ECON-C4100 - Capstone: Econometrics I

Otto Toivanen

Lecture 1

Teachers

Professor Otto Toivanen

Economics dept. Aalto U.

otto.toivanen@aalto.fi

Office hours: on appointment.

TA Jaakko Markkanen

Economics dept. Aalto U.

jaakko.m.markkanen@aalto.fi

Office hours: on appointment.

Objectives of this lecture

By the end of the lecture, you

- understand why we have econometrics as a mandatory part of B.Sc. studies in economics.
- know the learning objectives of the course(s).
- understand how the course is organized.
- understand how your performance on the course is assessed.

1. Interesting questions

- How should we price our (new) product(s)?
- Is our advertising working?
- Is our new incentive scheme delivering results?
- What affects the probability of defaulting on a mobile phone credit?

Interesting questions

- What is the return to education?
- What are the welfare effects of a merger?
- Does parental education affect children's job market outcomes?
- Do informational nudges affect tax avoidance?
- Who benefits from public transport investments?

How this course can help

- Oftentimes, you sit on the answer $\rightarrow$ data.
- Getting a roughly-right answer not that difficult: apply statistics / econometrics.
- When you don't sit on the answer, you may get it with a little work $\rightarrow$ experiments.
- To do all this intelligently, you need (economic) theory.

2. What an increasing # firms already do

- Collect data on their own performance – electricity companies.
- Collect data on their customers – retail stores, telecom operators.
- Sometimes, of markets or rivals - Nielsen...
- Analyze these data: A/B Testing
- Increasingly, experiments: Google, airlines, garment stores...

What an increasing # governments do

- Collect data.
- Plan in advance how to evaluate government policies.
- (Try to / claim to) base decisions on empirical evidence.

What an increasing # governments do

- They institute formal evaluations of implemented policies.
- Example: EU's competition authority (DG Competition) has the authority to block state aid unless they approve the evaluation plan of a member state for that state aid policy (example: Finnish R&D subsidies).

3. Two megatrends

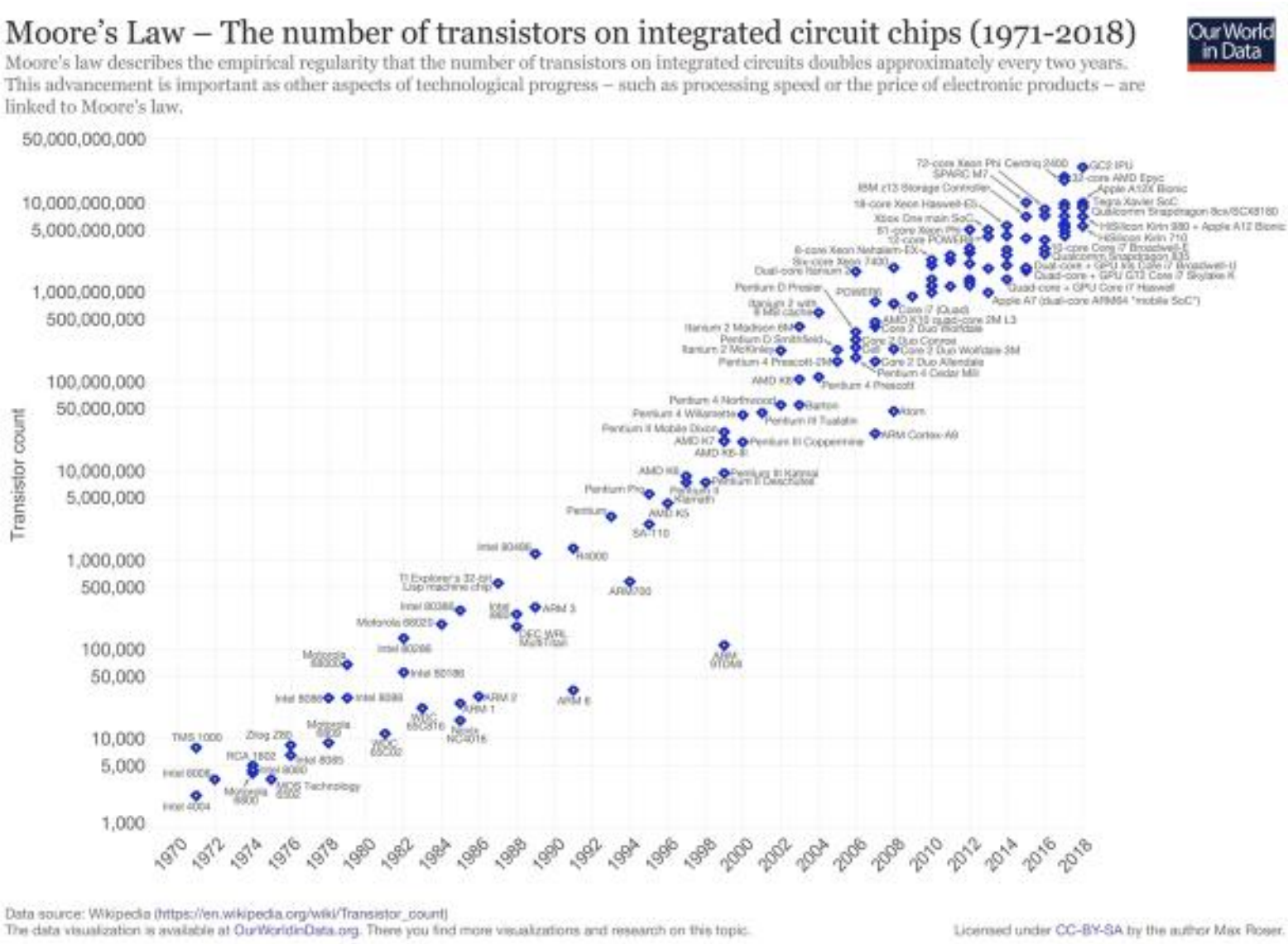
1. Increasing amount of (digital) data available.
2. Huge increases in computing power / decline in cost of computing.

- These trends are reflected in how economics is useful in the real world.
- They are reflected in what economics research is about and how it is conducted.

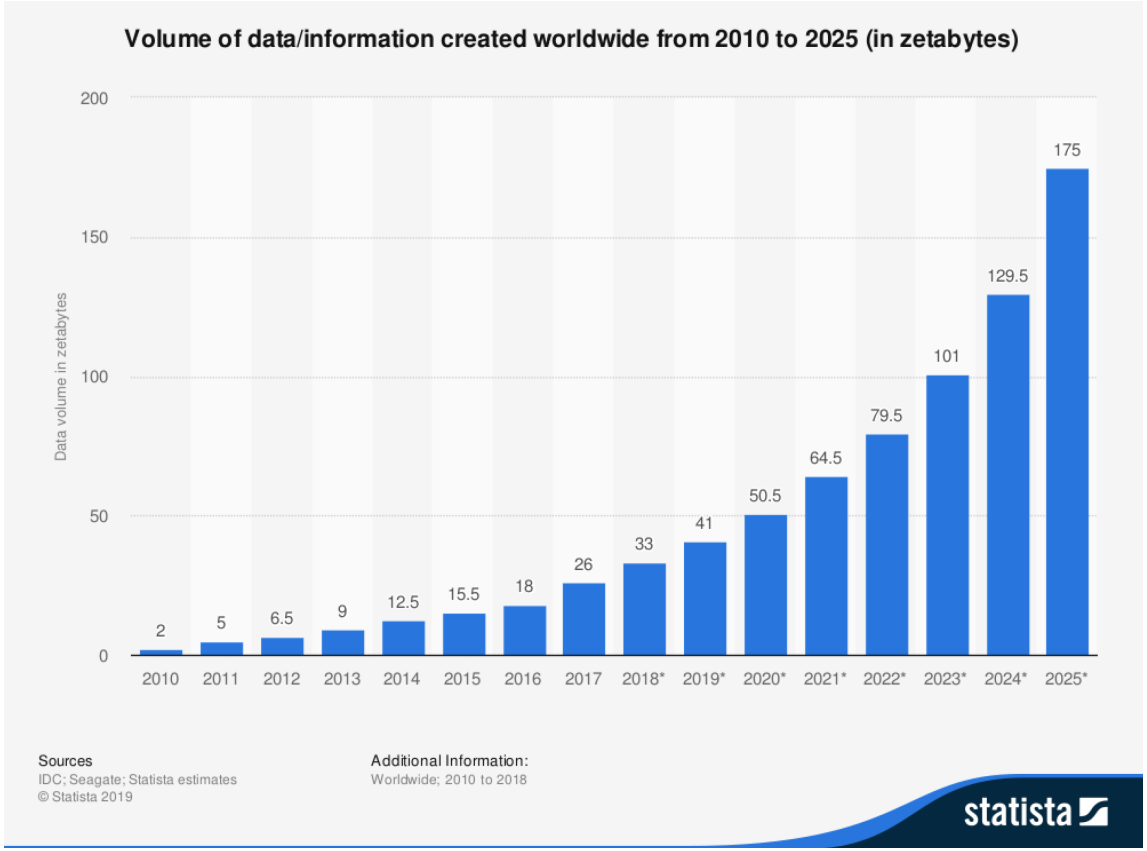
Some statistics about statistics / data

- Facebook stores, accesses, and analyzes **30+ Petabytes** of user generated data. (peta $= 10^{15} = 1000\,TB$).
- Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data.

Computing power has exploded



How much information is out there?



Source: Statista.com. 1 zetabite = 1 billion terabytes = 1000^7 bytes.

Even text has become data...

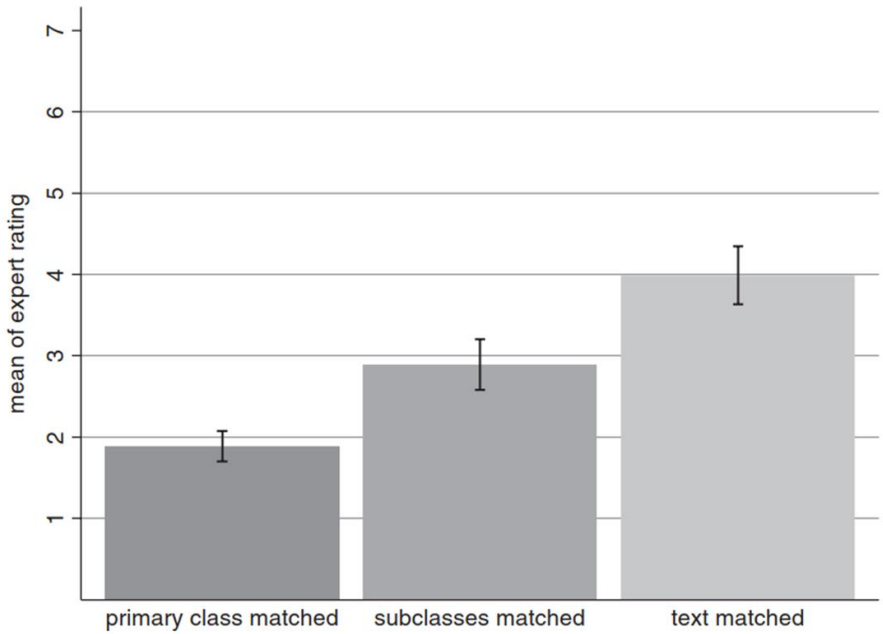


FIGURE 5 Means of expert ratings of technological similarity of primary-class-matched, subclasses-matched, and text-matched patent pairs. *Notes:* The sample is constructed by selecting for each field of expertise a random sample of baseline patents, and for each baseline patent a primary-class-matched, a subclasses-matched, and a text-matched patent. The order of the patent pairs is randomized, and the experts rate the similarity of the patent pairs in their field on a Likert scale from 1 to 7. Matched patents from the same patent family are excluded. The sample consists of 297 ratings conducted by 11 experts from three different fields. The figure displays the means of expert ratings and the corresponding 95% confidence intervals for the primary-class-matched, subclasses-matched, and text-matched patent pairs respectively

Arts, S., Cassiman, B. & Gomez, J. (2018). Text matching to measure patent similarities. *Strategic Management Journal*, *39*, 62–84

4. Trends in research in economics

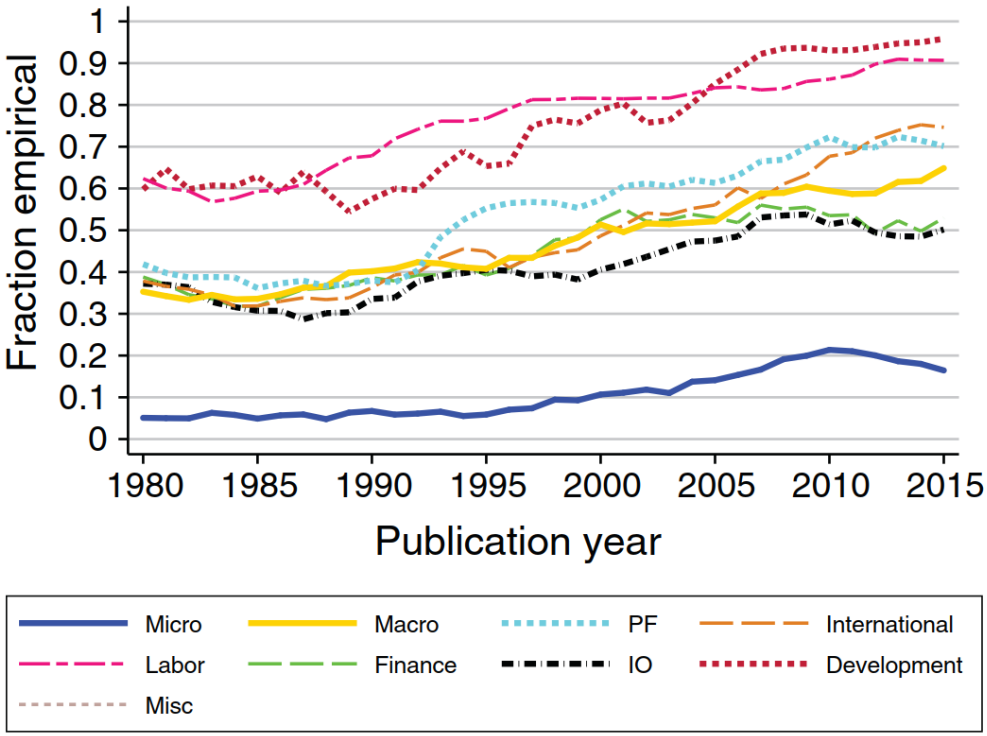


FIGURE 4. WEIGHTED FRACTION EMPIRICAL BY FIELD

Note: Five-year moving averages of the weighted fraction of publications in each field that are empirical.

Angrist, J., Azoulay, P., Ellison, G., Hill, R. & Lu, S. F. (2017). Economic research evolves: Fields and styles. *American Economic Review*, *107*(5), 293–97

Trends in research in economics

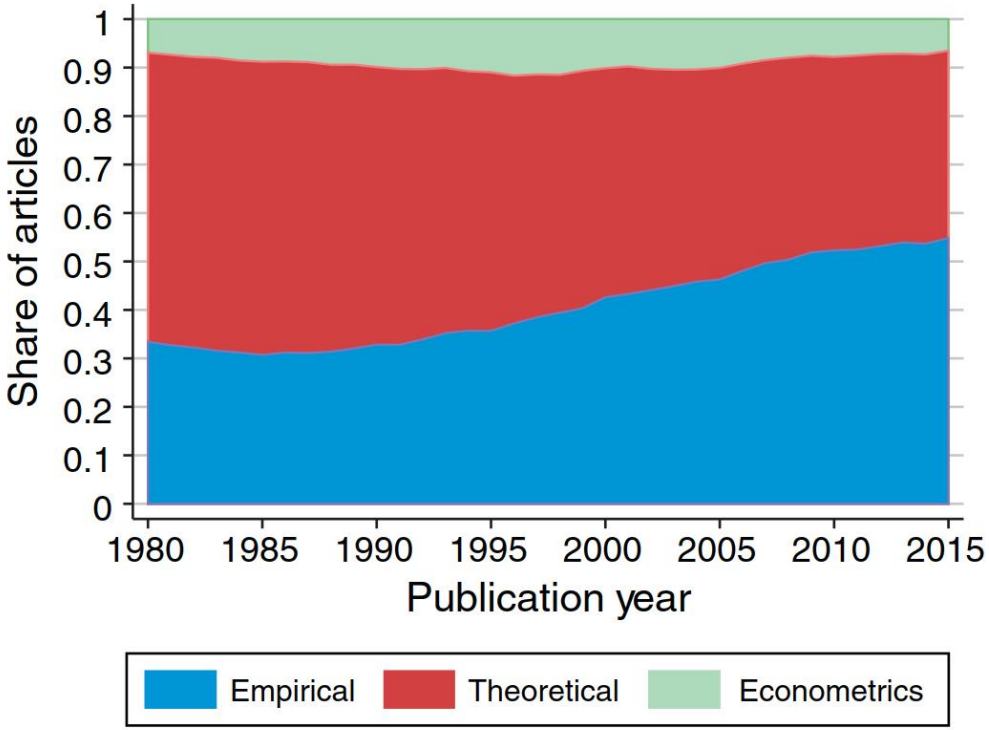


FIGURE 6. WEIGHTED PUBLICATIONS BY STYLE

Note: Five-year moving averages of weighted publication shares in each style.

Angrist, J., Azoulay, P., Ellison, G., Hill, R. & Lu, S. F. (2017). Economic research evolves: Fields and styles. *American Economic Review*, *107*(5), 293–97

These trends are reflected in how economics is useful in the real world

- Economics research is more and more policy-oriented.
- Economics is more and more used by private companies (e.g. Google, Amazon, Nokia, ...)
- Watch "Letting the data take you in new directions" by Rachel Griffith (U. of Manchester).
- Watch "Fixing society with data" by Richard Blundell (UCL).

5. What this course is about

- The ABC of how to
  1. understand the answers to
  2. evaluate the quality of answers to and
  3. provide an answer to

  the type of questions posed above.

What this course is about

- Tools: economic theory + statistical tools + data + knowledge. In short: econometrics.
- Learning outcomes: Students
  1. are acquainted with the principles of empirical methods in economics.
  2. know how to perform descriptive analysis of data.
  3. are acquainted with econometrics methods for cross-section data.
  4. understand the difference between descriptive and causal analysis.
  5. have basic knowledge of the econometrics software package Stata.
  6. know the basics of how to program, how to document and how to ensure replicability of their econometric analysis.

What this course is about

- Econometrics:

  A branch of economics in which economic theory and statistical methods are fused in the analysis of numerical and institutional data

  Hood, W. & Koopmans, T. (1953). Studies in econometric method. *Cowles Commission Monograph no. 14, Wiley*

What this course is about

- **Economic theory:** modeling economic decisions requires understanding how and why they are made.

What this course is about

- **Economic theory:** modeling economic decisions requires understanding how and why they are made.
- **Statistical methods:** tools to derive numbers from numbers to add to knowledge.

What this course is about

- **Economic theory:** modeling economic decisions requires understanding how and why they are made.
- **Statistical methods:** tools to derive numbers from numbers to add to knowledge.
- **Numerical data:** the raw material to be explained, and to be used to explain.

What this course is about

- **Economic theory:** modeling economic decisions requires understanding how and why they are made.
- **Statistical methods:** tools to derive numbers from numbers to add to knowledge.
- **Numerical data:** the raw material to be explained, and to be used to explain.
- **Institutional data:** the environment in which the numerical data arises.

What this course is also about

- How to interpret research results.
- How to evaluate research.
- How to conduct (good) research (thesis?).

6. Approach

1. Practical rather than theoretical, though both covered.
2. Hands-on learning of each step.
3. Use of (business-) relevant data for illustration & learning.
4. Own (group) work.
5. Emphasis on methods for **cross-section** data.

The Main Data

- Pharmaceutical price data (Salvia).
- Prices, product characteristics etc.
- Twice a month cross-sections from 2002 to 2020.
- Source: The Association of Finnish Pharmacies.

7. Contents & what affects grading

- Book: Stock, J. H. & Watson, M. W. (3rd Edition). *Introduction to econometrics*.
- **Lectures & exam:** follow the book, though we try to make as much use of the pharmaceuticals data as possible.
- **Exercises:** some surely dull, some hopefully interesting, all useful.
- **Capstone work.**

Course evaluation

- Exercises 20%
- Capstone 35%
- Exam 45%
  - Course exam 25.02.2021
  - Retake exam 07.05.2021

7.1 Lectures – current plan

12.1 Lecture 1&2 introduction, descriptive statistics and estimation of the mean ch2&3 of Stock and Watson

14.1 Lecture 2 statistics recap - estimation of the mean ch2&3

19.1 Lecture 3 univariate regression #1 ch4&5

21.1 Lecture 4 univariate regression #2 ch5

Lectures – current plan

26.1 Lecture 5 univariate regression c'ed

28.1 Lecture 6 multiple regression #1: estimation ch6

2.2 Lecture 7 multiple regression #2: interpretation, testing ch7

4.2 Lecture 8 multiple regression #3: problems ch8

Lectures – current plan

9.2 Lecture 9 causal parameters #1: experiments & problems with observational data ch13.1, 13.2, 6.1

11.2 Lecture 10 causal parameters #2: instrumental variables ch12

16.2 Lecture 11 causal parameters #3: IV c'ed

18.2 Lecture 12 recap

7.2 Exercises and Problem Sets

- 5 graded problem sets and 6 exercise sessions.
- Problem sets are published a week before the deadline. All deadlines are before the start of the next exercise session (14:00 EET).
- Problem sets have equal weight and include both analytical and empirical problems.
- You need at least 50% of points to pass the course.

Exercises and Problem Sets

- Deadlines are strict - do not email us your solutions.
- Plagiarism is strictly forbidden. Do not share your answers or code. You can discuss the exercises in small groups but all answers must be self-written.
- Detailed instructions are found on MyCourses.

Exercises and Problem Sets

Introduction to Stata 15.01.

Problem Set 1 - 22.01.

Problem Set 2 - 29.01.

Problem Set 3 - 05.02.

Problem Set 4 - 12.02.

Problem Set 5 - 19.02.

Software

- This course uses Stata software but you can use any software you like. However, support is only given in Stata.
- Aalto U. has a campus license for Stata, so it is available for all students.
- The Department of Economics policy is to use Stata in all course-related empirical work.
- Applied Econometrics I and II, Labour Economics, Empirical Industrial Organization, ...

7.3 Capstone work

- Objective: students learn to
  1. identify an empirical research question
  2. relate it to what is already known
  3. analyze it using econometric tools
  4. present their analysis.
- In other words, students learn how to conduct and evaluate empirical research.
- These skills are valuable not only in academia, but in an increasing number of "real-life" jobs.

Capstone work

- Capstone group work: 2 - 3 students.
  1. Formulate your research question
  2. Collect your (preferably) own data
  3. Analyze
  4. Report
- Report, length 5 - 10 pages. An appendix of max. 5 pages allowed.
- Assessment based on report and presentation.

7.3 Capstone work

- Two modes of handing in:
  1. by end of March
     1. tutoring session: Th. March 4, 14 - 15.
     2. presentations: Th. March 25, 9 - 11.
  2. by end of April
     1. tutoring session: Th. April 8, 13.30 - 15
     2. presentations: Wed. April 21 and Th. April 22, 9 - 15.

Capstone FAQ

- Can I work alone? No.
- Can the report be longer? No, but an appendix of reasonable size (= shorter than the report) allowed.
- What is a good topic?
  1. Something that interests you.
  2. Something you can find data on.
- Does it matter whether the work is handed in in March or April? Not for grading, but the April deadline allows for a deeper and wider toolbox.

Key ingredients of the Capstone report

1. Research question. This needs to be specific enough.
2. Motivation of the research question - why should one bother?
3. What is already known about the topic?
4. Data.
5. Economic Theory.

8. Relation to Econometrics II

- The follow-up course of this course is ECON-C4200, Econometrics II.
- It takes place 2.3 - 12.4.2021.
- This course is the prerequisite for Econometrics II.
- In Econometrics II, the emphasis is on methods designed for **panel** and **time-series** data whereas in this course we concentrate on **cross-section** data and methods.