THE BUSINESS SCHOOL

Research Methods For Economists

Introduction

The major objective of research in economics is to find out the truth about economic questions that is bothering, individual households, communities, policy makers in the local and national governments or the international community as a whole. Some questions are quantitative by nature such as the distribution of income, employment level by sectors, prices and costs of commodities, demand and supply of various goods and services in the economy, international trade, growth rates of output employment, capital stock, investment, rate of returns on financial assets, primary secondary or the university level education. Some other questions are qualitative. These range about the welfare and just society, philosophical questions including behavioural and psychological analysis of decision making process by individuals and firms or the policy makers that often involve abstract reasoning.

Economists have developed many theories regarding how the various market function or should function. How the various pieces of economic activities make the national or international economy. Economic research therefore can be divided into two main groups 1) theoretical research 2) applied. Theoretical research often involves derivation of demand supply or equilibrium conditions using some sort of optimising process. Diagrams, equations or simply the logical statements can be used for theoretical deduction. Many of the standard micro or macro economic models or extension of those in various fields of economics example of theories, consumer optimisation, producer optimisation, determination of prices in a particular market for goods and services or factors of production or the general equilibrium in the entire economy or intertemporal models for accumulation, investment and growth or marginal or cumulative distributions of populations or samples or law of large numbers are example of theoretical research. These can be abstract models and requires using algebra, calculus, matrix or real analysis or stochastic probability theory to represent these theoretical ideas.

Theories need to be applied in practice to make them useful for improvement in the welfare of human society. The application involves systematic collection of information on variables identified by particular theory. Testing the claims made by those theories using linear or non-linear functions or various estimation or computation techniques. As amount of information has grown so has the need to processing those information. The applied research is basically about processing these information consistently, coherently, systematically using inductive methods. Applied research can also vary according to the nature of method used in analysis. There are mainly four categories of applied research: 1) statistical and econometric analysis 2) calibration and computations of system of equations 3) strategic analysis 4) experimental analysis. Statistical analysis involves designing, implementing and collecting data on economic variables scientifically in an unbiased manner. This also involves determining the properties of distribution of those variables, collecting information on central tendencies, correlation and pattern of relations among them. Econometric analysis further involves using data for testing various economic theories based on cross sections and time series data. Calibration and computation of system of equations involves solving n number of equations on the basis of certain assumption about their behaviour, such as market demand and supply functions, or input-output analysis or a general equilibrium system. Linear, non-linear or dynamic programming is often used to determine such a system. Game theory is becoming increasingly popular tool to analyse inter-dependence among economic agents where the action to be taken by one is determined by the beliefs or perception of that individual about the action taken by other people in the economy. They are applied to analyse the process and outcome of bargaining, strategic contingency planning or just in describing the behaviours of economic agents. Experimental analysis has the concept of using control groups for testing economic theories, such as impacts of certain policy in economic stability, such as the adoption of euro, effect of certain drugs, or certain measures on productivity, health or educational attainment.

Economics Subject group on the Research Methods module is to introduce students to quantitative and analytical tools required to prepare research proposals in the second year and for the Independent Study or the Dissertation modules in economics in the third year and to provide basic skills required to execute research programs as a professional economist taking account of the most relevant economic theory and available primary or secondary dataset in economic issue after graduation. Skills learnt in all other modules including the Empirical and Micro Economics in the first semester might be very useful in doing tutorials and assignments in this module. A special tutorial group will be formed for students who have not done Empirical economics in the first semester.

This work book aims to illustrate various techniques that can be useful for economics students in conducting their research.

Issues and topics
Basic Statistical analysis

Basics: Research questions, theory and applications and surveys
Sampling and distributions: mean and variance, covariance and correlation, Frequency distributions, Test of stationarity of time series.
Testing the properties of a random variable: Use of Normal, Standard Normal, t, F and chi-square distributions in decision making.

Basic econometric analysis

Properties of an estimator: unbiasedness, consistency and asymptotic efficiency
Choice of models: Interpreting regression results: slopes and elasticities
Errors in statistical decisions: Type I and II errors,
Transformation of variables for analysis.
Prediction and forecasting
Heteroscedasticity and Autocorrelation.

Mathematical techniques

Theoretical research: Models of demand and supply and strategic decision optimisation basics of linear and non-linear programming
Questions about the whole Economy: Introduction to the Input-Output and general equilibrium models for an economy.

Game theory

Experimental techniques

Issues and topics

Choice of techniques is the most important part of the research process. There are millions of topics that might be interesting, see the hard or electronic copies of journals available through the library (JSTOR, Econlit, Business source premier, SSRN). Hundreds of thousands of economists have written so many thing in so many subject. For instance I can sit down and make a list of topics that I see needs some further research as following:

Timely Issues for Research in Economics

UK Economy

1. Energy prices and economy
2. Policy rules for economic stability and growth
3. Reform in public policy: tax, spending, trade, regulations, redistribution
4. Can tax cuts finance budget deficit? (endogenous growth model)
5. Productivity growth in manufacturing and services sectors
6. Human capital, research and development and economic growth
7. Reasons for decline in public sector investment over years
8. Role of public and private sector in funding of education and health sectors
9. Impact of volatility of exchange and interest rate on exports
10. Unemployment and inflation: in the long and the short run
11. Two speed economy: growth of income of skilled and unskilled workers
12. Provision for pension and social security
13. Impact in the economy of rising oil and energy prices
14. Liberalisation of the financial sector and private sector investment
15. Regulations of market for certain products (e.g. carpets, mobile phones, banana, cars, cosmetics, drugs, cloths, furniture, nursing home, houses,)
16. Arguments for and against privatisation of semi-public goods (i.e. railways, airlines, telecommunications)
17. Determinants of wage and earning by professions, skills and regions
18. Wage and income of sport clubs and top quality sport men and women
19. Factors contributing to variation in growth of regional and local economies
20. Equity and redistribution aspects of council tax, income tax or direct and indirect taxes
21. Can growth occur with redistribution
22. New deal and public and private sector partnership
23. Role of demand side and supply side policies in the economy
24. Patterns of consumption and saving by categories of households
25. Employment and output multipliers with Input-output model of the UK economy
26. Assessment of the reliability of macroeconomic forecasts
27. London Stock Exchange and global economy
28. Evaluation of economic costs and benefits of environmental levy

Economic Growth and Development Issues

29. Why the four-fifth of the World is still underdeveloped?
30. Why there is a North-South divide in per-capita income?
31. Story of productivity growth: impact of industrial to internet revolutions
32. Examination of poverty alleviation and economic growth
33. How much can human capital contribute towards economic growth?
34. Problems in transfer and adoption of technology
35. Why cannot all countries grow at the same rate?
36. What is the best technology to achieve higher rate of growth?
37. Balanced versus unbalanced growth
38. Does economic growth promote economic inequality?
39. Do higher environmental standards reduce the rate of growth?
40. Conflict, coalition and economic growth
41. Economic costs of conflicts and HIV in Africa
42. How much spending on research and development promote economic growth?
43. Infrastructure and economic growth.

Macroeconomic issues

44. Why are Keynesian models applicable more in some countries than in others?
45. Why the rates of unemployment are higher in rigid labour markets?
46. Examination economic problems when savings are not equal to investment
47. Should households save more to make economy grow faster?
48. Resource imbalances and economic crises
49. First, second and third theories of economic crises.
50. What are the best policy rules for stability and growth?
51. Trade-off between unemployment and inflation?
52. Can independent central banks do better than government controlled ones?
53. How can exchange rate instability be harmful for an economy?
54. Credibility of public policy and market reactions

Microeconomic issue

55. Determinants of consumption and saving.
56. Are consumers sovereign in market for goods and services?
57. Income and substitution effects of price changes
58. Analysis of short and long run cost of a certain firm or industry
59. Consequences of factor and product taxes in a competitive market.
60. Impacts of new technology in costs of production and supply.
61. Market imperfections, inefficiency and regulation
62. Is there any evidence for income and substitution effects in agriculture, manufacturing or engineering sectors?
63. What are the welfare consequences of duopoly or oligopoly in the energy markets?
64. Does deregulation and privatisation bring efficiency in allocation of resources?
65. Analysis of expenditure pattern of households
66. How elasticities of supply and demand affect the burden to a consumer?
67. Examination of benefits and costs of privatisation
68. Application of utility maximisation hypothesis under uncertainty.

Trade issues

69. Examination of Tariff and non-tariff barriers of trade
70. Does global free trade reduce or increase income and wage inequality?
71. Does the direct foreign investment promote economic growth?
72. Who benefits and who loses from regional economic cooperation?
73. Assessment of impact of increase in oil prices in the global income
74. Enlargement of EU and Economic prospects of its new members
75. How can liberal trade reduce pressure of illegal immigration to rich countries?
76. Evaluation of achievements of the WTO and the Doha rounds of trade talk
77. Leontief paradox or factor price equalisation?

Public Policy issues

78. Should budget be balanced all the time?
79. Should government subsidise education or pay more unemployment benefit?
80. How can budget deficit create external and internal imbalances?
81. Can lower taxes reduce budget deficit?
82. Does the Ricardian equivalence apply in modern economies?
83. Examination of optimal tax rate and evidence
84. Optimal amount of public services?
85. Optimal allocation of public funds between local and central authorities

Households and labour market

86. Why current economic policies have created pension crisis in the West?
87. Income dynamics and life-cycle profiles of income
88. Determinants of wage and labour supply
89. Gender inequality in wage and earning
90. Link between educational qualification and earning
91. New technology, redundancy and structural transformation of labour market
92. Social safety net and unemployment: re-examination of Beveridge provisions.

Environment and natural resources

93. Economic impacts of Kyoto agreement

94. Consumption and production side externalities and social welfare
95. Double dividend hypothesis of environmental taxes
96. Do tight environmental regulations reduce economic growth?
97. Valuation and optimal use of non-renewable resources

Financial market and economy

98. Over or under investment, Value of a firm and the optimal stock of capital
99. Why banks tend to accumulate non-performing debt with weak monitoring?
100. Best way of financing economic development.
101. Analysis of risks and return in the financial market?
102. Volatility of financial markets and economy
103. Can Tobin tax (transaction cost) deter financial crisis?

Education economics

104. Impacts of universal primary education in skill formation
105. Who should pay tuitions: students or the government?
106. Matching education and job market

Country specific studies

107. Analysis of markets using microeconomic models
108. Model for macroeconomic policy evaluation and forecasting
109. Evaluation of impacts of tax reforms
110. Forecasting various policy scenarios
111. Movement in commodity prices and terms of trade

Commodity Markets:

112. agricultural goods: sugar, potato, cotton, rubber, green vegetables, tomato
113. fruit: apples, banana, pears, grapes, oranges, mango, jackfruit , coconut, nuts
114. grains: rice, corn, millet, wheat, maize, palm oil, peanuts
115. meat market: fish, beef, pork, lamb
116. drinks: wine, beer, whiskey, martini
117. metals and minerals: gold, silver, aluminium, steel, iron, copper, tin, zinc, oil

Uncertainty and asymmetric information

118. risks and uncertainty and markets for insurance
119. moral hazards and adverse selection
120. principle agent problem and monitoring
121. efficient contract and incentives
122. provisions for contingency

Energy sectors

123. Energy prices and trade in the global economy
124. Generation and distribution of electricity and pollution
125. Kyoto agreement on climate change
126. Trade-off between trade and environment
127. OPEC effect on oil and energy prices
128. Renewable energy and exhaustion of non-renewable energy
129. Fuel poverty
130. Role of energy sector in the growth of economy
131. Technological factors in promotion of the energy sector

Basic Statistical analysis

Basics: Research questions, theory and applications and surveys
Sampling and distributions: mean and variance, covariance and correlation, Frequency distributions, Test of stationarity of time series.
Testing the properties of a random variable: Use of Normal, Standard Normal, t, F and chi-square distributions in decision making.

Descriptive Statistics in Economic Research

1. Design a pilot for a consumer satisfaction survey for a supermarket in your local area with a questionnaire that includes about 20 questions. Generate a tabulation plan for the information from the survey. Formulate any five useful hypotheses for the supermarket. Discuss how you could test these hypotheses.

2. Do appropriate bar, line, pie, area charts using Excel data files: MBA.XLS, Students.xls, grow_lowinca.xls and income_distribution.xls.

3. Prepare a panel data on growth rates, ratios of investment, saving, exports and imports and inflation and population growth for any five low income economies using data on grow_lowinca.xls.

4. Read the panel data file you generated in (3) above and find the mean, variance, Skewness and Kurtosis for the growth rates, ratios of investment , saving and trade and population growth rates using the descriptive statistics package in PcGive.

5. Is there any correlation between growth rates and saving, investment, trade and population?

Brief Instructions to the use of software

1. **Excel** Spreadsheets are very user friendly and could be used for algebraic calculations and statistical analyses for many kinds of economic models. First prepare an analytical solution by hand then use Excel formula to compute. Excel has constrained optimiser routine at tool/goal seek and solver commend. Koop (2000) is a brilliant text for analysis of economic data using excel.
Koop G (2000) Analysis of Economic Data, Wiley, UK.

2. **OX-GiveWin-PcGive-STAMP** (www.oxmetrics.net) is a very good econometric software for analysing time series and cross section data. This software is available in all labs in the network of the university by sequence of clicks Start/applications/economics/givewin. Following steps are required to access this software.
a. save the data in a standard **excel file**.
b. start give win at start/applications/economics/**givewin**.
c. open the data file using **file/open datafile** command.
d. choose **PcGive module** for econometric analysis.
e. select the package such as **descriptive statistics**, **econometric modelling** or **panel data** models.
d. choose **dependent and independent** variables as asked by the menu. Choose options for output.
e. do the **estimation** and analyse the results, generate **graphs** of actual and **predicted** series.
Consult manuals by Doornik J A and D.F. Hendry ((2003) PC-Give Volume I-III, GiveWin Timberlake Consultants Limited, London or by visiting the web http://www.oxmetrics.net.
A **Batch file** can be written in OX for more complicated calculations using a text editor in GiveWin or such as pfe32.exe. Such file contains instructions for computer to compute several tasks in a given sequence.

Tentative Answers for Tutorial 1

Q1.

Consumer Satisfaction Survey

I. Background - questions

A. Age

1. below 20 2. 20-29 3. 30-39 4. 40-49 5. 50-59 6. 60 or above

B. Gender 1. Male 2. Female

C. Employment 1. Employed 2. Unemployed

D. If Unemployed 1. Student 2. housewife 3. job seeker 4. disabled

E. Employed 1. self employed 3. semi skilled manual 4. Professional

F. Accommodation 1. rent 2. owns a house 3. Council house

G. Education 1. below GCSE 2. GCSE 3. A-level 4. College 5. University

II. Income and expenditure

What is your average monthly income? £

What is your average monthly spending?

1. Housing cost £ 2. Food and beverage £

3. Transportation £ 4. Entertainment £

5.Other £

III. How frequently do you shop in the following superstores in month and how much do you spend at one time?

	Tesco	ASDA	Morrison	Jackson	Sainsbury
Shopping Time					
Amount spent					

IV. Rank these stores according to your level of satisfaction

	Tesco	ASDA	Morrison	Jackson	Sainsbury
Excellent					
Good					
Average					
Bad					
Very bad					

V. Rank any one of the above stores on the basis of following criteria

	Excellent	Good	Average	bad	Very bad
Choices					
Staff					
Other services					
Adverts					
Store Credits					

Sample size: 100.

Potential hypotheses

a. Does expenditure in a superstore depend on income of an individual?
b. Is there any link between the average spending and frequency of shopping time?
c. Is there a positive correlation among people spending on housing, food and beverage, transportation, entertainment and other spending?
d. Does the level of spending depend on the quality of store credit?
e. Do unemployed people like more ASDA than Tesco?

f. Find the Spearman rank coefficient between the rating of superstores made by male/female, employed/unemployed, those who own a house/those who live in council houses.
g. Test whether expenditure differs by the level of education and age?
h. How can you show an interaction effect of age and education and gender on average monthly spending?

Testing these hypotheses requires some understanding of descriptive statistics and use of Normal, T, F and chi-square distributions.

Tabulation plan (Statistical software such as SPSS is very good in tabulationa)

Gender based analysis (structure of the sample)

	Male	Female
Below GCSE		
GDSE		
A Level		
College		
University		

Average income (spending)

	Male	Female
Below GCSE		
GDSE		
A Level		
College		
University		

Average income (spending) by gender and age

	Male	Female
Below 20		
20-29		
30-39		
40-49		
50-59		
60 or above		

Average satisfaction rates for a particular store by gender and age

	Male	Female
Below 20		
20-29		
30-39		
40-49		
50-59		
60 or above		

Average satisfaction rates for a particular store by employment category

	Employed	Unemployed
Below 20		
20-29		
30-39		
40-49		
50-59		
60 or above		

By accommodation status

	Male	Female
Renting a house		
Owns a house		
Stays in a council house		

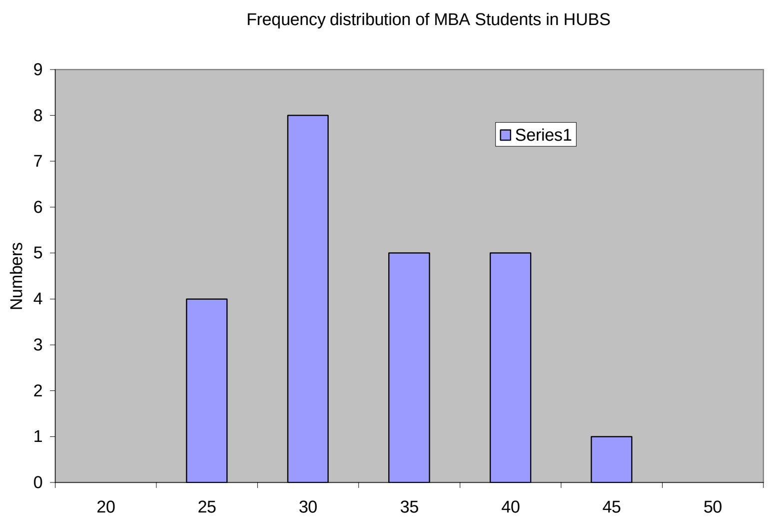
Satisfaction rating of stores by age

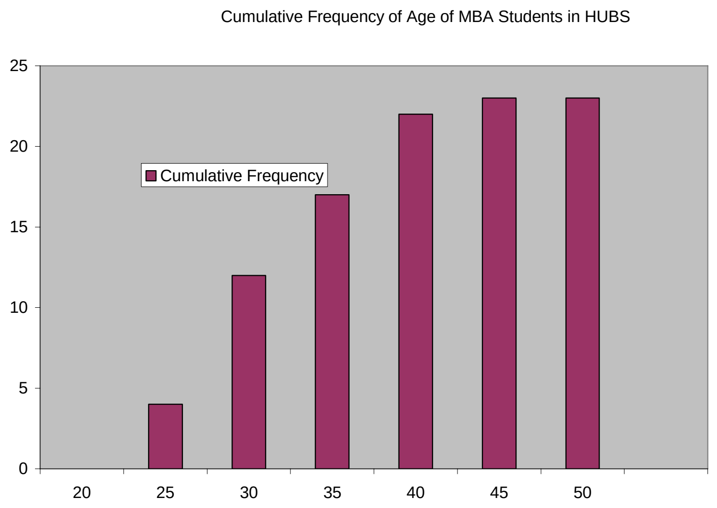
	Below 20	20-29	30-39	40-49	50-59	60 or above
Tesco						
ASDA						
Morrison						
Jackson						
Sainsbury						

Satisfaction rating of stores by the level of education

	Tesco	ASDA	Morrison	Jackson	Sainsbury
Below GCSE					
GCSE					
A Level					
College					
University					

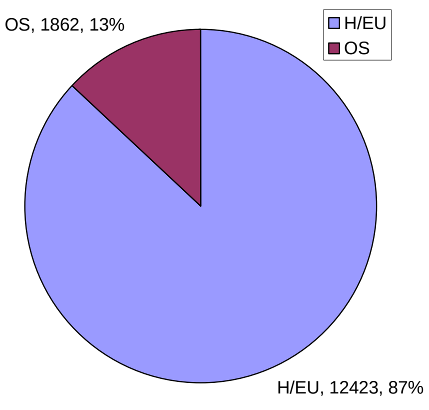
Q2. B



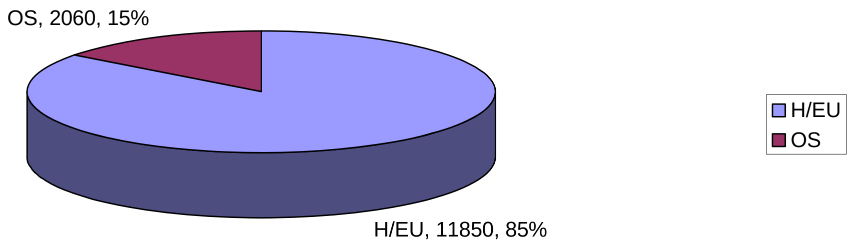

Cumulative Frequency of Age of MBA Students in HUBS
Cumulative Frequency
25
20
15
10
5
0
20
25
30
35
40
45
50

Q2.

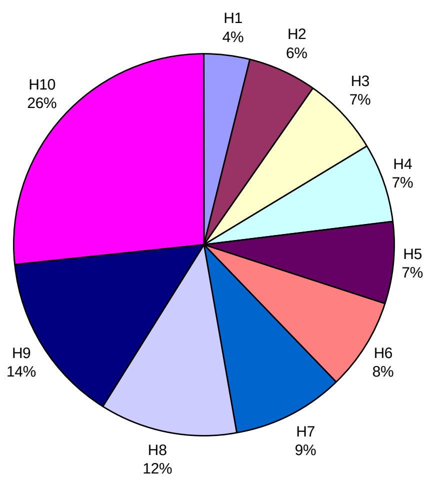
Compostion and number of Home, EU and Overseas Students in Hull, 2004



Composition of Home and EU and Overseas Students in the University of Hull in 2001



Consumption of Households in the UK

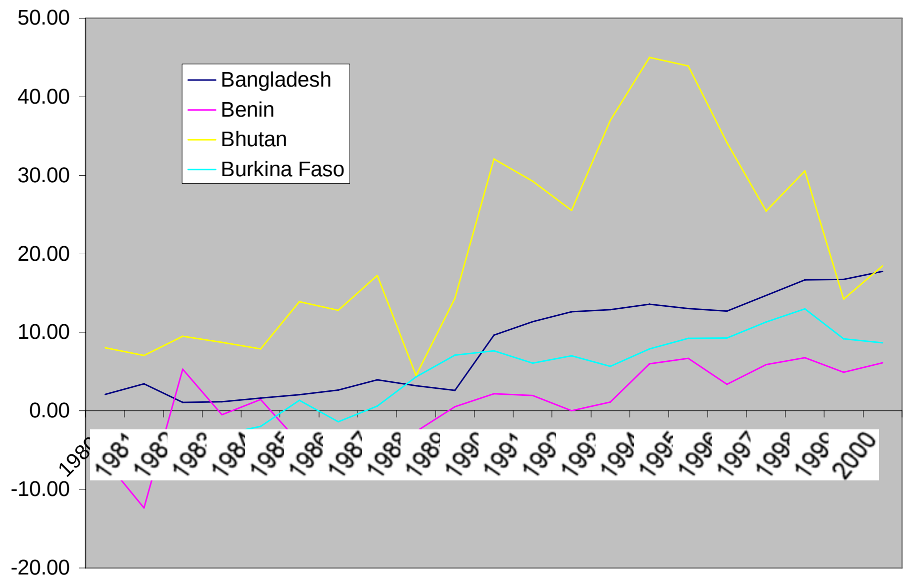


Steps to make a pie chart: 1. highlight data of consumption distribution by hh 2. Click on chart Wizard in excel 3. put data labels, titles and legends as appropriate

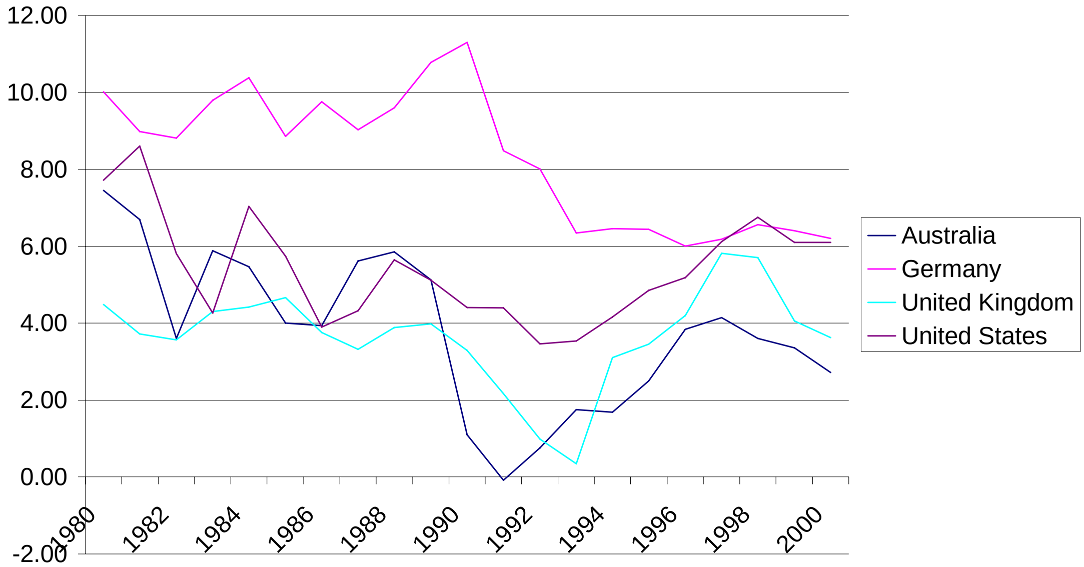
4. Save the chart in as a new worksheet to look nicer 5. copy and past in a word document like above.

Q2. Line graphs.

Saving ratio in Four Low Income Countries



Saving ratios in selected OECD economies



Q3. Preparation of the panel data

Year	gy	inv	rint	ppg	sv	pcg	trd	gdppc
1970	2.36	19.54	-0.03	0.31	21.18	2.04	43.86	11827
1971	2.12	19.60	-1.62	0.53	21.18	1.58	43.55	12014
1972	3.63	19.21	-0.55	0.30	19.14	3.31	42.40	12412
1973	7.31	20.58	0.65	0.22	20.20	7.07	48.47	13290
1974	-1.68	21.69	-5.16	0.02	17.94	-1.70	59.61	13064
1975	-0.68	20.72	-14.20	-0.02	17.68	-0.66	52.57	12978
1976	2.80	20.43	-5.36	-0.02	19.97	2.82	57.24	13344
1977	2.36	19.39	-3.53	-0.05	21.29	2.41	58.75	13666
1978	3.40	19.19	-2.06	-0.02	21.55	3.43	55.19	14134
1979	2.75	19.36	-0.51	0.11	20.77	2.63	55.20	14505
1980	-2.18	18.76	-2.69	0.16	19.83	-2.33	52.03	14167
1981	-1.27	17.11	1.78	0.04	18.81	-1.31	50.37	13982
1982	1.80	17.11	4.12	-0.06	18.47	1.86	50.63	14242
1983	3.75	17.02	4.29	0.10	18.27	3.64	51.94	14760
1984	2.45	18.11	4.97	0.23	18.18	2.22	56.79	15087
1985	3.78	18.14	6.35	0.32	19.27	3.46	56.47	15608
1986	4.21	18.00	7.49	0.29	17.36	3.90	52.03	16217
1987	4.43	18.82	4.22	0.28	17.88	4.14	51.94	16889
1988	5.17	20.54	4.01	0.26	17.79	4.89	49.62	17715
1989	2.11	21.63	6.01	0.35	18.10	1.75	51.55	18026
1990	0.66	20.55	6.59	0.35	17.57	0.30	50.65	18081

1991	-1.47	17.91	4.58	0.43	16.07	-1.89	47.30	17739
1992	0.07	16.49	5.24	0.34	14.96	-0.27	48.34	17692
1993	2.33	15.83	3.10	0.32	14.84	2.00	51.75	18046
1994	4.39	15.85	3.89	0.35	15.74	4.02	53.44	18772
1995	2.79	16.30	4.07	0.36	16.52	2.42	57.09	19226
1996	2.55	16.62	2.61	0.33	16.28	2.21	58.83	19651
1997	3.51	16.66	3.57	0.35	17.27	3.15	56.88	20270
1998	2.64	17.39	4.06	0.42	17.03	2.21	53.88	20718
1999	2.29	17.69	2.99	0.41	15.83	1.87	53.53	21105
2000	3.07	17.77	4.10	0.40	15.99	2.66	56.32	21667
1970	0.22	17.83	2.70	1.17	18.35	-0.94	11.26	16985
1971	2.88	18.34	0.11	1.26	19.10	1.58	11.24	17254
1972	5.55	19.11	1.05	1.07	19.53	4.43	11.83	18018
1973	5.90	19.51	2.31	0.95	20.87	4.89	13.57	18900
1974	-0.56	18.94	1.64	0.91	19.75	-1.46	17.08	18623
1975	-0.34	17.75	-1.37	0.99	18.36	-1.32	16.15	18378
1976	5.64	18.09	1.14	0.95	18.94	4.64	16.63	19230
1977	4.70	19.40	0.37	1.01	19.37	3.65	16.97	19933
1978	5.57	20.66	1.81	1.06	20.70	4.46	17.52	20822
1979	3.21	21.23	3.96	1.10	21.03	2.08	18.94	21254
1980	-0.24	20.24	5.57	0.96	19.57	-1.19	20.69	21001
1981	2.45	19.96	8.72	0.98	20.63	1.45	20.01	21306
1982	-2.07	18.88	8.16	0.95	17.92	-3.00	18.16	20667
1983	4.33	18.62	6.57	0.91	17.14	3.38	17.30	21367
1984	7.28	19.55	8.03	0.87	18.62	6.35	18.18	22724
1985	3.82	19.58	6.52	0.89	17.45	2.90	17.25	23384
1986	3.37	19.32	5.97	0.92	16.64	2.42	17.51	23950
1987	3.36	18.60	5.01	0.89	16.16	2.44	18.59	24535
1988	4.16	18.36	5.70	0.91	16.49	3.22	19.76	25324
1989	3.50	18.05	6.79	0.94	17.07	2.53	20.20	25965
1990	1.74	17.37	5.87	1.06	16.35	0.68	20.62	26141
1991	-0.50	16.16	4.65	1.10	15.81	-1.59	20.64	25725
1992	3.06	16.20	3.72	1.15	16.00	1.89	20.80	26211
1993	2.67	16.72	3.52	1.20	16.13	1.45	20.93	26592
1994	4.08	17.27	4.96	1.24	16.92	2.79	21.99	27334
1995	2.70	17.68	6.51	1.29	16.98	1.39	23.46	27713
1996	3.61	18.19	6.20	1.27	17.42	2.30	23.71	28351
1997	4.47	18.72	6.36	1.25	18.40	3.18	24.50	29251
1998	4.40	19.52	7.00	1.23	18.67	3.13	23.87	30166
1999	4.26	20.17	6.40	1.21	17.91	3.01	24.19	31073
2000	4.20	20.17	6.87	1.19	17.91	2.97	24.19	31996
1970	4.20	20.17	6.87	0.74	17.91	2.97	24.19	31996
1971	4.20	20.17	6.87	0.83	17.91	2.97	24.19	31996
1972	4.20	25.92	6.87	0.45	28.09	2.97	37.99	19088
1973	4.77	24.38	6.87	0.31	28.27	4.45	39.47	19936
1974	0.19	22.05	6.87	0.03	26.39	0.17	46.86	19969
1975	-1.25	20.80	6.87	-0.38	22.68	-0.87	44.99	19795
1976	5.32	20.54	6.87	-0.46	23.82	5.81	47.55	20945
1977	2.85	20.68	6.87	-0.19	23.26	3.04	47.03	21582
1978	3.00	21.05	2.94	-0.11	23.73	3.11	45.59	22253
1979	4.22	22.13	4.66	0.03	24.11	4.20	47.93	23186
1980	0.98	23.01	6.71	0.25	22.99	0.72	51.67	23354
1981	0.49	22.19	10.63	0.15	21.85	0.34	55.06	23435
1982	-0.56	20.95	9.00	-0.11	21.81	-0.45	55.74	23328

1983	2.05	20.94	6.92	-0.27	22.37	2.33	54.27	23872
1984	3.04	20.41	7.71	-0.35	22.74	3.40	57.27	24684
1985	2.35	19.94	7.42	-0.19	22.97	2.55	59.97	25313
1986	2.52	19.92	5.84	0.04	24.48	2.48	54.09	25942
1987	1.69	20.02	6.58	0.14	24.15	1.54	52.13	26341
1988	3.65	20.27	6.78	0.39	25.03	3.25	53.01	27196
1989	3.48	20.86	7.63	0.78	25.97	2.68	56.52	27925
1990	3.23	22.32	7.91	0.86	25.71	2.35	54.27	28581
1991	2.84	23.76	7.40	0.73	24.06	2.09	52.85	29179
1992	2.24	24.04	8.15	0.76	23.57	1.47	49.32	29608
1993	-1.09	23.03	8.86	0.66	22.67	-1.74	45.36	29094
1994	2.35	23.13	8.75	0.44	23.51	1.89	46.84	29645
1995	1.73	22.44	8.74	0.15	23.31	1.57	48.33	30110
1996	0.77	21.76	8.90	0.33	22.69	0.43	49.57	30241
1997	1.40	21.40	8.24	0.19	22.94	1.20	54.33	30604
1998	2.05	21.32	7.79	-0.03	23.36	2.08	56.21	31241
1999	1.56	21.33	7.86	0.05	23.16	1.51	57.92	31712
2000	2.95	21.40	10.06	0.08	23.06	2.87	66.33	32623
1970	5.73	24.90	2.22	0.90	27.34	4.79	30.40	16412
1971	4.78	25.29	0.54	0.94	27.19	3.80	31.05	17037
1972	4.43	25.34	-0.57	0.87	27.32	3.52	31.67	17637
1973	5.44	25.85	0.10	0.80	27.96	4.60	33.52	18447
1974	3.11	26.42	0.12	0.65	26.87	2.44	41.50	18897
1975	-0.28	24.72	-2.63	0.45	24.40	-0.73	36.10	18759
1976	4.24	24.53	-1.85	0.40	24.40	3.83	39.05	19477
1977	3.22	23.46	0.22	0.45	24.26	2.76	39.97	20015
1978	3.35	22.92	-0.91	0.43	24.17	2.90	38.61	20596
1979	3.31	22.99	-0.18	0.43	23.84	2.87	40.63	21186
1980	1.61	23.82	1.26	0.51	22.66	1.09	43.20	21418
1981	1.22	23.11	2.93	0.56	20.57	0.65	45.10	21558
1982	2.63	22.48	1.90	0.55	19.73	2.07	44.82	22005
1983	1.49	21.20	2.98	0.45	19.75	1.03	44.28	22232
1984	1.65	20.38	4.80	0.40	19.54	1.24	46.95	22508
1985	1.45	20.30	5.37	0.41	19.26	1.04	46.80	22742
1986	2.41	20.38	4.56	0.41	20.32	1.99	40.90	23196
1987	2.53	20.98	6.55	0.43	20.21	2.10	40.47	23682
1988	4.61	21.95	6.22	0.46	21.73	4.14	41.60	24663
1989	4.17	22.54	6.67	0.98	22.45	3.16	44.36	25441
1990	2.61	22.56	7.45	0.53	22.39	2.07	43.46	25967
1991	1.00	21.95	7.06	0.43	21.99	0.57	43.47	26114
1992	1.49	20.94	7.85	0.46	21.44	1.02	42.48	26381
1993	-0.89	19.38	6.43	0.40	19.74	-1.28	39.97	26043
1994	2.07	19.08	6.11	0.33	20.30	1.73	41.63	26493
1995	1.67	18.79	6.34	0.32	20.61	1.35	43.64	26850
1996	1.10	18.48	5.24	0.31	19.97	0.79	44.50	27061
1997	1.90	17.95	4.99	0.31	20.84	1.59	48.02	27490
1998	3.40	18.43	5.56	0.33	21.73	3.06	49.58	28333
1999	2.92	19.06	5.87	0.38	21.86	2.53	49.69	29049
2000	3.10	19.69	5.73	0.46	21.96	2.62	55.90	29811
1970	10.71	35.63	1.13	1.13	40.44	9.46	20.05	20465
1971	4.70	34.36	2.38	1.29	38.58	3.36	20.42	21153
1972	8.41	34.23	1.36	1.40	37.95	6.91	18.60	22614
1973	8.03	36.51	-4.90	0.83	38.25	7.14	19.77	24229
1974	-1.23	34.90	-9.68	1.91	36.72	-3.09	27.57	23479

1975	3.09	32.56	1.79	1.60	32.94	1.45	25.20	23821
1976	3.98	31.29	0.23	0.74	32.75	3.21	25.95	24585
1977	4.39	30.26	0.76	0.96	32.58	3.39	24.22	25418
1978	5.27	30.51	1.74	0.90	32.74	4.32	20.22	26517
1979	5.48	31.78	3.52	0.84	31.72	4.60	23.71	27737
1980	2.82	31.68	2.76	0.78	31.48	2.01	27.90	28296
1981	2.85	30.75	3.38	0.74	32.07	2.09	28.26	28888
1982	3.14	29.62	5.43	0.68	30.77	2.45	27.93	29595
1983	2.27	28.14	5.14	0.68	29.94	1.57	25.70	30060
1984	3.84	27.88	3.83	0.63	30.88	3.18	26.82	31016
1985	4.36	27.66	4.10	0.61	31.73	3.73	24.96	32172
1986	2.97	27.53	4.32	0.61	31.91	2.34	18.43	32927
1987	4.46	28.63	5.31	0.49	31.81	3.95	17.37	34228
1988	6.51	30.00	4.32	0.43	32.89	6.06	17.46	36301
1989	5.28	30.97	3.27	0.41	33.24	4.85	19.12	38063
1990	5.33	32.19	4.42	0.34	33.71	4.97	19.81	39955
1991	3.12	31.76	4.43	0.31	34.05	2.80	18.28	41074
1992	0.93	30.48	4.39	0.25	32.85	0.68	17.48	41351
1993	0.42	29.20	3.78	0.25	31.45	0.17	15.92	41422
1994	1.00	28.20	4.03	0.34	30.07	0.65	15.99	41693
1995	1.57	27.75	3.78	0.38	29.59	1.18	16.78	42186
1996	3.47	28.45	3.50	0.26	29.63	3.20	18.91	43538
1997	1.80	28.09	2.08	0.26	29.80	1.53	20.39	44206
1998	-1.10	26.89	2.38	0.25	28.72	-1.35	19.51	43609
1999	0.76	26.19	3.60	0.19	27.55	0.57	18.42	43856
2000	2.40	26.19	2.67	0.17	27.55	2.22	18.42	44830

Q4.

Computing descriptive statistics using Excel (Tools/data analysis/descriptive statistics)

gy		*Inv*		*rint*	
Mean	2.792468	Mean	22.23586	Mean	3.964101
Standard Error	0.168905	Standard Error	0.381073	Standard Error	0.294195
Median	2.8495	Median	20.684	Median	4.4157
Mode	4.2	Mode	20.169	Mode	6.8743
Standard Deviation	2.102854	Standard Deviation	4.744315	Standard Deviation	3.662698
Sample Variance	4.421994	Sample Variance	22.50852	Sample Variance	13.41536
Kurtosis	1.079932	Kurtosis	0.447801	Kurtosis	4.120547
Skewness	0.159391	Skewness	1.100491	Skewness	-1.47578
Range	12.8881	Range	20.681	Range	24.822
Minimum	-2.1791	Minimum	15.833	Minimum	-14.197
Maximum	10.709	Maximum	36.514	Maximum	10.625
Sum	432.8325	Sum	3446.558	Sum	614.4356
Count	155	Count	155	Count	155
Largest(1)	10.709	Largest(1)	36.514	Largest(1)	10.625
Smallest(1)	-2.1791	Smallest(1)	15.833	Smallest(1)	-14.197
Confidence Level(95.0%)	0.33367	Confidence Level(95.0%)	0.752804	Confidence Level(95.0%)	0.581178

Descriptive statistics calculated from the PcGive

Normality test and descriptive statistics using Pcive

Normality tests and descriptive statistics (using growth5panel_gw.xls)

The sample is 1 - 155
Normality test for gy
Observations 155
Mean 2.7925
Std.Devn. 2.0961
Skewness 0.15784
Excess Kurtosis 1.0069
Minimum -2.1791
Maximum 10.709
Asymptotic test: Chi^2(2) = 7.1920 [0.0274]*
Normality test: Chi^2(2) = 8.1183 [0.0173]*

Normality test for rint
Observations 155
Mean 3.9641
Std.Devn. 3.6509
Skewness -1.4615
Excess Kurtosis 3.9504
Minimum -14.197
Maximum 10.625
Asymptotic test: Chi^2(2) = 155.96 [0.0000]**
Normality test: Chi^2(2) = 42.597 [0.0000]**

Normality test for sv
Observations 155
Mean 22.947
Std.Devn. 5.7244
Skewness 0.92077
Excess Kurtosis 0.10999
Minimum 14.836
Maximum 40.438
Asymptotic test: Chi^2(2) = 21.980 [0.0000]**
Normality test: Chi^2(2) = 49.977 [0.0000]**

Normality test for trd
Observations 155
Mean 36.875
Std.Devn. 15.186
Skewness -0.089341
Excess Kurtosis -1.5212
Minimum 11.239
Maximum 66.325
Asymptotic test: Chi^2(2) = 15.150 [0.0005]**
Normality test: Chi^2(2) = 36.316 [0.0000]**

Normality test for gdppc
Observations 155
Mean 24469.
Std.Devn. 7377.5
Skewness 0.78516

Normality test for inv
Observations 155
Mean 22.236
Std.Devn. 4.7290
Skewness 1.0898
Excess Kurtosis 0.39502
Minimum 15.833
Maximum 36.514
Asymptotic test: Chi^2(2) = 31.690 [0.0000]**
Normality test: Chi^2(2) = 75.142 [0.0000]**

Normality test for ppg
Observations 155
Mean 0.53548
Std.Devn. 0.42381
Skewness 0.32474
Excess Kurtosis -0.11963
Minimum -0.46116
Maximum 1.9090
Asymptotic test: Chi^2(2) = 2.8167 [0.2445]
Normality test: Chi^2(2) = 3.3174 [0.1904]

Normality test for pcg
Observations 155
Mean 2.2317
Std.Devn. 1.9980
Skewness 0.0016827
Excess Kurtosis 1.0749
Minimum -3.0929
Maximum 9.4642
Asymptotic test: Chi^2(2) = 7.4614 [0.0240]*
Normality test: Chi^2(2) = 9.1553 [0.0103]*

Excess Kurtosis 0.49015
Minimum 11827.
Maximum 44830.
Asymptotic test: Chi^2(2) = 17.477 [0.0002]**
Normality test: Chi^2(2) = 20.451 [0.0000]**

Q5. Correlation coefficients from Excel

	gy	*inv*	*rint*	*ppg*	*sv*	*pcg*	*trd*	*gdppc*
Gy	1							
Inv	0.274538	1						
Rint	-0.01514	-0.23112	1					
Ppg	0.283211	0.175482	-0.10285	1				
Sv	0.289307	0.954287	-0.18066	0.083435	1			
Pcg	0.977493	0.247064	0.002356	0.077893	0.283346	1		
Trd	-0.25733	-0.39036	0.115557	-0.66774	-0.32154	-0.12148	1	
Gdppc	-0.00773	0.441057	0.368287	0.032412	0.44054	-0.01598	-0.38251	1

From PcGive

	gy	inv	rint	ppg	sv	pcg	trd	gdppc
Gy	1	0.27454	-0.01514	0.28321	0.28931	0.97749	-0.25733	-0.00773
Inv	0.27454	1	-0.23112	0.17548	0.95429	0.24706	-0.39036	0.44106
Rint	-0.01514	-0.23112	1	-0.10285	-0.18066	0.002356	0.11556	0.36829
Ppg	0.28321	0.17548	-0.10285	1	0.083435	0.077893	-0.66774	0.032412
Sv	0.28931	0.95429	-0.18066	0.083435	1	0.28335	-0.32154	0.44054
Pcg	0.97749	0.24706	0.002356	0.077893	0.28335	1	-0.12148	-0.01599
Trd	-0.25733	-0.39036	0.11556	-0.66774	-0.32154	-0.12148	1	-0.38251
Gdppc	-0.00773	0.44106	0.36829	0.032412	0.44054	-0.01599	-0.38251	1

Survey on Study and Expenses

I would appreciate if you could help me by answering the following survey questions. This study is just for academic purpose and information provided will strictly be kept confidential.

1. Your gender
   Male   Female
2. Age group
   a.15-20   b.20-25 c. 25-30   d. above 30
3. How many hours do you study and sleep now-a-days?

   Study: a. less than three b. 3-5   c. 5-8   d. 8-10

   Sleep: a. less than 5 b. 5-7 c. 7-8   d. 8-10 e. more than 10

4. How many pages do you read everyday?
   a. less than 10 b. 10-20   c. 20-40   d. above 40

5. How many pages do you write everyday?
   a. less than 5 b. 5-10   c. 10-20   d. above 20

6. How many years have you worked so far?

   a. less than 1 Yr b. 1-3   c. 3-5   d. 5-8   e. above 8

7. What is your total spending?

   a. rent ……   b. drinks (tea,coffee or milk) …… c. all other items …..

8. Do you smoke?   a. Yes   b. No

Possible hypotheses:

1. Is number of years of work significant determinants of how many pages does one read, or write or his or her hours of study?

2. Is the amount of spending in drinks linked to numbers of hours of study?

3. Are hours of study significantly different between males and females?

4. Do pages read everyday vary according to age group?

5. Do smokers study longer hours?

Statistical analysis:

a. proportion of male and female by age group and smoking habits in the sample
b. Mean, variance, Skewness and Kurtosis and correlation of
   i. age, study hours, pages written and read and years of work
   ii. expenses on rents, drinks and other things

6. Calculate the conditional probability of :
   a. a person is smoker given that he is male.
   b. at least 10 page writing given that she is female.
   c. the poor student (who spends the least) studying longer hours

7. Write a probit model for decision to smoke as a function of sex, spending, study hours and pages written and read.

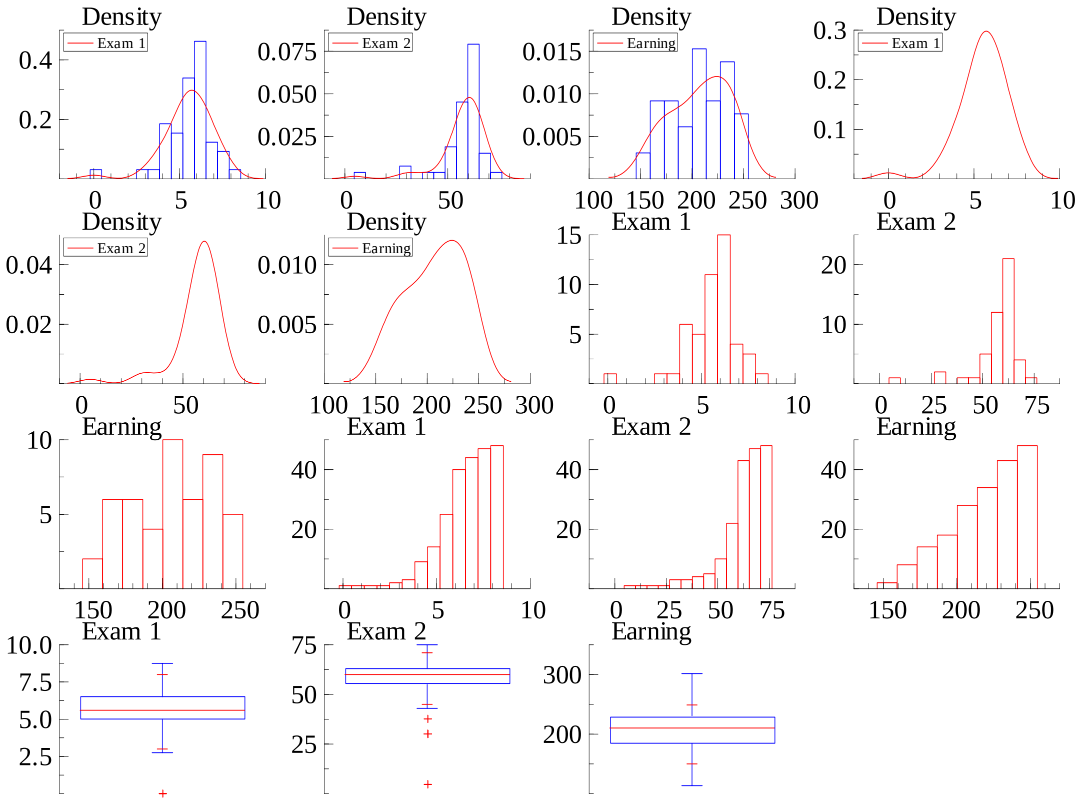
Tutorial 2
Statistical Analysis of cross sections

Prepare an excel file for the following data on marks scored by students in two exams and their monthly earnings from part time jobs. Use Excel to calculate various statistics as asked below.

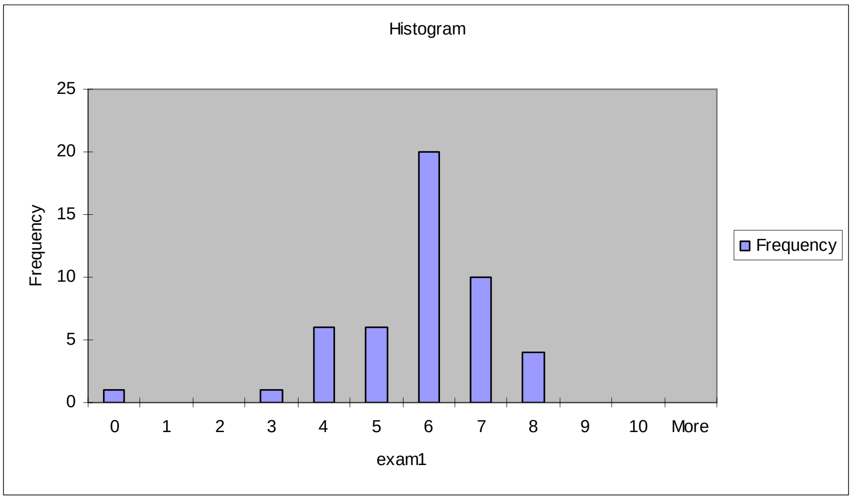
Scores in Exams and Earnings

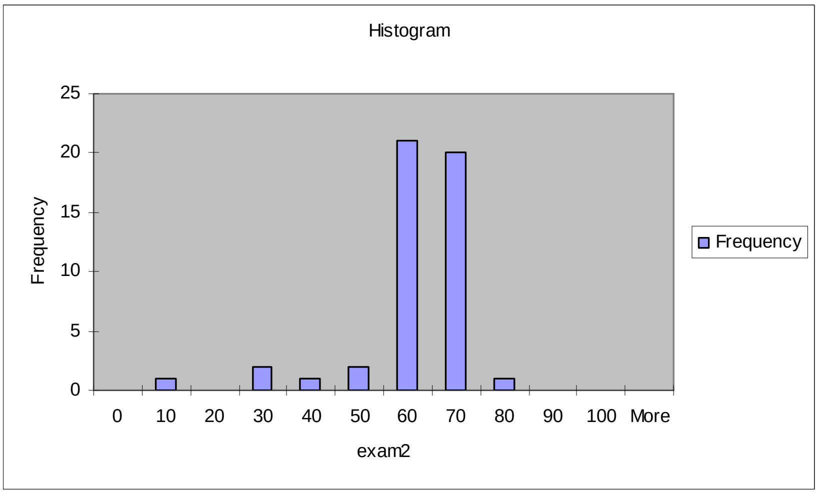
Observation	Exam 1	Exam 2	Earning	observation	Exam 1	Exam 2	Earning
1	0	5	248	25	5.2	45	196
2	6	55	161	26	4.8	55	208
3	5.6	60	213	27	7	63	245
4	5.6	54	222	28	7.5	62	155
5	4	56	180	29	5.2	65	168
6	6	65	184	30	5.2	66	211
7	6	58	249	31	6.4	60	208
8	5.2	61	191	32	4	58	238
9	3.2	30	213	33	6	68	172
10	6.5	61	186	34	5.2	64	172
11	7.5	65	232	35	6	57	219
12	4	30	235	36	7	60	202
13	5	58	231	37	5.6	63	178
14	7	68	242	38	5.6	61	175
15	6.5	68	209	39	7	62	163
16	4.5	60	230	40	6.5	65	193
17	8	71	238	41	6.5	53	200
18	6	62	150	42	7.5	62	206
19	6.5	55	237	43	5	58	241
20	6	50	227	44	6	63	227
21	5.2	51	184	45	6	58	245
22	5.2	55	229	46	5	57	161
23	4	64	205	47	3	38	213
24	4	51	221	48	4.8	60	227

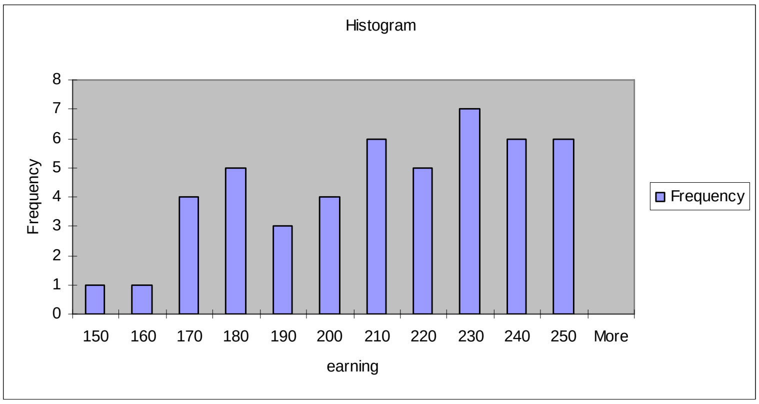
1. Represent the data on scores in exams and earning using marginal and cumulative frequency diagrams.
2. What are means and variances of scores in exam 1 and exam 2? What are the coefficients of variation of scores in exams1 and exam 2?
3. What is the covariance of marks in exams 1 and 2?
4. What is the correlation coefficient of scores between exam 1 and 2?
5. If exam 1 weighs 100 percent but the scores in exam 2 weigh only 10 percent what would be the weighted aggregate mean score in these two exams? What would be the variance of weighted scores?
6. Exam 1 took place before exam 2. Test whether scores in exam 1 can predict scores in exam 2?
7. Predict scores in exam 2 for students who scored 6 and 8 in exam 1.
8. Test hypothesis whether scores in exam 1and exam 2 are significant determinants of earning. Why may earnings be negatively related with their score in the exams for full times students?
9. How can behaviours of teachers and students change the distribution of marks?
10. If the true mean was 6 for score 1 and 58 for score 2 find whether the current sample reflects the population using t-test.
11. Derive the standard normal distribution for score 2 and construct a 99 percent confidence interval for it.



Exam1	Frequency	exam2	Frequency	earning	Frequency
0	1	0	0	150	1
1	0	10	1	160	1
2	0	20	0	170	4
3	1	30	2	180	5
4	6	40	1	190	3
5	6	50	2	200	4
6	20	60	21	210	6
7	10	70	20	220	5
8	4	80	1	230	7
9	0	90	0	240	6
10	0	100	0	250	6
More	0	More	0	More	0


Histogram
Frequency
25
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
10
More
exam1
Frequency


Histogram
Frequency
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
More
exam2
Frequency


Histogram
Frequency
8
7
6
5
4
3
2
1
0
150
160
170
180
190
200
210
220
230
240
250
More
earning
Frequency

1.b

Sample Estimators of Population Parameters

http://europa.eu.int/comm/trade/issues/bilateral/dataxls.htm

Mean $$\bar{X} = \frac{\sum X_i}{N}$$ First Moment: μ_1

Variance $$\mathrm{var}(X) = \frac{\sum (X_i - \bar{X})^2}{N-1}$$ Second Moment: μ_2

Standard Dev $$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N-1}}$$ Skewness (Pearson)= (Mean-Mode)/stand.dev.

Skewness: $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \mu_3 = \frac{\sum (X_i - \bar{X})^3}{N}$$

Kurtosis: $$\beta_2 = \frac{\mu_4}{\mu_2^2} \qquad \mu_4 = \frac{\sum (X_i - \bar{X})^4}{N}$$

Normality implies: $\sqrt{\beta_1} = 0 \quad \beta_2 = 3$

http://www.stats4schools.gov.uk/ http://www.eustatistics.gov.uk/yearbook.asp

Means, standard deviations and correlations (using score06.xls)
The sample is 1 - 48
Means

Exam 1	Exam 2	Earning
57.000	5.5313	207.08

Standard deviations (using T-1)

Exam 1	Exam 2	Earning
11.420	1.3968	28.158

Correlation matrix:

	Exam 1	Exam 2	Earning
Exam 1	1.0000	0.75784	-0.25572
Exam 2	0.75784	1.0000	-0.16489
Earning	-0.25572	-0.16489	1.0000

Coefficient of variation (CV) is the standard deviation relative to the mean.

$CV = \frac{\sigma}{\bar{X}} \times 100$. This is useful for comparing variable in different series.

1.c Coefficient of variation: ratio of standard deviation/mean

	Exam 1	*Exam 2*
Exam 1	1.910482	
Exam 2	11.8375	127.7083

1.d Correlation matrix:

	Exam 1	Exam 2
Exam 1	1.0000	0.75784
Exam 2	0.75784	1.0000

1.e
Weighted average:

$\overline{X}_w = \dfrac{\sum X_w = \sum (w_1 X_1 + w_2 X_2)}{N}$ where $w_1 = 1, \; w_1 = 1 \; w_2 = 0.1$ X_1 and X_2 are scores in exam1 and exam2 respectively. The weighted average: $\overline{X}_w = 11.23125$

Variance of the weighted score:
5.673257979

1.f

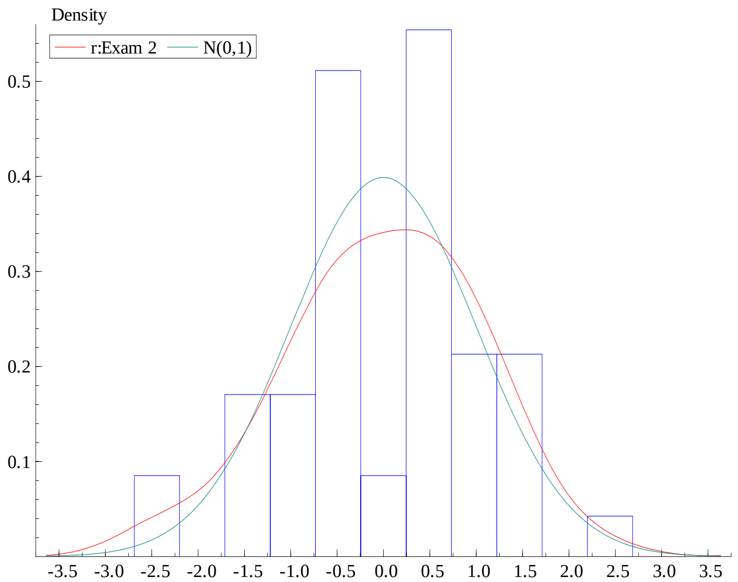
Exam 2 = + 22.73 + 6.196*Exam 1
(SE) (4.48) (0.786)

EQ(1) Modelling Exam 2 by OLS-CS (using Score06new.xls)
The estimation sample is: 1 to 48

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	22.7279	4.484	5.07	0.000	0.3584
Exam 1	6.19608	0.7865	7.88	0.000	0.5743

sigma	7.53165	RSS	2609.38672
R^2	0.574325	F(1,46) =	62.06 [0.000]**
log-likelihood	-164.005	DW	1.65
no. of observations	48	no. of parameters	2
mean(Exam 2)	57	var(Exam 2)	127.708

Normality test: Chi^2(2) = 0.86583 [0.6486]
Null of the normality of errors is not rejected.



hetero test: F(2,43) = 15.956 [0.0000]**
hetero-X test: F(2,43) = 15.956 [0.0000]**
RESET test: F(1,45) = 15.797 [0.0003]**
There is statistical evidence for hetroscedasticity.

For instance see below that the error term is significantly varies with x1 and x1square.

EQ(2) Modelling residuals by OLS-CS (using Score06new.xls)
The estimation sample is: 1 to 48

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	**-18.8723**	**6.145**	**-3.07**	**0.004**	**0.1733**
ex1square	**-0.997558**	**0.2510**	**-3.97**	**0.000**	**0.2598**
Exam 1	**9.27424**	**2.432**	**3.81**	**0.000**	**0.2443**

sigma 6.55131 RSS 1931.38663
R^2 0.259831 F(2,45) = 7.898 [0.001]**
log-likelihood -156.784 DW 1.79
no. of observations 48 no. of parameters 3
mean(residuals) 2.66454e-015 var(residuals) 54.3622

Normality test: Chi^2(2) = 4.4373 [0.1088]
hetero test: F(3,41) = 3.3671 [0.0275]*
hetero-X test: F(4,40) = 2.4964 [0.0579]
RESET test: F(1,44) = 2.4058 [0.1281]

residuals = - 18.87 - 0.9976*ex1square + 9.274*Exam 1
(SE) (6.14) (0.251) (2.43)

1.g. Score in exam 2: $X_2 = \alpha + \beta X_1 + \varepsilon$

If 6 : Exam 2 = + 22.73 + 6.196*Exam 1 = 22.73 + 6.196*6 = 59.906
If 8: Exam 2 = + 22.73 + 6.196*Exam 1 = 22.73 + 6.196*8 =72.298

1.h The estimation sample is: 1 to 48

Earning = + 242.7 + 1.369*Exam 1 - 0.7574*Exam 2
(SE) (20.7) (4.45) (0.544)

EQ(3) Modelling Earning by OLS-CS (using Score06new.xls)
The estimation sample is: 1 to 48

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	242.683	20.66	11.7	0.000	0.7541
Exam 1	1.36870	4.448	0.308	0.760	0.0021
Exam 2	-0.757373	0.5440	-1.39	0.171	0.0413

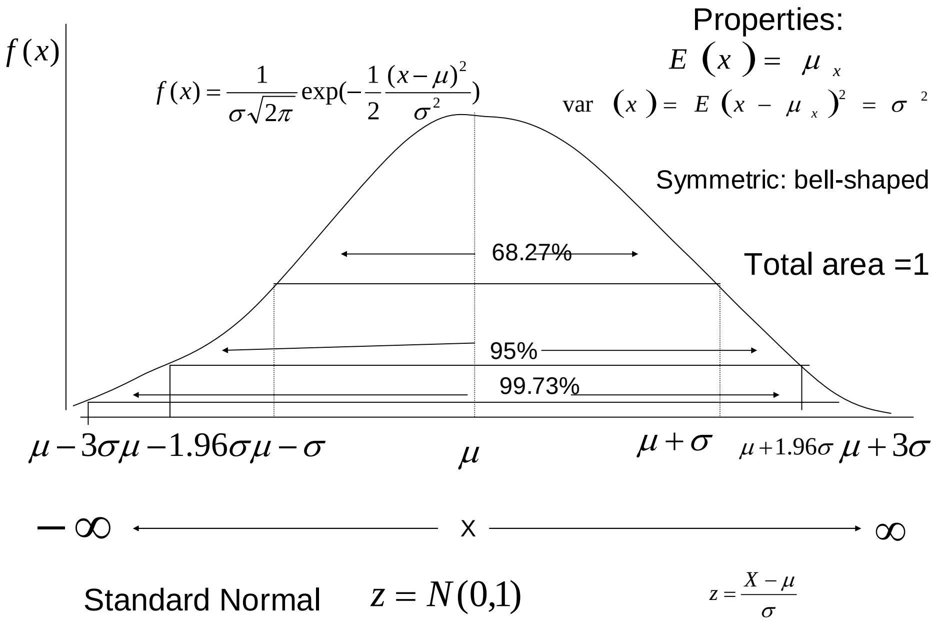
sigma 27.7911 RSS 34755.6345
R^2 0.0673551 F(2,45) = 1.625 [0.208]
log-likelihood -226.147 DW 2.07
no. of observations 48 no. of parameters 3
mean(Earning) 207.083 var(Earning) 776.368

Normality test: Chi^2(2) = 3.2243 [0.1995]
hetero test: F(4,40) = 1.3676 [0.2624]
hetero-X test: F(5,39) = 1.2220 [0.3172]
RESET test: F(1,44) = 0.19637 [0.6598]

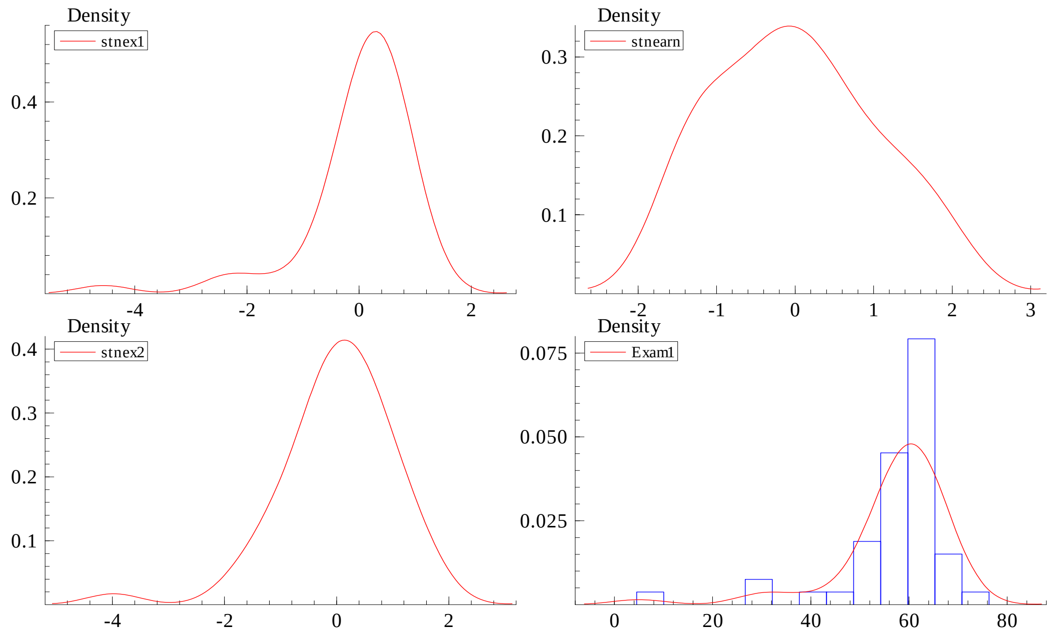
Any random distribution can be modified to standard normal distribution:

$$z = \frac{X - \overline{X}}{\sigma}$$

Normal Distribution of X: Bell Shaped Distribution



Modifying the scores to a standard normal distribution.



1.i student with more efforts can earn more marks and teachers can have more liberal or less liberal approaches to marking.

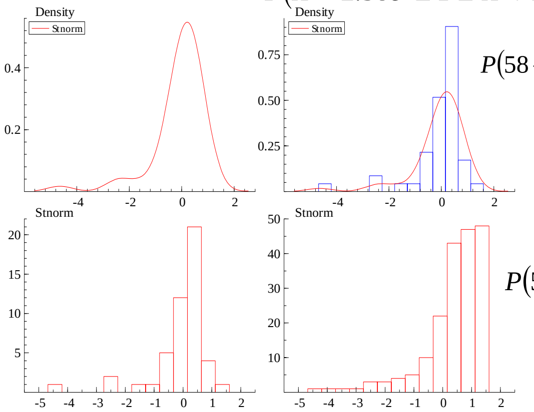
Hypotheses mean score in exam 1 is 6 and 58 in exam 2. The degrees of freedom $(n-1)$ is 48-1= 47. Critical or theoretical value for this is $t_{45,0.01} \approx 2.65$.

For score in exam 1: $t_{n-1} = \frac{\overline{X} - \mu}{s_x / \sqrt{n}} = \frac{5.5313 - 6}{1.3968} = \frac{0.4687}{1.3968} = 0.03355 \leq 2$

For score in exam 2: $t_{n-1} = \frac{\overline{X} - \mu}{s_x / \sqrt{n}} = \frac{57.0 - 58}{11.420} = \frac{-1}{11.420} = -0.0876 \leq |2|$

This test does not reject the null. There we can accept true mean scores in these exams are 6 and 58.

$$P\left(\overline{X} - 2.58\sigma \leq z \leq \overline{X} + 2.58\sigma\right) = (1-\alpha) = 0.99$$



$$P(58 - 2.58 \times 1 \leq z \leq 58 + 2.58 \times 1) = (1-\alpha) = 0.99$$

$$P(55.42 \leq z \leq 60.58) = (1 - 0.01) = 0.99$$

Thus the true score lies between 55.42 and 60.58 with probability of error of 1 percent.

Research Methods (26343): Economics
Test Questions

Student number: BA or B.Sc. Date: Feb 7, 2006

1. Draw a Venn diagram and show how any four samples of 3 elements can be generated from a population that contains 6 elements. [8 marks]

2. Illustrate how *F* statistics can be used to test whether variances in two different samples are statistically different or not. [8 marks]

   [hint: $F = \frac{\mathrm{var}(X_1)/m_1}{\mathrm{var}(X_2)/m_2}$ with m_1 and m_2 degrees of freedom].

3. Show unbiasedness, efficiency and consistency for an estimator using the following diagram [8 marks]

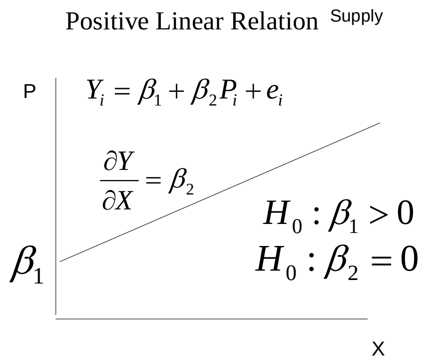
Marginal probability density for a random variable

4. Construct a 95 percent confidence interval for a Normally distributed variable whose mean 25 and standard deviation is 2.

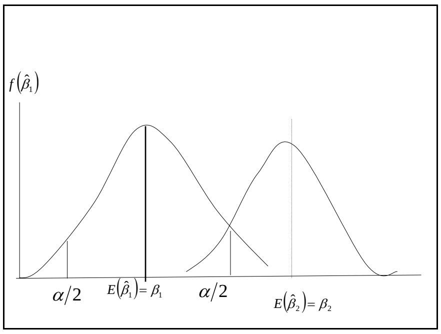
[Hint: $P(\bar{X} - z_\alpha \sigma < \mu < \bar{X} + z_\alpha \sigma) = (1-\alpha)$; $z_\alpha = 1.96$]. [8 marks]

5. Draw a density function and show how one can do two sided or one sided tests on whether a variable belongs to a normal distribution? [8 marks]

6. Show how one can compute the price elasticity of supply from the estimates of the following regression model. [8 marks]



7. Use following diagrams to show how a researcher can raise probability of making type II error while trying to minimise type I error. [8 marks]



8. Observe the following statistics on the trade ratio (trd) and GDP per capita (gdppc) for 155 countries. Are both of them normally distributed base on skewness, Kurtosis statistics and Chi-square statistic? [8 marks]

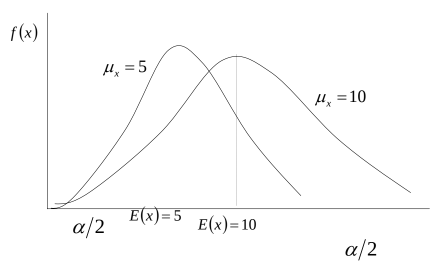
Normality test for trd		Normality test for gdppc	
Observations	155	Observations	155
Mean	36.875	Mean	24469.
Std.Devn.	15.186	Std.Devn.	7377.5
Skewness	-0.089341	Skewness	0.78516
Excess Kurtosis	-1.5212	Excess Kurtosis	0.49015
Minimum	11.239	Minimum	11827.
Maximum	66.325	Maximum	44830.
Asymptotic test: Chi^2(2) =	15.150 [0.0005]**	Asymptotic test: Chi^2(2) =	17.477 [0.0002]**
Normality test: Chi^2(2) =	36.316 [0.0000]**	Normality test: Chi^2(2) =	20.451 [0.0000]**

9. In a sample of 48 employees average number of times an employee takes a public transport was found to be 5.53 out of ten possible trips with standard error of 1.4. Hypothetically this mean is believed to be 6. Determine using t-test whether sample mean 5.53 is statistically significant or not. [8 marks]

Critical t-value for the 1% level of significance for 48 degrees of freedom is 2.

10. Consider a situation where you assumed that true mean of a population was 5 instead of 10 which was its true mean. If you infer from sample assuming that 5 was indeed a true mean what short of error are you making? [8 marks]

What happens when the hypothesis is wrong?



11. Study the following estimate of demand for a product estimated using Excel. Based on *t*, *F* tests and *R*-square statistics determine whether the coefficients and the overall model is statistically significant or not. [8 marks]
Critical value of F for (9,1) degrees of freedom is 9.78.

Estimation of a linear demand Model in Excel

SUMMARY OUTPUT				
Regression Statistics				

Multiple R	0.993505637			
R Square	0.987053452			
Adjusted R Square	0.986918592			
Standard Error	6.504250497			
Observations	98			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	1	309636.7	309636.7	7319.104
Residual	96	4061.306	42.30527	
Total	97	313698		
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	318.8933546	1.526653	208.884	2E-129
Price	-3.96756482	0.046376	-85.5518	1.98E-92

12. X is a random variable, μ is its mean, and σ its standard deviation. What are the mean and variance of a normalized random variable? [6 marks]
    a. mean is zero and variance is 1
    b. mean is 1 and variance is zero
    c. both mean and variance are zero
    d. both mean and variance are 1

13. Income is normally distributed with mean μ and variance σ^2. A surveyor can be 99.73 percent certain about the mean income within
    a. one standard errors $(\mu \pm \sigma)$
    b. two standard errors $(\mu \pm 2\sigma)$
    c. three standard errors $(\mu \pm 3\sigma)$
    d. four standard errors $(\mu \pm 4\sigma)$

[6 marks]

Research Methods Economics

1. X is a random variable, μ is its mean, and σ its standard deviation. What are the mean and variance of a normalized random variable?
   a. mean is zero and variance is 1
   b. mean is 1 and variance is zero
   c. both mean and variance are zero
   d. both mean and variance are 1

Answer: a

2. What is the level of significance α in the confidence interval $P(-1.96 \le z \le 1.96) = (1-\alpha) = 0.95$ for a randomized normal variable, $z \sim N(0,1)$?
   a. 1% b. 5% c. 2.5% d. 10%

Answer: b.

3. What are the values of skewness and kurtosis for a normally distributed random variable?
   a. 0 and 3 b. 3 and 0 c. 3 and 3 d. 0, 0

Answer: a

4. There are 11 regions in the UK. You want to design a sample taking four regions at a time. How many samples can you design from these 11 regions?

   a. 462 b. 330 c. 360 d. 400

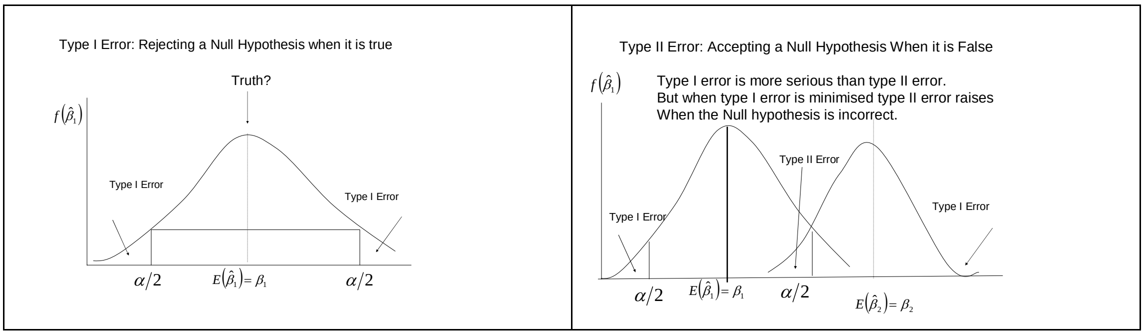
   Hints: $Cn = \frac{n!}{(n-r)!r!}$

Answer: b

5. Income is normally distributed with mean μ and standard error σ. A surveyor can be 99.73 percent certain about the mean income within
   a. one standard errors $(\mu \pm \sigma)$ b. two standard errors $(\mu \pm 2\sigma)$
   c. three standard errors $(\mu \pm 3\sigma)$ d. four standard errors $(\mu \pm 4\sigma)$

Answer: c

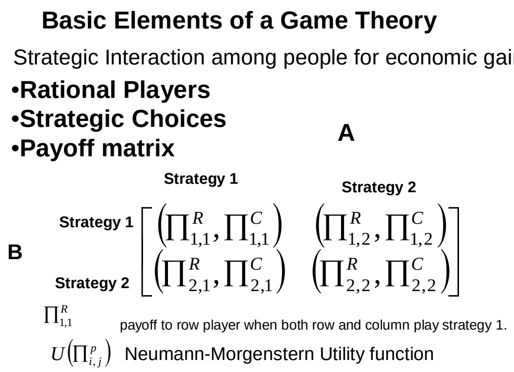
6. What happens when you increase the level of significance in a test?



   a. Type I error increases and Type II error falls.
   b. Type I error falls but Type II error raises
   c. Both Type I and Type II error raise
   d. Only Type I error falls but the Type II error remains the same.

Answer: b

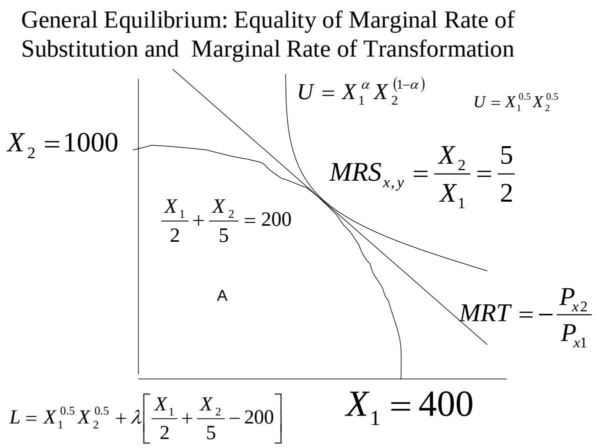
7. When is a strategy dominant for both A and B players in a game?



a. When A gains more than B
b. When B gains more than A
c. When none of them does better by choosing any other strategy
d. When A agrees to compensate for loss of B

Answer: c

8. General equilibrium in the economy is the system of prices where



a. where consumer maximize their utility but the producers don't
b. where producers maximize profits but consumers don't
c. where some markets still do not clear
d. where all markets clear and both consumer and producer optimise given the constraints they face

Answer: d.

9. What best describes the t-test, $t_{n-k-1} = \frac{\overline{X} - \mu}{s_x / \sqrt{n}}$?

a. It tests whether the sample mean compares to population mean
b. It tests whether the sample variance compares the population variance
c. It tests whether both the sample mean variance compare to the population mean and variance
d. It tests whether either of the sample mean variance compare to the population mean or the variance

Answer: a.

10. What is the χ^2 test used for? $\chi^2 = \sum_k X_k^2 \quad X_i \sim N(0,1)$

a. It tests whether the sample mean of a square of a normally distributed variable compares to its population counterpart

b. It tests whether the sample variance of a square of a normally distributed variable compares to its population counterpart
c. It tests whether the sample sum of a square of a normally distributed variable compares to its population counterpart
d. It tests whether either of the sample mean variance compare to the population mean or the variance

Answer: c

11. What does the F test, $F = \frac{V_1/m_1}{V_2/m_2} \sim F_{(m_1,m_2)}$ tests for?

a. It tests whether the means are different between two samples
b. It tests whether the variances differ between two samples
c. It tests whether both the sample mean variance differ between two samples
d. It tests whether the mean in one sample compares to variance in another sample

Answer: b.

Problem 3

Estimation, Hypothesis Testing and Confidence Intervals

1. Numbers of years of work among employees in a local labour market is assumed to be normally distributed. The mean years of work was found to be 25 with a standard deviation 2 in a survey of 200 individuals. Using this information construct 90, 95 and 99 percent confidence intervals for years of work among employees in this market (Hints, Z = 1.645, Z = 1.960, Z = 2.576 for 10, 5 and 1 percent level of significance respectively).

2. Formulate a hypothesis regarding possible relation between consumption, income and the interest rate. Regression estimate of consumption on income generated slope coefficient 0.95 and its standard error was 0.33 in a survey of 40 students.

i) Do you accept or reject your hypothesis with 1 percent level of significance?
ii) Construct 90, 95 and 99 percent confidence intervals for this slop coefficient.
iii) What are type I and type II errors in this example?

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

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**

(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 and expected signs of the parameters in this model.

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

(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 $\overline{R}^2$.

(d) Estimate the variance of the error term and the slope coefficient. Explain its importance.

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

(f) Build 95 percent confidence interval for estimate of slope and intercept terms.

(g) Discuss how reducing type I error may cause increase in type II errors.

(h) Calculate the elasticity of demand for football around the mean of Y and X_1.

(i) 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 while introducing an additional variable.

(j) Estimate the parameters of that multiple regression model.

(k) What is your prediction of the number of tickets sold if X_1 = 5 and X_2 =4?

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.

Answers to Problem 3

1. For a variable with standard normal distribution, $z = \frac{\overline{X} - \mu}{\sigma}$, where $\overline{X}$ is the sample mean σ sample standard deviation and μ the population mean the confidence interval for μ for α level of significance is given by

$$P\left(\overline{X} - z_{\alpha}\sigma < \mu < \overline{X} + z_{\alpha}\sigma\right) = (1 - \alpha)$$

The 90 percent confidence interval: $P(25 - 1.645 \times 2 < \mu < 25 + 1.645 \times 2) = (1.0 - 0.1) = 0.9$

The 95 percent confidence interval:
$P(25 - 1.196 \times 2 < \mu < 25 + 1.196 \times 2) = (1.0 - 0.05) = 0.95$

The 99 percent confidence interval:
$P(25 - 2.576 \times 2 < \mu < 25 + 2.576 \times 2) = (1.0 - 0.01) = 0.99$

2.

$$t_{calc} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\text{var}(\hat{\beta}_2)}} = \frac{0.95}{0.33} = 2.879 \quad t_{40,0.01} = 2.704$$

$H_0 : \beta = 0$ $H_1 : \beta \neq 0$; $t_{calc} > t_{40,0.01}$ Therefore reject the null. The statistical evidence does not support the hypothesis that the slope coefficient equals zero. In other words the estimated coefficient is statistically significant.

$$P\left(t_c \leq \frac{\hat{\beta} - \beta}{\sqrt{\text{var}(\hat{\beta})}} \leq t_c \right) = (1 - \alpha)$$

$$P\left(\hat{\beta} - t_c SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_c SE(\hat{\beta})\right) = (1 - \alpha)$$

From t-table critical values of t $t_{40,0.01} = 2.704$; $t_{40,0.05} = 2.201$ and $t_{40,0.1} = 1.684$.

The 90 percent confidence interval:

$P(0.95 - 1.684 \times 0.33 \leq \beta \leq 0.95 + 1.684 \times 0.33) = (1 - 0.1) = 0.9$

The 95 percent confidence interval:

$P(0.95 - 2.201 \times 0.33 \leq \beta \leq 0.95 + 2.201 \times 0.33) = (1 - 0.05) = 0.95$

The 90 percent confidence interval:

$P(0.95 - 2.704 \times 0.33 \leq \beta \leq 0.95 + 2.704 \times 0.33) = (1 - 0.01) = 0.90$

3. 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.

Observations	Y	X_1	X_2	YX_1	YX_2	X_1X_2	X_1^2	X_2^2	Y^2
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.

(l) 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?

Answer:

- A simple regression Model

$Y_i = \beta_1 + \beta_2 X_{1,i} + e_i$

Main assumptions about the error term e_i are following:

- Mean of e_i is zero for every value of x_i , $E[e_i] = 0$
- variance of e_i is constant $\text{var}[e_i] = \sigma^2$ for every ith observation

- $\text{cov}(e_i e_j) = 0 \quad for\ all \quad i \neq j$; this also means there is no autocorrelation or heteroscedasticity; errors are homoscedatic and independent of each other
- there is no correlation between e_i and the explanatory variable x_i ; $E[e_i x_i] = 0$
- explanatory variable, x_i , is exogenous or non-stochastic or it is not random
- variance of the dependent variable is equal to the variance of the error term $\text{var}(y_i) = \text{var}[e_i] = \sigma^2$

Expected sign of is negative ($\frac{\partial Y}{\partial X_1} = \beta_2 < 0$) and β_1 is intercept term expected to be positive. This would show the demand for the number of tickets if the price level was set to zero.

(m) Estimate the slope and intercept parameters. Use cross products and squared terms provided for you in the above table.
Answer: The easiest way to estimate regression with one explanatory variable is by using the equation in deviation form. $y_i = \beta_2 x_{1,i} + e_i$. This implies

$$\hat{\beta}_2 = \frac{\sum_i x_{1i} y_i}{\sum_i x_{1i}^2}$$

where using the deviation formula and the information given in the table

$\sum_i x_{1i} y_i = \sum_i (X_{1i} - \overline{X}_1)(Y_i - \overline{Y}) = \sum_i X_{1i} Y_i - N\overline{X}_1\overline{Y}$ = 97-7(4) (5)= 97-140 = -43

where $\overline{X}_1 = \frac{\sum X_1}{N} = \frac{35}{7} = 5$ and $\overline{Y} = \frac{\sum Y_i}{N} = \frac{28}{7} = 4$

and $\sum_i x_{1i}^2 = \sum_i (X_{1,i} - \overline{X}_1)^2 = \sum_i X_{1,i}^2 - N\overline{X}_1^2$ = 245 – 7(5)^2 = 245-175 = 70;

$$\hat{\beta}_2 = \frac{-43}{70} = -0.6143$$

Use this value for $\hat{\beta}_2$ and the means of Y (in 10 thousands), price of tickets (X_1) to estimate $\hat{\beta}_1$ as following:

$$\beta_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_1 = 4 - (-0.6143)5 = 4 + 3.0715 = 7.07$$

This the estimated relation can be written as:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{1,i} = 7.07 - 0.6143 X_{1,i}$$

$\frac{\partial Y}{\partial X_1} = \beta_2 = -0.6143 < 0$ as we expected. This gives us a downward sloping linear demand for tickets.

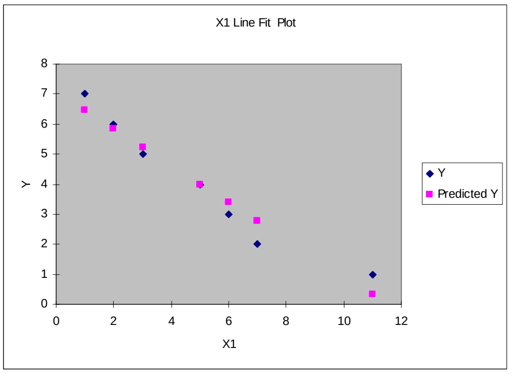
See estimates form Excel

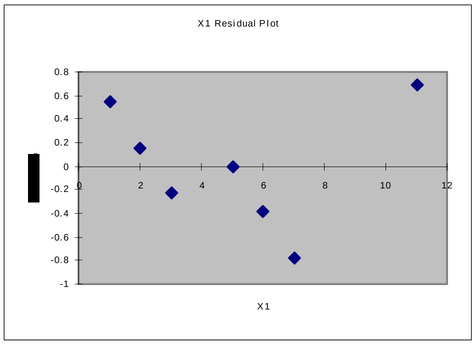
Regression Statistics	
Multiple R	0.971271
R Square	0.943367
Adjusted R Square	0.932041
Standard Error	0.563154
Observations	7

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	26.41429	26.41429	83.28829	0.000265
Residual	5	1.585714	0.317143		
Total	6	28			

	Coefficients	tandard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	7.071429	0.39821	17.75803	1.04E-05	6.047798	8.095059
X1	-0.61429	0.06731	-9.12624	0.000265	-0.78731	-0.44126





(n) 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]

Answer:

We can use the estimated for $\hat{\beta}_2$ to estimate the sum squared errors and sum squared for explained variation as following:

$$y_i = \beta_2 x_{1,i} + e_i \quad y_i = \hat{y}_i + e_i \Rightarrow \sum y_i^2 = \sum_i \hat{y}_i^2 + \sum_i e_i^2$$

$$\sum_i y_i^2 = \hat{\beta}_2^2 \sum_i x_{1,i}^2 + \sum_i \hat{e}_i^2 \Rightarrow 28 = (-0.6143)2 * 70 + \sum_i \hat{e}_i^2$$

$$\sum_i \hat{y}_i^2 = (-0.6143)^2 * 70 = 26.41 \text{ and } \sum_i \hat{e}_i^2 = 28 - 26.41 = 1.584$$

$$R^2 = \frac{\sum_i \hat{y}_i^2}{\sum y_i^2} = \frac{26.41}{28} = 94.32\%;$$

$$R^2 = 1 - \frac{\sum_i \hat{e}_i^2 \Big/ (N-K)}{\sum y_i^2 \Big/ (N-1)} = 1 - \frac{1.584/(7-2)}{28/(7-1)} = 1 - \frac{0.3168}{4.667} = 93.21\%$$

(o) Estimate the variance of the error term and the slope coefficient in this simple model. [10]

The unbiased estimator of the variance of the error term is given by

$$\text{var}(e_i) = \hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{N-K} = \frac{1.584}{5} = 0.3168;$$

The variance of the slope parameter is given by:

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum_i x_{1,i}^2} = \frac{0.3168}{70} = 0.00453$$

(p) Test whether the slope term is significant at 5% confidence level. [10]

Answer:

Calculate t value for the given estimates to test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$.

$t_{calc} = \frac{\beta_2 - \hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2)}} = \frac{-0.6143}{\sqrt{0.0043}} = -9.367$; The theoretical t value for $t_{5,0.05} = 2.571$.

Therefore calculate t-value is larger than the table t-value. This implies that the estimated coefficient $\hat{\beta}_2$ is statistically significant. Based on this evidence we can reject the null hypothesis which states that $H_0 : \beta_2 = 0$.

f. A 95 percent confidence interval means accepting the errors of 5 percent. Such confidence interval is given by the following formula

$$P\left(t_c \leq \frac{\hat{\beta}_k - \beta}{\sqrt{\text{var}(\hat{\beta}_k)}} \leq t_c \right) = 1 - \alpha$$

Or by reorganizing this

$$P\left(\hat{\beta}_k - t_c SE(\hat{\beta}_k) \leq \beta_k \leq \hat{\beta}_k + t_c SE(\hat{\beta}_k)\right) = 1 - \alpha = (1 - 0.05) = 0.95$$

$$P\left[\hat{\beta}_k - t_c SE(\hat{\beta}_k), \beta_k + t_c SE(\hat{\beta}_k)\right]$$

As shown above $\hat{\beta}_2 = \frac{-43}{70} = -0.6143$ $\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum_i x_{1,i}^2} = \frac{0.3168}{70} = 0.00453$,

$$SE(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = \sqrt{\frac{\hat{\sigma}^2}{\sum_i x_{1,i}^2}} = \sqrt{\frac{0.3168}{70}} = \sqrt{0.00453} = 0.06731$$

degrees of freedom $(N-K) = 7-2 = 5$ and $t_{5,0.05} = 2.571$

$$P\left(\hat{\beta}_k - t_c SE(\hat{\beta}_k) \le \beta_k \le \hat{\beta}_k + t_c SE(\hat{\beta}_k)\right) =$$

$$P(-0.6143 - 2.571 \times 0.06731 \le \beta_k \le -0.6143 + 2.571 \times 0.06731)$$

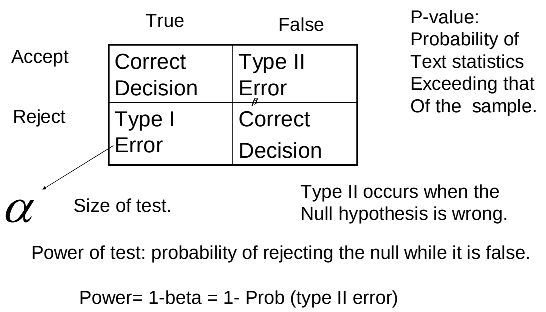
$$P(-0.6143 - 2.571 \times 0.06731 \le \beta_k \le 0.6143 + 2.571 \times 0.06731)$$

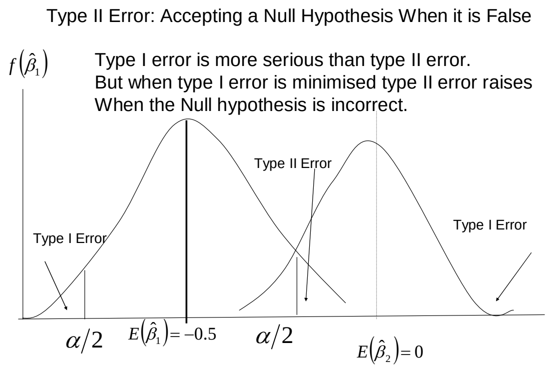
$$P(-0.78731 \le \beta_k \le -0.44126) = 1 - \alpha = (1 - 0.05) = 0.95$$

Thus we are 95 percent confident that the true β_k lies between -0.44126 and -0.78731.

(h). Consider following two diagrams to think about type I and Type II error. The type I error equals the level of significance chosen by the researcher. When one is happy being 95 percent confident then Type I error is 5 percent. The Type II error occurs when the Null hypothesis in incorrect. In the above test the Null tested was $H_0 : \beta_2 = 0$. If the right null was instead say $H_0 : \beta_2 = -0.5$ then having $H_0 : \beta_2 = 0$ instead of the true null will create a Type II error.

Type I and Type II Error in Hypothesis Testing





(h) elasticity of demand

The definition of elasticity of food expenditure on income is given by

$$\eta = \frac{\partial Y/Y}{\partial X/X} = \frac{\partial Y}{\partial X}\frac{X}{Y} = \hat{\beta}_2 \frac{\overline{X}}{\overline{Y}} = -0.6143\frac{7}{4} = -1.075$$

i. 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.

Answer :

This multiple regression model is obtained by adding the income term in the above model:

$$Y_i = \beta_1 + \beta_2 X_{1,i} + \beta_3 X_{2,i} + e_i$$

Expected sign of the slope coefficient in prices β_2 is negative as before($\frac{\partial Y}{\partial X_1} = \beta_2 < 0$) and that of income coefficient is positive $\frac{\partial Y}{\partial X_2} = \beta_3 > 0$, β_1 is intercept term expected to be positive. Additional assumption required is that the price and income variables should not be collinear. Otherwise it will create a Multicollinearity problem. No estimate of β_2 and β_3 is possible when X_1 and X_2 are exactly correlated.

(j) This model need to be modified to estimate the parameters of the multiple regression model given in (i). Again it is convenient to use the regression model in the deviation form.

$$\sum_i x_{1,i} y_i = \beta_2 \sum_i x_{1,i}^2 + \beta_3 \sum_i x_{1,i} x_{2,i} \qquad \text{(g.1)}$$

$$\sum_i x_{2,i} y_i = \beta_2 \sum_i x_{1,i} x_{2,i} + \beta_3 \sum_i x_{2,i}^2 \qquad \text{(g.2)}$$

$$\hat{\beta}_2 = \frac{\sum_i x_{1,i}x_{2,i}\sum_i x_{2,i}y_i - \sum_i x^2_{2,i}\sum_i x_{1,i}y_i}{\left(\sum_i x_{1,i}x_{2,i}\right)^2 - \sum_i x^2_{2,i}\sum_i x^2_{1,i}} \quad \text{(g.3)}$$

$$\hat{\beta}_3 = \frac{\sum_i x_{1,i}x_{2,i}\sum_i x_{1,i}y_i - \sum_i x_{2,i}y_i\sum_i x^2_{1,i}}{\left(\sum_i x_{1,i}x_{2,i}\right)^2 - \sum_i x^2_{2,i}\sum_i x^2_{1,i}} \quad \text{(g.4)}$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}_1 + \hat{\beta}_3\overline{X}_2 \quad \text{(g.5)}$$

$\sum_i x_{1,i}y_i = \sum_i \left(X_{1,i} - \overline{X}_1\right)\left(Y_i - \overline{Y}_i\right) = \sum_i X_{1,i}Y_i - N\overline{X}_1\overline{Y}_i$ = 97-7(4) (5)= 97-140 = -43

and $\sum_i x^2_{1i} = \sum_i \left(X_{1,i} - \overline{X}_1\right)^2 = \sum_i X^2_{1,i} - N\overline{X}^2_1$ = 245 – 7(5)^2 = 245-175 = 70;

$\sum_i x_{2,i}y_i = \sum_i \left(X_{2,i} - \overline{X}_2\right)\left(Y_i - \overline{Y}_i\right) = \sum_i X_{2,i}Y_i - N\overline{X}_1\overline{Y}_i = 126 - 7(4)(4) = 126 - 112 = 14$

$\sum_i x^2_{2,i} = \sum_i \left(X_{21,i} - \overline{X}_2\right)^2 = \sum_i X^2_{2,i} - N\overline{X}^2_2 = 126 - 7(4)^2 = 126 - 112 = 14$;

$\sum_i x_{1,i}x_{2,i} = \sum_i \left(X_{1,i} - \overline{X}_1\right)\left(X_{2,i} - \overline{X}_2\right) = \sum_i X_{1,i}X_{2,i} - N\overline{X}_1\overline{X}_2 = 117 - 7(5)(4) = 117 - 140 = -23$

$$\hat{\beta}_2 = \frac{\sum_i x_{1,i}x_{2,i}\sum_i x_{2,i}y_i - \sum_i x^2_{2,i}\sum_i x_{1,i}y_i}{\left(\sum_i x_{1,i}x_{2,i}\right)^2 - \sum_i x^2_{2,i}\sum_i x^2_{1,i}} = \frac{(-23)14 - 14(-43)}{(-23)^2 - (70)(14)} = \frac{280}{-451} = -0.6208$$

$$\hat{\beta}_3 = \frac{\sum_i x_{1,i}x_{2,i}\sum_i x_{1,i}y_i - \sum_i x_{2,i}y_i\sum_i x^2_{1,i} -}{\left(\sum_i x_{1,i}x_{2,i}\right)^2 - \sum_i x^2_{2,i}\sum_i x^2_{1,i}} = \frac{(-23)(-43) - 14(70)}{(-23)^2 - (14)70} = \frac{9}{-451} = -0.01996$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}_1 - \hat{\beta}_3\overline{X}_2 = 4 - (-0.6208)5 - (-0.01996)4 = 7.024$$

Estimated multiple regression line is

$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{1,i} + \hat{\beta}_3 X_{2,i} = 7.024 - 0.6208X_1 - 0.01996X_2$

Compare this result from the results obtained from the regression routine contained in Excel (tools/data analysis/regression) as given below.

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.971318					
R Square	0.943459					
Adjusted R Square	0.915188					
Standard Error	0.629116					
Observations	7					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	26.41685	13.20843	33.37255	0.003197	
Residual	4	1.583149	0.395787			
Total	6	28				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	7.184035	1.467634	4.894978	0.008073	3.109222	11.25885
X1	-0.62084	0.110843	-5.60112	0.004989	-0.92859	-0.31309
X2	-0.01996	0.247852	-0.08051	0.939696	-0.7081	0.668192
RESIDUAL OUTPUT						
Observation	*Predicted Y*	*Residuals*				
1	0.314856	0.685144				
2	2.798226	-0.79823				
3	3.379157	-0.37916				
4	3.980044	0.019956				
5	5.201774	-0.20177				
6	5.842572	0.157428				
7	6.48337	0.51663				

Compare this with the results in the PcGive

```
( 1) Modelling Y by OLS-CS (using Q3_football_gw.xls)
     The estimation sample is: 1 to 7

              Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant          7.18404      1.468     4.89   0.008   0.8569
X1              -0.620843     0.1108    -5.60   0.005   0.8869
X2             -0.0199557     0.2479  -0.0805   0.940   0.0016

sigma              0.629116  RSS              1.58314856
R^2                0.943459  F(2,4) =      33.37 [0.003]**
log-likelihood     -4.72984  DW                      1.8
no. of observations       7  no. of parameters         3
mean(Y)                   4  var(Y)                    4

Normality test:   Chi^2(2) =  0.11831 [0.9426]
Hetero test: not enough observations
Hetero-X test: not enough observations
RESET test:       F(1,3)   =   18.074 [0.0239]*

  Y =  + 7.184 - 0.6208*X1 - 0.01996*X2
(SE)    (1.47)  (0.111)    (0.248)
```

Matrix method could be used to find the solution as following:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} N & \sum_i X_{1,i} & \sum_i X_{21,i} \\ \sum_i X_{1,i} & \sum_i X_{1,i}^2 & \sum_i X_{1,i} X_{2,i} \\ \sum_i X_{21,i} & \sum_i X_{1,i} X_{2,i} & \sum_i X_{1,i}^2 \end{bmatrix}^{-1}}_{X'X} \underbrace{\begin{bmatrix} \sum_i Y_i \\ \sum_i Y_i X_{1,i} \\ \sum_i Y_i X_{2,i} \end{bmatrix}}_{X'Y}$$

$$= \left(X'X\right)^{-1} X'Y$$

(k) What is your prediction of the number of tickets sold if X_1 = 5 and X_2 =4? [5]

Answer:

$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{1,i} + \hat{\beta}_3 X_{2,i} = 7.024 - 0.6208 X_1 - 0.01996 X_2$

$7.024 - 0.6208(5) - 0.01996(4) = 4.00$

(l) 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.

Answer:

$$Y_i = \beta_1 + \beta_2 X_{1,i} + \beta_3 X_{2,i} + \gamma G + \theta G X_{2,i} + \psi_1 S_1 + \psi_2 S_2 + \psi_3 S_3 + e_i$$

$$G = \begin{cases} 1\ if\ \ male \\ 0\ otherwise \end{cases}$$

$$S_1 = \begin{cases} 1\ if\ \ summer \\ 0\ otherwise \end{cases}$$

$$S_2 = \begin{cases} 1\ if\ \ Autumn \\ 0\ otherwise \end{cases}$$

$$S_3 = \begin{cases} 1\ if\ \ \ w\text{int}\,er \\ 0\ otherwise \end{cases}$$

- Answer any five questions from this part.

PART B

(m) 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]

Answer

$$\hat{\beta}_2 = \frac{\sum_i x_{1,i}x_{2,i}\sum_i x_{2,i}y_i - \sum_i x_{2,i}^2 \sum_i x_{1,i}y_i}{\left(\sum_i x_{1,i}x_{2,i}\right)^2 - \sum_i x_{2,i}^2 \sum_i x_{1,i}^2} = \frac{0.01\sum_i x_{1,i}^2 \sum_i x_{1,i}y_i - 0.01\sum_i x_{1,i}^2 \sum_i x_{1,i}y_i}{0.01\left(\sum_i x_{1,i}\right)^2 - 0.01\left(\sum_i x_{1,i}\right)^2} = \frac{0}{0}$$

Thus the coefficient becomes indeterminate.

Problem 4
Time Series Analysis

1. Study monthly data on unemployment, retail price index and value of transactions in the stock market as contained in stocksandprice.xls file for the UK economy from 1973:4 to 2004:8. It is taken from the www.statistics.gov.uk using the Navidata. Answer following questions using PcGive.

a. Represent unemployment, retail price and value of stock transactions using line graphs.
b. Check whether these series are stationary using the unit root test. What will be the consequence of running a regression if variables are non-stationary?
c. Use GiveWin calculator to compute the first difference of the unemployment rate, retail price index and value of stocks.
d. Test for unit root in the first differences. Are they stationary?
e. Determine whether volume of transaction in stocks can be explained by retail prices and unemployment rates. Do all this using first differences.
f. Fit an autoregressive model to forecast volume of stocks in for next twenty months.
g. Write 95 percent confidence intervals for these forecasts.

2. There is a concern about the disparity of income among rich and poor households in the UK. Study the distributions on income and consumption by deciles of households for 2003 as given below.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	Total
Gross income	2587	3803	6699	11796	18627	26837	36106	46189	60047	111296	323987
Consumption	10286	15935	18247	18395	18631	21032	25662	31639	39886	72299	272012

Source: Economic Trends, 2003.

a. Use Excel to plot consumption and income in (i) a pie chart and (ii) a column or a bar chart.
b. Compute means and variances of income and consumption.
c. Calculate the individual and cumulative shares of income deciles of households.
d. Represent this income distribution using a Lorenz curve.
e. What is a Gini (G) coefficient? Which values of G represent perfect equality and which one represents perfect inequality?
f. Calculate the Gini-coefficients for above distribution of income and consumption. (hint: area of a triangle = 0.5× base × hight).

g. Calculate the benefit and transfers implicit in the above table. Briefly explain how such redistribution of income occurs in the real world situation.

3. Study the income distribution pattern of the United States as given in the following table and answer b-d of parts of the above questions.

Average income	Population Quintile	Income share
10186	1st (lowest)	0.035
25321	2nd	0.087
42492	3rd	0.146
66939	4th	0.23
145811	5th (highest)	0.501

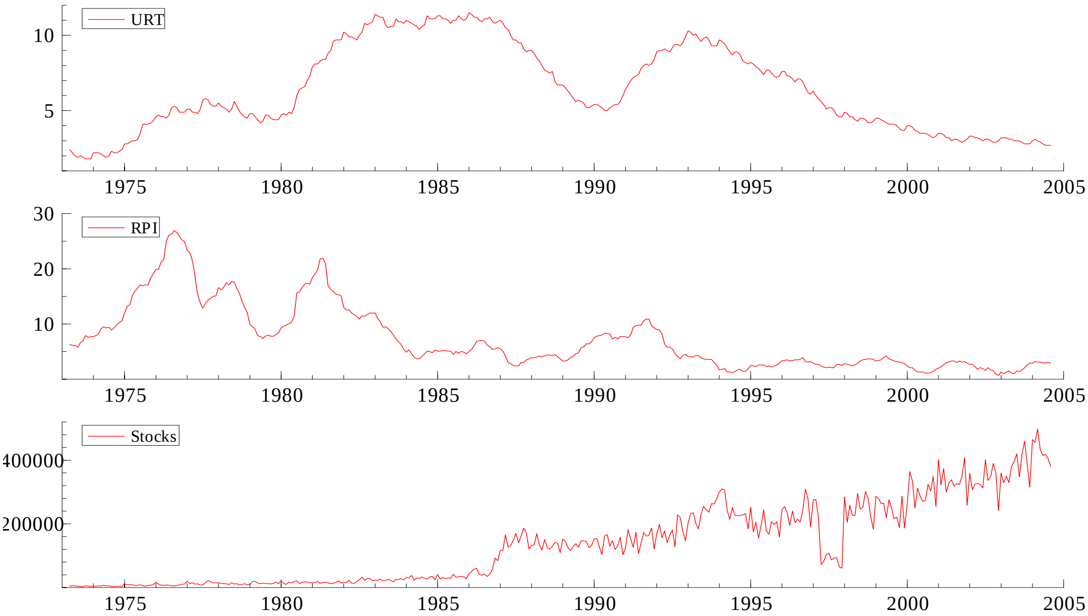
Source: DeNavas-Walt and Cleveland (2002) and Weal (2005) Economic Growth, p. 366.

How much difference do you observe in Gini coefficients between the US and the UK? Why?

Answers

1.a Unemployment, retail price and the value of stock in the UK 1973:4 to 2004:8.

A time series has four components: trend, cycle, season and random element.



Unit Root Tests of Unemployment Rate

Unit root exists in the level of unemployment rate

URT: ADF tests (T=373, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-1.143	0.99595	0.1969	-0.4586	0.6468	-3.237	
2	-1.165	0.99588	0.1967	-0.9016	0.3679	-3.242	0.6468
1	-1.209	0.99573	0.1966	6.955	0.0000	-3.245	0.6005
0	-0.9868	0.99630	0.2088			-3.127	0.0000

There is no unit root in the first difference

DURT: ADF tests (T=372, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-7.625**	0.40441	0.1948	-3.144	0.0018	-3.258	
2	-10.17**	0.28958	0.1971	0.4874	0.6263	-3.237	0.0018
1	-11.59**	[illegible]	[illegible]	[illegible]	[illegible]	-3.242	[illegible]
0	-13.52**	0.33903	0.1969			-3.245	0.0125

Output from the PcGive

1.b A series is non-stationary when its mean and variance are not constant over time or when these vary over time.

These series are non-stationary in their levels but stationary in the first differences. Use the ADF test to determine this.

Stocks: ADF tests (T=373, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-0.3988	0.99458	3.203e+004	-4.522	0.0000	20.76	
2	-0.8823	0.98777	3.287e+004	-5.243	0.0000	20.81	0.0000
1	-1.495	0.97873	3.402e+004	-8.258	0.0000	20.88	0.0000
0	-2.657	0.95946	3.698e+004			21.04	0.0000

URT: ADF tests (T=373, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-1.143	0.99595	0.1969	-0.4586	0.6468	-3.237	
2	-1.165	0.99588	0.1967	-0.9016	0.3679	-3.242	0.6468
1	-1.209	0.99573	0.1966	6.955	0.0000	-3.245	0.6005
0	-0.9868	0.99630	0.2088			-3.127	0.0000

RPI: ADF tests (T=373, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-2.175	0.98866	0.5856	0.8751	0.3821	-1.057	
2	-2.089	0.98918	0.5854	3.513	0.0005	-1.060	0.3821
1	-1.729	0.99095	0.5943	10.04	0.0000	-1.033	0.0016
0	-0.9820	0.99422	0.6695			-0.7971	0.0000

DURT: ADF tests (T=372, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-7.625**	0.40441	0.1948	-3.144	0.0018	-3.258	
2	-10.17**	0.28958	0.1971	0.4874	0.6263	-3.237	0.0018
1	-11.59**	0.30717	0.1969	0.9270	0.3545	-3.242	0.0068
0	-13.52**	0.33903	0.1969			-3.245	0.0125

DRPI: ADF tests (T=372, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-6.943**	0.56614	0.5882	-8.517e-005	0.9999	-1.048	
2	-7.458**	0.56614	0.5874	-0.6487	0.5170	-1.054	0.9999
1	-8.435**	0.55099	0.5869	-3.297	0.0011	-1.058	0.8108
0	-11.73**	0.45967	0.5947			-1.034	0.0112

DStocks: ADF tests (T=372, Constant; 5%=-2.87 1%=-3.45)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
3	-12.79**	-1.1570	3.208e+004	-0.4319	0.6661	20.77	
2	-18.02**	-1.2069	3.204e+004	4.594	0.0000	20.76	0.6661
1	-21.24**	-0.78935	3.290e+004	5.390	0.0000	20.81	0.0000
0	-29.65**	-0.40820	3.413e+004			20.88	0.0000

How to make a Non-Stationary Series to a Stationary Series

- Logs
- ratios
- First difference
- Second difference
- Third or higher order difference
- Cointegration
- Error Correction

1.d A spurious regression occurs in time series. It happens when the coefficients in a regression appear to be significant when there is no such relation. Such coefficients may just be showing the trend, cycle or season element inherent in a time series.
Make series stationary and run the regression in stationary series to find whether there is any significant relation.

See the following estimates on whether stock prices are determined by unemployment rate and the retail price indices. Coefficients are significant. Tempting to state that there is a significant relationship. But from above Stocks, URT and RPI are non-stationary variables. **This is an example of a spurious regression**.

Modelling Stocks by OLS (using stocksandprice.xls)
The estimation sample is: 1973 (5) to 2004 (7)

Stocks = + 3.271e+005 - 1.301e+004*URT - 1.452e+004*RPI
(SE) (1.23e+004) (1.53e+003) (751)

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	327122.	1.227e+004	26.7	0.000	0.6566
URT	-13010.5	1525.	-8.53	0.000	0.1636
RPI	-14521.9	750.9	-19.3	0.000	0.5013

sigma	85430.1	RSS	2.71496802e+012
R^2	0.540399	F(2,372) =	218.7 [0.000]**
log-likelihood	-4788.89	DW	0.206
no. of observations	375	no. of parameters	3
mean(Stocks)	136840	var(Stocks)	1.57526e+010

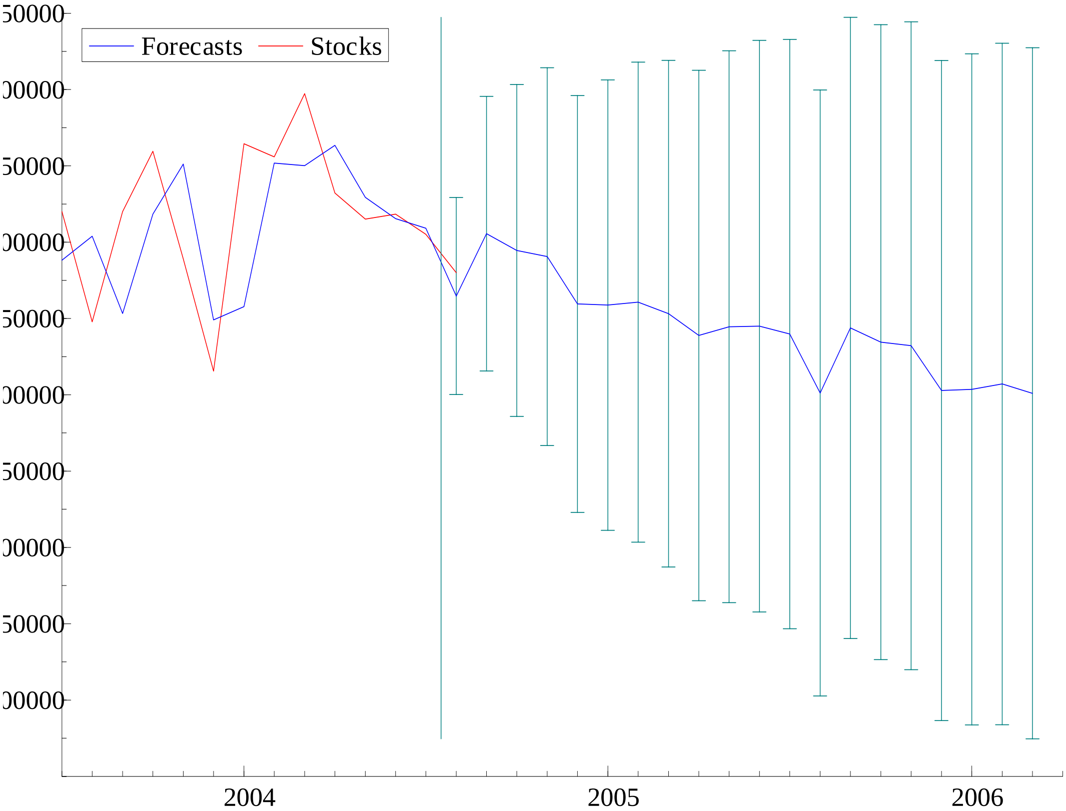
Check whether this relation is true for when series are made stationary taking the first differences. Now it is clear that these relations are not significant.

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	1052.82	1920.	0.548	0.584	0.0008
DURT	11591.2	9243.	1.25	0.211	0.0042
DRPI	-628.760	2890.	-0.218	0.828	0.0001

sigma 37173.2 RSS 5.14046688e+011
R^2 0.00423779 F(2,372) = 0.7916 [0.454]
log-likelihood -4476.85 DW 2.82
no. of observations 375 no. of parameters 3
mean(Dstocks) 1067.46 var(Dstocks) 1.37663e+009

This proves that apparent relation seen above is **indeed a spurious relation**.

In PcGive Choose econometric modelling/single equation dynamic model/ chose lag length of 1/ regress STOCKS on constant and Stocks_1/ choose OLS estimation./ Test graphics for actual and fitted. Forecast 20 periods/ dynamic forecasts and graph. To write results/



1.g Fitting an autoregressive model for stock price with seasonal elements.
Stocks = + 0.9721*Stocks_1 - 2.91e+004 + 8.015e+004*Seasonal
(SE) (0.0134) (6.09e+003) (8.21e+003)
+ 2.935e+004*Seasonal_1 + 3.609e+004*Seasonal_2 + 9003*Seasonal_3
(8.2e+003) (8.19e+003) (8.2e+003)
+ 3.835e+004*Seasonal_4 + 4.106e+004*Seasonal_5 + 3.158e+004*Seasonal_6
(8.13e+003) (8.13e+003) (8.13e+003)
+ 2.468e+004*Seasonal_7 + 4.427e+004*Seasonal_8 + 3.915e+004*Seasonal_9
(8.19e+003) (8.2e+003) (8.19e+003)
+ 3.35e+004*Seasonal_10
(8.19e+003)

Modelling Stocks by OLS (using stocksandprice.xls)
The estimation sample is: 1973 (5) to 2004 (7)

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Stocks_1	0.972102	0.01341	72.5	0.000	0.9355
Constant	-29102.2	6092.	-4.78	0.000	0.0593
Seasonal	80154.7	8205.	9.77	0.000	0.2086
Seasonal_1	29351.1	8196.	3.58	0.000	0.0342
Seasonal_2	36094.6	8195.	4.40	0.000	0.0509
Seasonal_3	9003.28	8196.	1.10	0.273	0.0033
Seasonal_4	38350.8	8132.	4.72	0.000	0.0579
Seasonal_5	41064.9	8130.	5.05	0.000	0.0658
Seasonal_6	31577.5	8129.	3.88	0.000	0.0400
Seasonal_7	24676.7	8195.	3.01	0.003	0.0244
Seasonal_8	44270.9	8197.	5.40	0.000	0.0746
Seasonal_9	39145.6	8194.	4.78	0.000	0.0593
Seasonal_10	33504.3	8194.	4.09	0.000	0.0442

sigma	32258	RSS	3.76689266e+011
R^2	0.936232	F(12,362) =	442.9 [0.000]**
log-likelihood	-4418.56	DW	2.74
no. of observations	375	no. of parameters	13
mean(Stocks)	136840	var(Stocks)	1.57526e+010

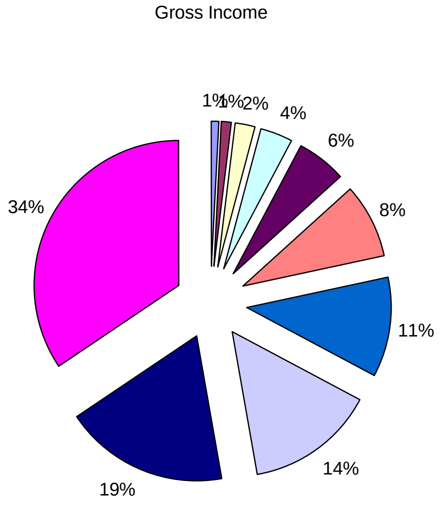
1-step (ex post) forecast analysis 2004 (8) to 2004 (8)
Parameter constancy forecast tests:
Forecast Chi^2(1) = 0.086143 [0.7691]
Chow F(1,362) = 0.082404 [0.7742]

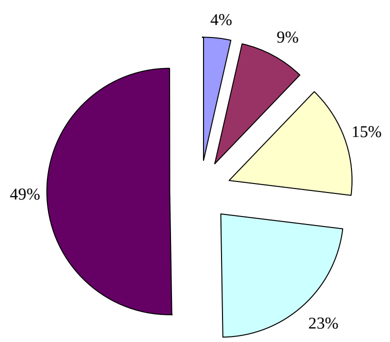
Dynamic (ex ante) forecasts for Stocks (SE based on error variance only)

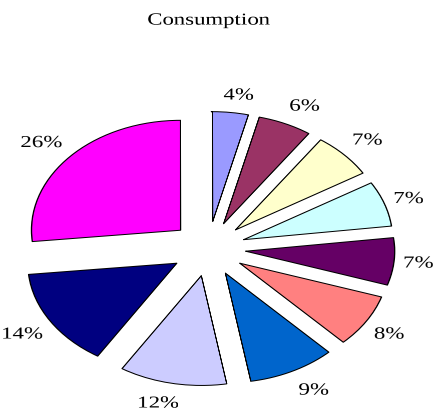
Horizon	Forecast	SE	Actual	Error	t-value
2004-8	364781.	3.226e+004	379990.	15209.0	0.471
2004-9	405657.	4.499e+004	.NaN	.NaN	.NaN
2004-10	394589.	5.434e+004	.NaN	.NaN	.NaN
2004-11	390573.	6.190e+004	.NaN	.NaN	.NaN
2004-12	359578.	6.827e+004	.NaN	.NaN	.NaN
2005-1	358796.	7.379e+004	.NaN	.NaN	.NaN
2005-2	360749.	7.865e+004	.NaN	.NaN	.NaN
2005-3	353160.	8.298e+004	.NaN	.NaN	.NaN
2005-4	338882.	8.688e+004	.NaN	.NaN	.NaN
2005-5	344597.	9.041e+004	.NaN	.NaN	.NaN
2005-6	345027.	9.362e+004	.NaN	.NaN	.NaN
2005-7	339803.	9.655e+004	.NaN	.NaN	.NaN
2005-8	301222.	9.925e+004	.NaN	.NaN	.NaN
2005-9	343871.	1.017e+005	.NaN	.NaN	.NaN
2005-10	334526.	1.040e+005	.NaN	.NaN	.NaN
2005-11	332186.	1.061e+005	.NaN	.NaN	.NaN
2005-12	302820.	1.081e+005	.NaN	.NaN	.NaN
2006-1	303621.	1.099e+005	.NaN	.NaN	.NaN
2006-2	307113.	1.116e+005	.NaN	.NaN	.NaN
2006-3	301021.	1.132e+005	.NaN	.NaN	.NaN

For Ex-Ante forecast choose test/forecast periods (20 above)/option/ write result instead of graphing.

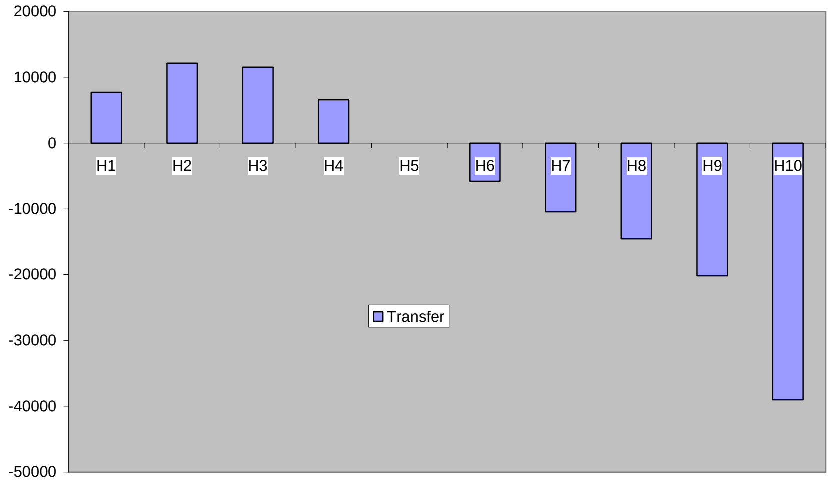
Distribution of Income and Consumption and Transfers from High to Low Income Households in the UK





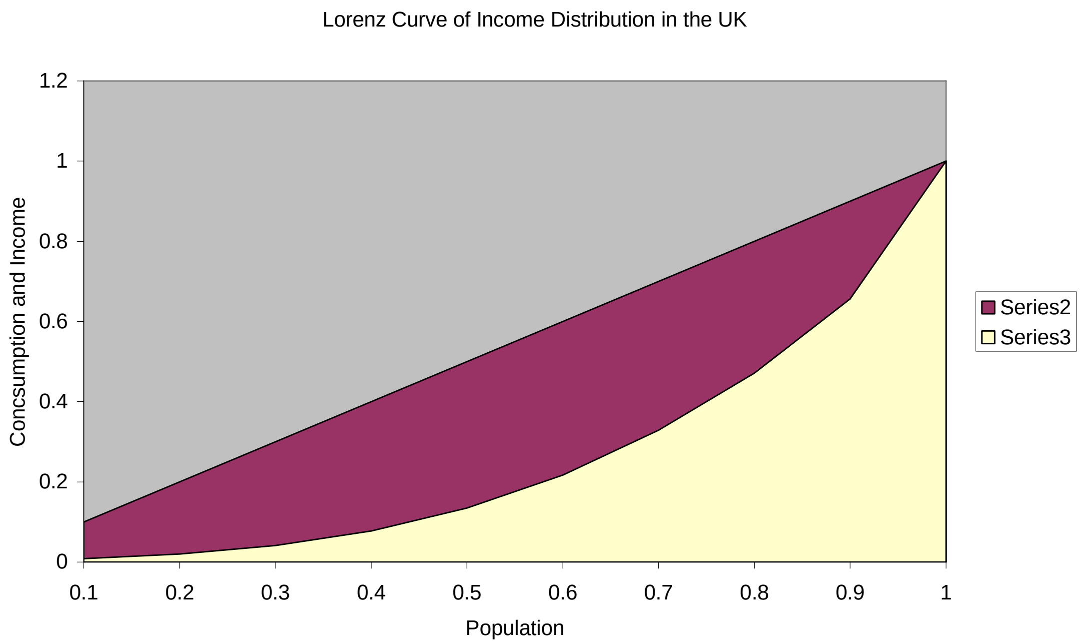


Benefit Transfer and Taxes by Households in the UK



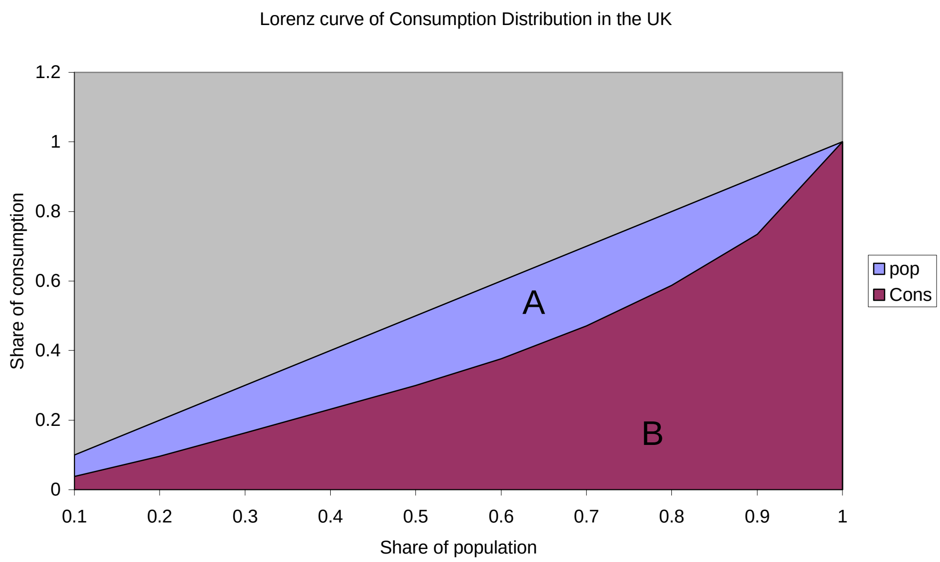
	B	C	D	E	F	G	H	I	J
	Consumption	Share by H	Cumshare	Difference	Population	Triangle	Rectangle	Area	equality
H1	10286	0.03781	0.03781		0.1	0.00189		0.00189	0.1
H2	15935	0.05858	0.0964	0.05858	0.2	0.00293	0.00378	0.00671	0.2
H3	18247	0.06708	0.16348	0.06708	0.3	0.00335	0.00964	0.01299	0.3
H4	18395	0.06763	0.2311	0.06763	0.4	0.00338	0.01635	0.01973	0.4
H5	18631	0.06849	0.2996	0.06849	0.5	0.00342	0.02311	0.02654	0.5
H6	21032	0.07732	0.37692	0.07732	0.6	0.00387	0.02996	0.03383	0.6
H7	25662	0.09434	0.47126	0.09434	0.7	0.00472	0.03769	0.04241	0.7
H8	31639	0.11631	0.58757	0.11631	0.8	0.00582	0.04713	0.05294	0.8
H9	39886	0.14663	0.73421	0.14663	0.9	0.00733	0.05876	0.06609	0.9
H10	72299	0.26579	1	0.26579	1	0.01329	0.07342	0.08671	1
Total	272012					0.05	0.29983	0.34983	
			Gini	G=(A/(A+B))	0	perfect equality		A	B
			Gini		0.30033			0.15017	0.34983
			Gini		1	perfect inequality			

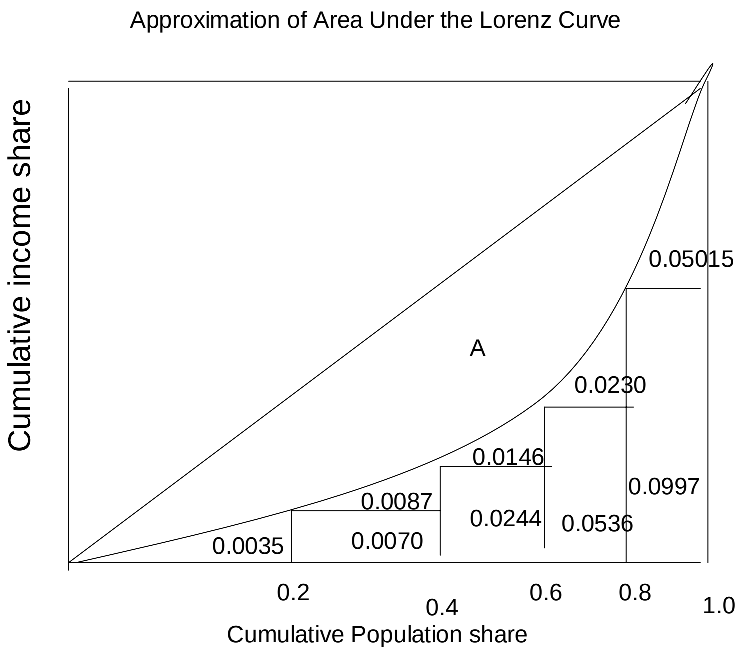
Excel formulas for calculations
Share of income by deciles for H2: +B2/B12; repeat this for all ten households.
Cumulative share for H2: +D2+C3
Difference of the cumulative share of H2: =+D3-D2
Population share of H2 =0.1+F2
Triangle for H2=+E3*(F3-F2)*0.5
Area of rectangle for H2=+(F3-F2)*D2
Area under the Lorenz curve = +G3+H3 ; Gini coefficient G =+(0.5-I12)/0.5



Steps for making a Lorenz curve:

1. prepare data by households 2. find cumulative share of income and population by deciles 3. select XY plot choosing income share and population shares 4. Select series and x axis. 4. change to area type graph as above.
2. steps to calculate Gini coefficient: a. divide the are below the Lorenz curve in ten equal parts b. divide each part into triangles and rectangles c. find the areas of rectangles and triangles d. sum up for the total area under the Lorenz curve area B E. Deduct B from 0.5 to find A. then use G = A/(A+B). A=0 in perfect equality and A=1 in perfect inequality; most other cases lie between these extremes.





Income group	Income	Ishare	CMIshare	Equality	pop	cmpop	Rectangle	Triangle	Total area
1st (lowest)	10186	0.035034	0.035034	0.2	0.2	0.2	0	0.003503	0.003503
2nd	25321	0.087089	0.122123	0.4	0.2	0.4	0.007007	0.008709	0.015716
3rd	42492	0.146147	0.268269	0.6	0.2	0.6	0.024425	0.014615	0.039039
4th	66939	0.23023	0.498499	0.8	0.2	0.8	0.053654	0.023023	0.076677
5th (highest)	145811	0.501501	1	1	0.2	1	0.0997	0.05015	0.14985
Total	290749	1				Total area under Lorenz			0.284785
							Gini		0.4304

Gini Coefficient for UK , 0.3003, is a lot lower than 0.4304 for the US economy. Therefore this data shows UK to have a lot more equal distribution of income than the US. Is this a good thing?

Problem 5
Heteroscedasticity and Autocorrelation

1. Take a simple linear regression model of the following form.

$$Y_i = \beta_1 + \beta_2 X_i + e_i \qquad (4)$$

Where the variance of the error term differs for different observations of X_i.

(a) Discuss how the graphical method be used to detect the heteroscedasticity?

(b) Analyse consequences of heteroscedasticity on the BLUE properties of the OLS estimators.

(c) Discuss how the Goldfeld and Quandt and Glesjer tests can be used to determine existence of the heteroscedasticity problem?

(d) Illustrate any two remedial measures of removing the heteroscedasticity when the variance $\sigma_i^2 = \sigma^2 X_i$ and σ is known and when it is unknown.

(e) From a sample of 6772 observations on pay work-hours and taxes contained in PAYHRTX.XLS determine whether heteroscedasticity exists or not on the basis of cross section estimates from the the PcGive. Feel free to use Shazam if you know and prefer it.

- (B.Sc.) Suggest remedial measures to remove heteroscedasticity in a model like above.

2. Consider a simple linear regression model.

$$Y_t = \beta_1 + \beta_2 X_t + e_t$$

Now assume that errors are correlated to each other over time with AR(1) process as:

$$e_t = \rho e_{t-1} + v_t$$

where v_t is identically and normally distributed error term with zero mean and constant variance, $v_t \sim N\left(0, \sigma^2\right)$.

(a) Illustrate how the graphical method can be applied to detect autocorrelation in a simple regression model like above?

(b) What are consequences of autocorrelation in a regression model? Show how the existence of such autocorrelation among the error terms affects the BLUE properties of the OLS estimators.

(c) Define and derive the Durbin-Watson test statistics. Show how it can test for existence or non existence of autocorrelation in a given estimation?

(d) How the autocorrelation can be removed if the ρ is known?

(e) What is a spurious regression? Why does it arise and how does it affect the usefulness of estimation from an OLS regression? What can be done to correct it?

Application:
Read data on growth rate of per capita GDP, exchange rate and inflation rates from the www.imf.org for year 1980 to 2003 for China, India, South Africa, UK, USA and Brazil as contained in PERCAP6.GLS. Test whether inflation and the exchange rate are the significant variables in explaining the growth rate of per capita output (in PPP) in these economies. Determine whether heteroscedasticity and autocorrelation exist in this regression using PcGive. Feel free to use Shazam if you know and prefer it.

suggest remedy for autocorrelation in a model like this.

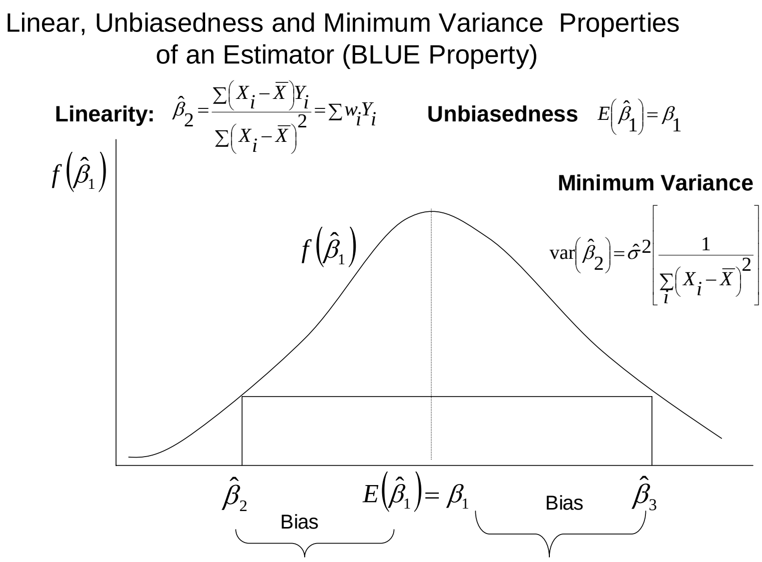
Tentative Answers - Tutorial 5
Some experiments with the Heteroscedasticity Problem in Work Hours and Pay Dataset

. Take a simple linear regression model of the following form.

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

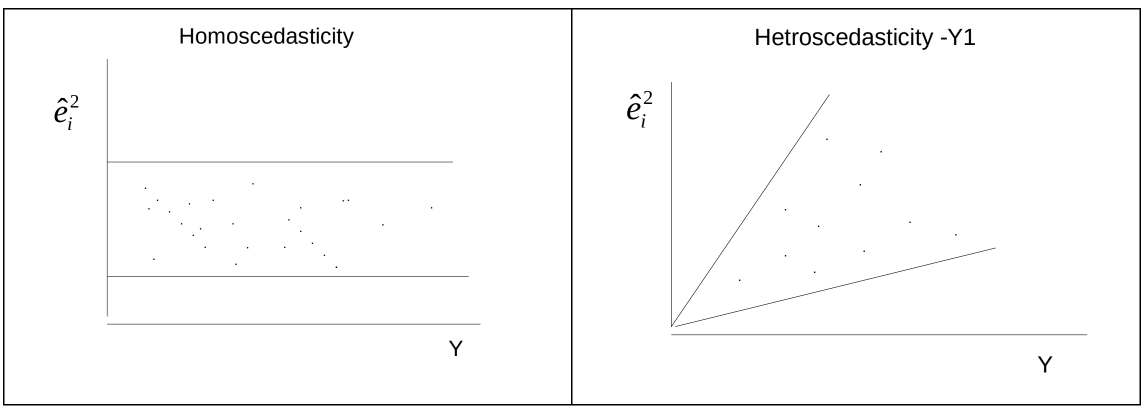
Where the variance of the error term differs for different observations of X_i.

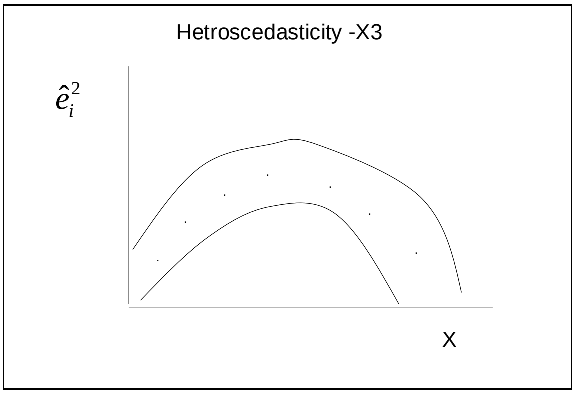
(f) When is the parameter β_2 unbiased? When is it efficient? When is it consistent?

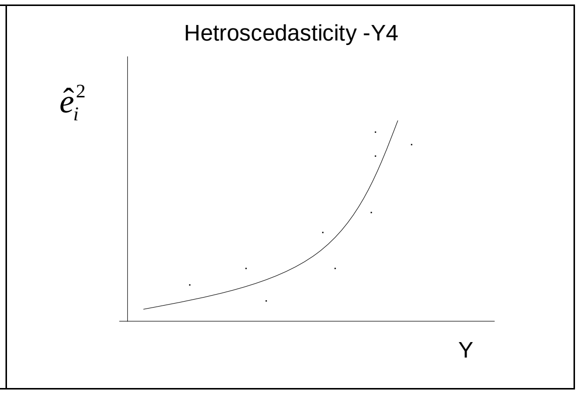


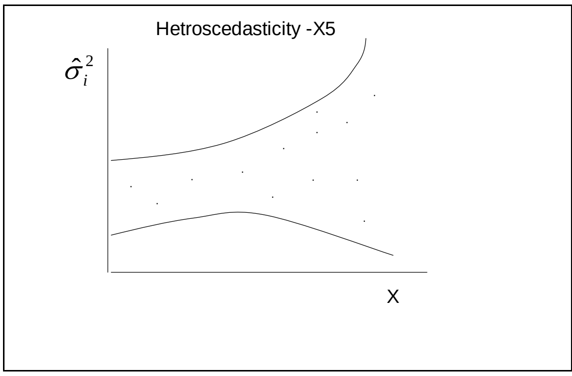
An estimator is consistent when the bias in estimation disappears as the sample size goes to infinity as following: $\underset{N \to \infty}{Bias}(\hat{\beta}) = 0$.

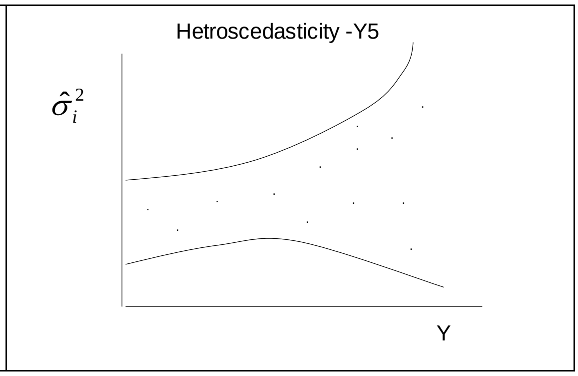
(g) Discuss how the graphical method be used to detect the heteroscedasticity?

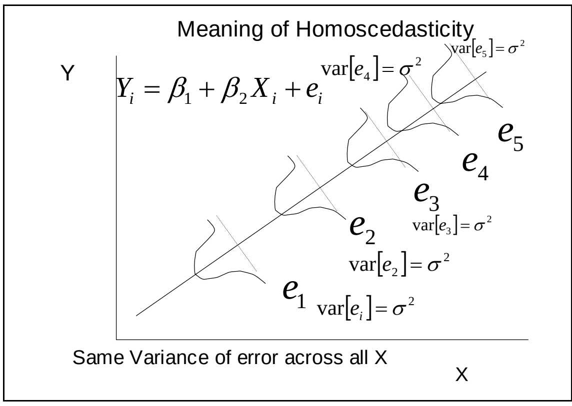


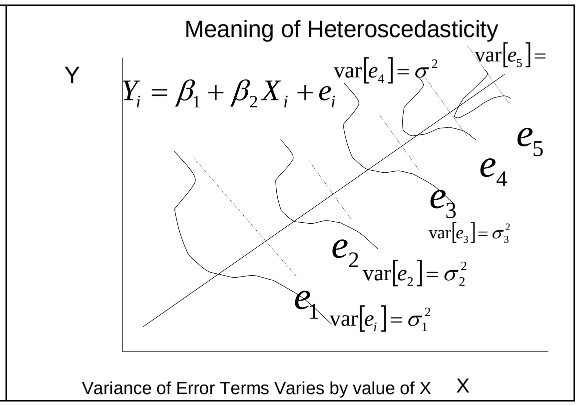












Causes:

- Learning: reduces errors; driving practice, driving errors and accidents
  typing practice and typing errors, defects in productions; improved machines
- Growth: saving and variance of saving increases with income
- Improved data collection: better formulas and good software
- Outliers affect the value of estimates
- Specification Errors and omitted variables:- in a demand model if you regress demand of a product to only its own price, there is a danger variables such as the prices of complements and income may appear in the error term.
- More heteroscedasticity exists in cross section than in time series data.

(h) Analyse consequences of heteroscedasticity on the BLUE properties of the OLS estimators.

Consequences of Hetroscedasticity

OLS Estimate is Unbiased.
But it is no longer efficient.

$$E\left(\hat{\beta}_2\right) = \beta_2$$

$$E\left(\hat{\beta}_2\right) = E\left[\sum w_i y_i\right] = E\left[\sum w_i\left(\beta_1 + \beta_2 x_i + e_i\right)\right]$$

$$= E\left[\sum w_i \beta_1 + \beta_2 \sum w_i x_i + \sum w_i e_i\right] = \beta_2$$

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sum_i \left(x_i - \bar{x}\right)^2 \sigma_i^2}{\left[\sum_i \left(x_i - \bar{x}\right)^2\right]^2} \qquad \text{not} \qquad \operatorname{var}\left[e_i\right] = \sigma^2$$

- OLS assumption: variance of e_i is constant $\operatorname{var}\left[e_i\right] = \sigma^2$ for every ith observation, $\operatorname{var}\left(\hat{\beta}_2\right) = \hat{\sigma}^2 \left[\frac{1}{\sum_i \left(x_i - \bar{x}\right)^2}\right]$
- but in case of heteroscedasticity variance of error is not constant:

$$\sigma_i^2 = \sigma^2 x_i$$

(i) Discuss how the Goldfeld and Quandt and Glesjer tests can be used to determine existence of the heteroscedasticity problem?

Goldfeld-Quandt test

Model $Y_i = \beta_1 + \beta_2 x_i + e_i$ (1)

Steps:

1. Rank observations in ascending order of one of the x variable
2. Omit c numbers of central observations leaving two groups with $\frac{n-c}{2}$ number of observations

3. Fit OLS to the first $\frac{n-c}{2}$ and the last $\frac{n-c}{2}$ observations and find sum of the squared errors from both of them.
4. Set hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ against } H_A : \sigma_1^2 \neq \sigma_2^2 .$$

5. compute $\lambda = \frac{ERSS_2 / df2}{ERSS_1 / df1}$ it follows F distribution.

Glejser test

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

There are several tests

$$|e_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|e_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|e_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|e_i| = \sqrt{\beta_1 + \beta_2 X_{i_i}^2} + v_i$$

In each case do t-test H_0: $\beta = 0$ against H_A: $\beta \neq 0$. If β is significant then that is the evidence of heteroscedasticity.

(j) Illustrate any two remedial measures of removing the heteroscedasticity when the variance $\sigma_i^2 = \sigma^2 X_i$ and σ is known and when it is unknown.

Remedial measures

Weighted Least Square and GLS when σ_i^2 known, divide the whole equation by σ_i

Apply OLS to transformed variables.

$$\frac{Y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{X_1}{\sigma_i} + \beta_2 \frac{X_2}{\sigma_i} + \beta_3 \frac{X_3}{\sigma_i} + \beta_4 \frac{X_4}{\sigma_i} + ... + \beta_k \frac{X_k}{\sigma_i} + \frac{e_i}{\sigma_i}$$

Variance of this transformed model equals 1.

Other examples: $Y_i = \beta_1 + \beta_2 x_i + e_i$ and assume $e_i^2 = \sigma^2 x_i^2$

$$\frac{Y_i}{x_i} = \frac{\beta_1}{x_i} + \beta_2 + \frac{e_i}{x_i}; \quad E\left(\frac{e_i}{x_i}\right)^2 = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2$$

(k) From a sample of 6772 observations on pay work-hours and taxes contained in PAYHRTX.XLS determine whether heteroscedasticity exists or not on the basis of cross section estimates from the the PcGive. Feel free to use Shazam if you know and prefer it.

Modelling PAY by OLS-CS (using PAYHRTX_sort.xls)
The estimation sample is: 1 to 6780
Dropped 8 observation(s) with missing values from the sample

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	307.966	10.63	29.0	0.000	0.1103
WHRS	0.209125	0.3788	0.552	0.581	0.0000
TAX	2.56524	0.02006	128.	0.000	0.7073

sigma 620.249 RSS 2.60409025e+009
R^2 0.707293 F(2,6769) = 8178 [0.000]**
log-likelihood -53152.3 DW 1.96
no. of observations 6772 no. of parameters 3
mean(PAY) 809.312 var(PAY) 1.31373e+006

Normality test: Chi^2(2) =3.3707e+005 [0.0000]**
hetero test: F(4,6764)= 696.91 [0.0000]**
hetero-X test: F(5,6763)= 564.35 [0.0000]**
RESET test: F(1,6768)= 306.40 [0.0000]**
These test reject homoscedastity.

PAY = + 308 + 0.2091*WHRS + 2.565*TAX
(SE) (10.6) (0.379) (0.0201)

Heteroscedasticity consistent standard errors

	Coefficients	SE	HACSE	HCSE	JHCSE
Constant	307.97	10.633	58.432	58.086	64.117
WHRS	0.20913	0.37878	0.28303	0.29554	0.29859
TAX	2.5652	0.020059	0.31737	0.31539	0.34842

	Coefficients	t-SE	t-HACSE	t-HCSE	t-JHCSE
Constant	307.97	28.965	5.2705	5.3019	4.8032
WHRS	0.20913	0.55210	0.73887	0.70760	0.70039
TAX	2.5652	127.88	8.0829	8.1337	7.3624

HACSE = heteroscedasticity and autocorrelation consistent standard error.
This corrected standard errors then can be applied to determine whether a coefficient is significant. This is also called White test.

(l) Suggest remedial measures to remove heteroscedasticity in a model like above.

Transform the model to purge the heteroscedastic errors as discussed in the remedial measure above.
when σ^2 unknown estimate σ^2 using the sample information and do the above procedures (**Gujarati** is a good text for Heteroscedasticity).

Shazam progam and results for heteroscedasticity.
ols pay whours tax/ resid=res1
diagnos/chowone=3000

```
      6772 OBSERVATIONS     DEPENDENT VARIABLE= PAY
 ...NOTE..SAMPLE RANGE SET TO:      1,   6772

  R-SQUARE =   0.7073     R-SQUARE ADJUSTED =   0.7072
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.38471E+06
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   620.25
                      ANALYSIS OF VARIANCE - FROM MEAN
                      SS          DF            MS                 F
 REGRESSION       0.62925E+10      2.       0.31462E+10          8178.239
 ERROR            0.26041E+10   6769.       0.38471E+06           P-VALUE
 TOTAL            0.88966E+10   6771.       0.13139E+07             0.000

VARIABLE    ESTIMATED   STANDARD   T-RATIO          PARTIAL STANDARDIZED ELASTICITY
   NAME     COEFFICIENT    ERROR    6769 DF    P-VALUE CORR.  COEFFICIENT  AT MEANS

WHOURS    0.20913     0.37878      0.55210     0.5809 0.0067  0.36310E-02  0.43316E-02
 TAX        2.5652     0.20059E-01   127.88     0.0000 0.8410  0.84106      0.61514
 CONSTANT   307.97      10.633       28.965     0.0000 0.3321   0.0000      0.38053

SEQUENTIAL CHOW AND GOLDFELD-QUANDT TESTS
    N1   N2    SSE1        SSE2       CHOW    PVALUE    G-Q       DF1  DF2 PVALUE
  3000 3772 0.93722E+09 0.14489E+10  205.98     0.000 0.8135     2997 3769 0.000
  CHOW TEST - F DISTRIBUTION WITH DF1=   3 AND DF2=6766
```

AS the Goldfeld and Quandt test shows variance of the two groups of the sample are not the same.

Transform the model by dividing every variable by the square of the residual
genr res2 =res1*res1
ols res2 whours tax
genr pay1 =pay/res2
genr whour1=whours/res2
genr tax1=tax/res2
*there is no heteroskedasticity in the transformed model
ols pay1 whour1 tax1/

```
      6772 OBSERVATIONS     DEPENDENT VARIABLE= PAY1
 ...NOTE..SAMPLE RANGE SET TO:      1,   6772

  R-SQUARE =   0.0435     R-SQUARE ADJUSTED =   0.0432
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.10677E+08
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   3267.6
 SUM OF SQUARED ERRORS-SSE=  0.72275E+11
 MEAN OF DEPENDENT VARIABLE =   54.610
 LOG OF THE LIKELIHOOD FUNCTION = -64405.4

 MODEL SELECTION TESTS - SEE JUDGE ET AL. (1985,P.242)
  AKAIKE (1969) FINAL PREDICTION ERROR - FPE =     0.10682E+08
     (FPE IS ALSO KNOWN AS AMEMIYA PREDICTION CRITERION - PC)
  AKAIKE (1973) INFORMATION CRITERION - LOG AIC =   16.184
  SCHWARZ (1978) CRITERION - LOG SC =              16.187
 MODEL SELECTION TESTS - SEE RAMANATHAN (1998,P.165)
  CRAVEN-WAHBA (1979)
     GENERALIZED CROSS VALIDATION - GCV =          0.10682E+08
  HANNAN AND QUINN (1979) CRITERION =              0.10693E+08
  RICE (1984) CRITERION =                          0.10682E+08
  SHIBATA (1981) CRITERION =                       0.10682E+08
  SCHWARZ (1978) CRITERION - SC =                  0.10714E+08
  AKAIKE (1974) INFORMATION CRITERION - AIC =      0.10682E+08

                     ANALYSIS OF VARIANCE - FROM MEAN
                      SS          DF            MS                 F
 REGRESSION       0.32879E+10      2.       0.16440E+10           153.966
 ERROR            0.72275E+11   6769.       0.10677E+08           P-VALUE
 TOTAL            0.75563E+11   6771.       0.11160E+08             0.000

                     ANALYSIS OF VARIANCE - FROM ZERO
                      SS          DF            MS                 F
```

```
 REGRESSION       0.33081E+10      3.       0.11027E+10           103.274
 ERROR            0.72275E+11   6769.       0.10677E+08           P-VALUE
 TOTAL            0.75583E+11   6772.       0.11161E+08             0.000

 VARIABLE    ESTIMATED   STANDARD   T-RATIO         PARTIAL STANDARDIZED ELASTICITY
   NAME     COEFFICIENT    ERROR    6769 DF    P-VALUE CORR.  COEFFICIENT  AT MEANS
 WHOUR1     7.8874     0.53429       14.762     0.0000 0.1766  0.17550      0.15532
 TAX1       3.1025     0.33363       9.2992     0.0000 0.1123  0.11055      0.10924
 CONSTANT   40.163      39.717       1.0112     0.3119 0.0123   0.0000      0.73544

 DURBIN-WATSON = 2.0003    VON NEUMANN RATIO = 2.0006    RHO = -0.00015
 RESIDUAL SUM =  0.78227E-09  RESIDUAL VARIANCE =  0.10677E+08
 SUM OF ABSOLUTE ERRORS=  0.54446E+06
```

Diagnositic suggests that there is no heteroscedasticity.

```
|_diagnos/acf/het

 REQUIRED MEMORY IS PAR=    1061 CURRENT PAR=    2000
 DEPENDENT VARIABLE = PAY1          6772 OBSERVATIONS
 REGRESSION COEFFICIENTS
    7.88736128123       3.10251000217       40.1627933487

 HETEROSKEDASTICITY TESTS
                            CHI-SQUARE     D.F.   P-VALUE
                          TEST STATISTIC
 E**2 ON YHAT:                   0.000     1    0.98357
 E**2 ON YHAT**2:                0.000     1    0.98708
 E**2 ON LOG(YHAT**2):           0.005     1    0.94533
 E**2 ON LAG(E**2) ARCH TEST:    0.000     1    0.99030
 LOG(E**2) ON X (HARVEY) TEST:   0.834     2    0.65892
 ABS(E) ON X (GLEJSER) TEST:     0.003     2    0.99853
 E**2 ON X                TEST:
           KOENKER(R2):          0.000     2    0.99977
           B-P-G (SSR) :         1.547     2    0.46150

 E**2 ON X X**2    (WHITE) TEST:
           KOENKER(R2):          0.001     4    1.00000
           B-P-G (SSR) :         4.182     4    0.38199

 E**2 ON X X**2 XX (WHITE) TEST:
           KOENKER(R2):          0.002     5    1.00000
           B-P-G (SSR) :         6.332     5    0.27521
Also note that there is no autocorrelation based on Durbin Watson or LM t-tests.

RESIDUAL CORRELOGRAM
  LM-TEST FOR HJ:RHO(J)=0, STATISTIC IS STANDARD NORMAL
  LAG     RHO       STD ERR     T-STAT     LM-STAT    DW-TEST BOX-PIERCE-LJUNG
    1    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0002
    2    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0003
    3    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0005
    4    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0006
    5    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0008
    6    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0009
    7    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0011
    8    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0012
    9    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0014
   10    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0016
   11    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0017
   12    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0019
   13    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0020
   14    -0.0002     0.0122    -0.0124     0.0124     2.0003     0.0022
   15     0.0001     0.0122     0.0099     0.0099     1.9998     0.0023
   16    -0.0001     0.0122    -0.0116     0.0116     2.0003     0.0024
   17    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0026
   18    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0027
   19    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0029
   20    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0030
   21    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0032
   22    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0033
   23    -0.0002     0.0122    -0.0125     0.0125     2.0003     0.0035
 LM CHI-SQUARE STATISTIC WITH  23  D.F. IS     0.003
Residuals do not depend on work hours and tax
```

genr res2 =res1*res1
ols res2 whours tax
het whour1 pay1 tax1/rstat

2. Consider a simple linear regression model.

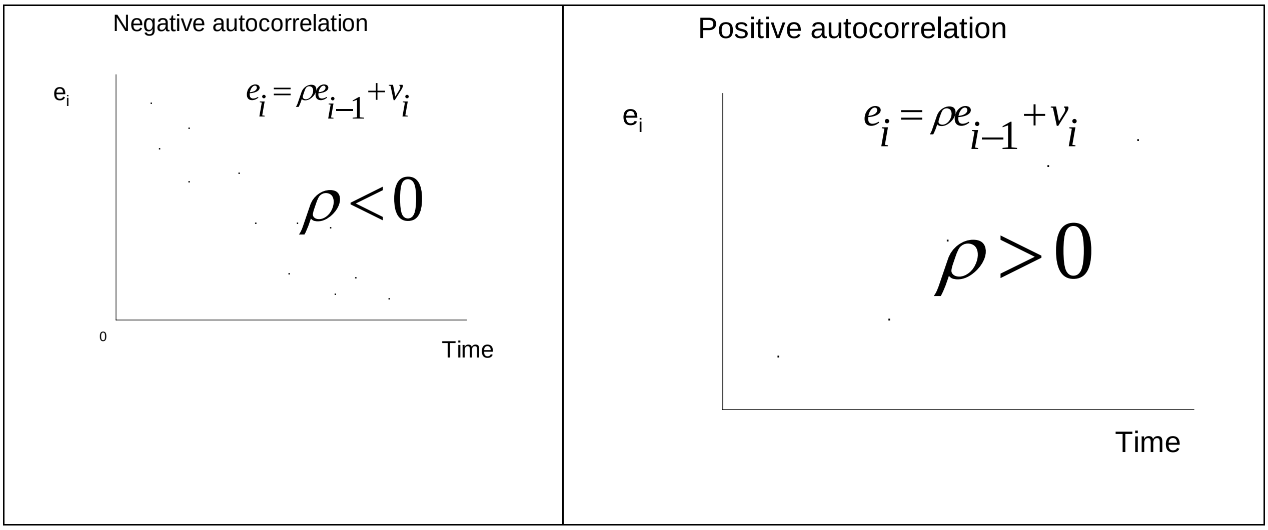
$$Y_t = \beta_1 + \beta_2 X_t + e_t$$

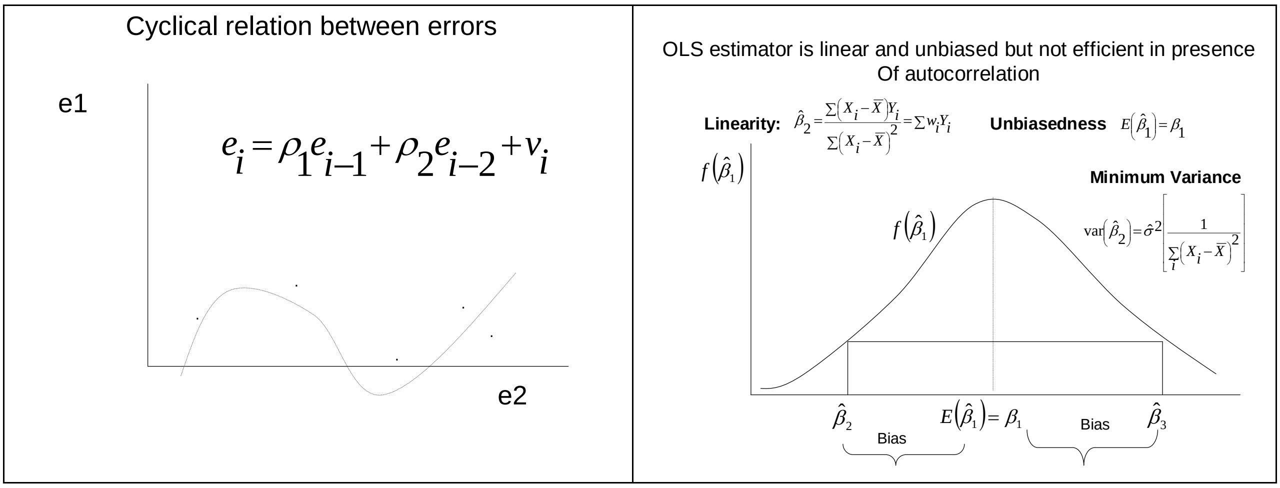
Now assume that errors are correlated to each other over time with AR(1) process as:

$$e_t = \rho e_{t-1} + v_t$$

where v_t is identically and normally distributed error term with zero mean and constant variance, $v_t \sim N(0, \sigma^2)$.

(f) Illustrate how the graphical method can be applied to detect autocorrelation in a simple regression model like above?



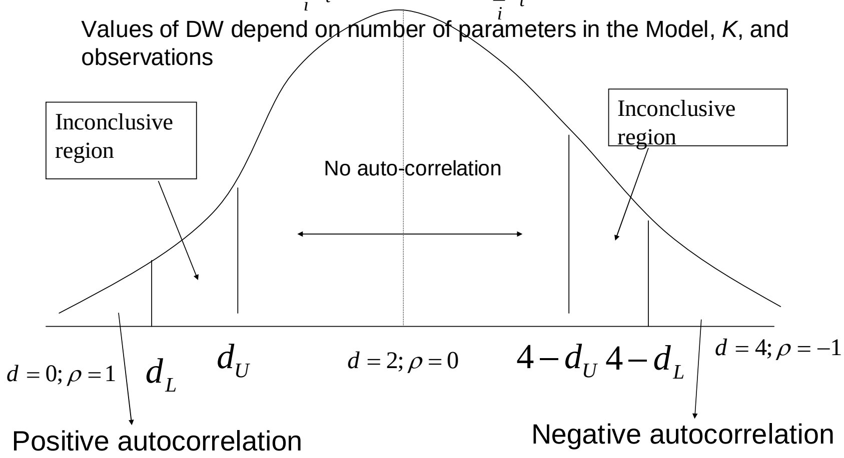


(g) What are consequences of autocorrelation in a regression model? Show how the existence of such autocorrelation among the error terms affects the BLUE properties of the OLS estimators.

(h) Define and derive the Durbin-Watson test statistics. Show how it can test for existence or non existence of autocorrelation in a given estimation?

Durbin-Watson (DW) test

$$\hat{d} = \frac{\sum_{i}^{T}\left(\hat{e}_t - \hat{e}_{t-1}\right)^2}{\sum_i \hat{e}_t^2} = \frac{\sum_i \hat{e}_t^2 - 2\sum_i \hat{e}_t \hat{e}_{t-1} + \sum_i \hat{e}_{t-1}^2}{\sum_i \hat{e}_t^2} \approx 2\left(1-\hat{\rho}\right)$$



(i) How the autocorrelation can be removed if the ρ is known?

When ρ is known –transform the model
Take a lag of the original model; and multiply it by ρ; and subtract from the original model to find a transformed model.

Given model: $Y_t = \beta_1 + \beta_2 x_t + e_t$

Where $e_t = \rho e_{t-1} + v_t$ and $v_t \sim N\left(\sigma^2\right)$

Multiply by ρ and lag by one period.

$$\rho e_{t-1} = \rho Y_{t-1} - \rho\beta_1 - \rho\beta_2 x_{t-1}$$

Subtract this from the original model

$$Y_t - \rho Y_{t-1} = \beta_1 - \rho\beta_1 + \beta_2 x_{t-1} - \rho\beta_2 x_{t-1} + e_t - \rho e_{t-1}$$

Transformed model

$$Y_t^* = \beta_1^* + \beta_2^* x_t^* + v_t$$

Where

$$Y_t^* = Y_t - \rho Y_{t-1}; \; X_t^* = X_t - \rho X_{t-1};$$

$$e_t^* = e_t - \rho e_{t-1};\ \beta^* = \beta_1 - \rho\beta_1.$$

OLS estimates unknown parameters of this transformed model will be Best Linear and Unbiased (BLUE). Retrieve β_1 of the original model from estimates of β^*.

(j) What is a spurious regression? Why does it arise and how does it affect the usefulness of estimation from an OLS regression? What can be done to correct it?

Application of OLS among non-stationary variable may generate a spurious regression, with a high R^2 and very low Durbin-Watson statistics (R^2>d).
A given time series $\{y_t\}$ is stationary when mean and variance are constant or independent of time.

$$E(y_t) = \mu \qquad \text{constant mean} \quad (1)$$

$$\text{var}(y_t) = \sigma^2 \qquad \text{constant variance} \quad (2)$$

$$\text{cov}(y_t y_{t-s}) = \text{cov}(y_t y_{t+s}) = \gamma_s \quad \text{time independent covariance (3)}$$

Time series y_t is non-stationary if the mean and variance of it is not constant. It is non-stationary if the variance is changing over time even though mean is constant.

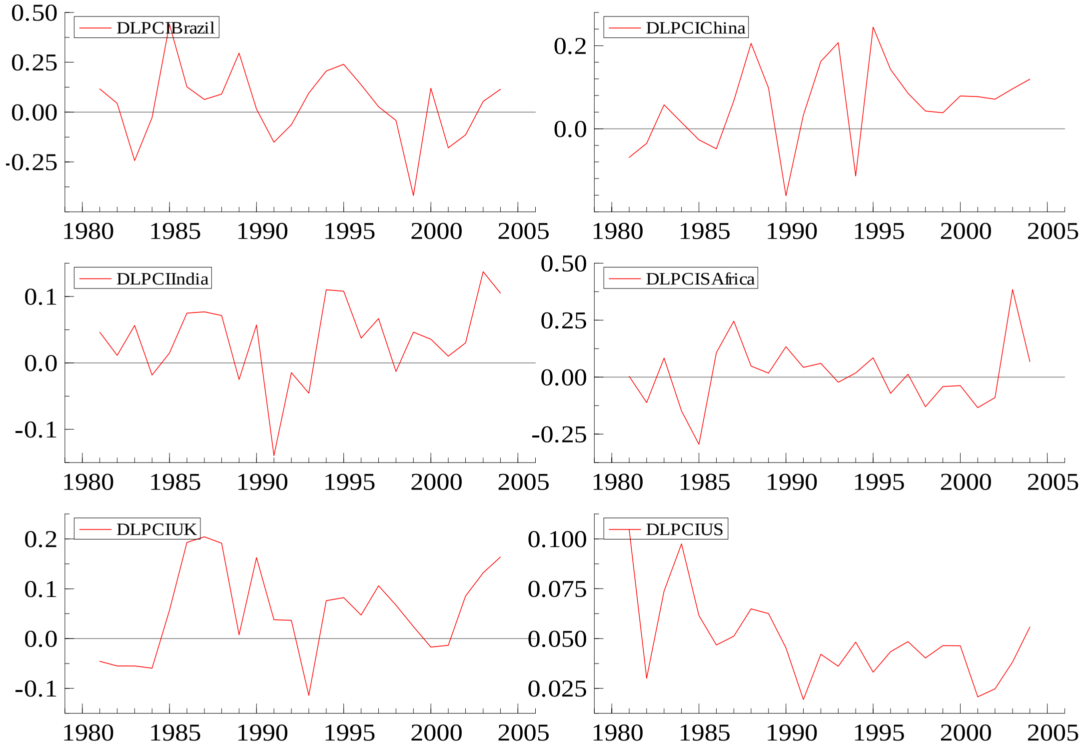
Many economic variables such as GDP, GDP components (C, I, G and X), inflation, exchange rates, labour force evolve over time. It is important to check whether these series have a constant mean and constant variance before they can be used in regression analysis. A meaningful cause and effect relationship requires that the concerned series are stationary. Application of OLS procedure in non stationary series produces a spurious relationship. A spurious relationship implies significant test-statistics (t, f, chi-square, R-square) even though there is no relationship among the variables. Econometric estimation using non-stationary variables may generate meaningless result though it may apparently seem statistically significant.

Application:
Read data on growth rate of per capita GDP, exchange rate and inflation rates from the www.imf.org for year 1980 to 2003 for China, India, South Africa, UK, USA and Brazil as contained in PERCAP6.GLS. Test whether inflation and the exchange rate are the significant variables in explaining the growth rate of per capita output (in PPP) in these economies. Determine whether heteroscedasticity and autocorrelation exist in this regression using PcGive. Feel free to use Shazam if you know and prefer it. Suggest a remedy for autocorrelation in a model like this.

A PcGive Batch File for Estimation

```
module("PcGive");
package("PcGive");
usedata("Percap6.in7");
system
{
    Y = DLPCIBrazil;
    Z = Constant, INFBrazil, EXRBrazil;
}
estimate("OLS", 1981, 1, 2004, 1);


module("PcGive");
package("PcGive");
usedata("Percap6.in7");
system
{
    Y = DLPCIChina;
    Z = Constant, INFChina, EXRChina;
}
estimate("OLS", 1981, 1, 2004, 1);

module("PcGive");
package("PcGive");
usedata("Percap6.in7");
system
{
    Y = DLPCIIndia;
    Z = Constant, INFIndia, EXRIndia;
}
estimate("OLS", 1981, 1, 2004, 1);

module("PcGive");
package("PcGive");
usedata("Percap6.in7");
system
{
    Y = DLPCISAfrica;
    Z = Constant, INFSAfrica, EXRSAfrica;
}
estimate("OLS", 1981, 1, 2004, 1);

module("PcGive");
package("PcGive");
usedata("Percap6.in7");
system
{
    Y = DLPCIUK;
    Z = Constant, INFUK, EXRUK;
```

```
}
estimate("OLS", 1981, 1, 2004, 1);

module("PcGive");
package("PcGive");
usedata("Percap6.in7");
system
{
    Y = DLPCIUS;
    Z = Constant, INFUS, EXRUS;
}
estimate("OLS", 1981, 1, 2004, 1);
```



The above figures are the log-differences, which give the growth rates of per capita income in these six countries.

Study and explain following results.

```
EQ( 1) Modelling DLPCIBrazil by OLS (using Percap6.in7)
       The estimation sample is: 1981 to 2004

                Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant          0.0454869    0.06514    0.698   0.493   0.0227
INFBrazil       3.11460e-005 5.318e-005   0.586   0.564   0.0161
EXRBrazil        -0.0553425    0.09525   -0.581   0.567   0.0158

sigma               0.184488  RS             0.714754398
R^2                0.0542803  F(2,21) =     0.6027 [0.557]
log-likelihood       8.11192  DW                   1.73
no. of observations       24  no. of parameters       3
mean(DLPCIBrazil)    0.03968  var(DLPCIBrazil)  0.0314908

AR 1-2 test:      F(2,19)  =  0.66903 [0.5239]
ARCH 1-1 test:    F(1,19)  =   1.1407 [0.2989]
Normality test:   Chi^2(2) =   2.1332 [0.3442]
hetero test:      F(4,16)  =  0.18293 [0.9439]
hetero-X test:    F(5,15)  =  0.14676 [0.9780]
RESET test:       F(1,20)  =  0.63338 [0.4355]

DLPCIBrazil = + 0.04549 + 3.115e-005*INFBrazil - 0.05534*EXRBrazil
(SE)          (0.0651)  (5.32e-005)           (0.0953)
```

EQ(2) Modelling DLPCIChina by OLS (using Percap6.in7)
The estimation sample is: 1981 to 2004

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	-0.105935	0.07798	-1.36	0.189	0.0808
INFChina	0.00361121	0.002744	1.32	0.202	0.0762
EXRChina	0.0953272	0.04890	1.95	0.065	0.1533

sigma	0.0936002	RSS	0.183981092
R^2	0.198224	F(2,21) =	2.596 [0.098]
log-likelihood	24.3972	DW	2.34
no. of observations	24	no. of parameters	3
mean(DLPCIChina)	0.0579471	var(DLPCIChina)	0.00956112

AR 1-2 test: F(2,19) = 5.5751 [0.0124]*
ARCH 1-1 test: F(1,19) = 0.18950 [0.6682]
Normality test: Chi^2(2) = 6.2565 [0.0438]*
hetero test: F(4,16) = 5.6332 [0.0050]**
hetero-X test: F(5,15) = 7.7464 [0.0009]**
RESET test: F(1,20) = 2.7909 [0.1104]

DLPCIChina = - 0.1059 + 0.003611*INFChina + 0.09533*EXRChina
(SE) (0.078) (0.00274) (0.0489)

EQ(3) Modelling DLPCIIndia by OLS (using Percap6.in7)
The estimation sample is: 1981 to 2004

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	0.0450041	0.06950	0.648	0.524	0.0196
INFIndia	-0.00394464	0.004581	-0.861	0.399	0.0341
EXRIndia	0.00361901	0.006730	0.538	0.596	0.0136

sigma	0.0600267	RSS	0.0756672753
R^2	0.087486	F(2,21) =	1.007 [0.382]
log-likelihood	35.059	DW	1.86
no. of observations	24	no. of parameters	3
mean(DLPCIIndia)	0.0350292	var(DLPCIIndia)	0.00345507

AR 1-2 test: F(2,19) = 0.18337 [0.8339]
ARCH 1-1 test: F(1,19) = 0.13395 [0.7184]
Normality test: Chi^2(2) = 1.8807 [0.3905]
hetero test: F(4,16) = 3.4801 [0.0316]*
hetero-X test: F(5,15) = 2.7614 [0.0581]
RESET test: F(1,20) = 0.077791 [0.7832]

DLPCIIndia = + 0.045 - 0.003945*INFIndia + 0.003619*EXRIndia
(SE) (0.0695) (0.00458) (0.00673)

EQ(4) Modelling DLPCISAfrica by OLS (using Percap6.in7)
The estimation sample is: 1981 to 2004

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	-0.362469	0.2466	-1.47	0.156	0.0933
INFSAfrica	0.0189345	0.01323	1.43	0.167	0.0889
EXRSAfrica	0.119475	0.07853	1.52	0.143	0.0993

sigma	0.136579	RSS	0.391732239
R^2	0.101023	F(2,21) =	1.18 [0.327]
log-likelihood	15.3282	DW	2
no. of observations	24	no. of parameters	3
mean(DLPCISAfrica)	0.00970972	var(DLPCISAfrica)	0.0181564

AR 1-2 test: F(2,19) = 2.3885 [0.1188]
ARCH 1-1 test: F(1,19) =0.0066752 [0.9357]
Normality test: Chi^2(2) = 5.7202 [0.0573]
hetero test: F(4,16) = 1.9400 [0.1528]
hetero-X test: F(5,15) = 1.9076 [0.1527]
RESET test: F(1,20) = 0.11214 [0.7412]

DLPCISAfrica = - 0.3625 + 0.01893*INFSAfrica + 0.1195*EXRSAfrica
(SE) (0.247) (0.0132) (0.0785)

```
EQ( 5) Modelling DLPCIUK by OLS (using Percap6.in7)
       The estimation sample is: 1981 to 2004

                Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant         0.00892220     0.3044   0.0293   0.977   0.0000
INFUK           -0.00513451    0.01022   -0.502   0.621   0.0119
EXRUK              0.108989     0.4540    0.240   0.813   0.0027

sigma              0.0911772  RSS              0.174578938
R^2                0.0497486  F(2,21) =      0.5497 [0.585]
log-likelihood       25.0267  DW                      1.09
no. of observations       24  no. of parameters          3
mean(DLPCIUK)      0.0548144  var(DLPCIUK)      0.00765494

AR 1-2 test:      F(2,19)  =   2.3882 [0.1188]
ARCH 1-1 test:    F(1,19)  = 0.083304 [0.7760]
Normality test:   Chi^2(2) =  0.23930 [0.8872]
hetero test:      F(4,16)  =  0.72920 [0.5850]
hetero-X test:    F(5,15)  =  0.63048 [0.6795]
RESET test:       F(1,20)  =   1.7078 [0.2061]

DLPCIUK =  + 0.008922 - 0.005135*INFUK + 0.109*EXRUK
(SE)        (0.304)    (0.0102)         (0.454)
```

```
EQ( 6) Modelling DLPCIUS by OLS (using Percap6.in7)
       The estimation sample is: 1981 to 2004

                Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant          -0.882863      1.540   -0.573   0.573   0.0154
INFUS            0.00786667   0.001854     4.24   0.000   0.4615
EXRUS              0.883411      1.504    0.587   0.563   0.0162

sigma              0.0160444  RSS            0.00540590909
R^2                 0.461572  F(2,21) =     9.001 [0.002]**
log-likelihood       66.7253  DW                      1.31
no. of observations       24  no. of parameters          3
mean(DLPCIUS)      0.0492795  var(DLPCIUS)      0.00041834

AR 1-2 test:      F(2,19)  =   1.6105 [0.2259]
ARCH 1-1 test:    F(1,19)  = 0.043763 [0.8365]
Normality test:   Chi^2(2) =   7.5455 [0.0230]*
hetero test:      F(4,16)  =  0.78401 [0.5520]
hetero-X test:    F(4,16)  =  0.78401 [0.5520]
RESET test:       F(1,20)  = 0.084660 [0.7741]

DLPCIUS =  - 0.8829 + 0.007867*INFUS + 0.8834*EXRUS
(SE)        (1.54)   (0.00185)        (1.5)
```

Problem 6
Demand and Supply analysis Using Input Output Table for a Hypothetical Economy

1. Find the determinant and inverse of matrix $A = \begin{bmatrix} 5 & 4 \\ 7 & 8 \end{bmatrix}$.

2. An economy, with two sectors, X_1 and X_2, has following input-output table.

	X_1	X_2	F	Total
X_1	10	20	70	100
X_2	30	20	150	200
Labour input	40	50		90
Capital input	20	110		130
Total	100	200	220	

Where X_1 is the gross production of sector 1 and X_2 is the gross production of sector 2, *F* is the final demand that includes consumption, investment, government spending and next exports.

a. Write equations to represent demands by sector for two sectors of the economy.

b. Check how demand and supply, income and expenditure accounts are balanced for this economy.

c. Find technical coefficients, $a_{i,j} = X_{i,j} \big/ X_j$ and share of primary inputs for both sectors. Here $X_{i,j}$ represents intermediate demand for sector *i* good by sector *j*; $X_{1,1} = 10$ $X_{1,2} = 20$ $X_{2,1} = 30$ and $X_{2,2} = 20$ represent intermediate inputs.

d. Put the technological coefficients in a Leontief matrix $A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$.

e. Propose an input-output model for this economy (using the matrix format).

f. With an identity matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $(I - A)$?

a. Express gross output in terms of final demand and inverse of $(I - A)$ matrix.

b. Find out the impact of a 30 percent increase in the final demand of sector 2 in outputs of sector 1 and sector 2.

General Equilibrium in Markets

2. An economy produces two goods; apples and oranges and has two households, $h = \{A, B\}$. Household *A* owns the apple farm and produces 100 quintals of apples and household *B* owns the orange farm and produces 200 quintals of oranges. Both households like to consume apples and oranges. Their consumption preferences are given by Cobb-Douglas Utility functions, $U^h = X_{1,h}^{\alpha_h} X_{2,h}^{(1-\alpha_h)}$ where U^h is the utility to household *h*, $X_{1,h}$ and $X_{2,h}$ are consumptions of apples and oranges and α_h represents the weight of each good in the utility functions, $0 < \alpha_h < 1$. Household *A* spends 40 percent of income in apple and 60 percent in oranges and household *B* spends 60 percent in apples and 40 percent in oranges. Market structure is competitive.

a. Represent the supply of goods using an appropriate Edgeworth box diagram.

b. Find the relative price in this economy that is consistent with maximization of utility (satisfaction) by both households. Choose price of commodity 1 as a numeraire.

c. Determine the income for each household.

d. What are demands for apples and oranges by households *A* and *B*?

e. Check whether the market clearing conditions for equilibrium are fulfilled. Find their levels of utility at equilibrium

f. Represent both consumption and supply and relative price of goods in equilibrium using an appropriate Edgeworth box diagram.

How would you estimate α_h in a real world situation?

Answer 6

1. Economic theory and economic models: Illustration of demand and supply curves.

$D = a_0 - a_1P$; $S = b_0 + b_1P$

Simple market:

$$D = 300 - 2P$$

$$S = 20 + 3P$$

$$P = \frac{a_0 - b_0}{a_1 - b_1} = \frac{300 - 20}{2 + 3} = \frac{280}{5} = 56$$

$$D = 300 - 2P$$

$$Q = a_0 - a_1 \frac{a_0 - b_0}{a_1 - b_1} = 300 - 2(56) = 300 - 112 = 188$$

What is the deadweight loss of 20 percent advalorem tax ?

$S = b_0 + b_1(1-t)P$

$$D = 300 - 2P$$

$$S = 20 + 3(1 - t)P$$

$$Pt = \frac{a_0 - b_0}{a_1 - b_1(1 - t)} = \frac{300 - 20}{2 + 3(1 - 0.2))} = \frac{280}{4.4} = 63.64$$

$$Ps = Pt(1 - t) = 63.63 \times 0.8 = 50.90$$

$$Qt = a_0 - a_1 \frac{a_0 - b_0}{a_1 - b_1(1 - t)} = 300 - 2(63.64) = 300 - 127.3 = 172.2$$

Revenue of the supplier before tax: $PQ = 56 \times 188 = 10528$

Revenue of the supplier after tax: $Pt \times Qt = 50.90 \times 172.2 = 8764.98$

Revenue for the government: $R = t \times Pt \times Qt = 0.2 \times 63.64 \times 172.2 = 2191.76$

Consumers' loss: $\frac{1}{2} \times (Pt - P) \times \Delta Q = 0.5 \times (63.63 - 56) \times (172.2 - 188) = -60.35$

Producers' loss: $\frac{1}{2} \times (Ps - P) \times \Delta Q = 0.5 \times (50.90 - 56) \times (172.2 - 188) = 40.29$

Deadweight loss of taxes: Consumers loss + Producer's loss = 60.3+40.3 = 100.6

Variables in this model: P, Q

Parameters: a_0, a_1, b_0 , b_1 ,t. How to estimate them?

Lecture 8 - 9

Output, Price, Welfare and Profit in an imperfectly competitive market

Market demand function: $P = 30 - q_1 - q_2$

Cost: $C_i = 6q_i$

Profit: $\Pi_i = Pq_i - C_i$

Duopoly: Cournot Model

Firm 1: $\Pi_1 = Pq_1 - C_1 = (30 - q_1 - q_2)q_1 - 6q_1 = 30q_1 - q_1^2 - q_1q_2 - 6q_1$

Firm 2: $\Pi_2 = Pq_2 - C_2 = (30 - q_1 - q_2)q_2 - 6q_2 = 30q_2 - q_2^2 - q_1q_2 - 6q_2$

$$\frac{\partial \Pi_1}{\partial q_1} = 30 - 2q_1 - q_2 - 6 = 0 \Rightarrow 24 = 2q_1 + q_2$$

$$\frac{\partial \Pi_2}{\partial q_2} = 30 - 2q_2 - q_1 - 6 = 0 \Rightarrow 24 = q_1 + 2q_2$$

Solving these two equation $q_1 = q_2 = 8$

$P = 30 - q_1 - q_2 = 30 - 8 - 8 = 14$

$\Pi_1 = Pq_1 - C_1 = 14 \times 8 - 6 \times 8 = 64$

$\Pi_2 = Pq_2 - C_2 = 14 \times 8 - 6 \times 8 = 64$

Consumer surplus from firm 1 product: $CS_1 = \frac{1}{2} \times 16 \times 8 = 64$

Consumer surplus from firm 2 product: $CS_2 = \frac{1}{2} \times 16 \times 8 = 64$

Total welfare: $\Pi_1 + \Pi_2 + CS_1 + CS_2 = 64 + 64 + 64 + 64 = 256$

Stackleberg equilibrium

When the Firm 1 is the leader and firm 2 is the follower.

$\Pi_1 = Pq_1 - C_1 = 30q_1 - q_1^2 - 6q_1$

$$\frac{\partial \Pi_1}{\partial q_1} = 30 - 2q_1 - 6 = 0 \Rightarrow 24 = 2q_1 \Rightarrow q_1 = 12$$

$\Pi_2 = Pq_2 - C_2 = 30q_2 - q_2^2 - q_1q_2 - 6q_2$

$$\frac{\partial \Pi_2}{\partial q_2} = 30 - 2q_2 - q_1 - 6 = 0 \Rightarrow 24 - q_1 = 2q_2 \Rightarrow 2q_2 = 24 - 12 \Rightarrow q_2 = 6$$

$P = 30 - q_1 - q_2 = 30 - 12 - 6 = 12$

$\Pi_1 = Pq_1 - C_1 = 12 \times 12 - 6 \times 12 = 72$

$\Pi_2 = Pq_2 - C_2 = 12 \times 6 - 6 \times 6 = 36$

Consumer surplus from firm 1 product: $CS_1 = \frac{1}{2} \times 18 \times 12 = 108$

Consumer surplus from firm 2 product: $CS_2 = \frac{1}{2} \times 18 \times 6 = 54$

Total welfare: $\Pi_1 + \Pi_2 + CS_1 + CS_2 = 72 + 36 + 108 + 54 = 270$

Collusion (cartel): maximise industry profit.

Market demand function: $P = 30 - q_1 - q_2 = 30 - Q$

Cost: $C = 6Q$

Profit: $\Pi = PQ - C = (30 - Q)Q - 6Q = 30Q - Q^2 - 6Q$

$\frac{\partial \Pi}{\partial Q} = 30 - 2Q - 6 = 0$ ➔ $2Q = 24$ ➔ $Q = 12 \quad q_1 = 6 \quad q_2 = 6$

$P = 30 - q_1 - q_2 = 30 - Q = 30 - 12 = 18$

Consumer surplus from firm 1 product: $CS_1 = \frac{1}{2} \times 12 \times 6 = 36$

Consumer surplus from firm 2 product: $CS_2 = \frac{1}{2} \times 12 \times 6 = 36$

$\Pi_1 = Pq_1 - C_1 = 18 \times 6 - 6 \times 6 = 72$

$\Pi_2 = Pq_2 - C_2 = 18 \times 6 - 6 \times 6 = 72$

Total welfare: $\Pi_1 + \Pi_2 + CS_1 + CS_2 = 72 + 72 + 36 + 36 = 216$

Comparison of solutions in an imperfectly competitive market

	Cournot	Stackleberg	Collusion
P	14	12	18
q_1	8	12	6
q_2	8	6	6
Π_1	64	72	72
Π_2	64	36	72
CS_1	64	108	36
CS_2	64	162	36
TW	256	270	216

Problem 7

Models of demand and supply and strategic decisions

1. Market demand and supply for a normal good are $D = a - bP$ and $S = -c + dP$ respectively, where *D* and *S* are quantities demanded and supplied and *a, b, c* and *d* are parameters representing the behaviour of buyers and suppliers in the market.

   a. What are theoretical assumptions about the sign of parameters *a*, *b*, *c* and *d* in this model? How would you obtain their numerical values? How demand and supply elasticities can be calculated with knowledge of *b* and *d* respectively?
   b. The demand for and supply of surfing boards in a market were given by $D = 300 - 3P$ and $S = -5 + 5P$ respectively. Determine the price and quantity in equilibrium.
   c. Show demand and supply schedules and equilibrium in a properly labelled and scaled diagram.
   d. The government introduces a 20 percent sales tax rate on sale of surfing boards. Show equilibrium prices and quantities before and after this sales tax in another diagram.
   e. Calculate the deadweight loss of taxes to consumers, to producers and to the entire economy using partial equilibrium analysis contained in this model.

2. There are two cinema halls in Hull. Objective of each is to maximise profit. The market demand curve for movies is given by $Q = 50 - 2P$, where *Q* is demand and *P* is the price. The price per ticket depends on total sales $P = 10 - (q_1 + q_2)$, where q_1 and q_2 are quantities sold by each hall. The cost function for hall *i* is $C_i = 2q_i$ for i = 1, 2.

   a) How many tickets does each hall sell to maximise its profit taking sales of another hall as fixed and how much should has one to pay to go to a movie? (hint: Cournot model).
   b) Calculate the consumers' surplus at that price.
   c) How big is the producers' surplus?
   d) What is the size of welfare to the entire economy at those prices and quantities?
   e) Answer questions (a) to (d) when Hall 1 acts as a Stackleberg leader in the market.
   f) Answer questions (a) to (d) when both of these halls collide to maximise joint profits.
   g) Put results on sales, price, profit and welfare for all three markets in one table.
   h) Make a pay-off matrix of profits in all three strategic markets for Cinema 1 and Cinema 2.
   i) What price would have prevailed if this market was perfectly competitive? What would have been the value of welfare then?

3. Two firms are in the telecom market. Their pay-offs by advertising or not advertising are as listed in the following matrix.

		Firm 2	
		Adv	DnotAdv
Firm 1	*ADV*	(10,5)	(15,0)
	DnotAdv	(6,8)	(10,2)

   Solve this game using a dominant strategy. What are the pay-offs for firm 1 and firm 2?

4. (B.Sc.) Consider a zero sum game for a competitive market where one can benefit only at the expense of another.

		Player 1	
		S1	***S2***
Player 2	*S1*	(10,–10)	(–10,10)
	S2	(–10,10)	(10,–10)

The probability of player 2 playing S1 strategy is given by π . Represent expected pay-off of this game in a diagram. Find the optimal mix of strategies for this player, i.e. find the optimal value of π under the mixed strategy.

Input-output Model

Learning Matrix

Revenues of two stores are 200 and 300 respectively. The sell two different products X_1 and X_2. Firm 1 sells 50 units of X_1 and 40 units of X_2. Firm 2 sells 70 units of X_1 and 80 units of X_2. Formulate this information fist in equation and then put it in the matrix form. Then find the prices of X_1 and X_2.

$$R_1 = P_1 X_{1,1} + P_2 X_{1,2}$$
$$R_2 = P_1 X_{2,1} + P_2 X_{2,2}$$

$X_{i,j}$ is the sell by firm i of product j.

$$\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

$$200 = 50P_1 + 40P_2$$
$$300 = 70P_1 + 80P_2$$

From the given information

$$\begin{bmatrix} 200 \\ 300 \end{bmatrix} = \begin{bmatrix} 50 & 40 \\ 70 & 80 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}; \quad \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 50 & 40 \\ 70 & 80 \end{bmatrix}^{-1} \begin{bmatrix} 200 \\ 300 \end{bmatrix}$$

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \frac{\begin{bmatrix} 80 & -40 \\ -70 & 50 \end{bmatrix}}{(50 \times 80 - 40 \times 70)} \begin{bmatrix} 200 \\ 300 \end{bmatrix} = \frac{\begin{bmatrix} 80 & -40 \\ -70 & 50 \end{bmatrix}}{(1200)} \begin{bmatrix} 200 \\ 300 \end{bmatrix} = \frac{1}{1200} \begin{bmatrix} 80(200) - 40(300) \\ -70(200) + 50(300) \end{bmatrix}$$

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \frac{1}{1200} \begin{bmatrix} 16000 - 12000 \\ -14000 + 15000 \end{bmatrix} = \frac{1}{1200} \begin{bmatrix} 4000 \\ 1000 \end{bmatrix} = \begin{bmatrix} 40/12 \\ 10/12 \end{bmatrix} = \begin{bmatrix} 10/3 \\ 5/6 \end{bmatrix}$$

These prices satisfy above revenue equations.

$$200 = 50P_1 + 40P_2 \rightarrow 200 = 50 \times \frac{10}{3} + 40 \times \frac{5}{6} = \frac{1000 + 200}{6} = 200$$

$$300 = 70P_1 + 80P_2 \rightarrow 200 = 70 \times \frac{10}{3} + 80 \times \frac{5}{6} = \frac{1400 + 400}{6} = 300$$

Demand and Supply: Input Output Analysis

An economy, with two sectors, X_1 and X_2, has following input-output table.

	X_1	X_2	F	Total
X_1	10	20	70	100
X_2	30	20	150	200
Labour input	40	50		90
Capital input	20	110		130
Total	100	200	220	

Propose an input-output model for this economy.

Gross Supply = intermediate demand plus final demand.
Income (labour income plus capital income) = expenditure (F1 +F2) = GDP

$$X_1 = X_{1,1} + X_{1,2} + F_1$$
$$X_2 = X_{2,1} + X_{2,2} + F_2$$

$X_{i,j}$ is intermediate input from sector i to sector j.

$$a_{1,1} = \frac{X_{1,1}}{X_1} \quad a_{1,2} = \frac{X_{1,2}}{X_2} \quad a_{2,1} = \frac{X_{2,1}}{X_2} \quad a_{2,2} = \frac{X_{2,2}}{X_2}$$

$a_{i,j}$ share of input from row sector i to sector j.

$$X_1 = a_{1,1}X_1 + a_{1,2}X_2 + F_1$$
$$X_2 = a_{2,1}X_1 + a_{2,2}X_2 + F_2$$

How does the final demand affect the level of gross output?

$$(1 - a_{1,1})X_1 - a_{1,2}X_2 = F_1$$
$$-a_{2,1}X_1 + (1 - a_{2,2})X_2 = F_2$$

Better to use matrix algebra to solve the problem.

$$\begin{bmatrix} (1 - a_{1,1}) & -a_{1,2} \\ -a_{2,1} & (1 - a_{2,2}) \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} (1 - a_{1,1}) & -a_{1,2} \\ -a_{2,1} & (1 - a_{2,2}) \end{bmatrix}^{-1} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$$

$$X = (I - A)^{-1} F$$

$$(I - A)^{-1} = \frac{Adj(A)}{|A|} = \frac{[C_{i,j}]'}{|A|} = \frac{\begin{bmatrix} (1 - a_{2,2}) & a_{2,1} \\ a_{1,2} & (1 - a_{1,1}) \end{bmatrix}}{(1 - a_{1,1})(1 - a_{2,2}) - a_{1,2}a_{2,1}}$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \frac{\begin{bmatrix} (1 - a_{2,2}) & a_{2,1} \\ a_{1,2} & (1 - a_{1,1}) \end{bmatrix}}{(1 - a_{1,1})(1 - a_{2,2}) - a_{1,2}a_{2,1}} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$$

$$a_{1,1} = \frac{X_{1,1}}{X_1} = \frac{10}{100} = 0.1 \quad a_{1,2} = \frac{X_{1,2}}{X_2} = \frac{20}{200}$$
$$a_{2,1} = \frac{X_{2,1}}{X_2} = \frac{30}{100} = 0.3 \quad a_{2,2} = \frac{X_{2,1}}{X_2} = \frac{20}{200} = 0.1$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \frac{\begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.9 \end{bmatrix}}{(0.9)(0.9) - 0.3 \times 0.1} \begin{bmatrix} 70 \\ 50 \end{bmatrix} = \frac{\begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.9 \end{bmatrix}}{0.81 - 0.03} \begin{bmatrix} 70 \\ 50 \end{bmatrix} = \begin{bmatrix} \frac{78}{0.78} \\ \frac{156}{0.78} \end{bmatrix} = \begin{bmatrix} 100 \\ 200 \end{bmatrix}$$

Model is calibrated.
Final demand includes demand for consumption, investment, government spending and net exports. If any of these change it has economy wide impacts, that such impact can be evaluated using such a model. For instance if the final demand in sector 1 rises by 15 percent what will be the change in total output.

$$\begin{bmatrix} \Delta X_1 \\ \Delta X_2 \end{bmatrix} = \frac{\begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.9 \end{bmatrix}}{0.81 - 0.03} \begin{bmatrix} 0.15 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{0.135}{0.78} \\ \frac{0.045}{0.78} \end{bmatrix} = \begin{bmatrix} 0.1731 \\ 0.0577 \end{bmatrix}$$

Thus a 15 percent increase in the final demand of sector 1 raises output in sector 1 by 17.3 percent and in sector 2 by 5.8 percent.
Price is constant in this input-output model but prices are flexible in the economy.
Lecture 9

General equilibrium in a pure exchange economy with flexible prices.

2. An economy produces apples and oranges. Household A owns the apple farm and produces 100 quintal of apples and household B owns the orange farm and produces 200 quintals of oranges. Both households like to consume apple and oranges. Their consumption preferences by given by Cobb-Douglas Utility functions, household A spends 40 percent of income in apple and 60 percent in oranges and household B spends 60 percent in apples and 40 percent in oranges.
Market structure is competitive. Find the relative price in this economy that is consistent with maximization of utility (satisfaction) by both households. Choose price of commodity 1 as a numeraire. Find the income of each households, their demands for both apples and oranges. Check whether the conditions for equilibrium are fulfilled. Find their levels of utility at equilibrium.

Demand functions:

$$X_1^A = \frac{\alpha I^A}{P_1} \quad X_2^A = \frac{(1-\alpha) I^A}{P_2}$$

$$X_1^B = \frac{\beta I^B}{P_1} \quad X_2^B = \frac{(1-\beta) I^B}{P_2}$$

Supply of good 1: $\omega_1 = \omega_1^A + \omega_1^B$ for simplicity $\omega_1 = \omega_1^A$ and $\omega_1^B = 0$

Supply of good 2: $\omega_2 = \omega_2^A + \omega_2^B$ for simplicity $\omega_2^A = 0$ and $\omega_2 = \omega_1^B$

Further assume that $\omega_1^A = 100 \quad \omega_2^B = 200 \quad \alpha = 0.4 \quad \beta = 0.6$

Market clearing condition

$$X_1^A + X_1^B = \omega_1$$
$$X_2^A + X_2^B = \omega_2$$
$$I^A = P_1 \omega_1$$
$$I^B = P_2 \omega_2$$

Using above demand functions

$$\frac{\alpha I^A}{P_1} + \frac{\beta I^A}{P_1} = \omega_1$$

$$\frac{(1-\alpha)I^A}{P_2} + \frac{(1-\beta)I^B}{P_2} = \omega_2$$

Substituting above

$$\frac{\alpha P_1\omega_1}{P_1} + \frac{\beta P_2\omega_2}{P_1} = \omega_1 \Rightarrow \alpha_1\omega_1 + \frac{\beta P_2\omega_2}{P_1} = \omega_1$$

$$100(0.4) + 200(0.6)\frac{P_2}{P_1} = 100 \Rightarrow 0.4 + 1.2\frac{P_2}{P_1} = 1$$

$$\frac{P_2}{P_1} = \frac{0.6}{1.2} = 0.5$$

You get the same relative price even if you use the second market clearing condition. Normalise price of good 1; $P_1 = 1$.

$$I^A = P_1\omega_1 = 1 \times 100 = 100$$

$$I^B = P_2\omega_2 = 0.5 \times 200 = 100$$

$$X_1^A = \frac{\alpha I^A}{P_1} = \frac{0.4 \times 100}{1} = 40 \quad X_2^A = \frac{(1-\alpha)I^A}{P_2} = \frac{0.6 \times 100}{0.5} = 120$$

$$X_1^B = \frac{\beta I^B}{P_1} = \frac{0.6 \times 100}{1} = 60 \quad X_2^B = \frac{(1-\beta)I^B}{P_2} == \frac{0.4 \times 100}{0.5} = 80$$

$$X_1^A + X_1^B = 40 + 60 = 100 = \omega_1$$

$$X_2^A + X_2^B = 120 + 80 = 200 = \omega_2$$

The market is cleared.
Change of the numeraire does affect the optimal allocation. Try normalising $P_2 = 1$.

3. **GAMS** is good particularly in solving a general equilibrium model with many linear or non-linear equations on continuous or discrete variables. It comes with a number of solvers that are useful for numerical analysis. For economic modelling it can solve very large scale models using detailed structure of consumption, production and trade arrangements on unilateral, bilateral or multilateral basis in the global economy where the optimal choices of consumers and producers are constrained by resources and production technology or arrangements for trade. It is a user friendly software. Any GAMS programme involves declaration of set, parameters, variables, equations, initialisation of variables and setting their lower or upper bounds and solving the model using Newton or other methods for linear or non-linear optimisation and reporting the results in tables or graphs (e.g. ISLM.gms). GAMS/MPSGE program is good for large scale standard general equilibrium models. GAMS programme is located at N:\special\ec\gams\gams in the university network and can be used by going through following steps.
First, create a directory called models in G:drive G:\> md Models then G:\> cd Models. Then write or copy a GAMS program file in that directory such as G:\models\islm.gms .
Type N:\special\ec\gams\gams islm.gms to run a GAMS program of a model in the network. The results of the model computations can be seen in the list file called ISLM.LST.

The check whether the results are consistent with the economic theory underlying the model such as ISLM-ASAD analysis for evaluating the impacts of expansionary fiscal and monetary policies. Use knowledge of growth theory to explain results of the Solow growth model from Solow.gms.
Consult GAMS and GAMS/MPSGE User Manuals, GAMS Development Corporation, 1217 Potomac Street, Washington D.C or www.gams.com.

Additional Problem
An Introduction to the Matrix Algebra

Find the determinant of the following matrix.

a) $A = \begin{bmatrix} 50 & 40 \\ 70 & 80 \end{bmatrix}$ b) A= $\begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ c) B = $\begin{bmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{bmatrix}$

1. Find the inverse of the following matrix

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

2. Prove that following matrix is a positive definite matrix

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

3. Solve following equations system using Cramer's rule

$$x_1 + 2x_2 + 2x_3 = 1 \qquad (1)$$
$$2x_1 + 2x_2 + 3x_3 = 3 \qquad (2)$$
$$x_1 - x_2 + 3x_3 = 5 \qquad (3)$$

1. Find the determinant of the following matrix.

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

(a)

$$|A| = \begin{vmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}$$ = (2.5.9)+(1.6.7)+(3.8.4)-(7.5.3)-(8.6.2)-(9.4.1)=90+42+96-105-96-36=228-237=-9

(b)

$$|B| = \begin{vmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{vmatrix}$$ =(-7.1.5)+(0.4.0)+(3.6.9)-(0.1.3)-(6.4.(-7))-(5.9.0)=-35+0+162-0+168-0=

-35+330=295.

- If the determinant of a matrix is zero then that matrix is called a singular matrix. Singularity reflects linear dependence among explanatory variables.

2. Find the inverse of the following matrix

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

Determinant of this matrix = $|A|$ =99≠0 => the inverse A^{-1} exists.

The cofactor matrix $C_{i,j} = (-1)^{i+j}|M_{i,j}|$ where $M_{i,j}$ are minors of each element.

$$C = \begin{bmatrix} \begin{vmatrix} 3 & 2 \\ 0 & 7 \end{vmatrix} & -\begin{vmatrix} 0 & 2 \\ 3 & 7 \end{vmatrix} & \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} \\ -\begin{vmatrix} 1 & -1 \\ 0 & 7 \end{vmatrix} & \begin{vmatrix} 4 & -1 \\ 3 & 7 \end{vmatrix} & -\begin{vmatrix} 4 & 1 \\ 3 & 0 \end{vmatrix} \\ \begin{vmatrix} 1 & -1 \\ 3 & 2 \end{vmatrix} & -\begin{vmatrix} 4 & -1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} 4 & 1 \\ 0 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 21 & 6 & -9 \\ -7 & 31 & 3 \\ 5 & -8 & 12 \end{bmatrix}$$

$$C' = Adj(A) = \begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

The desired inverse matrix is $A^{-1} = \frac{1}{|A|} Adj(A) = \frac{1}{99} \begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$

3. Solve following equations system using Cramer's rule

$$x_1 + 2x_2 + 2x_3 = 1 \qquad (1)$$
$$2x_1 + 2x_2 + 3x_3 = 3 \qquad (2)$$
$$x_1 - x_2 + 3x_3 = 5 \qquad (3)$$

Answer: First trite this system of three equations and three unknowns in matrix format as following.

$$\begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 3 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \quad ;$$

A x b

$$|A| = \begin{vmatrix} 1 & 2 & 2 \\ 2 & 2 & 3 \\ 1 & -1 & 3 \end{vmatrix}$$ =1.2.3+2.3.1+2(-1).2-2.2.1-(-1).3.1-3.2.2=6+6-4-4+3-12=12-8+3-12=-5

The solutions for x1, x2, and x3 is defined as following: $x_1 = \dfrac{\begin{vmatrix} 1 & 2 & 2 \\ 3 & 2 & 3 \\ 5 & -1 & 3 \end{vmatrix}}{|A|} = \dfrac{-5}{-5} = 1$

$$x_2 = \frac{\begin{vmatrix} 1 & 1 & 2 \\ 2 & 3 & 3 \\ 1 & 5 & 3 \end{vmatrix}}{|A|} = \frac{5}{-5} = -1 \quad x_3 = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 3 \\ 1 & -1 & 5 \end{vmatrix}}{|A|} = \frac{-5}{5} = -1 \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Textbooks:

Research requires **thinking and solving problems** more than reading books. Text books, nevertheless, can expose one to popular tools used by professional researchers. **Koop (2000)** brilliantly shows how to use **Excel** for data analysis. Dougherty (2002), Hill-Griffith and Judge (2001) and Studenmund (2001) are good introductory texts in econometrics, any one of these is enough this module. **Excel**, **GiveWin PcGive, STAMP** and **Shazam** are very useful software for handling data and economic analysis. They are available through the **Start and Applications/Economics menu** in the university network. The GAMS is useful for solving general equilibrium or linear and non-linear programming models. Texts in **micro** and **macro** economics are helpful in generating ideas about economic issues.

- Koop G. (2000) Analysis of Economic Data, Wiley, UK.
- Dougherty C. (2002) Introduction of Econometrics by, Second Edition, Oxford University Press.

Softwares:

- Doornik J A and D.F. Hendry ((2003) PC-Give Volume I-III, GiveWin Timberlake Consultants Limited, London.
- Shazam (1997) User's Reference Manual, Version 8.0. http://shazam.econ.ubc.ca/
- GAMS Users Manual, GAMS Development Corporation, Washington DC. www.gams.com.

Problem 1

(q)

a.

Cross Section Analysis

The marks scored by students in two exams and their monthly earnings from a part time jobs are as given in the following table.

Scores in Exams and Earnings

Observation	Exam 1	Exam 2	Earning	observation	Exam 1	Exam 2	Earning
1	0	5	248	25	5.2	45	196
2	6	55	161	26	4.8	55	208
3	5.6	60	213	27	7	63	245
4	5.6	54	222	28	7.5	62	155
5	4	56	180	29	5.2	65	168
6	6	65	184	30	5.2	66	211
7	6	58	249	31	6.4	60	208
8	5.2	61	191	32	4	58	238
9	3.2	30	213	33	6	68	172
10	6.5	61	186	34	5.2	64	172
11	7.5	65	232	35	6	57	219
12	4	30	235	36	7	60	202
13	5	58	231	37	5.6	63	178
14	7	68	242	38	5.6	61	175
15	6.5	68	209	39	7	62	163
16	4.5	60	230	40	6.5	65	193
17	8	71	238	41	6.5	53	200
18	6	62	150	42	7.5	62	206
19	6.5	55	237	43	5	58	241
20	6	50	227	44	6	63	227
21	5.2	51	184	45	6	58	245

22	5.2	55	229	46	5	57	161
23	4	64	205	47	3	38	213
24	4	51	221	48	4.8	60	227

12. Represent scores in exams and earning data using frequency table with ten intervals.

13. What are means and variances in those exams?

14. What is the coefficient of variations for scores in exams1 and 2?
15. What is the covariance of marks in exams 1 and 2?
16. What is correlation coefficient between scores in exam 1 and 2?
17. If exam 1 only counts for 10 percent but the scores in exam 2 weigh 100 percent what would be the weighted mean score in these two exams?
18. Exam 1 took place before exam 2. Test whether scores in exam 2 can predict scores in exam 1?
19. Find predicted scores in exam 2 for students who scored 6 and 8 in exam 1.

20. Test hypothesis whether scores in exam 1 and exam 2 are significant determinants of earning. Why may earnings be negatively related with test scores?