A heroic attempt to describe best practices for

“Natural” “Experiments” as (one) cure for (some types of) Endogeneity

In only 15 minutes

Victor Bennett
Duke and USC
Endogeneity is a big problem...

• ...in that the word “endogeneity” includes many smaller problems
  – Omitted variables
  – Simultaneous causation
  – Measurement error

• ...and those can make a liar out of you
  – OLS coefficients are unbiased if
    • Model is the right one
    • X is full rank
    • Strict exogeneity

So you can’t even trust correlations, let alone claim causation
So how do I fix it?

That depends on what exactly the problem is!

Editors and reviewers: Please be specific about what problem you want fixed

Authors: Please be specific about the problem you’re trying to solve

A suggestion: Explicitly state your ideal experiment
An example: Bennett, Seamans, Zhu (*Forthcoming, SMJ*)

Make customers willing to pay more knowing they can resell

Competes with news goods

What I’d like to do (but can’t):

$$\text{Price}_{it} = \alpha + \beta \text{ (resellable}_{it}) + \varepsilon_{it}$$
Estimating causal impact of resale market

• Need a setting where resale market exists for reasons unrelated to price

• *Ideal experiment:*

  For a set of identical items, randomly make some of them resellable

  If randomly selected, but not identical:
  • Controls
  • Matching
  • Synthetic controls

  If not randomly selected generally, see if there is a setting where they are

Word of warning: there generally isn’t one. Better for “Will I” than “Do I” have an endogeneity problem
Our “experiment”: Craigslist entry on concert tickets

• Random?
  – Of course...not

So how is this an “experiment?”

Just has to be unrelated to the *specific problem you are trying to solve*

*Most important task is to convince us of that*
From example to exemplar

One of the fundamental questions in development economics is why we see persistent development differences among seeming similar cities.

One proposed mechanism:
Have to be physically close enough to other cities to make use of networks.

What I’d like to do (but can’t):
Development of city_{it} = \alpha + \beta (close to another city_{it}) + \varepsilon_{it}
• **Ideal experiment:**
  
  For a set of identical cities, randomly drop a city next to it

---

What if division was determined based on development rate?

Good etiquette: A graph showing pre- and post-treatment trends

Doesn’t look like it
What if something else going along with treatment caused effect?

- Close to border = Fear of war
  - City-level data on rubble per capita
  - Percentage of dwellings destroyed
  - Count of refugees from DDR

- Far from border = Closer to Netherlands
  - Interacted measure with distance from western border

- Etc.
Lessons

• Need to know threats to fix them
• Need to convince reader threats you’re fixing are the right ones
• With a natural experiment, all the work is in building confidence in the experiment
  – Perfect experiment needs no controls, let alone fancy stuff
So how do I get one?

Steal it!

What strategy questions depend on endogenous geography?

Innovation spillovers – Do proximate firms cite you because they’re near, or are they near because they do similar things?
Innovation spillovers – Do proximate firms cite you because they’re near, or are they near because they do similar things?
Take these “experiments”...please!

• Access to labor
  – Mariel Boatlift (Card 1990), Rider and Tan (Forthcoming)

• Competition
  – China Syndrome: (Autor, Dorn, Hanson 2013), etc.

• Customers served
  – Distance: Rubin (2008), Bennett (2010, unpublished)

• Status and reputation

• Networks
  – Access: Catalini (unpublished), September 11, Visa Limits