Econometrics in Economics / Introduction to Econometrics

Intro

Patrick Schmidt (based on slides by **Simon Heß** and **Daniel Gutknecht**)

Winter 23

Welcome to econometrics

Econometrics?

- Economics without data?

Econometrics?

- Economics without data? **No, thank you.**

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:
  - How to analyze data

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:
  - How to analyze data
  - How to test economic models

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:
  - How to analyze data
  - How to test economic models
  - How to understand causality ($\neq$ correlation)

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:
  - How to analyze data
  - How to test economic models
  - How to understand causality ($\neq$ correlation)
  - How to forecast

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:
  - How to analyze data
  - How to test economic models
  - How to understand causality ($\neq$ correlation)
  - How to forecast
  - Statistics for regression models

Econometrics?

- Economics without data? **No, thank you.**
- Econometrics: Connects economic theory to data.
- We will learn:
  - How to analyze data
  - How to test economic models
  - How to understand causality ($\neq$ correlation)
  - How to forecast
  - Statistics for regression models

Roadmap for today

- Formalities
- Why should I study econometrics?
- Big picture: Causality and data
- Refresher on statistics and probability (partly in exercise)

Formalities

Two rules, one course:

- lectures (weekly), exercise (bi-weekly), and computer sessions (bi-weekly)

Two rules, one course:

- lectures (weekly), exercise (bi-weekly), and computer sessions (bi-weekly)
- **Econometrics in Economics** (BOEE):
  - all are relevant for the exam
- **Introduction to Econometrics** (EOEK):
  - only lectures and exercise relevant for the exam
  - computer sessions still recommended

Two rules, one course:

- lectures (weekly), exercise (bi-weekly), and computer sessions (bi-weekly)
- **Econometrics in Economics** (BOEE):
  - all are relevant for the exam
- **Introduction to Econometrics** (EOEK):
  - only lectures and exercise relevant for the exam
  - computer sessions still recommended

OLAT

- slides and problem sets will be distributed via OLAT

OLAT

- slides and problem sets will be distributed via OLAT
- for content questions, please use the **OLAT forum**

OLAT

- slides and problem sets will be distributed via OLAT
- for content questions, please use the **OLAT forum**
- my **office hours**: Wednesday 11pm by email appointment

OLAT

- slides and problem sets will be distributed via OLAT
- for content questions, please use the **OLAT forum**
- my **office hours**: Wednesday 11pm by email appointment
- recordings on OLAT (if technology allows)

OLAT

- slides and problem sets will be distributed via OLAT
- for content questions, please use the **OLAT forum**
- my **office hours**: Wednesday 11pm by email appointment
- recordings on OLAT (if technology allows)

Computer sessions:

- we use **R** and **RStudio** in the computer sessions
  - free, powerful (for statistics and data science), and open source software

Computer sessions:

- we use **R** and **RStudio** in the computer sessions
    - free, powerful (for statistics and data science), and open source software
    - common alternatives: Stata (commercial), Matlab (commercial), Python (more general), Julia (new)
- if possible, bring a laptop with R and RStudio installed (installation tutorial on OLAT)

Computer sessions:

- we use **R** and **RStudio** in the computer sessions
  - free, powerful (for statistics and data science), and open source software
  - common alternatives: Stata (commercial), Matlab (commercial), Python (more general), Julia (new)
- if possible, bring a laptop with R and RStudio installed (installation tutorial on OLAT)
- coding can be fun, potentially useful for Bachelor's Thesis, and for your career

Computer sessions:

- we use **R** and **RStudio** in the computer sessions
  - free, powerful (for statistics and data science), and open source software
  - common alternatives: Stata (commercial), Matlab (commercial), Python (more general), Julia (new)
- if possible, bring a laptop with R and RStudio installed (installation tutorial on OLAT)
- coding can be fun, potentially useful for Bachelor's Thesis, and for your career
- Tech Academy offers courses on data science (Flyer on OLAT)

Computer sessions:

- we use **R** and **RStudio** in the computer sessions
  - free, powerful (for statistics and data science), and open source software
  - common alternatives: Stata (commercial), Matlab (commercial), Python (more general), Julia (new)
- if possible, bring a laptop with R and RStudio installed (installation tutorial on OLAT)
- coding can be fun, potentially useful for Bachelor's Thesis, and for your career
- Tech Academy offers courses on data science (Flyer on OLAT)

Organisation

- **Cannon Warren** will hold the computer tutorials

Organisation

- **Cannon Warren** will hold the computer tutorials
- final exam (in person, 90 minutes)

Organisation

- **Cannon Warren** will hold the computer tutorials
- final exam (in person, 90 minutes)
  - For BOEE: exam will contain questions on R code

Organisation

- **Cannon Warren** will hold the computer tutorials
- final exam (in person, 90 minutes)
  - For BOEE: exam will contain questions on R code
- there will be a mock exam

Organisation

- **Cannon Warren** will hold the computer tutorials
- final exam (in person, 90 minutes)
  - For BOEE: exam will contain questions on R code
- there will be a mock exam
- exercises will not be handed in. Solutions will be discussed in the exercise classes and uploaded to OLAT.

Organisation

- **Cannon Warren** will hold the computer tutorials
- final exam (in person, 90 minutes)
  - For BOEE: exam will contain questions on R code
- there will be a mock exam
- exercises will not be handed in. Solutions will be discussed in the exercise classes and uploaded to OLAT.
- I appreciate active participation: **ask** if stuff is unclear

Organisation

- **Cannon Warren** will hold the computer tutorials
- final exam (in person, 90 minutes)
  - For BOEE: exam will contain questions on R code
- there will be a mock exam
- exercises will not be handed in. Solutions will be discussed in the exercise classes and uploaded to OLAT.
- I appreciate active participation: **ask** if stuff is unclear

Useful textbooks

1. J. Wooldridge (2019), 7th edition: **Introductory Econometrics**, Cengage Learning.
2. J. Stock and M.Watson (2019): **Introduction to Econometrics**, Pearson.
3. J. Angrist and S. Pischke (2008): **Mostly Harmless Econometrics**, Princeton Univ. Press.

Motivation: Why econometrics?

Motivation: Why econometrics?

In a nutshell, **econometrics** can be characterized by using statistics to address economic questions.

Motivation: Why econometrics?

In a nutshell, **econometrics** can be characterized by using statistics to address economic questions.

Economics:

- explains relationships between economic variables, or the behavior of economic agents or markets through models

Motivation: Why econometrics?

In a nutshell, **econometrics** can be characterized by using statistics to address economic questions.

Economics:

- explains relationships between economic variables, or the behavior of economic agents or markets through models

Econometrics:

- studies the related quantitative questions using economic data and statistical methods

Motivation: Why econometrics?

In a nutshell, **econometrics** can be characterized by using statistics to address economic questions.

Economics:

- explains relationships between economic variables, or the behavior of economic agents or markets through models

Econometrics:

- studies the related quantitative questions using economic data and statistical methods

Question:

So what are these quantitative questions?

1. **Testing and falsifying economic theories** empirically:
    - Are exchange rate markets for freely-traded currencies rational?
    - Are stock prices unpredictable?

1. **Testing and falsifying economic theories** empirically:
   - Are exchange rate markets for freely-traded currencies rational?
   - Are stock prices unpredictable?

2. **Quantifying relationships** between economic variables:
   - does education have a positive (causal) effect on wages? And if yes, what is the magnitude?
   - what is the average gender pay gap?

1. **Testing and falsifying economic theories** empirically:
   - Are exchange rate markets for freely-traded currencies rational?
   - Are stock prices unpredictable?

2. **Quantifying relationships** between economic variables:
   - does education have a positive (causal) effect on wages? And if yes, what is the magnitude?
   - what is the average gender pay gap?

3. **Program evaluation** for policy analysis:
   - does a specific job training shorten unemployment duration on average?
   - does a reduction in class size have a differential effect on male and female students?

1. **Testing and falsifying economic theories** empirically:
   - Are exchange rate markets for freely-traded currencies rational?
   - Are stock prices unpredictable?

2. **Quantifying relationships** between economic variables:
   - does education have a positive (causal) effect on wages? And if yes, what is the magnitude?
   - what is the average gender pay gap?

3. **Program evaluation** for policy analysis:
   - does a specific job training shorten unemployment duration on average?
   - does a reduction in class size have a differential effect on male and female students?

4. **Prediction (forecasting)** of economic variables:
   - by how much will GDP grow next year?
   - how volatile will stock markets be over the next week?
   - How large will be the demand for specific jobs in 10 years?

Example

A policy maker wonders how much money to invest in additional teachers and asks you to investigate empirically (i.e., using data):

- how does **class size influence school performance**?

Example

A policy maker wonders how much money to invest in additional teachers and asks you to investigate empirically (i.e., using data):

- how does **class size influence school performance**?

First, you develop a model for **test scores**, say y, as a function of **class size**, say x.

Example

A policy maker wonders how much money to invest in additional teachers and asks you to investigate empirically (i.e., using data):

- how does **class size influence school performance**?

First, you develop a model for **test scores**, say y, as a function of **class size**, say x.

Can we learn anything about this model from **the data**?

Example

A policy maker wonders how much money to invest in additional teachers and asks you to investigate empirically (i.e., using data):

- how does **class size influence school performance**?

First, you develop a model for **test scores**, say y, as a function of **class size**, say x.

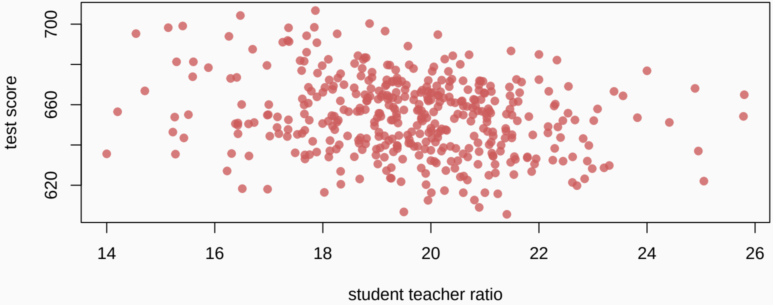
Can we learn anything about this model from **the data**?

Let's look at `CASchools`

- **data set** on **math and reading test performance**
- across 420 Californian school districts in the year 1999
- plot **test score** against the **class size** (both measured as averages in each district)

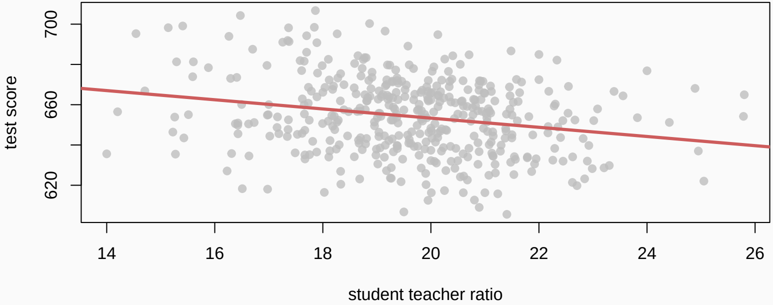
Scatter plot

Relationship is **not determinstic**: But is there some connection between class size and test scores?



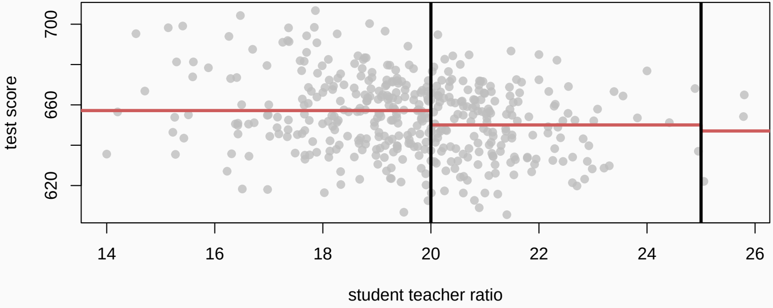
Scatter plot

A **fitted linear line**, indicates a negative relation



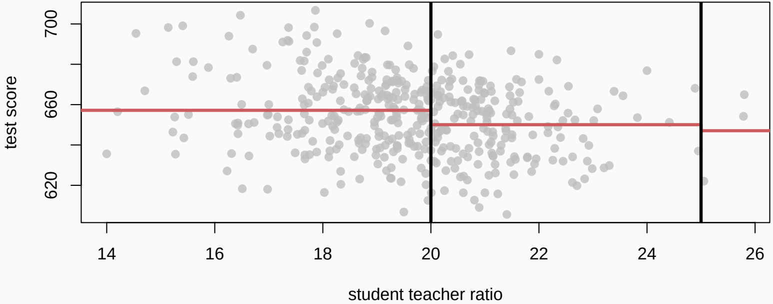
Scatter plot

Another approach: Divide observations in groups and calculate test score averages by group:



Scatter plot

- Group **small** (class size ≤ 20): average score $= 657$
- Group **medium** (class size > 20 & ≤ 25): average score $= 650$
- Group **large** (class size > 25): average score $= 647$



R code

```
library(AER)     # provides sample data sets
library(dplyr)   # provides data analyses tools
library(scales) # for nicer plots/transparency

data("CASchools")

CASchools <- CASchools |>
  mutate( student_teacher_ratio = students / teachers,
          test_score = (read + math)/2)

plot(CASchools$student_teacher_ratio,CASchools$test_score,
     xlab="student teacher ratio",
     ylab="test score",
     col=alpha("grey",0.7), pch=19)

abline( lm(test_score ~ student_teacher_ratio, dat=CASchools),
        col="indianred",
        lw=3)
```

R code

```
library(AER)    # provides sample data sets
library(dplyr)  # provides data analyses tools
library(scales) # for nicer plots/transparency

data("CASchools")

CASchools <- CASchools |>
  mutate( student_teacher_ratio = students / teachers,
          test_score = (read + math)/2)

plot(CASchools$student_teacher_ratio,CASchools$test_score,
     xlab="student teacher ratio",
     ylab="test score",
     col=alpha("grey",0.7), pch=19)

abline( lm(test_score ~ student_teacher_ratio, dat=CASchools),
        col="indianred",
        lw=3)
```

Question

Would you recommend increasing gov't spending to reduce class sizes based on these numbers?

Potential issues (that we can't solve yet)

Potential issues (that we can't solve yet)

Estimation uncertainty:

- the **size of subgroups** differs substantially
  - 243 districts have *small* and 174 have *medium* avg. class sizes

Potential issues (that we can't solve yet)

Estimation uncertainty:

- the **size of subgroups** differs substantially
  - 243 districts have *small* and 174 have *medium* avg. class sizes
  - only 3 districts have *large* avg. class size $\Rightarrow$ substantial **estimation uncertainty**

Potential issues (that we can't solve yet)

Estimation uncertainty:

- the **size of subgroups** differs substantially
  - 243 districts have *small* and 174 have *medium* avg. class sizes
  - only 3 districts have *large* avg. class size $\Rightarrow$ substantial **estimation uncertainty**

Causality:

- can we infer that smaller class sizes **caused** higher test scores
  - would large-class districts have performed better with smaller classes?

Potential issues (that we can't solve yet)

Estimation uncertainty:

- the **size of subgroups** differs substantially
  - 243 districts have *small* and 174 have *medium* avg. class sizes
  - only 3 districts have *large* avg. class size $\Rightarrow$ substantial **estimation uncertainty**

Causality:

- can we infer that smaller class sizes **caused** higher test scores
  - would large-class districts have performed better with smaller classes?
- or are we **comparing apples and oranges**?
  - districts with larger classes may be special in **other dimensions** (income, government spending, socio-economic status, school quality)

These factors (some of which may be **unobserved**) could be the actual **drivers** of the negative relationship between test scores and class size.

These factors (some of which may be **unobserved**) could be the actual **drivers** of the negative relationship between test scores and class size.

- i.e., class size may not be the **cause** of lower test scores, but just a **correlate**
- so a policy that changes class sizes would not actually affect test scores

These factors (some of which may be **unobserved**) could be the actual **drivers** of the negative relationship between test scores and class size.

- i.e., class size may not be the **cause** of lower test scores, but just a **correlate**
- so a policy that changes class sizes would not actually affect test scores

Econometrics is not just about applying **statistical tools** to **economic data**

These factors (some of which may be **unobserved**) could be the actual **drivers** of the negative relationship between test scores and class size.

- i.e., class size may not be the **cause** of lower test scores, but just a **correlate**
- so a policy that changes class sizes would not actually affect test scores

Econometrics is not just about applying **statistical tools** to **economic data**

- but also about rigorous **thinking**
    - to interpret the **empirical evidence**

These factors (some of which may be **unobserved**) could be the actual **drivers** of the negative relationship between test scores and class size.

- i.e., class size may not be the **cause** of lower test scores, but just a **correlate**
- so a policy that changes class sizes would not actually affect test scores

Econometrics is not just about applying **statistical tools** to **economic data**

- but also about rigorous **thinking**
  - to interpret the **empirical evidence** and
  - to choose the right statistical tool

These factors (some of which may be **unobserved**) could be the actual **drivers** of the negative relationship between test scores and class size.

- i.e., class size may not be the **cause** of lower test scores, but just a **correlate**
- so a policy that changes class sizes would not actually affect test scores

Econometrics is not just about applying **statistical tools** to **economic data**

- but also about rigorous **thinking**
  - to interpret the **empirical evidence** and
  - to choose the right statistical tool

Nobel Prices

- 2021: Joshua D. Angrist and Guido W. Imbens
    - “for their methodological contributions to the analysis of causal relationships”
    - related to our chapter on **instrumental variables**

Nobel Prices

- 2021: Joshua D. Angrist and Guido W. Imbens
  - “for their methodological contributions to the analysis of causal relationships”
  - related to our chapter on **instrumental variables**
- 2019: Abhijit Banerjee, Esther Duflo and Michael Kremer
  - “for their experimental approach to alleviating global poverty”
  - related to our chapter on **causality**

Nobel Prices

- 2021: Joshua D. Angrist and Guido W. Imbens
  - “for their methodological contributions to the analysis of causal relationships”
  - related to our chapter on **instrumental variables**
- 2019: Abhijit Banerjee, Esther Duflo and Michael Kremer
  - “for their experimental approach to alleviating global poverty”
  - related to our chapter on **causality**
- 2003: Robert F. Engle III
  - “for methods of analyzing economic time series with time-varying volatility (ARCH)”
  - related to our chapter on **time series**

Nobel Prices

- 2021: Joshua D. Angrist and Guido W. Imbens
  - "for their methodological contributions to the analysis of causal relationships"
  - related to our chapter on **instrumental variables**
- 2019: Abhijit Banerjee, Esther Duflo and Michael Kremer
  - "for their experimental approach to alleviating global poverty"
  - related to our chapter on **causality**
- 2003: Robert F. Engle III
  - "for methods of analyzing economic time series with time-varying volatility (ARCH)"
  - related to our chapter on **time series**

Causality, data types, data structures

Causality

Economists are often concerned with understanding causal effects

- of one variable (a price, a policy change, ...)
- on an outcome variable (demand, wages, unemployment spell, ...).

Causality

Economists are often concerned with understanding causal effects

- of one variable (a price, a policy change, . . .)
- on an outcome variable (demand, wages, unemployment spell, . . .).

Informal definition: Causality

We speak of ‘causality’ if **manipulating one variable** in isolation has a direct **consequence on another variable**.

Causality

Economists are often concerned with understanding causal effects

- of one variable (a price, a policy change, ...)
- on an outcome variable (demand, wages, unemployment spell, ...).

Informal definition: Causality

We speak of 'causality' if **manipulating one variable** in isolation has a direct **consequence on another variable**.

Ex.: Fertilizer on crop yield

Classical example from the early statistics literature: How does the use of fertilizer *affect* crop yields?

Causality

Economists are often concerned with understanding causal effects

- of one variable (a price, a policy change, ...)
- on an outcome variable (demand, wages, unemployment spell, ...).

Informal definition: Causality

We speak of 'causality' if **manipulating one variable** in isolation has a direct **consequence on another variable**.

Ex.: Fertilizer on crop yield

Classical example from the early statistics literature: How does the use of fertilizer *affect* crop yields?

What kind of **statistical experiment** could we design to infer such a causal relationship?

The gold standard evidence: Experiments

1. We want to generate variation: So we could divide the field into 100 equally spaced plots.
2. We then **randomly select**
   - 50 plots were we use fertilizer (**treated**)
   - 50 plots that we leave without fertilizer (**control**)
3. We measure crop yield on each plot

Estimating causal effects

These data allow us to estimate the **(average) causal effect** of fertilizer on crop yields:

- compare the average yield on **treated plots** with the **control plots**
- assumption: fertilizer only affects plots it is applied to (“no interference/spillovers’ ’)

Estimating causal effects

These data allow us to estimate the **(average) causal effect** of fertilizer on crop yields:

- compare the average yield on **treated plots** with the **control plots**
- assumption: fertilizer only affects plots it is applied to ("no interference/spillovers")

Such an experiment is called **Randomized Control Trial (RCT)**

Estimating causal effects

These data allow us to estimate the **(average) causal effect** of fertilizer on crop yields:

- compare the average yield on **treated plots** with the **control plots**
- assumption: fertilizer only affects plots it is applied to ("no interference/spillovers' ')

Such an experiment is called **Randomized Control Trial (RCT)**

If carried out properly, it comes as close as possible to an **ideal experiment** in the natural sciences

- data collected from RCTs are usually called **experimental data**
- Yet, in **social sciences** it is often **difficult** to obtain **experimental data**.

Estimating causal effects

These data allow us to estimate the **(average) causal effect** of fertilizer on crop yields:

- compare the average yield on **treated plots** with the **control plots**
- assumption: fertilizer only affects plots it is applied to (“no interference/spillovers’ ’)

Such an experiment is called **Randomized Control Trial (RCT)**

If carried out properly, it comes as close as possible to an **ideal experiment** in the natural sciences

- data collected from RCTs are usually called **experimental data**
- Yet, in **social sciences** it is often **difficult** to obtain **experimental data**.

Experiments in social sciences

Ex.: The causal effect of class size on test scores

Suppose, we study the **causal effect** of changing class sizes on test scores. An **RCT** would have to:

- randomly reallocate teachers
- prohibit students from changing schools
- and establish identical tests

- Questions that we cannot tackle with experiments:
  - "What is the effect of a school closures on student happiness?"
  - "What is the effect of a gas import stop from Russia for Germany?"

For those questions, we cannot rely on experimental data

Experiments in social sciences

Ex.: The causal effect of class size on test scores

Suppose, we study the **causal effect** of changing class sizes on test scores. An **RCT** would have to:

- randomly reallocate teachers
- prohibit students from changing schools
- and establish identical tests

- Questions that we cannot tackle with experiments:
  - "What is the effect of a school closures on student happiness?"
  - "What is the effect of a gas import stop from Russia for Germany?"

Experiments in social sciences

Ex.: The causal effect of class size on test scores

Suppose, we study the **causal effect** of changing class sizes on test scores. An **RCT** would have to:

- randomly reallocate teachers
- prohibit students from changing schools
- and establish identical tests

- Questions that we cannot tackle with experiments:
  - "What is the effect of a school closures on student happiness?"
  - "What is the effect of a gas import stop from Russia for Germany?"
- For those questions, we cannot rely on experimental data

Experiments in social sciences

Ex.: The causal effect of class size on test scores

Suppose, we study the **causal effect** of changing class sizes on test scores. An **RCT** would have to:

- randomly reallocate teachers
- prohibit students from changing schools
- and establish identical tests

- Questions that we cannot tackle with experiments:
  - "What is the effect of a school closures on student happiness?"
  - "What is the effect of a gas import stop from Russia for Germany?"
- For those questions, we cannot rely on experimental data

Experimental vs observational data

- Data can be obtained from a variety of sources/methods/settings.
  - These determine what we are able to learn from data.

Experimental vs observational data

- Data can be obtained from a variety of sources/methods/settings.
  - These determine what we are able to learn from data.
- **Experimental data became increasingly important**, but most data in the social sciences are **observational data**.

Experimental vs observational data

- Data can be obtained from a variety of sources/methods/settings.
  - These determine what we are able to learn from data.
- **Experimental data became increasingly important**, but most data in the social sciences are **observational data**.
  - is **not experimental**, i.e. has not been generated from an RCT
  - usually from **surveys** or **administrative records**

Experimental vs observational data

- Data can be obtained from a variety of sources/methods/settings.
  - These determine what we are able to learn from data.
- **Experimental data became increasingly important**, but most data in the social sciences are **observational data**.
  - is **not experimental**, i.e. has not been generated from an RCT
  - usually from **surveys** or **administrative records**

Experimental vs observational data

Downsides of observational data:

Experimental vs observational data

Downsides of observational data:

- more difficult to **infer the causal effect**

Experimental vs observational data

Downsides of observational data:

- more difficult to **infer the causal effect**
    - needs more "econometric work" than just comparing averages

Advantages:

- often larger in **scope** (exists where no experiments are done)
- often larger **scale** (may comprise the entire population of a country)
- typically reflects **real-life behavior** rather than artificial experimental setting

Experimental vs observational data

Downsides of observational data:

- more difficult to **infer the causal effect**
  - needs more "econometric work" than just comparing averages

Advantages:

- often larger in **scope** (exists where no experiments are done)
- often larger **scale** (may comprise the entire population of a country)
- typically reflects **real-life behavior** rather than artificial experimental setting

Experimental vs observational data

Downsides of observational data:

- more difficult to **infer the causal effect**
  - needs more “econometric work” than just comparing averages

Advantages:

- often larger in **scope** (exists where no experiments are done)
- often larger **scale** (may comprise the entire population of a country)
- typically reflects **real-life behavior** rather than artificial experimental setting

This class focuses on **observational data** and studies the **assumptions and situations** under which we can **learn about causal effects**.

Data structures – Cross-sectional data

1. **Cross-sectional data:**
   - **We observe** multiple individual units (individuals, firms, . . .) at one point in time

Data structures – Cross-sectional data

1. **Cross-sectional data:**
   - **We observe** multiple individual units (individuals, firms, ...) at one point in time
   - Usually units are denoted with subscript i, e.g., we record the income of n individuals and income of individual i is given by y_i with $i = 1, .., n$

Data structures – Cross-sectional data

1. **Cross-sectional data:**
   - **We observe** multiple individual units (individuals, firms, . . .) at one point in time
   - Usually units are denoted with subscript i, e.g., we record the income of n individuals and income of individual i is given by y_i with $i = 1, .., n$
   - Often units are assumed to be randomly sampled from a population

Data structures – Cross-sectional data

1. Cross-sectional data:

- **We observe** multiple individual units (individuals, firms, ...) at one point in time
- Usually units are denoted with subscript i, e.g., we record the income of n individuals and income of individual i is given by y_i with $i = 1, .., n$
- Often units are assumed to be randomly sampled from a population

Table 1: Mock data on income of individuals

id	income	age
1	0	68
2	384	23
3	0	34
4	1795	32

Data structures – time series data

2. Time series

Data structures – time series data

2. Time series

- **We observe** one unit at many different points in time (frequency: hourly, monthly, etc.)

Data structures – time series data

2. Time series

- **We observe** one unit at many different points in time (frequency: hourly, monthly, etc.)
- Usually units are denoted with subscript t, e.g. y_t with $t = 1, ..., T$, and T is the number of time points

Data structures – time series data

2. Time series

- **We observe** one unit at many different points in time (frequency: hourly, monthly, etc.)
- Usually units are denoted with subscript t, e.g. y_t with $t = 1, ..., T$, and T is the number of time points
- Typically dependence over time plays a central role, which complicates statistical inference

Data structures – time series data

2. Time series

- **We observe** one unit at many different points in time (frequency: hourly, monthly, etc.)
- Usually units are denoted with subscript t, e.g. y_t with $t = 1, ..., T$, and T is the number of time points
- Typically dependence over time plays a central role, which complicates statistical inference

Table 2: GDP of The Gambia

year	GDPpc	growth
2017	680	0.98
2018	733	1.08
2019	773	1.05
2020	757	0.98
2021	836	1.10

Data structures – pooled cross-sections and panel data

3. **Panel data** (sometimes: "longitudinal")

Data structures – pooled cross-sections and panel data

3. **Panel data** (sometimes: "longitudinal")

- **We observe** multiple units across multiple points in time
- Observations typically indexed by it. So y_{it} with, $i = 1, ..., n$; and $t = 1, ..., T$

Data structures – pooled cross-sections and panel data

3. **Panel data** (sometimes: “longitudinal”)

- **We observe** multiple units across multiple points in time
- Observations typically indexed by it. So y_{it} with, $i = 1, ..., n$; and $t = 1, ..., T$
- Observations are, e.g, student-semester, country-year, or city-month

Data structures – pooled cross-sections and panel data

3. **Panel data** (sometimes: "longitudinal")

- **We observe** multiple units across multiple points in time
- Observations typically indexed by it. So y_{it} with, $i = 1, ..., n$; and $t = 1, ..., T$
- Observations are, e.g, student-semester, country-year, or city-month

Table 3: City populations over time

year	city	population
2018	Vienna	1,889,000
2019	Vienna	1,890,000
2020	Vienna	1,911,000
2018	Frankfurt/Main	769,000
2019	Frankfurt/Main	777,000
2020	Frankfurt/Main	785,000

Recap

Recap

- Questions to be answered using econometrics:
  - Falsifying theories, quantifying relationships, treatment evaluation, forecasting
- Example of Californian school districts:
  - Visual analysis, basic quantitative analysis
  - Issues: Estimation uncertainty, causality
- Causality
- Experimental vs observational data
- Data structures (cross-sectional, time series, panel)

Course outline

Course outline (1)

- will be updated

1. Introduction
   - Motivation
   - Econometric data, problems, and analyses (W ch. 1, SW ch. 1)
   - Review of probability & statistics (W appendix B-C, SW ch. 2+3)
2. Bivariate regression (W ch. 2, SW ch. 4)
   - The simple regression model
   - OLS estimator
   - Basic properties of OLS
   - Variance estimation

Course outline (2)

3. Multivariate regression (W ch. 3, SW ch. 6)
   - Matrix notation
   - Derivation and mechanics of OLS
   - Omitted variable bias
   - Gauss-Markov
   - Inference and testing (W ch. 4, SW ch. 7)
4. Further topics in regression analyses
   - Functional form, dummy variables (W ch. 6–7, SW ch. 8)
   - Testing (SW ch. 5)
   - Heteroscedasticity and generalized error term structures (W ch. 8, SW ch. 5)
   - Measurement error, missing data (W ch. 9)

Course outline (3)

5. Time series (SW ch. 15)
6. Causality
   - Instrumental variable estimation (W ch. 15, SW ch. 12)
7. Machine Learning (SW ch. 14)
   - Lasso

W: J. Wooldridge (2019), 7th edition: **Introductory Econometrics**, Cengage Learning.

SW: J. Stock and M.Watson (2019): **Introduction to Econometrics**, Pearson.

How to study effectively?

How to study effectively?

- Comes last, because it is the most important point today
- Econometrics is relatively abstract
  - do not binge, study regularly
  - Reiterate. Rest and sleep inbetween. It get's easier each time.
  - Your mind is not a computer. Reiterate in different forms.
  - If you are stuck, change perspective and try to find the simplest thing you can understand. Ask. Sometimes just formulating the question is helpful.
  - Do not just sit there and read/watch, instead summarize slides, try exercises without solutions, etc.

How to study effectively?

- Comes last, because it is the most important point today
- Econometrics is relatively abstract
  - do not binge, study regularly
  - Reiterate. Rest and sleep inbetween. It get's easier each time.
  - Your mind is not a computer. Reiterate in different forms.
  - If you are stuck, change perspective and try to find the simplest thing you can understand. Ask. Sometimes just formulating the question is helpful.
  - Do not just sit there and read/watch, instead summarize slides, try exercises without solutions, etc.

If you go on a journey, take someone with you.

- Form study groups (2-5 persons)
    - Easier to get started and more fun
    - Meet and discuss lectures or try exercises
    - Ask on OLAT or write me a mail. I will connect you randomly.
    - Study groups also make sense if experience differs: Explaining and asking is also very effective.
    - Fixed time day before or after the lecture makes sense

Appendix