

Introductory Econometrics

Lecture 1: What is Econometrics?

Alex Alekseev

October 17, 2022

University of Regensburg, Department of Economics

Introduction

Why Study Econometrics?

1. It is a primary methodological tool for empirical economics

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021



© Nobel Prize Outreach. Photo: Paul Kennedy

David Card

Prize share: 1/2



© Nobel Prize Outreach. Photo: Risdon Photography

Joshua D. Angrist

Prize share: 1/4



© Nobel Prize Outreach. Photo: Paul Kennedy

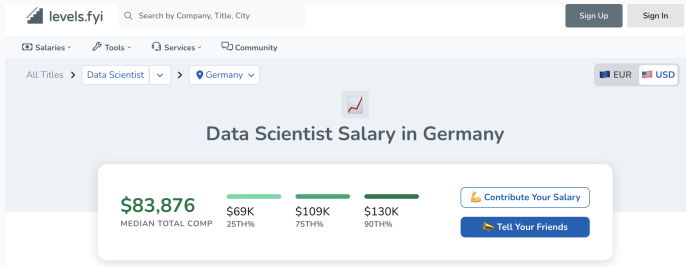
Guido W. Imbens

Prize share: 1/4

David Card (*“for his empirical contributions to labour economics”*)
and to Josh Angrist + Guido Imbens (*“for their methodological contributions to the analysis of causal relationships”*)

Why Study Econometrics?

1. It is a primary methodological tool for empirical economics
2. Data science is a good-paying job



“discover empirical relation between economic variables, provide forecast of various economic quantities of interest”

First issue of volume 1, Econometrica, 1933

"The science of model building consists of a set of quantitative tools which are used to construct and then test mathematical representations of the real world. The development and use of these tools are subsumed under the subject heading of econometrics"

Pindyck and Rubinfeld (1998)

Definitions of Econometrics

"At a broad level, econometrics is the science and art of using economic theory and statistical techniques to analyze economic data. Econometric methods are used in many branches of economics, including finance, labor economics, macroeconomics, microeconomics, marketing, and economic policy. Econometric methods are also commonly used in other social sciences, including political science and sociology"

Stock and Watson (2007)

“Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy”

Wooldridge (2009)

A Trade Example: What Determines Trade Flows?

Research Question

Identify the factors that affect **imports** to Germany from other countries and quantify their impact

Basic Questions

1. Which **economic theories** are relevant for this question?
2. What **data** can be used to check these economic theories?
3. How to **decide** which economic theories are supported or rejected by the data?

Let's have a look at the data set on the imports (in current US dollars) to Germany from other countries in 2004

Can you guess what are the **top exporters** to Germany?

Imports to Germany in 2004 by Country



First Attempt

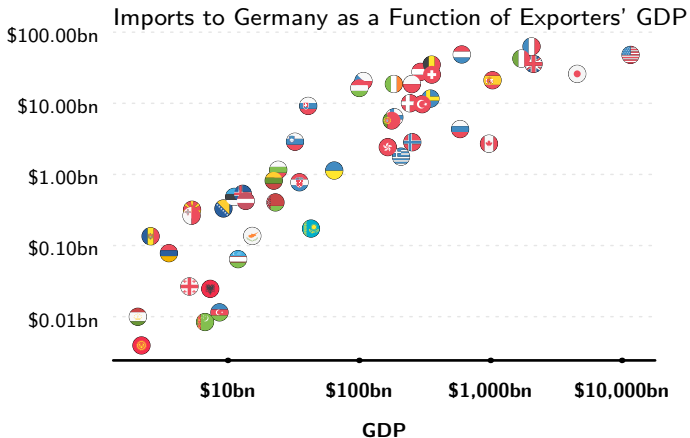
1. Which **economic theories** are relevant for this question?
 - Ignore, for the moment, all economic theory and simply hypothesize that observed imports depend on the GDP of the exporting country

1. Which **economic theories** are relevant for this question?
 - Ignore, for the moment, all economic theory and simply hypothesize that observed imports depend on the GDP of the exporting country
2. What **data** can be used to check these economic theories?
 - Collect the GDP data for exporters

First Attempt

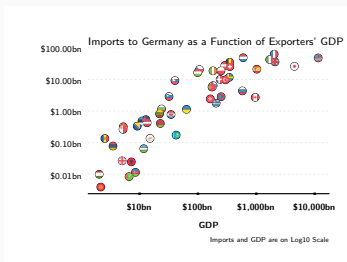
1. Which **economic theories** are relevant for this question?
 - Ignore, for the moment, all economic theory and simply hypothesize that observed imports depend on the GDP of the exporting country
2. What **data** can be used to check these economic theories?
 - Collect the GDP data for exporters
3. How to **decide** which economic theories are supported or rejected by the data?
 - Plot the data, e.g., using a scatter plot

Can you predict the **relationship** between the GDP of an exporter and the value of its import to Germany?



Imports and GDP are on Log10 Scale

Some Questions



- Is there a relationship between imports and GDP?
- Is it positive or negative?
- What is the direction of causality?
- How can we quantify this relationship?
- Are there other potential factors affecting imports?
- Is it possible to forecast future trade flows?

What We Have Done So Far

We tried to simplify the reality by building a **model**

An Economic Model

- Reduces the complexity of the reality such that the model is **useful** for answering the research question of interest

*"All models are wrong, but
some are useful"*

George Box

An Economic Model

- Reduces the complexity of the reality such that the model is **useful** for answering the research question of interest
- Is built on **assumptions** from which implications can be logically deduced

An Economic Model

- Reduces the complexity of the reality such that the model is **useful** for answering the research question of interest
- Is built on **assumptions** from which implications can be logically deduced
- Should be as **simple** as possible and as complex as necessary

"Everything should be made as simple as possible, but not simpler"

Albert Einstein

An Economic Model

- Reduces the complexity of the reality such that the model is **useful** for answering the research question of interest
- Is built on **assumptions** from which implications can be logically deduced
- Should be as **simple** as possible and as complex as necessary
- Should be **falsifiable**

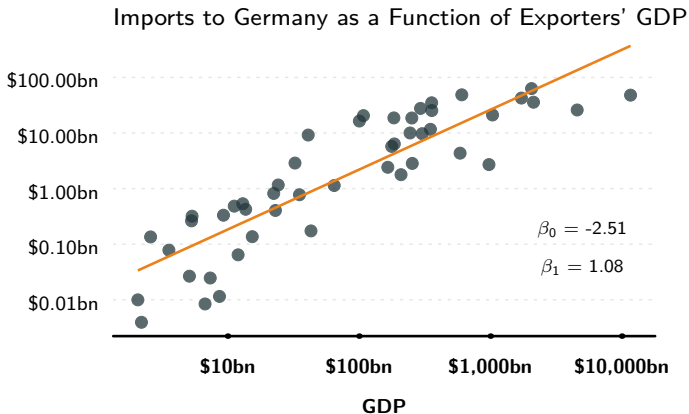
A Model of Imports

- Let's consider a simple formal model of the relationship between imports to Germany and the exporters' GDP

$$\log(\text{Imports}) = \beta_0 + \beta_1 \log(\text{GDP})$$

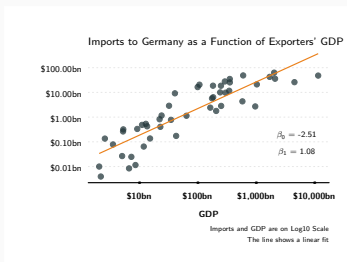
- Does it make sense?
- How can we determine the parameters of the model?

Fitting a Straight Line



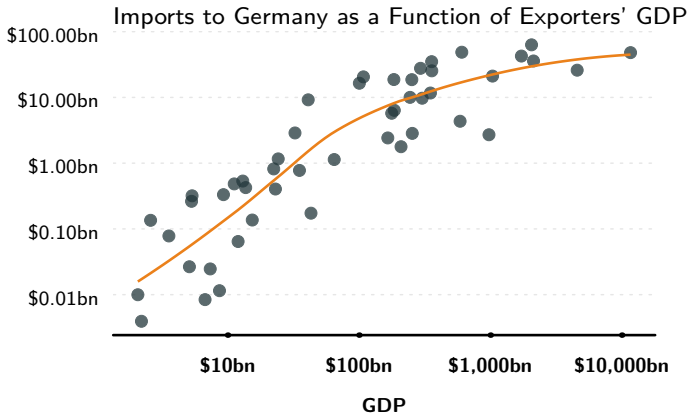
Imports and GDP are on Log10 Scale
The line shows a linear fit

Some Questions



- How to fit a line like this?
- What properties should it have?
- What to do with other potential factors that are currently omitted from the model?
- Is the relationship linear?
- How may the results change in a different sample, say, for 2022?

Fitting a Non-Straight Line



Imports and GDP are on Log10 Scale

The line shows a loess fit

Economic Models and the Need for Econometrics

Standard Problems of Economic Models

- A model is likely to neglect some **relevant factors**
- Numeric (estimation) results generally **depend** on a data set (sample)
- A different data set will lead to different estimates
 - Estimates will have some **uncertainty**

How Does Econometrics Help?

- Offers solutions for dealing with **unobserved factors** in economic models
- Gives “both a numerical answer to the question and a measure how precise the answer is” (Stock and Watson, 2007, p. 7)
- Provides us with tools that allow to reject **hypotheses** using statistical techniques by confronting theory with data...
- ...and to quantify the **probability** of such decisions to be wrong
- Allows us to quantify the **uncertainty** of forecasts
- How likely is it that the actual value will be close to the predicted value?

The regression model

Quantifies the effect of a change in one variable on the outcome while holding other variables constant (*ceteris paribus*)

Steps of an Econometric Analysis

1. Formulate a research question
2. Specify an economic model
3. Select the class of econometric models
4. Collect the data
5. Estimate the econometric model
6. Check for the correct specification
7. Use the model

The Uses of Econometric Methods

- Providing a formal framework for falsifying postulated economic relationships by confronting an economic theory with data using statistical methods
- Analyzing the effects of policies
- Forecasting

Causation vs. Correlation

Correlation is not causation

Just because two variables are correlated does not necessarily mean that one causes the other

Examples

- **Obvious:** When a rooster crows, the sun soon after rises, but does it mean that the rooster causes the sun to rise?
- **Less obvious:** A central bank raises interest rates after which a recession follows. Does it mean that raising the interest rates caused the recession?

Who Cares?

Not interested	Interested
if countries with higher minimum wages have less poverty	if raising the minimum wage reduces poverty
if people who take a popular common-cold-shortening medicine get better	if the medicine made them get better more quickly
if the central bank raising interest rates was shortly followed by a recession	if the interest rate increase caused the recession

What is Causality?

A **useful** way to think about it is this:

We can say that X causes Y if, were we to intervene and change the value of X , then Y would also change as a result

Using the Definition

- If a fox eats the rooster, would it mean that the sun would no longer rise?
- If we were to stop the central bank and not let it raise the interest rates, would it prevent the recession?

Types of Economic Data

Experimental vs. Observational

"The word experiment is used ... to mean an investigation where the system under study is under the control of the investigator. This means that the individuals or material investigated, the nature of the treatments or manipulations under study and the measurement procedures used are all selected, in their important features at least, by the investigator. By contrast in an observational study some of these features, and in particular the allocation of individuals to treatment groups, are outside the investigator's control"

Cox and Reid (2000)

Our focus in this course is **observational** data

Cross-Sectional Data

- Collected across several units at a single **point in time**
- Units: “economic agents,” e.g., individuals, households, investors, firms, economic sectors, cities, countries
- The **order** of observations has no meaning
- Popular to use index i
- The data are a **random sample** of the underlying population
- Rely on the variation **across** units
- Example: our trade data

The focus of this course

Time Series Data

- Collected across **differing** points or periods of time
- Popular to use index t
- The **order** of observations matters
- Sampling frequency
 - Fixed: annually, quarterly, monthly, weekly, daily,
 - Variable: ticker data, duration data (e.g., unemployment spells)
- Allow the analysis of **dynamic** effects
- Example: trade flow from the US to Germany from 1990 to 2022

- A **collection** of cross-sectional data for at least two different points/periods of time
- Individual units remain **identical** in each cross-sectional sample
- Popular to index as *it*
- Example: imports from 54 different countries to Germany from 1990 to 2022 where all 54 countries remain in the panel for all subsequent years ($T = 32$, $N = 54$)

Pooled Cross Section

- Like panel data but units can **change** across time
- Example: in 2021 the countries of origin are France, Italy, the US, but in 2022 the countries of origin are the Netherlands, Portugal, and Spain