INTRODUCTION TO ECONOMETRICS (EC212)

Course duration: 54 hours lecture and class time (Over three weeks)

LSE Teaching Department: Department of Economics

Lead Faculty (session two): Dr Taisuke Otsu and Dr Marcia Schafgans

Lead Faculty (session three): Dr Christopher Dougherty

Pre-requisites: An introductory statistics course and at least one semester of multivariate calculus, passed with respectable grades.

Course Objective:

The objective of this course is to provide the basic knowledge of econometrics that is essential equipment for any serious economist or social scientist, to a level where the participant would be competent to continue with the study of the subject in a graduate programme. While the course is ambitious in terms of its coverage of technical topics, equal importance is attached to the development of an intuitive understanding of the material that will allow these skills to be utilised effectively and creatively, and to give participants the foundation for understanding specialized applications through self-study with confidence when needed.

Content:

The first part of the course introduces the statistical tool known as regression analysis applied to cross-sectional data. It begins with the use and properties of the classical linear regression model and then discusses how various technical problems should be handled. Initially ordinary least squares is the standard technique, but eventually the focus shifts to instrumental variables estimation. The second part of the course discusses the application of the regression model to time series data.

Although the course does not make use of matrix algebra, it is technically rigorous, giving emphasis to the analysis of the finite sample and asymptotic properties of least squares and instrumental variables estimators under different assumptions concerning the data generation process and to the accompanying implications for statistical inference. Participants are expected to be able to provide proofs of the unbiasedness or biasedness, and consistency or inconsistency of least squares and instrumental variables estimators in simple models where appropriate. Examples of applications in economics are used throughout.

In addition to the formal theory sessions, participants take part in daily workshops using Stata to fit educational attainment and wage equation models with cross-sectional data and EViews to fit demand functions with time series data.

Teaching:

The course is taught by different faculty in the two sessions. While the majority of the content will overlap, there are differences in the coverage on some topics as indicated below and the text books on which the sessions are based differ.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

EC212 - SESSION TWO

Lead Faculty: Dr Taisuke Otsu and Dr Marcia Schafgans

Course Text:

The course text will be the fifth (or sixth) edition of J.M. Wooldridge, *Introductory Econometrics: A Modern Approach,* International Edition (or, alternatively, the Emea 1st Edition + its online content)

Topics:

1. Simple Regression Analysis

It will be shown how a hypothetical linear relationship may be quantified using appropriate data. The principle of least squares analysis will be explained and expressions for coefficients derived. Conditions for the unbiasedness of the regression coefficients and factors governing their variance will be discussed.

2. Multiple Regression Analysis: Estimation

The regression analysis is generalized to cover the case where there are several or many explanatory variables in the regression model, rather than just one. The problem of multicollinearity and the use of restrictions will also be discussed. This discussion also covers: the consequences of including in the regression a variable which should not be there; the consequences of leaving out a variable that should be included; and methods for testing whether a restriction is valid.

3. Multiple Regression Analysis: Inference

Based on the properties of the estimated regression coefficients, hypothesis testing and the construction of confidence intervals of the regression model will be discussed. Various inference methods including single and multiple hypotheses testing will be covered.

4. Multiple Regression Analysis: Further Topics

To deal with the situation where economic data follow non-normal distributions, some asymptotic properties of the regression coefficients will be discussed. Additional topics related to goodness-of-fit, functional form, and prediction will be covered.

5. Specification of Regression Variables

Many economic processes are best modelled by nonlinear relationships: demand functions and production functions, for example. Linear regression analysis will be extended to cover nonlinear models. It frequently happens that some of the factors which should be taken account of in a regression model are qualitative in

nature and therefore not measurable in quantitative terms. It will be explained how such factors can be handled with dummy variables.

6. Heteroscedasticity

The course now turns to a common problem encountered in regression analysis, discussing its detection and consequences, and means to alleviate it.

7. Efficiency (OLS versus WLS) and More on Specification and Data Issues

We discuss an optimality property of OLS (Gauss-Markov theorem) and introduce the WLS estimator (a Generalized Least Squares estimator) that is efficient in the presence of heteroscedasticity.

Functional misspecification, proxy variables and missing data are discussed.

8. Time series data Modelling Dynamic Processes

Many economic processes require time to work themselves out and accordingly there is a need to introduce a time dimension into econometric modelling. The autoregressive distributed lag (ADL) model, an important method of modelling dynamics, is discussed and the asymptotic and finite-sample properties of estimators are investigated. Important concepts introduced are stationarity and weak dependence

9. Autocorrelation

In time series models it is common for the errors to exhibit autocorrelation. We will discuss its detection, consequences, and means to alleviate it.

10. Endogeneity – omitted variables, measurement error and simultaneity

One further complication of the regression model discussed is that of correlatedness between the errors and regressors. We discuss various examples where this occurs: omitting relevant variables, measurement errors, lagged endogenous variables in the presence of autocorrelation, and simultaneity. We discuss the undesirability of using OLS in these settings and discuss a simple test for endogeneity.

11. Instrumental variable estimation and two stage least squares

The IV estimator (and the two stage least squares estimator) is an estimator that provides consistent estimates in the presence of endogeneity. In the presence of more instruments then we need (over-identification), the two stage least squares estimator is introduced as the optimal IV estimator in the presence of homoscedasticity.

12. Non Stationary Time Series

Many economic processes exhibit trending behaviour, such as GDP and consumption, that render them non-stationary. We will consider two types of non-stationary processes: trend stationary processes and difference stationary processes. While the former allows us to use standard regression analysis (weak dependence), the latter gives rise to persistence (strong dependence) that renders standard regression analysis invalid. As shocks have very different consequences in these settings, it is important to distinguish these processes and we will consider the DF and ADF test (tests for unit roots) for this purpose.

Conducting regression analysis with variables that exhibit a unit root can give rise to a spurious, or meaningless, relationship. We need to distinguish this from a cointegrating relationship, where our regression analysis is reflective of a long run relationship between the variables. We will show how to test for the presence of a spurious relationship.

Computer Exercises

Participants will be given homework assignments every day and some of these will involve the use of the computing facilities at LSE. The course will use the Stata regression application. No prior knowledge of this application is required. Assistance with the practical work will be provided.

Summer School

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

EC212 - SESSION THREE

Lead Faculty: Dr Christopher Dougherty

Course Text:

The course text will be the fifth edition of C. Dougherty, *Introduction to Econometrics*, Oxford University Press, 2016

Presentation:

Computer video graphics will be used to present the analysis.

Topics:

1. Simple Regression Analysis

It will be shown how a hypothetical linear relationship may be quantified using appropriate data. The principle of least squares analysis will be explained and expressions for coefficients derived.

2. Properties of Regression Coefficients and Hypothesis Testing

Conditions for the unbiasedness of the regression coefficients and factors governing their variance will be discussed, together with hypothesis testing and the construction of confidence intervals in the context of the regression model.

3. Multiple Regression Analysis

Least squares regression analysis is generalized to cover the case where there are several or many explanatory variables in the regression model, rather than just one. The problem of multicollinearity and the use of restrictions will also be discussed.

4. Transformation of Variables

Many economic processes are best modelled by nonlinear relationships: demand functions and production functions, for example. Linear regression analysis will be extended to cover nonlinear models.

5. Dummy Variables

It frequently happens that some of the factors which should be taken account of in a regression model are qualitative in nature and therefore not measurable in quantitative terms. It will be explained how such factors can be handled with dummy variables.

6. Specification of Regression Variables

This preliminary discussion of specification issues covers: the consequences of including in the regression a variable which should not be there; the consequences of leaving out a variable that should be included; what happens if there is difficulty finding data on a variable and use a proxy for it instead; and methods for testing whether a restriction is valid.

7. Heteroscedasticity

The course now turns to a common problem encountered in regression analysis, discussing its detection and consequences, and means to alleviate it.

8. Stochastic Regressors and Measurement Errors

Two further complications of the regression model are discussed, and the Instrumental Variable approach to the fitting of a regression model is introduced. The use of the Durbin-Wu-Hausman specification test is described.

9. Simultaneous Equations Estimation

The next topic is the estimation of relationships that jointly form a simultaneous equations model (for example, a simple macroeconomic model). It will be shown that Ordinary Least Squares is in general not a suitable regression technique and that Instrumental Variables should be used instead.

10. Modelling Dynamic Processes

Many economic processes require time to work themselves out and accordingly there is a need to introduce a time dimension into econometric modelling. The autoregressive distributed lag (ADL) model, an important method of modelling dynamics, is discussed and the asymptotic and finite-sample properties of estimators are investigated. The course concludes with a treatment of two special issues that arise in regressions using time series data. One relates to the use of simulation. In many situations, mathematical analysis allows only the asymptotic properties of estimators to be established It will be shown how simulation can be used to investigate their finite-sample properties as well.

11. Autocorrelation

Autocorrelation of the disturbance term gives rise to a further set of issues that are often important in time series regressions. The most common tests for autocorrelation—the Breusch–Godfrey test, the Durbin–Watson test, and the Durbin *h* test—are presented and methods of rectifying the problem are discussed. The treatment leads to a more general discussion of model specification.

12. Panel data regressions

The final mainstream topic of the course is the application of regression analysis to panel (longitudinal) data sets where one has repeated observations on the same set of agents (for example, individuals, households, or enterprises). It will be shown how some important econometric problems may be overcome through the use of fixed effects or random effects models. (This set of topics, being relatively technical, will not be included in the syllabus for the final examination.)

Optional Topics

The last teaching day will be used to give an outline of some advanced topics that are not part of the syllabus for the examinations: binary choice models (logit and probit), maximum likelihood estimation, and an introduction to regression models with nonstationary time series.

Computer Exercises

Participants will be given homework assignments every day and some of these will involve the use of the computing facilities at LSE. The course will use the Stata and EViews regression applications. No prior knowledge of these applications is required. Assistance with the practical work will be provided.

Credit Transfer: If you are hoping to earn credit by taking this course, please ensure that you confirm it is eligible for credit transfer well in advance of the start date. Please discuss this directly with your home institution or Study Abroad Advisor.

As a guide, our LSE Summer School courses are typically eligible for three credits within the US system and 7.5 ECTS in Europe. Different institutions and countries can, and will, vary. You will receive a digital transcript and a printed certificate following your successful completion of the course in order to make arrangements for transfer of credit.

If you have any queries, please direct them to summer.school@lse.ac.uk