

**The Government of the Russian Federation the Federal State Autonomous Educational
Institution of Higher Education «National Research University «Higher School of
Economics»**

Sociological department

Course syllabus

Panel data Analysis and Applications for the Social Sciences

The author of the course: Kseniya Abanokova, PhD in Economics
Email: kabanokova@hse.ru

Moscow, 2018

This program cannot be used by other departments of the University and other universities or colleges without a permission from Economic Sociology department

Overview:

The purpose of this course is to present the background necessary to understand the applications of panel data and to provide students with the practical skills which could be applied to analyse a variety of research questions in social sciences. The wide range of topics will be covered: continuous and discrete panel models, survival analysis. The seminars are designed to give the students some hands-on experience with data manipulation through a statistical software package Stata. The course is designed with a focus on students as the persons doing research and involves reading, writing and in class practice. Previous experience in basic statistics and econometrics is expected.

Aims of the course:

This course will give students a practical grounding in the panel data analysis. It has the following key aims

1. Introduce advantages and limitations of panel data and consider what sort of questions may be answered by using this data.
2. Review some panel data models commonly used in social sciences and demonstrate how longitudinal analysis contributes to the social processes.
3. Show how to handle panel data and provide students with the skills to manipulate panel datasets and choose the right technique for research question.

Learning Goals:

Having completed the essential practical assignments students are expected to leave the course with:

1. An understanding of basic concepts of panel data analysis
2. Developing practical skills in selecting and conducting different types of panel data analysis
3. Confidence to manipulate panel datasets with different panel methods, choose the rights data technique and interpret the results

Course Structure:

The course will cover one major topic each week (see course content below). The time spent on each topic may be adjusted. The course consists of lectures and practical seminars with Stata implementations. Each lecture is followed by a practical session where students will use Stata to implement the methods covered in the lectures. Main data will be from Russian Longitudinal Monitoring Survey - HSE (RLMS-HSE). Also, there will be Problems sets to be discussed in class. The practical seminars and lectures will have been illustrated the value of panel surveys, using findings from the RLMS-HSE as the primary example.

Course Content:

Week 1. The main concepts and definitions of the panel analysis. We will discuss many types of longitudinal data first and then we will discuss why panel data are so useful (benefits and drawbacks of using panel data), what kind of data are required for panel analysis and how to get to know panel data. Applications of panel data to wide areas of empirical studies will be reviewed. Basic data manipulation and descriptive techniques, simple summary statistics (within and between variation, transition tables) will be covered this week. We will also look at some empirical applications of using panel data to mobility analysis and labor market studies.

Week 2. Introduction to linear panel data models. This week will begin by giving a brief overview of the basic panel data models and discussing various reasons for endogeneity and ways of solving the endogeneity with panel data. We will discuss how to eliminate unobserved heterogeneity with standard fixed- and random effects models and when to use these models. Some empirical applications of these models will be discussed.

Week 3. Extensions to panel data models: dynamic models. We will keep exploring different forms of endogeneity with fixed and random effect models. Basic fixed and random effects regressions with continuous dependent variables will be estimated and main assumptions will be tested. We will discuss unobserved heterogeneity and state dependence issues in panel data and will know when it is necessary to use dynamic models. We will look at the applications of these models to empirical studies focusing on the correct specification and interpretation.

Week 4. Introduction to nonlinear panel data models. We will discuss discrete outcomes in panel data, types of discrete variables and will know how nonlinear models are estimated. We will compare linear probability model and probit/logit models (random effect probit/logit model,

conditional fixed effect logit model) and will estimate predicted probabilities and marginal effects. Applications of these models to health and well-being studies will be covered.

Week 5. Extensions to nonlinear panel data models: selection models. We will also discuss non-response, sample selection and attrition in panel data and their consequences. We will explore qualitative response models and limited dependent variables and will look at how tobit and Heckman selection models can be estimated with panel data. Applications of these methods to labor supply model will be made.

Week 6. Policy evaluation. We will discuss how to evaluate the impact of various policies with panel data using difference-in-difference and matching estimators

Week 7. Introduction to survival analysis.

Course Plan:

Week №	Topics	Total number of hours	Credit hours		Selfstudy hours
			Lectures	Seminars	
1	The main concepts and definitions of the panel analysis	10	2	2	6
2	Introduction to linear panel data models	10	2	2	6
3	Extensions to panel data models: dynamic models	10	2	2	6
4	Introduction to nonlinear panel data models	10	2	2	6
5	Extensions to nonlinear panel data models: selection models	10	2	2	6
6	Policy evaluation	10	2	2	6
7	Introduction to survival analysis	10	2	2	6
	Project			4	28
	Overall	102	14	18	70

Reading list:

- Wooldridge, J. M. (2001), *Econometric Analysis of Cross Section and Panel Data*, MIT Press
- Cameron A. C and Trivendi P. K. (2010), *Microeconometrics Using Stata*. STATA Press
- Halaby, C. N. (2004). Panel data models in sociological research: Theory into practice, *American Review of Sociology* 30, 507-544

- Lecture notes, problem sets as well as relevant articles (all in pdf format) will be available from LMS website. Each presentation contains suggested reading on the topic. Other material will be distributed in class

Course Grading:

1. Tests (20%) usually include a theoretical section (see below)
2. Research project (40%) - see below
3. Final exam (40%) usually includes an applied section (see below)

A course grade is a weighted sum of rounded grades for the three components: tests, project, exam.

Total grade= Tests (20%) +Project (40%) + Exam (40%)

A passing grade for course is 3.5. Grades under 3.5 are failing.

Rounding rules: each component of the total grade (tests, projects, exam) is rounded to tenths and the final grade is further rounded to ones.

1. Tests

Weekly tests consist of around 8-12 multiple-choice questions on lecture content with questions on data analysis and econometrics. Tests are open-book, any materials can be used. Tests can be completed only during the 10-15 min.

2. Research project

As a part of this course students are expected to prepare the presentation of research project (in PowerPoint slides and in Word texts in Eng language) using panel data and estimation methods that have been studying during the course. This assessment counts as 40% of the final mark for the course on Panel Data.

Project outcome is a presentation+text (60%), discussant's activity (20%) and peer reviewing (20%). Deadline for the text is 1 day before scheduled presentation day. Text size ~10 pages with abstract and list of references (Times New Roman, 12)

For this assignment, students are limited to 4 persons in a group and up to 15 min for each presentation. Each group will have up to 4 discussants (will be chosen by lector) who will be responsible for asking the questions (1 question per one presentation). Students will be responsible to review the quality of each presentation according to peer assessment form.

Grades are published after all presentations have been made. The topic is the choice of group and need not (but may be) approved by the lector. Each group should prepare a presentation of research proposal, which includes the following elements:

1. Introduction

1.1 Underline the reasons why the proposed area of research is important, and what previous studies have found. You are not required to present a comprehensive literature review, but try to concentrate on several studies that have appeared as peer-reviewed publications from 2000 onwards. Use on-line services such as Google Scholar, EconLit to investigate past work on this topic.

1.2 Describe the details of the research questions which you propose to address, together with associated hypotheses. Work through the theoretical issue.

2. Data availability and measurement issues

2.1 Explain how panel data can inform the research question in ways which cross-sectional data could not. Describe the “perfect” data you will need. Investigate what data sets are available now, its limitations and what type of data sets have been used in the past literature.

2.2 Describe the variables appropriate for analysis (consider reasonable interactions, quadratics, logs of the variables)

2.3 Describe the empirical model you will use for the analysis

3. Methodology

3.1 Describe the estimation methods which you propose to use for the analysis (which would include the methods taught in this course, but not be restricted to it), and the particular aspects of the research question which will be addressed by each of these methods.

3.2 Describe the potential strengths and limitations of each of the methods you choose.

Examples of the topics:

The determinants of divorce and separation

The effects of shift-working on mental health

The effect of education on health

The dynamic of poverty

The effect of health on labor market outcomes

3. Exam

Examination lasts 1.5 hours and consists of seminars content in STATA calculations. Exam is part open-book, only course materials are allowed at the exam. Exam attendance at the specified time is required. Students who arrive late will have less time to finish, finishing at the same time as other students.

4. Special Equipment and Software Support (if required).

Lecture: computer connected to Internet and projector

Seminar: computer connected to Internet and installed STATA package (V12 and above)
and projector