UNIVERSITY OF LONDON

Course information 2023-24
EC2020 Elements of econometrics

General information

COURSE LEVEL: 5

CREDIT: 30

NOTIONAL STUDY TIME: 300 hours

Summary

Econometrics is the application of statistical methods to the quantification and critical assessment of hypothetical economic relationships using data. This course gives students an opportunity to develop an understanding of econometrics to a standard that will equip them to understand and evaluate most applied analysis of cross-sectional data and to be able to undertake such analysis themselves.

Conditions

Prerequisite: If taken as part of a BSc degree, the following course(s) must be attempted before you can register on this course.

• EC1002 Introduction to economics **AND**

• ST104A Statistics 1 (half course) **AND** ST104B Statistics 2 (half course) **OR** ST1215 Introduction to Mathematical Statistics† **AND**

• MT105A Mathematics 1 (half course) **AND** MT105B Mathematics 2 (half course) **OR** MT1174 Calculus **OR** MT1186 Mathematical methods†

† You are advised to take a full unit mathematics course (MT1174 Calculus OR MT1186 Mathematical methods) and the full unit statistics course (ST1215 Introduction to mathematical statistics) as they provide a stronger foundation in concepts used in EC2020 Elements of econometrics than are provided by the half-unit combinations MT105A Mathematics 1 + MT105B Mathematics 2 and ST104A Statistics 1 + ST104B Statistics 2

Aims and objectives

The aims of this course are:

- To understand the nature and scope of econometrics as a social science.
- To develop an understanding of the use of regression analysis and related techniques for quantifying economic relationships and testing economic theories.
- Provide a good foundation for further study in econometrics and statistical techniques.
- To equip students to read and evaluate empirical papers in professional journals.
- To provide students with practical experience of using mainstream regression programmes to fit economic models.

Please consult the current EMFSS Programme Regulations for further information on the availability of a course, where it can be placed on your programme's structure, and other important details.

Learning outcomes

At the end of the course and having completed the essential reading and activities students should be able to:

- Describe and apply the classical regression model and its application to cross-section data.
- Describe and apply the:
  - Gauss-Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables.
- Recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models.
- Competently use regression, logit and probit analysis to quantify economic relationships using R.
- Describe and explain the principles underlying the use of maximum likelihood estimation.
- Apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time series applications (for example, autocorrelation) and remedial measures.
- Recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by co-integration.

Essential reading

For full details, please refer to the reading list: Wooldridge, J. *Introductory Econometrics: A Modern Approach.* (Cengage, 2020) seventh edition [ISBN 9781337558860]).

Assessment

This course is assessed by a three-hour and fifteen-minute closed-book written examination.

Syllabus

Review: Basic mathematical tools, fundamentals of probability and mathematical statistics. Summation operator. Linear, exponential and logarithmic function. Elasticity. Percentage change. Proportionate change. Probability distribution of a random variable. Continuous and discrete random variables. Expected value of a random variable. Expected value of a function of a random variable. Population variance and standard deviation. Expected value rules. Independence of two random variables. Population covariance, covariance and variance rules, and correlation. Conditional Expectation. Normal distribution. Sampling and estimators. Unbiasedness. Estimator of the population variance. Efficiency. Mean square error. Convergence in probability and plim rules. Consistency. Convergence in distribution (asymptotic limiting distributions) and the role of central limit theorems. Confidence intervals. Hypothesis testing. Type I error and the significance of a test. Type II error and the power of a test. T tests. One-sided tests.

Simple regression model. Derivation of linear regression coefficients. Interpretation of a regression equation. Regression coefficients as random variables. SLR Assumptions. Unbiasedness of the regression coefficients. Precision of the regression coefficients. Important results relating to OLS regressions. Goodness of fit. Units of measurement and functional form. Elasticities and double-logarithmic models. Semilogarithmic models.

Multiple regression analysis: Estimation. Multiple regression with two explanatory variables. Mechanics and interpretation of OLS. Ceteris Paribus. Partialling-out interpretation of OLS. MLR assumptions. Expected value of the multiple regression coefficients. Variance of the regression coefficients. Multicollinearity. Gauss-Markov Theorem. Effect of including irrelevant variables. Omitted Variable Bias.

Multiple regression analysis: Inference. Sampling Distribution of regression coefficients. Classical Linear Regression Model. *t* test of a hypothesis relating to a regression coefficient. Type I error and Type II error. Confidence intervals. One- and two-sided tests. P-values of a test. Test hypothesis about a single linear combination of parameters. Reparameterization of a regression model. Tests of multiple restrictions. *F* test.

Multiple regression analysis: Further Issues (Quadratics and Interactions). Models with quadratic and interactive variables. Adjusted R-squared.

Multiple regression analysis with Qualitative Information (Dummy Variables). Dummy variables. Dummy classification with more than two categories. The effects of changing the reference category. The dummy variable trap. Multiple sets of dummy variables. Interactions involving dummy variables. Allowing for different slopes. Chow test.

Multiple regression analysis: Asymptotics. Asymptotic properties of the regression coefficients, consistency and asymptotic normality. Law of large numbers and central limit theorem.

Heteroskedasticity: Meaning of heteroscedasticity. Consequences of heteroscedasticity. Breush-Pagan and White tests for heteroscedasticity. Elimination of heteroskedasticity using weighted or logarithmic regressions. Heteroskedasticity robust inference after OLS (White standard errors).

Instrumental Variable Estimation and Two-Stage Least Squares. Instrumental variables (IV). Three requirements of an instrument. Asymptotic properties of IV estimators. Multiple instruments. Two-stage least squares (TSLS). Testing for endogeneity.

Measurement Errors. Measurement error and its consequences.

Simultaneous Equation Models. Definitions of endogenous variables, exogenous variables, structural form equations and reduced form equations. Simultaneity bias: Inconsistency of OLS on structural form equations. Use of instrumental variables in SEMs. Two-stage least squares. Exact identification, under identification, and over identification. Order and rank condition for identification.

Binary Response Models and Maximum Likelihood Estimation. Linear probability model. Logit and probit analysis. Fundamentals of Maximum likelihood estimation. Asymptotic properties of MLE. Likelihood ratio tests and asymptotic t tests.

Regression Analysis with Time Series Data. Static models and distributed lag models (lagged variables). Long run and short run effects. Autoregressive distributed lag (ADL) model. Assumptions for regressions with time series data. Strict exogeneity. Stationary and weakly dependent timeseries. AR and MA processes. Asymptotic properties of OLS estimators of ADL models, including asymptotic limiting distributions.

Serial Correlation in Time Series Regression. Definition of autocorrelation. Consequences of autocorrelation. Test for autocorrelation. Autocorrelation and heteroscedasticity robust inference after OLS (Newey-West, HAC standard errors). Autocorrelation in the presence of lagged dependent variable. Apparent autocorrelation caused by variable or functional misspecification.

Trends, Seasonality, and Highly Persistent Time Series in Regression Analysis. Trends and seasonality. Stationary and nonstationary processes. Random walk (with drift). Difference-stationarity and Trend-stationarity. Integrated of order 1. Spurious regressions. Unit root tests. Co-integration. Error correction models. Engle-Granger two-step procedure.