Econometric Analysis and Applications
Econometric Analysis and Applications
Module Introduction and Overview

Contents

1 Introduction to the Module 2
2 The Module Authors 2
3 Study Materials 3
3 Module Overview 5
4 Learning Outcomes 7
6 Assessment 8
Specimen Examination 15
1 Introduction to the Module

*Econometric Analysis and Applications* is the second econometrics module offered to MSc students who need to broaden their understanding of the application of quantitative methods to inquiry in finance or economics. This module assumes that you have studied the classical linear regression model at an introductory level and that you are familiar with the assumptions that underlie that model. You will be aware that there are many cases in which these assumptions are not satisfied, and know how such problems as heteroscedastic disturbances and autocorrelated errors can be detected, and what can be done about them. It is assumed, too, that you have a basic working knowledge of the econometric software, *EViews*, which was introduced in the previous module, although basic instructions for using the program are provided in this module too.

The purpose of this module is to broaden your knowledge and extend your understanding of econometrics. In doing this, you will work with data. The first two units extend your knowledge of single equation methods. Unit 1 considers how to make progress with dummy – that is, qualitative – regressors. Unit 2 introduces dynamic models by showing how lags and expectations can be incorporated. The following three units focus on models that consist of two or more equations – simultaneous equation models. The nature of such systems is explained and their identification and estimation discussed. The analysis of dynamic models is extended in Unit 6, where the times series properties of variables are discussed. By understanding the times series properties of variables, you will be able to specify dynamic econometric models that capture both short- and long-run effects. These are discussed in Unit 7. The final unit, Unit 8, focuses on forecasting with both econometric and time series models.

2 The Module Authors

This module was originally conceived and written by Graham Smith with contributions from Caroline Dinwiiddy, Linda Hesselman and John Nankervis. The exercises were based on an econometrics package, *Microfit*, which has now been replaced by much more powerful software, *EViews*, and the module has been revised to take account of this change.

The module has been revised by Dr Jonathan Simms, who is a tutor for CeFiMS, and has taught at University of Manchester, University of Durham and University of London. He has contributed to development of various CeFiMS modules including *Econometric Principles & Data Analysis, Financial Econometrics, Risk Management: Principles & Applications, Public Financial Management: Reporting and Audit, and Introduction to Law and to Finance*.

The module, and its more basic predecessor, *Econometric Principles and Data Analysis*, were designed and written by Dr Graham Smith, who is Senior Lecturer in the Department of Economics, SOAS, where he teaches econometrics to MSc students and carries out research on empirical
finance. His main research interests focus on emerging stock markets and he has published extensively in international refereed journals. His recent research demonstrates that stock market efficiency is determined by market size, liquidity and the quality of markets.

3 Study Materials

These module units are your central learning resource; they structure your learning unit by unit. Each unit should be studied within a week. The module units are designed in the expectation that studying the unit and the associated readings in the textbook, and completing the exercises, will require 15 to 20 hours during the week.

📚 Textbook

In addition to the module units you must read the assigned sections from the textbook, which is provided with your module materials:


This textbook aims to teach econometric principles in ‘an intuitive and informative way without resorting to matrix algebra, calculus or statistics beyond the introductory level’. In the module units which draw on the textbook, there is a section, called Study Guide. This leads you through the relevant parts of the textbook, and helps you to read and understand the analysis presented there.

This is an excellent textbook on econometrics, and the quality and detail of explanation are well suited to this module. The examples in the textbook are drawn from finance, financial economics, and economics. In the module units the examples and end-of-unit exercises are drawn predominantly from applications in finance. Units 6, 7 and 8 provide an enhanced level of explanation, analysis and discussion, with optional readings from the textbook.

💻 EViews

This module will use EViews, Student Edition, econometrics software that you will use to do the exercises in the units, and also the data analysis part of your assignments. The results presented in the units are also from EViews.

If you have studied the previous econometric module, Econometric Principles and Data Analysis, you will already be well practised in using the software.

To download and install your copy of EViews Student Version you will be sent a licence number and access code URL. To use the 24-character licence number which will look something like:

E9000XXX – XXX2XX7X – YY5YY3Y

You will need to go to: http://www.eviews.com/download/student9
Instructions to install EViews, and to register your copy of the software, are included in the file, EViews 9 Student Version_Installation and registration instructions.pdf which is available on the VLE and also at http://www.eviews.com/download/student9/EViews%20Student%20Version.pdf

Your student edition of EViews will run for two years after installation, and you will be reminded of this every time you open the program.

You will be asked to register your copy of EViews the first time you use it, after installing it on your computer. You will also be invited to run EViews Update to check for any updates to the software. You can run EViews Update at any time using Help>EViews Update … from within EViews.

EViews is very easy to use and you can operate it in a number of ways:

- there are drop-down menus
- selecting an object and then right-clicking provides a menu of available operations
- double-clicking an object opens it
- keyboard shortcuts work.

There is also the option to work with Commands; these are short statements that inform the program what you wish to do, and once you have built up your own vocabulary of useful Commands, this can be a very effective way of working. You can also combine all of these ways of working with EViews.

In each unit there are instructions to help you use EViews to do the exercises. In addition, EViews includes help files, which you can read as pdf files, or navigate via the EViews help and search facility. Unit 1 includes a section introducing EViews.

Although easy to use, EViews is a very powerful program. There are advanced features that you will not use on this module, and you should not be worried if you see these, either in the menus or the help files. The best advice is to stay focused on the subject that is being studied in each unit, and to do the exercises for the unit; this will reinforce your understanding and also develop your confidence in using data and EViews.

**Exercises**

As already noted, there are exercises in every unit. These require you to work with EViews and data files, available from the VLE in the module area for this study session, to do your own econometric analysis. It is very important that you attempt these exercises, and do not just look at the Answers at the end of the units. Your understanding of the material you have studied in the unit will be greatly improved if you do the exercises yourself. You will also develop better understanding and confidence in using EViews.
Audio module guide

There is an audio guide to accompany this module, in which Professor Pasquale Scaramozzino, the Academic Programme Director for Quantitative Finance, and Jonathan Simms discuss how the module extends your understanding of econometrics. The guide also provides useful advice on how to study the module materials. The guide is available in the module area on your VLE.

We advise that you listen to the guide before you start your study of the module. You may also find it useful to listen to the guide again before you study Units 6 and 7, especially the sections on stationarity and cointegration. Here is a brief summary of the content, with timings (the total length is 18:53 minutes).

The guide begins with a discussion of how *Econometric Analysis & Applications* relates to your study of *Econometric Principles & Data Analysis*, and how the module extends and deepens your understanding of econometrics (00:44). Dr Simms then provides an intuitive explanation of stationarity, nonstationarity, and cointegration techniques (02:31). Following this there is a summary of the main points in relation to dynamic models, the short and long run, and applications in finance (07:12). There is a short introduction to the textbook, *Basic Econometrics* (08:25), followed by practical advice on how to study the module materials and how to work on your assignments, emphasising the importance of good record keeping and making notes as you proceed (09:37). As you study, it may be useful to attempt to apply the methods you are learning in contexts with which you are familiar, to improve your insight and interpretation, and this point is made in the guide (14:01). Professor Scaramozzino and Dr Simms then consider how the module enables you to develop a more critical understanding of econometrics (15:59), and provide some final advice to students who are about to study C332 (17:16).

4 Module Overview

The module follows the usual eight-unit structure, and the topics covered are noted below.

**Unit 1** Dummy Variables

1.1 Introduction
1.2 The Use of Dummy Variables
1.3 The Chow Test for Parameter Stability
1.4 Unit Study Guide
1.5 Example – Long-Term Trends in Terms of Trade
1.6 Summary
1.7 Exercises
1.8 Answers to Exercises

**Unit 2** Dynamic Models – Lags and Expectations

2.1 Ideas and Issues
2.2 Lags
5 Learning Outcomes

After studying this module you will be able to:

• specify dummy variables to measure qualitative influences in regression analysis
• explain the use of intercept and slope dummy variables
• use and interpret the Chow test of parameter stability
• explain the nature of the ‘dummy variable trap’ and how to avoid it
• explain finite distributed lag models, including immediate impact, long-run reactions and mean lag
• implement the Koyck transformation
• explain and discuss the adaptive expectations hypothesis and its limitations
• discuss the properties of estimators of distributed lag and autoregressive models
• implement both Durbin’s h test and the LM test of autocorrelation and interpret the results
• explain and implement the Granger test of causality
• explain ‘simultaneous equation bias’
• interpret in a model the behavioural equations, definitions or identities, and equilibrium conditions
• identify conditions for stability in dynamic simultaneous equation systems
• explain the identification problem
• discuss the implications of equations which are exactly identified, overidentified, and not identified
• explain and apply indirect least squares
• explain the properties of the OLS estimator of the slope coefficients of a structural equation from a simultaneous system
• explain the method of ILS, implement it in appropriate situations, and discuss the properties of ILS estimators
• explain and discuss the method of two-stage least squares (TSLS or 2SLS), implement it for an identified equation with EViews, and outline the properties of 2SLS estimators
• discuss what is meant by stationary and nonstationary time series and provide examples of each
• explain the Dickey-Fuller and Augmented Dickey-Fuller tests of the hypothesis that a series is I(1)
• using EViews, produce correlograms, implement Dickey-Fuller, Augmented Dickey-Fuller and Phillips-Perron unit root tests for a single series, and interpret the results
• explain the nature of cointegration and the relationship between spurious regression and cointegration
• discuss and implement tests of cointegration
• explain, interpret and estimate error correction models
• explain the nature of vector autoregressions (VARs)
• discuss and carry out Johansen cointegration tests
• explain the nature of autoregressive and moving average processes
• define an ARIMA model, and use it to forecast
• interpret the measures of forecast evaluation provided by EViews
• explain the relationship between a model in structural, restricted and unrestricted reduced form
• calculate forecasts using econometric models, including static and dynamic single equation and simultaneous systems.

6 Assessment

Your performance on each module is assessed through two written assignments and one examination. The assignments are written after Unit 4 and Unit 8 of the module session. Please see the VLE for submission deadlines. The examination is taken at a local examination centre in September/October.

Preparing for assignments and exams

There is good advice on preparing for assignments and exams and writing them in Chapter 8 of Studying at a Distance by Christine Talbot. We recommend that you follow this advice.

The examinations you will sit are designed to evaluate your knowledge and skills in the subjects you have studied: they are not designed to trick you. If you have studied the module thoroughly, you will pass the exam.

Understanding assessment questions

Examination and assignment questions are set to test your knowledge and skills. Sometimes a question will contain more than one part, each part testing a different aspect of your skills and knowledge. You need to spot
the key words to know what is being asked of you. Here we categorise the types of things that are asked for in assignments and exams, and the words used. All the examples are from the Centre for Financial and Management Studies examination papers and assignment questions.

Definitions
Some questions mainly require you to show that you have learned some concepts, by setting out their precise meanings. Such questions are likely to be preliminary and be supplemented by more analytical questions. Generally, ‘Pass marks’ are awarded if the answer only contains definitions. They will contain words such as:

- Describe
- Define
- Examine
- Distinguish between
- Compare
- Contrast
- Write notes on
- Outline
- What is meant by
- List

Reasoning
Other questions are designed to test your reasoning, by explaining cause and effect. Convincing explanations generally carry additional marks to basic definitions. They will include words such as:

- Interpret
- Explain
- What conditions influence
- What are the consequences of
- What are the implications of

Judgement
Others ask you to make a judgement, perhaps of a policy or of a course of action. They will include words like:

- Evaluate
- Critically examine
- Assess
- Do you agree that
- To what extent does

Calculation
Sometimes, you are asked to make a calculation, using a specified technique, where the question begins:

- Use indifference curve analysis to
- Using any economic model you know
- Calculate the standard deviation
- Test whether

It is most likely that questions that ask you to make a calculation will also ask for an application of the result, or an interpretation.

Advice
Other questions ask you to provide advice in a particular situation. This applies to law questions and to policy papers where advice is asked in relation to a policy problem.
Your advice should be based on relevant law, principles and evidence of what actions are likely to be effective. The questions may begin:

- Advise
- Provide advice on
- Explain how you would advise

**Critique**

In many cases the question will include the word ‘critically’. This means that you are expected to look at the question from at least two points of view, offering a critique of each view and your judgement. You are expected to be critical of what you have read.

The questions may begin:

- Critically analyse
- Critically consider
- Critically assess
- Critically discuss the argument that

**Examine by argument**

Questions that begin with ‘discuss’ are similar – they ask you to examine by argument, to debate and give reasons for and against a variety of options, for example

- Discuss the advantages and disadvantages of
- Discuss this statement
- Discuss the view that
- Discuss the arguments and debates concerning

---

**The grading scheme: Assignments**

The assignment questions contain fairly detailed guidance about what is required. All assignments are marked using marking guidelines. When you receive your grade it is accompanied by comments on your paper, including advice about how you might improve, and any clarifications about matters you may not have understood. These comments are designed to help you master the subject and to improve your skills as you progress through your programme.

**Postgraduate assignment marking criteria**

The marking criteria for your programme draws upon these minimum core criteria, which are applicable to the assessment of all assignments:

- understanding of the subject
- utilisation of proper academic [or other] style (e.g. citation of references, or use of proper legal style for court reports, etc.)
- relevance of material selected and of the arguments proposed
- planning and organisation
- logical coherence
- critical evaluation
- comprehensiveness of research
- evidence of synthesis
- innovation/creativity/originality.
The language used must be of a sufficient standard to permit assessment of these.

The guidelines below reflect the standards of work expected at postgraduate level. All assessed work is marked by your Tutor or a member of academic staff, and a sample is then moderated by another member of academic staff. Any assignment may be made available to the external examiner(s).

80+ (Distinction). A mark of 80+ will fulfil the following criteria:

- very significant ability to plan, organise and execute independently a research project or coursework assignment
- very significant ability to evaluate literature and theory critically and make informed judgements
- very high levels of creativity, originality and independence of thought
- very significant ability to evaluate critically existing methodologies and suggest new approaches to current research or professional practice
- very significant ability to analyse data critically
- outstanding levels of accuracy, technical competence, organisation, expression.

70–79 (Distinction). A mark in the range 70–79 will fulfil the following criteria:

- significant ability to plan, organise and execute independently a research project or coursework assignment
- clear evidence of wide and relevant reading, referencing and an engagement with the conceptual issues
- capacity to develop a sophisticated and intelligent argument
- rigorous use and a sophisticated understanding of relevant source materials, balancing appropriately between factual detail and key theoretical issues. Materials are evaluated directly and their assumptions and arguments challenged and/or appraised
- correct referencing
- significant ability to analyse data critically
- original thinking and a willingness to take risks.

60–69 (Merit). A mark in the 60–69 range will fulfil the following criteria:

- ability to plan, organise and execute independently a research project or coursework assignment
- strong evidence of critical insight and thinking
- a detailed understanding of the major factual and/or theoretical issues and directly engages with the relevant literature on the topic
- clear evidence of planning and appropriate choice of sources and methodology with correct referencing
- ability to analyse data critically
- capacity to develop a focussed and clear argument and articulate clearly and convincingly a sustained train of logical thought.
50–59 (Pass). A mark in the range 50–59 will fulfil the following criteria:

- ability to plan, organise and execute a research project or coursework assignment
- a reasonable understanding of the major factual and/or theoretical issues involved
- evidence of some knowledge of the literature with correct referencing
- ability to analyse data
- examples of a clear train of thought or argument
- the text is introduced and concludes appropriately.

40–49 (Fail). A Fail will be awarded in cases in which there is:

- limited ability to plan, organise and execute a research project or coursework assignment
- some awareness and understanding of the literature and of factual or theoretical issues, but with little development
- limited ability to analyse data
- incomplete referencing
- limited ability to present a clear and coherent argument.

20–39 (Fail). A Fail will be awarded in cases in which there is:

- very limited ability to plan, organise and execute a research project or coursework assignment
- failure to develop a coherent argument that relates to the research project or assignment
- no engagement with the relevant literature or demonstrable knowledge of the key issues
- incomplete referencing
- clear conceptual or factual errors or misunderstandings
- only fragmentary evidence of critical thought or data analysis.

0–19 (Fail). A Fail will be awarded in cases in which there is:

- no demonstrable ability to plan, organise and execute a research project or coursework assignment
- little or no knowledge or understanding related to the research project or assignment
- little or no knowledge of the relevant literature
- major errors in referencing
- no evidence of critical thought or data analysis
- incoherent argument.

The grading scheme: Examinations

The written examinations are ‘unseen’ (you will only see the paper in the exam centre) and written by hand, over a three-hour period. We advise that you practise writing exams in these conditions as part of your examination preparation, as it is not something you would normally do.

You are not allowed to take in books or notes to the exam room. This means that you need to revise thoroughly in preparation for each exam.
This is especially important if you have completed the module in the early part of the year, or in a previous year.

Details of the general definitions of what is expected in order to obtain a particular grade are shown below. These guidelines take account of the fact that examination conditions are less conducive to polished work than the conditions in which you write your assignments. Note that as the criteria of each grade rises, it accumulates the elements of the grade below. Assignments awarded better marks will therefore have become comprehensive in both their depth of core skills and advanced skills.

**Postgraduate unseen written examinations marking criteria**

**80+ (Distinction).** A mark of 80+ will fulfils the following criteria:
- very significant ability to evaluate literature and theory critically and make informed judgements
- very high levels of creativity, originality and independence of thought
- outstanding levels of accuracy, technical competence, organisation, expression
- outstanding ability of synthesis under exam pressure.

**70–79 (Distinction).** A mark in the 70–79 range will fulfil the following criteria:
- clear evidence of wide and relevant reading and an engagement with the conceptual issues
- develops a sophisticated and intelligent argument
- rigorous use and a sophisticated understanding of relevant source materials, balancing appropriately between factual detail and key theoretical issues
- direct evaluation of materials and their assumptions and arguments challenged and/or appraised;
- original thinking and a willingness to take risks
- significant ability of synthesis under exam pressure.

**60–69 (Merit).** A mark in the 60–69 range will fulfil the following criteria:
- strong evidence of critical insight and critical thinking
- a detailed understanding of the major factual and/or theoretical issues and directly engages with the relevant literature on the topic
- develops a focussed and clear argument and articulates clearly and convincingly a sustained train of logical thought
- clear evidence of planning and appropriate choice of sources and methodology, and ability of synthesis under exam pressure.

**50–59 (Pass).** A mark in the 50–59 range will fulfil the following criteria:
- a reasonable understanding of the major factual and/or theoretical issues involved
- evidence of planning and selection from appropriate sources
- some demonstrable knowledge of the literature
- the text shows, in places, examples of a clear train of thought or argument
- the text is introduced and concludes appropriately.
40–49 (Fail). A Fail will be awarded in cases in which:
- there is some awareness and understanding of the factual or theoretical issues, but with little development
- misunderstandings are evident
- there is some evidence of planning, although irrelevant/unrelated material or arguments are included.

20–39 (Fail). A Fail will be awarded in cases which:
- fail to answer the question or to develop an argument that relates to the question set
- do not engage with the relevant literature or demonstrate a knowledge of the key issues
- contain clear conceptual or factual errors or misunderstandings.

0–19 (Fail). A Fail will be awarded in cases which:
- show no knowledge or understanding related to the question set
- show no evidence of critical thought or analysis
- contain short answers and incoherent argument.

Specimen exam papers

CeFiMS does not provide past papers or model answers to papers. Modules are continuously updated, and past papers will not be a reliable guide to current and future examinations. The specimen exam paper is designed to be relevant and to reflect the exam that will be set on this module.

Your final examination will have the same structure and style and the range of question will be comparable to those in the Specimen Exam. The number of questions will be the same, but the wording and the requirements of each question will be different.

Good luck on your final examination.

Further information

Online you will find documentation and information on each year’s examination registration and administration process. If you still have questions, both academics and administrators are available to answer queries.

The Regulations are also available at www.cefims.ac.uk/regulations/, setting out the rules by which exams are governed.
Econometric Analysis and Applications

Specimen Examination

This is a specimen examination paper designed to show you the type of examination you will have at the end of this module. The number of questions and the structure of the examination will be the same, but the wording and requirements of each question will be different.

The examination must be completed in THREE hours. Answer FOUR questions.

The examiners give equal weight to each question; therefore, you are advised to distribute your time approximately equally over four questions.

Statistical tables are provided at the end of this examination paper.

Candidates may use their own electronic calculators in this examination provided they cannot store text; the make and type of calculator MUST BE STATED CLEARLY on the front of the answer book.
Answer **FOUR** questions.

1. **Answer **ALL** parts of this question.**

a) Explain the Chow test of parameter stability.

b) Using weekly data, a single-index model has been estimated for Kraft Foods Inc. over two consecutive subsamples. The estimated equations and associated sums of squared residuals (RSS) are:

\[ Y_t = 0.00034 + 0.526685X_t + \hat{u}_t \quad N_1 = 260 \quad RSS_1 = 0.185856 \]
\[ (0.001666) \quad (0.080533) \]

\[ Y_t = 0.001139 + 0.526145X_t + \hat{u}_t \quad N_2 = 263 \quad RSS_2 = 0.163028 \]
\[ (0.001541) \quad (0.046636) \]

in which \( Y \) is the weekly return on the stock of Kraft Foods Inc., \( X \) is the weekly return on the NYSE Composite Index, and standard errors are in brackets. When the equation is estimated for the combined data set, \( RSS_{1+2} = 0.349044 \).

Using the Chow test at the 0.05 significance level, test the hypothesis that parameters are stable over time.

c) How might dummy variables be used to test the hypothesis that both the intercept and slope coefficients have changed?

2. **Answer **ALL** parts of this question.**

a) Explain the adaptive expectations hypothesis.

b) The relation between the one-month interest rate and the expectation concerning the one-week interest rate may be modelled as

\[ Y_t = \alpha + \beta E_{t-1}X_{t+1} + u_t \]

where \( Y_t \) is the one-month interest rate at time \( t \); \( E_{t-1}X_{t+1} \) is the expectation, formed at time \( t \), of the one-week interest rate in the next period, that is at \( t + 1 \); and \( u_t \) is a disturbance term which satisfies the assumptions of the classical linear regression model.

Suppose expectations concerning future one-week interest rates are formed as follows:

\[ E_t X_{t+1} - E_{t-1}X_t = \gamma \left( X_t - E_{t-1}X_t \right) \]

Derive a regression equation relating \( Y_t \) to \( X_t, Y_{t-1}, u_t \) and \( u_{t-1} \).
c) The following equation for the one-month interest rate was estimated with monthly data over the period 1991M1-2011M6 (240 observations after adjustments), with standard errors in brackets
\[
\hat{Y}_t = 0.081 + 0.779X_t + 0.218Y_{t-1} \quad R^2 = 0.997 \quad DW = 1.131 \\
(0.021) \quad (0.029) \quad (0.029)
\]
i) Test the hypothesis that the disturbances in this regression do not have first-order serial correlation.

ii) Derive an estimate of the expectations adjustment parameter and interpret it.

iii) What is the long-run effect of a change in the one-week interest rate on the one-month interest rate? Comment on this result.

iv) Given your answer to part c.i., comment on the properties of the OLS estimators when used to estimate this autoregressive specification.

3. Answer ALL parts of this question.

a) In the context of a simultaneous equation model, explain clearly the differences between:

i) endogenous, exogenous and predetermined variables

ii) behavioural equations, equilibrium conditions and identities

iii) the structural form and reduced form of the model.

b) Consider the following market model
\[
Q_t^d = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t \\
Q_t^s = \beta_1 + \beta_2 P_t + \beta_3 P_{t-1} \\
Q_t^d = Q_t^s
\]
in which \(Q_t^d\), \(Q_t^s\) and \(P_t\) are endogenous and \(Y_t\) is exogenous.

i) Find the reduced form of the model;

ii) Find the final form of the model;

iii) Explain what determines the stability of the model.

PLEASE TURN OVER
4. Answer ALL parts of this question.

a) Explain clearly the meaning and significance of saying that an equation is ‘identified’.

b) Explain carefully the order and rank conditions for identification. Consider the following simultaneous equations system

\[
\begin{align*}
Y_{1t} &= \alpha_1 + \alpha_2 Y_{3t} + u_{1t} \\
Y_{2t} &= \beta_1 + \beta_2 X_{1t} + \beta_3 Y_{3t} + \beta_4 Y_{1t} + u_{2t} \\
Y_{3t} &= \gamma_1 + \gamma_2 X_{2t} + \gamma_3 X_{3t} + \gamma_4 Y_{2t} + u_{3t}
\end{align*}
\]

in which the \( Y \) are endogenous variables, the \( X \) are exogenous variables and the \( u \) are disturbances.

c) Use the order and rank conditions to examine the identification of the equations.

d) What are the implications for identification if \( \beta_4 = 0 \)?

5. Answer ALL parts of this question.

Comment on the results in parts (a), (b), and (c).

In your answers, explain and interpret the Figures and the estimated equations, and discuss the overall method being used. Conduct, explain, and report any hypothesis tests that you consider are relevant.

Part (a) – Figures 5.1 to 5.4, part (b) – equation (5.1), and part (c) – equation (5.2) relate to the natural log of a stock market index \( X \), for 200 observations. The equations have been estimated by OLS. \( N \) indicates the number of observations used in estimation, and \( t \)-ratios are in brackets.
b) \[ \Delta X_t = 0.035349 - 0.005568X_{t-1} + e_t \quad N = 199 \]  
\hspace{1cm} (t = 0.666700) \ (t = -0.634310)  
where \( \Delta X_t = X_t - X_{t-1} \), and \( e_t \) is the OLS residual in estimated equation (5.1). Using the Schwarz information criterion, no lagged values of \( \Delta X_t \) are included in equation (5.1). The appropriate critical values for the test represented in equation (5.1) are \(-3.463235\) (1% significance level); \(-2.875898\) (5%); \(-2.574501\) (10%).

c) \[ \Delta^2 X_t = 0.001575 - 0.920952\Delta X_{t-1} + \omega \quad N = 198 \]  
\hspace{1cm} (t = 0.535097) \ (t = -12.91881)  
where \( \Delta^2 X_t = \Delta X_t - \Delta X_{t-1} \), and \( \omega \) is the OLS residual in estimated equation (5.2). Using the Schwarz information criteria, no lagged values of \( \Delta^2 X_t \) are included in equation (5.2). The appropriate critical values for the test represented in equation (5.2) are \(-3.463405\) (1% significance level); \(-2.875898\) (5%); \(-2.574501\) (10%).
6. **Answer ALL parts of this question.**
   a) What are the principal characteristics of a recursive model? How would you obtain estimates of the parameters?
   b) Explain carefully one method for estimating the parameters of an overidentified equation in a system of simultaneous equations.
   c) What are the main features of two-stage least squares (2SLS) estimators?

7. **Answer ALL parts of this question.**
   a) What is meant by a spurious regression?
   b) What is the relationship between spurious regression and cointegration?
   c) Explain a test of cointegration based on OLS residuals.
   d) Explain the nature of a first-order error correction model (ECM).

8. **Answer ALL parts of this question.**
   a) Explain carefully how single equation econometric models can be used to generate forecasts.
   b) The following ARIMA(1, 1, 0) model has been estimated to period $T$.
      \[ \Delta Y_t = \hat{u} + \rho \Delta Y_{t-1} + \hat{u}_t \quad t = 1, 2, \ldots, T. \]
      How can this model be used to create forecasts for observations after period $T$?
   c) Explain any three measures of forecast accuracy and discuss their relative merits.

[END OF EXAMINATION]
Unit Content

This first unit shows how explanatory variables that are qualitative can be included in regression analysis, using dummy variables. In addition, two tests of parameter stability are explained and discussed – the Chow test and the dummy variable alternative to the Chow test.

You may remember from the Introduction that there is an audio guide to accompany this module. The guide is available in the course area on your VLE, and if you have not already done so then we advise that you listen to the guide now.

The guide begins with a discussion of how Econometric Analysis & Applications relates to your study of Econometric Principles & Data Analysis, and how the module extends and deepens your understanding of econometrics (00:44). Dr Simms then provides an intuitive explanation of stationarity, nonstationarity, and cointegration techniques (02:31). There is then a summary of the main points in relation to dynamic models, the short and long run, and applications in finance (07:12). There is a short introduction to the textbook (08:25) and practical advice on how to study the course materials and to work on your assignments (09:37). As you study, it may be useful to attempt to apply the methods you are learning in contexts with which you are familiar, to improve your insight and interpretation (14:01). We then consider how the module enables you to develop a more critical understanding of econometrics (15:59), and provide some final advice to students studying C332 (17:16).

Learning Outcomes

After studying this unit, the readings, and completing the exercises, you will be able to:

- specify dummy variables to measure qualitative influences in regression analysis
- include dummy variables in regression equations
- explain the use of intercept and slope dummy variables
- interpret regression output that includes dummy variables
- use and interpret the Chow test of parameter stability
- use and interpret the dummy variable test of parameter stability
- explain the nature of the ‘dummy variable trap’ and how to avoid it.

📖 Reading for Unit 1

1.1 Introduction

By the time you’ve reached this course, you should have developed a clear understanding of what econometrics is about. The econometrician is interested in analysing relationships between variables. The particular area of research is usually suggested by some aspect of finance or economic theory, and the contribution of econometrics is to provide a set of tools with which to analyse hypothesised relationships using data from the real world. Interpreting data can be quite difficult, and trying to identify relationships between variables successfully is where much of the skill – and interest – of econometrics lies.

In the hunt for relationships between variables we have so far developed two complementary techniques: firstly, we can use graphs or scatter plots to illuminate the relationships between pairs of variables; secondly, we can use the techniques of linear regression to form estimates of the numerical relationships between variables. As you will already have gathered, much econometric theory is concerned with procedures for assessing the significance, accuracy and precision of these numerical estimates.

An assumption that has been implicit in the examples discussed so far is that we can always obtain a set of numerical values for the variables we are interested in – whether these come from time series or cross-section data. Examples of regression analysis discussed in the previous course, Econometric Principles and Data Analysis, included models of stock returns, spot and forward exchange rates, and price-earnings ratios. In all these examples, the analysis was based on numerical information about prices, quantities, exchange rates, rates of return etc.

However, it is easy to think of important explanatory variables that are not numerical or quantitative in nature. For example, stock returns may be influenced by monthly effects, or the particular day of the week on which trading is taking place; yields on bonds may be influenced by whether the bonds are considered to be investment grade or not; we may wish to take into account the effect of being in one sector or industry compared to others; many economic and financial relationships are influenced by different government and regulatory policies; and the prices of many commodities are influenced by war.

The object of this unit is to show how explanatory variables that are qualitative, rather than quantitative, in nature (such as seasonality, industry grouping, regulatory policy or war) can be included in regression analysis. This is done by the use of what are known as dummy variables.
1.2 The Use of Dummy Variables

Intercept Dummies

To illustrate the way in which a dummy variable is used, let us consider the supply function for an export commodity. Assume that we wish to estimate a simple regression equation relating the supply of exports ($Y$) to the price of export crops relative to domestic food crops ($X$). Assume further that data are available for 1980–2009 (30 observations). The regression model is

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$  \hspace{1cm} (1.1)

However, it is known that for five years within the sample period (1994–98) there was a civil war in progress, which seriously disrupted the planting and marketing of export crops. Assume the scatter plot of exports against relative price shows the relationship in Figure 1.1, with the observations for the five war years 1994–98 falling below the other observations.

Figure 1.1 Scatter plot of exports against relative price

If we ignore the evidence of the graph and simply fit a regression to the complete set of numerical data on export levels and prices, we shall have an average relationship which does not describe the normal peacetime relationship as well as it would if we had taken the effect of the war years into account. On the evidence of the scatter plot, there are two separate relationships between $X$ and $Y$, one for the peacetime years and one for the years of civil war. Graphically, it looks as though we should fit two lines to the data depending on whether we are estimating a peacetime or a wartime export function. These two regression lines are illustrated in Figure 1.2. Note that the two lines are drawn with the same slope but different intercepts.
How can we estimate two linear relationships from one set of data? It is possible to allow for the influence of the civil war on the export supply function in the following way: we introduce a new dummy variable, $D$, into the regression model

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + u_i$$  \hspace{1cm} (1.2)

where $D = 0$ in years of peace, 1980-93 and 1999-2009,

and $D = 1$ in years of civil war, 1994-98

This means that when using Eviews a variable $D$ is included which takes the following values:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

What are the consequences for the regression? When $D = 0$ (i.e. in peacetime years), we shall be estimating

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 0 + u_i$$  \hspace{1cm} (1.3)

which reduces to

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$  \hspace{1cm} (1.4)

When $D = 1$ we shall be estimating the relationship

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 1 + u_i$$  \hspace{1cm} (1.5)

or

$$Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + u_i$$  \hspace{1cm} (1.6)
In other words, including the dummy variable has the effect of altering the value of the intercept and shifting the function up or down, depending on whether the observation in question falls in peacetime or not.

When the regression equation is estimated, the coefficient $\beta_3$ will be tested in the usual way; if $\beta_3$ is significantly different from zero, then the dummy variable representing the effect of the war years would appear to be a relevant explanatory variable; if $\beta_3$ is not significantly different from zero, we would conclude that the dummy variable should not be included in our export supply function.

With two categories, peace/war, we use one dummy variable taking two values 0/1 – not two dummy variables. A linear combination of two dummy variables results in a column of 1s, which would be perfectly collinear with the constant (a column of 1s). Including as many dummies as there are categories for the variables would lead to perfect multicollinearity and the breakdown of the regression procedure – the so called ‘dummy variable trap’.

It is normal, but not essential, for $D$ to take the value 0 or 1. The values (0, 1) can be mapped using

$$E = \alpha_1 + \alpha_2 D$$

When $D = 0$, $E = \alpha_1$

$D = 1$, $E = \alpha_1 + \alpha_2$

If $\alpha_1 = 1$, $\alpha_2 = 1$, then (0, 1) maps into (1, 2) and

if $\alpha_1 = 1$, $\alpha_2 = 2$, then (0, 1) maps into (1, 3).

That is, dummy variables are qualitative variables that do not have a natural scale of measurement. Usually, they take values of 0 when a characteristic is absent and 1 when it is present.

Dummy variables are often used to pick up seasonal influences on variables. In finance we often observe a year-end effect, and, depending on the timing of the tax year-end, this is referred to as the January effect. Towards the end of the tax year, portfolio managers sell shares that have declined in value, so that the losses reduce the taxes payable. However, the selling of shares lowers stock returns. As the new tax year begins, investors begin buying shares and stock returns increase. To take another example, an economy with a large agricultural sector is likely to show considerable fluctuations in output from one season to another, and these fluctuations in output may well be reflected in other statistical series such as those for employment, tax revenues and balance of payments. Consumption patterns, too, may show seasonal fluctuations. These may be related to the weather – demand for ice-cream and soft drinks is likely to be higher in hotter seasons of the year – or they may be related to cultural factors – patterns of expenditure may be affected by, for example, the month of Ramadan or preparations for Christmas.
For an example of the analysis of seasonal factors let us consider a model of daily stock returns. To keep the analysis brief we will examine the hypothesis that there are quarterly effects (if we were considering the hypothesis of a January effect we would examine monthly effects). Assume that the daily return on a stock market index, $Y_t$, is affected by the return on the index on the previous day, $Y_{t-1}$, and it is also known that seasonal factors are important. To examine the effects of four quarters in the year we would have to include three dummies in the model

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \gamma_4 D_{4t} + u_t$$  \hspace{1cm} (1.8)$$

where $D_2 = 1$ if the trading day is in the second quarter and $D_2 = 0$ otherwise;

$D_3 = 1$ if the trading day is in the third quarter and $D_3 = 0$ otherwise; and

$D_4 = 1$ if the trading day is in the fourth quarter and $D_4 = 0$ in all other quarters.

In the first quarter, $D_2 = D_3 = D_4 = 0$ and the restricted model of stock market returns is

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t.$$ \hspace{1cm} (1.9)$$

The three restrictions are $\gamma_2 = \gamma_3 = \gamma_4 = 0$.

For the second quarter, $D_2 = D_4 = 0$ and we have

$$Y_t = (\beta_1 + \gamma_2) + \beta_2 Y_{t-1} + u_t.$$ \hspace{1cm} (1.10)$$

In the third quarter, $D_2 = D_3 = 0$ and

$$Y_t = (\beta_1 + \gamma_3) + \beta_2 Y_{t-1} + u_t.$$ \hspace{1cm} (1.11)$$

Finally, for the fourth quarter, $D_2 = D_3 = 0$ and the regression function for stock market returns is

$$Y_t = (\beta_1 + \gamma_4) + \beta_2 Y_{t-1} + u_t.$$ \hspace{1cm} (1.12)$$

Testing the significance of seasonal effects is straightforward. Estimate (1.8) and (1.9) by OLS and carry out an $F$ test on the set of $\gamma$s. The null hypothesis is $\gamma_2 = \gamma_3 = \gamma_4 = 0$, and so the intercept $\beta_1$ is appropriate for all four quarters. That is, the set of seasonal dummy variables does not ‘improve’ the model. The alternative hypothesis is that at least one of the $\gamma_j \neq 0$. The test is carried out by viewing the regression (1.9) as the restricted version of (1.8) with $r = 3$ being the number of restrictions. Both models are estimated using the full sample. Regression (1.8) has residual sum of squares $RSS_0$ (and $k = 5$ parameters) and the restricted model (1.9) has residual sum of squares $RSS_r$.

The test statistic

$$\frac{(RSS_r - RSS_0) / r}{RSS_0 / (N-k)} \sim F_{(r,N-k)}$$ \hspace{1cm} (1.13)$$
if the null hypothesis is true. If the calculated value of the test statistic is greater than the critical value at a predetermined significance level, say 0.05, then the null hypothesis is rejected.

Just as in our previous example, the number of dummies included in the regression equation must always be one less than the number of categories. In our first example, there were two categories – ‘peacetime’ and ‘wartime’ – so one dummy was included; in the second example we were considering four quarters of the year so three dummies were included. If we wished to consider the seasonal impact of every month of the calendar year, we would have had to include eleven dummy variables to avoid the dummy variable trap.

Occasionally in time series work, a dummy variable is included which takes the value 1 for one time period and 0 for all other time periods – in effect, a dummy variable for a single observation. This is equivalent to deleting that particular row of data, and often the resulting distribution of residuals is much better behaved and more precise estimates are obtained. This technique is useful in the case of strikes, natural disasters, the impact effect of policy changes, etc. A typical example would be to use such a dummy to capture the quarter in which the Organization of Petroleum Exporting Countries (OPEC) oil embargo occurred, 1973Q4. The estimated coefficient on this dummy variable is the prediction error for the period. The null hypothesis that the single observation is ‘consistent’ with the estimated relationship can be readily examined with a t test.

Consider the following example of the use of a dummy variable taken from the field of financial economics, discussed by HH Kelejian and WE Oates in their textbook Econometrics: Principles and Applications:

The range of application of dummy variables is virtually unlimited. An example [...] comes from a recent study by one of the authors.1 The issue under study was whether or not the formal character of a country’s political constitution has a systematic impact on the extent of decentralisation in the nation’s public finances. Or, in short, is the constitution itself important in determining the relative share of fiscal activity of the central government in the public sector as a whole? Having already ascertained the importance of certain other variables (such as population size and the level of per capita income), the procedure was to introduce a dummy variable with a value of one if the country had a ‘federal’ constitution (i.e., one guaranteeing some political autonomy to decentralised levels of government) or a value of zero in the absence of a federal constitution (i.e., where the scope of authority of decentralised levels of government is determined by the central government itself). Using cross-sectional data for a sample of 53 countries, the estimated equation was

\[
\hat{G} = 96 + 1.21 \ln P - 0.004X - 0.60Z - 15.9F
\]

\[
(1.14)
\]

\[
N = 53, R^2 = 0.95 \text{ and (absolute value of the } t \text{ ratio)}
\]

\[
(12.1)(1.3) \quad (2.3) \quad (5.3) \quad (4.7)
\]

\[1 \text{ Oates (1972) Chapter 5.}\]
where
\( \hat{G} \) is central-government share of total public revenues (as a percentage);
\( \ln P \) is the logarithm of population size (in thousands);
\( Y \) is per capita income in 1965 US$s;
\( Z \) is social security contributions as a percentage of total public current revenue;
\( F \) is a dummy variable taking the value 1 for countries with federal constitutions and 0 for countries with non-federal constitutions.

The results in the equation are clearly consistent with the hypothesis that the existence of a federal constitution contributes to an increased degree of decentralisation in the public finances. The coefficient of the dummy variable, \( F \), is negative and possesses a t ratio in excess of 4, so that we can easily reject the null hypothesis of no association between \( G \) and \( F \) at the usual 0.05 significance level. The magnitude of the coefficient suggests that, after allowing for the effects of population size, income, and so on, the central government in federal countries collects, on average, about 16 percentage points less of total public revenues than do the central governments in countries without federal constitutions. The constraints imposed formally by political constitutions would thus appear to be of considerable importance in determining the degree of decentralisation in public fiscal activity.


**Slope Dummies**

The dummy variables discussed so far have all been intercept dummies, which had the effect of shifting the function up or down by causing the intercept or constant term in the regression to take different values. The implicit assumption was that the relationships between the explanatory variables and the dependent variable (represented by the derivative or slope of the function in simple linear regression and by the partial derivatives in multiple regression) were not affected by the inclusion of a qualitative explanatory variable.

Sometimes, however, it might be thought that slope coefficients (for example, company beta coefficients, marginal propensities or elasticities) are likely to change in response to changes in circumstances.

Consider a single-index model relating the return on the shares of a particular company (\( Y \)) to the returns on a relevant stock market index income (\( X \)). The simple regression model has the form
\[
Y_t = \beta_1 + \beta_2 X_t + u_t
\]  
(1.15)
and the company’s beta coefficient (the extent to which the return on the company stock varies in response to changes in the return on the index) is measured by the regression coefficient \( \beta_2 \). Suppose that the sample data come from time series observations and that a change in the company’s beta coefficient is thought to have occurred after a particular year because of a change in the economic and financial environment. Such a change might be
noticed in a number of countries following oil price shocks, or shocks in financial markets which have lasting effects, for example the financial crisis of 2008. The longer-term effects of a financial crisis could be modelled by including a slope dummy variable in the regression in the following manner:

\[ Y_t = \beta_1 + \beta_2 X_t + \beta_3 D_t X_t + u_t \]  \hspace{1cm} (1.16)

where

\[ D_t = 0 \text{ for sample observations before 2008} \]

\[ D_t = 1 \text{ for sample observations for 2008 and subsequent years.} \]

The effect of including a slope dummy variable is that two separate relationships are estimated. If \( D_t = 0 \), then we have (equation 1.15):

\[ Y_t = \beta_1 + \beta_2 X_t + u_t \]

and if \( D_t = 1 \) then the model becomes

\[ Y_t = \beta_1 + (\beta_2 + \beta_3) X_t + u_t \]  \hspace{1cm} (1.17)

and the new, post-2008, beta coefficient is given by the coefficient \( \beta_2 + \beta_3 \). Note that to estimate model (1.16) in Eviews, it is possible to specify an equation including \( D_t \times X_t \) as a regressor, and estimate (1.16) directly, or you can define a new variable, \( Z_t = D_t \times X_t \) and estimate

\[ Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + u_t \]  \hspace{1cm} (1.18)

Whether you estimate (1.16) directly, or estimate (1.18), if the null hypothesis that \( \beta_3 = 0 \) is not rejected, the beta coefficient, represented by the slope of the linear single-index model, is not affected by the dummy variable in question.

The above analysis of the single-index model was based on time series analysis, but dummy variables can be used in regressions based on either time series or cross-section data.

Rachev et al. (2007) consider a model of corporate bond spreads for a cross-section of 100 firms. The spread for a corporate bond is the difference between the interest rate on the corporate bond and the interest rate on government bonds of equivalent maturity. Interest rates on corporate bonds are higher than those on government bonds to reflect expected default loss; different tax treatments (interest payments on government bonds are often not taxable); and the riskier return associated with corporate bonds. The analysis considers whether there is a difference in the regression explaining bond spreads for two cases: investment-grade bonds and bonds that are not investment grade (rated CCC+ and below). The following regression model is used

\[ S_i = \beta_0 + \beta_1 D_i + \beta_2 C_i + \beta_3 D_i C_i + \beta_4 CR_i + \beta_5 D_i CR_i + \beta_6 \ln EBIT_i + \beta_7 D_i \ln EBIT_i + u_i \]  \hspace{1cm} (1.19)
\( S_i \) is the option-adjusted spread for the bond issue of company \( i \), measured in basis points (100 basis points = 1 per cent).

\( C_i \) is the coupon rate for the bond of company \( i \), expressed in percentage terms; a higher coupon rate increases default risk, and increases the spread.

\( CR_i \) is the coverage ratio, which is earnings before interest, taxes, depreciation and amortisation, divided by interest expenses; a higher coverage ratio reduces the risk of default, and therefore is associated with a lower spread.

\( EBIT_i \) is earnings before interest, measured in millions of dollars, for the preceding 12 months; higher earnings are expected to reduce default risk, and therefore reduce the spread.

\( D_i \) is a dummy variable which equals zero if the bonds of company \( i \) are investment grade, and equals 1 if the bonds of company \( i \) are considered not to be investment grade, that is they are rated as CCC+ and lower.

Notice in (1.19) that the dummy variable affects the intercept and the three slope coefficients. Data are available for 100 companies on two dates, giving 200 observations in total. (The authors also present results for a separate equation which instead includes a dummy variable to test the hypothesis that the regression estimation will be affected by the two different dates. Can you think how that could be done?)

The estimated coefficients for equation (1.19) obtained by Rachev et al. (2007, p. 137) are as follows.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Standard error</th>
<th>t-statistic</th>
<th>Prob. value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>284.52</td>
<td>73.63</td>
<td>3.86</td>
<td>0.00</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>597.88</td>
<td>478.74</td>
<td>1.25</td>
<td>0.21</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>37.12</td>
<td>7.07</td>
<td>5.25</td>
<td>0.00</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>-45.54</td>
<td>38.77</td>
<td>-1.17</td>
<td>0.24</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>-10.33</td>
<td>1.84</td>
<td>-5.60</td>
<td>0.00</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>50.13</td>
<td>40.42</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td>( \beta_6 )</td>
<td>-83.76</td>
<td>13.63</td>
<td>-6.15</td>
<td>0.00</td>
</tr>
<tr>
<td>( \beta_7 )</td>
<td>-0.24</td>
<td>62.50</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Review Question**

Considering the intercept and each of the slope parameters separately, is there a significant effect from a corporate bond not having an investment-grade rating?

The answer is no. The Prob. values for the effect of the dummy variable on the intercept, and the interactions with the coupon rate, coverage ratio, and
earnings, are all greater than 0.05. We would not reject the null hypothesis that, individually, these estimated coefficients are not different from zero. However, the authors do find that the dummy variable terms are jointly significant, using an $F$-test.

### 1.3 The Chow Test for Parameter Stability

The consequence of including dummy variables in regression is essentially that we estimate two or more regressions simultaneously. There is, incidentally, no reason why several dummy variables should not be included in a regression and both intercept and slope dummies can be used in the same regression. The only problem (which is not specific to the inclusion of dummy variables) is that adding too many additional explanatory variables reduces the degrees of freedom and this may be a serious problem with a small data set.

An alternative approach to the question of whether the same parameters are appropriate for the entire data set is to test for parameter instability. The usual test is the Chow test, which is an $F$-test. Consider, for example, a test on the single-index model discussed in the last section. The data before and after 2008 are considered as two periods. The Chow test assumes that the disturbance terms for the first and second periods are both normally distributed with zero mean and the same variance and also independently distributed. Three separate regressions are run, as summarised below.

<table>
<thead>
<tr>
<th>Observations</th>
<th>$RSS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>the entire data set</td>
<td>$N$</td>
</tr>
<tr>
<td>the period before any parameters are believed to change</td>
<td>$n_1$</td>
</tr>
<tr>
<td>the period after at least one parameter is believed to change</td>
<td>$n_2$</td>
</tr>
</tbody>
</table>

Each regression equation includes $k$ parameters.

The Chow test compares the residual sum of squares from a regression run on the entire data set with the residual sums of squares resulting from two separate regressions on two sub-groups within the sample. That is the test compares $RSS_X$ with $RSS_{n_1} + RSS_{n_2}$.

If the two values are close, the same parameters are appropriate for the entire data set; the parameters are stable. The point of the $F$-test is to see whether the residual sum of squares (which measures the variation in the data not explained by the regression) is significantly reduced by fitting two separate regressions rather than just one.

Given for period 1: \[ Y_t = \alpha_1 + \alpha_2 X_t + u_{1t} \] (1.20)

and for period 2: \[ Y_t = \beta_1 + \beta_2 X_t + u_{2t} \] (1.21)
we have
\[ H_0 : \beta_1 = \alpha_1, \quad \beta_2 = \alpha_2, \]
that is, the same parameters are appropriate for the entire data set, and
\[ H_1 : \text{at least one of } \beta_j \neq \alpha_j, \]
at least one parameter differs between the two sub-periods.

*Eviews* will carry out a Chow test for you, but you need to consider how the test is constructed. Let the total number of observations in the sample be \( N \), and consider two sub-groups of size \( n_1 \) and \( n_2 \) (with \( n_1 + n_2 = N \)). We are interested in comparing \( RSS_N \) with \(( RSS_{n_1} + RSS_{n_2} )\) and the relevant \( F \)-ratio has the difference between these two quantities \([ RSS_N - (RSS_{n_1} + RSS_{n_2})]\) in the numerator, and \(( RSS_{n_1} + RSS_{n_2} )\) in the denominator. To form the \( F \)-statistic we need the degrees of freedom associated with numerator and denominator. Recalling that the degrees of freedom associated with any regression is given by \( n - k \), where \( n \) is the sample size and \( k \) is the number of parameters being estimated, it is easy to show that the degrees of freedom in the numerator will be given by \( k \), and in the denominator by \( (n_1 + n_2 - 2k) \).

The test statistic is
\[
\frac{[RSS_N - (RSS_{n_1} + RSS_{n_2})]/k}{[RSS_{n_1} + RSS_{n_2}]/(n_1 + n_2 - 2k)} \sim F_{k, n_1 + n_2 - 2k} \text{ if } H_0 \text{ is true.} \quad (1.22)
\]

If the calculated value of the test statistic is greater than the critical value at a predetermined significance level, say 0.05, reject \( H_0 \); the same parameters are not appropriate for the entire sample period. So the two separate regressions of the two sub-groups give a better fit to the data than the single regression using the whole data set. In other words, we conclude that there has been a significant change in the parameters of the regression equation between the two periods. Note that while the use of the Chow test can suggest that there is parameter instability, it does not to tell us which parameters may have changed – it may be the intercept term, or one or more of the regression coefficients. For this reason, dummy variables provide a more direct way of examining the stability of the individual parameters.

### 1.4 Unit Study Guide

The main reading for this unit is Chapter 9 and Chapter 8, Section 8.7, of Gujarati and Porter. If you have understood the material covered so far in this unit, the material from Gujarati and Porter should not present any difficulty. As you read, do not be put off by the authors’ mention of ANOVA (analysis of variance) and ANCOVA (analysis of covariance) models. This course has not used that terminology, but the discussion of dummy variables in this unit has introduced you to all the relevant concepts.
Chapter 9 begins by introducing the concept, and use of, an intercept dummy variable. Most of the material covered will be familiar to you already; the only points made by Gujarati and Porter that were not discussed specifically in Section 1.2 above are to be found in section 9.1 of the textbook, where the authors discuss the possibility of a regression where the only explanatory variable is a dummy variable, and in section 9.2 where they explain the ‘dummy variable trap’ (that is, including the same number of dummy variables as there are categories–instead of one less) in terms of the multicollinearity problem that would arise if the number of dummies was exactly equal to the number of categories. Otherwise, there are no new ideas for you in these sections and, I hope, no problems.

Reading
Please read sections 9.1–9.4 of Gujarati and Porter, pp 277–85, now.

Gujarati and Porter discuss the Chow test in section 8.7 of Chapter 8. They describe this section as dealing with comparisons between two regressions. Strictly speaking, the comparisons are between a regression on ‘pooled’ data – that is, all the observations – and separate regressions on two sub-groups of the data. You should work carefully through their explanation of the Chow test and the example on pp 254–59 to satisfy yourself that the procedures described by Gujarati and Porter are the same as those presented in Section 1.3 of this unit. Note, too, the assumptions on which the Chow test is based, page 256.

Reading
Please read Chapter 8 section 8.7, pp 254–59, now.

In Chapter 9, Gujarati and Porter explain carefully the usefulness of dummy variables in tests of structural stability. They do not have a separate section discussing the use of a dummy variable which alters the slope coefficient, but introduce this concept in the context of a single regression equation on ‘pooled’ data – equation (9.5.1) – which has one dummy variable used to test for a different intercept and a different slope for the chosen sub-groups. This is a very useful general formulation, and allows direct comparison with the Chow test. Gujarati and Porter list the advantages of using dummy variables rather than the Chow test in example 9.4, pp 287–88.

Reading

The next four sections provide examples of the use of dummy variables. You should read these sections carefully. It is important to be aware of these further applications of dummy variable techniques.
Reading
Please read sections 9.6–9.9, pp 288–97.

How should we interpret dummy variables in regressions where the dependent variable is defined in logarithms? This is explained at the start of Section 9.10. Both the Chow test and the use of dummy variables with pooled data assume homoscedastic disturbances. It is also assumed that the disturbances are not autocorrelated. The implications of (i) heteroscedasticity and (ii) the elimination of first-order autocorrelation in the presence of autocorrelation are also discussed briefly in Gujarati and Porter.

Reading
Please read pp 297–98 on The Interpretation of Dummy Variables in Semilogarithmic Regression; pp 298–99 on Dummy Variables and Heteroscedasticity and Dummy Variables and Autocorrelation; and finally, please read the chapter summary and conclusions, pp 304–05.

1.5 Example – Long-Term Trends in Terms of Trade

An issue that has attracted a considerable amount of interest in the literature on commodity prices is the question of terms of trade between countries producing primary products and countries producing manufactured goods. Much of the interest in this question has arisen because primary producers have been loosely classified as ‘poor’ or less-developed countries and manufacturing producers as ‘rich’ countries. A widely held view (sometimes referred to as the Prebisch-Singer hypothesis) is that the terms of trade have moved and are likely to continue to move against primary producers. This view can be supported either from the point of view of dependency theorists who argue that countries of the ‘periphery’ are disadvantaged in their relations with countries of the ‘centre’; or from the point of view of neo-classical theorists who base their arguments on perceived differences in the elasticities of demand and supply for the two types of commodity.

How can the hypothesis that the terms of trade move against primary producers be tested? Before answering this question we note the important distinction to be made between a country’s net barter terms of trade, NBTT, and a country’s income terms of trade, YTT. The net barter terms of trade are measured by the ratio of the average price of its exports, $P_X$, to the average price of its imports, $P_M$.

$$NBTT = \frac{P_X}{P_M} \quad (1.23)$$

This expression provides a measure of the quantity of exports required to finance a given quantity of imports. (If a country exported only coffee and
imported only cars, the NBTT would tell us how many tons of coffee had to be exported to purchase a given quantity of cars.) If the ratio \( P_x/P_M \) rises there is said to have been a favourable movement in the (net barter) terms of trade; if the ratio falls, the terms of trade are said to have turned against the country in question. The country’s income terms of trade, \( YTT \), are measured by its income from exports (the average price of exports multiplied by the quantity of exports, \( Q_X \), relative to the average price of imports):

\[
YTT = \frac{P_x Q_X}{P_M}
\]

Equation (1.24) provides a measure of the purchasing power of exports, and it is obviously possible for the income terms of trade to improve even if there has been an adverse movement in the NBTT; this would happen if the increase in the quantity of goods exported more than compensated for the fall in unit price.

The Prebisch-Singer hypothesis is concerned only with changes in the net barter terms of trade between primary product prices and manufactured product prices. To test the hypothesis that the terms of trade have moved against primary producers, we need time series data on the NBTT between primary products and manufactures, and a regression model can then be used to relate changes in the NBTT to time. This is an example of an econometric problem where most of the difficulty lies in collecting satisfactory data; the regression analysis is very straightforward. Spraos (1980), whose regression results we will be discussing below, presents several index number series that between them cover the years 1871 to 1970. It is not possible here to discuss at length the data problems, but they include:

- shortage of data for the early years (UK balance of payments statistics were used by Prebisch in his original study)
- changes in the quality and composition of goods
- changes in transport costs
- differing data sources and definitions,

and in recent years:

- the presence or absence of petroleum exports in the statistics.

Finally, and very importantly, if evidence of decline in the terms of trade is to be taken as evidence of a movement in the terms of trade against developing countries, account has to be taken of the fact that rich countries are themselves large producers of primary products and developing countries increasingly export manufactured products. Leaving aside data problems, the hypothesis of a decline in the terms of trade can be tested using the regression model.

\[
\ln NBTT_t = \beta_1 + \beta_2 T_t + u_t
\]
where $T_t$ is time. In this semi-logarithmic regression model, $\beta_2$ represents the rate of change of $NBTT$ over time.

---

### Study Note: Interpretation of a Semi-logarithmic Regression

Underlying the model in (1.25) is the relation

$$NBTT_t = NBTT_0 e^{r T_t}$$

where $NBTT_0$ is the value of $NBTT$ in the initial year, $r$ is the rate of growth of $NBTT$ over time $T_t$. If you take natural logarithms of this expression you will derive

$$\ln NBTT_t = \ln NBTT_0 + r T_t$$

which can be compared with equation (1.25).

---

Spraos ran a number of regressions using different series and different sample periods. We will look at two regression results based on his UN data; the first covers the years 1900–38:

$$\ln NBTT_t = 4.572 - 0.00725 T_t + \hat{u}_t$$

$(0.00188)$

$R^2 = 0.33185$ (1.26)

and the second covers the years 1900–70:

$$\ln NBTT_t = 4.438 - 0.00134 T_t + \hat{u}_t$$

$(0.00096)$

$R^2 = 0.03687$ (1.27)

The figures under the estimated regression coefficients are the standard errors. Calculating the $t$ statistics for the test of the null hypothesis that the coefficient $\beta_2 = 0$, we find for (1.26) $t = 3.856$ and for (1.27) $t = 1.396$.

---

**Exercise**

Before you read on, can you give the economic interpretation of these regression results?

---

The conclusions drawn by Spraos are that in the period up to 1938 there was a clearly discernible deterioration in the terms of trade for primary producers (averaging 0.7% each year), but that when the sample period is extended to 1970, this conclusion can no longer be drawn because the coefficient $\beta_2$ is not significantly different from zero.

Let us now consider a criticism of Spraos’s work made by Sapsford (1985) who argued that there was a significant shift in the series in the post-war period (defined as 1950 and thereafter), so that the parameters in the regression (1.27) cannot be treated as constants over the whole period. Using the new techniques introduced in this unit, we can examine this claim. Let us first consider the evidence of the Chow test, using the data from 1900–38 and 1950–70 as our two sub-periods. (There is a gap in the series for the years 1939–49.) The following equations give the regression results and the residual sums of squares needed for the Chow test:
The test statistic is

\[ F_{\text{calc}} = \frac{[1.2457 - (0.51683 + 0.00874)]/2}{(0.51683 + 0.00874)/56} = 38.37 \]

From Table D.3 on p 884 of Gujarati and Porter, the critical value for \( F_{(2, 56)} \) at the 0.05 significance level is approximately 3.15. As 38.37 > 3.15 we conclude that the null hypothesis of parameter stability is rejected; there is evidence of parameter instability over this period.

Now let us use a dummy variable to test for a break in the series from 1950. We estimate a regression for the entire period, 1900-70, including an intercept dummy, \( D_t \), defined as follows

\[ D = 0 \text{ for the period before 1950} \]
and
\[ D = 1 \text{ for 1950 and thereafter.} \]

The estimated model is

\[ \ln NBTT_t = 4.594 - 0.00836 T_t + 0.5003 D_t + \hat{u}_t \]

\[ R^2 = 0.55957 \quad (1.31) \]

**Exercise**

What are the econometric and economic implications of (1.31)?

If we compare the regression (1.31), which includes the dummy variable with (1.27), which is based on the same period but does not include a dummy variable, we can see a number of important changes. \( R^2 \) has risen from 0.03687 to 0.56. The coefficient on the dummy variable has a \( t \) statistic of 7.709, so the dummy variable is certainly significant. The coefficient of time, which was insignificant in (1.27), is now negative and has a highly significant \( t \) statistic of –5.887. (The implication of the coefficient –0.00836 is that the rate of deterioration in the net barter terms of trade was approximately 0.8% per year.) The introduction of the dummy variable representing the break in the series after 1950 has made a big difference to the regression results and throws doubt on Spraos’s conclusion that there is no evidence for a deterioration in the terms of trade over the period 1900–70.

Note that this is not the last word on the subject. The debate continues! A later comprehensive survey in the *World Bank Economic Review* by Grilli and Yang (1988) used a newly constructed index of commodity prices and two indexes of manufactured goods prices to show that over the period 1900 to 1986 the relative prices of all primary commodities fell by 0.5 per cent a year and those of non-fuel primary commodities by 0.6 per cent a year, thus
confirming the downward trend (though not the figures) postulated by Prebisch.

However, they then make two very important comments on these results. The first is that the terms of trade of non-fuel primary commodities are not the same as the net barter terms of trade of non-oil-exporting developing countries (i.e. these countries do not just export primary products and import manufactures). Secondly, they emphasise that because of increased export quantities there has been a growth in the purchasing power of non-fuel commodity exports and they conclude (p.46) that in the post-World War II period the ‘available evidence on the income terms of trade of non-oil-exporting developing countries indicates that consistent and substantial gains were obtained by them’.

### 1.6 Summary

This unit has introduced the concept of the dummy variable. A dummy variable can be used in regression analysis to represent the effect of an explanatory variable that is qualitative or categorical in nature.

In cross-section data, variables such as sector, industry or regulatory regime can be represented by dummy variables; in time series data, dummies can be used to represent seasonal influences on the dependent variable, or a break in the series caused by a change in the economic and financial environment. Dummy variables can only take two values, 0 or 1, and they can be used to model changes in the intercept term and/or changes in the regression coefficients for different sub-groups within the sample. When including dummy variables in a regression, the number of dummy variables must always be one less than the number of categories taken by the variable in question; in this way the ‘dummy variable trap’ is avoided.

The use of the Chow test to test for parameter instability was also introduced in this unit.

Notice, however, that the use of a qualitative variable as a dependent variable in a regression has not been discussed.

### 1.7 Exercises

1a The UN data provided by Spraos and used for section 1.5 in this unit are in a text file called C232C332_U1_Q1.txt. This has 53 annual (although undated) observations on the following variables:

- $NBTT$ is the net barter terms of trade
- $T$ is time: $T = 1$ in 1900; $T = 71$ in 1970; gaps in numbering correspond to war years with no observations.
Opening the file

The tab-delimited text file C232C332_U1_Q1.txt contains three columns: Observation, NBTT, and T. For observation 1, NBTT = 80 and T = 1. For observation 53, NBTT = 85 and T = 71.

To open the file in *Eviews* go to File/Open/Foreign Data as Workfile … then browse to select the file. After you have found and opened the file, you will get the dialogue box ‘Text Read – Step 1 of 4’. This shows the preview window – how *Eviews* will interpret the data in the file. You can check that the first values are as noted above. Click Next.

Step 2 of 4 asks about the delimiter between entries; this is a single tab, as indicated, so click Next.

Step 3 of 4 identifies that the column headers (the names of the variables) are in line 1. Click Next.

Step 4 of 4 concerns the Import method (the default is to Create a new workfile) and the structure of the workfile. Although the data in this example are in a chronological sequence, the data themselves are undated, and the unstructured/undated workfile structure is appropriate, so click on Finish.

You should now see the Workfile window (C232C332_U1_Q1) with a list of variables. *Eviews* has allocated the data to the appropriate observations and also created the series observation.

Create a dummy variable $D$ where $D = 0$ for observations 1–32 (years 1900–1938) and $D = 1$ for observations 33–53 (years 1950–1970).

Creating dummy variables

You can create dummy variables in *Eviews* in a number of ways.

Using the drop-down menus, select Quick/Generate Series… and then enter the equation: ‘d1950=0’ and click on OK. This will generate the series d1950, which has a value 0 for every observation. Why not d for dummy? In *Eviews* d is an illegal or reserved name. To set the value 1 for observations 33 to 53, you need to change the sample. Select Quick/Sample and type in 33 53 to replace 1 53. Click on OK. To set the values, select Quick/Generate Series… and type d1950=1 for the equation, and click on OK. Finally, reset the sample by selecting Quick/Sample… then replace 33 53 with 1 53 (or ‘@all’ will also select all the observations) then click on OK.

Alternatively, the commands to create the dummy are as follows.

```
series d1950=0
smpl 33 53
d1950.fill(s,l) 1
smpl @all
```

The third line uses the fill command. In the brackets the l (not one) specifies a loop repeating the fill over the list of values as many times as it takes to fill the series, although in this case we have used only one fill value. (Instead of using the fill command in the above sequence, you can use the command ‘series 1950=1’.)
As an alternative to the above two methods, after you have created the series d1950 = 0, open the series d1950 in spreadsheet view by double-clicking on d1950, then click on the Edit +/- toggle button, and then type in the 1 values for observations 33 to 53 (this method is probably the quickest if the dummy variable you are creating takes the value 1 for only one observation or a few observations).

To check the series d1950, double click on it, or use the command 'show d1950'.

i) Regress lnNBTT on a constant, T and D. This should reproduce the result in (1.31).

**OLS estimation**

To estimate the regression equation using dropdown menus, select Quick/Estimate Equation … and then in the equation specification type the variables to be included in the regression, with the dependent variable first. What is the specification for Q1a part i)? The dependent variable is the natural logarithm of NBTT, there is a constant, and the independent variables are T and d1950. So the specification is log(nbtt) c t d1950. Then click on OK.

*Eviews* now shows the estimation results. To save this output in *Eviews*, click on the Name button, and provide a name for the output. Note that *Eviews* does not allow spaces in the names of objects. Try Inbtt_c_t_d1950 as a name for this equation. You can also copy the output to a word processor. Select the parts of the output you wish to copy, then rightclick and Copy. Keep the default at Formatted, and click on OK. Then you can paste into your word processor (it works in Microsoft Word).

To estimate the equation using a command: ‘equation lnbtt_c_t_d1950.ls log(nbtt) c t d1950’. This command estimates the equation with the specification log(nbtt) c t d1950 using least squares, and names the results Inbtt_c_t_d1950.

Note that it is not necessary to create a new series that is the natural logarithm of nbtt. We can specify an equation in which one of the variables is log(nbtt). If you prefer, you could create a new series that is the logarithm of nbtt, but note that you will need a different specification for the equation to estimate.

ii) Regress lnNBTT on a constant, T, D, and D × T. Interpret this model by writing out the sample regression lines implied when D = 0 and D = 1 and comparing them with (1.29) and (1.30).

What is the specification for this equation? It is not necessary to create a new series which is equal to d1950 × t. Instead, the transformation can be incorporated into the equation specification. Use the following specification: ‘log(nbtt) c t d1950 d1950*t’.

b) Use *Eviews* to plot NBTT and time, T. Comment on any features of the data (apart from the break after 1950) that seem to you to be significant. Select Quick/Graph… and in the series list, type nbtt t and click on OK. The line-symbol type of graph is required here, so click on OK.

Or, using a command: ‘line nbtt t’. 
You can alter how the graph appears using the Options button. Name the graph to save it in *Eviews*. Selecting the plot area selects all of the graph, then right-click and Copy, and you can copy the graph to Word. Or you can right-click the graph and Save graph to disk… which gives the option of eps, jpg, etc.

To save your workfile, select the workfile window so that it is highlighted, then click on File/Save As … and browse to the folder where you want to save the workfile. Keep the saving precision defaults if prompted.

2 A researcher is investigating the influence of the coverage ratio, credit rating and industrial sector on bond spreads.

The following model is proposed:

\[ Y = \beta_1 + \beta_2 X + \beta_3 J + \beta_4 M + u_i \]

where

- \( Y \) is the bond spread in basis points;
- \( X \) is the coverage ratio;
- \( J = 0 \) if the bond is rated as investment grade and \( J = 1 \) if the rating is CCC+ and below; and
- \( M = 0 \) if the company is not in the manufacturing sector and \( M = 1 \) if the company is in the manufacturing sector.

**a)** Write down the regression equations corresponding to each of the following:

i) The bond spread for an investment grade bond for a company in the non-manufacturing sector;  

ii) The bond spread for a non-investment grade bond for a company in the non-manufacturing sector;  

iii) The bond spread for an investment grade bond for a company in the manufacturing sector;  

iv) The bond spread for a non-investment grade bond for a company in the manufacturing sector.

**b)** How would you interpret the estimated regression equation if \( \beta_3 \) is significantly different from zero and \( \beta_4 \) is not significantly different from zero?

3 This question examines whether the parameters of the single-index model for Goldman Sachs are the same in the periods before and after the financial crisis of 2008, and in particular the bankruptcy of Lehman Brothers in the week ending 15 September 2008.

The text file C232C332_U1_Q3.txt contains weekly data on the weekly proportionate return on the shares of Goldman Sachs Group Inc \( (Y) \), and weekly return on the New York Stock Exchange US 100 index \( (X) \), for the period 4 April 2005 to 28 March 2011. The return on Goldman Sachs shares are calculated as

\[ Y_i = \frac{P_t - P_{t-1}}{P_{t-1}} \]
where $P_t$ is the share price at time $t$. The return on the NYSE index is calculated in the same way. For reference (to four decimal places), on 4 April 2005, $X = 0.0083$ and $Y = 0.0332$; on 11 April 2005 $X = -0.0236$ and $Y = -0.0570$; and on 28 March 2011 $X = 0.0147$ and $Y = 0.0143$.

a) Use the Chow test at the 0.05 significance level to test the hypothesis that the parameters of the single-index model for Goldman Sachs

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

are stable between the two sub-periods 4 April 2005-8 September 2008 and 15 September 2008-28 March 2011.

Be careful how you set the dates of the sub-periods in Eviews. Two options are available for the treatment of the Month/Day order in dates: Month/Day/Year and Day/Month/Year. To set these options go to Options/General Options … and in the tree on the left-hand side, select Date representation.

If Month/Day/Year is being used, then 8 September 2008 is represented as 9/8/2008, and if Day/Month/Year is being used, 8 September 2008 is 8/9/2008.

b) Does the Chow test provide any information about the separate changes in the intercept coefficient and slope coefficient?

c) Use a dummy variable to test the hypothesis that both the slope and the intercept have changed between the two sub-periods.

d) What procedure would you use if you were only interested in testing the hypothesis that the beta coefficient for Goldman Sachs was changed between the two sub-periods?

4 The following model represents the log of earnings for investment managers, $Y$, as a function of experience (in years), $X$, and a dummy variable, $G$, which takes the value of 0 if the investment manager is male, and 1 if the investment manager is female.

$$\ln Y_t = \beta_1 + \beta_2 X_t + \beta_3 G_t + u_t$$

For each of the following statements, say whether it is true or false and give a brief explanation of your answer.

a) The coefficient $\beta_2$ measures the proportional change in $Y_t$ for a given absolute change in $X_t$.

b) The coefficient $\beta_3$ measures the proportional change in $Y$ when the head of the investment fund is female.

c) Because $G_t$ is a qualitative variable, not a quantitative variable, its inclusion in the model will not affect the value of $R^2$.

d) If regression equations are estimated using the two sub-samples corresponding to male investment managers, and female investment managers, the regression coefficients will be the same as those obtained from the above model with $G = 0$ and $G = 1$.

e) If observations for each investment manager were available over a ten-year period, a regression for the pooled data set could be modelled by including nine year dummies in the regression.
1.8 Answers to Exercises

1 a) i) The *Eviews* output is

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>4.594213</td>
<td>0.034621</td>
<td>132.6993</td>
<td>0.0000</td>
</tr>
<tr>
<td>T</td>
<td>-0.008364</td>
<td>0.001420</td>
<td>-5.889133</td>
<td>0.0000</td>
</tr>
<tr>
<td>D1950</td>
<td>0.500266</td>
<td>0.064943</td>
<td>7.703150</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.559565

If you are in any doubt about the interpretation of this equation, look back at the comments on equation (1.31) which you have just reproduced.

ii) The *Eviews* output is

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>4.571526</td>
<td>0.035407</td>
<td>129.1143</td>
<td>0.0000</td>
</tr>
<tr>
<td>T</td>
<td>-0.007254</td>
<td>0.001483</td>
<td>-4.892043</td>
<td>0.0000</td>
</tr>
<tr>
<td>D1950</td>
<td>0.951932</td>
<td>0.231510</td>
<td>4.111834</td>
<td>0.0001</td>
</tr>
<tr>
<td>D1950*T</td>
<td>-0.008143</td>
<td>0.004016</td>
<td>-2.027499</td>
<td>0.0481</td>
</tr>
</tbody>
</table>

R-squared 0.593655

If you are in any doubt about the interpretation of this equation, look back at the comments on equation (1.31) which you have just reproduced.

\[ \ln NBTT_i = 4.594 - 0.00836T_i + 0.5003D_{1950} + \hat{\epsilon}_i \]

\[
\begin{pmatrix}
0.001420 \\
0.064943
\end{pmatrix}
\]

\[ R^2 = 0.55957 \]

\[ (0.001420) \quad (0.064943) \]
ln NBTT$_i = 4.572 - 0.00725T_i + 0.952D_i - 0.00814 \left( D_i \times T_i \right) + \hat{u}_i$

For the period 1900-1938 we have $D = 0$ and
\[
\ln NBTT_i = 4.572 - 0.00725T_i + \hat{u}_i
\]
For the period 1950-1970 we have $D = 1$
\[
\ln NBTT_i = 4.572 - 0.00725T_i + 0.952(1) - 0.00814(1)T_i + \hat{u}_i
\]
\[
\therefore \ln NBTT_i = 5.524 - 0.0154T_i + \hat{u}_i
\]
You will see from the comparison with (1.29) and (1.30) that by introducing the two dummy variables, you have been able to estimate the intercepts and the slopes of the two regressions covering 1900-38 and 1950-70, respectively, with a single OLS regression.

b) See Figure 1.3.

**Figure 1.3 Plot of NBTT and T 1900–1970**

A really good answer to this question might have started: ‘Surely we should have graphed the data before embarking on any numerical analysis!’ If you had started by looking at the graph, would you ever have thought it sensible to represent the relationship between the net barter terms of trade and time by a simple linear regression? It is obvious that the period divides into three quite distinct sub-periods. Before the First World War the terms of trade for primary producers improved steadily. After the Second World War they deteriorated. Between the wars they oscillated with no obvious trend. In our analysis we picked up the second break in the series by use of the dummy variable $D$ and $D \times T$. However, the graph suggests that the earlier break in the series at the time of the First World War should also be taken into account. On the basis of the graph, it would be hard to support the view that there was any discernible long-term trend in the net barter terms of trade between 1900 and 1970.

2 a) i) $J = 0, \ M = 0$
\[
Y_i = \beta_1 + \beta_2 X_i + u_i
\]
ii) $J = 1, \ M = 0$
$Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + u_i$

iii) $J = 0, \ M = 1$

$Y_i = (\beta_1 + \beta_4) + \beta_2 X_i + u_i$

iv) $J = 1, \ M = 1$

$Y_i = (\beta_1 + \beta_3 + \beta_4) + \beta_2 X_i + u_i$

b) If $\beta_3$ is statistically significant, but $\beta_4$ is not statistically significant, this suggests that the bond spread is affected by the credit rating (investment or non-investment grade), but not by the sector in which the company operates.

3a) For the Chow test we need to estimate the restricted model, which is the same model over the complete data period, and an unrestricted model, which is the SIM estimated over each of the two sub-periods. The estimation results from *Eviews* are as follows (using the month/day/year date representation).

### Complete period

Dependent Variable: Y  
Method: Least Squares  
Included observations: 313

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.002692</td>
<td>0.002583</td>
<td>1.042166</td>
<td>0.2981</td>
</tr>
<tr>
<td>X</td>
<td>1.698934</td>
<td>0.095451</td>
<td>17.79904</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.504624</td>
<td></td>
<td></td>
<td>0.003414</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.503031</td>
<td></td>
<td></td>
<td>0.064810</td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.045689</td>
<td></td>
<td></td>
<td>-3.327558</td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>0.649202</td>
<td></td>
<td></td>
<td>3.303621</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>522.7629</td>
<td></td>
<td></td>
<td>3.317992</td>
</tr>
<tr>
<td>F-statistic</td>
<td>316.8057</td>
<td></td>
<td></td>
<td>2.064138</td>
</tr>
<tr>
<td>Prob(F-statistic)</td>
<td>0.000000</td>
<td></td>
<td></td>
<td>0.000000</td>
</tr>
</tbody>
</table>

### Sub-period 1

Dependent Variable: Y  
Method: Least Squares  
Included observations: 180

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.002485</td>
<td>0.002276</td>
<td>1.091546</td>
<td>0.2765</td>
</tr>
<tr>
<td>X</td>
<td>1.660661</td>
<td>0.129396</td>
<td>12.83390</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.480609</td>
<td></td>
<td></td>
<td>0.002949</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.477691</td>
<td></td>
<td></td>
<td>0.042250</td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.045689</td>
<td></td>
<td></td>
<td>-4.128873</td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>0.649202</td>
<td></td>
<td></td>
<td>4.093996</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>522.7629</td>
<td></td>
<td></td>
<td>4.114488</td>
</tr>
<tr>
<td>F-statistic</td>
<td>316.8057</td>
<td></td>
<td></td>
<td>1.996187</td>
</tr>
<tr>
<td>Prob(F-statistic)</td>
<td>0.000000</td>
<td></td>
<td></td>
<td>0.000000</td>
</tr>
</tbody>
</table>
Sub-period 2

Dependent Variable: Y
Method: Least Squares
Included observations: 133

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.002979</td>
<td>0.005267</td>
<td>0.565633</td>
<td>0.5726</td>
</tr>
<tr>
<td>X</td>
<td>1.711145</td>
<td>0.145826</td>
<td>11.73417</td>
<td>0.000</td>
</tr>
</tbody>
</table>

R-squared 0.512451
Adjusted R-squared 0.508729
S.E. of regression 0.060728

The Chow test uses the $F$-statistic

$$F = \frac{[RSS_n - (RSS_n + RSS_{n_2})]/k}{[RSS_n + RSS_{n_2}]/(n_1 + n_2 - 2k)}$$

Substituting in this formula, we have

$$F_{calc} = \frac{[0.649202 - (0.165959 + 0.483115)]/2}{0.165959 + 0.483115}/(180 + 133 - 4) = 0.030468$$

The null hypothesis is $H_0$: the parameters are stable. The alternative hypothesis is $H_1$: at least one parameter is different. From Table D.3 on page 884 of the textbook, the 0.05 critical value of $F_{(2,309)} = 3.00$. The calculated value of the test statistic is less than the critical value; we therefore do not reject the null hypothesis and conclude that there is no significant change in the parameters between the two periods.

b) No. The Chow test shows only whether the parameters are significantly different between sub-samples; it does not reveal which parameters (if any) have changed.

c) To test hypotheses about both the slope and the intercept, the single-index model for the entire period could be represented by

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 D_t + \beta_4 (D_t X_t) + u_t,$$

where $D_t = 0$ for 4 April 2005 to 8 September 2008 and $D_t = 1$ for 15 September 2008 to 28 March 2011. In the following Eviews output, the dummy variable $D_t$ is named D2008.

Dependent Variable: Y
Method: Least Squares
Included observations: 313

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.002485</td>
<td>0.003417</td>
<td>0.727218</td>
<td>0.4676</td>
</tr>
<tr>
<td>X</td>
<td>1.660661</td>
<td>0.194223</td>
<td>8.550291</td>
<td>0.0000</td>
</tr>
<tr>
<td>D2008</td>
<td>0.000494</td>
<td>0.005241</td>
<td>0.094324</td>
<td>0.9249</td>
</tr>
<tr>
<td>D2008*X</td>
<td>0.050484</td>
<td>0.223237</td>
<td>0.226145</td>
<td>0.8212</td>
</tr>
</tbody>
</table>
An $F$-test of the joint significance of the two dummy variable terms suggests we would not reject the null-hypothesis that these two variables are not significant. How do we know this? Note that the unrestricted regression with the two-dummy variable terms is equivalent to the unrestricted regression over the two separate sub-periods in part a). In the regression with two dummy variable terms, the residual sum of squares is 0.649074 which equals 0.165959 + 0.483115 from part a. And the relevant restricted regression involving no dummy variables is identical to the restricted regression in part a).

However, we have the additional information given by the estimated coefficients for the dummy variable affecting the constant and slope.

d) Including only the slope dummy $D_tX_t$ in the model and excluding the intercept dummy $D_t$ in the model given in the answer to (c) above, would mean that the dummy variable would influence only the parameter $\beta_2$ which is the coefficient that measures the sensitivity of the return on Goldman Sachs shares to the return on the stock market index.

4 a) True. The model is in semi-log form, and in particular, the log-lin form (the dependent variable is measured in logs, but the regressors are linear). For the continuous regressor $X_i$, $\beta_2$ measures the constant proportional change in $Y_i$ for a given absolute change in $X_i$. To see this note that (ignoring the error term)

$$Y_i = e^{\beta_1 + \beta_2 X_i + \beta_3 G_i}$$

so

$$\frac{\partial Y_i}{\partial X_i} = e^{\beta_1 + \beta_2 X_i + \beta_3 G_i} \times \beta_2 = Y_i \times \beta_2 \times w$$

so

$$\frac{\partial Y_i}{\partial X_i} / Y_i = \beta_2$$

and the percentage change in $Y_i$ (or semi-elasticity with respect to $X_i$) is $100 \times \beta_2$.

(See Gujarati and Porter pp. 162-163 if you require more explanation of this point.)

b) False. The coefficient on the dummy variable in this semi-logarithmic model does not measure the proportionate change in $Y_i$ for an absolute change in the dummy variable. The dummy variable is not continuous: it takes only values 0 and 1. Consider the values for $Y_i$ when the dummy variable takes the values of 0 and 1 (again, ignoring the error term)

When $G_i = 0$, $Y_i = e^{\beta_1 + \beta_2 X_i}$
and when \( G_i = 1 \), \( Y_i = e^{\beta_1 x_i + \beta_3} 

So the proportionate change in \( Y_i \) is 

\[
\frac{e^{\beta_1 x_i + \beta_3} - e^{\beta_1 x_i}}{e^{\beta_1 x_i}} = e^{\beta_3 x_i} - 1 = e^{\beta_3} - 1
\]

and the percentage change in \( Y_i \) is \( 100 \times (e^{\beta_3} - 1) \).

(See Gujarati and Porter p. 314 Appendix 9A for more explanation of this point.)

c) False. If the gender of the investment manager is a relevant explanatory variable in determining the earnings of investment managers, its inclusion in the model will increase the value of \( R^2 \) (which measures the ratio of the explained variation to the total variation in the dependent variable).

d) False. The two separate regressions can be expected to show variation in both the intercept and slope coefficients. The model in this question only allows for variation in the intercept term.

e) True. One year would act as base year and nine year dummies would represent the remaining years. This is one way of combining cross-section and time-series data in a single model.

References


