

Syllabus

Panel Data: Analysis and Applications for the Social Sciences

(3 ECTS)

Daria Salnikova, lecturer (e-mail: dsalnikova@hse.ru, web-page: <https://www.hse.ru/staff/salnikova>)

Department of higher mathematics, NRU Higher School of Economics

Master's programme: Applied Methods of Social Analysis of Markets, Faculty of Social Sciences

Meeting Minute # ___ dated _____ 2019

1. Course Description

a) Pre-requisites

This course is suitable for those with a background in basic statistical inference. Experience in applying linear regression models is a plus. If this is not the case, students need not worry as the first section of this course gives a gentle introduction to regression analysis.

b) Abstract

“Panel data: Analysis and Applications for the Social Sciences” is a blended-learning course. The online course “Getting and Cleaning Data” (<https://www.coursera.org/learn/data-cleaning>) covers the basics of data manipulation in R. The first part of the course gives an overview of multiple regression models. The second part of the course focuses on the methodological tools necessary to succeed in handling panel data, namely, regression models with interaction terms and exploratory longitudinal data analysis. The third part covers fixed-effects and random-effects models. Lectures provide students with the theoretical foundations of panel data analysis. Practical sessions develop data analysis and data visualization skills. Students use RStudio for statistical analysis. At the practical sessions, students discuss the key approaches to handling panel data and illustrate them with different examples from social science research, in particular, economic sociology. Students are given datasets from original studies to replicate the findings and change the model specifications if needed.

2. Learning Objectives

The course aims to provide students with the theoretical background and practical skills in conducting panel data analysis. Specifically, the learning objectives are as follows:

- to enable students to choose appropriate models for panel data analysis
- to develop data manipulation and visualization skills
- to enable students to implement linear panel models in RStudio

3. Learning Outcomes

By the end of the course students are expected to apply fixed- and random- effects models to analyze panel data, to interpret the results, to have data visualization skills and skills in implementing the afore-mentioned methods by using RStudio in the context of panel data analysis. Students will learn the advantages and limitations of different approaches to panel data analysis. This knowledge will help students choose a set of appropriate statistical tools to test their research hypotheses.

4. Course Plan

Distribution of hours

Theme	Lectures	Seminars	Self-study	Total
Introduction. Linear regression analysis	4	4	10	18
Data manipulation. Supplementary tools for panel data analysis	0	4	16	20
Interaction terms in regression analysis	4	4	20	28
Fixed-effects models	6	6	22	34
Random-effects models VS Fixed-effects models	2	2	14	18
	Contact hours: 36		82	118

Theme 1. Introduction. Linear regression analysis

Types of data structures. Multiple linear regression models with their applications to cross-sectional data. Assumptions. Model specification. Interpretation of regression analysis results. Model diagnostics.

Theme 2. Data manipulation. Supplementary tools for panel data analysis

Panel VS Time-series cross-section (TSCS) VS Time-series data. Exploratory data analysis and visualization of panel data. Within- and between-group variation. Reshaping data. Merging data. Students are required to listen to the following lectures online (Week 3, Week 4, “Getting and Cleaning Data”. Available at: <https://www.coursera.org/learn/data-cleaning>) before the given practical session.

Theme 3. Interaction terms in regression analysis

Moderation VS Mediation. Conditional hypotheses with examples from social science research. Multiple linear regression models with interaction terms. Model specification. Interpretation of interaction effects. Interaction between binary predictors. Interaction between binary and continuous predictors. Marginal effects. Visualization of interaction effects.

Theme 4. Fixed-effects models

Fixed-effects model VS pooled model. Least-squares dummy-variable models. Within-group transformation. The technique underlying the estimation of coefficients in fixed-effects models. Aggregation bias. Model diagnostics.

Theme 5. Random-effects models VS Fixed-effects models

Random-effects models: assumptions, model estimation, generalized least-squares method and feasible generalized least-squares method. Hausman test and its limitations.

5. Reading List

Required reading list

Basic econometrics / D. N. Gujarati, D. C. Porter. – 5th. – Boston [etc.]: McGraw-Hill, 2009. – 921 p. – (Economics series) .

Halaby, C. (2004). Panel Models in Sociological Research: Theory into Practice. *Annual Review of Sociology*, 30(1), 507 – 544.

Introductory econometrics: a modern approach / J. M. Wooldridge. – 4th ed. – Mason: South-Western Cengage Learning, 2009. – 865 p.

Анализ панельных данных и данных о длительности состояний: учеб. пособие / Т. А. Ратникова, К. К. Фурманов. – М.: Изд. дом Высшей школы экономики, 2014. – 373 с.

Optional reading list

Econometric analysis of cross section and panel data / J. M. Wooldridge. – Cambridge: The MIT Press, 2002. – 752 p.

Эконометрика: начальный курс: учебник для вузов / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий . – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 575 с.

6. Grading System

Students' performance is assessed on the basis of

- 1) Seminar activity – 3 quizzes (The value of each quiz is 10% of the final grade)
- 2) 3 home assignments (The value of each home assignment is 10% of the final grade)
- 3) Quantitative research essay (The essay contributes 40% to the final grade)

All quizzes are closed book. Students arriving late will not receive extra time to write a quiz. All the quizzes, home assignments and the essay are graded on a 10-point scale. If a student submits his or her work (home assignment or essay) after the deadline, the mark will be reduced by 1 score for each day of delay.

Students have the possibility to gain additional scores for their active engagement in seminars, in particular, computer labs (problem solving) and bonus home assignments. The final grade is rounded arithmetically.

7. Examination Type

There is no exam, students are assessed on their seminar performance, home assignments and the quantitative research essay.

Quizzes include the following types of assessment: interpretation of model estimates and graphs, problem solving, open theoretical questions, a research critique (determining if methods applied in a study are appropriate). Examples of quiz problems are available at: https://yadi.sk/i/_wYUvPU8MWQyzQ

The research essay is intended to demonstrate students' skills in data analysis and their ability to apply appropriate statistical methods to a given research task.

The parts of the quantitative research essay are as follows:

1. Abstract
2. Research puzzle
3. A concise literature review
4. Hypotheses
5. Data
6. Methods with a detailed explanation of their choice, model specifications
7. Results and their interpretation
8. Robustness checks
9. Discussion and conclusion
10. Literature list

8. Methods of Instruction

The course is taught in the form of lectures (theoretical background), seminars (problem solving, discussion of studies conducted on panel data analysis, interpretation of model estimates), in particular, computer labs (data visualization, data analysis using Rstudio, replication of study results). Additionally, the online course “Getting and Cleaning Data” (<https://www.coursera.org/learn/data-cleaning>) is used.

9. Special Equipment and Software Support (if required)

The course requires R (Rstudio) as software support. Seminars take place in a computer class with a projector.