (stable unit treatment value assumption or SUTVA).

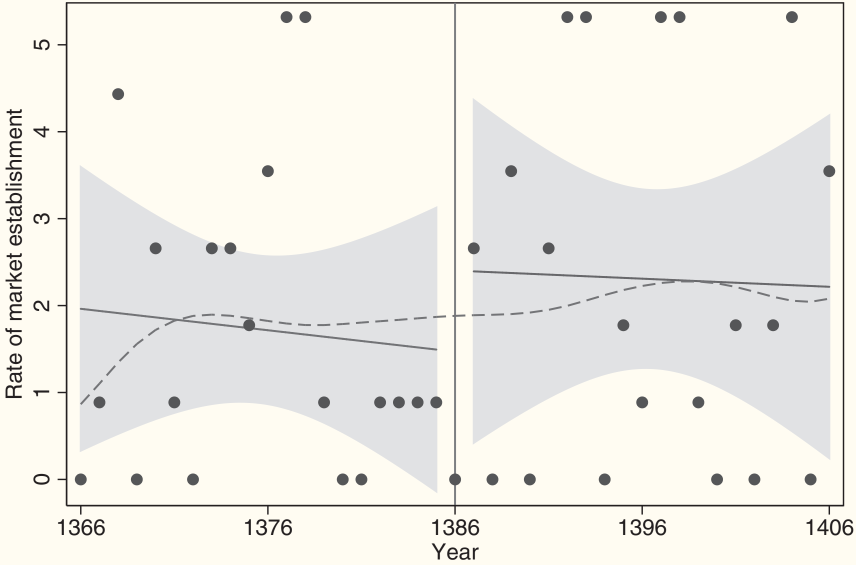
## Quasi-randomization

- Also sometimes known as natural experiments.
- Product of a random event, historical accident, geographical feature,...
- Not fully random, but if sufficiently uncorrelated with possibly omitted variables, they can be a good approximation.
- In particular, if we can use additional controls while we estimate the causal effects.
- However, it requires a judgment call to assess whether selection bias is being avoided.
- Example: Medieval Universities, Legal Institutions, and the Commercial Revolution by Davide Cantoni and Noam Yuchtman (2014).

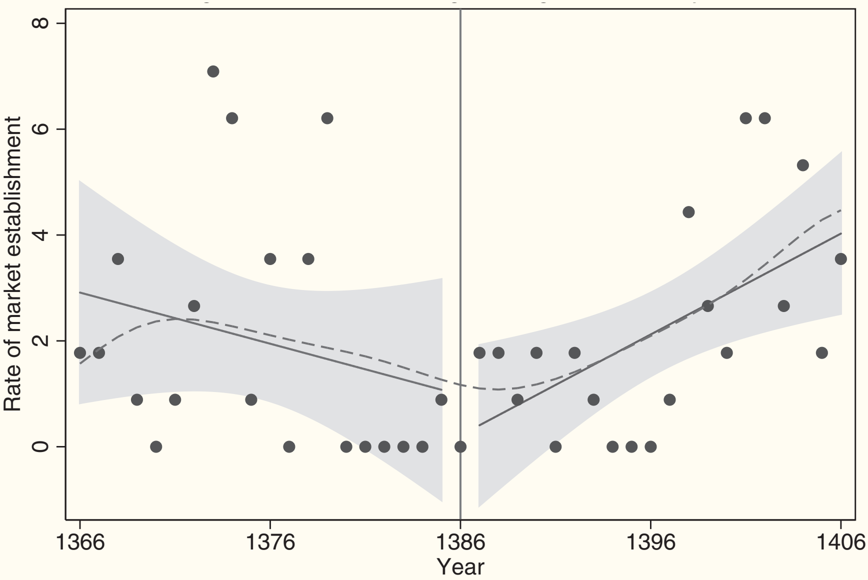
# Western Schism, 1378-1417



# Cities with a small change in distance to universities



# Cities with a large change in distance to universities

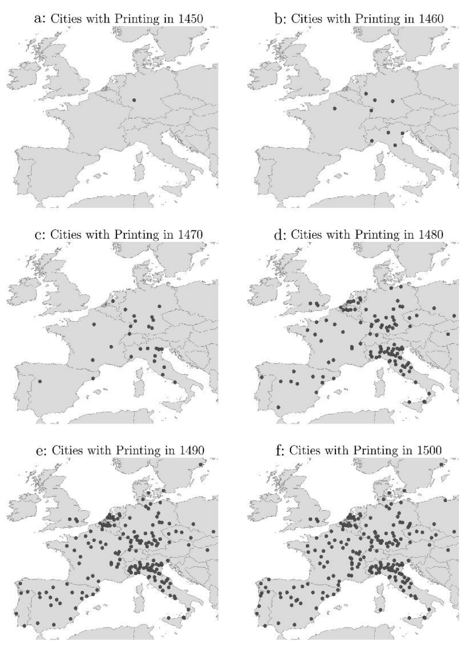


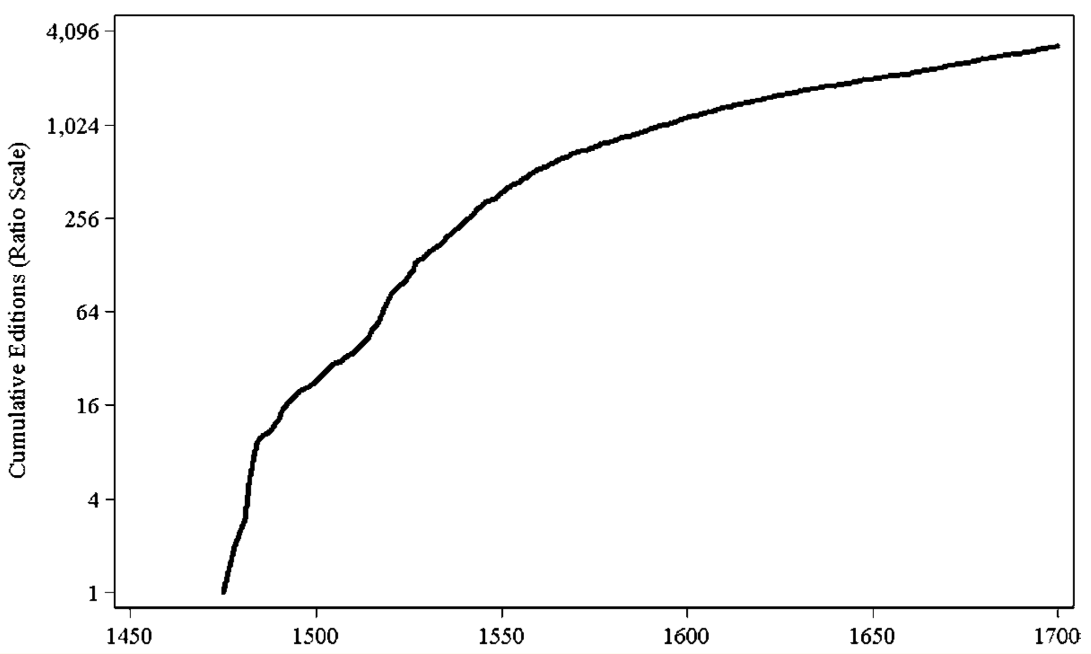
## Regression

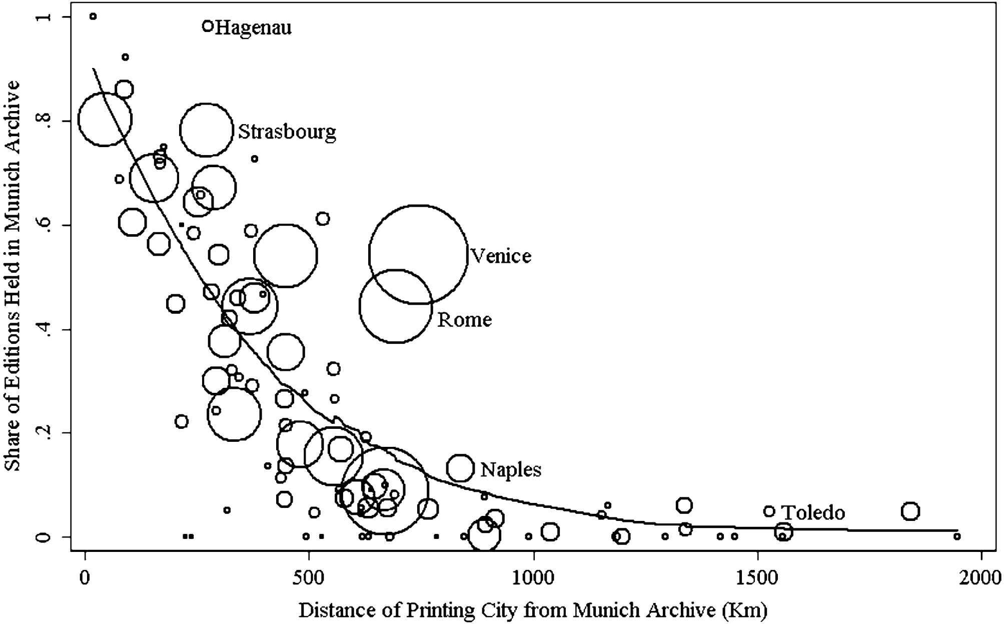
- Selection bias does not necessarily need to be unobservable.
- Additional controls in our regression may remove much (most?) of the bias.
- Difficulty in selecting regressors.
- In economic history, it often requires clustering errors (and possibly bootstrapping).

## A question: the impact of the printing press

- Printing press is one of the most important inventions in history.
- No doubt about its impact in cultural life (i.e., Protestant reformation, middle class literacy, ...).
- What was the impact of the printing press on economic growth?
- Information Technology and Economic Change: The Impact of the Printing Press by Jeremiah E. Dittmar (2011).
- Take advantage of a historical accident: when was the printing press introduced in a town?


a: Cities with Printing in 1450
b: Cities with Printing in 1460
c: Cities with Printing in 1470
d: Cities with Printing in 1480
e: Cities with Printing in 1490
f: Cities with Printing in 1500

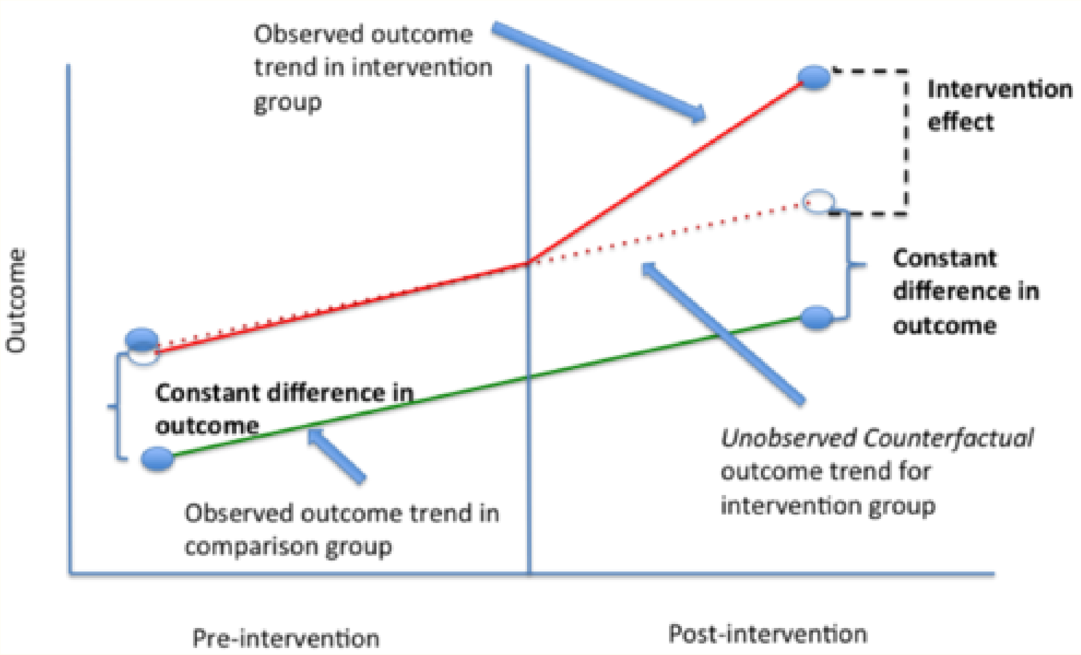

Cumulative Editions (Ratio Scale)
4,096
1,024
256
64
16
4
1
1450
1500
1550
1600
1650
1700


Hagenau
Strasbourg
Venice
Rome
Naples
Toledo
1
.8
.6
.4
.2
0
0
500
1000
1500
2000
Share of Editions Held in Munich Archive
Distance of Printing City from Munich Archive (Km)

| | Dependent Variable Is Log City Growth | | | |
|---|---|---|---|---|
| | Pre-Adoption | Post-Adoption | | |
| (1) | (2) | (3) | (4) | (5) |
| Independent Variable | Growth 1400–1500 | Growth 1500–1600 | Growth 1500–1700 | Growth 1500–1800 |
| Print Adoption 1450–1500 | 0.07 | 0.19*** | 0.26*** | 0.30*** |
| | (0.08) | (0.06) | (0.08) | (0.09) |
| Editions Per Capita | 0.03 | 0.03* | 0.04 | 0.05 |
| | (0.03) | (0.02) | (0.03) | (0.03) |
| University | −0.12 | 0.02 | 0.17* | 0.17* |
| | (0.11) | (0.07) | (0.09) | (0.09) |
| Roman Site | 0.08 | −0.01 | 0.09 | 0.04 |
| | (0.06) | (0.05) | (0.08) | (0.07) |
| Capital | 0.31** | 0.95*** | 1.46*** | 1.98*** |
| | (0.13) | (0.16) | (0.20) | (0.27) |
| Freedom Index | −0.23 | 0.27*** | 0.29** | −0.07 |
| | (0.14) | (0.10) | (0.13) | (0.14) |
| Atlantic Port | 0.16 | 0.34*** | 0.64*** | 0.76*** |
| | (0.18) | (0.09) | (0.14) | (0.12) |
| Mediterranean Port | 0.21* | 0.15 | 0.57*** | 0.65*** |
| | (0.13) | (0.12) | (0.15) | (0.17) |
| Baltic Port | −0.16 | 0.25** | 0.55** | 0.37 |
| | (0.18) | (0.12) | (0.22) | (0.24) |
| Navigable River | 0.14* | 0.18*** | 0.23*** | 0.39*** |
| | (0.08) | (0.06) | (0.09) | (0.09) |
| Log Population | −0.22*** | −0.30*** | −0.42*** | −0.64*** |
| | (0.04) | (0.04) | (0.05) | (0.05) |
| Country FE | Yes | Yes | Yes | Yes |
| Observations | 291 | 495 | 515 | 622 |
| *R* Squared | 0.33 | 0.32 | 0.35 | 0.47 |

# Difference-in-differences

- OLS is often applied in combination with difference-in-differences.
- What is difference-in-differences?
- Parallel trend assumption.
- Assumption difficult to verify. One can use pre-treatment data to show that the trends were the same.


Observed outcome trend in intervention group
Intervention effect
Outcome
Constant difference in outcome
Constant difference in outcome
Unobserved Counterfactual outcome trend for intervention group
Observed outcome trend in comparison group
Pre-intervention
Post-intervention

## Practical implementation

- It is important to be careful with standard errors (Bertrand, Dufflo, and Mullainathan, 2004):
    1. Block bootstrapping standard errors.
    2. Clustering standard errors at the group level.
- In our example of the printing press:

$$Y_{i,t} = \theta_i + \delta_t + \sum_{t=1300}^{1700} \alpha_t D_t T_i + X'_{i,t}\gamma + \varepsilon_{i,t}$$

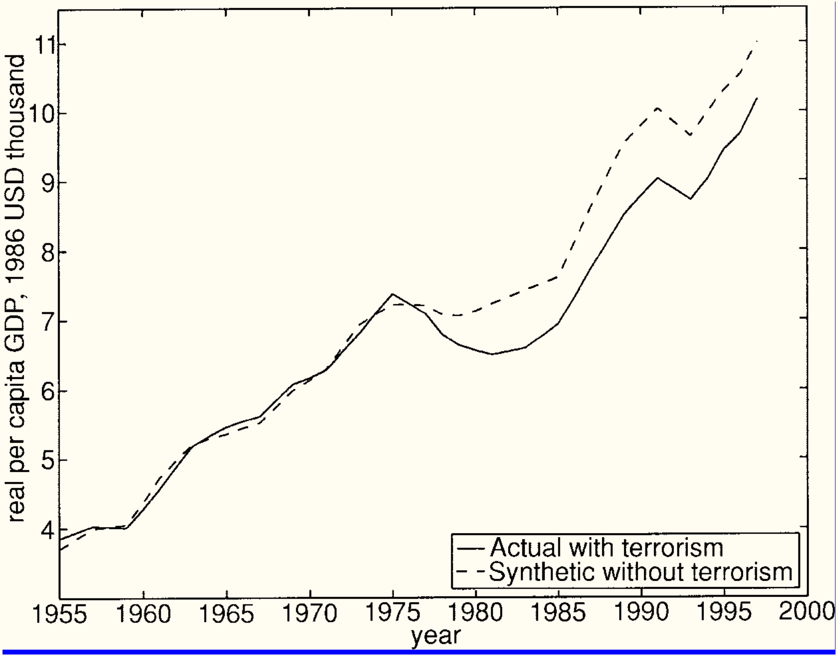
Log City Growth: The Timing of the Print Advantage

| (1) Variable | (2) All Cities Balanced Sample | (3) Exclude German Cities | (4) Exclude Italian & Dutch Cities | (5) Exclude If East of Elbe River | (6) Only Port Cities | (7) Only Cities Without Ports |
|---|---|---|---|---|---|---|
| Print × Yr1400 | 0.09 | 0.10 | 0.09 | 0.11 | 0.27 | −0.04 |
| | (0.16) | (0.18) | (0.20) | (0.17) | (0.38) | (0.16) |
| Print × Yr1500 | 0.34** | 0.39** | 0.41** | 0.34** | 1.39*** | 0.10 |
| | (0.15) | (0.17) | (0.18) | (0.16) | (0.42) | (0.15) |
| Print × Yr1600 | 0.13 | 0.22 | 0.08 | 0.16 | 0.73** | −0.01 |
| | (0.16) | (0.17) | (0.20) | (0.16) | (0.34) | (0.17) |
| Print × Yr1700 | 0.19 | 0.25 | 0.16 | 0.22 | 0.84** | 0.00 |
| | (0.14) | (0.16) | (0.17) | (0.14) | (0.42) | (0.15) |
| Atlantic × Yr1400 | 0.12 | 0.27 | 0.13 | 0.12 | −0.32 | — |
| | (0.31) | (0.33) | (0.37) | (0.31) | (0.52) | — |
| Atlantic × Yr1500 | 0.43* | 0.55** | 0.38 | 0.44* | −0.24 | — |
| | (0.25) | (0.28) | (0.28) | (0.25) | (0.52) | — |
| Atlantic × Yr1600 | 0.42* | 0.49* | 0.33 | 0.45** | 0.47 | — |
| | (0.22) | (0.25) | (0.24) | (0.22) | (0.38) | — |
| Atlantic × Yr1700 | 0.60*** | 0.73*** | 0.64*** | 0.62*** | 0.32 | — |
| | (0.19) | (0.20) | (0.21) | (0.19) | (0.38) | — |
| *R* squared | 0.55 | 0.57 | 0.58 | 0.54 | 0.77 | 0.53 |
| Observations | 1,010 | 875 | 710 | 850 | 225 | 785 |
| Adopting Cities | 83 | 71 | 53 | 78 | 16 | 67 |
| Nonadopting Cities | 119 | 104 | 89 | 92 | 29 | 90 |

## Synthetic controls

- If we do not have a good comparison, we can build synthetic controls.
- The Economic Costs of Conflict: A Case Study of the Basque Country by Abadie and Gardeazabal (2003).
- What were the effects of terrorism?
- Problem: Basque country was, in economic terms, quite different than the rest of Spain.
- We can re-weight the other 16 Spanish regions to create a country that resembles the Basque country as much as possible.
- Optimal weights were Catalonia: 0.8508, Madrid: 0.1492, and all other regions: 0.

| | Basque Country (1) | Spain (2) | "Synthetic" Basque Country (3) |
|---|---|---|---|
| Real per capita GDP[a] | 5,285.46 | 3,633.25 | 5,270.80 |
| Investment ratio (percentage)[b] | 24.65 | 21.79 | 21.58 |
| Population density[c] | 246.89 | 66.34 | 196.28 |
| Sectoral shares (percentage)[d] | | | |
| Agriculture, forestry, and fishing | 6.84 | 16.34 | 6.18 |
| Energy and water | 4.11 | 4.32 | 2.76 |
| Industry | 45.08 | 26.60 | 37.64 |
| Construction and engineering | 6.15 | 7.25 | 6.96 |
| Marketable services | 33.75 | 38.53 | 41.10 |
| Nonmarketable services | 4.07 | 6.97 | 5.37 |
| Human capital (percentage)[e] | | | |
| Illiterates | 3.32 | 11.66 | 7.65 |
| Primary or without studies | 85.97 | 80.15 | 82.33 |
| High school | 7.46 | 5.49 | 6.92 |
| More than high school | 3.26 | 2.70 | 3.10 |



real per capita GDP, 1986 USD thousand
11
10
9
8
7
6
5
4
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
year
Actual with terrorism
Synthetic without terrorism

# Instrumental variables I

- Previous analysis has a problem: what if the printing press was introduced in cities that were poised to grow? (rhetorical question: Dittmar is actually careful about this).

- Instrumental variables (IVs) is one of the most popular techniques in applied empirical analysis.

- Wright (1928) and Reiersol (1941).

- IV use a variable $z$ that predicts independent variable $x$ in regression of interest, but it is uncorrelated with dependent variable $y$ to produce quasi-experimental variation in $x$.

- IVs can be rigorously shown through a GMM (generalized method of moments) argument.

# Instrumental variables II

- Imagine that we have a linear regression:

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

  but $Cov(x, \varepsilon) \neq 0$. Thus, standard OLS estimate is biased.

- However, we have a variable $z$ such that

$$\begin{aligned} Cov(x, z) &\neq 0 \\ Cov(z, \varepsilon) &= 0 \end{aligned}$$

- The first assumption is testable. The second one is not.

## Why do IVs work?

- We can find:

$$\begin{aligned} Cov(y,z) &= Cov(\beta_0 + \beta_1 x + \varepsilon, z) \\ &= Cov(\beta_0, z) + \beta_1 Cov(x,z) + Cov(\varepsilon, z) \end{aligned}$$

- Therefore:

$$\beta_1 = \frac{Cov(y,z)}{Cov(x,z)}$$

- First stage vs. second stage.

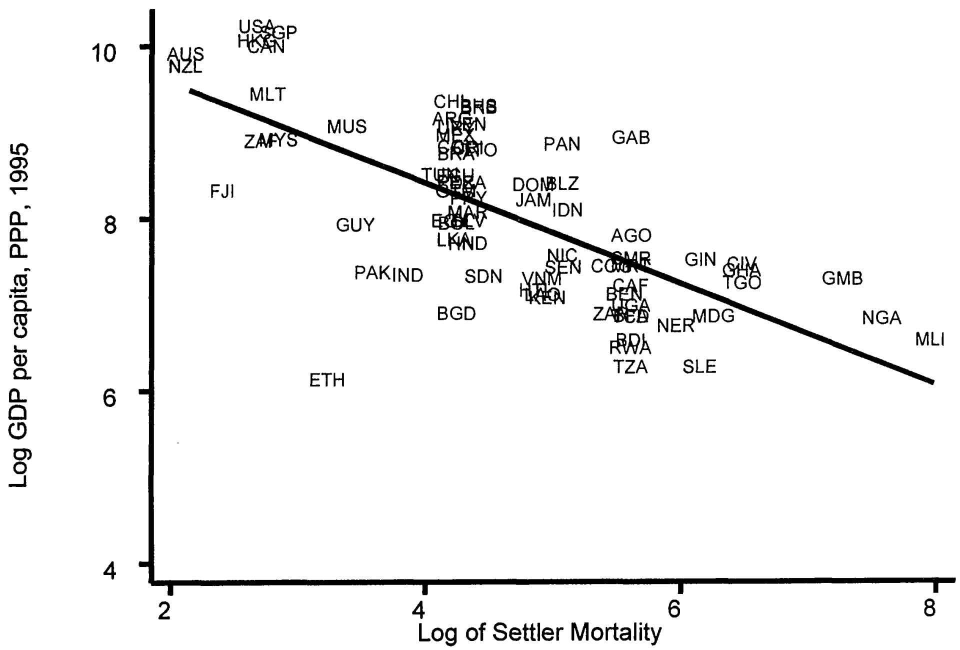

Log GDP per capita, PPP, 1995
Log of Settler Mortality
10
8
6
4
2
4
6
8
AUS
NZL
MLT
MUS
FJI
GUY
PAK
IND
ETH
SDN
BGD
PAN
GAB
JAM
IDN
AGO
GIN
GMB
TGO
NER
MDG
NGA
MLI
TZA
SLE
RWA

TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

| | Base sample (1) | Base sample (2) | Base sample without Neo-Europes (3) | Base sample without Neo-Europes (4) | Base sample without Africa (5) | Base sample without Africa (6) | Base sample with continent dummies (7) | Base sample with continent dummies (8) | Base sample, dependent variable is log output per worker (9) |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: Two-Stage Least Squares | | | | | | | | | |
| Average protection against expropriation risk 1985–1995 | 0.94 | 1.00 | 1.28 | 1.21 | 0.58 | 0.58 | 0.98 | 1.10 | 0.98 |
| | (0.16) | (0.22) | (0.36) | (0.35) | (0.10) | (0.12) | (0.30) | (0.46) | (0.17) |
| Latitude | | −0.65 | | 0.94 | | 0.04 | | −1.20 | |
| | | (1.34) | | (1.46) | | (0.84) | | (1.8) | |
| Asia dummy | | | | | | | −0.92 | −1.10 | |
| | | | | | | | (0.40) | (0.52) | |
| Africa dummy | | | | | | | −0.46 | −0.44 | |
| | | | | | | | (0.36) | (0.42) | |
| "Other" continent dummy | | | | | | | −0.94 | −0.99 | |
| | | | | | | | (0.85) | (1.0) | |
| Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995 | | | | | | | | | |
| Log European settler mortality | −0.61 | −0.51 | −0.39 | −0.39 | −1.20 | −1.10 | −0.43 | −0.34 | −0.63 |
| | (0.13) | (0.14) | (0.13) | (0.14) | (0.22) | (0.24) | (0.17) | (0.18) | (0.13) |
| Latitude | | 2.00 | | −0.11 | | 0.99 | | 2.00 | |
| | | (1.34) | | (1.50) | | (1.43) | | (1.40) | |
| Asia dummy | | | | | | | 0.33 | 0.47 | |
| | | | | | | | (0.49) | (0.50) | |
| Africa dummy | | | | | | | −0.27 | −0.26 | |
| | | | | | | | (0.41) | (0.41) | |
| "Other" continent dummy | | | | | | | 1.24 | 1.1 | |
| | | | | | | | (0.84) | (0.84) | |
| $R^2$ | 0.27 | 0.30 | 0.13 | 0.13 | 0.47 | 0.47 | 0.30 | 0.33 | 0.28 |
| Panel C: Ordinary Least Squares | | | | | | | | | |
| Average protection against expropriation risk 1985–1995 | 0.52 | 0.47 | 0.49 | 0.47 | 0.48 | 0.47 | 0.42 | 0.40 | 0.46 |
| | (0.06) | (0.06) | (0.08) | (0.07) | (0.07) | (0.07) | (0.06) | (0.06) | (0.06) |
| Number of observations | 64 | 64 | 60 | 60 | 37 | 37 | 64 | 64 | 61 |

# Regression discontinuity design

- We take advantage of a sudden change (threshold effect) on a treatment.
- Treatments are often somewhat arbitrary.
- Proposed by Thistlewaithe and Campbell (1960).
- Key: precise knowledge of the rules determining treatment and willingness to extrapolate across covariates locally.
- Often called RDD.

# Sharp and fuzzy RDD

- Sharp regression discontinuity (RD):

$$D_i = \begin{cases} 1 \text{ if } x_i \geq x_0 \\ 0 \text{ if } x_i < x_0 \end{cases}$$

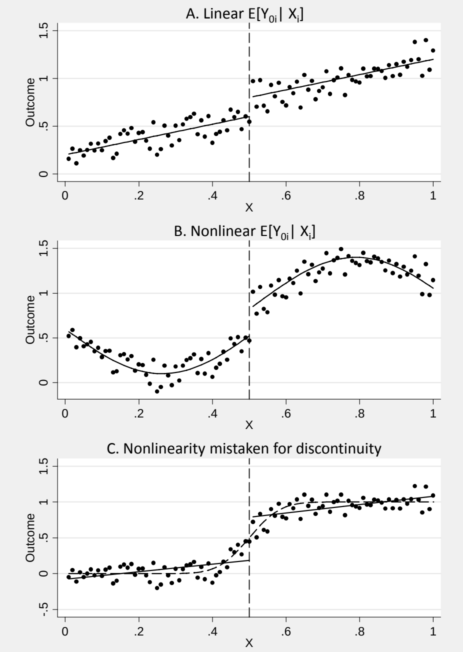
- Original example: merit scholarships.

- Fuzzy regression discontinuity (RD):

$$P\left(D_i = 1 | x_i\right) = \begin{cases} g_1(x_i) \text{ if } x_i \geq x_0 \\ g_0(x_i) \text{ if } x_i < x_0 \end{cases}$$

- Sharp RDD is a selection-in-observables while fuzzy RDD is an IV.

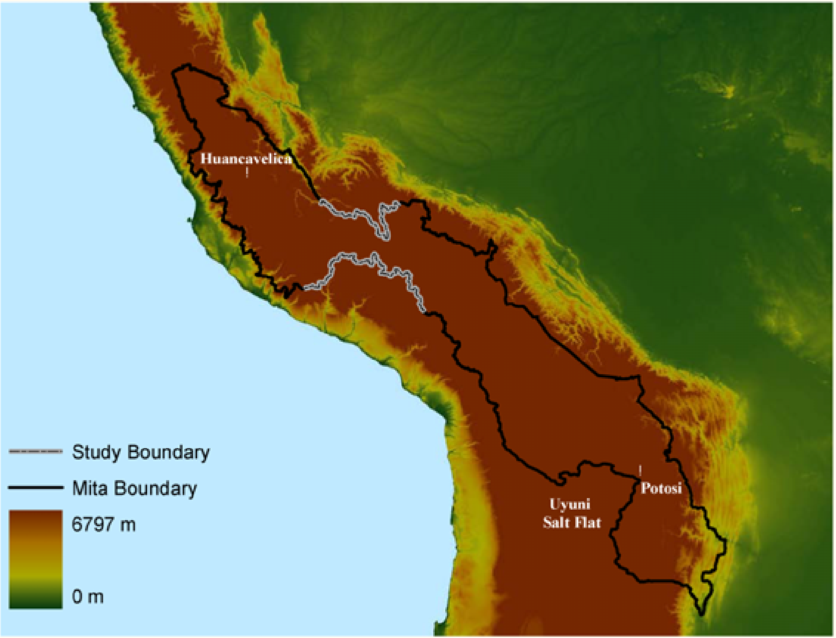
- Nonparametric vs. parametric specifications of distance to threshold.

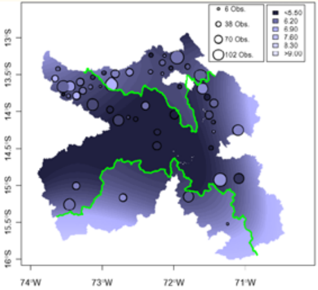

A. Linear E[Y0i | Xi]
Outcome
X
B. Nonlinear E[Y0i | Xi]
Outcome
X
C. Nonlinearity mistaken for discontinuity
Outcome
X

# A question: the impact of the Mita

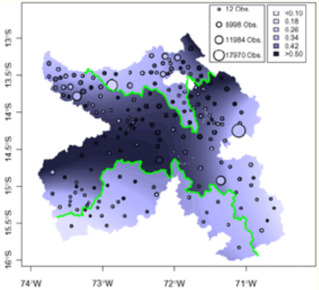
- *Mita*: an extensive forced mining labor system in current-day Peru and Bolivia between 1573 and 1812.
- What is its legacy?
- The Persistent Effects of Peru's Mining Mita by Dell (2010).
- RDD:

$$c_{ibd} = \alpha + \gamma * mita_d + X'_{ib}\beta + f(\text{geographical location}_d) + \phi_b + \varepsilon_{ibd}$$

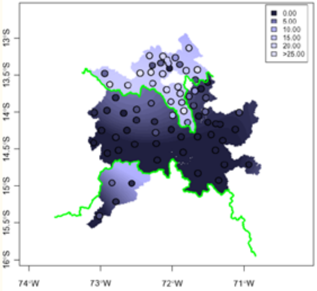

Huancavelica
Potosi
Uyuni
Salt Flat
Study Boundary
Mita Boundary
6797 m
0 m



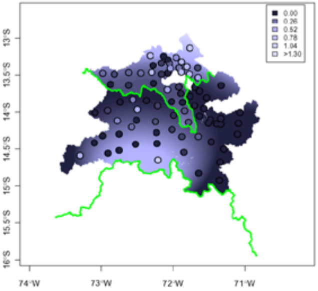

(a) Consumption (2001)



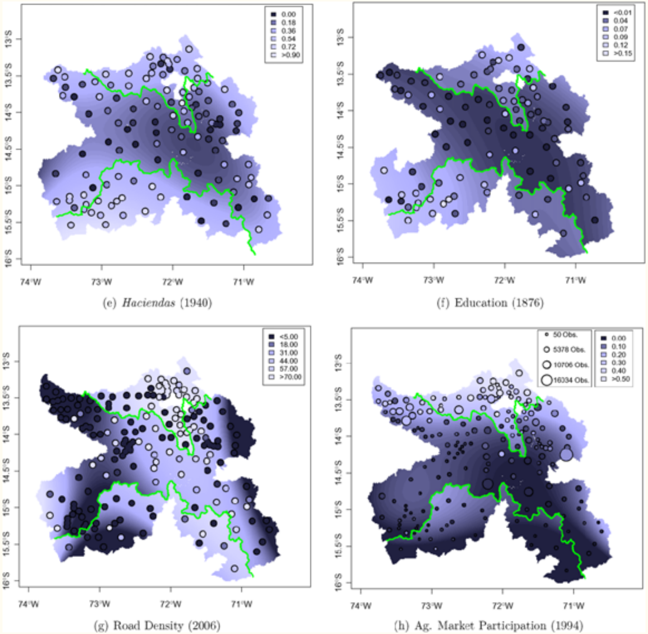

(b) Stunting (2005)




(c) *Haciendas* (1689)




(d) *Haciendas* (1845)




(e) *Haciendas* (1940)

(f) Education (1876)

(g) Road Density (2006)

(h) Ag. Market Participation (1994)

| | Dependent Variable | | | | | | |
|---|---|---|---|---|---|---|---|
| | Log Equiv. Hausehold Consumption (2001) | | | Stunted Growth, Children 6–9 (2005) | | | |
| Sample Within: | <100 km of Bound. | <75 km of Bound. | <50 km of Bound. | <100 km of Bound. | <75 km of Bound. | <50 km of Bound. | Border District |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Panel A. Cubic Polynomial in Latitude and Longitude | | | | | | | |
| *Mita* | −0.284 | −0.216 | −0.331 | 0.070 | 0.084* | 0.087* | 0.114** |
| | (0.198) | (0.207) | (0.219) | (0.043) | (0.046) | (0.048) | (0.049) |
| $R^2$ | 0.060 | 0.060 | 0.069 | 0.051 | 0.020 | 0.017 | 0.050 |
| Panel B. Cubic Polynomial in Distance to Potosí | | | | | | | |
| *Mita* | −0.337*** | −0.307*** | −0.329*** | 0.080*** | 0.078*** | 0.078*** | 0.063* |
| | (0.087) | (0.101) | (0.096) | (0.021) | (0.022) | (0.024) | (0.032) |
| $R^2$ | 0.046 | 0.036 | 0.047 | 0.049 | 0.017 | 0.013 | 0.047 |
| Panel C. Cubic Polynomial in Distance to *Mita* Boundary | | | | | | | |
| *Mita* | −0.277*** | −0.230** | −0.224** | 0.073*** | 0.061*** | 0.064*** | 0.055* |
| | (0.078) | (0.089) | (0.092) | (0.023) | (0.022) | (0.023) | (0.030) |
| $R^2$ | 0.044 | 0.042 | 0.040 | 0.040 | 0.015 | 0.013 | 0.043 |
| Geo. controls | yes | yes | yes | yes | yes | yes | yes |
| Boundary F.E.s | yes | yes | yes | yes | yes | yes | yes |
| Clusters | 71 | 60 | 52 | 289 | 239 | 185 | 63 |
| Observations | 1478 | 1161 | 1013 | 158,848 | 115,761 | 100,446 | 37,421 |

## RDD in economic history

- In economic history, political frontiers are particularly popular.
- Frontiers define different policies applied to often similar environments.
- But other RDDs are possible (time, cohort, ethnicity...).
- Generalization: regression kink design (Card, Lee, Pei and Weber, 2012).
- Often called RDD.

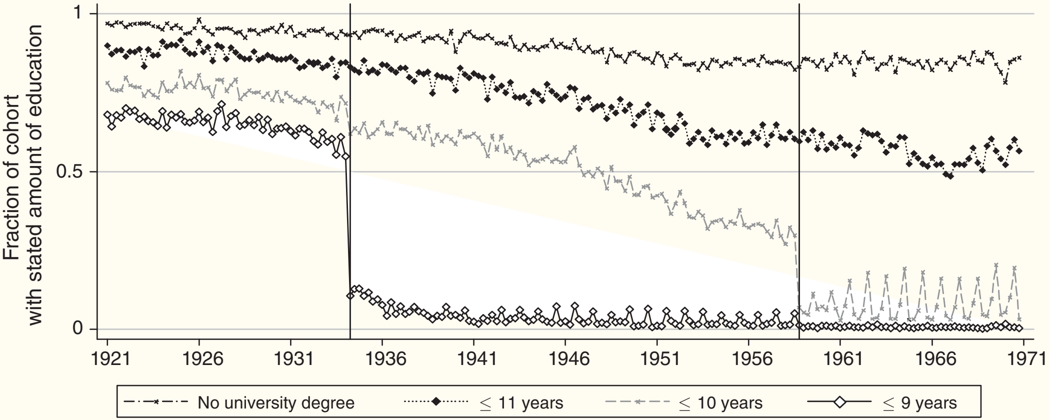



FIGURE 1. YEARS OF FULL-TIME EDUCATION BY QUARTER OF BIRTH

# Figure 3: Daily UI Benefits

Bottom Kink Sample


# Figure 4: Daily UI Benefits

Top Kink Sample

Figure 5: Log Time to Next Job

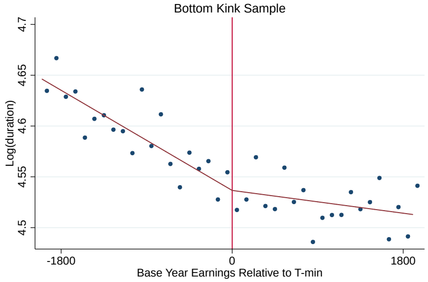
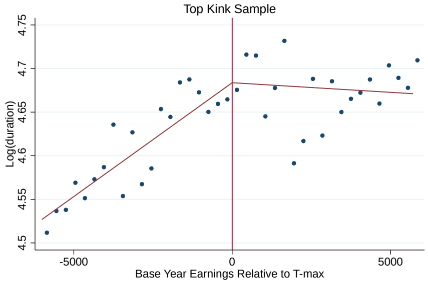



Figure 6: Log Time to Next Job



## Structural breaks in a time series

- Similar to RDD, but in a time series context.
- Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany by Fabian Waldinger (2010).
- Law for the Restoration of the Professional Civil Service on April 7, 1933.
- Effects concentrated among top professors.

# Reichsgesetzblatt

Teil I

Ausgegeben zu Berlin, den 7. April 1933

s zur Wiederherstellung des Berufsbeamtentums. Vom 7. April 1933. ..........

derherstellung des Berufsbeamtentums.
Vom 7. April 1933.

regierung hat das folgende Gesetz be-
hiermit verkündet wird:

§ 1

Viederherstellung eines nationalen Be-
ums und zur Vereinfachung der Ver-
ien Beamte nach Maßgabe der folgen-
ungen aus dem Amt entlassen werden,
ie nach dem geltenden Recht hierfür
Voraussetzungen nicht vorliegen.

eamte im Sinne dieses Gesetzes gelten
und mittelbare Beamte des Reichs,

des jeweiligen Grundgehalts der vo
bekleideten Stelle bewilligt werden;
sicherung nach Maßgabe der reichsgese
versicherung findet nicht statt.

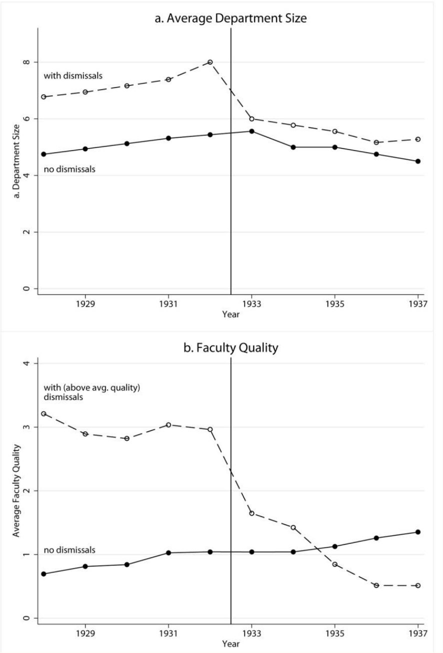
(4) Die Vorschriften der Abs. 2 un
Personen der im Abs. 1 bezeichneten
vor dem Inkrafttreten dieses Gesetzes
stand getreten sind, entsprechende An

§ 3

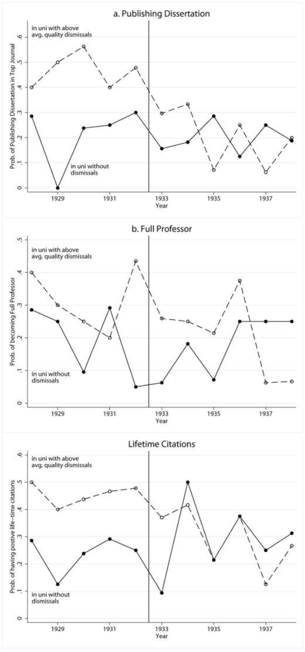
(1) Beamte, die nicht arischer Abs
sind in den Ruhestand (§§ 8ff.) zu v
es sich um Ehrenbeamte handelt, si
Amtsverhältnis zu entlassen.

TABLE 1
Number of Dismissed Mathematics Professors

| Year of Dismissal | Number of Dismissed Professors | Percentage of All Mathematics Professors in 1933 |
|---|---|---|
| 1933 | 35 | 15.6 |
| 1934 | 6 | 2.7 |
| 1935 | 5 | 2.2 |
| 1936 | 1 | .4 |
| 1937 | 2 | .9 |
| 1938 | 1 | .4 |
| 1939 | 1 | .4 |
| 1940 | 1 | .4 |
| 1933–34 | 41 | 18.3 |



a. Average Department Size
a. Department Size
with dismissals
no dismissals
0
2
4
6
8
1929
1931
1933
1935
1937
Year
b. Faculty Quality
Average Faculty Quality
with (above avg. quality) dismissals
no dismissals
0
1
2
3
4
1929
1931
1933
1935
1937
Year



a. Publishing Dissertation
in uni with above avg. quality dismissals
in uni without dismissals
Prob. of Publishing Dissertation in Top Journal
Year
b. Full Professor
in uni with above avg. quality dismissals
in uni without dismissals
Prob. of becoming Full Professor
Year
Lifetime Citations
in uni with above avg. quality dismissals
in uni without dismissals
Prob. of having postive life−time citations
Year

# Other methods

- Matching estimators (pure and propensity).
- Heckman's selection model.
- Quantile Regression.