1. Introduction

Econometrics II
Winter 2020
Osaka U

Shuhei Kitamura

About the course

Econometrics II (B & M)

- Date & time: Monday 2nd, Wednesday 1st
- Location

  (onsite): Hōkei 3
  (online): Blackboard Collaborate Ultra
- Language: Japanese (lecture materials in English)
- Instructor: Shuhei KITAMURA (kitamura@osipp.osaka-u.ac.jp)
- Office hours: By appointment
- TA: Mayuko ABE (u656691k@ecs.osaka-u.ac.jp)
- Office hours: Tuesday 9-10 (Zoom ID: 886 872 0129, password: osipp_econ)

About the course (cont.)

Objective: Learn together how to properly conduct causal analysis using modern econometric techniques

In particular, your are expected to:

- Learn the basic knowledge and techniques of causal analysis
- *Critically* apply them to conduct a rigorous analysis

What I mean by "critically" is that you are able to:

- Detect endogeneity problems
- Argue how to alleviate them (and possibly implement it)

Intuition is more important than mathematical explanations

- There will be mathematical explanations too

You will learn R to conduct analysis

About the class (cont.)

Prerequisites

- Econometrics I by Prof. Kohara, and/or
- Econometric Methods I by Prof. Matsubayashi

Use your laptop to run R code

- We will install them on your computer at the end of this lecture

No textbook. This course content is based on the following books:

- Angrist, J. D. and J.-S. Pischke (2015). *Mastering 'Metrics The Path From Cause to Effect*. Princeton: Princeton University Press.
- Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach*. Boston: Cengage Learning.
- (Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.)

R beginner's guide:

- Using R for Introductory Econometrics
- Course materials from my data management course

About the class (cont.)

Course materials and important notices are available on CLE

Outline of the course

1. Introduction
2. Treatment effects
   - Randomized Control Trial (RCT)
   - Regression-control analysis
3. Instrumental Variables (IV)
4. Regression Discontinuity Design (RDD)
5. Difference-in-Differences (DID)
   - Fixed effects models
6. Propensity Score Matching (PSM)
   - Probit and Logit models
7. Clustered standard errors

Grading

Five assignments (40%)

Written exam (60%)

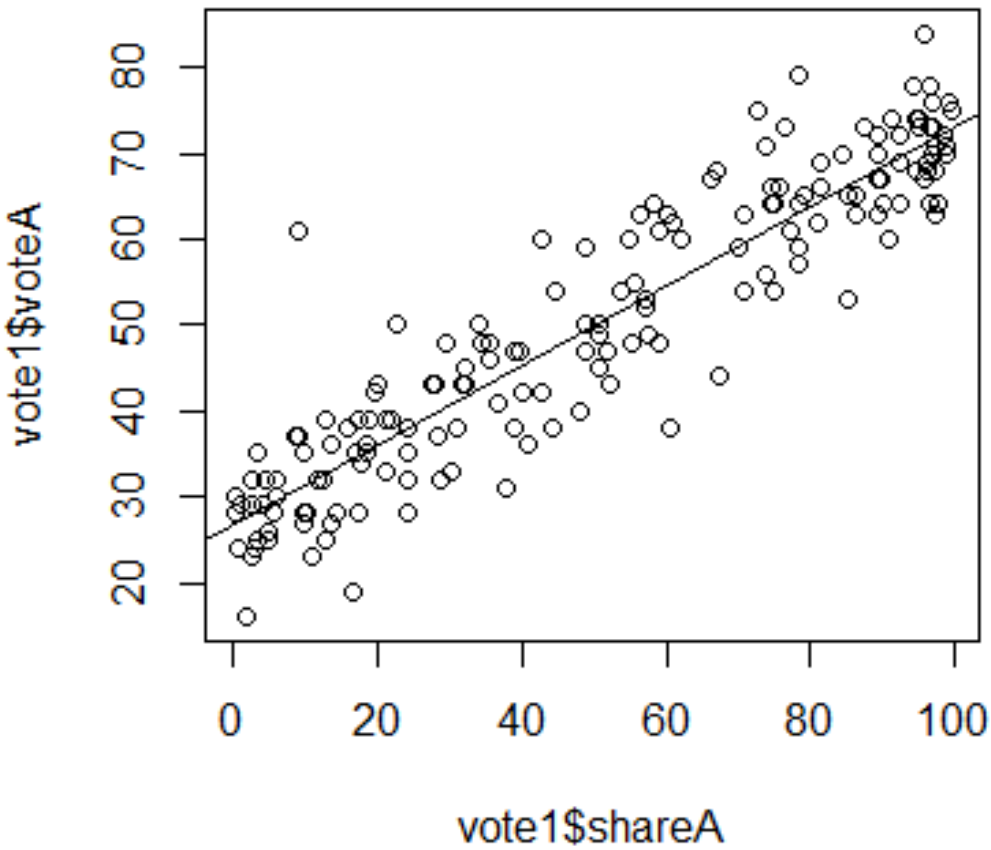
Causation vs. correlation

Let's start from distinguishing causation from correlation

- Causation: $X \rightarrow Y$
- Correlation: There is some relationship between X and Y

An example: X is the percentage of total campaign expenditure and Y is the vote share

Voting outcomes and campaign expenditures



Data: U.S. House of Representatives 1988.

Voting outcomes and campaign expenditures (cont.)

Using the data of the U.S. House of Representatives, we get the positive and significant relationship between the vote share and the share of campaign expenditures

$$\text{vote}_i = 26.81 + 0.46\text{exp\_share}_i, \tag{1}$$

Where i is a particular politician.

Suppose you are a consultant agent for a particular political candidate, and are writing a report to him/her.

Can you write to him/her "if you increase the share of campaign expenditures by one percentage point, vote share will increase by 0.46 percentage point"?

Causation vs. correlation (cont.)

Practitioners want to know a causal effect.

- Does increasing campaign expenditures increase the vote share? If so, how much?
- Does tax cut improve firm performance? If so, how much?
- Does the TV advertisement of particular goods increase the sales of those goods? If so, how much?
- Does distributing mosquito nets decrease mortality rates? If so, how much?

A causal effect is about the quantity, while a causal relationship is about the relationship.

- You can at least pin down a causal effect if you properly apply the econometric method.

A causal effect is the effect of X on an outcome Y as measured in an ideal setting.

- We will learn what the *ideal* settings are.

Endogeneity problem

What kind of endogeneity problems can you spot in the electoral campaign-voting example?

- What is the endogeneity problem?

There are several types of endogeneity problems.

Let's see them using examples.

Case I: Institutions and GDP per capita

Societies with a social organization that provides encouragement for investment will prosper (Locke, Smith, Hayek, etc.).

- Developed countries have become rich because they have good institutions such as secure property rights and checks and balances.
- Is this true?

Suppose you run the following regression:

$$\text{LogGDPpc}_c = \alpha + \beta \text{QualityOfInstitutions}_c + \varepsilon_c, \qquad (2)$$

Where LogGDPpc_c is log GDP per capita and $\text{QualityOfInstitutions}_c$ is the quality of institutions as measured by the average expropriation risk.

Case I: Institutions and GDP per capita (cont.)

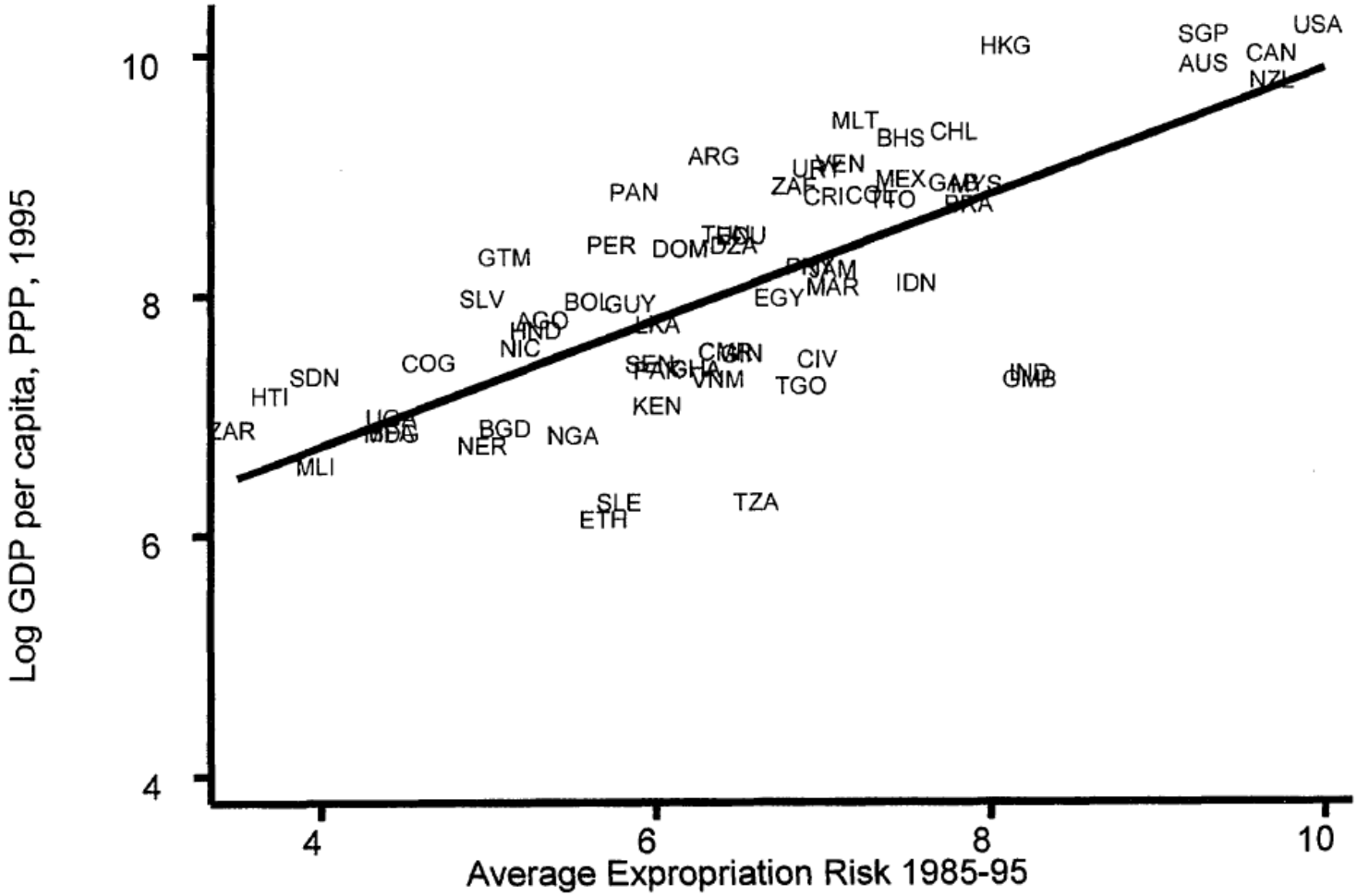


FIGURE 2. OLS RELATIONSHIP BETWEEN EXPROPRIATION RISK AND INCOME

Source: Acemoglu et al. (2001).

Case I: Institutions and GDP per capita (cont.)

Based on this regression analysis, can you conclude that better institutions *do* increase GDP per capita?

Case I: Institutions and GDP per capita (cont.)

There are several endogeneity problems:

- Reverse causality/simultaneity: Rich economies might be able to afford better institutions
- Omitted-variable bias: Unobserved factors might affect both institutions and economic performance
- Measurement error: Institutions variable might be measured with considerable errors

Case II: Advertisements and sales

Suppose you are a data analyst working in a private company.

You find a positive relationship between an advertisement of an ice cream product and its sales.

Can you say "the advertisement increased the ice cream product's sales"?

What are potential endogeneity issues here?

- Reverse causality?
- Omitted-variable bias?
- Measurement error?

Case III: Hospitals and health outcomes

Consider the question: "Does hospitalization make people healthier?"

Let's use the following survey question in the 2005 National Health Interview Survey (NHIS) to find an answer:

- "Would you say your health in general is excellent, very good, good, fair, or poor?"

The answer for the question takes a value between 1 (poor health) and 5 (excellent health).

Comparing those who have been hospitalized and have not gives:

	Sample size	Mean health status
Hospital	7,774	2.07
No hospital	90,049	2.79

Case III: Hospital and health outcomes (cont.)

The results imply that if you run the following regression:

$$\text{Health}_i = \alpha + \beta \text{Hospitalized}_i + \varepsilon_i, \tag{3}$$

You should get a negative significant estimate of β (≈ -0.72).

Does this mean that going to a hospital makes people sicker?

What are potential endogeneity issues here?

Hospital and health outcomes

This is an example of selection bias, yet another endogeneity issue.

- Sick patients are *selected* into the treatment group (in this case, individuals who are hospitalized).

True effects are likely to be positive, but since selection bias is so big, you observe a negative effect.

- Effects you observe $(-)$ = True effects $(+)$ + Selection bias $(-)$

Important: Selection bias $\neq$ sample selection

- What is sample selection?

Case IV: Mosquito nets and malaria infection rates

Suppose you are working in a country in Sub-Saharan Africa as an officer of an international NGO which distributes mosquito nets in villages with high malaria infection rates.

After giving mosquito nets, you find that mortality rates are higher in villages where the NGO distributed the nets than other villages.

Do you conclude that mosquito nets increase mortality rates?

What are potential endogeneity issues here?

How to get a causal effect: Intuition

In order to get a causal effect, you should get rid of the endogeneity problem.

How to solve the problem?

You need to compare similar groups of which some of them are treated (getting good institutions, advertisements, medical care, mosquito nets, etc.) while the others are not. = An ideal setting

- The former group is often called a treatment group, while the latter group is called a control group.

To make treatment and control groups very similar, you need to (quasi-)randomly divides individuals into groups.

How to get a causal effect: In practice

Use experimental data (field or lab experiment)

- Individual units (e.g., people) are randomly assigned to either the treatment or control group.
- Can be expensive, time consuming, or ethically impossible

Use exogenous variation (natural experiment)

- You rely on natural events that randomly assign individual units to either group.
- Not always as clean as experimental data

Use observational data

- You rely heavily on econometric techniques

$\rightarrow$ You will learn all cases in this course.

Endogeneity problem (revisited)

What kind of endogeneity problems can you spot in the electoral campaign-voting example?

How can you possibly alleviate them?

Whenever you see an empirical statement (e.g., an advertisement of an ice cream product increases its sales), think critically about the credibility of that statement (e.g., is there any endogeneity?) and possible solutions (e.g., what about using a natural experiment?)

Summary

Correlation doesn't imply causation

Endogeneity problem

- Reverse causality/simultaneity
- Omitted-variable bias
- Measurement error
- Selection bias