Econometrics II

Week 1

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Fall 2022

Welcome :-)

- This is **Econometrics II** course
- Second in the undergraduate Econometrics sequence
- Taught by **Barbara Pertold-Gębicka**
- Seminar leaders:
  - Kseniya Bortnikova
  - Jan Mošovský
  - Mathieu Petit
  - Salim Turdaliev
- The course webpage: Econometrics II in SIS
  - All course information
  - Materials for lectures & seminars

Outline of Today's talk

- General Remarks about the course
- Syllabus
- Goals of Econometrics II
- Repetition: OLS and the underlying assumptions

The course

- Econometrics I $\rightarrow$ Econometrics II
- We will use the book (available in the IES library):

Text book

Jeffrey M. Wooldridge: Introductory Econometrics. A Modern Approach. Boston: Cengage Learning, edition fourth or higher.

- For those keen to study Econometrics in more detail or from another angle, these are the recommended complimentary books
  - "Introductory Econometrics for Finance" by Chris Brooks
  - "Mostly Harmless Econometrics: An Empiricist's Companion" by Angrist and Pischke
  - "Microeconometrics: Methods and Applications" by Cameron and Triveldi

Course organization

- Lectures
  - Wednesdays, 2:00-3:20 p.m.
  - Students required to **prepare for the lecture**
    - Go through the associated reading
    - Revise the pre-required material
  - During lectures I will give examples and explain more difficult material
  - There will be a short **online quiz** active on Monday & Tuesday from time to time. You can gain bonus points in such quiz!

- Seminars
  - Wednesdays: 15:30-16:50 (room 016); Wednesdays: 18:30-19:50 (room 016); Thursdays: 14:00-15:20 (room 016)
  - Each seminar relates to a lecture
  - Students **are required** to attend a lecture or carefully read associated reading before participating in a seminar
  - You will solve sample problems illustrating the topic; problems are similar to exam questions

Requirements

Course requirements

- Assignments: 0 - 40% (10 points each)
- Midterm Exam: 0 - 20%
- Final Exam: 0 - 40% (at least 21 points to pass the Course)
- Up to 6 bonus points in quizzes.

Grading

- 91% - 100% Grade A
- 81% - 90% Grade B
- 71% - 80% Grade C
- 61% - 70% Grade D
- 51% - 60% Grade E
- 0% - 50% Grade F (Fail)

Course organization - further remarks

- Software
  - All empirical exercises solved using one software
  - R - please, install it on your computers!
  - In home assignments you are free to use any software
- Assignments
  - 4 home assignments
  - Posted in SIS two weeks before the deadline
  - Work in groups of 2 students
  - You can earn up to 10 points for each home assignment

Home assignments

- There will be four home assignments during the semester
- Dates
  - First assignment announced Oct 19, deadline Nov 2
  - Second assignment announced Nov 2, deadline Nov 16
  - Third assignment announced Nov 23, deadline Dec 7
  - Fourth assignment announced Dec 7, deadline Dec 21
- Contents
  - Each assignment related to one topic
  - One theoretical problem and one empirical problem
  - The empirical problem should be approached as a small project. Students are expected to write a short report.

Exams

- Midterm exam
  - One-shot exam, no additional terms
  - **November 23, 2022, 2:00-3:30pm.**
  - Written during lecture time (room 314)
- Final exam
  - must-take exam, three trials possible
  - During the exam period (dates announced soon)
  - **Must obtain more than 50% points to pass the Course!**

Goals of the Econometrics II course

- **Econometrics I** taught you the basic tools
  - Ordinary Least Squares (OLS) with cross-sectional data
  - Assumptions necessary for OLS estimates to be BLUE
  - Evaluation of regression analysis - R-squared, hypothesis testing
  - Detecting and dealing with heteroskedasticity
  - Model specification and data issues
- **Econometrics II** will extend your knowledge
  - Regression analysis with time series data
  - Regression analysis with panel data
  - What happens and what can we do when OLS assumptions are not satisfied?
  - Regression analysis with limited dependent variables (count variables, binary variables)
  - **Practical application of econometric analysis**

Goals of the Econometrics II course

- After taking **Econometrics II** you should be able to independently conduct an empirical analysis
  - Properly formulate the research question
  - Chose the best econometric approach
  - Find suitable data
  - Conduct regression analysis using a statistical software
  - Verify whether the obtained estimates can be trusted
  - Interpret your results and/or propose improvements
- While the statistics behind econometric methods is an exact science, the data and processes we analyze are not exact.
- One needs to understand the data generating processes in order to apply econometrics.
- We will put a lot of attention to this part of the **art of econometrics**.

How to achieve this goal

- Material covered by Econometrics II is **more difficult** than Econometrics I
- We will learn many new approaches
- This requires systematic work during the whole semester
  - Students go through the relevant reading **before** each lecture
    (information about reading for next week given at the end of each lecture)
  - Students revise previously learned skills when necessary
    (information about revisions given at the end of preceding lecture)
  - During the lecture I will quickly go through the basic/easy stuff
  - We will devote most of the time to practical examples and more difficult material

For the next lecture

- Revise basics of Econometrics I with special attention to
  - Gauss-Markov assumptions for OLS regression in cross-sectional data
  - Properties of OLS estimator for cross-sectional data
  - Logarithmic variables
  - Dummy variables
- Read chapter 10 of Wooldrige textbook

Let me help you with some revision

Given the following regression model:

$$y_i = \beta_0 + \beta_1 \cdot x_i + u_i$$

The OLS estimator of the slope parameter (β_1) is:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Note: All sums go from i=1 to i=n, i.e. sum over all observations.
Note: A simple (one-explanatory-variable) regression model is given as an example. For multiple regression model it's more convenient to work with matrix notation - why?

Error/disturbance vs. residual

The outcome data we observe is 'produced' by some processes, e.g. interaction between buyers and sellers, firms' profit maximization, individuals' rational decision-making

- We call them **data generating processes**
- Data generating process can be summarized in an equation
- There is usually some uncertainty in such process → **the disturbance(error)**

We collect a sample of data produced by a data generating process

- We use econometric techniques to estimate parameters of the data generating process
- We do not observe the disturbances(errors)
- We observe just a sample
- ⇒ our estimates are imprecise, they do not exactly coincide with the true data generating process
- The difference between the true outcome and the outcome we estimate is **the residual**

Unbiasedness

Theorem: Unbiasedness of OLS

Under Assumptions MLR.1 through MLR.4, the OLS estimators are unbiased estimators of the population parameters: $E(\hat{\beta}_j) = \beta_j$, $j = 0, 1, \ldots, k$.

- What does it mean in practice?
- What are the assumptions MLR.1 - MLR.4?

Classical Linear Model Assumptions

- Assumption MLR.1 (Linear in Parameters)
- Assumption MLR.2 (Random Sampling)
  - This assumption is very important! When it is satisfied, we have data that can be used to estimate model parameters, the data is representative of the studied population.
- Assumption MLR.3 (No Perfect Colinearity)
- Assumption MLR.4 (Zero Conditional Mean)
  - We can also talk about the **exogeneity** assumption

Classical Linear Model Assumptions

- Assumption MLR.1 (Linear in Parameters)
- Assumption MLR.2 (Random Sampling)
- Assumption MLR.3 (No Perfect Colinearity)
- Assumption MLR.4 (Zero Conditional Mean)
- Assumption MLR.5 (Homoskedasticity)

Variance of the OLS Estimator

Theorem: OLS Sampling Variances

Under Assumptions MLR.1 through MLR.5, the variance of $\hat{\beta}_j$, conditional on X (conditional on sample values of the independent variables) is:

$$Var(\hat{\beta}_j|X) = \sigma^2/[SST_j(1-R_j^2)], \qquad j = 1, \ldots, k,$$

where σ^2 is the variance of disturbance term u_i, SST_j is the total sum of squares of x_{ij} and R_j^2 is the R-squared.

Theorem: Unbiased Estimation of σ^2

Under MLR.1 through MLR.5, the estimator $\hat{\sigma}^2 = SSR/df$ is an unbiased estimator of σ^2, where $df = n - k - 1$.

Gauss-Markov Theorem

Gauss-Markov Theorem

Under MLR.1 through MLR.5, the OLS estimators are the best linear unbiased estimators (BLUE) conditional on X.

Classical Linear Model Assumptions

- Assumption MLR.1 (Linear in Parameters)
- Assumption MLR.2 (Random Sampling)
- Assumption MLR.3 (No Perfect Colinearity)
- Assumption MLR.4 (Zero Conditional Mean)
- Assumption MLR.5 (Homoskedasticity)
- Assumption MLR.6 (Normality)

Note: Never refer to these assumptions by numbers when no context is given! Rather describe them, refer to their names.

Inference Under the Classical Linear Model (CLM) Assumptions

Theorem: Normal Sampling Distribution

Under MLR.1 through MLR.6, the OLS estimators are normally distributed, conditional on X. Further, under the null hypothesis, each t statistic has a t distribution, and each F statistic has an F distribution.

What will we do in Econometrics II

- How can we adjust the OLS estimation method to time-series data?
- How can we adjust the OLS estimation method to panel data?
- What to do when some of the CLM assumptions are not satisfied?

Cross Sectional Data vs. Time Series Data

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i \qquad u_i \sim N(0, \sigma^2)$$

$$i = 1, 2, \ldots, n$$

vs.

$$y_t = \beta_0 + \beta_1 x_{t1} + u_t \qquad u_t \sim N(0, \sigma^2)$$

$$t = 1, 2, \ldots, T$$

The Nature of Time Series Data

- Time series vs. cross-sections: Temporal ordering.
- Until now, we have studied properties of OLS estimator based on the assumption that samples are random
- Time series data are not random samples (why?)

Instead, we deal with Stochastic Processes

- "stochastic" from the Greek "stochos": aim, guess, or characterized by conjuncture and randomness.
- The observed data is one realization of a stochastic process.

- How does it challenge the CLM model?

This is all for today!

- Tutorial sessions this week
  - practical revision of Econometrics I
  - how to produce a report based on econometric analysis

Thank you for your attention!

... and do not forget to read Chapter 10 for the next week!