Putting it all together

Paul Schrimpf

UBC
Economics 326

March 27, 2017
Overview

Step 0: ?

Step 1: before looking at data

Step 2: data
  Updated data
  Canada

Step 3: OLS

Step 4: IV
References

- Canada: Bunge et al. (2005)
Section 1

Overview
Plan

- This week and next: how to use the methods we’ve studied
- This week: walk-through research project
- Step-by-step:
  - Discuss what to do in groups
  - Decide what to do as a class
  - I’ll implement it (and you can too if you bring a computer)
  - Discuss results in groups and all together
  - Repeat
Section 2

Step 0: ?
We want to start a research project, but we’re not sure what it should be about. How do we decide?
Choosing a topic

- Ideal: choose something you’re passionate about
  - If you think it is important, you can convince others it is
  - Could be:
    - Policy relevant
    - “Deep” economic issue
    - Just something neat

- Okay, but not ideal: Build on what you know
  - Past research, other courses, other knowledge

- Common, but not something anyone admits to: Have a hammer, look for a nail
  - Version 1: have some data, think of a question the data can answer
  - Version 2: take a favorite method and find a problem for it
Section 3

Step 1: before looking at data
We’ve chosen a (vague) topic: crime and police. What should we do before we look at any data or even gather data?
Step 1

We’ve chosen a (vague) topic: crime and police. What should we do before we look at any data or even gather data?

Most important: clearly state the research question.
Define the question

- Question: what is the effect of changing the number of police officers on the crime rate?
  - Ideal experiment: randomly assign different numbers of police officers to different areas and compare the crime rates.
  - Economic theory: commit crime if

\[
\mathbb{E}[U(\text{crime})] > \mathbb{E}[U(\text{no crime})]
\]

\[
U(\text{crime}) + P(\text{arrested and convicted})U(\text{jail}) > \mathbb{E}[U(\text{no crime})]
\]

More police decrease crime by increasing \(P(\text{arrested and convicted})\), but how much and how much it matters to criminals is an empirical question.
Section 4

Step 2: data
Step 2: data

What data do we need?
Data summary

- Crime and number of police officers: FBI *Uniform Crime Reports*
  - Crime rates per 100k for murder, rape, robbery, assault, burglary, larceny, auto theft
  - 59 U.S. cities 1960-1992

- City demographics and economic conditions (potentially affect $U(\text{crime})$ and $U(\text{no crime})$)
  - City level population, portion black, portion female, portion various age ranges (some of these are interpolated, some are SMSA instead of city)
  - State unemployment, income per capita, education spending, welfare spending

- Political indicators:
  - Mayoral and gubernatorial election years
  - Term of mayors
Step 2: data

We have gathered our data. What is the first thing we should do with it?
Step 2: data

We have gathered our data. What is the first thing we should do with it?

We should check that it makes sense. Let’s look at some tables and figures. What tables and figures should we create?
Step 2: data

We have gathered our data. What is the first thing we should do with it?

We should check that it makes sense. Let’s look at some tables and figures. What tables and figures should we create?

Which if any of these tables and figures should we include in our table and/or presentation?
Interactive figures:
- http://faculty.arts.ubc.ca/pschrimpf/326/crime-rate.html
- http://faculty.arts.ubc.ca/pschrimpf/326/crime-scatter.html
- Code

Map:
- http://faculty.arts.ubc.ca/pschrimpf/326/crime-map.html
- Code
Updated data

- Have not yet downloaded newer data on number of police officers, income, elections, demographics
- Interactive figure:
  - Code
Canadian data

- We will not use this for further analysis because we do not have the other variables for Canada (but you could get them)
- Code
Canadian data

**All**

Year | Total
--- | ---
1980 | 5000
1990 | 6000
2000 | 7000
2010 | 8000

**Homicide**

Year | Total
--- | ---
1980 | 1.5
1990 | 2.0
2000 | 2.5
2010 | 3.0

**Violent**

Year | Total
--- | ---
1980 | 700
1990 | 1000
2000 | 1250
2010 | 1500

**Property**

Year | Total
--- | ---
1980 | 3000
1990 | 4000
2000 | 5000
2010 | 6000
Canadian data

Drug

Cannabis Trafficking

Heroin Trafficking

Cocaine Trafficking
Section 5

Step 3: OLS
Step 3: OLS - specification

Let’s use OLS to examine the relationship between crime and police. How should we specify the regression(s)? What is the dependent variable(s)? What is the regressor of interest? What controls should we include? Should we use the log of any of the variables?
How can we interpret the OLS estimates? Does the coefficient of interest have the expected sign? Is it small or large?

Do we think OLS answers our question? What variables might be omitted? What is the likely sign of the bias?
Step 3: OLS - inference

How should we calculate standard errors? What hypothesis(es) should we test?
Section 6

Step 4: IV
Step 4: IV

We need an instrument for police. That is, we need something that is correlated with the number of police but has no direct relationship with crime. What could be a good instrument?
We need an instrument for police. That is, we need something that is correlated with the number of police but has no direct relationship with crime. What could be a good instrument?

We will use mayor and governor elections as an instrument for the number of police. What is the rationale for this instrument? Are there any reasons to think it is not exogenous?

http://www.youtube.com/watch?v=lJbRVX1pFEc
Step 4: 2SLS - specification

How should we specify the 2SLS estimator? What is the dependent variable(s)? What is the endogenous regressor of interest? What is the instrument? What controls should we include?

What is the first stage? What is the reduced form? What should we check in these regressions?
Step 4: 2SLS - interpretation and inference

How can we interpret the estimates? Does the coefficient of interest have the expected sign? Is it small or large?

How should we calculate standard errors? What hypothesis(es) should we test?
Step 4: 2SLS - threats to validity

Why might elections not be a valid instrument? Is there anything we can check to reassure us that elections are exogenous?

http://www.youtube.com/watch?v=xH_6_8N0fwI
Assuming we have correctly estimated the effect of police on crime, what could we do with our estimate? What is a relevant policy question? What else do we need to know to answer that question?
Code

- Code for main results
- Code for calculating heteroskedasticity robust clustered standard errors


References


