UNIVERSITY OF
CAMBRIDGE
Judge Business School

MPO1 Workbook 2
Michaelmas 2011

- Answer all questions in no more than 11 pages (excluding appendix). Please be brief. There is no credit at all for extended answers.
- The weighting of Workbook 2 is 60% of the total for this module.
- The figure in square brackets after each question is the weight carried in Workbook 2.
- Documentation of your analysis: Your empirical analysis must be capable of being replicated. For this reason, you should annotate and include your $R-script$ file (or batch file in any other statistical package that you prefer to use) in the appendix. If you use packages that do not have batch command facility, you should report the sequence of your computational steps, such as data transformations etc in the appendix. This document / your batch file / your $R-script$ file) should be printed and attached as the **only** appendix.
- Note: http://www.intranet.jbs.cam.ac.uk/students/plagiarism.html
- Two copies of the completed work, together with a standard signed cover-sheet must be submitted to Ms. Nena Nelson at the MPhil Office by noon on 17 January 2012. Electronic submission is not permitted.
- If work is submitted after the deadline without valid tutorial reasons the mark for the submission will be reduced by $10n$ percentage points, where n is the number of whole, or part, weeks late.

Exercise 1

Consider the data *ex1.csv* on gasoline consumption, price of gasoline, and real income over the years 1970-99. Estimate the following model using the data over the period 1970-95:

$$gc = c + \beta_1 pg + \beta_2 ri + error$$

Where:

gc	log of nominal retail sales of gasoline service stations in millions of USD (gasoline retail price, city average, cents per gallon)
pg	log of gasoline retail price, city average, cents per gallon (consumer price index, indexed so that 1984=100)
ri	log of nominal personal disposable income in billions of USD (consumer price index, 1984=100)

(i) Carry out a test for heteroskedasticity.

(ii) Carry out a test on normality of the residuals.

(iii) Comment very briefly on the results.

[7.5%]

Exercise 2

Data from a regular survey of international prices and wages in major cities around the world (57 cities in 2000) carried out by *UBS* is presented in *ex2.csv*. This contains the price of *BigMac* measured in terms of the natural logarithm of minutes of labour required by an average worker to buy a *Big Mac*. The issue is to explain the variation in this variable using the following predictor variables (all these variables are expressed in natural logarithms as the original variables have highly skewed distributions).

labour	log of the minutes of labour required by an average worker to buy a Big Mac
rel.wage	log of average net wage relative to Zurich = log(100)
rice	log of labour minutes required by an average worker to buy 1 kg rice
vacation	log of average paid vacation days per year
bread	log of labour minutes required by an average worker to buy 1 kg bread.
Asia*	Indicator variable: 1 if city is in Asia, 0 otherwise.
ENO*	Indicator variable: 1 if city is in Europe, North America or Oceania, 0 otherwise.

* The reference (omitted) region is Africa and South America.

(i) Fit a multiple regression model with all the above variables entering in linear form. Are the regression assumptions satisfied?

(ii) Consider interactions to examine whether the effects of the log transformed explanatory variables vary according to region. Can some set of interactions be removed from the model without reducing the ability of the model to explain the dependent variable?

(iii) Estimate the best model you can fit, and examine whether the regression assumptions are satisfied.

(iv) Test the conjecture that as vacation days increase, the cost of a Big Mac goes down, *ceteris paribus*.

(v) Test the conjecture that as the cost of bread increases, the cost of a Big Mac goes up, *ceteris paribus*.

[15%]

Exercise 3

The file *ex3.csv* contains data on 40 students that can be used to determine whether performance in terms of *GPA* can be predicted from:

verbal	verbal score in the entrance examination (percentile)
mathematics	mathematics score in the entrance examination (percentile)

(i) Fit a simple linear model and assess the zero conditional mean error assumption.

(ii) Fit the full quadratic model by including also the squares of the explanatory variables and the interaction between them and assess the regression assumptions.

(iii) Are there any outliers?

(iv) Are there any influential observations?

[12.5%]

Exercise 4

This exercise confronts you with a simple research problem to be attempted over a relatively short period of time, with the results presented in the form of a **very** short paper.

firms.csv contains data relating to a sample consisting of 678 medium and large firms drawn from four countries, for the year 2001. The variables are from the company accounts of these firms and include:

country	country	*sales*	sales
public	publicly listed firms	*tanfixassts*	tangible fixed assets
uncons	unconsolidated accounts	*materials*	costs of materials used
fmcn50_ceo	family owned with family CEO	*age*	firm age
roce	return on capital employed	*hours*	avg hours worked / week
sic2	SIC 2-digit industries	*employment*	employment

You are required to explain the determinants of corporate profitability. This note is intended to provide some guidance on this exercise. The basic structure of your paper should be as below.

1) *Introduction*: The purpose of the introduction is to introduce the reader to the objective of the paper. The introduction may contain

   (i) Motivation: A very brief introduction of the topic for the benefit of a general reader.

   (ii) The research objective: Explain the objective of your paper. Which research question(s) are to be answered?

2) *Framework*: Present a simple conceptual framework for this analysis. Cut to the chase and explain very briefly any conjecture(s) and implications. What is the intuition behind your research? Are there any testable implications? What is an appropriate regression model?

3) *Empirical analysis*: Present your empirical analysis. The first sub-section should describe the data briefly using useful descriptive statistics or graphs. You should use your judgement on how much and what summary statistics / graphs you present.

   Like all datasets, this dataset may have limitations from the point of view of answering the question of interest, and these can be usefully pointed out.

   Present your regression results. Explain the model you have estimated. Discuss your results first from a statistical point of view (e.g. diagnostics, significance of coefficients etc.) and then, from a conceptual point of view. Do you accept or reject the premise you set out to test?

   All figures and tables should be put in the text. Ensure that the reader can understand what is shown in the figure or table using labels.

4) *Summary and conclusions*: What are the major findings of your empirical analysis? What can we learn from them? Any ideas for extensions, modifications or further investigations?

What is expected from you?
I do not expect the application of any techniques we have not covered in the computer lab sessions. The goal is to illustrate that you are able to perform your own independent quantitative study and to document your results in an appropriate way.

Length and form of the paper
The paper should not exceed 1500 words. One page of figures and graphs will count as 250 words. Please use double-spacing and a 12 point font.

[65%]