Lecture 1 - Fundamental Hurdles

Topics in Econometrics - M2 ENS Lyon

Vincent Bagilet

2026-09-15

Introduction

Why this class?

How would you define economics research?

- To me, *de facto* delineated by:
  - A **subject**: The study of socio-economic phenomena …
  - A **framework**: … through the lens of economic theory …
  - A set of **methods** (specifically for applied economics): … using econometrics tools
- Produces theoretical models and confronts them to data
- Applied economics is concerned with **taking theory to the data** (and then informing theory)

Why do we need econometrics?

- First, how would you define metrics?
  - Originally, the application of statistical methods to economic data, through the prism of economic theory
  - Now, to me, also a set of specific methods and approaches, very much geared towards causal inference

- Metrics is somehow **the core** of applied economics
- Think of it as a **toolbox**


LIGHTYEAR
ECONOMETRICS, EVERYWHERE

Steps of Applied Economics Analyses

- Define question/topic
- Find, get, wrangle and clean data
- Summary statistics
- Define an identification strategy
- Build a regression model
- Estimate your model
- Specification checks
- Additional inference
- Robustness checks
- Communicate

A More Structured Version

- Design: decisions of data collection and measurement
  - *eg*, decisions related to sample size and ensuring exogeneity of the treatment
- Modeling: define statistical models
  - In between design and analysis
- Analysis: estimation and questions of statistical inference
  - *eg* standard errors, hypothesis tests, and estimator properties

Objectives of the class

- Form more **intuition** on what econometric tools actually do to our data
- Discuss **practical hurdles** and challenges we may face when doing applied economics:
  - Help us be aware of some of them
  - Provide us with some tools to be able to spot others by ourself
- Side quest: help you get started with your masters thesis
- *Topics* class: will not cover everything but instead pick points within topics

Logistics

Website

TOPICS IN ECONOMETRICS Syllabus Lectures Simulations Assignments Resources

Topics in Econometrics

MSc Advanced Economics - ENS Lyon

This website gathers all the materials and information for the course Topics in Econometrics (ECO5106) of the MSc Advanced Economics at the ENS Lyon.

https://vincentbagilet.github.io/metrics_m2/

A typical lecture

1. I introduce concepts and intuition
2. We discuss a paper together (when reading assigned)
3. Some R coding together *and* on your own

Grading and assignments

Final project

- In pairs
- Build a simulation to replicate an analysis you may do in your master thesis
- Generate realistic fake-data, run your analysis and discuss your results
- Project proposal, short report, presentation

Structure of the final project

- Structure it as short research paper (pitch):
  - Quick motivation and context
  - Research question
  - Data section: describe how you generate your data. Start very simple, complexify later
  - Modeling and analysis: describe your model, your choices and the outputs of your regressions
  - Discussion: what did you learn with this exercise

Hurdles to accurate estimates

The need for accurate estimates

- “Accurate” estimates have:
  - The same **sign** as the “true” underlying effect
  - A comparable **magnitude** as the “true” underlying effect
  - That are **precise** enough
- Assumes that there is a “true” underlying effect, a true DGP
- They can:
  - Help with our understanding of the world and inform our theories
  - Even inform policy

Improving our understanding of the world

- Accurate estimates tell us:
  - Is there an effect?
  - In which direction?
  - Is it large enough to matter? (practical significance)
- They help:
  - Test, form and compare theories
  - Calibrate models
  - Build future research on solid ground

Informing policy: air pollution in the US

- Clean Air Act:
  - The EPA (Environmental protection agency) sets air quality standards
  - Based on the costs and benefits of various standards
  - States implement the standards
- Health benefits: $\Delta Y \approx Pop \times Y_0 \times \hat{\beta} \times \Delta AQ$
  - $\hat{\beta}$: health effect **estimated in the literature**
  - Benefits **proportional to** $\hat{\beta}$
- *Example*: in 2024, change in PM2.5 standard (12 → 9 μg/m^3):
  - Health benefits computed from literature estimates: $22-46B per year
  - Costs: $0.6B per year

Typical sources of inaccurate estimates

- At the **design** stage:
  - *eg*, measurement error, low statistical power
- At the **identification** stage:
  - *eg*, reverse causality, confounders, SUTVA violations, (implicit) sample selection
- At the **modeling** stage:
  - *eg*, misspecification, bad controls, negative weights and implicit weighting,
- At the **inference** stage:
  - *eg*, incorrect standard errors, weak instruments
- Selective publication and generalization:
  - *eg*, specification search, publication bias, external validity

Order of concern

- There is some chronology in hurdles: each only matters to the extent that the previous one are addressed
- We need to have, in that order of concern:
    1. A good **research question**, grounded in theory
    2. A good **identification strategy** to avoid some fundamental hurdles (reverse causality, confounders, etc)
    3. A **specification** that allows us to estimate the quantity we want to estimate
    4. Make reliable inference based on **well-constructed standard errors** for instance

In this class we will explore

- Common hurdles and challenges in empirical work
- How regression works "**under the hood**"
- Causal identification strategies and their assumptions
- How design, modeling, and analysis **choices shape empirical results**
- How to use **simulations** to explore estimator behavior and diagnose potential problems specific to your own cases
- Existing references and where to find additional information on a specific topic

Outline of the class

1. Introduction and Fundamentals
2. Simulations
3. Linear regression and causal inference
4. Design: Fixed Effects
5. Design: IV
6. Modelling
7. Analysis
8. Communication

Research questions

What is a good research question?

- It **can be answered**
  - There is some sort of objective answer
- It should **improve our understanding of the world**
  - Should **inform theory** in some way
  - Takes us from theory to an hypothesis (statement about what we will observe in the world)
- A solid econometric analysis only matters to the extent that you have a good research question (but the opposite might be true as well)

Example

- Impact of the size of motors of boats in Norway and cod catch under a catch cap
- Not that interesting in itself, is it? Would be more interesting if, for instance:
  - Look at this from a game theory and forced technological adoption perspective
  - Find a way to use this case to say something *new* or *different* on management of renewable natural resources
- Can produce radically different papers on the same topic and setting
- **Use theory to put light on your specific case *and* your specific case to inform theory**

Identifying a research question

- Can start with a research question/hypothesis or from theory
- Or can find a natural experiment and come up with a question
- Know your literature to identify gaps
- We are interested in why and not what
- Avoid data mining: it can help but to identify questions to test on other data sets

Is your research question good?

- **Relevance**: is it interesting, important or policy relevant ?
- **Potential results**: what would any result tell you about your theory?
- **Feasibility**: is the right data available?
- **Scale**: how much resources would you need?
- **Research design**: is there a good one that would allow you to answer your question?
- **Keep it simple**: avoid building several questions into one

Avoiding hurdles

How to avoid hurdles?

Learn, understand metrics and applied research

The Barque of Dante by Manet, after a painting by Delacroix

Replication, a helpful learning tool

How to avoid hurdles?

Derive the maths

- Sometimes relatively straightforward and very illuminating
- *eg* drivers of the variance of your estimator: $\mathbb{V}_{\hat{\beta}} = \frac{\sigma_u^2}{n\sigma_x^2}$
- Deriving the maths can be more complex and time consuming

How to avoid hurdles?

Simulations can help

- Super easy to implement simple simulations
- Can be informative of **what does not work**
- Can help you identify **where the issue comes from**
- We will discuss that with an example in a second

How to avoid hurdles?

Implement checks

- Check if the model seems to represent the DGP
- Check if our identification hypotheses seem to hold
- Check if the hypotheses for estimation seem to hold

Look at the consequences if this does not hold

- Robustness checks
- Evaluate the design retrospectively

Objectives for this class

- Build a mindful mindset
- Help you be aware of some of them
- Provide you with some tools to be able to spot others by yourself
- Learn how to implement simulations

Simulations: usefulness through an example

A simple example: OVB

- How does an **omitted variable** affect our point estimate of interest? Why?
- Under which condition is an omitted variable an issue
- How does it affect the point estimate? The s.e.?
- How does that vary with various parameters? *eg* correlation between variable (sign and magnitude)
- Start very simple and complexify the process
- Let's move to R

Lecture summary

What did we do today?

- Discussed the pervasiveness of econometrics in applied research and therefore the **usefulness of studying** it
- Browsed logistics
- Reviewed some common **hurdles** encountered in applied research
- Explored how to implement a **simple simulation** to understand the impact of an omitted variable depending on the value of various parameters

A few random take-away points

- A great econometrics analysis only matters to the extent that you have **a good research question**
- Get inspiration from **papers you like**: structure, methods, questions, graphs and tables, etc
- Attend **seminars**! You can always grab something (at worst in terms of form)
- There are many reasons for our estimation to yield incorrect results:
  - Learn how to spot and understand them
  - Simulations can help!

What did you learn, like, dislike?

Thank you!