Methods for Applied Microeconomics

Sepehr Ekbatani
(TeIAS)

Spring 2022

Introduction & Objectives

Introduction to applied microeconomics and its methodologies.
Academic research in the field explored using tools from economic theory and econometrics; with emphasis on **research design** and **economic modeling**.
Topics include approaches to identification, and the economics of

- Education
- Health
- Crime
- Gender
- Urban
- Public

General Information

Email: sepehrekbatani@teias.institute

Lectures: Mon 10AM-13PM

Office Hours: Sat 3-4PM

Textbooks:

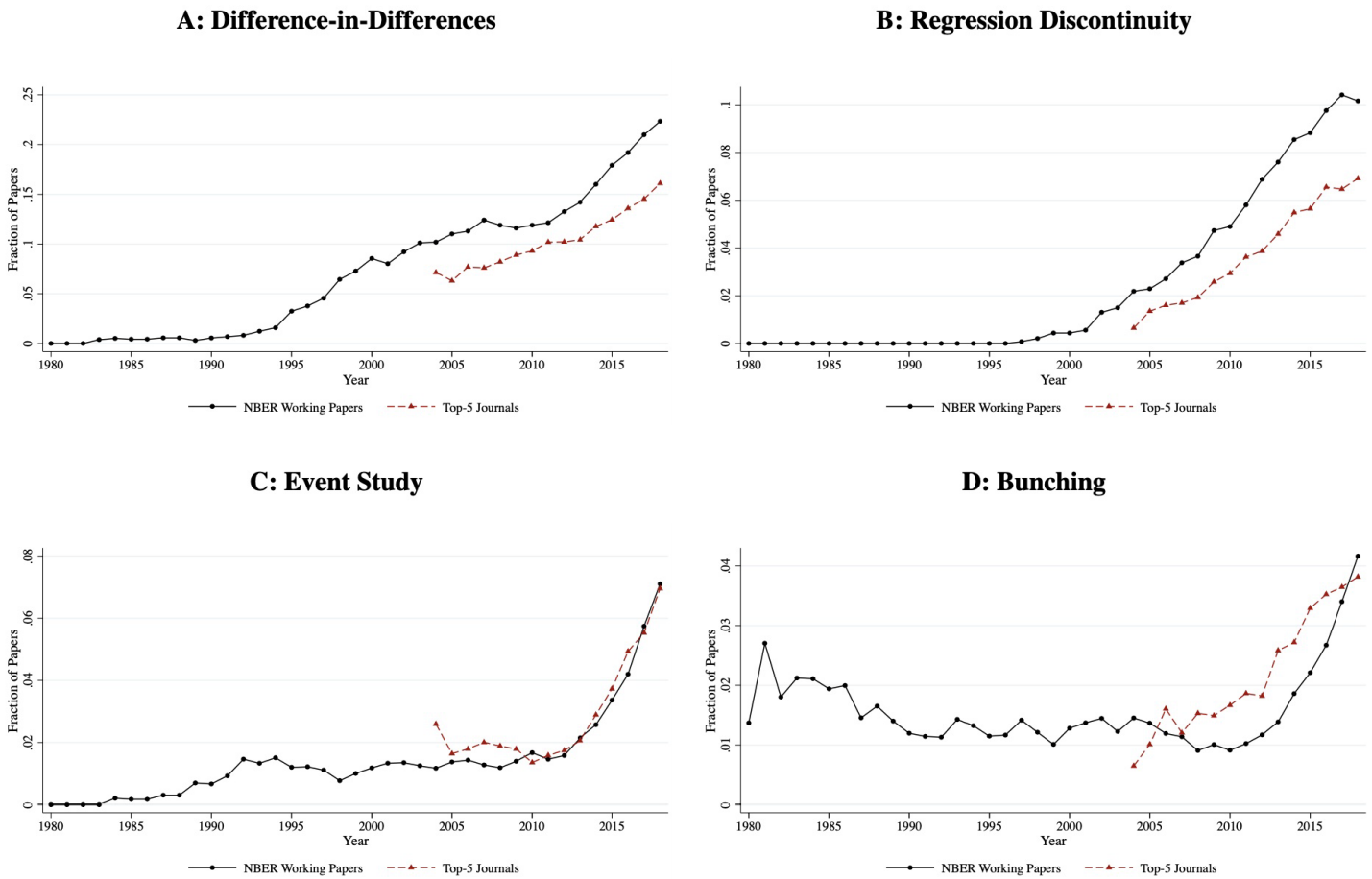
- Cunningham, Scott. "Causal Inference." The Mixtape 1 (2020).
- Angrist, Joshua D., and Jorn-Steffen Pischke. Mostly harmless econometrics. Princeton university press, 2008.

Course Outline

1. Introduction
2. Causal Inference
3. Randomized Controlled Trials
4. Instrumental Variables
5. Difference-in-Differences
6. Matching
7. Synthetic Controls
8. Regression Discontinuity
9. Event Studies
10. Choice Models
11. Weak and Bartik Instruments
12. Quantile Regressions

Sounds Boring?

Figure IV: Quasi-Experimental Methods



Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show 5-year moving averages.

Section 1

Introduction

First Words

"If applied econometrics were easy, theorists would do it. But it's not as hard as the dense pages of Econometrica might lead you to believe..."

– *Mostly Harmless Econometrics's last words*

Types of Empirical Economics?

1. Measurement
   - How much do prices change with the number of competitors?
   - What is the effect of college attendance on expected wage?
2. Model Testing
   - Is there moral hazard in auto insurance?
   - Does BNE do well in predicting bidding at oil auctions?
3. Model Estimation for Counterfactuals
   - How much would prices rise two firms merged? if the sales tax increased by 1%?
   - How would student outcomes differ under a different school choice mechanism?

Measurement vs. Model Estimation for Counterfactuals

This distinction is *usually* false.

- Most "measurement" questions in economics concern a counterfactual of the form "how would the world have been different if XXX were changed, all else fixed?"
- Returns to college: we can't just compare wages for those who attended and those who did not.

Challenges

Main challenges in answering empirical questions:

- Identification
- Causality

Identification Is Not Causality, and Vice Versa

- **Identification** means forming a unique mapping from features of data to quantities that are of interest to economists.
  - Whether/how the things we observe are capable of revealing the answers to the questions we care about.
- **Causality** is a concept defined by a *counterfactual*.
  - What would happen if certain things were changed while others were held fixed?

Remarks on Identification

- Identification cannot be defined without the notion of a true structure within a class defined by maintained hypotheses (*what we usually call a "model"*). The model may be simple or complicated, may involve economics or only hypothesized statistical relationships.
- Identification has nothing to do with a given sample or an estimator. In fact, strictly speaking it is not even about what one could learn from an infinitely large sample.

Research Design

- A **Research Design** consists of:
  - Causal question
  - Source of manipulation/identification
  - Data
  - Econometric approach
- The quality of a research design determines credibility of empirical research
  - Should also answer a question of economic or policy interest
  - Should also clearly advance the current literature on the topic
- Frequently Asked Questions (Krueger and Pischke)
  1. What is the causal relationship of interest?
  2. What is the experiment that could ideally be used to capture this causal effect?
  3. What is your identification strategy?
  4. What is your way of statistical inference?

Types of Empirical Work: A Proper Taxonomy

1. **Descriptive**: Estimate relationships between observables
   - Establish facts about the data, e.g.,
     - College grads earn 98% more per hour than others
     - Income inequality higher now than 30 years ago
     - Airline prices higher now than before merger wave
   - Facts sometimes suggest causal relationships
2. **Structural**: estimate features of a data generating process (i.e., a model) that are (assumed to be) invariant to the policy changes or other *counterfactuals* of interest
   - Estimate demand for schools $\rightarrow$ predict outcomes under a voucher system
   - Estimate model of schooling, marriage, and labor supply choices $\rightarrow$ measure specific notions of the male-female wage gap
   - Estimate demand and firm costs $\rightarrow$ predict the welfare effects of a merger.

Program Evaluation, Counterfactuals, Models

Any type of "causal inference" is always a form of **structural estimation**. It requires a set of maintained hypotheses about the world (i.e., a model) allowing one to *define* and *identify* a counterfactual quantity of interest.

TT, ATE, LATE, QTE, etc. are all precisely defined only under a well specified model of how the data are being generated.

What about Reduced Form?

Definition. A **reduced form** is a functional or stochastic mapping for which the inputs are (i) exogenous variables and (ii) unobservables ("structural errors"), and for which the outputs are endogenous variables. e.g., $Y = f(X, Z, U)$.

Where does it come from?

Formally, a reduced form is obtained by **solving a (structural) model** for each endogenous variable as a function of the exogenous observables and structural errors.

The classic example is perfectly competitive supply and demand:

$$Q = D(P, X, U_d) \text{ (demand)}$$

$$P = MC(Q, Z, U_s) \text{ (supply)}$$

Solving for equilibrium yields the reduced form relations

$$P = p(Z, X, U_s, U_d)$$

$$Q = q(Z, X, U_s, U_d).$$

Reduced Form

Reduced Form label often combined with use of IV, RD, etc. due to endogeneity/selection: e.g., "reduced-form demand model" or "reduced-form estimation of the LATE"

For this to make sense, one must mean that the model estiamted is "structural" in the econometrics sense but "reduced form" in yet another sense: one has collapsed a more complex model into a simpler representation, e.g., by-passing some underlying mechanisms. This would be a **reduced form structural model**.

But EVERY model involves collapsing a more complex world into to simpler representation! There is no coherent way to label only some models as reduced form in this alternative sense.

The Role of Economic Models in Empirical Work

We are economists, not statisticians.

- Statisticians are good at describing the data.
- Economists are good at interpreting it using formal logic:
  - given a set of maintained hypotheses, the data imply . . .
- Where do the maintained hypotheses come from? How can they be evaluated? How do we know which maintained hypotheses are useful?

Economic Models in Empirical Work

Many important questions can be answered only by exploiting economic models (vs. statistical or DAG models) to provide a logical framework for interpreting the data:

- To tell us what to look at: what are the structural features of interest for our questions?
- To define what it means to have a "valid" estimation method.
- To provide functional/probabilistic relationships that can be used to estimate of the structural features of interest; e.g.,
  - optimality conditions that relate observables to primitives
  - IV conditions (absent an experiment, what are valid instruments? this requires economic reasoning, which means at least an informal model).

The Role of Models

"all models are wrong, but some are useful." – George Box

Useful at a minimum because without a model of some kind there is typically only hand waving. Attempts to go beyond data description without a model are "not even wrong" — i.e., one cannot even define what "right" means.

The Role of Models

> *"Art is not truth. Art is a lie that makes us realize truth..."*
> – *Pablo Picasso*

The art of empirical work includes selecting a model that captures essential features for the purpose at hand and allows one to justify an interpretation of a measurement. This will involve assumptions that one could question, debate, reject, or improve upon.

But only by specifying a model can one speak coherently about whether the maintained assumptions are problematic, whether certain data allow measurement, what alternative assumptions might imply, and how science might progress.