Econometrics I

MF2 (Michaelmas Term)

Course leader: Paul Kattuman (p.kattuman@jbs.cam.ac.uk)

This module is intended to enable students to understand how linear regression and associated statistical techniques are used to estimate causal relationships from primarily cross-sectional, observational data.

Aims and objectives

To introduce students to the principles and methods in data analysis, and their application. By the end of the course students should:

- Understand how linear regression and associated statistical techniques are used to estimate *causal* relationships from cross-sectional data.
- Be able to specify, estimate, test, interpret and critically evaluate single equation regression models, with applications in a variety of subject areas of management, finance and business economics.
- Be equipped to read and critically evaluate empirical papers in research journals in management.

Content

There will be two 90 minute sessions each week. Friday sessions focus on concepts and examples. Wednesday sessions focus on estimation and involve hands-on computing. *R* is the primary computing platform used in this course. Corresponding *Stata* syntax and commands will also be available, so that students are able to use either or both for applied econometric analysis. Students are encouraged to bring their laptops to all Friday sessions.

Week 1:

- Concepts Lecture:
    - Causal effects, Randomised trials and the Potential outcomes framework
    - Regression modelling with observational data vs. Randomised experiments.
    - Ordinary Least Squares (OLS) estimators of regression model parameters. Interpretation of the regression model.
    - Evaluating the model.
    - Regression parameter estimators as random variables with sampling distributions. Desirable properties of estimators: unbiasedness, efficiency, consistency.
    - Statistical Inference.
- Applications Lecture:
    - Introduction to Software: R, Stata.
    - Data exploration.
- Quiz 1.

Week 2:

- Concepts Lecture:
    - Partial effect interpretation of `slope' coefficients in Multiple Linear Regression. .
    - The probability model underlying Linear Regression.
    - Assumptions underpinning properties (of the sampling distributions) of OLS estimators: unbiasedness, efficiency, consistency.

- o The importance of Conditional Mean zero / independence assumption in estimating Causal effects: Regression vis-à-vis Randomized experiments.
  - o Inference: post-regression tests of hypotheses.
- Applications Lecture:
  - o Data description with summary statistics.
  - o Data visualization.
- Quiz 2.

Week 3:

- Concepts Lecture:
  - o Multicollinearity, perfect and imperfect.
  - o Variance Inflation, consequence, diagnosis.
  - o Model specification issues: Exclusion of relevant variables and omitted variable bias;
  - o Inclusion of irrelevant variables and imperfect multicollinearity.
  - o Tests of joint hypotheses to improve model specification: Chi-squared distribution, F distribution, Tests for overall significance of the regression.
  - o Tests of linear restrictions on model coefficients.
- Applications Lecture:
  - o Probability Distributions.
  - o Testing Hypotheses.
  - o Estimating Regression models.
- Quiz 3.

Week 4:

- Concepts Lecture:
  - o Linear variable transformations – data scaling, standardized coefficients.
  - o Nonlinear variable transformations.
    - Quadratic and polynomial specifications.
    - Logarithmic transformations and elasticities.
    - Box-Cox transformation.
    - Predicting Y when log(Y) is the dependent variable.
  - o Models with interactions - mediation and moderation.
- Applications Lecture:
  - o Estimating Multiple Linear Regression models.
  - o Diagnosing multicollinearity, solutions
  - o Robust estimation – Heteroscedasticity consistent std. errors.
  - o Joint test for multiple exclusions and linear restrictions
- Quiz 4.

Week 5:

- Concepts Lecture:
  - o Incorporating qualitative information in regression analysis.
  - o Dummy (e.g., binary) explanatory variables.
    - o Single categorical explanatory variable.
  - o Effect of changing the reference category.
  - o Two or more non-overlapping categorical explanatory variables.
  - o Introduction to fixed effects Panel data models.
  - o Two or more overlapping categorical explanatory variables.
  - o Interactions among dummy variables.

- o Slope dummy variables.
- o Advantages over running separate regressions.

- Applications Lecture:
  - o Quadratic and Polynomial models.
  - o Log transformed models.
  - o Models with Interaction terms.
  - o Box-Cox transformation.
  - o Component + Residual plots.
- Quiz 5.

Mid-term review quiz.

<u>Week 6:</u>

- Concepts Lecture:
  - o Testing OLS assumptions and remedies for violation.
  - o Maximum Likelihood Estimation and Large sample tests. Likelihood ratio, Wald, and Lagrange Multiplier tests. Test for functional form.
  - o Information criteria for Model selection.
  - o Test for functional form.
  - o Heteroskedasticity – Tests. Robust estimation. (Feasible) Generalized Least Squares estimation.
  - o Autocorrelation – Consequences. Tests.
  - o Tests for Normality.
- Applications Lecture:
  - o Intercept dummy variables.
  - o Creating dummy variables.
  - o Multiple overlapping dummy variables.
  - o Changing the reference category.
  - o Multiple non-overlapping dummy variables.
  - o Slope dummy variables.
  - o Testing joint hypotheses.
- Quiz 6.

<u>Week 7:</u>

- Concepts Lecture:
  - o Model specification tests continued.
  - o Unusual and outlying observations.
  - o Influential observations: Consequences, tests, remedies.
  - o Regression based prediction and confidence intervals for the dependent variable.
  - o Control variables for conditional mean independence.
- Applications Lecture:
  - o Test for multiple exclusions.
  - o Test for functional form misspecification.
  - o Test for heteroscedasticity.
  - o Test for serial correlation.
  - o Test for normality.
  - o Information criteria.
- Quiz 7

<u>Week 8:</u>

- Concepts Lecture:
    - Threats to Internal and External validity and endogeneity of an explanatory variable – omitted variables, measurement error in the dependent variable, missing data, nonrandom samples, simultaneous causality. Other deviations from iid errors.
    - Remedies for endogeneity: Introduction to Instrumental Variables estimation.
    - Carrying out an Empirical Project and reporting regression results in convincing detail.

- Applications Lecture:
    - Test for Outliers.
    - Measuring Leverage.
    - Identifying influential observations.
    - Estimating prediction and confidence intervals.

Test: 3-hour open book test on ? December (tbc)

Teaching methods

Concepts Lectures and Applications. Quizzes to aid self-instruction

Mark scheme

One 3-hour open book test after week 8 for 100 % of the total mark in the module.

Recommended texts

Stock J. and Watson M. (2019) or Stock J. and Watson M. (2015)	*Introduction to Econometrics.* 4th ed. Boston, Mass: Pearson 3rd ed., updated.	E-book via iDiscover E-book via iDiscover
Wooldridge, J.M. (2016)	*Introductory Econometrics: A Modern Approach*. 6th ed. Mason, Ohio: South Western College	E-book via iDiscover
Gábor Békés and Gábor Kézdi (2021)	*Data Analysis for Business, Economics, and Policy.* Cambridge University Press	