Doctoral Course
Advanced Regression Modeling

University of Münster, June 15-18, 2020
ONLINE COURSE

Course description

This course provides an advanced treatment of econometric techniques necessary to conduct empirical research in business and economics. The focus of the course will be on microdata as collected, for example, in large individual or household surveys. Students will learn how to carry out an empirical study, going beyond basic inferential statistics and simple linear regression. The course will cover linear and nonlinear regression models for cross-sectional and panel data, and methods for causal inference in observational studies. Examples from the literature and computer tutorials offer hands-on experiences in utilizing the methods.

Lecturer

Prof. Dr. Stefan Boes contact by email: stefan.boes@unilu.ch

Learning objectives

The course has two main objectives:

(i) learn the methodology of modern econometric analysis
(ii) acquire the skills to plan and execute your own empirical project

The course focuses on applied econometric tools, i.e., the management and use of real data and the application of advanced statistical software (Stata) to implement the discussed methods will be an integral part of the learning experience. Please make sure that you have Stata installed on your laptop as we will go through various data examples in our practice sessions.

Background reading

The main references for this course are:

Angrist JD, Pischke JS (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. [AP1]

Angrist JD, Pischke JS (2014) *Mastering Metrics: The Path from Cause to Effect*, Princeton University Press. [AP2]

Stock JH, Watson MW (2019) *Introduction to Econometrics*, 4e, Pearson. [SW]

Winkelmann R, Boes S (2009) *Analysis of Microdata,* 2e, Springer. [WB]

Prerequisites

Students are assumed to be familiar with basic statistics, including probability theory, inferential statistics and simple linear regression on the undergraduate level. I recommend chapters 1-4 of SW as a refresher and preparatory reading.

Lectures and topics

Lecture 1 - Linear regression

- Regression basics
- Specification, estimation, inference
- Description versus causality
- Cross-sectional and panel data

Lecture 2 - Causal inference I

- Potential outcomes framework
- Randomized experiments
- Regression analysis of experiments
- Instrumental variables estimation

Lecture 3 - Causal inference II

- Difference-in-differences (DID)
- Two-period and multiple-period cases
- Regression discontinuity design (RDD)
- Sharp and fuzzy RDDs
- Sensitivity and robustness checks

Lecture 4 - Nonlinear regression

- From linear to nonlinear regression
- Binary probit and logit models
- Ordered logit and probit
- Multinomial logit and probit
- Count data models

All lectures will be provided as **video lectures** and are ideally watched **before** the topic is covered during the course (see also schedule below). The video lectures and accompanying lecture slides as well as supplementary material that will be covered during the interactive online sessions will be distributed via Learnweb latest until June 12, 2020.

Time schedule of zoom sessions

Monday, June 15, 2020 **09.30 – 11.30 Q&A and practice** • Q&A - Lecture 1 • General introduction to Stata • Data handling and description • Regression examples **13.30 – 15.00 Optional group meetings**	***Tuesday, June 16, 2020*** **09.30 – 11.30 Q&A and practice** • Q&A - Lecture 2 • Basic examples of matching • Examples of social experiments • Dealing with data issues I **13.30 – 15.00 Optional group meetings**
Wednesday, June 17, 2020 **09.30 – 11.30 Q&A and practice** • Q&A - Lecture 3 • DID applications • RDD applications • Dealing with data issues II **13.30 – 15.00 Optional group Meetings**	***Thursday, June 18, 2020*** **09.30 – 11.30 Q&A and practice** • Q&A - Lecture 4 • Examples of nonlinear regressions • Model specification, estimation, and interpretation **13.30 – 15.00 Optional group meetings**

Q&A and practice sessions from 9:30 to 11:30 are designed as interactive online sessions, organized via zoom (login via the link below), in which you will have the opportunity to ask questions on the lecture material, and we will go through some practice material in the form of daily work plans and empirical examples using data from the Swiss Household Panel.

The **group meetings** in the afternoon are **voluntary**. During these group meetings, we can discuss more in-depth questions you may have related to your research, which may be interesting for those of you working on empirical projects. For these discussions, a separate sign-up sheet will be circulated in which you can indicate your interest in such a meeting. Small groups will be formed to coordinate the discussions in groups with similar interests.

Link to zoom sessions

https://unilu.zoom.us/j/92606105616?pwd=bVRtYXJWZ1BEVjY2bTlvSWt5V2pWdz09
Meeting ID: 926 0610 5616
Password: 303670

Assessment

Grading for this course is based on an **empirical homework assignment**, which consists of carrying out your own analysis based on data from the Swiss Household Panel and writing a short research paper that motivates your work and presents the results. The data will be provided in class. Please follow the analysis steps as discussed in the computer labs. Document your work in a single do-file containing all commands used for the analysis. The contents of the do-file must be attached to the paper, see below. Please remember the principle of replicability, i.e., all steps/commands of your analysis must be documented, and the results shown in the paper must be reproducible when starting from the dataset that is provided.

The **empirical paper** must be structured as follows:

1) *Introduction:* Motivate your research and provide relevant references (max. 10 published articles) to embed your work in the literature.
2) *Description of the data:* Briefly describe the data source (see www.swisspanel.ch for more information on the SHP) and variables and provide some descriptive statistics.
3) *Empirical methods:* Briefly describe the regression equation of interest that you are estimating, i.e., what is the dependent variable, what is the main variable of interest, what are the control variables, and what is the estimation method.
4) *Results:* Present and describe your results, paying attention to the correct interpretation of estimated model coefficients. Compare the results of different models, e.g., pooled OLS, RE, FE, with and without controls.
5) *Discussion and conclusion:* Place your results in the literature. Do you find similar or other results than in the cited literature? Suggest a research design that would allow you to draw causal conclusions for the question of interest in your study. This should be a realistic proposal, i.e., suggest a design that could potentially be implemented.

The paper must be restricted to **2000 words** at maximum, including a title page with your name, affiliation and a short abstract of the paper, but excluding the references, tables, figures (if needed) and the appendix with the contents of the do-file (commands used for the analysis). Please use standard page layout, double line spacing, and a 12pt standard font.

The paper in **pdf format** is due on **August 30, 2020, 6pm** by email to stefan.boes@unilu.ch. Late submissions are strictly NOT accepted and will be graded with failed. Evaluation is based on the quality and logical consistency of the argumentation (what is the question of interest and how is it addressed in the analysis), the correct interpretation of results, and the suitability of the proposed research design for causal inference.

Possible **topics** for the homework are the relationships between

- *log earnings (y)* and *nationality (main x)*
- *log earnings (y)* and *gender (main x)*
- *life satisfaction (y)* and *stressful work conditions (main x)*
- *job security (y)* and *number of employees (main x)*
- *health satisfaction (y)* and *sleep disorders (main x)*
- *health satisfaction (y)* and *body mass index (main x)*
- *satisfaction with free time (y)* and *number of kids (main x)*
- *life satisfaction (y)* and *meeting friends (main x)*
- *risk of unemployment (y)* and *education (main x)*
- *risk of failure at school (y)* and *number of siblings (main x)*

This list is not exhaustive and own suggestions are highly welcome.