Санкт-Петербургский филиал федерального государственного автономного образовательного учреждения высшего профессионального образования «Национальный исследовательский университет «Высшая школа экономики»

Факультет Санкт-Петербургская школа экономики и менеджмента
Департамент экономики

Рабочая программа дисциплины
Панельные данные

(читается на английском языке)

для образовательной программы Экономика
направления 38.03.01 «Экономика»
уровень бакалавриат

Разработчик программы:
Муравьев А.А., PhD, доцент, amuravyev@hse.ru

Согласована начальником ОСУП
«_____»_____2018 г.
Л.А.Кежун _____

Утверждена академическим советом ОП «Экономика»
«31» августа 2018 г., № протокола 1/2018-2019

Академический руководитель ОП
С.Г.Коковин _____

Санкт-Петербург, 2018

Course Syllabus

Title of the course	**Panel data (applied econometrics of panel data)**		
Title of the Academic Programme	Bachelor's program "Economics"		
Type of the course	Obligatory		
Prerequisites	Students' knowledge of the foundations of statistics and econometrics is a key prerequisite for the successful completion of the course.		
ECTS workload	3		
Total indicative study hours	Directed Study	Self-directed study	Total
	36	78	114
Course Overview	The course is designed for third-year undergraduate students in Economics. Its main goal is to familiarize the students with contemporary methods of panel data analysis, starting with the pooled OLS model and ending with dynamic panel data models. The course is of applied nature: The material is presented, whenever possible, in a non-technical way, with examples of empirical studies published in leading international economics and finance journals discussed in class. Lectures are supplemented by computer labs, which ensures that the students get hands-on experience of analyzing real world panel data using Stata 14/15. The topics covered include: pooled OLS model, fixed- and random-effects models, dealing with unbalanced panels, measurement error in panel data, endogenous explanatory variables, the Hausman-Taylor model, and an overview of dynamic panel data models (the Nickel bias, Anderson-Hsiao estimator and Arellano-Bond estimator).		
Intended Learning Outcomes (ILO)	Students should have a firm grasp of the key methods of panel data analysis Students should be able to effectively apply these methods in own empirical research. Students should be familiar with and be able to use key capabilities of the statistical package "Stata", including its programming options (the so-called do-files).		
Teaching and Learning Methods	The course consists of lectures (20 hours) and computer labs (16 hours). The main teaching and learning methods include lectures, reading, exercises in the computer lab and an empirical project.		

Content and Structure of the Course

№	Topic / Course Chapter	Total	Directed Study		Self-directed Study
			Lectures	Tutorials	
1	**Panel data and pooled OLS model**	18	2	4	12
2	**The fixed effects model**	18	4	2	12

3	**The random-effects model**	18	4	2	12
4	**Further topics in the analysis of linear panel data models**	32	6	4	22
5	**Dynamic panel data**	28	4	4	20
Total study hours		114	20	16	78

Indicative Assessment Methods and Strategy	Students' progress will be evaluated using a **homework assignment**, a **midterm exam (test)** and a **final exam**. The **home assignment** includes problems and/or computer exercises in Stata. The assignment will be distributed in class and will be due in approximately two weeks. The homework (only paper versions, files sent by email will not be accepted!) is to be handed in before class on the day it is due. No late homework is accepted. The **midterm exam** is a closed book, closed notes test scheduled in the middle of the course. At the end of the course there is a **final exam**, which is a closed book, closed notes test to be held in the classroom. The duration of the final exam is two academic hours. The overall grade for the course is computed as follows: Home assignment 25% of the final grade Midterm test 25% of the final grade Final exam 50% of the final grade In order to pass the course, students should get a passing mark (at least 4 out of 10) at the final exam. Failure at the final exam implies failure at the whole course regardless of the performance during the course.
Readings / Indicative Learning Resources	Mandatory • Sevestre, Patrick, and Laszlo Matyas (2008). The econometrics of panel data. Springer-Verlag Berlin. (**available online from the campus via Springer**). • Stata Manual (2015). Stata Corporation. (**available online at https://www.stata-press.com/manuals/documentation-set/**) Optional • Matyas, Laszlo (2017). The econometrics of multi-dimensional panels. Springer-Verlag, Berlin. (**available online from the campus via Springer**). • Roodman, D. (2009). How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata. Stata Journal 9(1): 86-136 (**available online from the campus via Scopus**). • Wooldridge, J.M. Introductory econometrics: a modern approach / J.M.Wooldridge. – 6th ed . – Boston : Cengage Learning, 2016 . – 789 p. – На англ. яз. - ISBN 978-1-3052-7010-7

Indicative Self- Study Strategies	**Type**	**+/–**	**Hours**
	Reading for seminars / tutorials (lecture materials, mandatory and optional resources)	+	30
	Assignments for seminars / tutorials / labs	+	10
	E-learning / distance learning (MOOC / LMS)	-	0

	Fieldwork	-	0
	Project work	+	20
	Other (please specify)	-	0
	Preparation for the exam	+	18
Academic Support for the Course	Academic support for the course is provided via LMS, where students can find: guidelines and recommendations for doing the course; guidelines and recommendations for self-study; samples of assessment materials.		
Facilities, Equipment and Software	Stata 14/15		
Course Instructor	Alexander A. Muravyev, PhD, Associate Professor of the Department of Economics; Evgeniya Yu. Polyakova, Senior Lecturer of the Department of Economics.		

Annex 1

Content of classes

Lecture 1. Panel data and pooled OLS model (2 hours). Panel data and their advantages. Pooled OLS model. Cluster-robust estimator of variance. Main reading: Lecture notes. Sevestre and Matyas (2008).
Lecture 2. The fixed effects model (4 hours). General issues in modeling heterogeneity using panel data. The fixed effects model assumptions. Estimation: the within, LSDV and first differences estimators. Two-way fixed effects models. Measures of goodness of fit in panel data. Predicted values of fixed effects. Main reading: Lecture notes. Sevestre and Matyas (2008).
Computer lab 1. Introduction to econometric package Stata (4 hours). Basic capabilities of Stata. Basic commands and the grammar of Stata. Data management tools. Output tables. User written commands. Do and log files. Pooled OLS. Cluster-robust estimator of variance in Stata Main reading: Lecture notes. Stata manual (2015).
Lecture 3. The random-effects model (4 hours). Assumptions. Estimation. Testing for random effects (the Breusch and Pagan test for random effects). Testing the consistency of the RE model (the Hausman test). Main reading: Lecture notes. Sevestre and Matyas (2008).
Computer lab 2. Essentials of panel data analysis in Stata (4 hours). The fixed effects model. LSDV and the first differences model. The random-effects models. The Breusch and Pagan test. The Hausman test. Main reading: Lecture notes. Stata manual (2015).
Lecture 4. Further topics in the analysis of linear panel data models (6 hours). Measurement error in panel data. Random trend model. Correlated random effects. Endogenous explanatory variables. The Hausman-Taylor model. Unbalanced panels. Main reading: Lecture notes.

Sevestre and Matyas (2008).
Computer lab 3. Advanced issues in the analysis of panel data in Stata (4 hours). Correlated random effects. Random trend model. Endogenous explanatory variables. The Hausman-Taylor model. Unbalanced panels. Reading: Main reading: Lecture notes. Stata manual (2015).
Lecture 5. Dynamic panel data (4 hours). Dynamic panel data (DPD) setup. Correlation between the lagged dependent variable and error term. Nickell bias. Anderson and Hsiao estimator. GMM estimation notes. Arellano-Bond estimator. System DPD estimator. Main reading: Lecture notes. Sevestre and Matyas (2008). Roodman (2009).
Computer lab 4. Dynamic panel data (DPD) in Stata (4 hours). Anderson and Hsiao estimator. Arellano-Bond estimator. System DPD estimator. Main reading: Lecture notes. Stata manual (2015). Roodman (2009).

Annex 2

Assessment Methods and Criteria

Assessment Methods

Types of Assessment	Forms of Assessment	Modules			
		1	2	3	4
Formative Assessment	Test				*
	Essay				
	Report/Presentation				
	Project				
	In-class Participation				
	Other – problem sets				
Interim Assessment (if required)	Assignment (e.g. written assignment)				*
Summative Assessment	Exam				*

Assessment Criteria

In-class Participation

Grades	Assessment Criteria
«Excellent» (8-10)	A critical analysis which demonstrates original thinking and shows strong evidence of preparatory research and broad background knowledge.
«Good» (6-7)	Shows strong evidence of preparatory research and broad background knowledge. Excellent oral expression.
«Satisfactory» (4-5)	Satisfactory overall, showing a fair knowledge of the topic, a reasonable standard of expression. Some hesitation in answering follow-up questions and/or gives incomplete or partly irrelevant answers.
«Fail» (0-3)	Limited evidence of relevant knowledge and an attempt to address the topic. Unable to offer relevant information or opinion in answer to follow-up questions.

Project Work

Grades	Assessment Criteria
«Excellent» (8-10)	A well-structured, analytical presentation of project work. Shows strong evidence and broad background knowledge. In a group presentation all members contribute equally and each contribution builds on the previous one clearly; Answers to follow-up questions reveal a good range and depth of knowledge beyond that covered in the presentation and show confidence in discussion.
«Good» (6-7)	Clearly organized analysis, showing evidence of a good overall knowledge of the topic. The presenter of the project work highlights key points and responds to follow up questions appropriately. In group presentations there is evidence that the group has met to discuss the topic and is presenting the results of that discussion, in an order previously agreed.
«Satisfactory» (4-5)	Takes a very basic approach to the topic, using broadly appropriate material but lacking focus. The presentation of project work is largely unstructured, and some points are irrelevant to the topic. Knowledge of the topic is limited and there may be evidence of basic misunderstanding. In a group presentation, most of the work is done by one or two students and the individual contributions do not add up.
«Fail» (0-3)	Fails to demonstrate any appropriate knowledge.

Written Assignments (Essay, Test/Quiz, Written Exam, etc.)

Grades	Assessment Criteria
«Excellent» (8-10)	Has a clear argument, which addresses the topic and responds effectively to all aspects of the task. Fully satisfies all the requirements of the task; rare minor errors occur;
«Good» (6-7)	Responds to most aspects of the topic with a clear, explicit argument. Covers the requirements of the task; may produce occasional errors.
«Satisfactory» (4-5)	Generally addresses the task; the format may be inappropriate in places; display little evidence of (depending on the assignment): independent thought and critical judgement include a partial superficial coverage of the key issues, lack critical analysis, may make frequent errors.
«Fail» (0-3)	Fails to demonstrate any appropriate knowledge.

Recommendations for students about organization of self-study

Self-study is organized in order to:

- Systemize theoretical knowledge received at lectures;
- Extending theoretical knowledge;
- Learn how to use legal, regulatory, referential information and professional literature;
- Development of cognitive and soft skills: creativity and self-sufficiency;
- Enhancing critical thinking and personal development skills;
- Development of research skills;
- Obtaining skills of efficient independent professional activities.

Self-study, which is not included into a course syllabus, but aimed at extending knowledge about the subject, is up to the student's own initiative. A teacher recommends relevant resources for self-study, defines relevant methods for self-study and demonstrates students' past experiences. Tasks for self-study and its content can vary depending on individual characteristics of a student. Self-study can be arranged individually or in groups both offline and online depending on the objectives, topics and difficulty degree. Assessment of self-study is made in the framework of teaching load for seminars or tests.

Special conditions for organization of learning process for students with special needs

The following types of comprehension of learning information (including e-learning and distance learning) can be offered to students with disabilities (by their written request) in accordance with their individual psychophysical characteristics:

1) *for persons with vision disorders:* a printed text in enlarged font; an electronic document; audios (transferring of learning materials into the audio); an individual advising with an assistance of a sign language interpreter; individual assignments and advising.
2) *for persons with hearing disorders: a* printed text; an electronic document; video materials with subtitles; an individual advising with an assistance of a sign language interpreter; individual assignments and advising.
3) *for persons with muscle-skeleton disorders: a* printed text; an electronic document; audios; individual assignments and advising.

Annex 3

Sample exam questions

Exam questions (100 points in total)

Problem 1 (15 points). We have considered five basic estimators applicable to panel data (pooled OLS, between, within [or fixed effects], first differences, and random effects).
(15) What can be said about consistency of these estimators if the true model is/data generating process follows (a) pooled OLS, (b) random effects and (c) fixed effects?

Problem 2 (15 points).
(5) Formulate a random trend model.
(5) How would you estimate it?
(5) Can you give an example of the use of such models?

Problem 3 (25 points). Consider a simple cross-sectional model $y= b_0+ b_1x_1^*+ u$ with a measurement error in the regressor $e_1= x_1–x_1^*$ (x_1^* is an unobserved measure, x_1 is its proxy). Assume $E(e_1) = 0$, $E(y| x_1^*, x_1) = E(y| x_1^*)$. We are thus really estimating $y= b_0+ b_1x_1+ (u–b_1e_1)$.
(5) How can the classical measurement error (CME) be defined in this example?
(15) What is the main consequence of CME for the OLS estimate of b_1 (derive a formula)?
(5) In general, is it likely to be more severe in the fixed effects model as compared to the cross-sectional case? Provide intuition.

Problem 4 (20 points).
(5) Write down a dynamic panel data model in a general form.
(15) Describe the Anderson and Hsiao approach to estimating its parameters.

Problem 5 (25 points). Here is an excerpt from a paper that studies the impact of capital flows on investment in a panel dataset of 22 transition countries for 10 years (1995 – 2004).

```
xtabond2 inv l.inv fdi loans portfolio l.growth uncert tot dev_m2, gmm (inv fdi
loans portfolio, lag (2 2)) iv(fin_integr trans_index flows_eeca l.growth
uncert tot dev_m2) nolevel
```

Arellano-Bond dynamic panel-data estimation, one-step difference GMM results

Group variable: ctry_dum				Number of obs	=	165
Time variable : year				Number of groups	=	22
Number of instruments = 39				Obs per group: min	=	3
F(8, 157)	=	6.88		avg	=	7.50
Prob > F	=	0.000		max	=	8

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
inv						
L1.	.2922856	.111738	2.62	0.010	.0715819	.5129893
fdi	.5202847	.2094545	2.48	0.014	.1065725	.933997
loans	.2789421	.1638248	1.70	0.091	-.044643	.6025271
portfolio	-.0086876	.3376843	-0.03	0.980	-.6756779	.6583028
growth						
L1.	.1167961	.0555715	2.10	0.037	.0070319	.2265604
uncert	.0397982	.0673439	0.59	0.555	-.0932187	.172815
tot	.9193659	1.916147	0.48	0.632	-2.865388	4.704119
dev_m2	.0443079	.0760188	0.58	0.561	-.1058435	.1944594

Sargan test of overid. restrictions: chi2(31) = 36.42 Prob > chi2 = 0.231

Arellano-Bond test for AR(1) in first differences: z = -0.01 Pr > z = 0.992
Arellano-Bond test for AR(2) in first differences: z = -0.48 Pr > z = 0.628

Here, **inv** stands for gross fixed capital formation as a percentage of GDP, **fdi**, **loans** and **portfolio** measure foreign direct investments, loans and portfolio (equity and bonds) as percentage shares of GDP, **growth** is a real GDP growth; **uncert** is a measure of economic uncertainty; **tot** is the change in the log terms of trade to gauge the price of imported capital goods; and **dev_m2** is a proxy for the liquidity available to finance investment. Finally, **flows_eeca** stands for the aggregate long-term capital inflows to the countries in the sample as a group, **fin_integr** an index of financial openness and **trans_index** is the EBRD transition index (measuring reform progress in the transition countries).

(5) Describe (write down) the model (main equation) corresponding to the above table with regression results.

(10) What is being instrumented here and with what instruments?

(10) How does the model perform: does it pass the necessary specification tests? Explain.

Annex 4

Аннотация на русском языке

Панельные данные (прикладная эконометрика панельных данных)

Курс предназначен для студентов третьего курса бакалавриата направления обучения «Экономика». Его главная цель - познакомить студентов с современными методами анализа панельных данных, начиная с модели пула и заканчивая динамическими моделями панельных данных. Курс имеет прикладную природу: материал излагается, по мере возможности, максимально нетехнически, с примерами эмпирических исследований, опубликованных в ведущих международных журналах по экономике и финансам, обсуждаемыми в аудитории. Лекции дополнены компьютерными классами, что позволяет студентам получить практический опыт анализа реальных данных с использованием статистического пакета Stata 14/15.

Охватываемые темы включают: модель пула, модели с фиксированными и случайными эффектами, несбалансированные панели, ошибки измерения в панельных данных, эндогенные объясняющие переменные, модель Хаусмана-Тейлора и обзор динамических панельных моделей (смещение Никелля, модели Андерсона-Сяо и Ареллано-Бонда).

По завершении курса:

- Студенты должны иметь четкое представление о ключевых методах анализа панельных данных.
- Студенты должны иметь возможность эффективно применять эти методы в собственных эмпирических исследованиях.
- Студенты должны быть знакомы и уметь использовать ключевые возможности статистического пакета «Stata», включая его опции программирования (так называемые do-файлы).

Курс состоит из лекций (20 часов) и компьютерных классов (16 часов). Основные методы преподавания и обучения включают лекции, работа с литературой, упражнения в компьютерной лаборатории и эмпирический проект.

Усвоение материала курса оценивается с помощью домашнего задания, промежуточного экзамена (теста) и итогового экзамена.