Lecture 1:
Introduction

Jan Zouhar

Introductory Econometrics

2

Course Information

Course Information

- **Lecturer info**:
  - *name*: Jan Zouhar
  - *e-mail*: zouharj@vse.cz
  - *web*: nb.vse.cz/~zouharj
  - *office*: room NB 431
  - *office hours*: Tue 16:15 – 17:45
Fri 10:30 – 12:00
- **Course Requirements**:
  - **30 points**: home assignments
    - **3** problems sets, **10** points each
    - will be posted on the website above
    - due dates (tentative): weeks 4, 7, and 11 of the semester.
  - **20 points**: mid-term test (week 9)
  - **50 points**: final test (week 13 + three more dates in the exam period)

Course Information (cont'd)

- **Grading scale**: standard ECTS 100 points
  - **90 – 100** points: excellent (**1**)
  - **75 – 89** points: very good (**2**)
  - **60 – 74** points: good (**3**)
  - **0 – 59** points: failed (**4**)
- **Recommended reading**:
  - Lecture notes and presentations.
  - WOOLDRIDGE, J. M.: *Introductory econometrics: a modern approach*, 5e. Mason: South-Western, 2013.
  - GUJARATI, D. N.: *Basic econometrics*, Boston: McGraw-Hill, 2003.
  - Virtually any other book on econometrics.
- **Software**:
  - we'll use the freeware *Gretl* econometric software, available at gretl.sourceforge.net

Introducing Econometrics

What is econometrics?

Typical questions in econometrics.

Causality & econometrics.

What is Econometrics?

Definition 1: *econometrics* = *econo* + *metrics*

- **economics vs. econometrics**
  - *economics*: focus on "how" and "why"
  - *econometrics*: focus on "how much" and "by how much"
  - **example**:
    - *economist*: "If the government increases alcohol excise tax, consumers will cut down on their alcohol consumption."
    - *econometrician*: "If the government increases alcohol excise tax by 20%, consumers will reduce their alcohol consumption by 1%."

  → econometrics is absolutely vital in applying economic theories in practice
    - reflected in the number of econometricians among Nobel prize laureates

What is Econometrics? (cont'd)

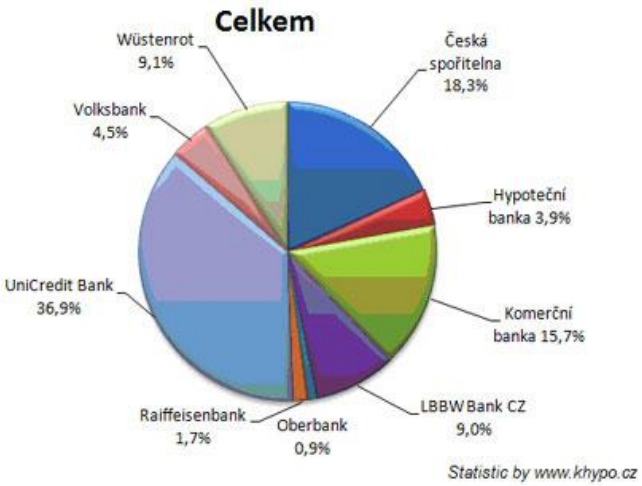
- econometrics is *not* concerned with the numbers themselves (the concrete information in the previous example), but rather with the *methods* used to obtain the information → *crucial role of statistics*

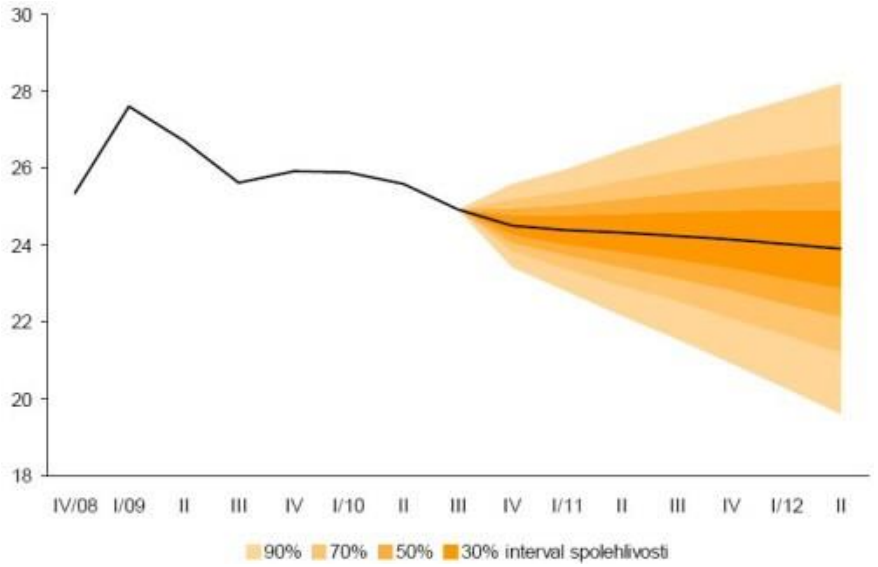
> Definition 2: *econometrics = statistics for economists*

- textbook definitions of econometrics:
  - "*application of mathematical statistics to economic data to lend empirical support to models constructed by mathematical economics and to obtain numerical estimates.*" (Samuelson et al., Econometrica, 1954)
  - "*application of mathematics and statistical methods to the analysis of economic data.*" (www.wikipedia.org)
- **econometrics vs. statistics**:
  - is econometrics a part of statistics? Not quite – economic data give rise to methods unparalleled in any branch of statistics.

What is Econometrics? (cont'd)

- typical econometricians' output for popular press





- three basic types of econometric questions
  - **descriptive**
  - **forecasting**
  - **causal** (or structural)

Descriptive Questions

- typical questions:
  - How much do men and women earn annually on average in the UK?
  - How long do recessions typically last?
  - How does medical insurance coverage vary with income?
  - How are amenities of a house reflected in the house's price?

  (*note: all these question mean "how much … on average* or *typically"*)
- descriptive type of questions is the simplest one
- main trait: *if we had enough data, we would know the answer for sure*
  - **example**: if you have a complete (and accurate) list of all UK citizens' income, you can answer the first question in the list above
- challenges:
  - *sampling*: how to make conclusions based on a sample rather than the whole population (→ *random sampling & statistical inference*)
  - *summary statistics*: how to summarize the (quantitative) answer in a nice, brief and comprehensible way

Forecasting Questions

- typical questions:
  - Who will win UEFA Euro 2016?
  - What will the global temperature be in 2040?
  - How long will the recession last this year?
  - What will be the Labour's vote share in the next election?
  - What will the stock price of Google be on 14th March?
  - Will you pass this course?
- we can never know the answers for sure in advance; however, there might be very high stakes behind these questions
  - good prediction = £1,000,000s
- common traits:
  - if we wait long enough, we'll know the answer
  - inferences based on time-related data (i.e., *time series*)
- highly visible applications of econometrics: forecasts of macroeconomic indicators (interest rates, inflation, GDP etc.)

Forecasting Questions (cont'd)

- alternative forecasting techniques:

- typical consequences:

Causal Questions

- typical questions:
  - If the federal reserve lowers interest rates today, what will happen to inflation tomorrow?
  - What is the effect of political campaign expenditures on voting outcomes?
  - How much more money will you earn as a result of taking this course?
  - Will spending a lot of money on highway construction get us out of the recession?
  - How would legalization of cannabis influence…
    - …the number of its users?
    - …tax revenues?
    - …citizens' overall happiness?
- note the *cause* and *effect* elements in the previous questions
- the presence of a causal link is suggested by economic theory (or common sense), the goal of econometric analysis is either to *empirically verify* or *quantify* this causal link

Causality, Ceteris Paribus, and Experiments

- in economic thinking, causal relations are strongly connected with the notion of **ceteris paribus** ("other things being equal")
  - **example**: consumer demand analysis – increasing a price makes consumers buy less *ceteris paribus* (however, if other factors change, anything can happen)
- therefore, if one could run an experiment with ceteris paribus conditions enforced, it would be easy to verify and evaluate the causal link
- this is the way things are done in natural sciences
  - **example**: with decreasing air pressure, lower water temperature is needed for it to boil and turn into steam

  → experiment: it's easy to provide for the ceteris paribus conditions in a laboratory setting
- in social sciences, such controlled experiments are either impossible, unethical or prohibitively expensive
  - **example**: political campaign expenditures – impossible to re-run the election with different campaign budgets

Causality, Ceteris Paribus, and Experiments (cont'd)

- we can distinguish between
  - **experimental data**: "created" in a laboratory experiment
  - **non-experimental / observational data**: researcher = passive collector of the data
- a large part of econometrics deals with how to get "correct" results despite working with non-experimental data.

Causality & Econometrics Sum-Up:

*Econometric tools cannot be used to **find** causal links; these have to be found in economic theory. Econometrics can help us **quantify** causal effects and/or **verify** their presence. The challenge in here consist in dealing with **non-experimental** data where **ceteris paribus** conditions cannot be established.*

A Note on Randomized Experiments

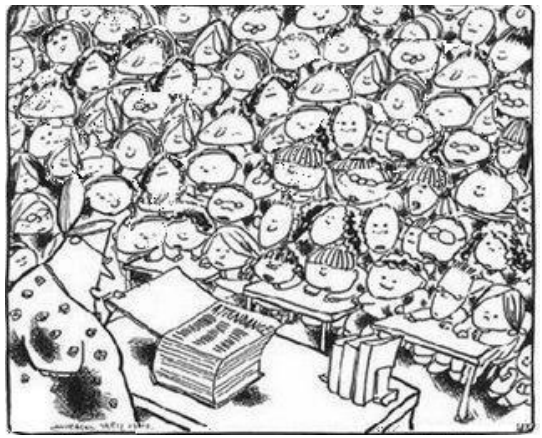
- sometimes, even in experiments related to natural sciences, it impossible to enforce ceteris paribus conditions
  - **example** (**crop yields**): assessing the effect of a new fertilizer on soybeans
    - ceteris paribus = ruling out other yield-affecting factors such as rainfall, quality of land, presence of parasites etc.
    - experimental design:
      1. Choose several one-acre plots of land.
      2. Apply different amounts of fertilizer to each plot.
      3. Use statistical methods to measure the association between yields and fertilizer amounts.
    - *drawback*: some of yield-affecting factors are not fully observed → impossible to choose "identical" plots of land
    - *solution*: statistical procedures still work correctly, if fertilizer amounts are *independent* of the other factors[1] – e.g., if we choose fertilizer amounts completely *at random* → hence **randomized experiments** ([1] we'll discuss this property in more detail later on)

A Note on Randomized Experiments (cont'd)

- **example** (**returns to education**): If a person is chosen from the population and given another year of education, by how much will his/her wage increase?
  - randomized experiment:
    1. Choose a group of people (children).
    2. Randomly assign a level of education to each person.
    3. After all of them have finished their schooling and got employed, measure their wages and use statistical methods.
    - would you let your child participate in such an experiment?
    - is it ethical to force people to participate?
  - it's fairly easy to collect non-experimental data on *wages* and *education*; however, ceteris paribus doesn't work here
    - *education* vs. *working experience* (easy to fix – collect data for *exp*.)
    - *education* vs. *ability* (difficult to fix – *ability* largely *unobservable*)
    - again, we'll cover this in more detail later on

A Note on Randomized Experiments (cont'd)

- **example (class sizes)**: does a kindergarten class size determine a pupil's performance in early years of study (and perhaps afterwards)?
  - randomized experiment:
    - Tennessee STAR programme (Student/Teacher Achievement Ratio), 1985–1989
    - kindergarten pupils randomly assigned to three different class modes:



      - 13–17 students, 1 teacher (*small*)
      - 22–26 students, 1 teacher (*regular*)
      - 22–26 students, 1 teacher + 1 teacher's aide (*regular* + *aide*)
    - students' performance tested throughout the following years (SAT)

A Note on Randomized Experiments (cont'd)

- extremely costly: budget = $12 million (for more info, see `STAR_Facts.pdf` from my website)
- even though the basic problem sounds fairly simple, and huge costs have been incurred in order to get everything done correctly, the are still doubts about the plausibility of the results (see `ClassSizeDebate.pdf`)

Randomized Experiments Sum-Up:

*If carried out properly, randomized experiments can **substitute** the **ceteris paribus** conditions. However, in social sciences, these experiments are typically either **impossible**, or at least **unethical** or **extremely costly** to conduct.*

Steps in Empirical Economic Analysis

General scheme

Step 1: Formulate the **question** of interest.

Step 2: Find a suitable **economic model**.

Step 3: Turn it into an **econometric model**.

Step 4: Obtain suitable **data**.

Step 5: Use econometric methods to **estimate** the econometric model.

Step 6: If needed, use **hypothesis tests** to answer the question from step 1.

Steps in Empirical Economic Analysis (cont'd)

Step 1: Formulate the question of interest.

- **example** (**crime vs. wage**): does the wage that can be earned in legal employment affect the decision to engage in criminal activity?

Step 2: Find a suitable **economic model**.

- formal relationships between economic variables
- **example** (**crime vs. wage**): Gary Becker (1968) – max. utility:

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

y	hours spent in criminal activity
x_1	criminal "hourly wage"
x_2	hourly wage in legal employment
x_3	income other than from crime or employment
x_4	probability of getting caught
x_5	probability of being convicted if caught
x_6	expected sentence if convicted
x_7	age

Steps in Empirical Economic Analysis (cont'd)

Step 3: Turn it into an **econometric model**.

- solve **quantification** issues
  - how can we measure *hours spent in criminal activity*?
  - how do we approximate the *probability of being caught* with an *observable* economic variable?
- specify the **functional form** of the economic relationships
  - **example** (**crime vs. wage**):

$$crime = \beta_0 + \beta_1 wage + \beta_1 oth\_inc + \beta_2 freq\_arr + \beta_3 freq\_conv + \beta_4 avg\_sen + \beta_5 age + u$$

    - u ... **error term** or **disturbance**, which contains:
      - unobserved factors ("criminal wage", moral character, family background)
      - measurement errors
      - random nature of human behaviour

The Structure of Econometric Data

Cross-sectional data.

Time series.

Pooled cross sections and panel data.

Cross-Sectional Data

obs	wage	educ	exper	age	female
1	18.10	12	17	35	1
2	36.87	18	27	51	0
…	…	…	…	…	…
526	61.45	20	3	29	1

- 1 **observation** = information about 1 cross-sectional unit
  - **cross-sectional units**: individuals, households, firms, cities, states
- data taken at a given point in time
- typical assumption: units form a **random sample** from the whole population → the notion of *independence* of the units' values
  - possible violations:
    - *censoring*: wealthier families are less likely to disclose their wealth
    - *small population*: neighboring states influence one another, their indicators are not independent

Time Series

year	T-bill	infl	displnc	C_ndur	popul
1994	4.95	2.6	4778.2	1390.5	260,660
1995	5.21	2.8	4945.8	1421.9	263,034
…	…	…	…	…	…

- observations on economic variables over time
  - stock prices, money supply, CPI, GDP, annual homicide rates, automobile sales
- **frequencies**: daily, weekly, monthly, quarterly, annually
- unlike cross-sectional data, ordering is important here!
  - behaviour of economic subject (and the resulting indicators) evolve in a gradual manner in time
  - lags in economic behaviour (oil prices today affect next month's actions)
- typically, observations **cannot** be considered **independent** across time

→ require more complex econometric techniques

Pooled Cross Sections

obs	year	hprice	sq_feet	bdrms	bthrms
1	2005	105,000	1400	3	1
.....	.....	.....	.....	.....	.....
250	2005	198,500	2350	5	3
251	2008	95,600	1800	3	2
.....	.....	.....	.....	.....	.....
550	2008	119,900	2150	4	2

- both cross-sectional and time-series features
- data collected in multiple (typically, two) points in time
- ordering is not crucial, *year* is recorded as an additional variable
- often used to evaluate the effect of a policy change
  - collect data *before* and *after* the policy change and see how the relationship between the variables changes
- *note*: in the second time period, the cross-sectional units need be neither distinct from nor identical to those in the first period

Panel (or Longitudinal) Data

	unit	year	popul	murders	unemp	police
Leicester	1	2008	293,700	5	6.3	358
	1	2010	299,500	7	7.4	396
Salisbury	2	2008	53,450	2	7.2	51
	2	2010	51,970	1	8.1	51
	.....	.....	.....	.....	.....	.....

- several cross-sectional units, a time series for each unit (time series with equal length)
- unlike with pooled cross sections, the same units are measured over time
  - more difficult /costly to obtain the data
  - have several advantages over (pooled) cross sections (for problem where panel data make sense)
  - can be treated as pooled cross section (but: loss of information)

Lecture 1:
Introduction

Jan Zouhar

Introductory Econometrics