Bachelor Thesis Seminar

Econometrics

Barbara Pertold-Gebicka

19.3.2024

A little about myself

Barbara Pertold-Gebicka

- PhD from CERGE-EI
- Assistant Professor at IES, teaching Econometrics II (JEB 110) and Applied Microeconometrics (JEM 007)
- Research in the fields of labor economics, economics of education, family economics, applied microeconometrics
- Office hours: Tuesday 14-15:15 (room #408)
- E-mail: gebicka@fsv.cuni.cz

What is econometrics?

"Econometrics is the quantitative application of statistical and mathematical models using data to [...] test existing hypotheses in economics and to forecast future trends from historical data."

[Adam Hayes, derivatives trader]

"Econometrics uses economic theory, mathematics, and statistical inference to quantify economic phenomena. In other words, it turns theoretical economic models into useful tools for economic policy making. “
[Sam Ouliaris, Senior Economist in the IMF Institute]

Statistics, Economics, and Data

- Statistics is the baseline for econometrics.
- Econometrics combines economic theory with statistics to analyze and test economic relationships.
- Proper usage of econometric tools requires understanding of statistics, economics, and the nature of the data.

Examples of econometrics usage

- Macroeconomic forecasting, e.g.
  what will be the inflation rate in the following quarter?
- Credit scoring, e.g.
  which personal characteristics predict high probability of not repaying a loan?
- Evaluation of policies, e.g.
  what is the effect of smoking ban on restaurant sales?
  what is the effect of covid-19 policies on countries' economic outcomes?
- Testing economic theories, e.g.
  does Phillips curve exist?

Econometrics

- Econometrics is a powerful tool
- It is easy to use since the development of fast-processing computers
- **But always remember about the underlying assumptions!!!**
- After several lectures of Econometrics I you will be able to
  - Run a multiple regression
  - Interpret the results
  - Test simple hypotheses about regression coefficients
- Why not to use econometric analysis in your bachelor thesis?

Sample theses written under my supervision

- Demand for gas: Evidence from the 2022 energy crisis. (2023)
- Gender gap in math score: does teacher gender matter? (2022)
- The Impact of COVID-19 on Students' Academic Performance. (2022)
- Do fringe benefits affect job satisfaction? (2020)
- Do family policies really affect fertility levels? (2019)
- Occupational regulation and its influence on the labor market: evidence from reforms in the Czech Republic. (2017)
- The relationship between capital structure and performance: empirical study of Czech manufacturing companies. (2016)
- Impact of metro station proximity on apartment value in Prague. (2015)

Let the data speak

- All these theses have one thing in common…
- …their authors learned how to “speak with the data”
- This involves:
  - Asking a question that can be tackled
  - Understanding which forces can affect the observed outcomes
  - Gathering the proper data
  - Performing econometric analysis
  - Interpreting the results

Econometric analysis: an example

Does lecture attendance affect exam performance?

Does lecture attendance affect exam performance?

Why might there be a relationship between lecture attendance and exam performance?

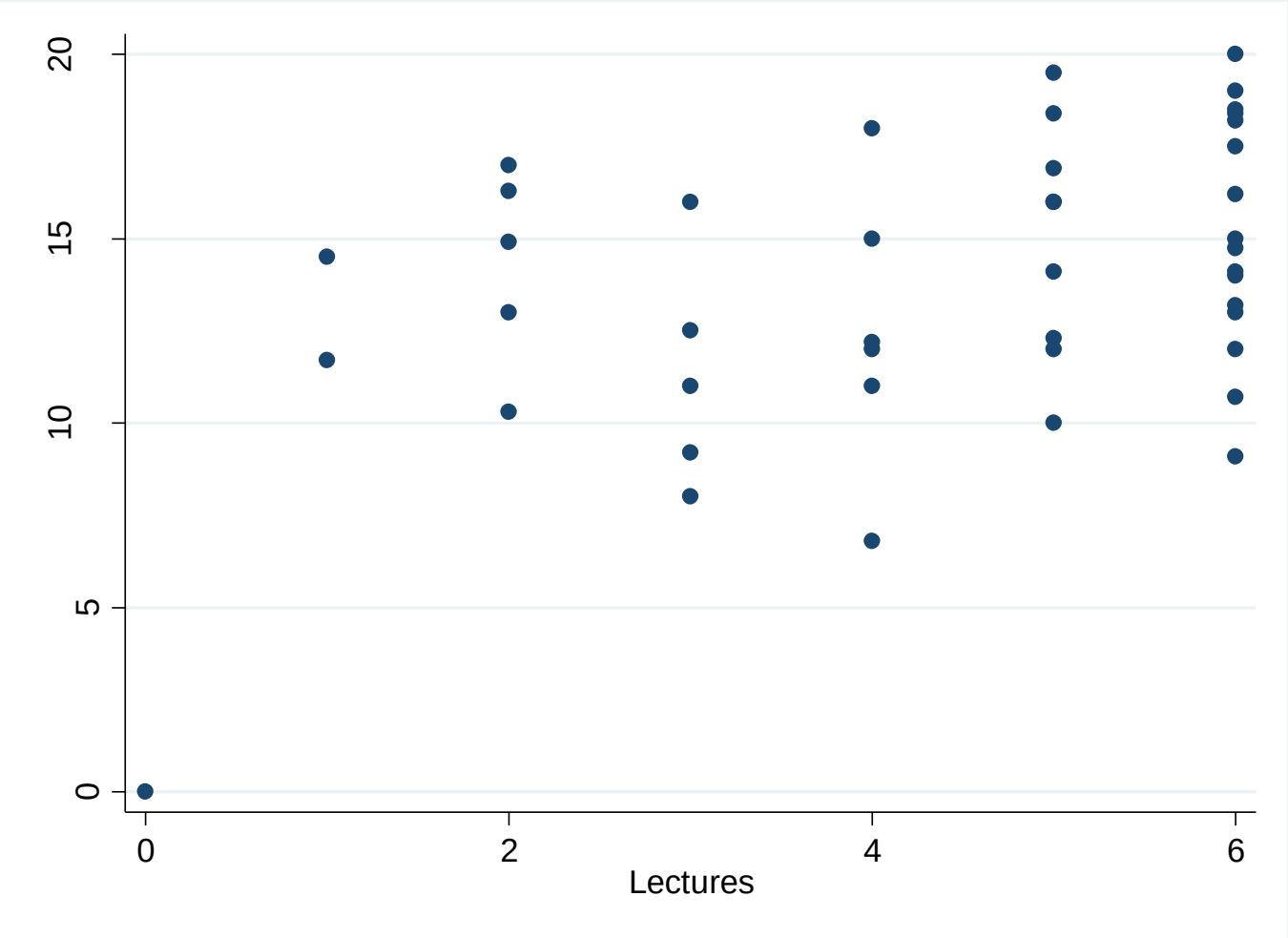
- Students learn during classes
- Students attending classes know what the teacher expects
- Highly motivated students attend more classes
- Smart students attend less classes (they don’t need to)
- Students who have a job attend less classes (they also have less time to study at home)
- ….

Does lecture attendance affect exam performance?

How to test this hypothesis?

- Collect data about students
  - Official school records
  - Questionnaire
- What do we need to know?
  - Number of lectures attended
  - Points gathered during an exam
  - Additional information, e.g.
    - how well a student performed in a related subject / prerequisite?
    - how easy/difficult it is for the student to attend classes?

Does lecture attendance affect exam performance?



Real data from the Econometrics II course a few years ago.

Does lecture attendance affect exam performance?

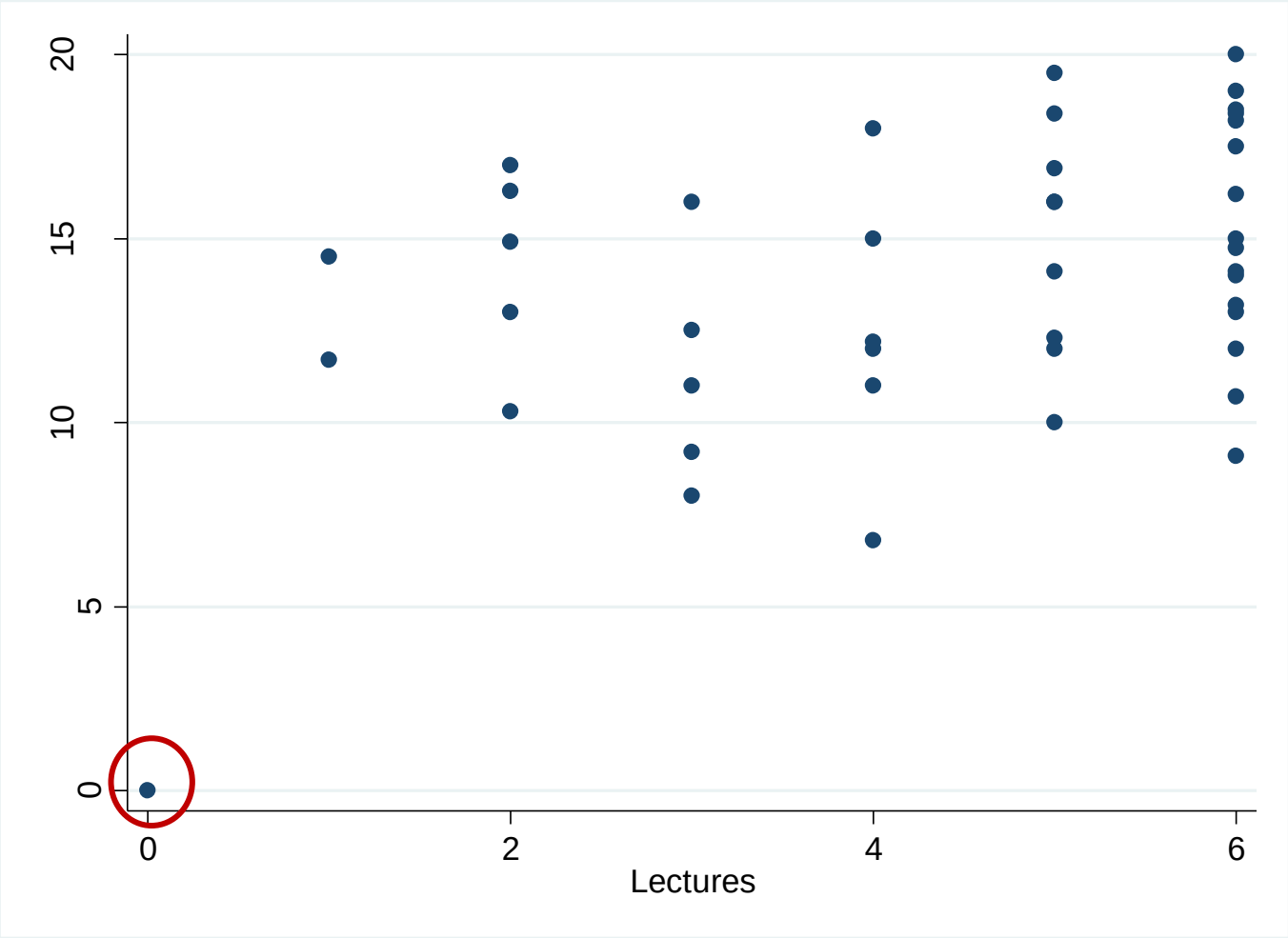
How to test this hypothesis?

- Build a regression model

$$midterm_i = \beta_0 + \beta_1 \cdot lectures_i + u_i$$

  where:
  - $midterm_i$ is the number of points obtained in the midterm exam
  - $lectures_i$ is the number of lectures attended
  - u_i is the random error (disturbance)
- The coefficient β_1 can be interpreted as follows:
  - When student attends one more lecture, his/her midterm score is higher by β_1.

Does lecture attendance affect exam performance?



Correlation coefficient: 0.428

Simple OLS regression:

$midterm_i = 9.570 + 0.975 \cdot lectures_i$

<u>After dropping the outlier</u>

Correlation coefficient: 0.277

Simple OLS regression:

$midterm_i = 11.580 + 0.574 \cdot lectures_i$

Does lecture attendance affect exam performance?

	(1)	(2)
Lectures	0.975	0.574
	(0.318)	(0.312)
Constant	9.570	11.580
	(1.489)	(1.476)
Observations	44	43
R-squared	0.183	0.076

Does lecture attendance affect exam performance?

	(1)	(2)
Lectures	0.975***	0.574**
	(0.318)	(0.312)
Constant	9.570***	11.580***
	(1.489)	(1.476)
Observations	44	43
R-squared	0.183	0.076

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Full sample (column 1), outliers dropped (column 2).

Does lecture attendance affect exam performance?

Which of these is actually the **effect** of lecture attendance?

- Students learn during classes
- Students attending classes know what the teacher expects
- Highly motivated students attend more classes
- Smart students attend less classes (they don’t need to)
- Students who have a job attend less classes (they also have less time to study at home)

Does lecture attendance affect exam performance?

Which of these is actually the **effect** of lecture attendance?

- Students learn during classes
- Students attending classes know what the teacher expects
- Highly motivated students attend more classes
- Smart students attend less classes (they don’t need to)
- Students who have a job attend less classes (they also have less time to study at home)

Does lecture attendance affect exam performance?

What relationship we observe if the following is true?

- Students learn during classes
  -> positive relationship
- Students attending classes know what the teacher expects
  -> positive relationship
- Highly motivated students attend more classes
  -> positive relationship
- Smart students attend less classes (they don’t need to)
  -> negative relationship
- Students who have a job attend less classes (they also have less time to study at home)
  -> positive relationship

Does lecture attendance affect exam performance?

What relationship we observe if the following is true?

- Students learn during classes
  -> positive relationship
- Students attending classes know what the teacher expects
  -> positive relationship
- Highly motivated students attend more classes
  -> positive relationship
- Smart students attend less classes (they don’t need to)
  -> negative relationship
- Students who have a job attend less classes (they also have less time to study at home)
  -> positive relationship

Does lecture attendance affect exam performance?

	(1)	(2)
Lectures	0.975***	0.574**
	(0.318)	(0.312)
Constant	9.570***	11.580***
	(1.489)	(1.476)
Observations	44	43
R-squared	0.183	0.076

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Full sample (column 1), outliers dropped (column 2).

Do these estimates mean that making lectures compulsory will improve students' grades?

Does lecture attendance affect exam performance?

- Basic regression model

$$midterm_i = \beta_0 + \beta_1 \cdot lectures_i + u_i$$

- OLS estimation results:

$$midterm_i = 11.580 + 0.574 \cdot lectures_i$$

$$i.e. \ \widehat{\boldsymbol{\beta}}_{\mathbf{1}} = \mathbf{0.574}$$

- This positive estimate might be driven by
  - Students really benefitting from lecture attendance
  - The fact that more motivated students (who put more effort into studying anyway) attend more lectures
- How to disentangle these two?
  - How to clear the estimations from the latter effect?

Does lecture attendance affect exam performance?

- Simple regression model

$$midterm_i = \beta_0 + \beta_1 \cdot lectures_i + u_i$$

- Augmented regression model

$$midterm_i = \beta_0 + \beta_1 \cdot lectures_i + \beta_2 \cdot Statistics_i + u_i$$

- The coefficient β_1 can now be interpreted as follows:
  - Given the Statistics grade, when a student attends one more Econometrics lecture, his/her midterm score is higher by β_1.
  - Conditional on the Statistics grade, when student attends one more Econometrics lecture, his/her midterm score is higher by β_1.
  - When we compare two random students with the same Statistics grade, the one who attended one more Econometrics lecture has, on average, the midterm score higher by β_1 points.

Does lecture attendance affect exam performance?

	(1)	(2)	(3)
Lectures	0.975***	0.574**	0.541***
	(0.318)	(0.312)	(0.249)
Statistics			-2.439***
			(0.495)
Constant	9.570***	11.580***	16.665***
	(1.489)	(1.476)	(1.567)
Observations	44	43	43
R-squared	0.183	0.076	0.425

Note: Dependent variable is the number of lectures attended; Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Full sample (column 1), outliers dropped (columns 2 and 3).

Does lecture attendance affect exam performance?

	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	2SLS
Lectures	0.975***	0.574**	0.541***	0.627
	(0.318)	(0.312)	(0.249)	(0.516)
Statistics			-2.439***	-2.434***
			(0.495)	(0.497)
Constant	9.570***	11.580***	16.665***	16.272***
	(1.489)	(1.476)	(1.567)	(2.596)
Observations	44	43	43	43
R-squared	0.183	0.076	0.425	0.423

Note: Dependent variable is the number of lectures attended; Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Full sample (column 1), outliers dropped (columns 2 and 3).

Does lecture attendance affect exam performance?

Interpreting the results

- The estimated coefficient is about 0.54 to 0.63. Is this a large effect?
- Summary statistics of the underlying data needed!

Variable	Mean	St. deviation
Lectures	4.465	1.594
Midterm	14.145	3.310

- Attending one additional lecture translates to the midterm score higher by ~0.6 points
- Increasing lecture attendance from zero to six corresponds to the midterm score increase by more than one standard deviation ($0.6 \cdot 6 = 3.6$).
- Attending one additional lecture translates to the midterm exam score increased by 4.2%, for an average student ($0.6 \div 14.145 = 0.042$).

Does lecture attendance affect exam performance?

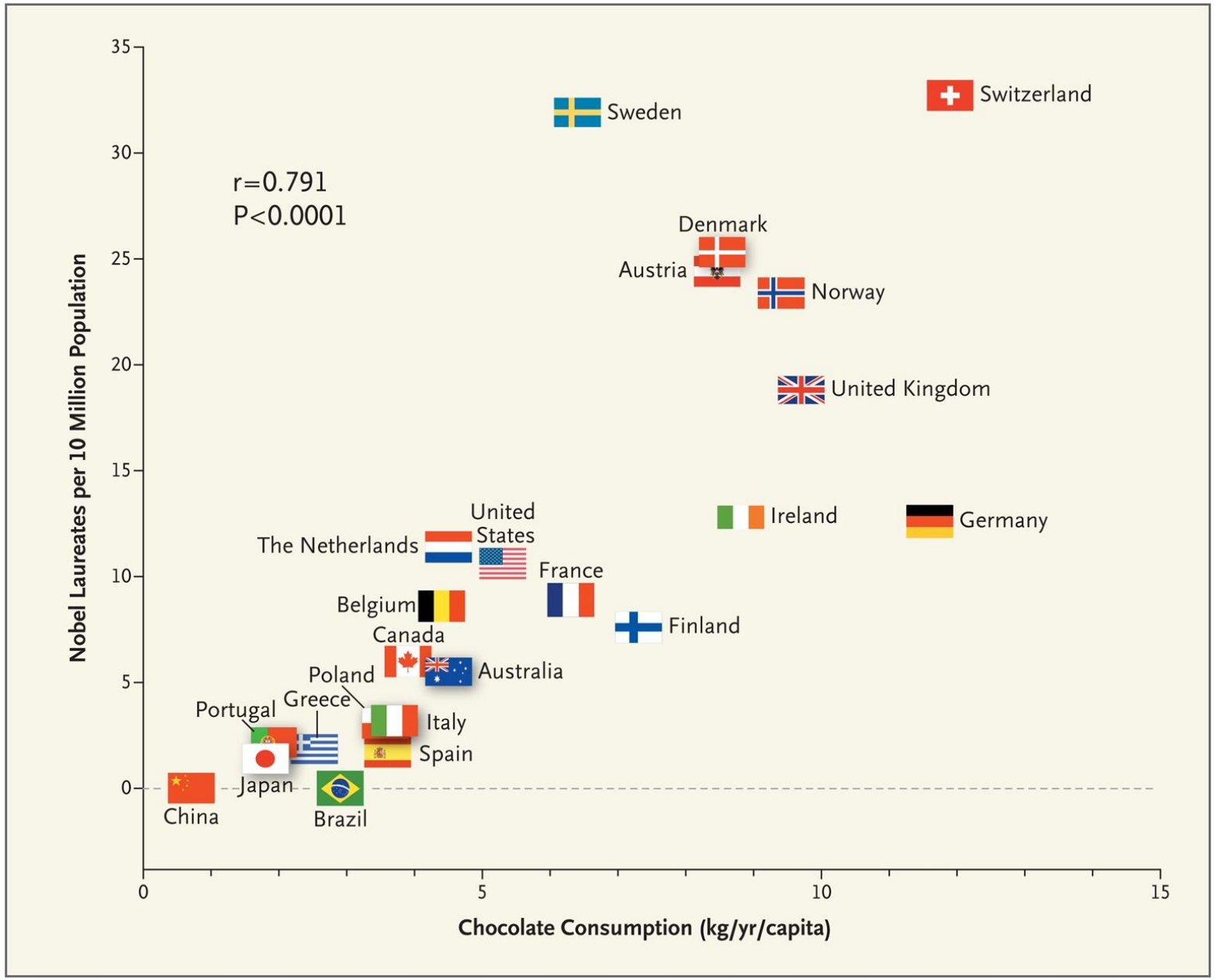
Robustness of the results

- Note that results in columns 2, 3, and 4 produce similar point estimate. This is a good sign!
- Inspection of the data suggests there might be heteroskedasticity (variation in midterm points is higher for many lectures attended than for few lectures attended).
  - Compute heteroskedasticity robust st. errors and see if estimates still significant!
- Functional form
  - Should we use midterm results in levels or maybe in logs?
- Need for more data?
  - Maybe collecting the data for the whole course (12 lectures) and performance in final exam will give more precise results?

Econometric analysis in your thesis

1. Choose a topic that you understand, carefully formulate the research question (testable hypothesis).
2. Think about a simple econometric model that could be used to tackle the research question (test the hypothesis).
3. Make sure that you have access to data that will be used to estimate the econometric model.
4. Now comes the fun part
   - Think deeper about the model formulated in point 2.
   - Is the model correctly specified?
   - Is the relationship between x and y driven only by your hypothesized channel? (will OLS provide unbiased estimates)
   - Can you disentangle between different channels?

$$Nobel_i = \beta_0 + \beta_1 \cdot chocolate_i + u_i$$



Source: [here]

Econometric analysis in your thesis

- Consult your ideas with the supervisor throughout the year.
- You can also consult with other IES lecturers.
  - We have different specializations.
  - Cross-sectional econometrics.
  - Time-series econometrics.
  - Check here: https://ies.fsv.cuni.cz/lide/interni-clenove
- What will you learn in Econometrics II
  - Basics of time-series econometrics.
  - Econometrics of panel data.
  - Dealing with violations of OLS assumptions.
  - Models for specific dependent variables (e.g. dummy dependent variable).

Thank you!

gebicka@fsv.cuni.cz