Approved

Academic Council of educational program “Applied Economics”
Protocol number 01, « 01» February 2018
Academic Council of educational program “Strategic Corporate Finance”
Protocol number n/a, « 26» June 2018

Author	Elena Kotyrlo, D.Sc., ekotyrlo@hse.ru
Credits	11
Contact Hours	144
Self-study Hours	274
Course	1

Section 1. General information about the course

The course “Advanced Econometrics” focuses on the estimation, inference and identification of regression models. Particular attention is paid to the econometric theory, to the application of econometrics to real-world problems, and to the interpretation of the estimation results. The first part of the course (Fall term) includes linear regressions and models with limited dependent data. Topics on Gauss-Markov theorem, endogeneity, instrumental variables, maximum likelihood estimation will be covered. The second part of the course (Spring term) is focused on issues in system of equations; time series models; panel data models; nonparametric and semiparametric models; Bayesian estimation.

The course will include the use of STATA and MS Excel. Use of R and other statistical analysis software is optional.

Section 2. Course goals, learning objectives, expected learning outcomes

The course aims to provide students with:

- knowledge on the fundamentals of econometrics and its application
- knowledge and proficiency on the use of statistical package STATA for econometric analysis
- practice in conducting data analysis and application of econometric tools in research and analytics

Prerequisites

Course requires knowledge of linear algebra, calculus, probability theory and mathematical statistics.

Section 3. Course Outline

No	Topic/Focus/Activity	Number of the week (lecture)	Readings and assignments
1.	Introduction	1	G, Ch. 1, V, Ch. 1
2.	Matrix algebra		V, Appendix A G, Appendix A
3.	Theory of probabilities and statistics. Estimation and inference.		V, Appendix B G, Appendices B,C
4.	The linear regression model. Least squares. Goodness-of-fit and analysis of variance.	2	V, Ch.2, 2.1, 2.2, 2.4 G, Ch. 2-3
5.	The Gauss-Markov theorem. Linear hypothesis testing.	3-5	V, Ch.2, 2.3, 2.5, 2.6, 2.8, 2.9 G, Ch. 4, 5
6.	Interpreting and comparing regression models. Functional form and structural change. Multicollinearity. Nonlinear models	6-7	V, Ch.2, 2.8 Ch.3 G, Ch. 4, 5, 6
7.	Heteroscedasticity. Generalized least squares.	8-9	V, Ch.4, 4.1-4.5 G, Ch.9
8.	Autocorrelation. Testing for first order autocorrelation.	10	V, Ch.4, 4.6-4.11 G, Ch.20
9.	Endogeneity, instrumental variables and GMM	11-12	V, Ch.5 G, Ch.13, C&T, Ch.6
10.	Panel data models. Introduction	13	V, Ch. 10, 10.1-10.3, B, Ch.2-3, G, Ch. 11,
11.	Maximum likelihood estimation and specification tests.	14-15	V, Ch.6, 6.1-6.2 G, Ch.14
12.	Binary choice models	16	V, Ch.7, 7.1, 7.2. G, Ch.17, 17.1-17.3
13.	Multi-response models. Models for count data	17-19	V, Ch.7, 7.3, 7.4, 7.5, 7.6. G, Ch.18, 18.4
14.	Tobit models. Sample selection bias	20-21	V, Ch. 7.4-7.6 G., Ch. 17.4-17.5
15.	Estimating Treatment Effects	22	V Ch.7, 7.7, G. Ch. 6.2.5, Ch. 19.6, C-T, Ch. 3.6.4
16.	Univariate time series models. General ARMA processes. Testing for unit roots	23-24	V, Ch.8.1-8.5 G, Ch.21, 21.1, 21.2
17.	Choosing ARMA model and its estimation	25	V, Ch.8, 8.6-8.9
18.	ARCH and GARCH models	26	V, Ch.8, 8.10

No	Topic/Focus/Activity	Number of the week (lecture)	Readings and assignments
19.	Multivariate time series models Models with stationary and nonstationary variables.	27-28	V, Ch.9 G, Ch.21, 21.3
20.	Dynamic panel data	29	V, Ch. 10.4-10.8, B, Ch. 8
21.	Non-parametric and semiparametric methods. Kernel density estimation	30	G, Ch.7, 7.5, Ch.12, 12.4 Ke
22.	Duration models	31	V, Ch.7, 7.8 G, Ch.19, 19.4
23.	Simulation-based estimation. Bootstrap standard errors	32	G, Ch.15, 15.2-15.4 C&T, Ch.11
24.	Bayesian estimation and inference	33	G, Ch.16 C&T, Ch.13
25.	Spatial econometrics	34	E, G Ch.11, 11.7
26.	Nonlinear regression models. Quantile regression	35	G, Ch.7, 7.3 Ko
27.	Big data. Shrinkage Methods. Ridge regression, Lasso, elastic Lasso.	36	H&T, Ch. 3

Description of course methodology and forms of assessment to be used:

Classes will meet twice per week, for a 2-hour lecture and a 2- hour tutorial. Students are expected to attend lectures and actively participate in all tutorial activities.

Individual meetings can be arranged during office hours (e-mail request for a meeting is desirable). The lecturer will hold office hours, except for breaks in the program and holidays. Additional hours will be held by the tutor.

Section 4. Course Overview

1. Introduction. About econometrics and the course program.

2. Vectors and matrices. Terminology. Matrix manipulations. Properties of matrices and vectors. Inverse matrices. Idempotent matrices. Eigenvalues and eigenvectors. Differentiation. Some least squares manipulations.

3. Statistical and distribution theory. Discrete random variables. Continuous random variables. Expectations and moments. Multivariate distributions. Conditional distributions. The normal distribution and its properties. t-distribution. Chi squared distribution. F-distribution.

Cross-sectional data

4. An introduction to linear regression. Ordinary least squares (OLS). Simple linear regression. Matrix notation. The linear regression model. Goodness-of-fit.

5. **The Gauss–Markov assumptions.** Small sample properties of the OLS estimator. Hypothesis testing. A simple *t* -test. Testing one linear restriction. A joint test of significance of regression coefficients. The general case. Size,

power and *p*-values. Asymptotic properties of the OLS estimator. Consistency. Asymptotic normality.

6. Interpreting and comparing regression models. Multicollinearity. Interpreting the linear model. Selecting the set of regressors. Misspecifying the set of regressors. Comparing non-nested models. Misspecifying the functional form. Nonlinear models. Testing the functional form. Testing for a structural break. Linear models. Loglinear models.

7. Heteroskedasticity. Heteroskedasticity. Introduction. Consequences for the OLS estimator. Deriving an alternative estimator. Estimator properties and hypothesis testing. Heteroskedasticity-consistent standard errors for OLS. A model with two unknown variances. Multiplicative heteroskedasticity. Testing for heteroskedasticity. Testing equality of two unknown variances. Testing for multiplicative heteroskedasticity. The Breusch–Pagan test. The White test. The Goldfeld-Quandt test.

8. Autocorrelation. Autocorrelation. First order autocorrelation. Estimation with a known ρ. Estimation with ρ unknown. Testing for first order autocorrelation. Asymptotic tests. The Durbin–Watson test. What to do when you find autocorrelation? Misspecification. Heteroskedasticity-and-autocorrelation-consistent standard errors for OLS.

9. Endogeneity, instrumental variables and GMM. Cases where the OLS estimator cannot be saved. Autocorrelation with a lagged dependent variable. Measurement error and simultaneity as a cause of endogeneity. The instrumental variables estimator. Estimation with a single endogenous regressor and a single instrument. Multiple endogenous regressors. The generalized instrumental variables estimator. Two-stage least squares. Specification tests. Weak instruments. The generalized method of moments (GMM).

Panel data

10. Models based on panel data. Advantages of panel data. Efficiency of parameter estimators. Identification of parameters. The fixed effects model. The random effects model. Fixed effects or random effects. The Hausman test. The Breush-Pagan test. F-test. Goodness-of-fit.

Estimation techniques

11. Maximum likelihood estimation and specification tests. An introduction to maximum likelihood (ML). General properties. The normal linear regression model. Specification tests: the Wald tests, likelihood ratio (LR), the Lagrange multiplier test (LM), Tests in the normal linear regression model. Testing for omitted variables. Testing for heteroskedasticity. Testing for autocorrelation.

Quasi-maximum likelihood and moment conditions tests. Quasi-maximum likelihood. Testing for normality.

Models with limited dependent variables

12. Binary choice models. Binary choice models. Using linear regression. Introducing binary choice models. An underlying latent model. Estimation. Goodness-of-fit. Specification tests in binary choice models. Relaxing some assumptions in binary choice models.

13. Multi-response models. Models for count data. Ordered response models. About normalization. Multinomial models. Models for count data. The Poisson and negative binomial models.

14. Tobit models. The standard Tobit model. Estimation. Specification tests in the Tobit model. Extensions of Tobit models. The Tobit II model. Estimation.

15. Estimating treatment effects. Sample selection bias. The nature of the selection problem. The average treatment effect. The average treatment effect for the treated. Difference-in-difference. Switching regression model.

Time series

16. Univariate time series models. Introduction. Stationarity and the autocorrelation function. General autoregressive-moving average (ARMA) processes. Formulating ARMA processes. Invertibility of lag polynomials. Common roots. Stationarity and unit roots. Testing for unit roots. Testing for unit roots in a first order autoregressive model. Testing for unit roots in higher order autoregressive models.

17. Choosing ARMA model and its estimation. The autocorrelation function. The partial autocorrelation function. Diagnostic checking. Criteria for model selection. Estimation of ARMA models by LS and ML. Predicting with ARMA models. The optimal predictor. Prediction accuracy.

18. Autoregressive conditional heteroskedasticity (ARCH). ARCH and GARCH models. Estimation and prediction.

19. Multivariate time series models. Dynamic models with stationary variables. Models with nonstationary variables. Spurious regressions. Cointegration. Cointegration and error-correction mechanisms. Vector autoregressive models (VAR). Cointegration: the multivariate case. Cointegration in a VAR. Testing for cointegration.

Panel data

20. Dynamic linear models. An autoregressive panel data model. Difference and system GMM. Dynamic models with exogenous variables. Nonstationarity, unit roots and cointegration. Panel data unit root tests. Panel data cointegration tests.

Estimation methodology

21. Non-parametric and semiparametric methods. Kernel density estimation. Statistical properties of estimators. Kernel functions. Nonparametric regression. Smoothing functions.

22. Duration models. Duration models as an example of semiparametric estimation. Hazard rates and survival functions. Proportional hazards models. Estimation.

23. Simulation-based estimation. Bootstrap standard errors. Random number generation. Generating pseudo-random numbers. Sampling from a standard uniform population. Sampling from continuous distributions. Sampling from a multivariate normal population. Sampling from discrete populations. Simulation-based statistical inference. Bootstrapping standard errors and confidence intervals.

24. Bayesian estimation and inference. Bayes theorem and the posterior density. Bayesian analysis of the classical regression model. Analysis with a noninformative prior. Estimation with an informative prior density. Bayesian inference. Point estimation. Interval estimation. Hypothesis testing. Large-sample results. Posterior distributions and the Gibbs sampler.

25. **Spatial econometrics.** Spatial autocorrelation. Contiguity matrix, W. A“time-space simultaneous” model. spatial lags. Spatial error correlation

26. Nonlinear regression models. Quantile regression. Nonlinear regression models. Assumptions of the nonlinear regression model. The nonlinear least squares estimator. Large sample properties of the nonlinear least squares estimator. Hypothesis testing and parametric restrictions. Quantile regression.

27. Shrinkage methods. Big data. Ridge regression, The Lasso, elastic net-penalty.

Section 5. Texts, readings and other informational resources

Required readings:

Verbeek M. A guide to modern econometrics. John Wiley & Sons, Ltd. 2004. [V]

Additional readings:

Baltagi B.H., Econometric analysis of panel data. John Wiley & Sons, Ltd. 2005 [B]

Cameron A. C., Trivedi P. K. Microeconometrics: Methods and applications. Cambridge university press, 2005. [C&T]

Elhorst J. P., Spatial econometrics: From cross-sectional data to spatial panels. Heidelberg, New York, Dordrecht, London : Springer, 2014. [E]

Greene W. H. Econometric analysis. 7 th. ed, Prentice Hall, NY, 2012 [G]

Hardle W., Muller M., Sperlich S., Werwatz A., Nonparametric and semiparametric models, Springer, 2004. [H]

Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer, 2008. [H&T]

Keele L., Semiparametric regression for the social sciences, John Wiley&Sons, Ltd., 2008. [Ke]

Koenker R., Quantile regression, Cambridge University Press, 2005. [Ko]

Section 6. Examination/Evaluation

Assessment and grade determination for the Fall term:

- The first intermediate test (IT1, Module 1) includes tests and problems on the topics 4-6.
- The first homework (HW1, Module 1): the course participants propose a hypothesis and collect their own cross sectional data for a regression model.
- The second homework (HW2, Module 2) is based on data collected in Module 1 and approved by a tutor. It imposes empirical justification of the stated hypothesis the use of the statistical package STATA (another software is optional) for data analysis on the base of the material of the topics 4-9. Students are expected to use statistical software STATA or another for data analysis.
- Midterm exam includes tests and problems on the topics 4-11.
- The cumulative score for the Fall term = 0.5 IT1 + 0.1 HW1 +0.2 HW2 + 0.2 Activity
- The final grade for the Fall term = 0.5 cumulative score for the Fall term + 0.5 midterm exam. A remarkable difference between the cumulative

score and the midterm exam score may be a reason for additional oral examination at the discretion of the tutor.

Assessment and grade determination for the Spring term:

- The second Intermediate Test (IT2, Module 3) includes test and problems on the topics 11-15.
- The third homework (HW3, Module 3). The course participants may collect their own data or relay on data given by tutors and use the statistical package STATA (another software is optional) for data analysis. Econometric techniques are based on the topics 10-15).
- The forth homework (HW4, Module 4) includes empirical justification of hypotheses relevant to time series analysis. It is based on the material of the topics 16-19.
- Final exam includes test and problems on the topics 16-27.
- The cumulative score for the Spring term = 0.5 IT2 +0.15 HW3+ 0.15 HW4 + 0.2 activity score
- The final grade for the Spring term = 0.2 final grade for the Fall term + 0.3 cumulative score for the Spring term + 0.5 final exam. A remarkable difference between the cumulative score and the final exam score may be a reason for additional oral examination at the discretion of the tutor.

Activity on classes is calculated as follows:

Maximum of activity score is 10 and ranged on the base of the average activity in the class.
Student is marked to be active on a class if he\she solves problems at home and at the (black-) whiteboard.

Section 7. Academic Integrity

The Higher School of Economics strictly adheres to the principle of academic integrity and honesty. Accordingly, in this course there will be a zero-tolerance policy toward academic dishonesty. This includes, but is not limited to, cheating, plagiarism (including failure to properly cite sources), fabricating citations or information, tampering with other students' work, and presenting a part of or the entirety of another person's work as your own. HSE uses an automated plagiarism-detection system to ensure the originality of students' work. Students who violate university rules on academic honesty will face disciplinary consequences, which, depending on the severity of the offense, may include having points deducted on a specific assignment, receiving a failing grade for the course, being expelled from the university, or other measures specified in HSE's Internal Regulations.