Causation versus Correlation: Tools and Solutions from Economics

Mungo Wilson

July 2020

Selected data

Imagine the editor of the Daily Mail decides to conduct a survey of voter intentions: He asks readers to write in (or text or email) and inform him how they intend to vote in the next election. Millions of readers respond.

Now imagine a small outfit selects a few hundred voters at random out of the whole electorate and obtains responses on their voting intentions.

Selected data

Imagine the editor of the Daily Mail decides to conduct a survey of voter intentions: He asks readers to write in (or text or email) and inform him of how they intend to vote in the next election. Millions of readers respond.

Now imagine a small outfit selects a few hundred voters at random out of the whole electorate and obtains responses on their voting intentions.

Something very like this happened in 1936 when Crossley, Roper and Gallup carried out a small random poll and called the presidential election for Roosevelt, contrary to a huge write-in survey conducted by the Literary Digest which called it for Landon.

Selected data

Ever since then we've known accurate polling must be done by random sampling (getting harder because of the decline of the land line), and this is the problem of bias.

How many people currently have coronavirus in the UK?

Many of us have been enjoined to download and use 'track and trace' apps such as the one in the King's College study. Is this random sampling? Or more like a write-in?

When I spoke to one leading pollster about the need for random testing to establish infection rates he said that the King's College app was a good example of this being put into practice.

Random sampling and random assignment

Random sampling (and its twin in RCTs, random assignment) is so powerful because in expectation every subject sampled is the SAME as the average subject in the population.

This is not true for data obtained in any way other than via random sampling. Inference from such ‘found’ data is usually almost impossible.

There is one known fix for selected data, if you can observe the population from which the selected data is drawn: the famous Heckman correction.

Random assignment

By randomly (sampling and then) assigning subjects to a control and to a treatment group, since the subjects in each group are in expectation the same as the average subject in the population, any difference in outcomes is almost surely due to the treatment of the treated and the non-treatment of the untreated.

Such a difference in outcomes can then reasonably be inferred to be caused by treatment of the treated.

The important point is counterfactuals: because they are in expectation the same, what happens to the treated is what would have happened to the controls, had they been treated, and vice versa.

Randomized controlled trials (RCTs)

Because of this feature, random assignment, RCTs are an incredibly powerful way of estimating effects of treatment.

There are still problems:

- External validity – the results are usually difficult to extrapolate to a different situation
- Especially if more widespread treatment would change the equilibrium (e.g. free university tuition; guaranteed minimum incomes)
- The Stable Unit Value Assumption is required: treating the treated has no effect on outcomes for the untreated (think vaccines).

Use of RCTs

RCTs are now increasingly used in business, development economics, public policy and behavioural economics to develop insights into causation.

Unfortunately, there are many situations, usually in the field rather than the lab, where RCTs are either:

- Unfeasible; or
- Unethical; or both.

For example, how can we tell if anti-COVID lockdown policies were effective or not in saving lives?

The identification revolution in economics

Most interesting and important questions in economics cannot be settled by running randomized controlled trials.

Economists were like resource-poor countries: they had to get smarter in order to make progress.

The techniques developed (IV, RDD and D-in-D) are ingenious and universally used in the discipline, but strangely economists don't talk to non-economists about them, and only in the last couple of years have they been taught outside Economics doctoral programs.

The recent pandemic has revealed the lack of knowledge of even the problems, let alone the solutions, outside the field of Economics. Economists should have been much more communicative!

Natural experiments and confounders

Fortunately for us, nature runs a lot of RCTs for us, we just have to know how to look for them.

We have an outcome Y and a treatment X. X is not in general randomly assigned. Therefore any correlation between X and Y cannot be interpreted as causal.

In particular, there be another unobserved variable W, called a confounder, which determines both X and independently, Y. The apparent relationship between X and Y is in fact due to their both depending W. X may have no effect on Y at all.

Confounders: C-section babies

It's probably time for an example. Many studies have shown a strong correlation between babies born by C-section, as opposed to natural birth, and a higher rate of childhood obesity later in childhood.

Here X is a treatment: birth by C-section (or not). Y is outcome: childhood obesity (or not). X and Y are strongly correlated in the data.

But W? What about mother's body-mass index at time of birth? That may have a strong influence on (a) probability of a C-section birth (it does) and (b) eventual obesity in children (it does).

Therefore the correlation between X and Y tells us **absolutely nothing** about the direct effect of being born by C-section on probability of developing obesity later in childhood.

Natural experiments

Fortunately for us, nature runs a lot of RCTs for us, we just have to know how to look for them.

What we need is another variable, Z, which **does** randomly assign different levels of X to different subjects, without otherwise having any effect on Y.

Such a variable is called an Instrumental Variable (IV) and has been the secret sauce of almost all empirical research in economics for the past 50 years. It is time its use became much more widespread.

C-section babies

Imagine there is a sub-group of expectant mothers who have a religious taboo against C-sections.

This religious taboo is not otherwise associated with childhood obesity.

Possession of this religious taboo, or not, would be a random way of determining whether or not a mother received a C-section delivery. By looking at that fraction of variation in C-section treatment explained by the presence or absence of religious taboo, we could indeed observe whether C-section babies were **thereby** more likely to become obese children.

Two-stage Least Squares (2SLS)

The cookbook recipe is then as follows:

- Regress X (C-section or not) on Z (religious taboo or not)
- Take the fitted value of this regression (the value of X predicted by Z, which depends only on Z, not on X), E[X|Z] and
- Regress Y (obesity as an older child) on this fitted value E[X|Z].
- The resulting estimate in an unbiased estimate of the causal effect of X on Y.

This has to be one of the smartest ideas ever. And it dates back to the 1940s!

Two-stage Least Squares (2SLS)

The cookbook recipe is then as follows:

- Regress X (C-section or not) on Z (religious taboo or not)
- Take the fitted value of this regression (the value of X predicted by Z, which depends only on Z, not on X), E[X|Z], and
- Regress Y (obesity as an older child) on this fitted value E[X|Z].
- The resulting estimate in an unbiased estimate of the causal effect of X on Y.

This has to be one of the smartest ideas ever. And it dates back to the 1940s!

Beautiful instruments and where to find them?

- How do we find instruments?
  - Inspiration
  - Brute force (no time in today's talk – there is a way! Google Bartik instruments.)

Inspiration

- How do we find instruments?
  - Inspiration
- Some suggested (non-exhaustive) headings:
  - Lotteries
  - Persistence
  - Natural (and non-natural) barriers
  - Climate, weather, and features of the environment

Lotteries

- Many situations involve lottery-based or lottery-like random assignments. Sometimes literally a lottery.
- Angrist and Krueger (1992): being drafted to serve in Vietnam. http://www.nber.org/papers/w4067.pdf)

Lotteries

- Many situations involve lottery-based or lottery-like random assignments. Sometimes literally a lottery.
- Angrist and Krueger (1992): being drafted to serve in Vietnam. http://www.nber.org/papers/w4067.pdf
- Angrist – charter schools lottery. (Textbook chapter 3.1)

Lotteries

- Many situations involve lottery-based or lottery-like random assignments. Sometimes literally a lottery.
- Angrist and Krueger (1992): being drafted to serve in Vietnam. http://www.nber.org/papers/w4067.pdf
- Angrist – charter schools lottery.
- Canayaz (2017) – seniority determination for Democrat representatives. https://sites.google.com/site/mehmetihsancanayaz/
- Kuhn, Kooreman, Soetevent, Kapteyn (2011): Study effects of money prizes on lottery winners and their neighbours.

Lotteries

- Many situations involve lottery-based or lottery-like random assignments. Sometimes literally a lottery. **Other times, something like gender of first or first two children is awfully lottery-like.** Look for lotteries!
- Angrist and Evans (1998) and Angrist, Lavy and Schlosser (2010) – two same versus two different children (not obviously a lottery, but it is one) or twins versus singleton (ditto). (Textbook chapter 3.3)

Lotteries

- Many situations involve lottery-based or lottery-like random assignments. Sometimes literally a lottery. Other times, something like gender of first or first two children is awfully lottery-like. Look for lotteries!
- Angrist and Evans (1998) and Angrist, Lavy and Schlosser (2010) – two same versus two different children (not obviously a lottery, but it is one) or twins versus singleton (ditto). (Textbook chapter 3.3)
- Family firm succession: https://doi.org/10.1162/qjec.122.2.647

Lottery = Something that's impossible to predict

- The weather (in Great Britain, anyway).
- Gender of nth-born child.
- Longevity of medieval monarchs.
- Whether the return on the stock market on a given day or week is above or below average.
- Earthquakes. (Morse http://faculty.haas.berkeley.edu/morse/research/papers/morsepayday_jfe2.pdf)
- Deaths. (CEOs. Presidents and Prime Ministers.) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=875808
- Sports games outcomes (sometimes).
- Elections/other votes (e.g. shareholder) which are too close to call.

Persistence

- General idea: Something that happened a long time ago can have very persistent long-term effects, while obviously not having any direct effect on the dependent variable.
- La Porta, Lopez de Silanes, Shleifer and Vishny (1998, 1999): legal origins and the development of local capital markets.
- Many observers (not generally French) have argued that British legal system treats equity investors more equitably than French legal system. Therefore countries who want to develop better decentralized sources of finance for firms should adopt British legal system.
- Obvious endogeneity problems.

Persistence

- Most countries have a legal system that mimics either the 19th century French or British legal system.
  - Colonial experience
  - Napoleon versus the British navy
  - Plus a desire to succeed like the Germans.
- For most such countries, decision was largely made by the end of the 19th century, or up to 300 years earlier (like the USA).
- Clearly no direct effect on capital market growth today.
- Therefore legal origin can be a valid instrument for exploring importance of different legal systems and laws for development of capital markets.

Persistence

- General idea: Something that happened a long time ago can have very persistent long-term effects, while obviously not having any direct effect on the dependent variable.
- Cohen, Coval and Malloy (2012): unemployment rates by city in the USA in 1931 explain unionization rates in 1931-2…and employment creation in the 1990s.
- They may also explain other things, such as political climate.
- Mechanism need not be clear to use as an instrument.

Why are some countries richer than others?

- Traditional answer: differences in natural resources, climate, vulnerability to natural disasters and endemic disease.
- New answer: differences in nature of institutions. For example, rule of law, ability of narrow elites to expropriate and the flip side, the protection of private property rights, democracy.
  - East vs West Germany
  - North vs South Korea
  - Northern Mexico vs South-Western USA

Richer because better institutions, or better institutions because richer?

- Maybe richer countries can afford better institutions
- Maybe we only think the particular institutions peculiar to richer countries are better because richer countries have them.
- And what determines what institutions a country gets anyway?

Richer because better institutions, or better institutions because richer?

- Maybe richer countries can afford better institutions
- Maybe we only think the particular institutions peculiar to richer countries are better because richer countries have them.
- And what determines what institutions a country gets anyway?
- Acemoglu et al: in most countries, institutions are extremely persistent. For example, differences in rule of law across countries little different from over 100 years ago.
  - Exceptions: South Korea, 1960s; Taiwan, Greece, Spain and Portugal in the 1970s.
  - Longer ago: Japan in 1860s, Germany in 1860s.

If institutions today are determined by institutions yesterday

- Maybe countries with ‘better’ institutions 100 years ago were already richer (they were), and being richer is persistent, like climate, so that doesn’t help much.
- And what determined who got better institutions 100 years ago?
- Answer: In most cases, local institutions were created by colonial powers. These powers were usually European nations.

If institutions today are determined by institutions yesterday

- Maybe countries with ‘better’ institutions 100 years ago were already richer (they were), and being richer is persistent, like climate, so that doesn’t help much.
- And what determined who got better institutions 100 years ago?
- Answer: In most cases, local institutions were created by colonial powers. These powers were usually European nations.
- Institutions created varied enormously, from Belgian Congo at one extreme to ‘New Europes’ of USA, Canada, Australia and New Zealand at the other.

If institutions today are determined by institutions yesterday

- Maybe countries with ‘better’ institutions 100 years ago were already richer (they were), and being richer is persistent, like climate, so that doesn’t help much.
- And what determined who got better institutions 100 years ago?
- Answer: In most cases, local institutions were created by colonial powers. These powers were usually European nations.
- Institutions created varied enormously, from Belgian Congo at one extreme to ‘New Europes’ of USA, Canada, Australia and New Zealand at the other.
- From the viciously exploitative to the (relatively) enlightened.

What determined local institutional choices of the colonial power?

- Acemoglu et al propose degree of permanent European settlement.
- Still not done: perhaps European settlers moved only to those colonies with a high probability of being rich (e.g. with better geographical factors); or maybe there just is a direct benefit of having large numbers of Europeans settle in your country.
- But there is something else: in the period of colonial settlement, European settlers tended to move mainly to where they could **survive**.

Settler mortality: death rates of soldiers, sailors and bishops in the early colonial period

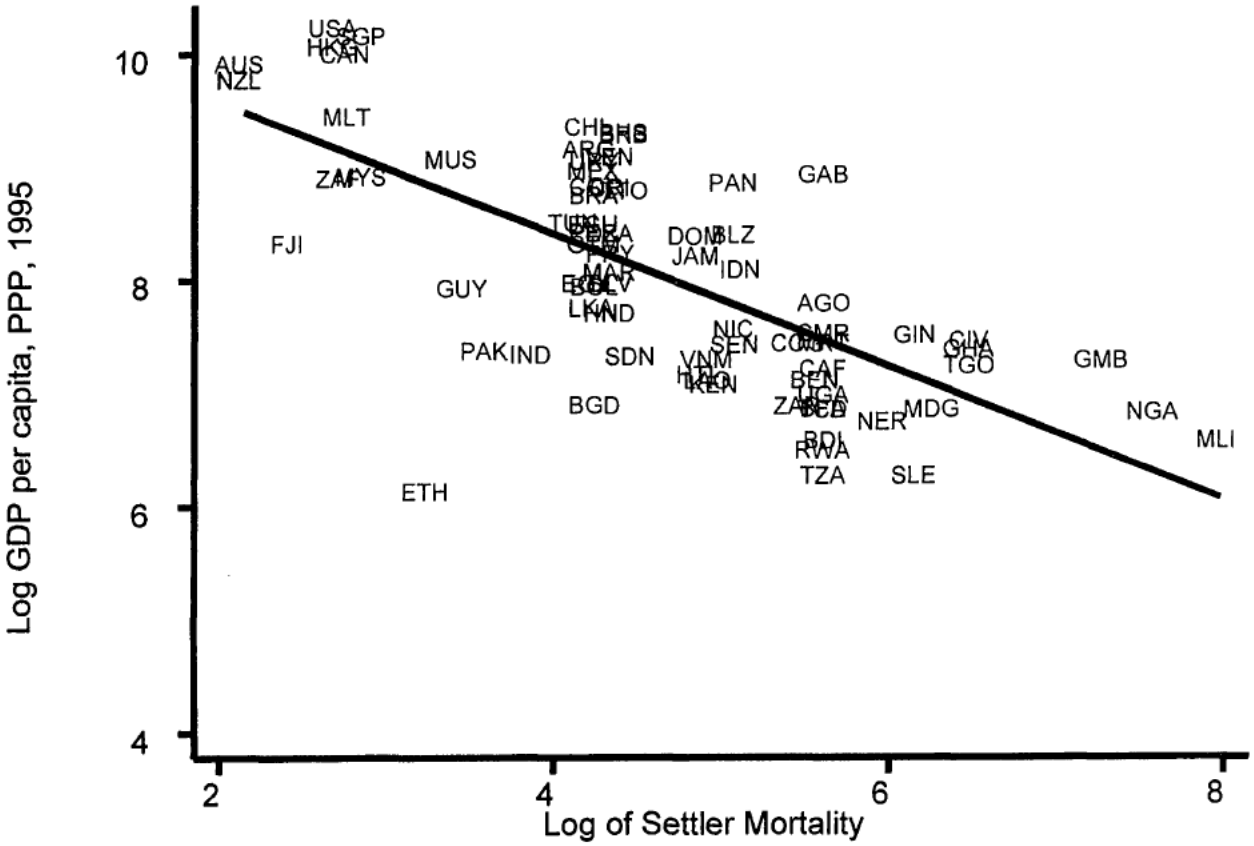


FIGURE 1. REDUCED-FORM RELATIONSHIP BETWEEN INCOME AND SETTLER MORTALITY

How does settler mortality in colonial times determine contemporary prosperity?

- If settlers could survive, many migrated.
- Colonies with many European settlers got institutions that benefitted local settlers (eventually): rule of law, private property protections. (E.g. Singapore.)
- Colonies with few European settlers were simply exploited. (‘Gold Coast’, ‘Ivory Coast’.) Forced labour, no property rights, arbitrary power of colonial government.
- After independence, most countries continued to have the institutions bequeathed them by their former colonial rulers…(E.g. Forced labour in Zaire under Mobutu, and in Guatemala until 1945.)
- And institutions determine prosperity!
- How would you prove it?

How does settler mortality in colonial times determine contemporary prosperity?

- Instrumental variable is European mortality rates by country in the early colonial period (e.g. 1600-1700s for Latin America, 1800s for Africa).
- Data from death rates of soldiers and sailors stationed in colonies (recorded by government of colonising country), and of Bishops in Latin America (kept by Catholic Church).

How does settler mortality in colonial times determine contemporary prosperity?

- Instrumental variable is European mortality rates by country in the early colonial period (e.g. 1600-1700s for Latin America, 1800s for Africa).
- Data from death rates of soldiers and sailors stationed in colonies (recorded by government of colonising country), and of Bishops in Latin America (kept by Catholic Church).
- The instrument needs to predict GDP per capita today only through its indirect relation to settlement rates, quality of early institutions, and therefore modern ones.
- Can you think of any problems?

What if endemic disease is the determinant?

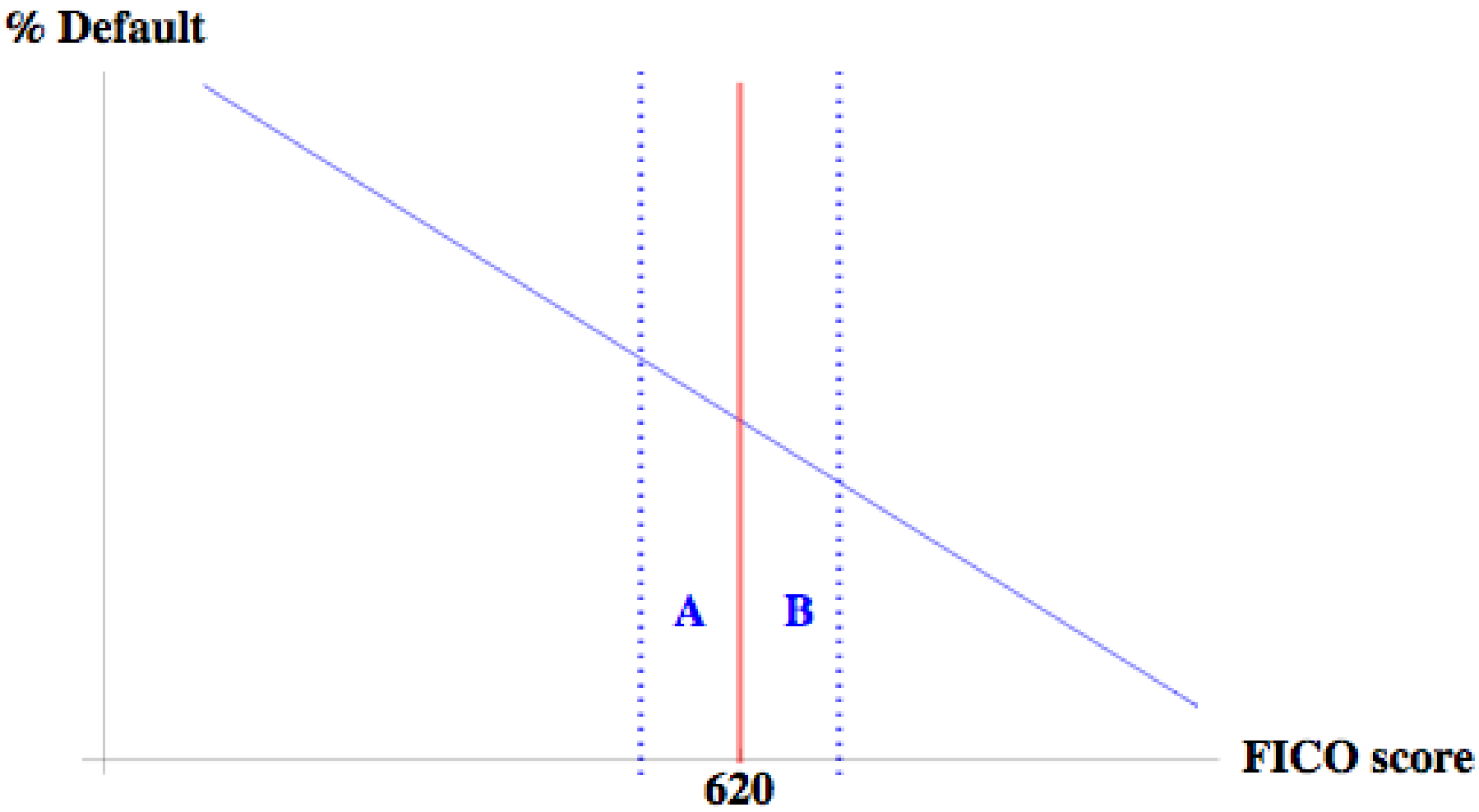
- If countries with high levels of endemic disease have low GDP per capita, then
  - Early European visitors' mortality rates higher
  - Today's GDP rates lower
  - And it's correlation, not causation.
- Principal killers were malaria and yellow fever
- These also killed native populations, but mainly in childhood. Death rates in adult populations much lower, due to acquired immunity.
- Many studies use latitude as measure for endemic disease.

What if endemic disease is the determinant?

TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

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

Hypothesis 1: Absent asymmetric information, relationship between FICO & default



Source: Dr. Debrah Noe

Regression discontinuity approach

* **Hard to compare borrowers with very different FICO scores.**

***However, borrowers with slightly different FICO scores should be very similar.**

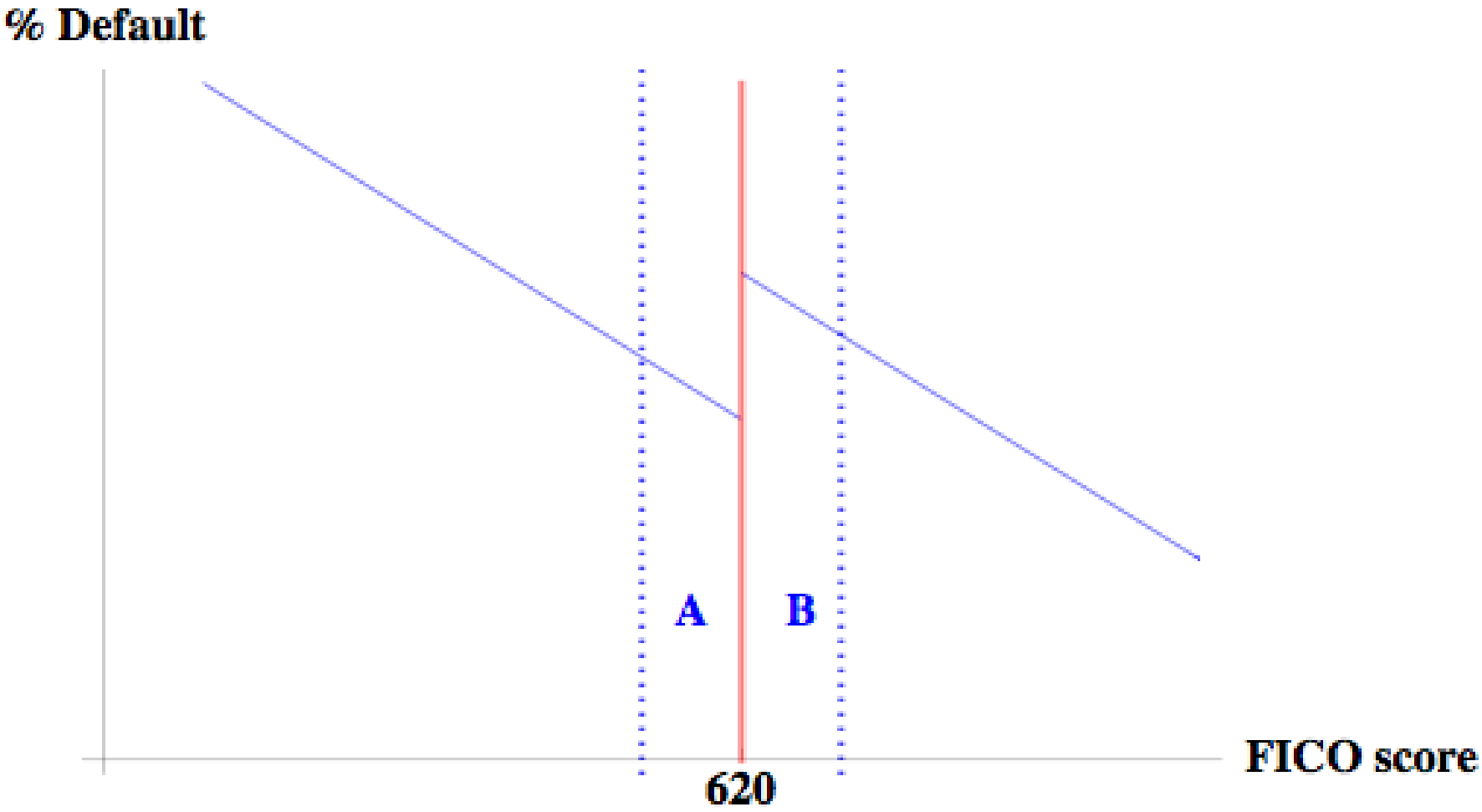
*** FICO < 620: loan not marketable & bank CAN'T sell lemons to 3rd parties**

*** FICO > 620: loan marketable & bank CAN sell lemon loans**

***If banks sell 'lemons' the % of default will jump up at the 620 FICO cutoff**

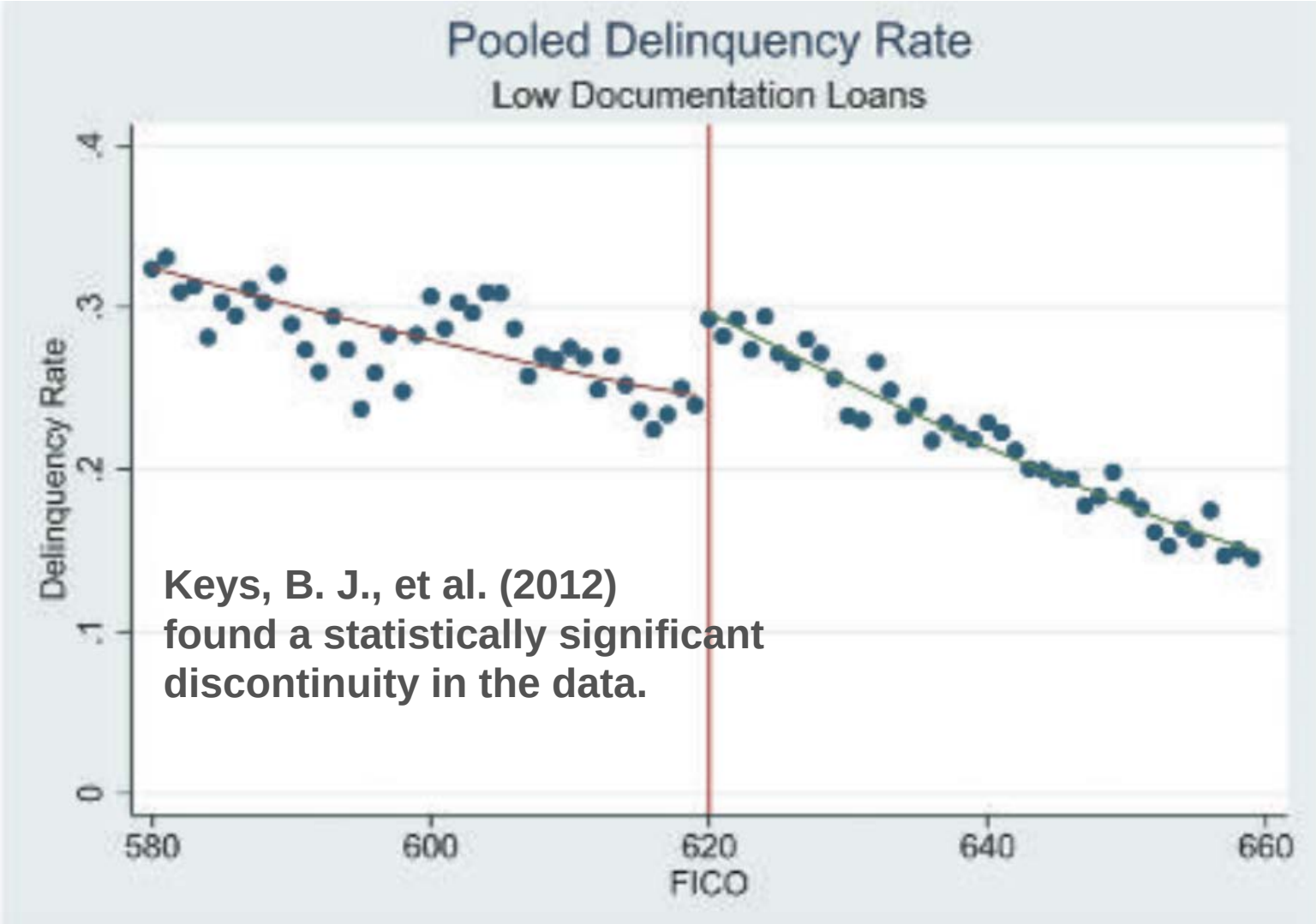
Source: Dr. Debrah Noe

Hypothesis 2: Asymmetric information: relationship between FICO & default



Source: Dr. Debrah Noe

Keys et al. (2012)



Compare two groups over time

- Treatment and control groups not the same, and subjects not usually randomly assigned to either.
- So difference between treatment and control group statistics (e.g. means) do not measure treatment effect, for all the usual reasons.
- However, if changes over time in treatment and control group would have been similar in absence of treatment of treated, then
  - Change in treatment group characteristics over time versus changes in control group over time measure treatment effect
    - Crucial assumption is ‘parallel trends’: without treatment, two groups would have changed by the same amount.
  - I.e. The **difference** in difference measures treatment effect.

Can be more powerful than previous slide suggests

- Compare:
  1. Springfield first got access to dial-up internet in March 1990. Forest Creek didn't get it until March 1992.
     - Divorce rate in Springfield in period March 1988-February 1990 was 23% per annum. For Forest Creek it was 18%.
     - Divorce rate in Springfield in period March 1990-February 1992 was 35% per annum. For Forest Creek it was unchanged at 18%.
     - Presumably Springfield didn't get the internet two years earlier than Forest Creek because more people in Springfield wanted to get divorced.
     - Check divorces rates in Springfield and Forest Creek moved in parallel prior to 1988 and after 1992.

Or not

- Compare:
  1. Springfield first got access to dial-up internet in March 1990. Forest Creek didn't get it until March 1992.
  2. CDS contracts written on British Airways bonds became centrally clearable in February 2014. CDS contracts on Scottish Power remained over-the-counter until 2018.
     - Both contracts had spreads of 1.5% over LIBOR before February 2014
     - After February 2014, BA CDS spreads fell to 1%, SP spreads unchanged.
     - Presumably BA contracts became centrally cleared first because more people *did* want to insure against default by BA than SP.
     - Check CDS spreads on BA and SP moved in parallel before 2014.

Effect on years of school and wages in 1995

‘In both [High and Low intensity] regions average educational attainment increased over time. However it increased by more in regions that received more schools. The difference in these differences can be interpreted as the causal effect of the program, under the assumption that in the absence of the program, the increase in educational attainment would not have been systematically different in low and high program regions. An individual young enough, born in a high program region, received on average 0.12 more years of education, and the logarithm of his wage in 1995 was 0.026 higher.’