Non-Random Exposure to Exogeneous Shocks: Theory and Applications

Paul Goldsmith-Pinkham
Yale SOM

October 1, 2021

- Goal today: present paper like I would teach it
- Consequence: some reordering, and reframing
- Criticism (constructive or otherwise) will be limited – this paper is R&R at ECMA!
- Definitely worth knowing this paper!

Economists have a nose for randomness

- Paraphrasing one of my statistics colleagues:

  Economists are really good at doing almost the right thing in empirical work.

  -Anonymous Yale Professor

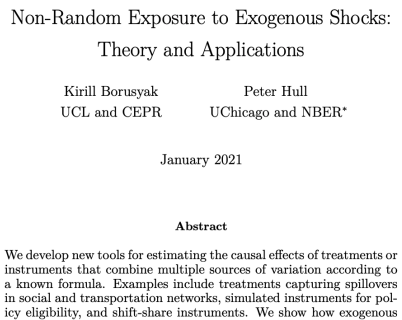
- Economists are clever at finding things that look convincingly "random"
  - Sometimes, it is easy to know how to use this randomness
  - **This paper is about when it is hard**

What is in this paper?

- Two key parts to this paper:
  1. Highlighting how seemingly complicated research designs can be framed as generalized propensity scores
  2. How complicated research designs can suffer from *interference*

- There are many interesting results that spiral out from these two insights, but these are the key kernels (third piece is thinking about uncertainty using randomization inference, but deeply tied to other pieces))

- Will first start with showing how complicated research designs $\rightarrow$ propensity scores
  - But first – what's a research design?



Non-Random Exposure to Exogenous Shocks: Theory and Applications

Kirill Borusyak, UCL and CEPR
Peter Hull, UChicago and NBER*

January 2021

Abstract

We develop new tools for estimating the causal effects of treatments or instruments that combine multiple sources of variation according to a known formula. Examples include treatments capturing spillovers in social and transportation networks, simulated instruments for policy eligibility, and shift-share instruments. We show how exogenous shocks to some, but not all, determinants of such variables can be leveraged while avoiding omitted variables bias. Our solution involves specifying counterfactual shocks that may as well have been realized and adjusting for a summary measure of non-randomness in shock exposure: the average treatment (or instrument) across such counterfactuals. We further show how to use shock counterfactuals for valid finite-sample inference, and characterize the valid instruments that are asymptotically efficient. We apply this framework to address bias when estimating employment effects of market access growth from Chinese high-speed rail construction, and to boost power when estimating coverage effects of expanded Medicaid eligibility.

Paul Goldsmith-Pinkham's definition of Research Design

- Research design shows up 69 times in Angrist and Pischke's JEP piece on the credibility revolution, but not defined
- A *(causal) research design* is a statistical and/or economic statement of how an empirical research paper will estimate a relationship between two (or more) variables that is causal in nature – X causing Y.
- The design should have a description for how some variation in X is either caused by or approximated by a randomized experiment.

Research designs from simple to complex

- Consider the trivial research design, following an RCT that randomly assigns $x_i \in \{0, 1\}$, and we want to estimate the effect of x_i on y_i:

$$y_i = \alpha + x_i\beta + \epsilon_i$$

- The research design is effectively a coin flip: $E(x_i) = p$, and each x_i is independent for each i
  - β is identified thanks to this coin flip design

- This is true even when we have covariates, w_i that stratify the experiment. We just need to control for w_i correctly: $E(x_i|w_i) = p(w_i)$ and we can estimate the ATE directly

- Effectively, the (potentially) endogenous w_i affects treatment, but if we condition correctly, we can still identify a causal effect

Research designs from simple to complex- Medicaid eligibility

- Now imagine the eligibility rules for Medicaid were being randomly assigned
  - Drawn from a bag just like marbles, completely randomly
- We can now estimate the effect of Medicaid eligibility on things like child mortality
  - Issue: eligibility is also a function of many *endogenous* features
- We consider a known function, f_i, and eligibility rules, g_i, such that $x_i = f(g_i, w_i)$ maps the w_i characteristics using the randomly drawn eligilibility rules
  - Much like w_i strata case, but more complex b/c can be high-dimensional / non-linear

Simulated instruments as a way to get a handle on this

- The challenge is that g_i is a complicated variable – it is a set of rules of that potentially complicated and hard to map to an "instrument" or "treatment"
- You don't want to just use x_i because it contains endogenous w_i
- Currie and Gruber (1996) solution: construct a variable $z_i = \sum_j f(g_i, w_j)$ which takes w from a random population (outside the state) and uses it to construct a "predicted" x
  - Intuitively, hold fixed the g_i and average over some distribution of w_j

Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women

Janet Currie

University of California, Los Angeles and National Bureau of Economic Research

Jonathan Gruber

Massachusetts Institute of Technology and National Bureau of Economic Research

A key question for health care reform in the United States is whether expanded health insurance eligibility will lead to improvements in health outcomes. We address this question in the context of the dramatic changes in Medicaid eligibility for pregnant women that took place between 1979 and 1992. We build a detailed simulation model of each state's Medicaid policy during this era and use this model to estimate (1) the effect of changes in the rules on the frac-

Simulated instruments as a way to get a handle on this

- The challenge is that g_i is a complicated variable – it is a set of rules of that potentially complicated and hard to map to an "instrument" or "treatment"
- You don't want to just use x_i because it contains endogenous w_i
- Currie and Gruber (1996) solution: construct a variable $z_i = \sum_j f(g_i, w_j)$ which takes w from a random population (outside the state) and uses it to construct a "predicted" x
  - Intuitively, hold fixed the g_i and average over some distribution of w_j

To the extent that relevant state- and year-specific characteristics are not captured by state and year dummies (i.e., they are not constant within a state or across states within a year), the coefficient on the fraction eligible will be biased by omitted variables. Suppose, for example, that a state recession is associated with both increases in eligibility and a higher incidence of low birth weight. Then this source of variation in eligibility could induce a spurious positive correlation between Medicaid eligibility and low birth weight.

In order to overcome this potential problem, we instrument the actual fraction eligible with a measure of the generosity of Medicaid in a state and year that depends only on the state's eligibility rules. To create our instrument, which we label the "simulated fraction eligible," we first take a sample of 3,000 women from the CPS in each year. We then calculate the fraction of this sample of women who would be eligible for Medicaid in each state. By using the same group of women in each state simulation, we obtain an estimate of the fraction eligible that depends only on the legislative environment and is independent of other characteristics of states. This measure can be thought of as a convenient parameterization of legislative differences affecting women in different states and years: the generosity of state Medicaid policy can be naturally summarized in terms of the effect it would have on a given, nationally representative, population. Furthermore, we reduce the sampling variability in our estimates that derives from having relatively small cells for some states in the CPS.[9]

This paper's approach vs. simulated instrument

- This is not the most efficient way to exploit this variation
- Remember our propensity score example: if we could just condition directly on w_i, then we would not worry about endogeneity
  - The solution, then, is to construct a propensity score and condition on that!
- Intuitively, "the eligibility rules for Medicaid were being randomly assigned"
  - In other words, we assert a counterfactual distribution over the policy rules $Pr(g)$
  - This allows us to construct the propensity score for a given individual

$$p(w_i) = Pr(x_i|w_i) = \sum_g f_i(g, w_i) Pr(g)$$

- With pscore in hand, estimation is straightforward, and known to be semiparametrically efficient!

Recentering vs. Controlling?

- The paper makes a big point about the recentering concept
  - I'm not sure why one would do this vs. just controlling
- This is particularly true in this setting (rather than with interference)
- My suggestion when reading the paper: just think of this as conditioning on propensity scores
  - Can reivsit this under interference

The return on propensity scores in an empirical example

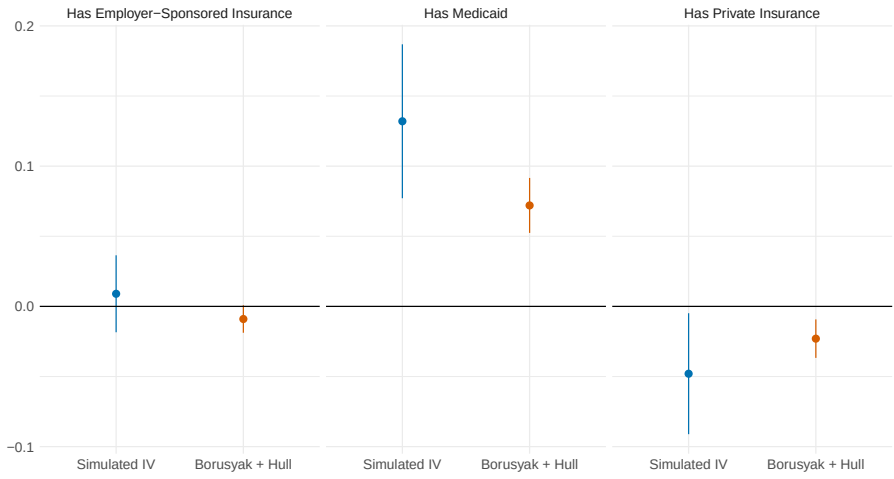
- Medicaid Empirical example in this paper: ACA medicaid expansion
- ACA expanded Medicaid in only some states thanks to NFIB v. Seblius allowing choice by states
- Interested in understanding compositional shifts in health care across states
  - Use ACS micro data and consider structural equation

$$y_{it} = x_{it}\beta + \alpha_{s(i)} + \alpha_t + \epsilon_{it}$$

  - y_{it} are different health insurance take-up; x_{it} is Medicaid eligibility for individual i
- "Simulated" IV: dummy for whether state expanded Medicaid z_{sim}
- Borusyak and Hull IV: construct a person-level indicator for whether a person is eligible under their state's law z_{bh}
  - Also identify the $p(w)$ that they are eligible on average across others states' laws
  - They recenter ($\tilde{z}_{bh} = z_{bh} - p(w)$ but you could just control...

The return on propensity scores in an empirical example

- **Much** more precise
- Makes sense!
- Seems like we should use it...



Discussion: When beautiful theory meets data roadblocks

- Many times, the data necessary to calculate eligibility and the outcomes of interest are not in the same dataset. This is especially true in simulated IV settings (I know from experience)
  - Appendix D.2 talks about these issues but it's a little vague
- Clarify ideas: Cohodes et al. (2016) consider effects of Medicaid in the 1980s on long-term education outcomes
  - Parental income (the relevant w_i for eligbility during childhood) is not known in the same dataset for y_i.
  - x_i is average eligibility for types born in state s at age a of race r in period t.
  - They just want the variation that comes from state laws, not the demographics
  - Construct Sim. IV that takes the average share of individuals in national population that would be eligible under states' laws
- Hopefully you're getting it now! Proposal from paper:
  - Construct $Pr(x_i|\tilde{w})$ that randomizes over *states* and control (or recenter)
- Note: this type of within-state demographic info is actually used in Mahoney (2015) for bankruptcy simulated instruments!

Discussion: Heterogeneity in the underlying data

- Something harder to take from the paper
  - How to consider aggregation issues
- State-level variation, but maybe some individuals experienced Medicaid expansion, while other experienced contraction
  - Less likely in Medicaid case, but possible in other settings
- E.g. monotonicity violation when there is heterogeneous treatment
  - Paper discusses some points on this in Appendix, but would be useful to be a bit more concrete in some examples
  - Something I tried to work on with Aronow and Sorkin but ran into serious data issues!

Second kernel of the paper: interference

- Medicaid example is simple to think about, and clarifies idea that:
  1. Can convert high-dimensional variation into simple treatment effects
  2. Can be more *efficient* (e.g. smaller s.e.)

- However, you can take this much further.

- Consider the design of a railroad. Imagine the world in which a railroad designer randomly threw darts on a map to decide where to construct train lines
  - Similar to the analogy of "drawing" the Medicaid eligibility rules
  - But now, how do we think about the "random" piece interacting with different places?
  - Let's start with something simple first

Interference in network settings

- Consider a setting where the researcher want to measure the impact of a randomized experiment on a network
  - In other words, for a given person i, and observed network W, we randomly treat some subset of individuals on the network.
  - We want to know what the effect of having more treated individuals connected to you x_i is on y_i
- Insight from paper: since the position in *network* affects probability of being connected to individuals, some individuals will inherently get more exposure!
  - Analogous to the friendship paradox
- Need to construct an analogous propensity score for the network setting, and control for that
  - Since we have a true RCT, this is not too hard!
  - (But we do have to make decisions about what the spillover is)
  - Aronow and Samii (2017) made serious progress on the network context

Interference in spatial settings

- Things can be more confusing than an RCT, but this same insight applies
  - Even with random shocks (darts on a board), some locations / people attract more treatment than others
  - Consider the application from the paper

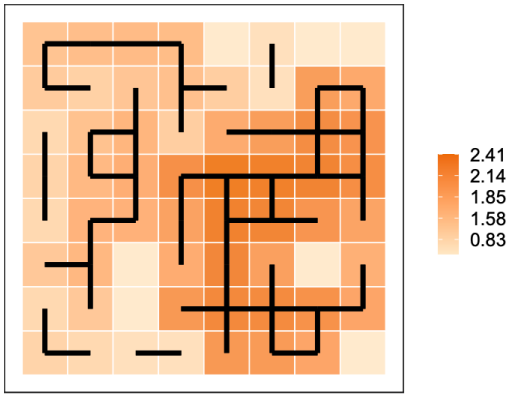
- Estimate the impact of market access growth (MA) on land values growth (V) in China
  - MA is influenced by transportation networks, and measures aggregated access to other populations

$$MA_{it} = \sum_j \tau(\mathbf{g}_t, loc_i, loc_j)^{-1} pop_j$$

- Want to estimate the effect of MA_{it} using "random" variation in network changes!
  - Can we just run the OLS? No!

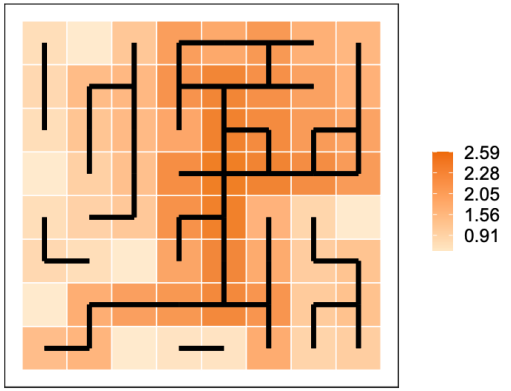
Stylized Example of Market Access on a Square Island

- Take a square with square villages and randomly assign roads
- How does market access change?



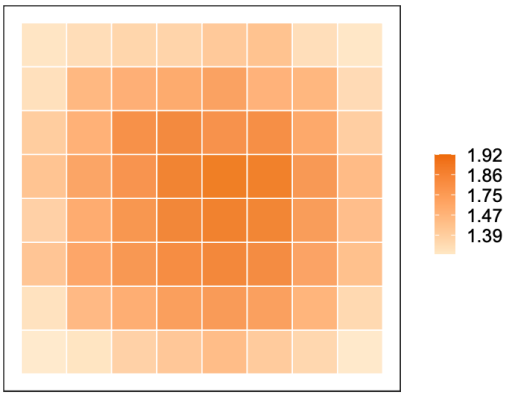
Stylized Example of Market Access on a Square Island

- Take a square with square villages and randomly assign roads
- How does market access change?
- If we rerandomize, does it look different?



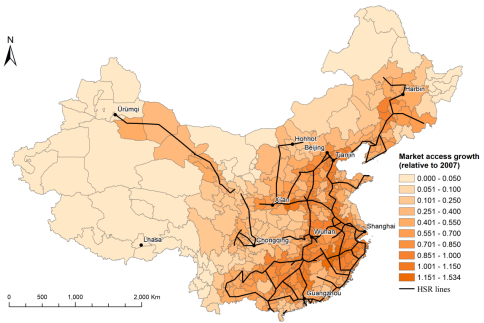
Stylized Example of Market Access on a Square Island

- Take a square with square villages and randomly assign roads
- How does market access change?
- If we rerandomize, does it look different?
- As with the network, some places get more market access than others on average!
- Need to account for this propensity difference



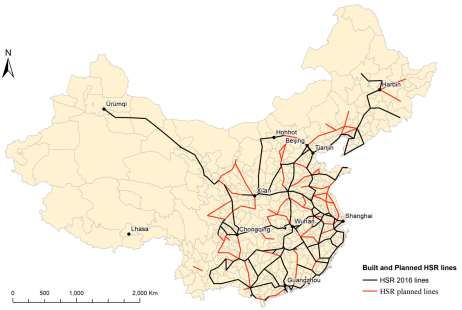
China: defining the counterfactual distribution

- In the stylized example, lines are laid randomly, making it easy to define the propensity scores
  - What about in China?
- What is the plausible counterfactual?



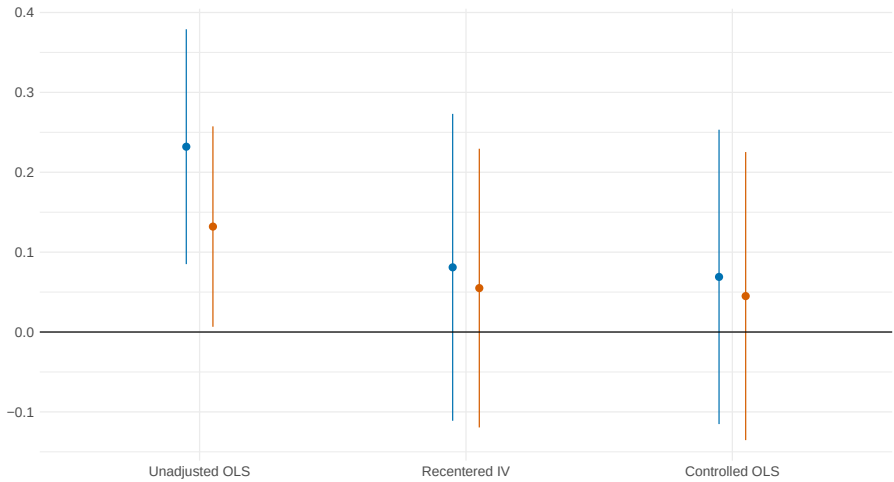
China: defining the counterfactual distribution

- In the stylized example, lines are laid randomly, making it easy to define the propensity scores
  - What about in China?
- What is the plausible counterfactual?
- Paper proposes an idea, and analagous to other examples
  - Use *planned* lines are randomized between unbuilt but planned, and built lines
  - Calculate distribution of propensity score by constructing MA_i under each counterfactual scenario



China railroads: the punchline

- There was substantial bias from using OLS!
- Makes sense – geography is king...
- No effect in randomized setting



Defining the counterfactual distribution

- If one takes issue with the counterfactuals, that is reasonable (but of course, challenging to prove one way or the other)
- Key issue: this paper is just making **text** what was already **subtext**
  - There was always an assumption about some counterfactual comparison in these designs!
- The issue is that many of these paper do not understand how to describe the randomization aspect of their research design
  - Consequentially, they cannot describe the "as-if random" component coherently
  - If a researcher has an alternative proposal, they should try that and see what estimates are available!
- Also suggests that reserchers can show a "range" of estimates under different scenarios

Key takeaways from paper

- Provide a toolbox for contexts when economists have found good "as-if" random variation (and can describe the counterfactual distribution)
- Show that in cases where treatment is not influenced by others' treatment status, approach maps very tightly with traditional propensity methods, and can be much more efficient
- In spatial and network cases where treatment spillovers exist, show how to adjust for bias arising from units location on network or graph (or relevant characteristic)