



THE BUSINESS SCHOOL

Econometrics I

**10635, 10972
(2003)**

Level: M

Semester: 2

Credits: 15

Lecturer: Dr. Keshab Bhattarai

Module Staff	Office No. and Location	Tel. Ext.	e.mail
Dr. Keshab Bhattarai	369 Wilberforce Building	6483	K.R.Bhattarai@hull.ac.uk

Econometrics I

Additional Notes

Introduction

Econometric analysis refers to systematic derivation and application of quantitative tools that economists use to test various hypotheses regarding cause and effect relationships between two or a set of random economic variables. These tools aim to derive the unknown parameters that minimise the errors of estimation and prediction about the unknown values of these variables. Econometric methods combine mathematics, statistics and probability theory to test various theories in economics.

Economic theory postulates quantitative relations among a set of economic variables. It is often expressed in terms of an economic model that may contain linear or non-linear relations between two variables, or a number of variables in one equation or in a system of equations. Economic theory only provides a qualitative statement about these variables; they may have positive or negative, linear or non-linear relations or no relation at all. Most of the time the theory is very imprecise about the numerical magnitude of these variables, which is very important in real life. Econometric techniques are designed to provide precise numerical estimates that are helpful either in prediction and forecasting of unknown values of these variables.

There are mainly two approaches to study econometrics. The classical approach assumes that values of random variables are derived from known distributions. These distributions are expressed in terms of standard parameters such as mean and variances or other moments of variable. Then econometric tests allow one to check whether a data series in sample are generated from such theoretical distributions. Most often these tests are based on normal distribution or other distributions such as t, F, Chi-square distributions that are derived from the normal distribution. Again there are cross section or time series methods depending upon the nature of observation. In contrast the Bayesian technique does not postulate any true theoretical distribution and uses sample information just to update priors that people have about the relationship among economic variables. Another distinction involves between parametric and non-parametric estimation. Parametric distributions assume underlying parameters defining a distribution whereas non-parametric methods do not have any underlying parameters to explain the generated series.

Econometrics is used in every branch of economics. Estimation of demand and supply for a particular product, or estimation of their price, income or cross price elasticities, or profits or shares of income are some examples from the micro economics. Estimation of consumption, investment, exports, imports or inflation, and unemployment, demand for money or money or income multipliers, estimation of determinants of economic growth in a cross section of countries over time are some examples from macro economics. Estimation of earnings, labour market participation or productivity or determinant of wage rates are some examples from the labour economics. Similarly trade economists would be interested in testing the terms of trade, determinants of real or nominal exchange rates. Estimation of revenue and cost function of a firm, or consumption or income functions of a household are other examples. Economic models may contain a single equation or a set of simultaneous equations. In summary we can say that econometric methods are applicable whenever there is a question of quantitative relationship between economic variables.

Econometric literature is vast and numerous theoretical as well as applied econometricians have contributed to it. It represents accumulated knowledge that has developed since the time of Newton Galton's, Bernoulli, Theil, Klein, Heckman, Hendry,

Peharan and Shin, Engle and Granger, Phillips, Davidson and McKinnon, Lancaster, Madala, Amemiya, Wallis among others have contributed significantly to the one or another branches of econometrics (see details in Handbook of Econometrics by Griliches and Intrilligator). There is a very strong tradition of econometrics studies in the UK as evidenced by the number of econometric journals edited by econometricians in the UK. It is ever growing field and it is impossible to list all contributors in a short pace.

The major objective of this module is to introduce students to the basic and standard econometric techniques that is often used for testing hypothesis following up the background knowledge that students have from the empirical methods or similar modules. It will derive and study properties of estimators in the classical linear regression analysis and analyse the problems caused by the violation of the classical regression technique such as the multicollinearity, heteroscedasticity, autocorrelation, diagnostic tests and remedial measures. It will then analyse simultaneity bias, identification problem, estimation of the system of equations.

Module Outline

The main aim of this module is to develop econometric techniques more formally and cover some advanced issues in multiple regression analysis including the detection, consequences and remedial measures to be taken in case the main assumptions behind a classical linear regression model are violated. It will build on Empirical Economics. A brief review of elementary matrix algebra and statistical inference will be provided to consolidate the background for the remaining lectures. After completing this module, students should be able to understand the logic behind econometric models and forecasts, of the sort discussed in applied economic journals and reports.

This module will have two main components. The first part will be on basic econometric methods. It will focus on derivation of properties of least square estimators in simple and multiple regression models more rigorously than done in the Empirical Economics. Some attention will be paid to application of these modelling techniques on economic analysis. A mid-term exam will be held after completing this part. The coverage on the second part, which aims to touch on additional topics in econometrics, will depend on progress made on the first part. The overall approach of this module will be to provide solid understanding of basic tools under discussion rather than taking up a new topic.

Aims

This module aims to consolidate understanding of standard econometrics techniques used in economic analysis. Teaching activities in this module will demonstrate

- (a) how to specify a econometric model to answer a economic question at hand
- (b) how to derive the standard estimators using econometric theory
- (c) how to estimate the above model using the sample data set and a standard software such as Shazam.
- (d) how select key macroeconomic variables and formulate a model for analysing cause and effect.
- (e) how to formulate an econometric model to explain the behaviour of the household, firm, government and the external sector in a given economy or over time.
- (f) how be able to diagnose a problem and analyse it critically.

- (g) how follow key articles in important econometric journals and be able to read more advanced text-books in the macroeconomics.

Learning Outcomes

After the successful completion of this module students should be able to

- (a) appreciate the major problems in the classical regression analysis and their consequences and remedial measures.
- (b) Set up the hypothesis to test an economic theory and use the available data set to test them.
- (c) be able to apply those estimates in writing standard economic reports based on empirical evidence in any aspect of economic analysis
- (d) do simple sensitivity tests based on model estimates.
- (e) Develop theoretical and empirical understanding of econometrics that is useful in following up more advanced text books or journals.
- (f) be familiar with the strengths and limitations of econometrics methods.

Skills

Module requires understanding of the theory, analytical approach using some basic mathematics, using the relevant cross section or time series for analysis. Students will learn simple matrix approach and constrained optimisation for derivation of best linear and unbiased estimators. They will be able to forecast, construct confidence intervals and test the reliability of parameters for use in economic analysis.

Significance of the Module

This module fits well with other modules in the economics B.A./ B.Sc.(Econ) degree programme. It is also useful to more students in the post-graduate programme who might have not done any econometrics before. It uses knowledge that students have from their previous semester.

The major aim of the module is to build confidence among students in using available data set in testing economic theories and be able to follow articles in various journal in economics and or be able to write professional reports based on empirical evidence..

Structure of Teaching & Learning Activities

The module will be delivered by two 50 minutes lecture sessions each week (20 lectures in total) and by 5 problem based tutorial sessions. Topics for lectures and tutorials are included at the end. Attendance on both lectures and classes is compulsory.

Assessment

The module will be assessed by following **three** components

Performance in this module will be assessed through: Assignments (20%), midterm examination (20%), and a final examination (60%). They contain theoretical as well as empirical parts.

Course work: Each week, starting from February 18th, there will be a lab based tutorial class to reinforce the theoretical explanations and proofs covered in the lectures. Attendance at the tutorial exercises is mandatory and you should make best efforts to do exercises assigned before you come to the class. These tutorials are designed to help you to prepare for assignments, the mid-term and final exams. Tentative list of lectures and tutorials is provided in the next page.

Assignments: Assignment for this module is given in a separate page at the end of this syllabus. Completed assignments should be submitted before the deadline or you will lose marks for that section of assignment except in extenuating circumstances.

Main text:

Hill, C.R., W.E.Griffiths and G.G. Judge (HGJ), Undergraduate Econometrics, 2nd edition, John Wiley and Sons Inc. New York, 2000.

Other useful texts (these can be found at HB.139 section in the 3rd floor of the library):

1. **Griffiths ,W.E., R.C. Hill and C. G. Judge (GHJ)** Learning and Practicing Econometrics, John Wiley and Sons Inc. New York, 1993.
2. **Gujarati (G), D. N.** (1995),Basic Econometrics, 3rd edition, McGraw Hill.
3. **Maddala (M) G. S.** (1992), Introduction to Econometrics, 2nd edition, Prentice Hall.
4. **Pindyck R.S. and Rubinfeld D.L.** (1998) Econometric Models and Econometric Forecasts, 4th edition, McGraw Hill.
5. **Greene, W. H.** (2000), Econometric Analysis, 4th edition, Macmillan.

Read one or two articles from one of the following journals to see how econometric results are discussed in the literature while reporting your results for the assignment: Economic Letters, Applied Economic Letters, The Economic Journal.

Submission date(s) for Coursework:

Essays are due by 5 PM on the Friday, 2003.

Assignment title(s):

- Essay should be based on one of the following topics and should contain a model and analysis and its application to the real world situation.

Assessment Criteria for Coursework:

Assessment criteria mentioned in pages 43-44 of the Undergraduate Handbook from the economics department applies to this essay.

Bear in mind that essays differ in style, motivation, originality, organisation, depth of analysis, clarity of presentation and facts and figures used to support the arguments. Essays that provide a good motivation, used one or two models (graphical and algebraic version as necessary) thoroughly to analyse the issue, present few empirical facts and relate them to the theory, provide a good conclusion and are original in their presentation may get first class marks. Essay that relies more on secondary rather than original thinking gets second class marks. Sloppy, incoherent and inconsistent presentation may lead to a low pass or a failing grade. Full citation and references should be provided whenever any idea is borrowed from

the literature. Copying without citation will be considered academic dishonesty and leads a failing grade (see below).

Submission of Coursework:

A Business School Cover Sheet must be attached to all Coursework and must be completed legibly and in full. Coursework must be submitted by the date and time stipulated in the essay box outside the school office.

Academic Dishonesty and Plagiarism

All work which is submitted for assessment must be your own work. Academic dishonesty is an attempt to engage in deception or fraud and will be penalised accordingly. It is a **very serious** offence.

* It is important that you have read and thoroughly understood the section entitled 'Plagiarism' in the *'University of Hull Business School Student Handbook'* and that you have read and understood the *'Code of Practice on the Use of Unfair Means'* which is published on the University of Hull website

Tentative Lecture Schedule and readings from the text book

Lecture #	Week	Day	Topics	Chap (HGJ)
1	1	Feb-03	Classical regression analysis –tests	3,5
2	1	Feb-03	BLUE Properties of OLS estimators	4,6
3	2	Feb-10	Statistical inference: review of t, F and chi-square distributions	2,5
4	2	Feb-10	Review of matrix for econometrics	5
5	3	Feb-17	Multiple Regression – derivations	7
6	3	Feb-17	Multiple Regression – tests and restrictions	8
7	4	Feb-24	Multicollinearity	8
8	4	Feb-24	Heteroskedasticity	11
9	5	Mar-03	Autocorrelation – causes and consequences	12
10	5	Mar-03	Autocorrelation –remedial measures	12
11	6	Mar-10	Non stationarity and spurious regression	16
12	6	Mar-10	Four ways of checking stationarity	16
13	7	Mar-17	Mid-term	
14	7	Mar-17	Solution of midterm	
15	8	Mar-24	Simultaneous Equation Models – identification	14
16	8	Mar-24	Simultaneous Equation Models –estimation	14
17	9	Apr-1	Pooling time-series and cross section data	17
18	9	Apr-1	Restrictions in multiple regression -in matrix form	8
19	10	Apr-28	Dummy variables and Large Sample Theory*	13, 15
20	10	Apr-28	Polynomial distributed lags*	9
21	12	May-12	Probability models*	18
22	12	May-12	Report writing using econometric analysis*	19

Note: I will provide my notes to each of the above topics but that should be complemented by readings as indicated above. The amount of coverage of topics after the mid-term will depend on progress made up to that time. Topics with * marks will be dropped if time is not enough for earlier topics.

Class Schedule and the end of the chapter exercises from the text

Class no.	Week	Topic	Homework exercises form HGJ
1	2	Hypothesis testing and confidence interval	2.9,2.22,5.1, 5.13
2	3	Proof of BLUE properties	4.13,4.15,4.6,4.17,4.9
		Preliminary Matrix Algebra	Class notes
3	4	Multiple regression –derivation and estimation	6.1-6.4, 6.15,6.19,6.20
4	5	Inference in multiple regression	7.4,7.7,7.6,7.10,7.11,
5	6	Multicollinearity and Heteroscedasticity	8.5,8.9,8.13,11.1,11.2,11.7 11.9
6	7	Autocorrelation	12.1,12.2,12.6,12.10,12.4
7	8	Simultaneous Equations system	14.5, 14.6
8	9	Stationarity and spurious regression	16.1,16.3,16.4,17.7,17.6
9	10	Dummy variables and large sample theory*	9.1,9.3,13.6
10	11	Polynomial distributed lags and probability models*	15.3, 15.6, 18.1,18.2,18.6

Problem 1
Simple and multiple regressions

1. Consider a simple supply function for a particular product.

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

Where Y_i is supply of that product and X_i is price received by suppliers it and the error term has a normal distribution with a zero mean and a constant variance as $e_i \sim N(0, \sigma^2)$.

(a) Suppose that you have a data set on quantity of supply and prices as given in the following table

Y_i	4	6	7	8	11	15	18	22
X_i	0.5	0.8	1.0	1.2	1.4	1.7	2.0	2.5

Use a pocket calculator to find $\sum X_i$, $\sum Y_i$, $\sum X_i Y_i$, $\sum X_i^2$ and $\sum Y_i^2$. Use these terms to find the estimates of β_1 and β_2 in the above regression line.

(b) Use estimated values b_1 and b_2 to predict the value of Y_i when $X_i = 3$

(c) Use estimate of b_2 to calculate the elasticity of supply around the mean values of Y_i and X_i .

(d) Calculate the variance of the error term. (hint: $\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - N\bar{Y}^2$)

(e) Use the variance of error calculated in (d) to construct a 95% confidence interval on b_2 and to judge significance and reliability of the parameter b_2 in this model?

2. Assume that expenditure in food (Y) depends upon total expenditure (X1) and food prices (X2). You have following observations on these three variables.

Y	3.5	4.5	5	6	7	9	8	10	12	14
X1	15	20	30	42	50	54	65	72	85	90
X2	16	13	10	7	7	5	4	3	3.5	2

- Construct a general least square model for this problem. State all underlying assumption behind such model.
- What are the expected signs of parameters and why?
- Find the least square normal regression equation of Y on X1 and X2.
- Estimate the parameters of the model. Comment on significance of t-test statistics.
- Construct the coefficient of multiple determination (R²) and standard errors of the estimated parameters and conduct test of significance.
- Construct 95 per cent confidence intervals for the population parameters.
- Find the explained and unexplained variations in.

Matrix for Econometrics

1. Find the determinant of the following matrix.

$$\text{a) } A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{b) } B = \begin{bmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{bmatrix}$$

2. Find the inverse of the following matrix

$$A = \begin{bmatrix} 4 & 1 & -1 \\ 0 & 3 & 2 \\ 3 & 0 & 7 \end{bmatrix}$$

3. Prove that following matrix is a positive definite matrix

$$A = \begin{bmatrix} 3 & 1 & -3 \\ -4 & 2 & 2 \\ 6 & -4 & 7 \end{bmatrix}$$

4. Solve following equations system using Cramer's rule

$$x_1 + 2x_2 + 2x_3 = 1 \quad (1)$$

$$2x_1 + 2x_2 + 3x_3 = 3 \quad (2)$$

$$x_1 - x_2 + 3x_3 = 5 \quad (3)$$

5. Find the characteristic root of the following matrix and calculate the characteristic vectors.

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 3 & -4 & 4 \end{bmatrix}$$

6. Define and provide examples of:

- Idempotent matrix
- Trace of a matrix
- Rank of matrix
- Kronecker products
- Jacobian matrix.

7. Check your hand calculations writing a Shazam Program to do above operations.

Problem 2

Use of Matrix in Regression Analysis

1. Consider a standard regression model of the following form

$$Y = X\beta + \varepsilon$$

where Y is $T \times 1$ vector of dependent variables, X is $T \times K$ matrix of explanatory variables, ε is $T \times 1$ vector of independently and identically distributed normal random variable with mean equal to zero and a constant variance, that is $\varepsilon \sim N(0, \sigma^2 I)$. β is a $K \times 1$ vector of unknown coefficients.

The idea of the ordinary least squares technique is to find an estimator for the unknown vector β by minimising the sum of error squares ($\varepsilon'\varepsilon$) using a sample data set on Y and X and from above $\varepsilon = Y - X\beta$.

- a) Show how you can derive the ordinary least square (OLS) estimator for β by minimising the sum of errors ($\varepsilon'\varepsilon$).
- b) Most often the variance-covariance matrix of the error term ($\sigma^2 I$) is unknown. Show how you could estimate this error variance, using your estimates in (a).
- c) How does an estimate of the variance of error term help to estimate the variance of estimated coefficients?
- d) What is the use of variances of the error term and of the estimated coefficients in testing a hypothesis of the form: $H_0: \beta = 0$ against $H_1: \beta \neq 0$? How can they be used to make a confidence interval on β ?
- e) Prove that the OLS estimator derived in (a) is the Best, Linear and Unbiased Estimator (BLUE).

2. Suppose that a leading supermarket in the city centre requests to estimate a demand function for beef. Your are considering estimating a model where demand for beef depends on price of beef, price of pork, price of chicken and consumer income as following:

$$D_t = \alpha_1 + \alpha_2 P_t^b + \alpha_3 P_t^p + \alpha_4 P_t^{ch} + \alpha_5 I_t + u_t$$

where D_t is demand for beef, P_t^b is price of beef, P_t^p is price of pork, P_t^{ch} is price of chicken, I_t is income of consumer and $u_t \sim N(0, \sigma^2)$ is a normally and identically distributed random variable.

- a) Using your knowledge of microeconomics, write down the expected signs of α_2 , α_3 , α_4 and α_5 in this model and explain why?
- b) Write major assumptions of the ordinary least square approach to this model.
- c) Suppose you have a data set on these variables over last 35 years and you want to estimate parameters α_1 , α_2 , α_3 , α_4 and α_5 . Derive normal equations that you will use get OLS estimators of these parameters?
- d) Compute the variances of parameters α_2 , α_3 , α_4 and α_5 .
- e) Compute variance-covariance matrix for the random term.
- f) Construct a confidence interval on α_2 , α_3 , α_4 and α_5 and D.
- g) How would your result be affected if you find that $P_t^b = 0.6P_t^p$?
- h) How would you modify your model to correct a problem in reported in (g)?

Problem 3
Confidence Interval of Prediction and Effect of Multicollinearity

1. One major use of an econometric model is prediction. Suppose that a local supermarket wants you to estimate a model that determines expenditure on food in terms of income, and to predict food demand next year. Consider a simple regression model of the following form:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

where Y_t is expenditure on food, X_t is income and u_t is independently and identically distributed random error term with a zero mean and a constant variance.

From the sample information on food expenditure and income contained in “food.dat” file find estimated values of β_0 , β_1 and σ^2 . Suppose that you want to predict the amount of expenditure on food y_0 next year using information on likely income next year, x_0 . You may safely assume that as before $u_0 \sim N(0, \sigma^2)$.

- a) Write down your prediction equation. Give an equation for the mean prediction, $E(y_0)$.
- b) What is your prediction of food expenditure if the income is £250? How can you compute your prediction error?
- c) What is the variance of prediction error?
- d) Construct a 95% confidence interval for your prediction. Explain what this interval means. How would you modify your model if the confidence interval of prediction is very large?
- e) Give a graphical explanation of your answers in (a)-(d), labelling your diagrams carefully.

2. Suppose that you are considering to study the relationship between aggregate consumption (C), and 3 components of income : wage income (W), nonwage-nonfarm income (P) and farm income (A) for an economy using a data set contained in income.dat. It can be expected that components of income move together – so multicollinearity may be a problem between the components on income. You specify the following regression equation:

$$C_t = \beta_0 + \beta_1 W_t + \beta_2 P_t + \beta_3 A_t + e_t$$

where $\beta_2 = 0.75\beta_1$ and $\beta_3 = 0.625\beta_1$

Estimate this equation with and without restrictions and then answer the following questions:

- a) what is the consequence of multicollinearity in this model?
- b) What is the effect of above restrictions on standard errors of estimates?
- c) Do your estimates support these restriction to avoid multicollinearity between the explanatory variables?

Problem 4 Autocorrelation and Heteroscedasticity

1. Suppose that you are estimating a log linear consumption function of the following form

$$\ln(C) = \alpha_0 + \alpha_1 \ln(Y) + \alpha_2 \ln(P) + e_i$$

where C , Y and P are consumption, income and prices and e_i is the random error term. Use information in *conyp.dat* to estimate unknown parameters α_0 , α_1 and α_2 and answer following questions using these results.

- What are the estimates of α_1 and α_2 ? Do these estimates have signs as you expected and why?
 - Does the Durbin-Watson Statistic show evidence of autocorrelation in the model? If so how does it affect the properties of the OLS estimators of α_1 and α_2 ?
 - What is the 95 and 90 percent of confidence interval estimate of α_1 and α_2 ?
 - How well does this model can explain variation in consumption? How do you decide overall fit of this model? What statistics do you use to decide at least there is one significant variable in the model?
 - What does the value of log likelihood function indicate? What is meaning of p-value of the estimates reported in the Shazam output?
2. Suppose that a monopolist's total revenue (tr), total cost (tc) and output is given by the following models

$$\text{Total revenue:} \quad tr = \beta_1 q + \beta_2 q^2 \quad (1)$$

$$\text{Total cost:} \quad tc = \alpha_1 + \alpha_2 q + \alpha_3 q^2 \quad (2)$$

where tr is total revenue, tc is the total cost, q is the output. Note that β_1 , β_2 , α_1 , α_2 and α_3 are unknown coefficients that you want to estimate. Use information in *mpoly.dat* to estimate those unknown coefficients.

- Find the profit maximising level of output as a function of unknown parameters β_1 , β_2 , α_1 , α_2 and α_3 in (1) and (2).
- Write down the least square estimates of the parameters in the total revenue and total cost functions.
- From the computer printout find out the profit maximising level of output deriving the profit maximising output using the first order condition.
- Is there any evidence of autocorrelation in any of these two equations? How do you detect this?
- Write down the generalised least square (GLS) estimates of parameters in the revenue and cost functions. Also show the autocorrelation coefficients used in GLS estimation?
- What is the profit maximising output with the generalised least square estimates?
- Use those estimates to make one period ahead forecast of the total cost, total revenue and the profit.

2. Suppose that you want to study the demand for money (M_t) as a function of the level of income (Y_t), interest rate (R_t), exchange rate (E_t), and wealth (W_t)

$$M_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 R_t + \gamma_4 E_t + \gamma_5 W_t + e_t$$

where e_t is a random error term with mean zero. The vector $e = \{e_t\}$ has variance-covariance matrix equal to:

$$Cov(e) = E(ee') = \begin{bmatrix} \text{var}(e_1) & \text{Cov}(e_1e_2) & \cdot & \cdot & \text{Cov}(e_1e_T) \\ \text{Cov}(e_1e_2) & \text{var}(e_2) & \cdot & \cdot & \text{Cov}(e_2e_T) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \text{Cov}(e_1e_T) & \text{Cov}(e_2e_T) & \cdot & \cdot & \text{var}(e_T) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \sigma_{1T} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \sigma_{33} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{T1} & \sigma_{T2} & \cdot & \cdot & \sigma_{TT} \end{bmatrix}$$

- What happens to the OLS estimators of $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ and γ_5 if the income and wealth are related to each other as $Y_t = 0.15W_t$? What would you do correct such a problem?
- The OLS method normally assumes that the condition $E(ee') = \sigma^2 I$ is satisfied. What does this imply for the values of σ_{ij} in the above matrix? Why may such an assumption not hold in many economic applications?
- Suppose that all off-diagonal elements are zero ($\sigma_{ij} = 0$ if $i \neq j$) in the above matrix but the diagonal elements are different from each other ($\sigma_{ij} \neq \sigma^2$ when $i = j$). What kind of problem is this? How does it affect the properties of the OLS estimators of $\gamma_2, \gamma_3, \gamma_4$ and γ_5 ?
- Now assume a variance-covariance matrix of the errors of the following form where diagonal elements are the same but the off-diagonal elements are different.

$$Cov(e) = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdot & \cdot & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \sigma^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{T1} & \sigma_{T2} & \cdot & \cdot & \sigma^2 \end{bmatrix}$$

What sorts of problem can you identify in this model? What kind of diagnostic test would you recommend? What values of this test would indicate the presence or absence of the problems you are testing for?

- Describe a generalised least square estimator and show how it can remove problems that you have encountered in (c) and (d)?

Problem 5 Lags and Functional Forms

1. Consider a production function of the following form:

$$Y_t = \beta_1 K_t^{\beta_2} L_t^{\beta_3} E_t^{\beta_4} M_t^{\beta_5} \exp\{e_t\}$$

where K is capital, L is labour, E is energy and M is other intermediate materials and $e_t \sim N(0, \sigma^2)$. Suppose you have an access to small data set with 25 observations on each of these variables.

- a. How can you apply OLS multiple regression technique to estimate unknown parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ in this model? What would be economic meaning of these parameters?
- b. From OLS regression run for the log-log specification on above variables one gets estimate of unknown parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ as presented in Table 1 below. Verify using information in "prod.dat" that you can get the following output also comment whether these estimates are economically and econometrically meaningful or not.

Table 1

Dependent variable ly							
VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO	PARTIAL CORR.		STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
			20 DF	P-VALUE			
LK	0.56070E-01	0.25927	0.21626	0.8310	0.0483	0.60541E-01	0.82304E-01
LL	0.22631	0.44269	0.51122	0.6148	0.1136	0.15352	0.13865
LE	0.43583E-01	0.38989	0.11178	0.9121	0.0250	0.48140E-01	0.48512E-01
LM	0.66962	0.36106	1.8546	0.0785	0.3831	0.71810	0.63733
CONSTANT	0.35163E-01	0.43932E-01	0.80040	0.4329	0.1762	0.0000	0.93210E-01
R-SQUARE = 0.9517 R-SQUARE ADJUSTED = 0.9421							
VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.34412E-02							
SUM OF SQUARED ERRORS-SSE= 0.68825E-01							
MEAN OF DEPENDENT VARIABLE = 0.37725							
LOG OF THE LIKELIHOOD FUNCTION = 38.2149							
ANALYSIS OF VARIANCE - FROM MEAN							
	SS	DF	MS	F			
REGRESSION	1.3565	4.	0.33913	98.550			
ERROR	0.68825E-01	20.	0.34412E-02	P-VALUE			
TOTAL	1.4254	24.	0.59390E-01	0.000			
CORRELATION MATRIX OF VARIABLES - 25 OBSERVATIONS							
LK	1.0000						
LL	0.94731	1.0000					
LE	0.98355	0.97159	1.0000				
LM	0.95875	0.98643	0.98271	1.0000			
	LK	LL	LE	LM			
VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS							
LK	0.67220E-01						
LL	0.44118E-02	0.19598					
LE	-0.79484E-01	-0.12991E-01	0.15202				
LM	0.12809E-01	-0.11339	-0.69033E-01	0.13036			
CONSTANT	-0.94659E-02	-0.15661E-02	0.79698E-02	0.12942E-02	0.19300E-02		
	LK	LL	LE	LM	CONSTANT		

Note: LK, LL, LE, LM are logs of K, L, E and M

c. After further consideration you wanted to modify the model as following.

$$\ln Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 L_{t-1} + \beta_3 M_{t-1} + \beta_4 K_{t-2} + \beta_5 M_{t-2} + \beta_6 E_{t-2} + \beta_7 K_{t-3} + e_t$$

Where subscript $t-i$ to a particular variable represents lag of i th order for that variable. Suppose that the OLS estimation now generates results as presented in Table 2. Again

verify that you can get this output. Comment on economic and econometric significance of this new model.

d. Build a 95% confidence interval for parameter β_7

e.

f. How could you test a restriction $\beta_4 + \beta_6 = 0$ in this model?

Table 2

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 17 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
LGY	-0.45589	0.15015	-3.0363	0.0075	-0.5930	-0.83272	-1.7097
LGL	-0.62831	0.17384	-3.6144	0.0021	-0.6592	-0.83696	-2.0243
LGM	1.0608	0.17239	6.1536	0.0000	0.8308	2.0049	3.9143
LG2K	-0.40452	0.22788	-1.7752	0.0938	-0.3954	-1.0097	-1.6958
LG2M	-0.36838	0.19086	-1.9301	0.0705	-0.4240	-0.78386	-1.2799
LG2E	0.83401	0.27783	3.0019	0.0080	0.5886	1.8572	3.0660
LG3K	0.19213	0.73210E-01	2.6244	0.0178	0.5370	0.51482	0.75396
CONSTANT	-0.93096E-02	0.52514E-01	-0.17728	0.8614	-0.0430	0.0000	-0.24678E-01

R-SQUARE = 0.9652 R-SQUARE ADJUSTED = 0.9509
F-value = 67.367
 VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.29174E-02
 SUM OF SQUARED ERRORS-SSE= 0.49596E-01
 MEAN OF DEPENDENT VARIABLE = 0.37725
 Note: LGY, LGK, LGL are lag of order of Y, K, L and M variables
 LG2K, LG2M, LG2E are second order lags of K, M and E variables and LG3K is third order lag of K

Problem 6

Validity of Restrictions in a Regression Model

1. Suppose that you are interested in estimating the demand for beer in Yorkshire pubs and consider the following multiple regression model:

$$\ln d_t = \beta_1 + \beta_2 \ln P_{bt} + \beta_3 \ln P_{Lt} + \beta_4 \ln P_{Rt} + \beta_5 \ln m_m + e_t$$

where d_t is the demand for beer, P_{bt} is the price of beer, P_{Lt} is the price of other liquor products, P_{Rt} is the price of food and other services, m_t is consumer income. $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are the set of unknown elasticity coefficients you would like to estimate. Again assume that errors are independently normally distributed, $e_t \sim N(0, \sigma^2)$.

- a) Estimate the unknown parameters of this model.
- b) How would you determine the overall significance of this model? Write down your test criterion. Compare that test statistic with another test statistic that you would use to test whether a particular coefficient, such as β_3 , is statistically significant or not.
- c) How would you establish whether a particular variable is helping to explain the variation in beer consumption?
- d) Further suppose that you have some non-sample information on the relation between the price and income coefficients as following:
 - sum of the elasticities equals zero: $\beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$
 - two cross elasticities are equal: $\beta_3 = \beta_4$ or $\beta_3 - \beta_4 = 0$
 - income elasticity is equal to unity: $\beta_5 = 1$

How do you test whether these restrictions are valid or not ?.

- e) In addition to the variables listed in the above model you suspect that gender and level of education of individuals are important determinants of beer consumption. Explain how you could incorporate these variables in this model.
 - f) The income of an individual also depends upon his/her age. Income in turn determines the consumption of beer. Thus age interacts with income. How would you introduce this age-income interaction effect in the above model?
2. A labour market study with a sample size of 5790 individuals explains variation in hours of work in terms a set of explanatory variables. These include ownership of house (HOUSE), non-labour income (NLI), marital status (MARRD), gender (SEX), alliance to a political party (CONS, LAB and LIB), importance given to health and money (HEALTH and MONEY) and presence of children at home (CHLDR), importance of having a job (JOB).
The major piece of result from the first set of estimate using Shazam is given in the following table 1. Verify that you can get this output using data in "hours.dat".
- a) explain the significance of various significance in this model.
 - b) The study presents another set of estimates following the above estimates as presented in Table 2. Compare and contrast results from Table 2 with those in Table 1. If you had to choose estimates between these two model results which one would you choose?
 - c) In order to improve the overall predictive power of the model the researcher still estimated another equation, by using one more variable LH which represents log hourly earning. The results of this model are given in Table 3. Are these results better than in Table 2 and Table 3? Give your reasons.

Table 1

R-SQUARE = 0.2587		R-SQUARE ADJUSTED = 0.2573	
ANALYSIS OF VARIANCE - FROM MEAN			
	SS	DF	MS
REGRESSION	0.31455E+06	11.	28596.
ERROR	0.90119E+06	5778.	155.97
TOTAL	0.12157E+07	5789.	210.01
			F
			183.342
			P-VALUE
			0.000
VARIABLE	ESTIMATED	STANDARD	T-RATIO
NAME	COEFFICIENT	ERROR	5778 DF
HOUSE	-0.33643	0.21990E-01	-15.299
NLI	-9.0724	0.79293	-11.442
MARRD	1.5640	0.35696	4.3815
SEX	9.3960	0.37980	24.740
CONS	1.8727	0.48295	3.8777
LAB	0.29375	0.40138	0.73184
LIB	0.56726	0.58512	0.96947
HEALTH	0.17540	0.16362	1.0720
MONEY	-0.11243	0.83475E-01	-1.3468
CHLDR	-0.26215	0.61933E-01	-4.2328
JOB	0.61733	0.94010E-01	6.5667
CONSTANT	31.492	1.7200	18.309
			P-VALUE CORR.
			0.0000-0.1973
			-0.20702
			0.0000-0.1488
			-0.13157
			0.0000 0.0575
			0.53799E-01
			0.22278E-01
			0.0000 0.3095
			0.32417
			0.12201
			0.0001 0.0509
			0.50979E-01
			0.95326E-02
			0.4643 0.0096
			0.99099E-02
			0.30642E-02
			0.3323 0.0128
			0.12165E-01
			0.16223E-02
			0.2837 0.0141
			0.12353E-01
			0.44175E-01
			0.1781-0.0177
			-0.15706E-01
			-0.19203E-01
			0.0000-0.0556
			-0.53495E-01
			-0.50425E-01
			0.0000 0.0861
			0.78766E-01
			0.13269
			0.0000 0.2342
			0.0000
			0.83167

Table 2

R-SQUARE = 0.2582		R-SQUARE ADJUSTED = 0.2573	
ANALYSIS OF VARIANCE - FROM MEAN			
	SS	DF	MS
REGRESSION	0.31392E+06	7.	44846.
ERROR	0.90183E+06	5782.	155.97
TOTAL	0.12157E+07	5789.	210.01
			F
			287.528
			P-VALUE
			0.000
VARIABLE	ESTIMATED	STANDARD	T-RATIO
NAME	COEFFICIENT	ERROR	5782 DF
HOUSE	-0.33578	0.21967E-01	-15.285
NLI	-8.9781	0.79123	-11.347
MARRD	1.6280	0.35515	4.5842
SEX	9.3394	0.37801	24.707
CONS	1.6293	0.41759	3.9017
CHLDR	-0.25687	0.61699E-01	-4.1633
JOB	0.60867	0.91478E-01	6.6537
CONSTANT	32.674	0.83165	39.288
			P-VALUE CORR.
			0.0000-0.1971
			-0.20663
			-0.13020
			-0.20718E-01
			0.0000 0.0602
			0.56002E-01
			0.23190E-01
			0.0000 0.3090
			0.32221
			0.12128
			0.0001 0.0512
			0.44353E-01
			0.82935E-02
			0.0000-0.0547
			-0.52418E-01
			-0.49410E-01
			0.0000 0.0872
			0.77660E-01
			0.13082
			0.0000 0.4590
			0.0000
			0.86287

Table 3

5790 OBSERVATIONS		DEPENDENT VARIABLE= HW	
R-SQUARE = 0.8572		R-SQUARE ADJUSTED = 0.8571	
ANALYSIS OF VARIANCE - FROM MEAN			
	SS	DF	MS
REGRESSION	0.10421E+07	4.	0.26053E+06
ERROR	0.17363E+06	5785.	30.014
TOTAL	0.12157E+07	5789.	210.01
			F
			8680.311
			P-VALUE
			0.000
VARIABLE	ESTIMATED	STANDARD	T-RATIO
NAME	COEFFICIENT	ERROR	5785 DF
LH	54.021	0.33968	159.04
HOUSE	-0.47294E-01	0.94036E-02	-5.0294
SEX	2.4138	0.17049	14.158
CONS	0.50673	0.18271	2.7735
CONSTANT	-45.713	0.53684	-85.152
			P-VALUE CORR.
			0.0000 0.9021
			0.87785
			2.1840
			0.0000-0.0660
			-0.29103E-01
			-0.10751E-01
			0.0000 0.1830
			0.83277E-01
			0.31344E-01
			0.0056 0.0364
			0.13794E-01
			0.25793E-02
			0.0000-0.7458
			0.0000
			-1.2072

d) One of the above estimates had used restriction in estimation as following:

Table 4

_test	
_test lib =0	
_test lab=0	
_test cons=0	
_end	
F STATISTIC = 6.2212663	WITH 3 AND 5777 D.F. P-VALUE= 0.00033
WALD CHI-SQUARE STATISTIC = 18.663799	WITH 3 D.F. P-VALUE= 0.00032
UPPER BOUND ON P-VALUE BY CHEBYCHEV INEQUALITY = 0.16074	

Comment on statistical significance of restriction using this result.

e) Comment on the value of F-statistics in all above models.

Problem 7

Estimation, Confidence Interval and Diagnostic Checks in a Simple Regression

Test yourself for a test –try to finish this exercise within 1 and half hours.

1. Suppose you have the following data set on number of tickets sold in a football match (Y in 10 thousands), price of tickets (X_1) and income of the customers (X_2 in 10 thousand pounds) as given in the following table. You want to find out the exact relation between tickets sold and prices and income of people watching football games.

Full marks [155]

Observations	Y	X_1	X_2	YX_1	YX_2	X_1X_2	X_1^2	X_2^2	Y^2
n1	1	11	2	11	2	22	121	4	1
n2	2	7	2	14	4	14	49	4	4
n3	3	6	4	18	12	24	36	16	9
n4	4	5	5	20	20	25	25	25	16
n5	5	3	6	15	30	18	9	36	25
n6	6	2	5	12	30	10	4	25	36
n7	7	1	4	7	28	4	1	16	49
Total	28	35	28	97	126	117	245	126	140

- Answer all questions from the part A and any five questions from the Part B.

PART A

- (a) Write a simple regression model to explain the number of tickets sold in terms of the price of the ticket. Explain briefly underlying assumptions of your model. What are the expected signs of the parameters in this model?

[5]

- (b) Estimate the slope and intercept parameters. Use cross products and squared terms provided for you in the above table.

[10]

- (c) Using your estimates in (b) find the explained squared sum $\sum_i \hat{Y}_i^2$, $\sum_i \hat{e}_i^2$ and the R^2 and \bar{R}^2 for this model.

[10]

- (d) Estimate the variance of the error term and the slope coefficient in this simple model.

[10]

- (e) Test whether the slope term is significant at 5% confidence level.

[10]

- (f) Write a multiple regression model to explain the number of tickets sold in terms of the price of the ticket and the income of individuals going to the football game. What additional assumption(s) do you need when you introduce one more variable.

[5]

- (g) Estimate the parameters of the multiple regression model given in (f).

[10]

- (h) What is your prediction of the number of tickets sold if $X_1 = 4$ and $X_2 = 3$?

[5]

- (i) Introduce dummy variables in your multiple regression model to show differences in demand for football ticket based on gender differences (1 for male and 0 for females),

four seasons (autumn, winter, spring and summer) and interaction between gender and income.

PART B

- (j) You want to introduce taste for the game as a new variable. Suppose that the variable giving the taste for football is exactly correlated with the income levels of individuals. Show how the OLS procedure breaks down if you regressed Y_i on X_2 and the taste variable (X_3) if the relation between X_3 and X_2 is given by $X_3 = 0.1X_2$.

[10]

- (k) It is likely that people with higher income go to football games more often than low income people. Demonstrate how estimates are unbiased but inefficient in presence of a heteroscedasticity in a simple regression of Y_i on X_1 and X_2 .

[10]

- (l) Suppose that the football game model that you are using here has an autoregressive error term that follows an AR(1) process. Why may the error term be correlated like this? Show how Durbin-Watson statistics could be used to detect autocorrelation among the error terms.

[10]

- (m) If the variance of the error terms is not constant show how one can apply GLS procedure to remove autocorrelation and heteroscedasticity problem from a regression model.

[10]

- (n) Suppose you would like to test a restriction that prices and income are not significant determinants of demand for football tickets. People go to games randomly. How can you test such restriction in the model discussed above?

[10]

- (o) A significant proportion of people who are going to a football game this year might also have gone to football games in past years. Show how could you construct an ARDL(2,2) model to explain demand for tickets for a football game during this year in terms of prices and income.

[10]

- (p) Suppose that the demand for tickets, Y and income X_2 are non-stationary I(1) variables but the price X_1 is a stationary or I(0) variable. What would happen if you apply OLS in your model above? Explain briefly how you can correct this problem.

[10]

- (q) Further suppose that demand for tickets depends on price of the ticket and price of the ticket depends on the demand for the tickets. There is an endogenous relation between quantity demanded and prices. How would this affect the efficiency of the estimated OLS parameters in your model?

[10]

Problem 8
Simultaneous Equation Models

1. Suppose that you have a simple model of consumption and income as following

Consumption function: $C_t = \beta_0 + \beta_1 Y_t + e_t$ (1)

National income identity: $Y_t = C_t + I_t$ (2)

- (a) Use rank and order conditions to find whether the consumption function is identified in this model.
- (b) Write a reduced form for this system. Show how you could retrieve the structural coefficients β_0 and β_1 if you applied OLS to this reduced form.
- (c) Show that application of OLS to (1) generates a biased estimate of β_1 .
- (d) What other method would you recommend to get an unbiased and best estimator for this model? Write steps to be followed until you get the structural coefficients β_0 and β_1 .
- (e) Write a short note on how this model could be used to make a historical simulation of consumption and income series.

2 Consider a market model for a particular product.

Demand: $Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t}$ (1)

Supply: $Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t}$ (2)

Here Q_t^d is quantity demanded and Q_t^s is quantity supplied, P_t is the price of commodity, P_{t-1} is price lagged by one period, I_t is income of an individual, $u_{1,t}$ and $u_{2,t}$ are independently and identically distributed (iid) error terms with a zero mean and a constant variance. Q_t and P_t are endogenous variables and P_{t-1} and I_t are exogenous variables, $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$ and β_2 are six parameters defining the system.

- (a) How can simultaneity bias occur if one tries to apply OLS to each of the above equations.
- (b) Use rank and order conditions to judge whether each of these two equations are over-, under- or exactly identified.
- (c) Write down the reduced form for this system.
- (d) How would you estimate the coefficients of the reduced form equations? Write down the estimator.

If equations are identified explain how you may retrieve the structural parameters, $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$ and β_2 from the coefficients of the reduced form equations.

Problem 9

Lags and Long-run Multipliers and Large Sample Tests in a Regression Model

1. Take a distributed lag model in which consumption in period t depends on current and past incomes as following:

$$C_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$$

where C_t is current consumption and X_{t-I} is income in period $t-I$, and, u_t is a normally distributed random disturbance term. Suppose from earlier estimates you have some information on parameters such as $\beta_0 = 5000$, $\beta_1 = 0.4$, $\beta_2 = 0.3$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$.

- a) What would be the immediate (short run) impact on consumption (C_t) of an increase in income (X_t) equal to £2000? What would be the intermediate impact on consumption after period two? What would be the long-run impact (after four periods) of this initial increase of income in this model?
- b) Initially you do not know how many lags of income are significant in determining current consumption. What is an ad-hoc procedure for determining an appropriate number of lags in such models? What are the potential drawbacks of estimating a distributed lag model by such procedure?
- c) Instead of assuming a four period lag model you prefer to use a general infinitely distributed lag model, such as:

$$C_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 X_{t-3} + \dots + \beta_k X_{t-k} + \dots + u_t$$

What is an apparent problem of estimating a distributed model like this?

- d) Further suppose that you follow Koyck's procedure $\beta_2 = \lambda\beta_1$, $\beta_3 = \lambda^2\beta_1$, ..., $\beta_k = \lambda^k\beta_1$ and so on. Show how the Koyck procedure helps to simplify the above model.
 - e) Why would you not want apply OLS in estimating an autoregressive model? How could an instrumental variable be used in place of an autoregressive term?
2. HGJ (14.10) In an attempt to explain annual income of a cross section of 45 people. A research a set up the model:

$$y_t = \beta_1 + \beta_2 a_t + \beta_3 g_t + \beta_4 m_t + e_t$$

where y is income (in thousand dollars), a is age, g is gender (0 for male and 1 for female), and m is a measure of the ability of the t th individual. Data on these variables, as well as on the number of years of schooling after high school s are given in table largesamp.dat.

- a) use least squares to estimate β_1 , β_2 , β_3 , β_4 .
- b) Use **Wald**, **LM**, **LR**, and **F** tests to test (i) the hypothesis that the gender does not influence income, and (ii) the hypothesis that gender and ability do not influence income.
- c) Using s as an instrument for m , find the instrumental variables estimate β_{iv} for β . Comment on the results relative to those in part (a)
- d) Using b and β_{iv} , test whether contemporaneous correlation between the regressors and the error exists.
- e) Suggest and carry through a Wald test that uses the instrumental variables estimator to test whether gender and ability do not influence income.

Problem 10

A Limited Dependent Variable (Binary Choice) Model

1. A university teacher is interested to estimate a model that explains whether a certain student gets a first class marks in an exam. He/she uses observations on limited dependent variable Y_i which takes value 1 ($Y_i = 1$) if a student gets a first class mark, value 0 ($Y_i = 0$) otherwise. Probability of getting a first class mark in an exam is a function of student effort index denoted by Z_i

$$Z_i = \beta_1 + \beta_2 H + \beta_3 E + \beta_4 A + \beta_5 P + e_t$$

where H = hours of study; E = hours spent on exercises, A = number of right answers you were able to write in exams, and P = number of pages of papers written. e_t is an iid random error term.

In a logit model the probability of getting a first class marks in an exam depends nonlinearly upon Z_i

as $P_i(Y_i = 1) = \frac{1}{1 + e^{-Z_i}}$, where $0 \leq P_i(Y_i = 1) \leq 1$. The log of the ratio of odds

$\left[\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \right]$ is a linear function of the index Z_i

$$\ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 H + \beta_3 E + \beta_4 A + \beta_5 P + e_t$$

- Explain how you can test the contribution of any one of the explanatory variables to the probability of getting a first class mark in the exam using this logit model.
- Discuss the estimation method that you will apply to estimate the parameters of a non-linear function with a limited dependent variable like this.
- A probit model, like a logit model, gives a probability of getting a first class marks as a function of an index of variables, Z_i . Its functional form takes a normal distribution such as

$$\Pr(Y_i = 1) = \Pr(Z_i^* \leq Z_i) = F(Z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} e^{-\frac{t^2}{2}} dt$$

where $t \sim N(0,1)$ is a normally distributed variable.

Give a graphical explanation of how a variable included in index Z_i would affect the probability of getting a first class mark.

- How would you test whether a particular variable such as hours spent on study is statistically significant or not in logit and probit models?
- Compare and contrast probit and logit models to a linear probability model in explaining probability of occurrence of an event as above?

Homework for Undergraduate Programme

A List of Recommended Exercises from the Text Book

I would like to recommend you following exercises from the end of the chapter exercises of the main text book **Undergraduate Econometrics** by Hill-Griffith and Carter (HGJ), 2nd edition, John Wiley and Sons, 2001. They are not compulsory and you will not be penalised for not doing them. If you tried them on your own and could not solve any of them answers can be discussed in tutorial sessions or through the blackboard.

Tutorial 1: Hypothesis testing and confidence interval

Exercises 2.9, 2.22, 5.1, 5.4, 5.13

Tutorial 2: Proof of BLUE properties

Exercises 4.6, 4.13, 4.15, 4.16, 4.17, 4.19

Tutorial 3: Multiple regression -derivation and estimation

Exercises 6.1, 6.2, 6.3, 6.4, 6.15, 6.20, 6.19

Tutorial 4: Inference in multiple regression

Exercises 7.4, 7.6, 7.7, 7.10, 7.11

Tutorial 5: Multicollinearity and Heteroscedasticity

Exercises 8.5, 8.9, 8.13, 11.1, 11.2, 11.7, 11.9

Tutorial 6: Autocorrelation

Exercises 12.1, 12.2, 12.4, 12.6, 12.10

Tutorial 7: Stationarity and spurious regression

Exercises 16.1, 16.3, 16.4, 17.6, 17.7

Tutorial 8: Simultaneous Equations system

Exercises 14.5 and 14.6 with the UK data

Tutorial 9: Dummy variables and large sample theory

Exercises 9.1, 9.3, 13.6,

Tutorial 10 Polynomial distributed lags and probability models

Exercises 15.3, 15.6, 18.1, 18.2, 18.

Homework for Post-graduate Programme
A List of Recommended Exercises from the Text Book

Useful links to web pages:

Links to economic journals, software and datasets relevant to this module are listed in the Blackboard for this module. All announcements for this module will be placed in the Blackboard. I would appreciate it if you give me your feedback on any aspect of the module also through Blackboard. There is a folder with all past module outlines, exercises and solutions in S:\econometrics1 if you would like to browse through materials covered last year. It also has a **shortcut** to “Shazam”. Students who do not have an “ec” username must contact me for access to this folder. **Manuals** for Shazam programme is available from the short loan section of the library.

Study Programme

The main contents of this module include formal analysis of econometric methods students will be provided standard sets of lecture handouts which should be complemented by reading text-books or journal articles listed below.

Journal List

The following Journals are appropriate to the module and will contain further articles which you may find helpful. You will find following journal in the current periodical section of library.

American Economic Review, Journal Political Economy, Journal of Economic Perspective, Journal Macroeconomics, Journal of Monetary Economics, Economic Journal, Journal of Money Credit and Banking, Economic Letters, Applied Economics, Applied Economics Letters International Economic Review, Journal of Policy Modelling.

WWW Sites

You will find these Web sites helpful for the module and are advised to consult them regularly. You may discover other Websites which are also helpful.

HM-Treasury UK <http://www.hm-treasury.gov.uk> ;

Bank of England <http://www.bankofengland.co.uk>

European Central Bank <http://www.ecb.int/>

International monetary fund: <http://www.imf.org>

Other Sources:

For example: Econlit JSTOR databases through the electronic services in the library should be useful for the literature search. Financial Times has plenty of day to day news and story about the macroeconomic events in real life.

The range of references and resources available throughout the University Library is increasing constantly on a daily basis. The list above should be thought of as an opening into the literature. You are strongly encouraged to browse through the stock and to pay particular attention to the New Periodicals shelves.

Programme Changes

Wherever possible, the programme timetables and content will be delivered as outlined in the Module Handbook. However, at times changes do have to be made but in the event of such changes occurring every effort will be made to re-schedule the activity, or replace it with work of an equivalent nature.

Attendance

The University has an Attendance Policy, which requires all students to attend all timetabled sessions for their programme of study. An Attendance Register will be kept for tutorial sessions and students with unauthorised absence will be subject to School and University disciplinary procedures. You are reminded that unauthorised absence may affect your course progress and LEA grant entitlement.

It is important that you have read and understood the section *entitled 'General Attendance'* in the Hull University Business School Student Handbook.

Student Support

All modules are supported by tutorial assistance. A list of Module Staff and their office numbers, telephone extensions and E.mail addresses is available on the inside front cover of this document. A specified time for Tutorials for this module can be seen in the Study Programme. Appointments may sometimes be made at other times but you are advised to contact your tutor to arrange an appointment. Arrange your tutorial by signing the Tutorial Sheet on your tutor's office door. This support is specifically for assistance with a named module upon which the tutor teaches.

Enquiries of a general nature, which may range over a number of modules, should be addressed to the Programme Leader of your registered degree.

Health and Safety

You are responsible for your own health and safety at all times. It is vitally important that you act sensibly and safely for both indoor and outdoor activities. You are required to follow all safety instructions and guidelines as laid down in your *'University of Hull Business School Student Handbook'*.

Module Evaluation

The module will be evaluated by means of the Business School Module Evaluation Questionnaire (MEQ), which all students are required to complete at the end of the module. The results of this formal evaluation will be forwarded to Student-Staff Committees and to Undergraduate Committee and will be used to make alterations and improvements to the delivery of the module next year if these are deemed to be necessary. Additional module evaluation techniques may also be employed. Issues concerning the module can be forwarded to the School Student-Staff Committee. You may also have the opportunity to make informal comments and suggestions concerning the module in tutorial sessions.

The Module Staff hope that you enjoy studying this module and that it makes a valuable educational contribution to your Course Programme.

THE UNIVERSITY OF HULL

Department of Economics

Level M Examination

May/June, 2002

ECONOMETRICS 1

May/June, 2 hours

Answer question **one** and any **two** of the remaining questions. All questions carry equal weight.

Do not open or turn over this exam paper, or start to write anything until told to by the Invigilator. Starting to write before being permitted to do so may be seen as an attempt to use Unfair Means.

3. Consider a regression model of the form,

$$Y = X\beta + \varepsilon$$

where Y is $T \times 1$ vector of dependent variables, X is $T \times K$ matrix of explanatory variables, ε is $T \times 1$ vector of independently and identically distributed Normal random variables with mean zero and constant variance, $\varepsilon \sim N(0, \sigma^2 I)$. β is a $K \times 1$ vector of unknown coefficients.

The idea of the ordinary least squares technique is to find an estimator for the unknown vector β by minimising the sum of error squares ($\varepsilon'\varepsilon$) using a sample data set of Y and X where, from above, $\varepsilon = Y - X\beta$.

- f) Derive the ordinary least squares (OLS) estimator for β by minimising the sum of errors ($\varepsilon'\varepsilon$).
- g) The variance-covariance matrix of the error term ($\sigma^2 I$) is typically unknown. Derive an expression for the estimator of the error variance, using the result obtained in (a).
- h) Using the estimator of the variance of error term, determine the variance of the estimated coefficients.
- i) Show how the variance of the error term and the variance of the estimated coefficients can be used;
 - (i) in testing a hypothesis of the form: $H_0: \beta = 0$ against $H_1: \beta \neq 0$.
 - (ii) to construct a confidence interval for β ?
- j) Prove that the OLS estimator derived in (a) is the Best, Linear and Unbiased Estimator (BLUE).

(2) Consider the following model relating aggregate consumption and national income,

$$\text{Consumption function:} \quad C_t = \beta_0 + \beta_1 Y_t + e_t \quad (1)$$

$$\text{National income identity:} \quad Y_t = C_t + I_t \quad (2)$$

- (f) Use the rank and order conditions to determine whether the consumption function is identified in this model.
 - (g) Write a reduced form for this system. Show how you could retrieve the structural coefficients β_0 and β_1 from OLS estimates of the reduced form model.
 - (h) Show that the estimation of equation(1) using OLS generates a biased estimate of β_1 .
 - (i) What other method could you use to obtain an unbiased and best estimator for this model? Outline the steps to be followed in order to obtain the structural coefficients β_0 and β_1 .
 - (j) Write a short note explaining how this model could be used to produce a historical simulation of the consumption and income series.
- (3) The following data set records the number of tickets sold for a football match (Y , in 10 thousands), the price of tickets (X_1) and income of the customers (X_2 , in 10 thousand pounds).

Observations	Y	X_1	X_2	YX_1	YX_2	X_1X_2	X_1^2	X_2^2	Y^2
n1	1	11	2	11	2	22	121	4	1
n2	2	7	2	14	4	14	49	4	4
n3	3	6	4	18	12	24	36	16	9
n4	4	5	5	20	20	25	25	25	16
n5	5	3	6	15	30	18	9	36	25
n6	6	2	5	12	30	10	4	25	36
n7	7	1	4	7	28	4	1	16	49
Total	28	35	28	97	126	117	245	126	140

An investigator wants to establish the exact relation between tickets sold, prices and the income of people watching football games.

- (a) Write a simple regression model explaining the number of tickets sold in terms of the price of the ticket. Explain, briefly, the underlying assumptions of your model. What are the expected signs of the parameters in this model?
- (b) Using the cross products and squared terms provided for you in the above table, estimate the slope and intercept parameters.
- (c) Using your estimates in (b), find the explained sum of squares $\sum_i \hat{Y}_i^2$, $\sum_i \hat{e}_i^2$, R^2 and \bar{R}^2 for this model.

- (d) Estimate the variance of the error term and the variance of the slope coefficient in this model.
- (e) Test whether the slope term is statistically significant at 5% significance level.
- (f) Write a multiple regression model to explain the number of tickets sold in terms of the price of the ticket and the income of individuals going to the football game. What additional assumption(s) do you need to make when introducing an additional variable?
- (g) Estimate the parameters of the multiple regression model given in (f).
- (h) What is your prediction of the number of tickets sold if $X_1 = 5$ and $X_2 = 4$?
- (i) Add dummy variables to your multiple regression model to determine the difference in the demand for football ticket associated with variables representing gender (1 for male and 0 for females), the four seasons (autumn, winter, spring and summer) and the interaction between gender and income.

- (4) Provide proofs or illustrations of any **four** of the following
- (a) Show that the OLS procedure breaks down if one regresses Y on X_1 and X_2 when there is exact collinearity between X_1 and X_2 .
 - (b) Demonstrate that estimates are unbiased, but inefficient, in the presence of heteroscedasticity in the error structure of a simple regression of Y_i on X_i .
 - (c) Consider a model in which the error term follows an AR(1) process. Why might the error term have this characteristic? Show that the Durbin-Watson statistic can be used to detect autocorrelation among the error terms.
 - (d) Show that if the variance of the error terms is not constant, a GLS procedure can be used to remove the autocorrelation and heteroscedasticity from a regression model.
 - (e) How can a unit-root test be used to check whether a particular series is stationary or non-stationary?
 - (f) Consider a simple wage-determination model relating the hourly wage rate and a number of explanatory variables, including years of schooling, age, age squared and on the job training.

Show that dummy variables can be used to capture;

- (a) differences across individuals with 4 different categories of education (no high school, high school, college and university)

and

- (b) the interaction between age and on the job training.

How would you determine the statistical significance of university education in explaining the hourly wage rate received by an individual?

- (5) A version of the Phillip's curve says that the rate of change of money wages is a function of the reciprocal of the unemployment rate.

$$\dot{w}_t = \beta_1 + \beta_2 \left(\frac{1}{u_t} \right) + e_t$$

Consider the following data set for a hypothetical economy:

Wage rate and unemployment rate from 1983 to 2000

Year	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Wage rate	1.62	1.65	1.79	1.94	2.03	2.12	2.26	2.44	2.57	2.66	2.73	2.8	2.92	3.02	3.13	3.28	3.43	3.58
Unemployment rate	1	1.4	1.1	1.5	1.5	1.2	1	1.1	1.3	1.8	1.9	1.5	1.4	1.8	2.1	1.5	1.3	1.4

- (a) From the computer output provided in the Appendix, test whether there is any relationship between \dot{w} and $\frac{1}{u}$.
- (b) What would be the change in the money wage rate if the unemployment rate were 1.7?
- (c) Using the results, estimate the natural rate of unemployment. (Remember that natural rate of unemployment corresponds to $\dot{w} = 0$.)
- (d) What is the economic meaning of β_1 ?
- (e) When does a change in the unemployment rate have the greatest impact on the rate of change in wages? When does it have the smallest impact?

Regression Results for the rate of change of wages and unemployment rate from 1983 to 2000

```

|_file l1 table8.6
UNIT 11 IS NOW ASSIGNED TO: table8.6
|_sample 1 18
|_read(l1) w u
  2 VARIABLES AND          18 OBSERVATIONS STARTING AT OBS      1

|_genr w1=lag(w)
..NOTE.LAG VALUE IN UNDEFINED OBSERVATIONS SET TO ZERO
|_sample 2 18
|_genr invu=1/u
|_genr wdot=(w-w1)/w1*100
|_genr one=1
1
|_ols wdot one invu / coef=b noconstant

REQUIRED MEMORY IS PAR=      5 CURRENT PAR=    500
OLS ESTIMATION
  17 OBSERVATIONS      DEPENDENT VARIABLE = WDOT
...NOTE..SAMPLE RANGE SET TO:      2,      18
...WARNING...VARIABLE ONE      IS A CONSTANT

R-SQUARE =    0.3926      R-SQUARE ADJUSTED =    0.3521
VARIANCE OF THE ESTIMATE-SIGMA**2 =    2.5877
STANDARD ERROR OF THE ESTIMATE-SIGMA =    1.6086
SUM OF SQUARED ERRORS-SSE=    38.815
MEAN OF DEPENDENT VARIABLE =    4.7926
LOG OF THE LIKELIHOOD FUNCTION = -31.1396
RAW MOMENT R-SQUARE =    0.9146

MODEL SELECTION TESTS - SEE JUDGE ET AL. (1985,P.242)
AKAIKE (1969) FINAL PREDICTION ERROR - FPE =    2.8921
(FPE IS ALSO KNOWN AS AMEMIYA PREDICTION CRITERION - PC)
AKAIKE (1973) INFORMATION CRITERION - LOG AIC =    1.0609
SCHWARZ (1978) CRITERION - LOG SC =    1.1589
MODEL SELECTION TESTS - SEE RAMANATHAN (1992,P.167)
CRAVEN-WAHBA (1979)
GENERALIZED CROSS VALIDATION - GCV =    2.9327
HANNAN AND QUINN (1979) CRITERION =    2.9173
RICE (1984) CRITERION =    2.9858
SHIBATA (1981) CRITERION =    2.8205
SCHWARZ (1978) CRITERION - SC =    3.1865
AKAIKE (1974) INFORMATION CRITERION - AIC =    2.8890

              ANALYSIS OF VARIANCE - FROM ZERO
              SS          DF          MS          F
REGRESSION    415.57         2.         207.78         80.297
ERROR         38.815        15.         2.5877         P-VALUE
TOTAL         454.38        17.         26.728         0.000

VARIABLE      ESTIMATED  STANDARD  T-RATIO      PARTIAL STANDARDIZED ELASTICITY
NAME          COEFFICIENT  ERROR    15 DF      P-VALUE CORR. COEFFICIENT AT MEANS
ONE          -1.4284     2.036   -0.7017    0.494-0.178    0.0000   -0.2980
INVU         8.7309     2.804    3.114     0.007 0.627    0.6266    1.2980
1
|_confid one invu / nofplot
USING 95% AND 90% CONFIDENCE INTERVALS

CONFIDENCE INTERVALS BASED ON T-DISTRIBUTION WITH 15 D.F.
- T CRITICAL VALUES =    2.131 AND    1.753
NAME  LOWER 2.5%  LOWER 5%  COEFFICIENT  UPPER 5%  UPPER 2.5%  STD. ERROR
ONE   -5.766     -4.997   -1.4284     2.140     2.909       2.036
INVU  2.756      3.816    8.7309     13.65    14.71       2.804
|_gen1 nu=-b:2/b:1
|_print nu
  NU
  6.112177
|_gen1 dwdu=-b:2/9
|_print dwdu
  DWDU
  -0.9700963
|_stop

```

