Course syllabus
«Panel Data Analysis and Applications for the Social Sciences»

Approved by
Academic Council
Protocol Nr. 03 from 05.06.2018

Author	Daria Salnikova, lecturer of the all-university department of higher Mathematics
Number of credits	3 credits
Contact work (hrs.)	32 hours
Self-study work (hrs.)	82 hours
Course	Master Course, 1 year
Course Format	Full time Course

1. Course Description

The course aims to provide students with the theoretical background and practical skills in conducting panel data analysis. The first 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 second part covers fixed-effects and random-effects models. The third part is devoted to mixed-effects models. Lectures provide students with the theoretical foundations of panel data analysis. Practical sessions develop data analysis and data visualization skills. Students use Stata software 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. Course Prerequisites

This course is suitable for those with a background in basic statistical inference and experience in applying linear regression models.

3. Learning Outcomes

By the end of the course students are expected to apply fixed-, random- and mixed-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 Stata 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

Theme 1. Introduction. Supplementary tools for panel data analysis

Types of data structures. Panel VS Time-series cross-section (TSCS) VS Time-series data. Exploratory data analysis and visualization of panel data. Within- and between-group variation.

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 2. Fixed-effects and random-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. Do we need time-invariant predictors in the context of panel data analysis? Post-treatment bias. Dealing with endogeneity. Model diagnostics.

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

Theme 3. Mixed-effects models

Repeated-measures ANOVA and its limitations. Mixed-effects models applied to panel data analysis. Comparison with fixed-effects and random-effects models. Assumptions of mixed-effects models. The intraclass correlation coefficient and its interpretation. Model specification: bottom-up strategy. A null model without any predictors as a preliminary step in data analysis. Linear change over time as a fixed effect and as a random effect. Non-linear effects of time. The polynomial function. Time-varying and time-invariant predictors. Cross-level interactions and their meaning. Interpretation of coefficient estimates in mixed-effects models.
Model estimation. Model testing. Comparing nested models. Likelihood-ratio tests. Non-nested models. The AIC (Akaike) and BIC (Schwarz) information criteria. Model diagnostics.
Data visualization tools for mixed-effects modeling.

5. Reading List

Compulsory reading

Brambor T., Clark W.R., Golder M. (2006). Understanding Interaction Models: Improving Empirical Analyses. Political Analysis, 14, 63 – 82.

Kam, C., Franzese, R. (2007). Modeling and Interpreting Interactive Hypotheses in Regression Analysis, UMich Press.

Rose, D., ed. 2004. *Researching social and economic change: The uses of household panel studies*. New York: Routledge.

Singer, J. D., Willett, J. B. (2003). *Applied Longitudinal Data Analysis. Modeling Change and Event Occurrence.* New York: Oxford University Press.

Stock J. H., Watson M. W. (2008). *Introduction to econometrics.* Pearson Addison Wesley.

Optional reading

Hainmueller, J., Mummolo, J., Xu, Y. (2018). How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice. Political Analysis, 1-30.

Hedeker, D., Gibbons R. D. (2006). *Longitudinal Data Analysis.* Hoboken, New Jersey: Wiley.

Hox, J. J. (2010). *Multilevel analysis: techniques and applications.* 2ndedition. New York and Hove: Routledge.

Jaccard J., Turrisi R. (2003). *Interaction effects in multiple regression.* 2nd edition. Thousand Oaks, London, Dehli: Sage publications.

Luke D. (2004). *Multilevel modeling.* Thousand Oaks, London, New Dehli: Sage Publications.

Ratnikova, T. A. Analiz panel'nyh dannyh v pakete "STATA": metodicheskie ukazaniya k komp'yuternomu praktikumu po kursu "Ekonometricheskij analiz panel'nyh dannyh" / T. A. Ratnikova. –M.: GU-VSHE, 2004.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. Cambridge, MA.

6. Grading System

Students' performance is assessed on the basis of

1) Seminar activity (30% of the final grade)
2) Quantitative research essay (30% of the final grade)
3) Exam (40 % of the final grade)

The seminar activity grade is calculated by averaging the score for class quizzes and the score for home assignments. Students have the possibility to gain additional scores for their active engagement in seminars and computer labs (problem solving). Three quizzes are expected in the course: one quiz at the end of each section. Scores for separate types of activity (seminar performance, essay and exam) are rounded. The final grade is rounded as well.

All quizzes and the final exam are closed book.

The final exam includes three 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).

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

Computer Applications

№	Application	**Access conditions**
1.	Stata	From the University's internal network

7. 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 models) and computer labs (data visualization, data analysis using Stata, replication of study results).