Empirical Strategies in Economic History

Jesús Fernández-Villaverde

November 17, 2019

¹University 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?
Causality

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).
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).
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.
“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.”.

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?
To estimate the impact of institutions on economic performance, we need a source of exogenous variation in institutions. In this paper, we propose a theory of institutional differences among countries colonized by Europeans, and exploit this theory to derive a possible source of exogenous variation. Our theory rests on three premises:

1. There were different types of colonization policies which created different sets of institutions. At one extreme, European powers set up “extractive states,” exemplified by the Belgian colonization of the Congo. These institutions did not introduce much protection for private property, nor did they provide checks and balances against government expropriation. In fact, the main purpose of the extractive state was to transfer as much of the resources of the colony to the colonizer. At the other extreme, many Europeans migrated and settled in a number of colonies, creating what the historian Alfred Crosby (1986) calls “Neo-Europes.” The settlers tried to replicate European institutions, with strong emphasis on private property and checks against government power. Primary examples of this include Australia, New Zealand, Canada, and the United States.

2. The colonization strategy was influenced by the feasibility of settlements. In places where the disease environment was not favorable to European settlement, the cards were stacked against the creation of Neo-Europes, and the formation of the extractive state was more likely.

3. The colonial state and institutions persisted even after independence.

Based on these three premises, we use the mortality rates expected by the first European settlers in the colonies as an instrument for current institutions in these countries. More specifically, our theory can be schematically summarized as

\[(potential \text{ settler mortality} \implies \text{settlements}) \implies \text{early institutions} \implies \text{current institutions} \implies \text{current performance.}\]

We use data on the mortality rates of soldiers, bishops, and sailors stationed in the colonies between the seventeenth and nineteenth centuries, largely based on the work of the historian Philip D. Curtin. These give a good indication of the mortality rates faced by settlers. Europeans were well informed about these mortality rates at the time, even though they did not know how to control the diseases that caused these high mortality rates.

Figure 1 plots the logarithm of GDP per capita today against the logarithm of the settler mortality rates per thousand for a sample of 75 countries (see below for data details). It shows a strong negative relationship. Colonies where Europeans faced higher mortality rates are today substantially poorer than colonies that were healthy for Europeans. Our theory is that this relationship reflects the effect of settler mortality working through the institutions brought by Europeans. To substantiate this, we regress current performance on current institutions, and instrument the latter by settler mortality rates. Since our focus is on property rights and checks against government power, we use the protection against “risk of expropriation” index from Political Risk Services as a proxy for institutions. This variable measures differences in institutions originating from different types of states and state policies.

1 By “colonial experience” we do not only mean the direct control of the colonies by European powers, but more generally, European influence on the rest of the world. So according to this definition, Sub-Saharan Africa was strongly affected by “colonialism” between the sixteenth and nineteenth centuries because of the Atlantic slave trade.

2 Note that although only some countries were colonized, there is no selection bias here. This is because the question we are interested in is the effect of colonization policy conditional on being colonized.

3 Government expropriation is not the only institutional feature that matters. Our view is that there is a “cluster of
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 1: 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)

Overall impact on GDP growth

Although the benchmark revision may have a significant impact on GDP levels, the growth rates are generally affected to a much lesser extent. Figure 4 compares economic growth for the OECD average according to the old methodologies with those according to the latest estimates. Over the period 1992 to 2012, the difference is within the boundaries of +/- 0.1 %-point, sometimes marginally above. The only exception is for the year 2009, for which the latest estimates show a decline in GDP growth of -5.6%, or 0.3 %-points less negative than the estimate according to the old methodologies (-5.9%).

Overall impact on NNI levels

The level changes in Net National Income (NNI) are generally more moderate than the changes in the levels of GDP. The difference between the changes in these indicators can be explained by two factors: (i) NNI is adjusted for depreciation, or consumption of fixed capital in national accounts terminology, which implies that the level shift of NNI is moderated by the higher levels of depreciation related to the increased levels of investments in R&D and military weapons systems; and

**OECD Total corresponds to available countries.
• 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-km² 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)
There is a common pattern in several series, but each contains some independent variation. EIC titles where the government is not the author, where trade is mentioned, and where stock, dividends, proprietor, and directors are mentioned are a high percentage in the 1690s, which was a period with much policy discussion, especially concerning the EIC's monopoly. Government authored EIC titles and mentions of stocks, dividends, proprietors, and directors are highest in the 1770s and 1780s, another period of much policy discussion.

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

Adding “tariff” and “war” to the P term set

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 ($\gamma=3$)
Some of the variables in AJR

- 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.
## TABLE 1—DESCRIPTIVE STATISTICS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whole world</th>
<th>Base sample</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log GDP per capita (PPP) in 1995</td>
<td>8.3</td>
<td>8.05</td>
<td>8.9</td>
<td>8.4</td>
<td>7.73</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>(1.1)</td>
<td>(1.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log output per worker in 1988 (with level of United States normalized to 1)</td>
<td>-1.70</td>
<td>-1.93</td>
<td>-1.03</td>
<td>-1.46</td>
<td>-2.20</td>
<td>-3.03</td>
</tr>
<tr>
<td></td>
<td>(1.1)</td>
<td>(1.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average protection against expropriation risk, 1985–1995</td>
<td>7</td>
<td>6.5</td>
<td>7.9</td>
<td>6.5</td>
<td>6</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>(1.8)</td>
<td>(1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraint on executive in 1990</td>
<td>3.6</td>
<td>4</td>
<td>5.3</td>
<td>5.1</td>
<td>3.3</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>(2.3)</td>
<td>(2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraint on executive in 1900</td>
<td>1.9</td>
<td>2.3</td>
<td>3.7</td>
<td>3.4</td>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(1.8)</td>
<td>(2.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraint on executive in first year of independence</td>
<td>3.6</td>
<td>3.3</td>
<td>4.8</td>
<td>2.4</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>(2.4)</td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Democracy in 1900</td>
<td>1.1</td>
<td>1.6</td>
<td>3.9</td>
<td>2.8</td>
<td>0.19</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(2.6)</td>
<td>(3.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European settlements in 1900</td>
<td>0.31</td>
<td>0.16</td>
<td>0.32</td>
<td>0.26</td>
<td>0.08</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.4)</td>
<td>(0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log European settler mortality</td>
<td>n.a.</td>
<td>4.7</td>
<td>3.0</td>
<td>4.3</td>
<td>4.9</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>163</td>
<td>64</td>
<td>14</td>
<td>18</td>
<td>17</td>
<td>15</td>
</tr>
</tbody>
</table>
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: $1, x, x^2, ...$ multiplied a vector of coefficients $\beta_0, \beta_1, \beta_2, ...$
- The Stone-Weierstrass theorem ensures us that this approximation converges in the “right” sense.
- Then,
  \[ \mathbb{E}[Y|X = x_i] = \beta_0 + \beta_1 x_i + \beta_2 x^2 + ... \]
- In practice, we want to truncate the approximation at a low degree of the polynomial. For example, linear:
  \[ \mathbb{E}[Y|X = x_i] \simeq \beta_0 + \beta_1 x_i \]
- And we get:
  \[ y_i = \mathbb{E}[Y|X = x_i] + \epsilon_i \simeq \beta_0 + \beta_1 x_i + \epsilon_i \]
Why OLS?

• We still need to determine $\beta = \{\beta_0, \beta_1\}$.

• We use some criteria 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:

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

• This is called “ordinary least squares” (or OLS for short).

• By construction, it is the MMSE linear estimator of the CFE.
III. Mortality of Early Settlers

A. Sources of European Mortality in the Colonies

In this subsection, we give a brief overview of the sources of mortality facing potential settlers. Malaria (particularly Plasmodium falciparum) and yellow fever were the major sources of European mortality in the colonies. In the tropics, these two diseases accounted for 80 percent of European deaths, while gastrointestinal diseases accounted for another 15 percent (Curtin, 1989 p. 30). Throughout the nineteenth century, areas without malaria and yellow fever, such as New Zealand, were more healthy than Europe because the major causes of death in Europe—tuberculosis, pneumonia, and smallpox—were rare in these places (Curtin, 1989 p. 13).

Both malaria and yellow fever are transmitted by mosquito vectors. In the case of malaria, the main transmitter is the Anopheles gambiae complex and the mosquito Anopheles funestus, while the main carrier of yellow fever is Aedes aegypti. Both malaria and yellow fever vectors tend to live close to human habitation. In places where the malaria vector is present, such as the West African savanna or forest, an individual can get as many as several hundred infectious mosquito bites a year. For a person without immunity, malaria (particularly Plasmodium falciparum) is often fatal, so Europeans in Africa, India, or the Caribbean faced very high death rates. In contrast, death rates for the adult local population were much lower (see Curtin [1964] and the discussion in our introduction above). Curtin (1998 pp. 7-8) describes this as follows:

Children in West Africa ... would be infected with malaria parasites shortly after birth and were frequently reinfected afterwards; if they lived beyond the age of about five, they acquired an apparent immunity. The parasite remained with them, normally in the liver, but clinical symptoms were rare so long as they continued to be infected with the same species of P. falciparum.
### Table 2—OLS Regressions

<table>
<thead>
<tr>
<th></th>
<th>Whole world (1)</th>
<th>Base sample (2)</th>
<th>Whole world (3)</th>
<th>Whole world (4)</th>
<th>Base sample (5)</th>
<th>Base sample (6)</th>
<th>Whole world (7)</th>
<th>Base sample (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>is log GDP per capita in 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average protection</strong></td>
<td>0.54 (0.04)</td>
<td>0.52 (0.06)</td>
<td>0.47 (0.06)</td>
<td>0.43 (0.05)</td>
<td>0.47 (0.06)</td>
<td>0.41 (0.06)</td>
<td>0.45 (0.04)</td>
<td>0.46 (0.06)</td>
</tr>
<tr>
<td>against expropriation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>risk, 1985–1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Latitude</strong></td>
<td>0.89 (0.49)</td>
<td>0.37 (0.51)</td>
<td>1.60 (0.70)</td>
<td>0.92 (0.63)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Asia dummy</strong></td>
<td></td>
<td>−0.62 (0.19)</td>
<td></td>
<td></td>
<td>−0.60 (0.23)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Africa dummy</strong></td>
<td></td>
<td>−1.00 (0.15)</td>
<td></td>
<td></td>
<td>−0.90 (0.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>“Other” continent dummy</strong></td>
<td>−0.25 (0.20)</td>
<td></td>
<td></td>
<td></td>
<td>−0.04 (0.32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.62</td>
<td>0.54</td>
<td>0.63</td>
<td>0.73</td>
<td>0.56</td>
<td>0.69</td>
<td>0.55</td>
<td>0.49</td>
</tr>
<tr>
<td><strong>Number of observations</strong></td>
<td>110</td>
<td>64</td>
<td>110</td>
<td>110</td>
<td>64</td>
<td>64</td>
<td>108</td>
<td>61</td>
</tr>
</tbody>
</table>

Notes: Dependent variable: columns (1)-(6), log GDP per capita (PPP basis) in 1995, current prices (from the World Bank's World Development Indicators 1999); columns (7)-(8), log output per worker in 1988 from Hall and Jones (1999). Average protection against expropriation risk is measured on a scale from 0 to 10, where a higher score means more protection against expropriation, averaged over 1985 to 1995, from Political Risk Services. Standard errors are in parentheses. In regressions with continent dummies, the dummy for America is omitted. See Appendix Table A1 for more detailed variable definitions and sources. Of the countries in our base sample, Hall and Jones do not report output per worker in the Bahamas, Ethiopia, and Vietnam.

Sachs and coauthors have argued for a direct effect of climate on performance, and Gallup et al. (1998) and Hall and Jones (1999) document the correlation between distance from the equator and economic performance. To control for this, in columns (3)-(6), we add latitude as a regressor (we follow the literature in using the absolute value measure of latitude, i.e., distance from the equator, scaled between 0 and 1). This changes the coefficient of the index of institutions little. Latitude itself is also significant and has the sign found by the previous studies. In columns (4) and (6), we also add dummies for Africa, Asia, and other continents, with America as the omitted group. Although protection against expropriation risk remains significant, the continent dummies are also statistically and quantitatively significant. The Africa dummy in column (6) indicates that in our sample African countries are 90 log points (approximately 145 percent) poorer even after taking the effect of institutions into account. Finally, in columns (7) and (8), we repeat our basic regressions using the log of output per worker from Hall and Jones (1999), with very similar results.

Overall, the results in Table 2 show a strong correlation between institutions and economic performance. Nevertheless, there are a number of important reasons for not interpreting this relationship as causal. First, rich economies may be able to afford, or perhaps prefer, better institutions. Arguably more important than this reverse causality problem, there are many omitted determinants of income differences that will naturally be correlated with institutions. Finally, the measures of institutions are constructed ex post, and the analysts may have had a natural bias in seeing better institutions in richer places. As well as these problems introducing positive bias in the OLS estimates, the fact that the institutions variable is measured with considerable error and corresponds poorly to the “cluster of institutions” that matter in practice creates attenuation and may bias the OLS estimates.
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.
• 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

Margarine consumed
Divorce rate in Maine

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

2lbs 4lbs 6lbs 8lbs

3.96 per 1,000 4.29 per 1,000 4.62 per 1,000 4.95 per 1,000

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:

  $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$

- 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).
• Counterfactuals: we do not observe both potential outcomes, only one of them.

• Furthermore, $Y_{1i} - Y_{0i}$, $Y_{1i}$, and $Y_{0i}$ are heterogeneous 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

\[ n = n_1 + n_1 \]

Total # agents = Total # treated agents + 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 \( X \):

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

- Decomposition:

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

\[
UT(X) = \mathbb{E}[Y_1|X, D_i = 0] - \mathbb{E}[Y_0|X, D_i = 0]
\]

- Then, using standard probability arguments:

\[
ATE(X) = Pr(D_i = 1|X) TT(X) + Pr(D_i = 0|X) UT(X)
\]
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 directly 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?
Imagine that an agent $i$ is assigned to a treatment randomly.

Recall that:

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

Then:

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

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 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 assignment (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.
  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 attainment.


- 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, ...).
outcomes apparently arises from the presence of an unobserved factor linked within families across generations and associated with better outcomes for parents and their children. Note that our results complement recent work by Cesarini et al. (2016) who examine the intergenerational effects of wealth shocks from Swedish lotteries circa 1990. They find little impact on child development (scholastic achievement, cognitive and noncognitive skills) in the next generation, although they do not follow the children into adulthood, nor do they observe grandchildren. They attribute their null results to a strong safety net. Though such social insurance was not present in antebellum Georgia, we nonetheless find similar noneffects of random wealth shocks on intergenerational outcomes.

FIGURE III
The Gradient in School Attendance by Father’s Percentile of Total Wealth, 1850
This figure displays the fraction of children attending school for each percentile of paternal total wealth (real estate plus slaves) in 1850. The base data are the main sample of households with lottery-eligible men, as used in Table II, for example. We then exclude from this sample those fathers matching to the Smith list to measure the gradient absent a direct lottery effect. Children are ages 5–17 (inclusive of endpoints). Approximately 22% of fathers have zero wealth, and such cases are coded to a percentile of 11. Lumpiness in the data yields uneven cell sizes. The dots are sized in proportion with the cell sizes.
Randomization as an answer


• 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.
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.
Figure 1: Old Cherokee County and the 1850 Locations of the Sample

Notes: This figure displays a map of the southeastern United States with information on the location (by county) in 1850 of the lottery-eligible households in our main sample. Black lines indicate the 1850 county boundaries, drawn from the NHGIS database. The area shaded in blue in northwest Georgia denotes old Cherokee County, which was allocated by the Cherokee Lottery of 1832. The sample consists of 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. If households in our sample are resident in a county in 1850, we place a red dot at the county centroid. The area of a dot is proportional to the number of sample households resident in that county. A minor fraction of sampled households resides in counties outside the frame of this map. Such households are included in the econometric analysis, but we chose to zoom in on this region to make the feature legible in this figure. Data sources and additional variable and sample definitions are found in the text.
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.
<table>
<thead>
<tr>
<th>Panel A: Lottery winner or loser</th>
<th>Whole Sample</th>
<th>Lottery “Losers”</th>
<th>Lottery “Winners”</th>
<th>p-Value, Mean Difference [N]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dummy for unique match to Smith (1838) list</td>
<td>0.124</td>
<td>0</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Dummy for match to Smith (1838), deflated to $\frac{1}{2}$ in case of ties</td>
<td>0.155</td>
<td>0.037</td>
<td>0.995</td>
<td>0.000 [14,375]</td>
</tr>
</tbody>
</table>

| Panel B: Predetermined outcomes | | | | |
| Age, in years | 51.2 | 51.3 | 50.9 | 0.122 [14,375] |
| Born in Georgia | 0.497 | 0.497 | 0.498 | 0.889 [14,375] |
| Born in South Carolina | 0.212 | 0.210 | 0.222 | 0.263 [14,375] |
| Born in North Carolina | 0.180 | 0.180 | 0.178 | 0.804 [14,375] |
| Number of Georgia-born children in the three years prior to the lottery | 1.333 | 1.333 | 1.332 | 0.910 [14,375] |
| Cannot read and write | 0.147 | 0.147 | 0.142 | 0.593 [14,340] |
| Number of letters in surname | 6.19 | 6.20 | 6.13 | 0.072 [14,375] |
| Frequency with which surname appears in sample | 36.2 | 36.3 | 35.3 | 0.389 [14,375] |
| Surname begins with “M” or “O” | 0.101 | 0.101 | 0.104 | 0.740 [14,375] |
| Mean wealth of families in the South with same surname | 1,203.4 | 1,204.5 | 1,195.7 | 0.373 [14,093] |
| Median wealth of families in the South with same surname | 184.1 | 184.6 | 181.0 | 0.276 [14,093] |
| Mean illiteracy of adults in the South with same surname | 0.175 | 0.175 | 0.176 | 0.124 [14,093] |
| Mean school attendance of children in the South with same surname | 0.323 | 0.323 | 0.323 | 0.998 [13,975] |

<p>| Panel C: Fertility and school attendance | | | | |
| Number of children in household born after the 1832 lottery | 3.965 | 3.930 | 4.135 | 0.002 [14,375] |
| School attendance among children aged 5–17, inclusive | 0.342 | 0.342 | 0.341 | 0.799 [47,749] |</p>
<table>
<thead>
<tr>
<th>Panel D: Other outcomes</th>
<th>(1) Spouse cannot read and write</th>
<th>(2) Resides in Georgia</th>
<th>(3) Resides in Alabama</th>
<th>(4) Resides in Old Cherokee County</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.235</td>
<td>0.723</td>
<td>0.144</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>(0.424)</td>
<td>(0.447)</td>
<td>(0.351)</td>
<td>(0.317)</td>
</tr>
<tr>
<td></td>
<td>0.236</td>
<td>0.722</td>
<td>0.144</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>(0.424)</td>
<td>(0.448)</td>
<td>(0.351)</td>
<td>(0.314)</td>
</tr>
<tr>
<td></td>
<td>0.231</td>
<td>0.729</td>
<td>0.145</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>(0.421)</td>
<td>(0.445)</td>
<td>(0.352)</td>
<td>(0.332)</td>
</tr>
<tr>
<td></td>
<td>0.676</td>
<td>0.548</td>
<td>0.935</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>[11,563]</td>
<td>[14,375]</td>
<td>[14,375]</td>
<td>[14,375]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel E: Measures of wealth in 1850 (18 years after the lottery)</th>
<th>(1) Real estate wealth</th>
<th>(2) Slave wealth</th>
<th>(3) Total wealth (sum of wealth in real estate and slaves)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,999.0</td>
<td>1,339.1</td>
<td>3,323.7</td>
</tr>
<tr>
<td></td>
<td>(4,694.2)</td>
<td>(5,761.0)</td>
<td>(8,691.0)</td>
</tr>
<tr>
<td></td>
<td>1,970.8</td>
<td>1,297.3</td>
<td>3,245.5</td>
</tr>
<tr>
<td></td>
<td>(4,422.0)</td>
<td>(5,329.7)</td>
<td>(7,952.9)</td>
</tr>
<tr>
<td></td>
<td>2,198.2</td>
<td>1,635.3</td>
<td>3,876.5</td>
</tr>
<tr>
<td></td>
<td>(6,290.1)</td>
<td>(8,189.0)</td>
<td>(12,734.4)</td>
</tr>
<tr>
<td></td>
<td>0.068</td>
<td>0.021</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>[13,094]</td>
<td>[14,375]</td>
<td>[13,094]</td>
</tr>
<tr>
<td></td>
<td>100,800</td>
<td>100,800</td>
<td>100,1,000</td>
</tr>
<tr>
<td></td>
<td>3,000</td>
<td>3,000</td>
<td>3,550</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel F: Select variables for those with below $300 in 1850 total wealth</th>
<th>(1) Number of children in household born after the 1832 lottery</th>
<th>(2) Number of slaves in 1840</th>
<th>(3) Has at least one slave in 1840</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.905</td>
<td>1.4</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>(2.471)</td>
<td>(6.7)</td>
<td>(0.392)</td>
</tr>
<tr>
<td></td>
<td>3.878</td>
<td>1.3</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>(2.453)</td>
<td>(6.6)</td>
<td>(0.384)</td>
</tr>
<tr>
<td></td>
<td>4.098</td>
<td>2.3</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td>(2.591)</td>
<td>(7.4)</td>
<td>(0.437)</td>
</tr>
<tr>
<td></td>
<td>0.063</td>
<td>0.074</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>[4,506]</td>
<td>[1,761]</td>
<td>[1,761]</td>
</tr>
</tbody>
</table>
The regression

- Regression:

\[ y_{ij} = \gamma T_j + \delta_{ai} + \beta x_{ij} + \epsilon_{ij} \]

- \( i \): 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.

- \( x_{ij} \): 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:

<table>
<thead>
<tr>
<th>Specification</th>
<th>Match to List of Winners:</th>
<th>None</th>
<th>Given Name</th>
<th>State of Residence</th>
<th>County of Residence</th>
<th>Urban Residence</th>
<th>Poisson (A) and Logit (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Post-1832 fertility of lottery-eligible men ([N = 14,306])</td>
<td>Basic</td>
<td>0.132</td>
<td>0.146</td>
<td>0.124</td>
<td>0.102</td>
<td>0.126</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.058)**</td>
<td>(0.061)**</td>
<td>(0.058)**</td>
<td>(0.059)*</td>
<td>(0.058)**</td>
<td>(0.014)**</td>
</tr>
<tr>
<td></td>
<td>(\frac{1}{n})</td>
<td>0.137</td>
<td>0.156</td>
<td>0.128</td>
<td>0.104</td>
<td>0.130</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.056)**</td>
<td>(0.060)**</td>
<td>(0.056)**</td>
<td>(0.058)*</td>
<td>(0.056)**</td>
<td>(0.014)**</td>
</tr>
<tr>
<td></td>
<td>Surname</td>
<td>0.184</td>
<td>0.137</td>
<td>0.106</td>
<td>0.090</td>
<td>0.089</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.073)**</td>
<td>(0.069)**</td>
<td>(0.065)</td>
<td>(0.067)</td>
<td>(0.066)</td>
<td>(0.016)*</td>
</tr>
<tr>
<td></td>
<td>(\frac{1}{n})</td>
<td>0.175</td>
<td>0.131</td>
<td>0.095</td>
<td>0.075</td>
<td>0.074</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.072)**</td>
<td>(0.068)*</td>
<td>(0.064)</td>
<td>(0.065)</td>
<td>(0.065)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Panel B: School attendance of children aged 5–17 ([N = 47,749])</td>
<td>Basic</td>
<td>-0.005</td>
<td>-0.002</td>
<td>-0.005</td>
<td>0.001</td>
<td>-0.004</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.051)</td>
</tr>
<tr>
<td></td>
<td>(\frac{1}{n})</td>
<td>-0.004</td>
<td>0.000</td>
<td>-0.004</td>
<td>0.004</td>
<td>-0.003</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.050)</td>
</tr>
<tr>
<td></td>
<td>Surname</td>
<td>-0.005</td>
<td>0.000</td>
<td>-0.005</td>
<td>0.002</td>
<td>-0.004</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.033)</td>
</tr>
<tr>
<td></td>
<td>(\frac{1}{n})</td>
<td>-0.006</td>
<td>-0.002</td>
<td>-0.006</td>
<td>0.004</td>
<td>-0.005</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.033)</td>
</tr>
</tbody>
</table>
### OUTCOMES OF GRANDCHILDREN OF LOTTERY-ELIGIBLE MEN IN 1880

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

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).
Cities with a small change in distance to universities

To analyze the importance of the change in distance to a university using the entire sample of cities, we now turn to the city-level panel data and estimate equation (3). As already noted, if improved access to universities drove market establishment after 1386, one would expect to see a positive coefficient on the triple interaction term $\text{DistUniv}_i \times \text{Year}_t \times \text{Post}_t$. At the same time, we should not see a statistically significant pre-1386 trend rate of market establishment in places with larger changes in distance to a university (i.e., the coefficient on $\text{DistUniv}_i \times \text{Year}_t$), nor should we see a significant change in the trend rate of market establishment in 1386 among cities with no change in distance.
Cities with a large change in distance to universities

In Table III, column (4), these predictions are confirmed: there is a significant, positive change in trend that is greater in cities experiencing a greater reduction in distance to a university, but no differential trends across these areas before 1386.

Our baseline panel specification used city-year-level data. A concern could be that this choice, resulting in over 90,000 observations in our baseline specification, distorts our statistical inferences, although standard errors are clustered at the city level. In addition, there might be general equilibrium effects of market establishment: a market in one city may replace (or simulate the creation of) a market in another, meaning the units of observation might not have been independent. To gauge the importance of the choice of city-year as the unit of analysis, we aggregate our data to larger units of observation.

We first consider as our outcome variable the number of markets established in a territorial lord's land in each year.

**FIGURE VIII**

Change in the Linear Trend Rate of Market Establishment, Cities with a Large (Above Median) Change in Distance to the Closest University and Nonparametric (Lowess) graph of market establishment rates

Corresponds to the regression in Table III, column (3)
• 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?


• Take advantage of a historical accident: when was the printing press introduced in a town?
As Nieto (2003) notes, faced with high travel costs and the uncertainties associated with the matching process, printers who established a profitable press in a given city had few incentives to leave.
Hundreds of commercial arithmetics were printed 1480–1550 (see Figure I below). Print media were also associated with the development of cutting-edge business practice. Social scientists have identified double-entry bookkeeping as an important technological innovation since the early twentieth century, when Weber (1927) and Sombart (1957) argued that it played a key role in the emergence of rational, optimizing business practice. The first published description of double-entry bookkeeping appeared in 1494 (Luca Pacioli’s Summa). Printed merchants' manuals then disseminated the key ideas. Generally, merchants' manuals combined instruction in accounting and arithmetic with non-quantitative guidance on business practice (Goldthwaite 1972; Hoock 2008). A subset contained tables that simplified the calculation of interest on loans, tariffs, and transport costs. Hoock (2008) observes that, “In some ways, [these handbooks] present the same characteristics as the modern pocket calculator with integrated routines.” Figure I documents that hundreds of different merchants’ manuals were printed 1480–1550. It shows that growth in the number of merchants’ manuals printed declined from high initial rates and...
This figure presents data for the 100 cities with the highest output of incunabula editions 1450–1500. For each city it shows what share of its editions are held in the Bayerische Staatsbibliothek in Munich and how far the city is from Munich. Markers are scaled to reflect the magnitude of city book production. Fitted values estimated with locally weighted regression. Data on total incunabula production from ISTC (1998).

IV. DATA

This article exploits data on the diffusion and output of printing presses over the technology's infant industry period (1450–1500). Between 1450 and 1500, entrepreneurs established printing presses across Europe and the real price of books fell by two-thirds (Zanden 2004; Clark 2004). Between 1500 and 1800, printing technology was largely unchanged and declines in the price of books were relatively modest (Febvre and Martin 1958; Füssel 2005).

Historical research emphasizes that the period 1450–1500 was the “first infancy” of printing. Books produced 1450–1500 are referred to as incunabula, from the Latin for

Clark (2004) finds that real book prices in England fell 75% between 1450 and 1530 and stabilized at one-third the pre-Gutenberg level through the late 1700s. Zanden (2009) examines Dutch data and estimates that real prices fell by two-thirds 1450–1500. Zanden estimates that between 1500 and 1800 book prices declined from approximately one-third to one-sixth of the pre-Gutenberg level.
## Table IV

Regression Analysis of Print Media and City Growth

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

Note: The dependent variable in column (2) is ln(Pop1500/Pop1400), where Pop_t is city population in year t. The dependent variable in column (3) is ln(Pop1600/Pop1500). The dependent variable in column (4) is ln(Pop1700/Pop1500). The dependent variable in column (5) is ln(Pop1800/Pop1500). Editions Per Capita are measured as editions published 1450–1500 per 10,000 inhabitants in 1500. University is an indicator for the presence of a historic university. Roman Site and Capital are indicators for cities located on sites of Roman settlement and historic capitals. Freedom Index is the DeLong and Shleifer (1993) index of regional institutions securing property rights. Atlantic Port, Mediterranean Port, and Baltic Port are indicators for historic port cities on these bodies of water. Navigable River is an indicator for cities on historically navigable inland waterways. Log Population measures the log of city population at the beginning of the relevant period. All variables are described in the Data Appendix. City growth 1400–1500 is taken as a placebo (the average date of adoption was 1476). Heteroskedasticity-robust standard errors are clustered at the country level and presented in parentheses. Significance at the 90%, 95%, and 99% confidence levels are indicated by *, **, and***, respectively.
• 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.
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:

\[ Y_{i,t} = \theta_i + \delta_t + \sum_{t=1300}^{1700} \alpha_t D_t T_i + X'_{i,t} \gamma + \varepsilon_{i,t} \]
<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
<tr>
<td>All Cities Balanced Sample</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td>0.11</td>
<td>0.27</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Exclude German Cities</td>
<td>(0.16)</td>
<td>(0.18)</td>
<td>(0.20)</td>
<td>(0.17)</td>
<td>(0.38)</td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td>Print × Yr1400</td>
<td>0.34**</td>
<td>0.39**</td>
<td>0.41**</td>
<td>0.34**</td>
<td>1.39***</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.17)</td>
<td>(0.18)</td>
<td>(0.16)</td>
<td>(0.42)</td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>Exclude Italian &amp; Dutch Cities</td>
<td>0.13</td>
<td>0.22</td>
<td>0.08</td>
<td>0.16</td>
<td>0.73**</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.17)</td>
<td>(0.20)</td>
<td>(0.16)</td>
<td>(0.34)</td>
<td>(0.17)</td>
<td></td>
</tr>
<tr>
<td>Print × Yr1600</td>
<td>0.19</td>
<td>0.25</td>
<td>0.16</td>
<td>0.22</td>
<td>0.84**</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.16)</td>
<td>(0.17)</td>
<td>(0.14)</td>
<td>(0.42)</td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>Exclude If East of Elbe River</td>
<td>0.12</td>
<td>0.27</td>
<td>0.13</td>
<td>0.12</td>
<td>-0.32</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.33)</td>
<td>(0.37)</td>
<td>(0.31)</td>
<td>(0.52)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Atlantic × Yr1400</td>
<td>0.43*</td>
<td>0.55***</td>
<td>0.38</td>
<td>0.44*</td>
<td>-0.24</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.28)</td>
<td>(0.28)</td>
<td>(0.25)</td>
<td>(0.52)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Atlantic × Yr1500</td>
<td>0.42*</td>
<td>0.49*</td>
<td>0.33</td>
<td>0.45**</td>
<td>0.47</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.25)</td>
<td>(0.24)</td>
<td>(0.22)</td>
<td>(0.38)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Atlantic × Yr1600</td>
<td>0.60***</td>
<td>0.73***</td>
<td>0.64***</td>
<td>0.62***</td>
<td>0.32</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.20)</td>
<td>(0.21)</td>
<td>(0.19)</td>
<td>(0.38)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Atlantic × Yr1700</td>
<td>0.55</td>
<td>0.57</td>
<td>0.58</td>
<td>0.54</td>
<td>0.77</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,010</td>
<td>875</td>
<td>710</td>
<td>850</td>
<td>225</td>
<td>785</td>
<td></td>
</tr>
</tbody>
</table>

Note: This table presents estimates of Equation (1) using the balanced panel of cities with population data observed every 100 years 1300–1800. The dependent variable is log population growth: \( \ln(\text{POP}_t + 100) \), where \( \text{POP}_t \) is city population in year \( t \), and \( t = 1300, \ldots, 1700 \). Print is an indicator variable for cities that adopted the printing press 1450–1500. The variables Yr1400, ..., Yr1700 are indicators for 100-year periods starting 1400, ..., 1700. Atlantic is an indicator variable for cities that were historic ports on the Atlantic Ocean. Regressions control for city, country, and year fixed effects; country cross year fixed effects; Mediterranean port cross-year fixed effects; and log population. See Data Appendix for details on the construction of the control variables. Heterskedasticity-robust standard errors clustered by city are in parentheses. Significance at the 90%, 95%, and 99% confidence levels are indicated by *, **, and ***, respectively.

The movable type printing press was developed by Johannes Gutenberg in Mainz around 1450. In subsequent decades entrepreneurial printers spread the technology to other European economies. It shows that distance from Mainz is a strong instrument for adoption and yields large, significant estimates of the technology’s impact. For analysis of factors associated with adoption see also the Online Appendix.
• If we do not have a good comparison, we can build synthetic controls.


• 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.
striking once the figures are expressed in per capita terms to reflect relative exposure to terrorism. During the period 1968–1997, ETA’s activity in the Basque Country, measured as the number of deaths per inhabitant per year, was 37 times as large as in the rest of Spain.7

II. Using Other Spanish Regions to Construct a "Synthetic" Basque Country Without Terrorism

A. Analytical Methods and Main Results

The goal of this section is to assess the impact that terrorism has had on economic growth for the Basque Country. Table 3, in columns (1) and (2), reports values of some variables typically associated with growth potential8 for the Basque Country and Spain for the immediate pre-terrorism years. During the 1960's, relative to the whole country, the Basque Country had higher per capita income, higher investment ratio (investment/production), was more densely populated, had a higher percentage of industrial production, and a better educated labor force. As a result, a simple comparison of the economic performance of the Basque Country and the rest of Spain during the terrorism years may not only reflect the impact of terrorism, but also other pre-terrorism differences which affected subsequent economic growth.

We approach this problem by comparing the economic evolution of the Basque Country during the terrorist era with that of a weighted combination of other Spanish regions chosen to resemble the characteristics of the Basque Country before terrorism. We conceptualize such a weighted average of other Spanish regions as a "synthetic" Basque Country without terrorism, against which we can compare the actual Basque Country with terrorism. Let \( J \) be the number of available control regions (the 16 Spanish regions other than the Basque Country), and \( W_j \) \( (j = 1, \ldots, J) \) a vector of nonnegative weights which sum to one. The scalar \( w_j \) \( (j = 1, \ldots, J) \) represents the weight of region \( j \) in the synthetic Basque Country. Each different value for \( W \) produces a different synthetic Basque Country, and therefore the choice of a valid subset of control regions is embedded in the choice of the weights \( W \).

As said above, the weights are chosen so that the synthetic Basque country most closely resembled the Basque Country before terrorism. We make several assumptions for Table 3:

- \( b = \text{Gross Total Investment/GDP, average for 1964}–\text{1969}. \)
- \( c = \text{Persons per square kilometer, 1969}. \)
- \( d = \text{Percentages over total production, 1961–1969}. \)
- \( e = \text{Percentages over working-age population, 1964–1969}. \)

### TABLE 3 — PRE-TERORISM CHARACTERISTICS, 1960’S

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

Sources: Authors’ computations from Matilde Mas et al. (1998) and Fundación BBV (1999).

---

7 See also Mark Kurlansky (1999) and CNN (2001) for additional background information on the Basque conflict.

expect terrorism to have a lagged negative effect on per capita GDP. In Figure 2, we plotted the per capita GDP gap, \( \frac{Y}{H} \), as a percentage of Basque per capita GDP, and the number of deaths caused by terrorist actions (used as a proxy for overall terrorist activity). As expected, spikes in terrorist activity seem to be followed by increases in the amplitude of the gap.
• 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.
• Imagine that we have a linear regression:

\[ y = \beta_0 + \beta_1 x + \varepsilon \]

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

• However, we have a variable \( z \) such that

\[
\begin{align*}
\text{Cov}(x, z) & \neq 0 \\
\text{Cov}(z, \varepsilon) & = 0
\end{align*}
\]

• The first assumption is testable. The second one is not.
Why do IVs work?

- We can find:
  \[
  \text{Cov}(y, z) = \text{Cov}(\beta_0 + \beta_1 x + \varepsilon, z) = \text{Cov}(\beta_0, z) + \beta_1 \text{Cov}(x, z) + \text{Cov}(\varepsilon, z)
  \]

- Therefore:
  \[
  \beta_1 = \frac{\text{Cov}(y, z)}{\text{Cov}(x, z)}
  \]

- First stage vs. second stage.
The relationship between settler mortality and current institutions is interesting in its own right. The regression shows that mortality rates faced by the settlers more than 100 years ago explain over 25 percent of the variation in current institutions. We also document that this relationship works through the channels we hypothesize: (potential) settler mortality rates were a major determinant of settlements; settlements were a major determinant of early institutions (in practice, institutions in 1900); and there is a strong correlation between early institutions and institutions today. Our two-stage least-squares estimate of the effect of institutions on performance is relatively precisely estimated and large. For example, it implies that improving Nigeria's institutions to the level of Chile could, in the long run, lead to as much as a 7-fold increase in Nigeria's income (in practice Chile is over 11 times as rich as Nigeria).

The exclusion restriction implied by our instrumental variable regression is that, conditional on the controls included in the regression, the mortality rates of European settlers more than 100 years ago have no effect on GDP per capita today, other than their effect through institutional development. The major concern with this exclusion restriction is that the mortality rates of settlers could be correlated with the current disease environment, which may have a direct effect on economic performance. We believe that this is unlikely to be the case and that our exclusion restriction is plausible. The great majority of European deaths in the colonies were caused by malaria and yellow fever. Although these diseases were fatal to Europeans who had no immunity, they had limited effect on indigenous adults who had developed various types of immunity. These diseases are therefore unlikely to be the reason why many countries in Africa and Asia are very poor today (see the discussion in Section III, subsection A). This notion is...
### TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

|                | Base sample | Base sample | Base sample | Base sample | Base sample | Base sample | Base sample | Base sample | Base sample |
|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                | (1)         | (2)         | (3)         | (4)         | (5)         | (6)         | (7)         | (8)         | (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        | 0.04        | -1.20       | -1.20       | -0.92       | -1.10       | -0.92       |
| (1.34)         | (1.46)      | (0.84)      | (0.84)      | (1.8)       | (1.8)       | (0.40)      | (0.52)      | (0.40)      |
| Asia dummy     | -0.92       | -1.10       | -0.46       | -0.46       | -0.99       | -0.99       | -0.94       | -0.99       | -0.94       |
| (0.40)         | (0.52)      | (0.36)      | (0.42)      | (0.85)      | (1.0)       | (0.85)      | (1.0)       |             |
| Africa dummy   | -0.46       | -0.44       | -0.27       | -0.27       | -0.46       | -0.46       | -0.27       | -0.27       | -0.27       |
| (0.36)         | (0.42)      | (0.41)      | (0.84)      | (0.84)      | (0.36)      | (0.42)      | (0.84)      | (0.84)      |
| "Other" continent dummy | -0.94       | -0.99       | -0.94       | -0.94       | -0.94       | -0.94       | -0.94       | -0.94       | -0.94       |
| (0.85)         | (1.0)       | (0.85)      | (1.0)       | (0.85)      | (1.0)       | (0.85)      | (1.0)       | (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        | 0.99        | 1.24        | 1.24        | 2.00        | 2.00        | 1.1         | 1.1         | 1.1         |
| (1.34)         | (1.50)      | (1.43)      | (0.84)      | (0.84)      | (1.34)      | (1.40)      | (0.84)      | (0.84)      |
| Asia dummy     | 0.33        | 0.47        | 1.24        | 1.24        | 0.33        | 0.47        | 1.1         | 1.1         | 1.1         |
| (0.49)         | (0.50)      | (0.84)      | (0.84)      | (0.84)      | (0.49)      | (0.50)      | (0.84)      | (0.84)      |
| Africa dummy   | -0.27       | -0.26       | 1.24        | 1.24        | -0.27       | -0.27       | 1.1         | 1.1         | 1.1         |
| (0.41)         | (0.41)      | (0.84)      | (0.84)      | (0.84)      | (0.41)      | (0.41)      | (0.84)      | (0.84)      |
| "Other" continent dummy | 0.27        | 0.30        | 0.13        | 0.13        | 0.47        | 0.47        | 0.30        | 0.33        | 0.28        |
| (0.06)         | (0.06)      | (0.08)      | (0.07)      | (0.07)      | (0.07)      | (0.07)      | (0.06)      | (0.06)      | (0.06)      |
| 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**

<table>
<thead>
<tr>
<th></th>
<th>Base sample</th>
<th>Base sample</th>
<th>Base sample</th>
<th>Base sample</th>
<th>Base sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Average protection against expropriation risk 1985–1995</td>
<td>0.52</td>
<td>0.47</td>
<td>0.49</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>64</td>
<td>64</td>
<td>60</td>
<td>60</td>
<td>37</td>
</tr>
</tbody>
</table>

Notes:
The dependent variable in columns (1)-(8) is log GDP per capita in 1995, PPP basis. The dependent variable in column (9) is log output per worker.

"Average protection against expropriation risk 1985–1995" is measured on a scale from 0 to 10, where a higher score means more protection against risk of expropriation of investment by the government, from Political Risk Services. Panel A reports the two-stage least-squares estimates, instrumenting for protection against expropriation risk using log settler mortality; Panel B reports the corresponding first stage. Panel C reports the coefficient from an OLS regression of the dependent variable against average protection against expropriation risk. Standard errors are in parentheses. In regressions with continent dummies, the dummy for America is omitted. See Appendix Table Al for more detailed variable descriptions and sources.
Regression discontinuity design

- We take advantage of a sudden change (threshold effect) on a treatment.
- Treatments are often somewhat arbitrary.
- Proposed by Thistlewaite 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):
  \[ P(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases} \]

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

- Nonparametric vs. parametric specifications of distance to threshold.
Figure 6.1.1: The sharp regression discontinuity design

A. Linear $E[Y_0 | X_i]$

B. Nonlinear $E[Y_0 | X_i]$

C. Nonlinearity mistaken for discontinuity
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:

  \[ c_{ibd} = \alpha + \gamma \times mita_d + X_{ib}^\prime \beta + f(\text{geographical location}_d) + \phi_b + \varepsilon_{ibd} \]
This discrete change suggests a regression discontinuity (RD) approach for evaluating the long-term effects of the mita, with the mita boundary forming a multidimensional discontinuity in longitude–latitude space. Because validity of the RD design requires all relevant factors besides treatment to vary smoothly at the mita boundary, I focus exclusively on the portion that transects the Andean range in southern Peru. Much of the boundary tightly follows the steep Andean precipice, and hence has elevation and the ethnic distribution of the population changing discretely at the boundary. In contrast, elevation, the ethnic distribution, and other observables are statistically identical across the segment of the boundary on which this study focuses. Moreover, specification checks using detailed census data on local tribute (tax) rates, the allocation of tribute revenue, and demography—collected just prior to the mita’s institution in 1573—do not find differences across this segment. The multidimensional nature of the discontinuity raises interesting and important questions about how to specify the RD polynomial, which will be explored in detail.

Using the RD approach and household survey data, I estimate that a long-run mita effect lowers equivalent household consumption by around 25% in 1864.
The results can be seen graphically in Figure 2. Each subfigure shows a district-level scatter plot for one of the paper's main outcome variables. These plots are the three-dimensional analogues to standard two-dimensional RD plots, with each district capital's longitude on the $x$ axis, its latitude on the $y$ axis, and the data value for that district shown using an evenly spaced monochromatic color scale, as described in the legends. When the underlying data are at the microlevel, I take district-level averages, and the size of the dot indicates the number of observations in each district. Importantly, the scaling on these dots, which is specified in the legend, is nonlinear, as otherwise some would be microscopic and others too large to display. The background in each plot shows predicted values, for a finely spaced grid of longitude–latitude coordinates.
ordinates, from a regression of the outcome variable under consideration on a cubic polynomial in longitude–latitude and the mita dummy. In the typical RD context, the predicted value plot is a two-dimensional curve, whereas here it is a three-dimensional surface, with the third dimension indicated by the color gradient. The shades of the data points can be compared to the shades of the predicted values behind them to judge whether the RD has done an adequate job of averaging the data across space. The majority of the population in the region is clustered along the upper segment of the mita boundary, giving these.

Three-dimensional surface plots of the predicted values are shown in Figure A2 in the Supplemental Material, and contour plots are available upon request.
## TABLE II

### Dependent Variable

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;100 km of Bound.</td>
<td>&lt;75 km of Bound.</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>Mita</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>−0.284</td>
<td>−0.216</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td>(0.207)</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.060</td>
<td>0.060</td>
</tr>
<tr>
<td><strong>Mita</strong></td>
<td>−0.337***</td>
<td>−0.307***</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.101)</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.046</td>
<td>0.036</td>
</tr>
<tr>
<td><strong>Mita</strong></td>
<td>−0.277***</td>
<td>−0.230**</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.089)</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.044</td>
<td>0.042</td>
</tr>
</tbody>
</table>

- Geo. controls: yes
- Boundary F.E.s: yes
- Clusters: 71 60 52 289 239 185 63
- Observations: 1478 1161 1013 158,848 115,761 100,446 37,421

The unit of observation is the household in columns 1–3 and the individual in columns 4–7. Robust standard errors, adjusted for clustering by district, are in parentheses. The dependent variable is log equivalent household consumption (ENAHO (2001)) in columns 1–3, and a dummy equal to 1 if the child has stunted growth and equal to 0 otherwise in columns 4–7 (Ministro de Educación (2005a)).

Mita is an indicator equal to 1 if the household's district contributed to the mita and equal to 0 otherwise (Saignes (1984), Amat y Juniet (1947, pp. 249, 284)). Panel A includes a cubic polynomial in the latitude and longitude of the observation's district capital, panel B includes a cubic polynomial in Euclidean distance from the observation's district capital to Potosí, and panel C includes a cubic polynomial in Euclidean distance to the nearest point on the mita boundary. All regressions include controls for elevation and slope, as well as boundary segment fixed effects (F.E.s). Columns 1–3 include demographic controls for the number of infants, children, and adults in the household. In columns 1 and 4, the sample includes observations whose district capitals are located within 100 km of the mita boundary, and this threshold is reduced to 75 and 50 km in the succeeding columns. Column 7 includes only observations whose districts border the mita boundary. 78% of the observations are in mita districts in column 1, 71% in column 2, 68% in column 3, 78% in column 4, 71% in column 5, 68% in column 6, and 58% in column 7. Coefficients that are significantly different from zero are denoted by the following system: *10%, **5%, and ***1%.
• 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...).


• Often called RDD.
shown below, finer month-year of birth comparisons and regression-based analyses point to similar conclusions. Both compulsory schooling changes were secured by an extensive program of school building and the key elements of the school system did not change between 1947 and 1972. The extra year of schooling created by the 1972 change kept students in high school courses for another year and meant that many more finished these and received formal qualifications. Ministry of Education reports and commentary from the time suggest that the extra year created by the 1947 change was used to introduce some students to more advanced materials and help other students master more basic material. Both changes have been found to increase the earnings of affected cohorts. In particular, Harmon and Walker (1995); Oreopoulos (2006); and Devereux and Hart (2010) all find that the 1947 change had statistically significant effects on male earnings; only Devereux and Hart (2010) fail to find statistically significant returns for women. Our own analysis of the earnings effects of the 1947 change (based on our RD model using month-year of birth) confirms this picture (see online Appendix C for analysis). Grenet (2013) finds that the 1972 reform had statistically significant effects on the earnings of men and women. Our analysis of

---

**Figure 1. Years of Full-Time Education by Quarter of Birth**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No university degree</td>
<td>0.90</td>
<td>0.85</td>
<td>0.80</td>
<td>0.75</td>
<td>0.70</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
</tr>
<tr>
<td>≤ 11 years</td>
<td>0.85</td>
<td>0.80</td>
<td>0.75</td>
<td>0.70</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
<td>0.35</td>
</tr>
<tr>
<td>≤ 10 years</td>
<td>0.80</td>
<td>0.75</td>
<td>0.70</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
<td>0.35</td>
<td>0.30</td>
</tr>
<tr>
<td>≤ 9 years</td>
<td>0.75</td>
<td>0.70</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
<td>0.35</td>
<td>0.30</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Figure 3: Daily UI Benefits
Bottom Kink Sample

Figure 4: Daily UI Benefits
Top Kink Sample
Figure 5: Log Time to Next Job
Bottom Kink Sample

Figure 6: Log Time to Next Job
Top Kink Sample
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.
derherstellung des Berufsbefugnisses. 
Bom 7. April 1933.

Regierung hat das folgende Gesetz bei
hiermit verkündet wird:

§ 1

Reichsgerichtsblatt

Teil I

Ausgegeben zu Berlin, den 7. April 1933


§ 1

Wiederherstellung eines nationalen Be-
stands und zur Vereinfachung der Ber-
ken Beamte nach Maßgabe der folgen-
ungen aus dem Amt entlassen werden,
je nach dem geltenden Recht hierfür
Berausgungen nicht vorgesehen.

§ 3

Nicht arischer Beamte, die nicht in den Ruhestand (§§ 8 ff.) zu
es sich um Ehrenämter handelt, sind vor dem Inkrafttreten dieses Gesetzes
getreten sind, entsprechende Am

mit gleichem Grundgehalt der von
besiedelten Stelle bewilligt werden;
sicherung nach Maßgabe der reichsge-
verpflichtung findet nicht statt.

(4) Die Vorschriften der Abs. 3 und
Personen der im Abs. 1 bezeichneten 
vor dem Inkrafttreten dieses Gesetzes
stand getreten sind, entsprechende Am
TABLE 1  
NUMBER OF DISMISSED MATHEMATICS PROFESSORS

<table>
<thead>
<tr>
<th>Year of Dismissal</th>
<th>Number of Dismissed Professors</th>
<th>Percentage of All Mathematics Professors in 1933</th>
</tr>
</thead>
<tbody>
<tr>
<td>1933</td>
<td>35</td>
<td>15.6</td>
</tr>
<tr>
<td>1934</td>
<td>6</td>
<td>2.7</td>
</tr>
<tr>
<td>1935</td>
<td>5</td>
<td>2.2</td>
</tr>
<tr>
<td>1936</td>
<td>1</td>
<td>.4</td>
</tr>
<tr>
<td>1937</td>
<td>2</td>
<td>.9</td>
</tr>
<tr>
<td>1938</td>
<td>1</td>
<td>.4</td>
</tr>
<tr>
<td>1939</td>
<td>1</td>
<td>.4</td>
</tr>
<tr>
<td>1940</td>
<td>1</td>
<td>.4</td>
</tr>
<tr>
<td>1933–34</td>
<td>41</td>
<td>18.3</td>
</tr>
</tbody>
</table>
Fig. 2.—Effect of dismissals on faculty in mathematics departments.

a, Effect on department size. Dashed line: departments with dismissals in 1933 and 1934; solid line: departments without dismissals.

b, Effect on average faculty quality. Dashed line: departments with dismissals of above-average department quality between 1933 and 1934; solid line: departments without dismissals.
Fig. 3.—Effect of dismissals on PhD student outcomes.

a, Effect on the probability of publishing dissertation in a top journal. Dashed line: departments with above-average quality dismissals between 1933 and 1934; solid line: departments without dismissals.

b, Effect on probability of becoming a full professor later in life. Dashed line: departments with above-average quality dismissals between 1933 and 1934; solid line: departments without dismissals.

c, Effect on the probability of having positive lifetime citations. Dashed line: departments with above-average quality dismissals between 1933 and 1934; solid line: departments without dismissals.
Other methods

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