# Contents

## Preface

0.1 Introduction ................................................................................. 1  
0.2 What is econometrics, and why study it? ................................. 1  
0.3 Aims ......................................................................................... 1  
0.4 Learning outcomes ..................................................................... 2  
0.5 How to make use of the textbook ............................................ 3  
0.6 How to make use of this subject guide ...................................... 3  
0.7 How to make use of the website ............................................... 4  
0.7.1 Slideshows ............................................................................. 4  
0.7.2 Data sets ............................................................................... 4  
0.8 Online study resources ........................................................... 5  
0.8.1 The VLE ............................................................................... 5  
0.8.2 Making use of the Online Library ........................................ 6  
0.9 Prerequisite for studying this subject ....................................... 6  
0.10 Application of linear algebra to econometrics ....................... 7  
0.11 The examination ...................................................................... 7  
0.12 Overview .................................................................................. 9  
0.13 Learning outcomes ............................................................... 10  
0.14 Additional exercises ............................................................. 10  
0.15 Answers to the starred exercises in the textbook .................... 11  
0.16 Answers to the additional exercises ....................................... 22  

## 1 Simple regression analysis

1.1 Overview .................................................................................. 27  
1.2 Learning outcomes ................................................................... 27  
1.3 Additional exercises ............................................................... 28  
1.4 Answers to the starred exercises in the textbook .................... 30  
1.5 Answers to the additional exercises ....................................... 35  

## 2 Properties of the regression coefficients and hypothesis testing

2.1 Overview .................................................................................. 41  

© Christopher Dougherty, 2016. All rights reserved.  
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
## Contents

2.2 Learning outcomes ........................................... 41
2.3 Further material ............................................. 42
2.4 Additional exercises ........................................ 43
2.5 Answers to the starred exercises in the textbook .......... 48
2.6 Answers to the additional exercises ....................... 53

3 Multiple regression analysis ................................. 59
  3.1 Overview .................................................. 59
  3.2 Learning outcomes ....................................... 59
  3.3 Additional exercises ..................................... 60
  3.4 Answers to the starred exercises in the textbook ........ 63
  3.5 Answers to the additional exercises ..................... 64

4 Transformations of variables ............................... 69
  4.1 Overview .................................................. 69
  4.2 Learning outcomes ....................................... 69
  4.3 Further material .......................................... 70
  4.4 Additional exercises ..................................... 72
  4.5 Answers to the starred exercises in the textbook ........ 74
  4.6 Answers to the additional exercises ..................... 77

5 Dummy variables ............................................... 85
  5.1 Overview .................................................. 85
  5.2 Learning outcomes ....................................... 85
  5.3 Additional exercises ..................................... 85
  5.4 Answers to the starred exercises in the textbook ........ 94
  5.5 Answers to the additional exercises ..................... 100

6 Specification of regression variables ..................... 115
  6.1 Overview .................................................. 115
  6.2 Learning outcomes ....................................... 115
  6.3 Additional exercises ..................................... 116
  6.4 Answers to the starred exercises in the textbook ........ 123
  6.5 Answers to the additional exercises ..................... 129

7 Heteroskedasticity .............................................. 145
  7.1 Overview .................................................. 145
  7.2 Learning outcomes ....................................... 145
## 7.3 Additional exercises ........................................ 145
## 7.4 Answers to the starred exercises in the textbook ............ 152
## 7.5 Answers to the additional exercises .......................... 159

## 8 Stochastic regressors and measurement errors ........................................ 169
## 8.1 Overview ........................................ 169
## 8.2 Learning outcomes ........................................ 169
## 8.3 Additional exercises ........................................ 170
## 8.4 Answers to the starred exercises in the textbook ............ 172
## 8.5 Answers to the additional exercises .......................... 180

## 9 Simultaneous equations estimation ........................................ 185
## 9.1 Overview ........................................ 185
## 9.2 Learning outcomes ........................................ 185
## 9.3 Further material ........................................ 186
## 9.4 Additional exercises ........................................ 187
## 9.5 Answers to the starred exercises in the textbook ............ 194
## 9.6 Answers to the additional exercises .......................... 199

## 10 Binary choice and limited dependent variable models, and maximum likelihood estimation ........................................ 213
## 10.1 Overview ........................................ 213
## 10.2 Learning outcomes ........................................ 213
## 10.3 Further material ........................................ 214
## 10.4 Additional exercises ........................................ 219
## 10.5 Answers to the starred exercises in the textbook ............ 225
## 10.6 Answers to the additional exercises .......................... 231

## 11 Models using time series data ........................................ 239
## 11.1 Overview ........................................ 239
## 11.2 Learning outcomes ........................................ 239
## 11.3 Additional exercises ........................................ 240
## 11.4 Answers to the starred exercises in the textbook ............ 245
## 11.5 Answers to the additional exercises .......................... 250

## 12 Properties of regression models with time series data ........................................ 261
## 12.1 Overview ........................................ 261
## 12.2 Learning outcomes ........................................ 261
Contents

12.3 Additional exercises ........................................... 262
12.4 Answers to the starred exercises in the textbook .............. 269
12.5 Answers to the additional exercises .............................. 273

13 Introduction to nonstationary time series .......................... 285
13.1 Overview ....................................................... 285
13.2 Learning outcomes .............................................. 285
13.3 Further material ................................................. 286
13.4 Additional exercises ............................................. 287
13.5 Answers to the starred exercises in the textbook ............. 291
13.6 Answers to the additional exercises .............................. 295

14 Introduction to panel data ........................................... 299
14.1 Overview ....................................................... 299
14.2 Learning outcomes .............................................. 299
14.3 Additional exercises ............................................. 300
14.4 Answer to the starred exercise in the textbook ............... 304
14.5 Answers to the additional exercises .............................. 306

15 Regression analysis with linear algebra primer ................... 313
15.1 Overview ....................................................... 313
15.2 Notation ......................................................... 314
15.3 Test exercises .................................................... 314
15.4 The multiple regression model .................................. 314
15.5 The intercept in a regression model ............................... 315
15.6 The OLS regression coefficients ................................ 316
15.7 Unbiasedness of the OLS regression coefficients ............ 317
15.8 The variance-covariance matrix of the OLS regression coefficients ........................... 317
15.9 The Gauss–Markov theorem ..................................... 319
15.10 Consistency of the OLS regression coefficients .............. 319
15.11 Frisch–Waugh–Lovell theorem .................................. 320
15.12 Exact multicollinearity .......................................... 323
15.13 Estimation of a linear combination of regression coefficients .............................................. 324
15.14 Testing linear restrictions ....................................... 325
15.15 Weighted least squares and heteroskedasticity ............ 325
15.16 IV estimators and TSLS ........................................ 327
15.17 Generalised least squares ....................................... 329
Contents
Preface

0.1 Introduction

0.2 What is econometrics, and why study it?

Econometrics is the application of statistical methods to the quantification and critical assessment of hypothetical economic relationships using data. It is with the aid of econometrics that we discriminate between competing economic theories and put numerical clothing onto the successful ones. Econometric analysis may be motivated by a simple desire to improve our understanding of how the economy works, at either the microeconomic or macroeconomic level, but more often it is undertaken with a specific objective in mind. In the private sector, the financial benefits that accrue from a sophisticated understanding of relevant markets and an ability to predict change may be the driving factor. In the public sector, the impetus may come from an awareness that evidence-based policy initiatives are likely to have the greatest impact.

It is now generally recognised that nearly all professional economists, not just those actually working with data, should have a basic understanding of econometrics. There are two major benefits. One is that it facilitates communication between econometricians and the users of their work. The other is the development of the ability to obtain a perspective on econometric work and to make a critical evaluation of it. Econometric work is more robust in some contexts than in others. Experience with the practice of econometrics and a knowledge of the potential problems that can arise are essential for developing an instinct for judging how much confidence should be placed on the findings of a particular study.

Such is the importance of econometrics that, in common with intermediate macroeconomics and microeconomics, an introductory course forms part of the core of any serious undergraduate degree in economics and is a prerequisite for admission to a serious Master’s level course in economics or finance.

0.3 Aims

The aim of EC2020 Elements of econometrics is to give you an opportunity to develop an understanding of econometrics to a standard that will equip you to understand and evaluate most applied analysis of cross-sectional data and to be able to undertake such analysis yourself. The restriction to cross-sectional data (data raised at one moment in time, often through a survey of households, individuals, or enterprises) should be emphasised because the analysis of time series data (observations on a set of variables over a period of time) is much more complex. Chapters 11 to 13 of the textbook, Introduction to econometrics, and this subject guide are devoted to the
analysis of time series data, but, beyond very simple applications, the objectives are confined to giving you an understanding of the problems involved and making you aware of the need for a Master’s level course if you intend to work with such data. Specifically the aims of the course are to:

- develop an understanding of the use of regression analysis and related techniques for quantifying economic relationships and testing economic theories
- equip you to read and evaluate empirical papers in professional journals
- provide you with practical experience of using mainstream regression programmes to fit economic models.

### 0.4 Learning outcomes

By the end of this course, and having completed the Essential reading and activities, you should be able to:

- describe and apply the classical regression model and its application to cross-sectional data
- describe and apply the:
  - Gauss–Markov conditions and other assumptions required in the application of the classical regression model
  - reasons for expecting violations of these assumptions in certain circumstances
  - tests for violations
  - potential remedial measures, including, where appropriate, the use of instrumental variables
- recognise and apply the advantages of logit, probit and similar models over regression analysis when fitting binary choice models
- competently use regression, logit and probit analysis to quantify economic relationships using standard regression programmes (Stata and EViews) in simple applications
- describe and explain the principles underlying the use of maximum likelihood estimation
- apply regression analysis to fit time-series models using stationary time series, with awareness of some of the econometric problems specific to time-series applications (for example, autocorrelation) and remedial measures
- recognise the difficulties that arise in the application of regression analysis to nonstationary time series, know how to test for unit roots, and know what is meant by cointegration.
0.5  How to make use of the textbook

The only reading required for this course is my textbook:


The syllabus is the same as that for EC220 Introduction to econometrics, the corresponding internal course at the London School of Economics. The textbook has been written to cover it with very little added and nothing subtracted.

When writing a textbook, there is a temptation to include a large amount of non-core material that may potentially be of use or interest to students. There is much to be said for this, since it allows the same textbook to be used to some extent for reference as well as a vehicle for a taught course. However, my textbook is stripped down to nearly the bare minimum for two reasons. First, the core material provides quite enough content for an introductory year-long course and I think that students should initially concentrate on gaining a good understanding of it. Second, if the textbook is focused narrowly on the syllabus, students can read through it as a continuous narrative without a need for directional guidance. Obviously, this is particularly important for those who are studying the subject on their own, as is the case for most of those enrolled on EC2020 Elements of econometrics.

An examination syllabus is provided as an appendix to this subject guide, but its function is mostly to indicate the expected depth of understanding of each topic, rather than the selection of the topics themselves.

0.6  How to make use of this subject guide

The function of this subject guide differs from that of other subject guides you may be using. Unlike those for other courses, this subject guide acts as a supplementary resource, with the textbook as the main resource. Each chapter forms an extension to a corresponding chapter in the textbook with the same title. You must have a copy of the textbook to be able to study this course. The textbook will give you the information you need to carry out the activities and achieve the learning outcomes in the subject guide.

The main purpose of the subject guide is to provide you with opportunities to gain experience with econometrics through practice with exercises. Each chapter of the subject guide falls into two parts. The first part begins with an overview of the corresponding chapter in the textbook. Then there is a checklist of learning outcomes anticipated as a result of studying the chapter in the textbook, doing the exercises in the subject guide, and making use of the corresponding resources on the website. Finally, in some of the chapters, comes a section headed ‘Further material’. This consists of new topics that may be included in the next edition of the textbook. The second part of each chapter consists of additional exercises, followed by answers to the starred exercises in the text and answers to the additional exercises.

You should organise your studies in the following way:

- first read this introductory chapter
Preface

- read the Overview section from the Review chapter of the subject guide
- read the Review chapter of the textbook and do the starred exercises
- refer to the subject guide for answers to the starred exercises in the text and for additional exercises
- check that you have covered all the items in the learning outcomes section in the subject guide.

You should repeat this process for each of the numbered chapters. Note that the subject guide chapters have the same titles as the chapters in the text. In those chapters where there is a ‘Further material’ section in the subject guide, this should be read after reading the chapter in the textbook.

0.7 How to make use of the website

You should make full use of the resources available at the Online Resource Centre maintained by the publisher, Oxford University Press (OUP): www.oup.com/uk/orc/bin/9780199567089. Here you will find PowerPoint slideshows that provide a graphical treatment of the topics covered in the textbook, data sets for practical work and statistical tables.

0.7.1 Slideshows

In principle you will be able to acquire mastery of the subject by studying the contents of the textbook with the support of this subject guide and doing the exercises conscientiously. However, I strongly recommend that you do study all the slideshows as well. Some do not add much to the material in the textbook, and these you can skim through quickly. Some, however, provide a much more graphical treatment than is possible with print and they should improve your understanding. Some present and discuss regression results and other hands-on material that could not be included in the text for lack of space, and they likewise should be helpful.

0.7.2 Data sets

To use the data sets, you must have access to a proper statistics application with facilities for regression analysis, such as Stata or EViews. The student versions of such applications are adequate for doing all, or almost all, the exercises and of course are much cheaper than the professional ones. Product and pricing information can be obtained from the applications’ websites, the URL usually being the name of the application sandwiched between ‘www.’ and ‘.com’.

If you do not have access to a commercial econometrics application, you should use gretl. This is a sophisticated application almost as powerful as the commercial ones, and it is free. See the gretl manual on the OUP website for further information.

Whatever you do, do not be tempted to try to get by with the regression engines built into some spreadsheet applications, such as Microsoft Excel. They are not remotely
adequate for your needs.

There are three major data sets on the website. The most important one, for the purposes of this subject guide, is the Consumer Expenditure Survey (CES) data set. You will find on the website versions in the formats used by Stata, EViews and gretl. If you are using some other application, you should download the text version (comma-delimited ASCII) and import it. Answers to all of the exercises are provided in the relevant chapters of this subject guide.

The exercises for the CES data set cover Chapters 1–10 of the text. For Chapters 11–13, you should use the Demand Functions data set, another major data set, to do the additional exercises in the corresponding chapters of this subject guide. Again you should download the data set in the appropriate format. For these exercises, also, answers are provided.

The third major data set on the website is the Educational Attainment and Earnings Function data set, which provides practical work for the first 10 chapters of the text and Chapter 14. No answers are provided, but many parallel examples will be found in the text.

0.8 Online study resources

In addition to the subject guide and the Essential reading, it is crucial that you take advantage of the study resources that are available online for this course, including the virtual learning environment (VLE) and the Online Library.

You can access the VLE, the Online Library and your University of London email account via the Student Portal at:

http://my.londoninternational.ac.uk

You should have received your login details for the Student Portal with your official offer, which was emailed to the address that you gave on your application form. You have probably already logged in to the Student Portal in order to register! As soon as you registered, you will automatically have been granted access to the VLE, Online Library and your fully functional University of London email account.

If you have forgotten these login details, please click on the ‘Forgotten your password’ link on the login page.

0.8.1 The VLE

The VLE, which complements this subject guide, has been designed to enhance your learning experience, providing additional support and a sense of community. It forms an important part of your study experience with the University of London and you should access it regularly.

The VLE provides a range of resources for EMFSS courses:

- Self-testing activities: Doing these allows you to test your own understanding of subject material.
Electronic study materials: The printed materials that you receive from the University of London are available to download, including updated reading lists and references.

Past examination papers and Examiners’ commentaries: These provide advice on how each examination question might best be answered.

A student discussion forum: This is an open space for you to discuss interests and experiences, seek support from your peers, work collaboratively to solve problems and discuss subject material.

Videos: There are recorded academic introductions to the subject, interviews and debates and, for some courses, audio-visual tutorials and conclusions.

Recorded lectures: For some courses, where appropriate, the sessions from previous years’ Study Weekends have been recorded and made available.

Study skills: Expert advice on preparing for examinations and developing your digital literacy skills.

Feedback forms.

Some of these resources are available for certain courses only, but we are expanding our provision all the time and you should check the VLE regularly for updates.

0.8.2 Making use of the Online Library

The Online Library contains a huge array of journal articles and other resources to help you read widely and extensively.

To access the majority of resources via the Online Library you will either need to use your University of London Student Portal login details, or you will be required to register and use an Athens login:

http://tinyurl.com/ollathens

The easiest way to locate relevant content and journal articles in the Online Library is to use the Summon search engine.

If you are having trouble finding an article listed in a reading list, try removing any punctuation from the title, such as single quotation marks, question marks and colons.

For further advice, please see the online help pages:

www.external.shl.lon.ac.uk/summon/about.php

0.9 Prerequisite for studying this subject

The prerequisite for studying this subject is a solid background in mathematics and elementary statistical theory. The mathematics requirement is a basic understanding of multivariate differential calculus. With regard to statistics, you must have a clear understanding of what is meant by the sampling distribution of an estimator, and of the principles of statistical inference and hypothesis testing. This is absolutely essential.
find that most problems that students have with introductory econometrics are not econometric problems at all but problems with statistics, or rather, a lack of understanding of statistics. There are no short cuts. If you do not have this background knowledge, you should put your study of econometrics on hold and study statistics first. Otherwise there will be core parts of the econometrics syllabus that you do not begin to understand.

In addition, it would be helpful if you have some knowledge of economics. However, although the examples and exercises relate to economics, most of them are so straightforward that a previous study of economics is not a requirement.

0.10 Application of linear algebra to econometrics

At the end of this subject guide you will find a primer on the application of linear algebra (matrix algebra) to econometrics. It is not part of the syllabus for the examination, and studying it is unlikely to confer any advantage for the examination. It is provided for the benefit of those students who intend to take a further course in econometrics, especially at the Master’s level. The present course is ambitious, by undergraduate standards, in terms of its coverage of concepts and, above all, its focus on the development of an intuitive understanding. For its purposes, it has been quite sufficient and appropriate to work with uncomplicated regression models, typically with no more than two explanatory variables.

However, when you progress to the next level, it is necessary to generalise the theory to cover multiple regression models with many explanatory variables, and linear algebra is ideal for this purpose. The primer does not attempt to teach it. There are many excellent texts and there is no point in duplicating them. The primer assumes that such basic study has already been undertaken, probably taking about 20 to 50 hours, depending on the individual. It is intended to show how the econometric theory in the text can be handled with this more advanced mathematical approach, thus serving as preparation for the higher-level course.

0.11 The examination

Important: the information and advice given here are based on the examination structure used at the time this subject guide was written. Please note that subject guides may be used for several years. Because of this we strongly advise you to always check both the current Regulations for relevant information about the examination, and the VLE where you should be advised of any forthcoming changes. You should also carefully check the rubric/instructions on the paper you actually sit and follow those instructions.

Candidates should answer eight out of ten questions in three hours: all of the questions in Section A (8 marks each) and three questions from Section B (20 marks each).

A calculator may be used when answering questions on this paper and it must comply in all respects with the specification given with your Admission Notice.

Remember, it is important to check the VLE for:
Preface

- up-to-date information on examination and assessment arrangements for this course
- where available, past examination papers and *Examiners’ commentaries* for the course which give advice on how each question might best be answered.
0.12 Overview

The textbook and this subject guide assume that you have previously studied basic statistical theory and have a sound understanding of the following topics:

- descriptive statistics (mean, median, quartile, variance, etc.)
- random variables and probability
- expectations and expected value rules
- population variance, covariance, and correlation
- sampling theory and estimation
- unbiasedness and efficiency
- loss functions and mean square error
- normal distribution
- hypothesis testing, including:
  - \( t \) tests
  - Type I and Type II error
  - the significance level and power of a \( t \) test
  - one-sided versus two-sided \( t \) tests
- confidence intervals
- convergence in probability, consistency, and plim rules
- convergence in distribution and central limit theorems.

There are many excellent textbooks that offer a first course in statistics. The Review chapter of my textbook is not a substitute. It has the much more limited objective of providing an opportunity for revising some key statistical concepts and results that will be used time and time again in the course. They are central to econometric analysis and if you have not encountered them before, you should postpone your study of econometrics and study statistics first.
0.13 Learning outcomes

After working through the corresponding chapter in the textbook, studying the corresponding slideshows, and doing the starred exercises in the textbook and the additional exercises in this subject guide, you should be able to explain what is meant by all of the items listed in the Overview. You should also be able to explain why they are important. The concepts of efficiency, consistency, and power are often misunderstood by students taking an introductory econometrics course, so make sure that you aware of their precise meanings.

0.14 Additional exercises

[Note: Each chapter has a set of additional exercises. The answers to them are provided at the end of the chapter after the answers to the starred exercises in the text.]

AR.1 A random variable $X$ has a continuous uniform distribution from 0 to 2. Define its probability density function.

AR.2 Find the expected value of $X$ in Exercise AR.1, using the expression given in Box R.1 in the text.

AR.3 Derive $E(X^2)$ for $X$ defined in Exercise AR.1, using the expression given in Box R.1.

AR.4 Derive the population variance and the standard deviation of $X$ as defined in Exercise AR.1, using the expression given in Box R.1.

AR.5 Using equation (R.9), find the variance of the random variable $X$ defined in Exercise AR.1 and show that the answer is the same as that obtained in Exercise AR.4. (Note: You have already calculated $E(X)$ in Exercise AR.2 and $E(X^2)$ in Exercise AR.3.)

AR.6 In Table R.6, $\mu_0$ and $\mu_1$ were three standard deviations apart. Construct a similar table for the case where they are two standard deviations apart.
AR.7 Suppose that a random variable \( X \) has a normal distribution with unknown mean \( \mu \) and variance \( \sigma^2 \). To simplify the analysis, we shall assume that \( \sigma^2 \) is known. Given a sample of observations, an estimator of \( \mu \) is the sample mean, \( \bar{X} \). An investigator wishes to test \( H_0 : \mu = 0 \) and believes that the true value cannot be negative. The appropriate alternative hypothesis is therefore \( H_1 : \mu > 0 \) and the investigator decides to perform a one-sided test. However, the investigator is mistaken because \( \mu \) could in fact be negative. What are the consequences of erroneously performing a one-sided test when a two-sided test would have been appropriate?

AR.8 Suppose that a random variable \( X \) has a normal distribution with mean \( \mu \) and variance \( \sigma^2 \). Given a sample of \( n \) independent observations, it can be shown that:

\[
\hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2
\]

is an unbiased estimator of \( \sigma^2 \). Is \( \sqrt{\hat{\sigma}^2} \) either an unbiased or a consistent estimator of \( \sigma \)?

0.15  **Answers to the starred exercises in the textbook**

R.2 A random variable \( X \) is defined to be the larger of the two values when two dice are thrown, or the value if the values are the same. Find the probability distribution for \( X \).

**Answer:**

The table shows the 36 possible outcomes. The probability distribution is derived by counting the number of times each outcome occurs and dividing by 36. The probabilities have been written as fractions, but they could equally well have been written as decimals.

<table>
<thead>
<tr>
<th>red</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value of X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>
R.4 Find the expected value of $X$ in Exercise R.2.

**Answer:**

The table is based on Table R.2 in the text. It is a good idea to guess the outcome before doing the arithmetic. In this case, since the higher numbers have the largest probabilities, the expected value should clearly lie between 4 and 5. If the calculated value does not conform with the guess, it is possible that this is because the guess was poor. However, it may be because there is an error in the arithmetic, and this is one way of catching such errors.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$p$</th>
<th>$Xp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>2</td>
<td>3/36</td>
<td>6/36</td>
</tr>
<tr>
<td>3</td>
<td>5/36</td>
<td>15/36</td>
</tr>
<tr>
<td>4</td>
<td>7/36</td>
<td>28/36</td>
</tr>
<tr>
<td>5</td>
<td>9/36</td>
<td>45/36</td>
</tr>
<tr>
<td>6</td>
<td>11/36</td>
<td>66/36</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>161/36 = 4.4722</td>
<td></td>
</tr>
</tbody>
</table>

R.7 Calculate $E(X^2)$ for $X$ defined in Exercise R.2.

**Answer:**

The table is based on Table R.3 in the text. Given that the largest values of $X^2$ have the highest probabilities, it is reasonable to suppose that the answer lies somewhere in the range 15–30. The actual figure is 21.97.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$X^2$</th>
<th>$p$</th>
<th>$X^2p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3/36</td>
<td>12/36</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>5/36</td>
<td>45/36</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>7/36</td>
<td>112/36</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>9/36</td>
<td>225/36</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>11/36</td>
<td>396/36</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>791/36 = 21.9722</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R.10 Calculate the population variance and the standard deviation of $X$ as defined in Exercise R.2, using the definition given by equation (R.8).

**Answer:**

The table is based on Table R.4 in the textbook. In this case it is not easy to make a guess. The population variance is 1.97, and the standard deviation, its square root, is 1.40. Note that four decimal places have been used in the working, even though the estimate is reported to only two. This is to eliminate the possibility of the estimate being affected by rounding error.
0.15. Answers to the starred exercises in the textbook

<table>
<thead>
<tr>
<th>$X$</th>
<th>$p$</th>
<th>$X - \mu_X$</th>
<th>$(X - \mu_X)^2$</th>
<th>$(X - \mu_X)^4p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1/36$</td>
<td>$-3.4722$</td>
<td>$12.0563$</td>
<td>$0.3349$</td>
</tr>
<tr>
<td>2</td>
<td>$3/36$</td>
<td>$-2.4722$</td>
<td>$6.1119$</td>
<td>$0.5093$</td>
</tr>
<tr>
<td>3</td>
<td>$5/36$</td>
<td>$-1.4722$</td>
<td>$2.1674$</td>
<td>$0.3010$</td>
</tr>
<tr>
<td>4</td>
<td>$7/36$</td>
<td>$-0.4722$</td>
<td>$0.2230$</td>
<td>$0.0434$</td>
</tr>
<tr>
<td>5</td>
<td>$9/36$</td>
<td>$0.5278$</td>
<td>$0.2785$</td>
<td>$0.0696$</td>
</tr>
<tr>
<td>6</td>
<td>$11/36$</td>
<td>$2.3341$</td>
<td>$2.1674$</td>
<td>$0.7132$</td>
</tr>
</tbody>
</table>

**Total** | 1.9715

R.12 Using equation (R.9), find the variance of the random variable $X$ defined in Exercise R.2 and show that the answer is the same as that obtained in Exercise R.10. (Note: You have already calculated $\mu_X$ in Exercise R.4 and $E(X^2)$ in Exercise R.7.)

**Answer:**

$E(X^2)$ is 21.9722 (Exercise R.7). $E(X)$ is 4.4722 (Exercise R.4), so $\mu_X^2$ is 20.0006. Thus the variance is $21.9722 - 20.0006 = 1.9716$. The last-digit discrepancy between this figure and that in Exercise R.10 is due to rounding error.

R.14 Suppose a variable $Y$ is an exact linear function of $X$:

$$Y = \lambda + \mu X$$

where $\lambda$ and $\mu$ are constants, and suppose that $Z$ is a third variable. Show that $\rho_{XZ} = \rho_{YZ}$

**Answer:**

We start by noting that $(Y_i - \bar{Y}) = \mu (X_i - \bar{X})$. Then:

$$\rho_{YZ} = \frac{E\left[ (Y_i - \bar{Y}) (Z_i - \bar{Z}) \right]}{\sqrt{E\left[ (Y_i - \bar{Y})^2 \right] E\left[ (Z_i - \bar{Z})^2 \right]}}$$

$$= \frac{\mu E\left[ (X_i - \bar{X}) (Z_i - \bar{Z}) \right]}{\sqrt{\mu^2 E\left[ (X_i - \bar{X})^2 \right] E\left[ (Z_i - \bar{Z})^2 \right]}}$$

$$= \frac{\mu E\left[ (X_i - \bar{X}) (Z_i - \bar{Z}) \right]}{\sqrt{\mu^2 E\left[ (X_i - \bar{X})^2 \right] E\left[ (Z_i - \bar{Z})^2 \right]}}$$

$$= \rho_{XZ}.$$

R.16 Show that, when you have $n$ observations, the condition that the generalised estimator $(\lambda_1 X_1 + \cdots + \lambda_n X_n)$ should be an unbiased estimator of $\mu_X$ is $\lambda_1 + \cdots + \lambda_n = 1$. 

© Christopher Dougherty, 2016. All rights reserved. Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
Answer:

\[ E(Z) = E(\lambda_1 X_1 + \cdots + \lambda_n X_n) \]
\[ = E(\lambda_1 X_1) + \cdots + E(\lambda_n X_n) \]
\[ = \lambda_1 E(X_1) + \cdots + \lambda_n E(X_n) \]
\[ = \lambda_1 \mu_X + \cdots + \lambda_n \mu_X \]
\[ = (\lambda_1 + \cdots + \lambda_n) \mu_X. \]

Thus \( E(Z) = \mu_X \) requires \( \lambda_1 + \cdots + \lambda_n = 1. \)

R.19 In general, the variance of the distribution of an estimator decreases when the sample size is increased. Is it correct to describe the estimator as becoming more efficient?

Answer:
No, it is incorrect. When the sample size increases, the variance of the estimator decreases, and as a consequence it is more likely to give accurate results. Because it is improving in this important sense, it is very tempting to describe the estimator as becoming more efficient. But it is the wrong use of the term. Efficiency is a comparative concept that is used when you are comparing two or more alternative estimators, all of them being applied to the same data set with the same sample size. The estimator with the smallest variance is said to be the most efficient. You cannot use efficiency as suggested in the question because you are comparing the variances of the same estimator with different sample sizes.

R.21 Suppose that you have observations on three variables \( X, Y, \) and \( Z, \) and suppose that \( Y \) is an exact linear function of \( Z: \)

\[ Y = \lambda + \mu Z \]

where \( \lambda \) and \( \mu \) are constants. Show that \( \hat{\rho}_{XZ} = \hat{\rho}_{XY}. \) (This is the counterpart of Exercise R.14.)

Answer:
We start by noting that \( (Y_i - \bar{Y}) = \mu (Z_i - \bar{Z}). \) Then:

\[ \hat{\rho}_{XY} = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \]
\[ = \frac{\sum (X_i - \bar{X}) \mu (Z_i - \bar{Z})}{\sqrt{\sum (X_i - \bar{X})^2 \sum \mu^2 (Z_i - \bar{Z})^2}} \]
\[ = \frac{\sum (X_i - \bar{X}) (Z_i - \bar{Z})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Z_i - \bar{Z})^2}} \]
\[ = \hat{\rho}_{XZ} \]
R.26 Show that, in Figures R.18 and R.22, the probabilities of a Type II error are 0.15 in the case of a 5 per cent significance test and 0.34 in the case of a 1 per cent test. Note that the distance between \( \mu_0 \) and \( \mu_1 \) is three standard deviations. Hence the right-hand 5 per cent rejection region begins 1.96 standard deviations to the right of \( \mu_0 \). This means that it is located 1.04 standard deviations to the left of \( \mu_1 \).

Similarly, for a 1 per cent test, the right-hand rejection region starts 2.58 standard deviations to the right of \( \mu_0 \), which is 0.42 standard deviations to the left of \( \mu_1 \).

**Answer:**

For the 5 per cent test, the rejection region starts \( 3 - 1.96 = 1.04 \) standard deviations below \( \mu_1 \), given that the distance between \( \mu_1 \) and \( \mu_0 \) is 3 standard deviations. See Figure R.18. According to the standard normal distribution table, the cumulative probability of a random variable lying 1.04 standard deviations (or less) above the mean is 0.8508. This implies that the probability of it lying 1.04 standard deviations below the mean is 0.1492. For the 1 per cent test, the rejection region starts \( 3 - 2.58 = 0.42 \) standard deviations below the mean. See Figure R.22. The cumulative probability for 0.42 in the standard normal distribution table is 0.6628, so the probability of a Type II error is 0.3372.

R.27 Explain why the difference in the power of a 5 per cent test and a 1 per cent test becomes small when the distance between \( \mu_0 \) and \( \mu_1 \) becomes large.

**Answer:**

The powers of both tests tend to one as the distance between \( \mu_0 \) and \( \mu_1 \) becomes large. The difference in their powers must therefore tend to zero.

R.28 A random variable \( X \) has unknown population mean \( \mu \). A researcher has a sample of observations with sample mean \( \bar{X} \). He wishes to test the null hypothesis \( H_0 : \mu = \mu_0 \). The figure shows the potential distribution of \( \bar{X} \) conditional on \( H_0 \) being true. It may be assumed that the distribution is known to have variance equal to one.

The researcher decides to implement an unorthodox (and unwise) decision rule. He decides to reject \( H_0 \) if \( \bar{X} \) lies in the central 5 per cent of the distribution (the tinted area in the figure).

(a) Explain why his test is a 5 per cent significance test.
(b) Explain in intuitive terms why his test is unwise.

(c) Explain in technical terms why his test is unwise.

Answer:

The following discussion assumes that you are performing a 5 per cent significance test, but it applies to any significance level.

If the null hypothesis is true, it does not matter how you define the 5 per cent rejection region. By construction, the risk of making a Type I error will be 5 per cent. Issues relating to Type II errors are irrelevant when the null hypothesis is true.

The reason that the central part of the conditional distribution is not used as a rejection region is that it leads to problems when the null hypothesis is false. The probability of not rejecting $H_0$ when it is false will be lower. To use the obvious technical term, the power of the test will be lower.

The figure shows the power functions for the test using the conventional upper and lower 2.5 per cent tails and the test using the central region. The horizontal axis is the difference between the true value and the hypothetical value $\mu_0$ in terms of standard deviations. The vertical axis is the power of the test. The first figure has been drawn for the case where the true value is greater than the hypothetical value. The second figure is for the case where the true value is lower than the hypothetical value. It is the same, but reflected horizontally.

The greater the difference between the true value and the hypothetical mean, the more likely is it that the sample mean will lie in the right tail of the distribution conditional on $H_0$ being true, and so the more likely is it that the null hypothesis will be rejected by the conventional test. The figure shows that the power of the test approaches one asymptotically. However, if the central region of the distribution is used as the rejection region, the probability of the sample mean lying in it will diminish as the difference between the true and hypothetical values increases, and the power of the test approaches zero asymptotically. This is an extreme example of a very bad test procedure.

![Figure 1: Power functions of a conventional 5 per cent test and one using the central region (true value $> \mu_0$).](image)
A researcher is evaluating whether an increase in the minimum hourly wage has had an effect on employment in the manufacturing industry in the following three months. Taking a sample of 25 firms, what should she conclude if:

(a) the mean decrease in employment is 9 per cent, and the standard error of the mean is 5 per cent

(b) the mean decrease is 12 per cent, and the standard error is 5 per cent

(c) the mean decrease is 20 per cent, and the standard error is 5 per cent

(d) there is a mean increase of 11 per cent, and the standard error is 5 per cent?

Answer:

There are 24 degrees of freedom, and hence the critical values of $t$ at the 5 per cent, 1 per cent, and 0.1 per cent levels are 2.06, 2.80, and 3.75, respectively.

(a) The $t$ statistic is $-1.80$. Fail to reject $H_0$ at the 5 per cent level.

(b) $t = -2.40$. Reject $H_0$ at the 5 per cent level but not the 1 per cent level.

(c) $t = -4.00$. Reject $H_0$ at the 1 per cent level. Better, reject at the 0.1 per cent level.

(d) $t = 2.20$. This would be a surprising outcome, but if one is performing a two-sided test, then reject $H_0$ at the 5 per cent level but not the 1 per cent level.

R3.33 Demonstrate that the 95 per cent confidence interval defined by equation (R.89) has a 95 per cent probability of capturing $\mu_0$ if $H_0$ is true.

Answer:

If $H_0$ is true, there is 95 per cent probability that:

$$\left| \bar{X} - \mu_0 \right| < t_{crit} \cdot \text{s.e.}(\bar{X})$$
Hence there is 95 per cent probability that $|\bar{X} - \mu_0| \leq t_{\text{crit}} \times \text{s.e.}(\bar{X})$. Hence there is 95 per cent probability that (a) $\bar{X} - \mu_0 < t_{\text{crit}} \times \text{s.e.}(\bar{X})$ and (b) $\mu_0 - \bar{X} < t_{\text{crit}} \times \text{s.e.}(\bar{X})$.

(a) can be rewritten $\bar{X} - t_{\text{crit}} \times \text{s.e.}(\bar{X}) < \mu_0$, giving the lower limit of the confidence interval.

(b) can be rewritten $\bar{X} - \mu_0 > -t_{\text{crit}} \times \text{s.e.}(\bar{X})$ and hence $\bar{X} + t_{\text{crit}} \times \text{s.e.}(\bar{X}) > \mu_0$, giving the upper limit of the confidence interval.

Hence there is 95 per cent probability that $\mu_0$ will lie in the confidence interval.

R.34 In Exercise R.29, a researcher was evaluating whether an increase in the minimum hourly wage has had an effect on employment in the manufacturing industry. Explain whether she might have been justified in performing one-sided tests in cases (a) – (d), and determine whether her conclusions would have been different.

**Answer:**

First, there should be a discussion of whether the effect of an increase in the minimum wage could have a positive effect on employment. If it is decided that it cannot, we can use a one-sided test and the critical values of $t$ at the 5 per cent, 1 per cent, and 0.1 per cent levels become 1.71, 2.49, and 3.47, respectively.

1. The $t$ statistic is $-1.80$. We can now reject $H_0$ at the 5 per cent level.

2. $t = -2.40$. No change, but much closer to rejecting at the 1 per cent level.

3. $t = -4.00$. No change. Reject at the 1 per cent level (and 0.1 per cent level).

4. $t = 2.20$. Here there is a problem because the coefficient has the unexpected sign. In principle we should stick to our guns and fail to reject $H_0$. However, we should consider two further possibilities. One is that the justification for a one-sided test is incorrect (not very likely in this case). The other is that the model is misspecified in some way and the misspecification is responsible for the unexpected sign. For example, the coefficient might be distorted by omitted variable bias, to be discussed in Chapter 6.

R.37 A random variable $X$ has population mean $\mu_X$ and population variance $\sigma_X^2$. A sample of $n$ observations $\{X_1, \ldots, X_n\}$ is generated. Using the plim rules, demonstrate that, subject to a certain condition that should be stated:

$$\text{plim} \left( \frac{1}{\bar{X}} \right) = \frac{1}{\mu_X}.$$  

**Answer:**

$\text{plim} \bar{X} = \mu_X$ by the weak law of large numbers. Provided that $\mu_X \neq 0$, we are entitled to use the plim quotient rule, so:

$$\text{plim} \left( \frac{1}{\bar{X}} \right) = \frac{\text{plim} \frac{1}{\bar{X}}}{\mu_X} = \frac{1}{\mu_X}.$$
R.39 A random variable $X$ has unknown population mean $\mu_X$ and population variance $\sigma^2_X$. A sample of $n$ observations $\{X_1, \ldots, X_n\}$ is generated. Show that:

$$Z = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3 + \cdots + \frac{1}{2^{n-1}}X_n$$

is an unbiased estimator of $\mu_X$. Show that the variance of $Z$ does not tend to zero as $n$ tends to infinity and that therefore $Z$ is an inconsistent estimator, despite being unbiased.

**Answer:**

The weights sum to unity, so the estimator is unbiased. However, its variance is:

$$\sigma^2_Z = \left(\frac{1}{4} + \frac{1}{16} + \cdots + \frac{1}{4^{n-1}} + \frac{1}{4^{n-1}}\right)\sigma^2_X.$$ 

This tends to $\sigma^2_X/3$ as $n$ becomes large, not zero, so the estimator is inconsistent.

Note: the sum of a geometric progression is given by:

$$1 + a + a^2 + \cdots + a^n = \frac{1 - a^{n+1}}{1 - a}.$$ 

Hence:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}} = \frac{1}{2} \left(1 + \frac{1}{4} + \cdots + \frac{1}{2^{n-2}}\right) + \frac{1}{2^{n-1}}$$

$$= \frac{1}{2} \left(\frac{1 - \left(\frac{1}{2}\right)^{n-1}}{1 - \frac{1}{2}}\right) + \frac{1}{2^{n-1}}$$

$$= 1 - \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}} = 1$$

and:

$$\frac{1}{4} + \frac{1}{16} + \cdots + \frac{1}{4^{n-1}} + \frac{1}{4^{n-1}} = \frac{1}{4} \left(1 + \frac{1}{4} + \cdots + \frac{1}{4^{n-2}}\right) + \frac{1}{4^{n-1}}$$

$$= \frac{1}{4} \left(\frac{1 - \left(\frac{1}{4}\right)^{n-1}}{1 - \frac{1}{4}}\right) + \frac{1}{4^{n-1}}$$

$$= \frac{1}{3} \left(1 - \left(\frac{1}{4}\right)^{n-1}\right) + \frac{1}{4^{n-1}} \rightarrow \frac{1}{3}$$

as $n$ becomes large.

R.41 A random variable $X$ has a continuous uniform distribution over the interval from 0 to $\theta$, where $\theta$ is an unknown parameter.
The following three estimators are used to estimate $\theta$, given a sample of $n$ observations on $X$:

(a) twice the sample mean
(b) the largest value of $X$ in the sample
(c) the sum of the largest and smallest values of $X$ in the sample.

Explain verbally whether or not each estimator is (1) unbiased, and (2) consistent.

Answer:

(a) It is evident that $E(X) = E(\bar{X}) = \theta/2$. Hence $2\bar{X}$ is an unbiased estimator of $\theta$. The variance of $\bar{X}$ is $\sigma^2_\bar{X}/n$. The variance of $2\bar{X}$ is therefore $4\sigma^2_\bar{X}/n$. This will tend to zero as $n$ tends to infinity. Thus the distribution of $2\bar{X}$ will collapse to a spike at $\theta$ and the estimator is consistent.

(b) The estimator will be biased downwards since the highest value of $X$ in the sample will always be less than $\theta$. However, as $n$ increases, the distribution of the estimator will be increasingly concentrated in a narrow range just below $\theta$. To put it formally, the probability of the highest value being more than $\epsilon$ below $\theta$ will be $(1 - \epsilon)^n$ and this will tend to zero, no matter how small $\epsilon$ is, as $n$ tends to infinity. The estimator is therefore consistent. It can in fact be shown that the expected value of the estimator is $\frac{n-1}{n}\theta$ and this tends to $\theta$ as $n$ becomes large.

(c) The estimator will be unbiased. Call the maximum value of $X$ in the sample $X_{\text{max}}$ and the minimum value $X_{\text{min}}$. Given the symmetry of the distribution of $X$, the distributions of $X_{\text{max}}$ and $X_{\text{min}}$ will be identical, except that that of $X_{\text{min}}$ will be to the right of 0 and that of $X_{\text{max}}$ will be to the left of $\theta$. Hence, for any $n$, $E(X_{\text{min}}) - 0 = \theta - E(X_{\text{max}})$ and the expected value of their sum is equal to $\theta$. The estimator will be consistent for the same reason as explained in (b).

The first figure shows the distributions of the estimators (a) and (b) for 1,000,000 samples with only four observations in each sample, with $\theta = 1$. The second figure shows the distributions when the number of observations in each sample is equal to

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
0.15. Answers to the starred exercises in the textbook

100. The table gives the means and variances of the distributions as computed from the results of the simulations. If the mean square error is used to compare the estimators, which should be preferred for sample size 4? For sample size 100?

<table>
<thead>
<tr>
<th></th>
<th>Sample size 4</th>
<th>Sample size 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>Mean</td>
<td>1.0000</td>
<td>0.8001</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0833</td>
<td>0.0267</td>
</tr>
<tr>
<td>Estimated bias</td>
<td>0.0000</td>
<td>−0.1999</td>
</tr>
<tr>
<td>Estimated mean square error</td>
<td>0.0833</td>
<td>0.0667</td>
</tr>
</tbody>
</table>

It can be shown (Larsen and Marx, *An Introduction to Mathematical Statistics and Its Applications*, p.382, that estimator (b) is biased downwards by an amount \( \frac{\theta}{n+1} \) and that its variance is:

\[
\frac{n\theta^2}{(n+1)^2(n+2)}
\]
while estimator (a) has variance $\theta^2 / 3n$. How large does $n$ have to be for (b) to be preferred to (a) using the mean square error criterion?

The crushing superiority of (b) over (a) may come as a surprise, so accustomed are we to finding that the sample mean in the best estimator of a parameter. The underlying reason in this case is that we are estimating a boundary parameter, which, as its name implies, defines the limit of a distribution. In such a case the optimal properties of the sample mean are no longer guaranteed and it may be eclipsed by a score statistic such as the largest observation in the sample. Note that the standard deviation of the sample mean is inversely proportional to $\sqrt{n}$, while that of (b) is inversely proportional to $n$ (disregarding the differences between $n$, $n + 1$, and $n + 2$). (b) therefore approaches its limiting (asymptotically unbiased) value much faster than (a) and is said to be superconsistent. We will encounter superconsistent estimators again when we come to cointegration in Chapter 13. Note that if we multiply (b) by $(n + 1)/n$, it is unbiased for finite samples as well as superconsistent.

## 0.16 Answers to the additional exercises

AR.1 The total area under the function over the interval $[0, 2]$ must be equal to 1. Since the length of the rectangle is 2, its height must be 0.5. Hence $f(X) = 0.5$ for $0 \leq X \leq 2$, and $f(X) = 0$ for $X < 0$ and $X > 2$.

AR.2 Obviously, since the distribution is uniform, the expected value of $X$ is 1. However we will derive this formally.

$$E(X) = \int_0^2 Xf(X) \, dX = \int_0^2 0.5X \, dX = \left[ \frac{X^2}{4} \right]_0^2 = \left[ \frac{2^2}{4} \right] - \left[ \frac{0^2}{4} \right] = 1.$$ 

AR.3 The expected value of $X^2$ is given by:

$$E(X^2) = \int_0^2 X^2f(X) \, dX = \int_0^2 0.5X^2 \, dX = \left[ \frac{X^3}{6} \right]_0^2 = \left[ \frac{2^3}{6} \right] - \left[ \frac{0^3}{6} \right] = 1.3333.$$ 

AR.4 The variance of $X$ is given by:

$$E \left( [X - \mu_X]^2 \right) = \int_0^2 [X - \mu_X]^2 f(X) \, dX = \int_0^2 0.5[X - 1]^2 \, dX$$

$$= \int_0^2 (0.5X^2 - X + 0.5) \, dX$$

$$= \left[ \frac{X^3}{6} - \frac{X^2}{2} + \frac{X}{2} \right]_0^2$$

$$= \left[ \frac{8}{6} - 2 + 1 \right] - [0] = 0.3333.$$ 

The standard deviation is equal to the square root, 0.5774.
AR.5 From Exercise AR.3, \( E(X^2) = 1.3333 \). From Exercise AR.2, the square of \( E(X) \) is 1. Hence the variance is 0.3333, as in Exercise AR.4.

AR.6 Table R.6 is reproduced for reference:

<table>
<thead>
<tr>
<th></th>
<th>One-sided test</th>
<th>Two-sided test</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 per cent significance test</td>
<td>0.09</td>
<td>0.15</td>
</tr>
<tr>
<td>2.5 per cent significance test</td>
<td>0.15</td>
<td>(not investigated)</td>
</tr>
<tr>
<td>1 per cent significance test</td>
<td>0.25</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Note: The distance between \( \mu_1 \) and \( \mu_0 \) in this example was 3 standard deviations.

Two-sided tests

Under the (false) \( H_0 : \mu = \mu_0 \), the right rejection region for a two-sided 5 per cent significance test starts 1.96 standard deviations above \( \mu_0 \), which is 0.04 standard deviations below \( \mu_1 \). A Type II error therefore occurs if \( \bar{X} \) is more than 0.04 standard deviations to the left of \( \mu_1 \). Under \( H_1 : \mu = \mu_1 \), the probability is 0.48.

Under \( H_0 \), the right rejection region for a two-sided 1 per cent significance test starts 2.58 standard deviations above \( \mu_0 \), which is 0.58 standard deviations above \( \mu_1 \). A Type II error therefore occurs if \( \bar{X} \) is less than 0.58 standard deviations to the right of \( \mu_1 \). Under \( H_1 : \mu = \mu_1 \), the probability is 0.72.

One-sided tests

Under \( H_0 : \mu = \mu_0 \), the right rejection region for a one-sided 5 per cent significance test starts 1.65 standard deviations above \( \mu_0 \), which is 0.35 standard deviations below \( \mu_1 \). A Type II error therefore occurs if \( \bar{X} \) is more than 0.35 standard deviations to the left of \( \mu_1 \). Under \( H_1 : \mu = \mu_1 \), the probability is 0.36.

Under \( H_0 \), the right rejection region for a one-sided 1 per cent significance test starts 2.33 standard deviations above \( \mu_0 \), which is 0.33 standard deviations above \( \mu_1 \). A Type II error therefore occurs if \( \bar{X} \) is less than 0.33 standard deviations to the right of \( \mu_1 \). Under \( H_1 : \mu = \mu_1 \), the probability is 0.63.

Hence the table is:

<table>
<thead>
<tr>
<th></th>
<th>One-sided test</th>
<th>Two-sided test</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 per cent significance test</td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td>1 per cent significance test</td>
<td>0.63</td>
<td>0.72</td>
</tr>
</tbody>
</table>

AR.7 We will assume for sake of argument that the investigator is performing a 5 per cent significance test, but the conclusions apply to all significance levels.

If the true value is 0, the null hypothesis is true. The risk of a Type I error is, by construction, 5 per cent for both one-sided and two-sided tests. Issues relating to Type II error do not arise because the null hypothesis is true.
If the true value is positive, the investigator is lucky and makes the gain associated with a one-sided test. Namely, the power of the test is uniformly higher than that for a two-sided test for all positive values of $\mu$. The power functions for one-sided and two-sided tests are shown in the first figure below.

If the true value is negative, the power functions are as shown in the second figure. That for the two-sided test is the same as that in the first figure, but reflected horizontally. The larger (negatively) is the true value of $\mu$, the greater will be the probability of rejecting $H_0$ and the power approaches 1 asymptotically. However, with a one-sided test, the power function will decrease from its already very low value. The power is not automatically zero for true values that are negative because even for these it is possible that a sample might have a mean that lies in the right tail of the distribution under the null hypothesis. But the probability rapidly falls to zero as the (negative) size of $\mu$ grows.

Figure 3: Power functions of one-sided and two-sided 5 per cent tests (true value > 0).

Figure 4: Power functions of one-sided and two-sided 5 per cent tests (true value < 0).
AR.8 We will refute the unbiasedness proposition by considering the more general case where $Z^2$ is an unbiased estimator of $\theta^2$. We know that:

$$E[(Z - \theta)^2] = E(Z^2) - 2\theta E(Z) + \theta^2 = 2\theta^2 - 2\theta E(Z).$$

Hence:

$$E(Z) = \theta - \frac{1}{2\theta}E[(Z - \theta)^2].$$

$Z$ is therefore a biased estimator of $\theta$ except for the special case where $Z$ is equal to $\theta$ for all samples, that is, in the trivial case where there is no sampling error. Nevertheless, since a function of a consistent estimator will, under quite general conditions, be a consistent estimator of the function of the parameter, $\sqrt{\sigma^2}$ will be a consistent estimator of $\sigma$. 


Chapter 1
Simple regression analysis

1.1 Overview

This chapter introduces the least squares criterion of goodness of fit and demonstrates, first through examples and then in the general case, how it may be used to develop expressions for the coefficients that quantify the relationship when a dependent variable is assumed to be determined by one explanatory variable. The chapter continues by showing how the coefficients should be interpreted when the variables are measured in natural units, and it concludes by introducing $R^2$, a second criterion of goodness of fit, and showing how it is related to the least squares criterion and the correlation between the fitted and actual values of the dependent variable.

1.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to explain what is meant by:

- dependent variable
- explanatory variable (independent variable, regressor)
- parameter of a regression model
- the nonstochastic component of a true relationship
- the disturbance term
- the least squares criterion of goodness of fit
- ordinary least squares (OLS)
- the regression line
- fitted model
- fitted values (of the dependent variable)
- residuals
- total sum of squares, explained sum of squares, residual sum of squares
- $R^2$. 
1. Simple regression analysis

In addition, you should be able to explain the difference between:

- the nonstochastic component of a true relationship and a fitted regression line, and
- the values of the disturbance term and the residuals.

### 1.3 Additional exercises

A1.1 The output below gives the result of regressing $FDHO$, annual household expenditure on food consumed at home, on $EXP$, total annual household expenditure, both measured in dollars, using the Consumer Expenditure Survey data set. Give an interpretation of the coefficients.

```stata
.reg FDHO EXP if FDHO>0
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>972602566</td>
<td>1</td>
<td>972602566</td>
<td>F( 1, 6332) = 3431.01</td>
</tr>
<tr>
<td>Residual</td>
<td>1.7950e+09</td>
<td>6332</td>
<td>283474.003</td>
<td>R-squared = 0.3514</td>
</tr>
<tr>
<td>Total</td>
<td>2.7676e+09</td>
<td>6333</td>
<td>437006.15</td>
<td>Root MSE = 532.42</td>
</tr>
</tbody>
</table>

| FDHO  | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-------|-------|-----------|------|-------|----------------------|
| EXP   | .0627099 | .0010706 | 58.57 | 0.000 | .0606112 .0648086    |
| _cons | 369.4418 | 10.65718 | 34.67 | 0.000 | 348.5501 390.3334    |

A1.2 Download the $CES$ data set from the website (see Appendix B of the text), perform a regression parallel to that in Exercise A1.1 for your category of expenditure, and provide an interpretation of the regression coefficients.

A1.3 The output shows the result of regressing the weight of the respondent, in pounds, in 2011 on the weight in 2004, using $EAWE$ Data Set 22. Provide an interpretation of the coefficients. Summary statistics for the data are also provided.

```stata
.reg WEIGHT11 WEIGHT04
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>769248.875</td>
<td>1</td>
<td>769248.875</td>
<td>F( 1, 498) = 1207.55</td>
</tr>
<tr>
<td>Residual</td>
<td>317241.693</td>
<td>498</td>
<td>637.031513</td>
<td>R-squared = 0.7080</td>
</tr>
<tr>
<td>Total</td>
<td>1086490.57</td>
<td>499</td>
<td>2177.33581</td>
<td>Root MSE = 25.239</td>
</tr>
</tbody>
</table>

| WEIGHT11 | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-----------|-------|-----------|------|-------|----------------------|
| WEIGHT04  | .9739736 | .0280281 | 34.75 | 0.000 | .9189056 1.029042    |
| _cons     | 17.42232 | 4.888091 | 3.56  | 0.000 | 7.818493 27.02614    |
1.3. Additional exercises

. sum WEIGHT04 WEIGHT11

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEIGHT04</td>
<td>500</td>
<td>169.686</td>
<td>40.31215</td>
<td>95</td>
<td>330</td>
</tr>
<tr>
<td>WEIGHT11</td>
<td>500</td>
<td>182.692</td>
<td>46.66193</td>
<td>95</td>
<td>370</td>
</tr>
</tbody>
</table>

A1.4 The output shows the result of regressing the hourly earnings of the respondent, in dollars, in 2011 on height in 2004, measured in inches, using EAWE Data Set 22. Provide an interpretation of the coefficients, comment on the plausibility of the interpretation, and attempt to give an explanation.

. reg EARNINGS HEIGHT

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1393.77592</td>
<td>1</td>
<td>1393.77592</td>
<td>F( 1, 498) = 9.23</td>
</tr>
<tr>
<td>Residual</td>
<td>75171.3726</td>
<td>498</td>
<td>150.946531</td>
<td>Prob &gt; F = 0.0025</td>
</tr>
<tr>
<td>Total</td>
<td>76565.1485</td>
<td>499</td>
<td>153.437171</td>
<td>R-squared = 0.0182</td>
</tr>
</tbody>
</table>

| EARNINGS | Coef. | Std. Err. | t     | P>|t| | 95% Conf. Interval |
|----------|-------|-----------|-------|-----|-------------------|
| HEIGHT   | .4087231 | .1345068 | 3.04  | 0.003 | .1444523 - .6729938 |
| _cons    | -9.26923 | 9.125089 | -1.02 | 0.310 | -27.19765 - 8.659188 |

A1.5 A researcher has data for 50 countries on $N$, the average number of newspapers purchased per adult in one year, and $G$, GDP per capita, measured in US $\$, and fits the following regression ($RSS = \text{residual sum of squares}$):

$$\hat{N} = 25.0 + 0.020G$$  \hspace{1cm} R^2 = 0.06, \hspace{1cm} RSS = 4,000.0

The researcher realises that GDP has been underestimated by $100 in every country and that $N$ should have been regressed on $G^* = G + 100$. Explain, with mathematical proofs, how the following components of the output would have differed:

- the coefficient of GDP
- the intercept
- $RSS$
- $R^2$.

A1.6 A researcher with the same model and data as in Exercise A1.5 believes that GDP in each country has been underestimated by 50 per cent and that $N$ should have been regressed on $G^* = 2G$. Explain, with mathematical proofs, how the following components of the output would have differed:

- the coefficient of GDP
- the intercept
- $RSS$
- $R^2$.
1. Simple regression analysis

A1.7 Some practitioners of econometrics advocate ‘standardising’ each variable in a regression by subtracting its sample mean and dividing by its sample standard deviation. Thus, if the original regression specification is:

\[ Y_i = \beta_1 + \beta_2 X_i + u_i \]

the revised specification is:

\[ Y_i^* = \beta_1^* + \beta_2^* X_i^* + v_i \]

where:

\[ Y_i^* = \frac{Y_i - \bar{Y}}{\hat{\sigma}_Y} \quad \text{and} \quad X_i^* = \frac{X_i - \bar{X}}{\hat{\sigma}_X} \]

\( \bar{Y} \) and \( \bar{X} \) are the sample means of \( Y \) and \( X \), \( \hat{\sigma}_Y \) and \( \hat{\sigma}_X \) are the estimators of the standard deviations of \( Y \) and \( X \), defined as the square roots of the estimated variances:

\[ \hat{\sigma}_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \quad \text{and} \quad \hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

and \( n \) is the number of observations in the sample. We will write the fitted models for the two specifications as:

\[ \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \]

and:

\[ \hat{Y}_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* \]

Taking account of the definitions of \( Y^* \) and \( X^* \), show that \( \hat{\beta}_1^* = 0 \) and that \( \hat{\beta}_2^* = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} \hat{\beta}_2 \). Provide an interpretation of \( \hat{\beta}_2^* \).

A1.8 For the model described in Exercise A1.7, suppose that \( Y^* \) is regressed on \( X^* \) without an intercept:

\[ \hat{Y}_i^* = \hat{\beta}_2^* X_i^* \]

Determine how \( \hat{\beta}_2^* \) is related to \( \hat{\beta}_2 \).

A1.9 A variable \( Y_i \) is generated as:

\[ Y_i = \beta_1 + u_i \quad (1.1) \]

where \( \beta_1 \) is a fixed parameter and \( u_i \) is a disturbance term that is independently and identically distributed with expected value 0 and population variance \( \sigma_u^2 \). The least squares estimator of \( \beta_1 \) is \( \bar{Y} \), the sample mean of \( Y \). Give a mathematical demonstration that the value of \( R^2 \) in such a regression is zero.

1.4 Answers to the starred exercises in the textbook

1.9 The output shows the result of regressing the weight of the respondent in 2004, measured in pounds, on his or her height, measured in inches, using EAWS Data Set 21. Provide an interpretation of the coefficients.
1.4. Answers to the starred exercises in the textbook

```
. reg WEIGHT04 HEIGHT

Source | SS      df   MS
-------------+------------------------------ F( 1, 498) = 176.74
 Model | 211309   1   211309  Prob > F = 0.0000
 Residual | 595389.95 498  1195.56215 R-squared = 0.2619
-------------+------------------------------ Adj R-squared = 0.2605
 Total | 806698.98 499  1616.63116 Root MSE = 34.577
-------------+------------------------------

WEIGHT04 | Coef.  Std. Err.   t    P>|t|   [95% Conf. Interval]
-------------+---------------------------------------------------------------
HEIGHT | 5.073711  .381639  13.29  0.000   4.32389  5.823532
_cons | -177.1703  25.93501 -6.83  0.000  -228.1258  -126.2147
-------------+---------------------------------------------------------------
```

**Answer:**

Literally the regression implies that, for every extra inch of height, an individual tends to weigh an extra 5.1 pounds. The intercept, which literally suggests that an individual with no height would weigh $-177$ pounds, has no meaning.

1.11 A researcher has international cross-sectional data on aggregate wages, $W$, aggregate profits, $P$, and aggregate income, $Y$, for a sample of $n$ countries. By definition:

$$Y_i = W_i + P_i.$$

The regressions:

$$\hat{W}_i = \hat{\alpha}_1 + \hat{\alpha}_2 Y_i$$
$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 Y_i$$

are fitted using OLS regression analysis. Show that the regression coefficients will automatically satisfy the following equations:

$$\hat{\alpha}_2 + \hat{\beta}_2 = 1$$
$$\hat{\alpha}_1 + \hat{\beta}_1 = 0.$$  

**Answer:**

$$\hat{\alpha}_2 + \hat{\beta}_2 = \frac{\sum (Y_i \bar{Y}) (W_i \bar{W})}{\sum (Y_i \bar{Y})^2} + \frac{\sum (Y_i \bar{Y}) (P_i \bar{P})}{\sum (Y_i \bar{Y})^2}$$

$$= \frac{\sum (Y_i \bar{Y}) (W_i + P_i \bar{W} - \bar{P})}{\sum (Y_i \bar{Y})^2}$$

$$= \frac{\sum (Y_i \bar{Y}) (Y_i \bar{Y})}{\sum (Y_i \bar{Y})^2}$$

$$= 1$$
1. Simple regression analysis

\[ \hat{\alpha}_1 + \hat{\beta}_1 = \left( \bar{W} - \hat{\alpha}_2 \bar{Y} \right) + \left( \bar{P} - \hat{\beta}_2 \bar{Y} \right) = \left( \bar{W} + \bar{P} \right) - (\hat{\alpha}_2 + \hat{\beta}_2) \bar{Y} = \bar{Y} - \bar{Y} = 0. \]

The intuitive explanation is that the regressions break down income into predicted wages and profits and one would expect the sum of the predicted components of income to be equal to its actual level. The sum of the predicted components is \( \hat{W}_i + \hat{P}_i = (\hat{\alpha}_1 + \hat{\alpha}_2 Y_i) + (\hat{\beta}_1 + \hat{\beta}_2 Y_i) \), and in general this will be equal to \( Y_i \) only if the two conditions are satisfied.

1.13 Suppose that the units of measurement of \( X \) are changed so that the new measure, \( X^* \), is related to the original one by \( X^*_i = \mu_2 X_i \). Show that the new estimate of the slope coefficient is \( \hat{\beta}_2^* / \mu_2 \), where \( \hat{\beta}_2^* \) is the slope coefficient in the original regression.

**Answer:**

\[ \hat{\beta}_2^* = \sum \left( X^*_i - \bar{X}^* \right) \left( Y_i - \bar{Y} \right) \frac{\sum \left( X^*_i - \bar{X}^* \right)^2}{\sum \left( \mu_2 X_i - \mu_2 \bar{X} \right)^2} = \frac{\mu_2 \sum \left( X_i - \bar{X} \right) \left( Y_i - \bar{Y} \right)}{\mu_2^2 \sum \left( X_i - \bar{X} \right)^2} = \frac{\hat{\beta}_2}{\mu_2}. \]

1.14 Demonstrate that if \( X \) is demeaned but \( Y \) is left in its original units, the intercept in a regression of \( Y \) on demeaned \( X \) will be equal to \( \bar{Y} \).

**Answer:**

Let \( X^*_i = X_i - \bar{X} \) and \( \hat{\beta}_1^* \) and \( \hat{\beta}_2^* \) be the intercept and slope coefficient in a regression of \( Y \) on \( X^* \). Note that \( \bar{X}^* = 0 \). Then:

\[ \hat{\beta}_1^* = \bar{Y} - \hat{\beta}_2^* \bar{X}^* = \bar{Y}. \]

The slope coefficient is not affected by demeaning:

\[ \hat{\beta}_2^* = \sum \left( X^*_i - \bar{X}^* \right) \left( Y_i - \bar{Y} \right) \frac{\sum \left( X^*_i - \bar{X}^* \right)^2}{\sum \left( X_i - \bar{X} \right)^2} = \hat{\beta}_2. \]

1.15 The regression output shows the result of regressing weight on height using the same sample as in Exercise 1.9, but with weight and height measured in kilos and centimetres: \( WMETRIC = 0.454 \times WEIGHT04 \) and \( HMETRIC = 2.54 \times HEIGHT \).

Confirm that the estimates of the intercept and slope coefficient are as should be expected from the changes in the units of measurement.
1.4. Answers to the starred exercises in the textbook

. gen WTMETRIC = 0.454*WEIGHT04
. gen HMETRIC = 2.54*HEIGHT

. reg WTMETRIC HMETRIC

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>43554.1641</td>
<td>1</td>
<td>43554.1641</td>
<td>F(  1,  498) = 176.74</td>
</tr>
<tr>
<td>Residual</td>
<td>122719.394</td>
<td>498</td>
<td>246.424486</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>166273.558</td>
<td>499</td>
<td>333.213544</td>
<td>R-squared = 0.2619</td>
</tr>
</tbody>
</table>

| WMETRIC | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|---------|-------|-----------|------|------|---------------------|
| HMETRIC | .9068758 | .0682142 | 13.29 | 0.000 | .7728527 1.040899 |
| _cons   | -80.43529 | 11.77449 | -6.83 | 0.000 | -103.5691 -57.30148 |

Answer:

Abbreviate WEIGHT04 to W, HEIGHT to H, WMETRIC to WM, and HMETRIC to HM. WM = 0.454W and HM = 2.54H. The slope coefficient and intercept for the regression in metric units, \( \hat{\beta}_2^M \) and \( \hat{\beta}_1^M \), are then given by:

\[
\hat{\beta}_2^M = \frac{\sum (H_i - \bar{H}) (W_i - \bar{W})}{\sum (H_i - \bar{H})^2} = \frac{\sum 2.54 (H_i - \bar{H}) 0.454 (W_i - \bar{W})}{\sum 2.54^2 (H_i - \bar{H})^2} = \frac{0.179 \sum (H_i - \bar{H}) (W_i - \bar{W})}{\sum (H_i - \bar{H})^2} = 0.179 \hat{\beta}_2 = 0.179 \times 5.074 = 0.908
\]

\[
\hat{\beta}_1^M = 0.454 \bar{H} - \hat{\beta}_2^M \bar{W} = 0.454 \bar{W} - \left( \frac{0.454 \hat{\beta}_2}{2.54} \right) (2.54 \bar{H}) = 0.454 (\bar{W} - \hat{\beta}_2 \bar{H}) = 0.454 \hat{\beta}_1 = 0.454 \times -177.2 = -80.4
\]
1. Simple regression analysis

The regression output confirms that the calculations are correct (subject to rounding error in the last digit).

1.16 Consider the regression model:

\[ Y_i = \beta_1 + \beta_2 X_i + u_i. \]

It implies:

\[ \bar{Y} = \beta_1 + \beta_2 \bar{X} + \bar{u} \]

and hence that:

\[ Y_i^* = \beta_2 X_i^* + v_i \]

where \( Y_i^* = Y_i - \bar{Y}, X_i^* = X_i - \bar{X} \) and \( v_i = u_i - \bar{u} \).

Demonstrate that a regression of \( Y^* \) on \( X^* \) using (1.49) will yield the same estimate of the slope coefficient as a regression of \( Y \) on \( X \). Note: (1.49) should be used instead of (1.35) because there is no intercept in this model.

Evaluate the outcome if the slope coefficient were estimated using (1.35), despite the fact that there is no intercept in the model.

Determine the estimate of the intercept if \( Y^* \) were regressed on \( X^* \) with an intercept included in the regression specification.

**Answer:**

Let \( \hat{\beta}_2 \) be the slope coefficient in a regression of \( Y^* \) on \( X^* \) using (1.49). Then:

\[
\hat{\beta}_2 = \frac{\sum X_i^* Y_i^*}{\sum X_i^{*2}} = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \hat{\beta}_2.
\]

Let \( \hat{\beta}_{2*} \) be the slope coefficient in a regression of \( Y^* \) on \( X^* \) using (1.35). Note that \( \bar{Y}^* \) and \( \bar{X}^* \) are both zero. Then:

\[
\hat{\beta}_{2*} = \frac{\sum (X_i^* - \bar{X}^*)(Y_i^* - \bar{Y}^*)}{\sum (X_i^* - \bar{X}^*)^2} = \frac{\sum X_i^* Y_i^*}{\sum X_i^{*2}} = \hat{\beta}_2.
\]

Let \( \hat{\beta}_{1*} \) be the intercept in a regression of \( Y^* \) on \( X^* \) using (1.35). Then:

\[
\hat{\beta}_{1*} = \bar{Y}^* - \hat{\beta}_{2*}\bar{X}^* = 0.
\]

1.18 Demonstrate that the fitted values of the dependent variable are uncorrelated with the residuals in a simple regression model. (This result generalises to the multiple regression case.)

**Answer:**

The numerator of the sample correlation coefficient for \( \hat{Y} \) and \( \hat{u} \) can be decomposed as follows, using the fact that \( \bar{u} = 0 \):

\[
\frac{1}{n} \sum (\hat{Y}_i - \bar{Y})(\hat{u}_i - \bar{u}) = \frac{1}{n} \sum \left( [\hat{\beta}_1 + \hat{\beta}_2 X_i] - [\bar{\beta}_1 + \bar{\beta}_2 \bar{X}] \right) \hat{u}_i \\
= \frac{1}{n} \hat{\beta}_2 \sum (X_i - \bar{X}) \hat{u}_i \\
= 0
\]
Demonstrate that, in a regression with an intercept, a regression of $Y$ on $X^*$ must have the same $R^2$ as a regression of $Y$ on $X$, where $X^* = \mu_2 X$.

**Answer:**
Let the fitted regression of $Y$ on $X^*$ be written $\hat{Y}_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^*$.

$$\hat{\beta}_1^* = \bar{Y} - \hat{\beta}_2^* \bar{X} = \bar{Y} - \frac{\hat{\beta}_2}{\mu_2} \bar{X} = \hat{\beta}_1.$$

Hence:

$$\hat{Y}_i^* = \hat{\beta}_1 + \frac{\hat{\beta}_2}{\mu_2} \mu_2 X_i = \hat{Y}_i.$$

The fitted and actual values of $Y$ are not affected by the transformation and so $R^2$ is unaffected.

1.25 The output shows the result of regressing weight in 2011 on height, using EAWE Data Set 21. In 2011 the respondents were aged 27–31. Explain why $R^2$ is lower than in the regression reported in Exercise 1.9.

```
. reg WEIGHT11 HEIGHT
Source | SS   df       MS
---------|--------+----------+----------
Model    | 236642.736  1 236642.736
Residual | 841926.912 498 1690.61629
---------|--------+----------+----------
Total    | 1078569.65 499 2161.46222
---------|--------+----------+----------
Number of obs = 500
F(  1, 498) = 139.97
Prob > F = 0.0000
R-squared = 0.2194
Adj R-squared = 0.2178
Root MSE = 41.117

WEIGHT11 | Coef.  Std. Err.  t   P>|t|  [95% Conf. Interval]
---------|--------+----------+----------+----------+------------------------------------------
HEIGHT   | 5.369246  .4538259  11.83  0.000   4.477597   6.260895
_cons    | -184.7802  30.8406  -5.99  0.000  -245.3739  -124.1865
---------|--------+----------+----------+----------+------------------------------------------
```

**Answer:**
The explained sum of squares is actually higher than that in Exercise 1.9. The reason for the fall in $R^2$ is the huge increase in the total sum of squares, no doubt caused by the cumulative effect of variations in eating habits.

### 1.5 Answers to the additional exercises

**A1.1** Expenditure on food consumed at home increases by 6.3 cents for each dollar of total household expenditure. Literally the intercept implies that $369 would be spent on food consumed at home if total household expenditure were zero. Obviously, such an interpretation does not make sense. If the explanatory variable were income, and household income were zero, positive expenditure on food at home would still be possible if the household received food stamps or other transfers, but here the explanatory variable is total household expenditure.
1. Simple regression analysis

A1.2 For each category, the regression sample has been restricted to households with nonzero expenditure. All the slope coefficients are highly significant. Housing has the largest coefficient, as one should expect. Surprisingly, it is followed by education. However, most households spent nothing at all on this category. For those that did, it was important.

<table>
<thead>
<tr>
<th>EXP</th>
<th>n</th>
<th>$\hat{\beta}_2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>0.0235</td>
<td>0.228</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>0.0316</td>
<td>0.176</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>0.0409</td>
<td>0.134</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>0.1202</td>
<td>0.241</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>0.0131</td>
<td>0.180</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>0.0527</td>
<td>0.354</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>0.0627</td>
<td>0.351</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>0.0058</td>
<td>0.082</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.0522</td>
<td>0.102</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>0.0373</td>
<td>0.278</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>0.0574</td>
<td>0.174</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>0.1976</td>
<td>0.469</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>0.0193</td>
<td>0.101</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.0068</td>
<td>0.059</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.0329</td>
<td>0.102</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>0.0069</td>
<td>0.213</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>0.0048</td>
<td>0.104</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>0.0045</td>
<td>0.034</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>0.0160</td>
<td>0.268</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.0040</td>
<td>0.051</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>0.0165</td>
<td>0.088</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>0.0145</td>
<td>0.076</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.0466</td>
<td>0.186</td>
</tr>
</tbody>
</table>

A1.3 The summary data indicate that, on average, the respondents put on 13 pounds over the period 2004–2011. Was this due to the relatively heavy becoming even heavier, or to a general increase in weight? The regression output indicates that weight in 2011 was approximately equal to weight in 2004 plus 17 pounds, so the second explanation appears to be the correct one. Note that this is an instance where the constant term can be given a meaningful interpretation and where it is as of much interest as the slope coefficient. The $R^2$ indicates that 2004 weight accounts for 71 per cent of the variance in 2011 weight, so other factors are important.

A1.4 The slope coefficient indicates that hourly earnings increase by 41 cents for every extra inch of height. The negative intercept has no possible interpretation. The interpretation of the slope coefficient is obviously highly implausible, so we know that something must be wrong with the model. The explanation is that this is a very poorly specified earnings function and that, in particular, we are failing to control for the sex of the respondent. Later on, in Chapter 5, we will find that
males earn more than females, controlling for observable characteristics. Males also tend to be taller. Hence we find an apparent positive association between earnings and height in a simple regression. Note that $R^2$ is very low.

A1.5 **The coefficient of GDP:** Let the revised measure of GDP be denoted $G^*$, where $G^* = G + 100$. Since $G^*_i = G_i + 100$ for all $i$ and $G^* = G + 100$ and so $G^*_i - G^* = G_i - G$ for all $i$. Hence the new slope coefficient is:

$$
\hat{\beta}^*_2 = \frac{\sum (G^*_i - G^*) (N_i - \bar{N})}{\sum (G^*_i - G^*)^2} = \frac{\sum (G_i - G) (N_i - \bar{N})}{\sum (G_i - G)^2} = \hat{\beta}_2.
$$

The coefficient is unchanged.

**The intercept:** The new intercept is:

$$
\hat{\beta}^*_1 = \bar{N} - \hat{\beta}^*_2 G^* = \bar{N} - \hat{\beta}_2 (G + 100) = \hat{\beta}_1 - 100\hat{\beta}_2 = 23.0.
$$

**RSS:** The residual in observation $i$ in the new regression, $\hat{u}^*_i$, is given by:

$$
\hat{u}^*_i = N_i - \hat{\beta}^*_1 - \hat{\beta}^*_2 G^*_i = N_i - (\hat{\beta}_1 - 100\hat{\beta}_2) - \hat{\beta}_2 (G_i + 100) = \hat{u}_i
$$

the residual in the original regression. Hence $RSS$ is unchanged.

**$R^2$:**

$$
R^2 = 1 - \frac{RSS}{\sum (N_i - \bar{N})^2}
$$

and is unchanged since $RSS$ and $\sum (N_i - \bar{N})^2$ are unchanged.

Note that this makes sense intuitively. $R^2$ is unit-free and so it is not possible for the overall fit of a relationship to be affected by the units of measurement.

A1.6 **The coefficient of GDP:** Let the revised measure of GDP be denoted $G^*$, where $G^* = 2G$. Since $G^*_i = 2G_i$ for all $i$, $G^* = 2G$ and so $G^*_i - G^* = 2(G_i - G)$ for all $i$. Hence the new slope coefficient is:

$$
\hat{\beta}^*_2 = \frac{\sum (G^*_i - G^*) (N_i - \bar{N})}{\sum (G^*_i - G^*)^2} = \frac{\sum 2(G_i - G) (N_i - \bar{N})}{\sum 4(G_i - G)^2} = \frac{2 \sum (G_i - G) (N_i - \bar{N})}{4 \sum (G_i - G)^2} = \frac{\hat{\beta}_2}{2} = 0.010
$$
1. Simple regression analysis

where $\hat{\beta}_2 = 0.020$ is the slope coefficient in the original regression.

The intercept: The new intercept is:

$$\hat{\beta}_1^* = \bar{N} - \hat{\beta}_2^* \bar{G} = \bar{N} - \frac{\hat{\beta}_2^*}{2} \bar{G} = \bar{N} - \frac{\hat{\beta}_2}{2} \bar{G} = \hat{\beta}_1 = 25.0$$

the original intercept.

RSS: The residual in observation $i$ in the new regression, $\hat{u}_i^*$, is given by:

$$\hat{u}_i^* = N_i - \hat{\beta}_1^* - \hat{\beta}_2^* G_i^* = N_i - \hat{\beta}_1 - \frac{\hat{\beta}_2}{2} G_i = \hat{u}_i$$

the residual in the original regression. Hence RSS is unchanged.

$R^2$:

$$R^2 = 1 - \frac{RSS}{\sum (N_i - \bar{N})^2}$$

and is unchanged since RSS and $\sum (N_i - \bar{N})^2$ are unchanged. As in Exercise A1.6, this makes sense intuitively.

A1.7 By construction, $\bar{Y}^* = \bar{X}^* = 0$. So $\hat{\beta}_1^* = Y^* - \hat{\beta}_2^* X^* = 0$.

$$\hat{\beta}_2^* = \frac{\sum (X_i^* - \bar{X}^*) (Y_i^* - \bar{Y}^*)}{\sum (X_i^* - \bar{X}^*)^2}$$

$$\quad = \frac{\sum X_i^* Y_i^*}{\sum X_i^2}$$

$$\quad = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\quad = \frac{\hat{\sigma}_Y \sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\quad = \frac{\hat{\sigma}_Y \hat{\beta}_2}{\hat{\sigma}_Y^2}.$$

$\hat{\beta}_2^*$ provides an estimate of the effect on $Y$, in terms of standard deviations of $Y$, of a one-standard deviation change in $X$.

A1.8 We have:

$$\hat{\beta}_2^{**} = \frac{\sum X_i^* Y_i^*}{\sum X_i^2} = \frac{\sum (X_i^* - \bar{X}^*) (Y_i^* - \bar{Y}^*)}{\sum (X_i^* - \bar{X}^*)^2} = \hat{\beta}_2^*.$$
1.5. Answers to the additional exercises

A1.9 We have:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

and \( \hat{Y}_i = \bar{Y} \) for all \( i \).
1. Simple regression analysis
Chapter 2  
Properties of the regression coefficients and hypothesis testing

2.1 Overview

Chapter 1 introduced least squares regression analysis, a mathematical technique for fitting a relationship given suitable data on the variables involved. It is a fundamental chapter because much of the rest of the text is devoted to extending the least squares approach to handle more complex models, for example models with multiple explanatory variables, nonlinear models, and models with qualitative explanatory variables. However, the mechanics of fitting regression equations are only part of the story. We are equally concerned with assessing the performance of our regression techniques and with developing an understanding of why they work better in some circumstances than in others. Chapter 2 is the starting point for this objective and is thus equally fundamental. In particular, it shows how two of the three main criteria for assessing the performance of estimators, unbiasedness and efficiency, are applied in the context of a regression model. The third criterion, consistency, will be considered in Chapter 8.

2.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to explain what is meant by:

- cross-sectional, time series, and panel data
- unbiasedness of OLS regression estimators
- variance and standard errors of regression coefficients and how they are determined
- Gauss–Markov theorem and efficiency of OLS regression estimators
- two-sided \( t \) tests of hypotheses relating to regression coefficients and one-sided \( t \) tests of hypotheses relating to regression coefficients
- \( F \) tests of goodness of fit of a regression equation

in the context of a regression model. The chapter is a long one and you should take your time over it because it is essential that you develop a perfect understanding of every detail.
2. Properties of the regression coefficients and hypothesis testing

2.3 Further material

Derivation of the expression for the variance of the naïve estimator in Section 2.3.

The variance of the naïve estimator in Section 2.3 and Exercise 2.9 is not of any great interest in itself, but its derivation provides an example of how one obtains expressions for variances of estimators in general.

In Section 2.3 we considered the naïve estimator of the slope coefficient derived by joining the first and last observations in a sample and calculating the slope of that line:

$$\hat{\beta}_2 = \frac{Y_n - Y_1}{X_n - X_1}.$$ 

It was demonstrated that the estimator could be decomposed as:

$$\hat{\beta}_2 = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

and hence that $E(\hat{\beta}_2) = \beta_2$.

The population variance of a random variable $X$ is defined to be $E([X - \mu_X]^2)$ where $\mu_X = E(X)$. Hence the population variance of $\hat{\beta}_2$ is given by:

$$\sigma^2_{\hat{\beta}_2} = E\left((\hat{\beta}_2 - \beta_2)^2\right) = E\left(\left(\beta_2 + \frac{u_n - u_1}{X_n - X_1}\right) - \beta_2\right)^2 = E\left(\left[\frac{u_n - u_1}{X_n - X_1}\right]^2\right).$$

On the assumption that $X$ is nonstochastic, this can be written as:

$$\sigma^2_{\hat{\beta}_2} = \left[\frac{1}{X_n - X_1}\right]^2 E\left([u_n - u_1]^2\right).$$

Expanding the quadratic, we have:

$$\sigma^2_{\hat{\beta}_2} = \left[\frac{1}{X_n - X_1}\right]^2 E\left(u_n^2 + u_1^2 - 2u_nu_1\right) = \left[\frac{1}{X_n - X_1}\right]^2 \left[E(u_n^2) + E(u_1^2) - 2E(u_nu_1)\right].$$

Each value of the disturbance term is drawn randomly from a distribution with mean 0 and population variance $\sigma^2_u$, so $E(u_n^2)$ and $E(u_1^2)$ are both equal to $\sigma^2_u$. $u_n$ and $u_1$ are drawn independently from the distribution, so $E(u_nu_1) = E(u_n)E(u_1) = 0$. Hence:

$$\sigma^2_{\hat{\beta}_2} = \frac{2\sigma^2_u}{(X_n - X_1)^2} = \frac{\sigma^2_u}{\frac{1}{2}(X_n - X_1)^2}.$$
Define $A = \frac{1}{2}(X_1 + X_n)$, the average of $X_1$ and $X_n$, and $D = X_n - A = A - X_1$. Then:

$$\frac{1}{2}(X_n - X_1)^2 = \frac{1}{2}(X_n - A + A - X_1)^2$$

$$= \frac{1}{2}[(X_n - A)^2 + (A - X_1)^2 + 2(X_n - A)(A - X_1)]$$

$$= \frac{1}{2}[D^2 + D^2 + 2(D)(D)] = 2D^2$$

$$= (X_n - A)^2 + (A - X_1)^2$$

$$= (X_n - A)^2 + (X_1 - A)^2$$

$$= (X_n - \bar{X} + \bar{X} - A)^2 + (X_1 - \bar{X} + \bar{X} - A)^2$$

$$= (X_n - \bar{X})^2 + (\bar{X} - A)^2 + 2(X_n - \bar{X})(\bar{X} - A) + (X_1 - \bar{X})^2 + (\bar{X} - A)^2 + 2(X_1 - \bar{X})(\bar{X} - A)$$

$$= (X_1 - \bar{X})^2 + (X_n - \bar{X})^2 + 2(\bar{X} - A)^2 + 2(X_1 + X_n - 2\bar{X})(\bar{X} - A)$$

$$= (X_1 - \bar{X})^2 + (X_n - \bar{X})^2 + 2(\bar{X} - A)^2 + 2(2A - 2\bar{X})(\bar{X} - A)$$

$$= (X_1 - \bar{X})^2 + (X_n - \bar{X})^2 - 2(\bar{X} - A)^2$$

$$= (X_1 - \bar{X})^2 + (X_n - \bar{X})^2 - 2(A - \bar{X})^2$$

$$= (X_1 - \bar{X})^2 + (X_n - \bar{X})^2 - \frac{1}{2}(X_1 + X_n - 2\bar{X})^2.$$

Hence we obtain the expression in Exercise 2.9. There must be a shorter proof.

### 2.4 Additional exercises

**A2.1** A variable $Y$ depends on a nonstochastic variable $X$ with the relationship:

$$Y = \beta_1 + \beta_2 X + u$$

where $u$ is a disturbance term that satisfies the regression model assumptions.

Given a sample of $n$ observations, a researcher decides to estimate $\beta_2$ using the expression:

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}.$$

(This is the OLS estimator of $\beta_2$ for the model $Y = \beta_2 X + u$.)

(a) Demonstrate that $\hat{\beta}_2$ is in general a biased estimator of $\beta_2$.

(b) Discuss whether it is possible to determine the sign of the bias.

(c) Demonstrate that $\hat{\beta}_2$ is unbiased if $\beta_1 = 0$.

(d) Demonstrate that $\hat{\beta}_2$ is unbiased if $\bar{X} = 0$.

**A2.2** A variable $Y_i$ is generated as:

$$Y_i = \beta_1 + u_i$$
2. Properties of the regression coefficients and hypothesis testing

where \( \beta_1 \) is a fixed parameter and \( u_i \) is a disturbance term that is independently and identically distributed with expected value 0 and population variance \( \sigma^2_u \). The least squares estimator of \( \beta_1 \) is \( \bar{Y} \), the sample mean of \( Y \). However, a researcher believes that \( Y \) is a linear function of another variable \( X \) and uses ordinary least squares to fit the relationship:

\[
\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X
\]

calculating \( \hat{\beta}_1 \) as \( \bar{Y} - \hat{\beta}_2 \bar{X} \), where \( \bar{X} \) is the sample mean of \( X \). \( X \) may be assumed to be a nonstochastic variable. Determine whether the researcher’s estimator \( \hat{\beta}_1 \) is biased or unbiased, and if biased, determine the direction of the bias.

A2.3 With the model described in Exercise A2.2, standard theory states that the population variance of the researcher’s estimator of \( \beta_1 \) is:

\[
\sigma^2_u \left[ \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right].
\]

In general, this is larger than the population variance of \( \bar{Y} \), which is \( \sigma^2_u / n \). Explain the implications of the difference in the variances.

In the special case where \( \bar{X} = 0 \), the variances are the same. Give an intuitive explanation.

A2.4 A variable \( Y \) depends on a nonstochastic variable \( X \) with the relationship:

\[
Y = \beta_1 + \beta_2 X + u
\]

where \( u \) is a disturbance term that satisfies the regression model assumptions. Given a sample of \( n \) observations, a researcher decides to estimate \( \beta_2 \) using the expression:

\[
\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}.
\]

It can be shown that the population variance of this estimator is \( \sigma^2_u / \sum X_i^2 \).

We saw in Exercise A2.1 that \( \hat{\beta}_2 \) is in general a biased estimator of \( \beta_2 \). However, if either \( \beta_1 = 0 \) or \( \bar{X} = 0 \), the estimator is unbiased. What can be said in this case about the efficiency of the estimator in these two cases, comparing it with the estimator:

\[
\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}.
\]

Returning to the general case where \( \beta_1 \neq 0 \) and \( \bar{X} \neq 0 \), suppose that there is very little variation in \( X \) in the sample. Is it possible that \( \hat{\beta}_2 \) might be a better estimator than the OLS estimator?

A2.5 Using the output for the regression in Exercise A1.1, reproduced below, perform appropriate statistical tests.
. reg FDHO EXP if FDHO>0

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>972602566</td>
<td>1</td>
<td>972602566</td>
<td>F( 1, 6332) = 3431.01</td>
</tr>
<tr>
<td>Residual</td>
<td>1.7950e+09</td>
<td>6332</td>
<td>283474.003</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2.7676e+09</td>
<td>6333</td>
<td>437006.15</td>
<td>R-squared = 0.3514</td>
</tr>
</tbody>
</table>

FDHO | Coef.  | Std. Err. | t    | P>|t| | [95% Conf. Interval]
-----|-------|-----------|------|------|----------------------|
EXP  | .0627099 | .0010706 | 58.57 | 0.000 | .0606112 .0648086 |
_cons | 369.4418 | 10.65718 | 34.67 | 0.000 | 348.5501 390.3334 |

A2.6 Using the output for your regression in Exercise A1.2, perform appropriate statistical tests.

A2.7 Using the output for the regression of weight in 2004 on height in Exercise 1.9, reproduced below, perform appropriate statistical tests.

. reg WEIGHT04 HEIGHT

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>211309</td>
<td>1</td>
<td>211309</td>
<td>F( 1, 498) = 176.74</td>
</tr>
<tr>
<td>Residual</td>
<td>595389.95</td>
<td>498</td>
<td>1195.56215</td>
<td>R-squared = 0.2619</td>
</tr>
<tr>
<td>Total</td>
<td>806698.95</td>
<td>499</td>
<td>1616.63116</td>
<td>Adj R-squared = 0.2605</td>
</tr>
</tbody>
</table>

WEIGHT04 | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval]
----------|-------|-----------|------|------|----------------------|
HEIGHT    | 5.073711 | .381639 | 13.29 | 0.000 | 4.32389 5.823532 |
_cons     | -177.1703 | 25.93501 | -6.83 | 0.000 | -228.1258 -126.2147 |

A2.8 Using the output for the regression of earnings on height in Exercise A1.4, reproduced below, perform appropriate statistical tests.

. reg EARNINGS HEIGHT

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1393.77592</td>
<td>1</td>
<td>1393.77592</td>
<td>F( 1, 498) = 9.23</td>
</tr>
<tr>
<td>Residual</td>
<td>75171.3726</td>
<td>498</td>
<td>150.946531</td>
<td>R-squared = 0.0182</td>
</tr>
<tr>
<td>Total</td>
<td>76565.1485</td>
<td>499</td>
<td>153.43717</td>
<td>Adj R-squared = 0.0162</td>
</tr>
</tbody>
</table>

EARNINGS | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval]
----------|-------|-----------|------|------|----------------------|
HEIGHT    | .4087231 | .1345068 | 3.04 | 0.003 | .1444523 0.6729938 |
_cons     | -9.286923 | 9.125089 | -1.02 | 0.310 | -27.19765 8.659188 |
2. Properties of the regression coefficients and hypothesis testing

A2.9 Explain whether it would be justifiable to use a one-sided test on the slope coefficient in the regression of the rate of growth of employment on the rate of growth of GDP in Exercise 2.20.

A2.10 Explain whether it would be justifiable to use a one-sided test on the slope coefficient in the regression of weight on height in Exercise 1.9.

A2.11 With the information given in Exercise A1.5, how would the change in the measurement of GDP affect:
- the standard error of the coefficient of GDP
- the $F$ statistic for the equation?

A2.12 With the information given in Exercise A1.6, how would the change in the measurement of GDP affect:
- the standard error of the coefficient of GDP
- the $F$ statistic for the equation?

A2.13 [This is a continuation of Exercise 1.16 in the text.] A sample of data consists of $n$ observations on two variables, $Y$ and $X$. The true model is:

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

where $\beta_1$ and $\beta_2$ are parameters and $u$ is a disturbance term that satisfies the usual regression model assumptions. In view of the true model:

$$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \bar{u}$$

where $\bar{Y}$, $\bar{X}$, and $\bar{u}$ are the sample means of $Y$, $X$, and $u$. Subtracting the second equation from the first, one obtains:

$$Y_i^* = \beta_2 X_i^* + u_i^*$$

where $Y_i^* = Y_i - \bar{Y}$, $X_i^* = X_i - \bar{X}$ and $u_i^* = u_i - \bar{u}$. Note that, by construction, the sample means of $Y^*$, $X^*$, and $u^*$ are all equal to zero.

One researcher fits:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X.$$  \hspace{1cm} (1)$$

A second researcher fits:

$$\hat{Y}^* = \hat{\beta}_1^* + \hat{\beta}_2^* X^*.$$  \hspace{1cm} (2)$$

[Note: The second researcher included an intercept in the specification.]

- Comparing regressions (1) and (2), demonstrate that $\hat{Y}_i^* = \hat{Y}_i - \bar{Y}$.
- Demonstrate that the residuals in (2) are identical to the residuals in (1).
- Demonstrate that the OLS estimator of the variance of the disturbance term in (2) is equal to that in (1).
- Explain how the standard error of the slope coefficient in (2) is related to that in (1).
2.4. Additional exercises

- Explain how $R^2$ in (2) is related to $R^2$ in (1).
- Explain why, theoretically, the specification (2) of the second researcher is incorrect and he should have fitted:
  $$\hat{Y}^* = \hat{\beta}_2^* X^*$$
  (3)
  not including a constant in his specification.
- If the second researcher had fitted (3) instead of (2), how would this have affected his estimator of $\beta_2$? Would dropping the unnecessary intercept lead to a gain in efficiency?

A2.14 For the model described in Exercise A1.7, show that $\hat{Y}^*_i = (\hat{Y}_i - \bar{Y})/\hat{\sigma}_Y$, and thus that $\hat{\bar{u}}_i = \hat{u}_i/\hat{\sigma}_Y$, where $\hat{Y}^*_i$ and $\hat{\bar{u}}_i$ are the fitted value of $Y^*_i$ and the residual in the transformed model.

Hence show that:

$$\text{s.e.}(\hat{\beta}_2^*) = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} \times \text{s.e.}(\hat{\beta}_2).$$

Hence find the relationship between the $t$ statistic for $\hat{\beta}_2^*$ and the $t$ statistic for $\hat{\beta}_2$ and the relationship between $R^2$ for the original specification and $R^2$ for the revised specification.

A2.15 A variable $Y_i$ is generated as:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
(1)

where $\beta_1$ and $\beta_2$ are fixed parameters and $u_i$ is a disturbance term that satisfies the regression model assumptions. The values of $X$ are fixed and are as shown in the figure. Four of them, $X_1$ to $X_4$, are close together. The fifth, $X_5$, is much larger.

The corresponding values that $Y$ would take, if there were no disturbance term, are given by the circles on the line. The presence of the disturbance term in the model causes the actual values of $Y$ in a sample to be different. The solid black circles depict a typical sample of observations.
2. Properties of the regression coefficients and hypothesis testing

Discuss the advantages and disadvantages of dropping the observation corresponding to $X_5$ when regressing $Y$ on $X$. If you keep the observation in the sample, will this cause the regression estimates to be biased?

2.5 Answers to the starred exercises in the textbook

2.1 Derive the decomposition of $\hat{\beta}_1$ shown in equation (2.29):

$$\hat{\beta}_1 = \beta_1 + \sum c_i u_i$$

where $c_i = \frac{1}{n} - a_i \bar{X}$ and $a_i$ is defined in equation (2.23).

Answer:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = (\beta_1 + \beta_2 \bar{X} + \bar{u}) - \bar{X} (\beta_2 + \sum a_i u_i)$$

$$= \beta_1 + \frac{1}{n} \sum u_i - \bar{X} \sum a_i u_i$$

$$= \beta_1 + \sum c_i u_i.$$

2.5 An investigator correctly believes that the relationship between two variables $X$ and $Y$ is given by:

$$Y_i = \beta_1 + \beta_2 X_i + u_i.$$

Given a sample of observations on $Y$, $X$, and a third variable $Z$ (which is not a determinant of $Y$), the investigator estimates $\beta_2$ as:

$$\frac{\sum (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}.$$

Demonstrate that this estimator is unbiased.

Answer:

Noting that $Y_i - \bar{Y} = \beta_2 (X_i - \bar{X}) + u_i - \bar{u}$, we have:

$$\hat{\beta}_2 = \frac{\sum (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}$$

$$= \frac{\sum (Z_i - \bar{Z}) \beta_2 (X_i - \bar{X}) + \sum (Z_i - \bar{Z}) (u_i - \bar{u})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}$$

$$= \beta_2 + \frac{\sum (Z_i - \bar{Z}) (u_i - \bar{u})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}.$$
2.5. Answers to the starred exercises in the textbook

Hence:

\[ \begin{align*}
E(\beta_2) &= \beta_2 + \frac{\sum (Z_i - \bar{Z}) E(u_i - \bar{u})}{\sum (Z_i - \bar{Z}) (X_i - \bar{X})} = \beta_2.
\end{align*} \]

2.8 Using the decomposition of \( \hat{\beta}_1 \) obtained in Exercise 2.1, derive the expression for \( \sigma^2_{\hat{\beta}_1} \) given in equation (2.42).

**Answer:**

\[ \begin{align*}
\hat{\beta}_1 &= \beta_1 + \sum c_i u_i, \text{ where } c_i = \frac{1}{n} - a_i \bar{X}, \text{ and } E(\hat{\beta}_1) = \beta_1. \text{ Hence:} \\
\sigma^2_{\hat{\beta}_1} &= E \left( \left( \sum c_i u_i \right)^2 \right) = \sigma^2_u \sum c_i^2 = \sigma^2_u \left( n \frac{1}{n^2} - 2 \frac{\bar{X}}{n} \sum a_i + \bar{X}^2 \sum a_i^2 \right).
\end{align*} \]

From Box 2.2, \( \sum a_i = 0 \) and:

\[ \sum a_i^2 = \frac{1}{\sum (X_i - \bar{X})^2}. \]

Hence:

\[ \sigma^2_{\hat{\beta}_1} = \sigma^2_u \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right). \]

2.9 Given the decomposition in Exercise 2.2 of the OLS estimator of \( \beta_2 \) in the model \( Y_i = \beta_2 X_i + u_i \), demonstrate that the variance of the slope coefficient is given by:

\[ \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2_u}{\sum X_j^2}. \]

**Answer:**

\[ \begin{align*}
\hat{\beta}_2 &= \beta_2 + \sum_{i=1}^n d_i u_i, \text{ where } d_i = X_i / \sum_{j=1}^n X_j^2, \text{ and } E(\hat{\beta}_2) = \beta_2. \text{ Hence:} \\
\sigma^2_{\hat{\beta}_2} &= E \left( \left( \sum_{i=1}^n d_i u_i \right)^2 \right) = \sigma^2_u \sum_{i=1}^n d_i^2 = \sigma^2_u \sum_{i=1}^n \left( \frac{X_i^2}{(\sum_{j=1}^n X_j^2)^2} \right) \\
&= \frac{\sigma^2_u}{\left( \sum_{j=1}^n X_j^2 \right)^2} \sum_{i=1}^n X_i^2 \\
&= \frac{\sigma^2_u}{\sum X_j^2}.
\end{align*} \]
2. Properties of the regression coefficients and hypothesis testing

2.12 It can be shown that the variance of the estimator of the slope coefficient in Exercise 2.5:

\[ \frac{\sum (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})} \]

is given by:

\[ \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2_u}{\sum (X_i - \bar{X})^2} \times \frac{1}{r_{XZ}^2} \]

where \( r_{XZ} \) is the correlation between \( X \) and \( Z \). What are the implications for the efficiency of the estimator?

**Answer:**

If \( Z \) happens to be an exact linear function of \( X \), the population variance will be the same as that of the OLS estimator. Otherwise \( 1/r_{XZ}^2 \) will be greater than 1, the variance will be larger, and so the estimator will be less efficient.

2.15 Suppose that the true relationship between \( Y \) and \( X \) is \( Y_i = \beta_1 + \beta_2 X_i + u_i \) and that the fitted model is \( \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \). In Exercise 1.13 it was shown that if \( X_i^* = \mu_2 X_i \), and \( Y \) is regressed on \( X^* \), the slope coefficient \( \hat{\beta}_2^* = \hat{\beta}_2/\mu_2 \). How will the standard error of \( \hat{\beta}_2^* \) be related to the standard error of \( \hat{\beta}_2 \)?

**Answer:**

In Exercise 1.23 it was demonstrated that the fitted values of \( Y \) would be the same. This means that the residuals are the same, and hence \( \sigma^2_u \), the estimator of the variance of the disturbance term, is the same. The standard error of \( \hat{\beta}_2^* \) is then given by:

\[
\text{s.e.}(\hat{\beta}_2^*) = \sqrt{\frac{\hat{\sigma}^2_u}{\sum (X_i^* - \bar{X})^2}} = \sqrt{\frac{\hat{\sigma}^2_u}{\sum (\mu_2 X_i - \mu_2 \bar{X})^2}} = \sqrt{\frac{\hat{\sigma}^2_u}{\mu_2^2 \sum (X_i - \bar{X})^2}} = \frac{1}{\mu_2} \text{s.e.}(\hat{\beta}_2).
\]

2.17 A researcher with a sample of 50 individuals with similar education, but differing amounts of training, hypothesises that hourly earnings, \( EARNINGS \), may be related to hours of training, \( TRAINING \), according to the relationship:

\[ EARNINGS = \beta_1 + \beta_2 \text{TRAINING} + u. \]
2.5. Answers to the starred exercises in the textbook

He is prepared to test the null hypothesis $H_0: \beta_2 = 0$ against the alternative hypothesis $H_1: \beta_2 \neq 0$ at the 5 per cent and 1 per cent levels. What should he report:

(a) if $\hat{\beta}_2 = 0.30$, s.e.($\hat{\beta}_2$) = 0.12?
(b) if $\hat{\beta}_2 = 0.55$, s.e.($\hat{\beta}_2$) = 0.12?
(c) if $\hat{\beta}_2 = 0.10$, s.e.($\hat{\beta}_2$) = 0.12?
(d) if $\hat{\beta}_2 = -0.27$, s.e.($\hat{\beta}_2$) = 0.12?

Answer:
There are 48 degrees of freedom, and hence the critical values of $t$ at the 5 per cent, 1 per cent, and 0.1 per cent levels are 2.01, 2.68, and 3.51, respectively.

(a) The $t$ statistic is 2.50. Reject $H_0$ at the 5 per cent level but not at the 1 per cent level.
(b) $t = 4.58$. Reject at the 0.1 per cent level.
(c) $t = 0.83$. Fail to reject at the 5 per cent level.
(d) $t = -2.25$. Reject $H_0$ at the 5 per cent level but not at the 1 per cent level.

2.22 Explain whether it would have been possible to perform one-sided tests instead of two-sided tests in Exercise 2.17. If you think that one-sided tests are justified, perform them and state whether the use of a one-sided test makes any difference.

Answer:
First, there should be a discussion of whether the parameter $\beta_2$ in:

$$EARNINGS = \beta_1 + \beta_2TRAINING + u$$

can be assumed not to be negative. The objective of training is to impart skills. It would be illogical for an individual with greater skills to be paid less on that account, and so we can argue that we can rule out $\beta_2 < 0$. We can then perform a one-sided test. With 48 degrees of freedom, the critical values of $t$ at the 5 per cent, 1 per cent, and 0.1 per cent levels are 1.68, 2.40, and 3.26, respectively.

(a) The $t$ statistic is 2.50. We can now reject $H_0$ at the 1 per cent level (but not at the 0.1 per cent level).
(b) $t = 4.58$. Not affected by the change. Reject at the 0.1 per cent level.
(c) $t = 0.83$. Not affected by the change. Fail to reject at the 5 per cent level.
(d) $t = -2.25$. Reject $H_0$ at the 5 per cent level but not at the 1 per cent level.

Here there is a problem because the coefficient has an unexpected sign and is large enough to reject $H_0$ at the 5 per cent level with a two-sided test.

In principle we should ignore this and fail to reject $H_0$. Admittedly, the likelihood of such a large negative $t$ statistic occurring under $H_0$ is very small, but it would be smaller still under the alternative hypothesis $H_1: \beta_2 > 0$.

However, we should consider two further possibilities. One is that the justification for a one-sided test is incorrect. For example, some jobs pay relatively low wages because they offer training that is valued by the employee.
2. Properties of the regression coefficients and hypothesis testing

Apprenticeships are the classic example. Alternatively, workers in some low-paid occupations may, for technical reasons, receive a relatively large amount of training. In either case, the correlation between training and earnings might be negative instead of positive.

Another possible reason for a coefficient having an unexpected sign is that the model is misspecified in some way. For example, the coefficient might be distorted by omitted variable bias, to be discussed in Chapter 6.

2.27 Suppose that the true relationship between \( Y \) and \( X \) is \( Y_i = \beta_1 + \beta_2 X_i + u_i \) and that the fitted model is \( \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \). In Exercise 1.13 it was shown that if \( X_i^* = \mu_2 X_i \), and \( Y \) is regressed on \( X^* \), the slope coefficient \( \hat{\beta}_2^* = \hat{\beta}_2 / \mu_2 \). How will the \( t \) statistic for \( \hat{\beta}_2^* \) be related to the \( t \) statistic for \( \hat{\beta}_2 \)? (See also Exercise 2.15.)

**Answer:**
In Exercise 2.15 it was shown that \( \text{s.e.}(\hat{\beta}_2^*) = \text{s.e.}(\hat{\beta}_2) / \mu_2 \). Hence the \( t \) statistic is unaffected by the transformation.

Alternatively, since we saw in Exercise 1.23 that \( R^2 \) must be the same, it follows that the \( F \) statistic for the equation must be the same. For a simple regression the \( F \) statistic is the square of the \( t \) statistic on the slope coefficient, so the \( t \) statistic must be the same.

2.30 Calculate the 95 per cent confidence interval for \( \beta_2 \) in the price inflation/wage inflation example:

\[
\hat{p} = -1.21 + 0.82w.
\]

(0.05) (0.10)

What can you conclude from this calculation?

**Answer:**
With \( n \) equal to 20, there are 18 degrees of freedom and the critical value of \( t \) at the 5 per cent level is 2.10. The 95 per cent confidence interval is therefore:

\[
0.82 - 0.10 \times 2.10 \leq \beta_2 \leq 0.82 + 0.10 \times 2.10
\]

that is:

\[
0.61 \leq \beta_2 \leq 1.03.
\]

We see that we cannot (quite) reject the null hypothesis \( H_0: \beta_2 = 1 \).

2.36 Suppose that the true relationship between \( Y \) and \( X \) is \( Y_i = \beta_1 + \beta_2 X_i + u_i \) and that the fitted model is \( \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \). Suppose that \( X_i^* = \mu_2 X_i \), and \( Y \) is regressed on \( X^* \). How will the \( F \) statistic for this regression be related to the \( F \) statistic for the original regression? (See also Exercises 1.23, 2.15, and 2.27.)

**Answer:**
We saw in Exercise 1.23 that \( R^2 \) would be the same, and it follows that \( F \) must also be the same.
2.6 Answers to the additional exercises

Note: Each of the exercises below relates to a simple regression. Accordingly, the $F$ test is equivalent to a two-sided $t$ test on the slope coefficient and there is no point in performing both tests. The $F$ statistic is equal to the square of the $t$ statistic and, for any significance level, the critical value of $F$ is equal to the critical value of $t$. Obviously a one-sided $t$ test, when justified, is preferable to either in that it has greater power for any given significance level.

A2.1 We have:

\[
\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i (\beta_1 + \beta_2 X_i + u_i)}{\sum X_i^2} = \frac{\beta_1 \sum X_i}{\sum X_i^2} + \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2}.
\]

Hence:

\[
E(\hat{\beta}_2) = \frac{\beta_1 \sum X_i}{\sum X_i^2} + \beta_2 + E \left( \frac{\sum X_i u_i}{\sum X_i^2} \right) = \frac{\beta_1 \sum X_i}{\sum X_i^2} + \beta_2 + \frac{\sum X_i E(u_i)}{\sum X_i^2}
\]

assuming that $X$ is nonstochastic. Since $E(u_i) = 0$, then:

\[
E(\hat{\beta}_2) = \frac{\beta_1 \sum X_i}{\sum X_i^2} + \beta_2.
\]

Thus $\hat{\beta}_2$ will in general be a biased estimator. The sign of the bias depends on the signs of $\beta_1$ and $\sum X_i$. In general, we have no information about either of these. However, if either $\beta_1 = 0$ or $\bar{X} = 0$ (and so $\sum X_i = 0$), the bias term disappears and $\hat{\beta}_2$ is unbiased after all.

A2.2 First we need to show that $E(\hat{\beta}_2) = 0$.

\[
\hat{\beta}_2 = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X}) (\beta_1 + u_i - \beta_1 - \bar{u})}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X}) (u_i - \bar{u})}{\sum (X_i - \bar{X})^2}.
\]

Hence, given that we are told that $X$ is nonstochastic:

\[
E(\hat{\beta}_2) = E \left( \frac{\sum (X_i - \bar{X}) (u_i - \bar{u})}{\sum (X_i - \bar{X})^2} \right) = \frac{1}{\sum (X_i - \bar{X})^2} E \left( \sum (X_i - \bar{X}) (u_i - \bar{u}) \right)
\]

\[
= \frac{1}{\sum (X_i - \bar{X})^2} \sum (X_i - \bar{X}) E (u_i - \bar{u})
\]

\[
= 0
\]

since $E(u) = 0$. Thus:

\[
E(\hat{\beta}_1) = E (\bar{Y} - \hat{\beta}_2 \bar{X}) = \beta_1 - \bar{X} E(\hat{\beta}_2) = \beta_1
\]

and the estimator is unbiased.
2. Properties of the regression coefficients and hypothesis testing

A2.3 In general, the researcher’s estimator will have a larger variance than \( \hat{Y} \) and therefore will be inefficient. However, if \( \bar{X} = 0 \), the variances are the same. This is because the estimators are then identical. \( \hat{Y} - \beta \bar{X} \) reduces to \( \hat{Y} \).

A2.4 The variance of the estimator is \( \sigma_u^2 / \sum X_i^2 \) whereas that of the estimator:

\[
\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}
\]

is:

\[
\frac{\sigma_u^2}{\sum (X_i - \bar{X})^2} = \frac{\sigma_u^2}{\sum X_i^2 - n\bar{X}^2}.
\]

Thus, provided \( \bar{X} \neq 0 \), \( \sigma_u^2 / \sum X_i^2 \) is more efficient than:

\[
\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}
\]

if \( \beta_1 = 0 \) because it is unbiased and has a smaller variance. It is the OLS estimator in this case.

If \( \bar{X} = 0 \), the estimators are equally efficient because the population variance expressions are identical. The reason for this is that the estimators are now identical:

\[
\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i(Y_i - \bar{Y})}{\sum X_i^2} = \frac{\sum X_iY_i}{\sum X_i^2} - \frac{\bar{Y}}{\sum X_i^2} \sum X_i = \frac{\sum X_iY_i}{\sum X_i^2}
\]

since \( \sum X_i = n\bar{X} = 0 \).

Returning to the general case, if there is little variation in \( X \) in the sample, \( \sum (X_i - \bar{X})^2 \) may be small and hence the population variance of \( \frac{\sum (X_i - \bar{X})(Y_i - Y)}{\sum (X_i - \bar{X})^2} \) may be large. Thus using a criterion such as mean square error, \( \beta_2 \) may be preferable if the bias is small.

A2.5 The \( t \) statistic for the coefficient of \( EXP \) is 58.57, very highly significant. There is little point performing a \( t \) test on the intercept, given that it has no plausible meaning. The \( F \) statistic is 3431.0, very highly significant. Since this is a simple regression model, the two tests are equivalent.

A2.6 The slope coefficient for every category is significantly different from zero at a very high significance level. (The \( F \) test is equivalent to the \( t \) test on the slope coefficient.)
2.6. Answers to the additional exercises

<table>
<thead>
<tr>
<th>EXP</th>
<th>n</th>
<th>( \hat{\beta}_2 )</th>
<th>s.e.(( \hat{\beta}_2 ))</th>
<th>t</th>
<th>( R^2 )</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>0.0235</td>
<td>0.0008</td>
<td>28.86</td>
<td>0.228</td>
<td>832.8</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>0.0316</td>
<td>0.0010</td>
<td>30.99</td>
<td>0.176</td>
<td>960.6</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>0.0409</td>
<td>0.0026</td>
<td>16.02</td>
<td>0.134</td>
<td>256.6</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>0.1202</td>
<td>0.0090</td>
<td>13.30</td>
<td>0.241</td>
<td>177.0</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>0.0131</td>
<td>0.0004</td>
<td>35.70</td>
<td>0.180</td>
<td>1274.8</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>0.0527</td>
<td>0.0010</td>
<td>52.86</td>
<td>0.354</td>
<td>2794.7</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>0.0627</td>
<td>0.0011</td>
<td>58.57</td>
<td>0.351</td>
<td>3431.0</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>0.0558</td>
<td>0.0005</td>
<td>12.78</td>
<td>0.082</td>
<td>163.4</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.0522</td>
<td>0.0070</td>
<td>7.44</td>
<td>0.102</td>
<td>55.3</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>0.0373</td>
<td>0.0008</td>
<td>46.89</td>
<td>0.278</td>
<td>2198.5</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>0.0574</td>
<td>0.0018</td>
<td>31.83</td>
<td>0.174</td>
<td>1013.4</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>0.1976</td>
<td>0.0027</td>
<td>74.16</td>
<td>0.469</td>
<td>5499.9</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>0.0193</td>
<td>0.0016</td>
<td>11.86</td>
<td>0.101</td>
<td>140.7</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.0068</td>
<td>0.0010</td>
<td>6.59</td>
<td>0.059</td>
<td>43.5</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.0329</td>
<td>0.0049</td>
<td>6.72</td>
<td>0.102</td>
<td>45.1</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>0.0069</td>
<td>0.0002</td>
<td>32.15</td>
<td>0.213</td>
<td>1033.4</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>0.0048</td>
<td>0.0003</td>
<td>16.28</td>
<td>0.104</td>
<td>265.1</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>0.0045</td>
<td>0.0007</td>
<td>6.03</td>
<td>0.034</td>
<td>36.4</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>0.0160</td>
<td>0.0003</td>
<td>46.04</td>
<td>0.268</td>
<td>2119.7</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.0040</td>
<td>0.0006</td>
<td>7.32</td>
<td>0.051</td>
<td>53.5</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>0.0165</td>
<td>0.0016</td>
<td>10.56</td>
<td>0.088</td>
<td>111.6</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>0.0145</td>
<td>0.0010</td>
<td>14.34</td>
<td>0.076</td>
<td>205.7</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.0466</td>
<td>0.0043</td>
<td>10.84</td>
<td>0.186</td>
<td>117.5</td>
</tr>
</tbody>
</table>

A2.7 The \( t \) statistic, 13.29, is very highly significant. (The \( F \) test is equivalent.)

A2.8 The \( t \) statistic for height, 3.04, suggests that the effect of height on earnings is highly significant, despite the very low \( R^2 \). In principle the estimate of an extra 41 cents of hourly earnings for every extra inch of height could have been a purely random result of the kind that one obtains with nonsense models. However, the fact that it is apparently highly significant causes us to look for other explanations, the most likely one being that suggested in the answer to Exercise A1.4. Of course, we would not attempt to test the negative constant.

A2.9 One could justify a one-sided test on the slope coefficient in the regression of the rate of growth of employment on the rate of growth of GDP on the grounds that an increase in the rate of growth of GDP is unlikely to cause a decrease in the rate of growth of employment.

A2.10 One could justify a one-sided test on the slope coefficient in the regression of weight on height in Exercise 1.9 on the grounds that an increase in height is unlikely to cause a decrease in weight.
2. Properties of the regression coefficients and hypothesis testing

A2.11 The standard error of the coefficient of GDP. This is given by:

\[
\sqrt{\frac{\hat{\sigma}_u^2}{\sum (G_i^* - \bar{G})^2}}
\]

where \(\hat{\sigma}_u^2\), the estimator of the variance of the disturbance term, is \(\sum \hat{u}_i^2/(n - 2)\). Since \(RSS\) is unchanged, \(\hat{\sigma}_u^2 = \hat{\sigma}_u^2\).

We saw in Exercise A1.6 that \(G_i^* - \bar{G} = G_i - \bar{G}\) for all \(i\). Hence the new standard error is given by:

\[
\sqrt{\frac{\hat{\sigma}_u^2}{\sum (G_i - \bar{G})^2}}
\]

and is unchanged.

\[
F = \frac{ESS}{RSS/(n - 2)}
\]

where:

\[
ESS = \text{explained sum of squares} = \sum (\hat{Y}_i^* - \bar{Y})^2.
\]

Since \(\hat{u}_i^* = \hat{u}_i\), \(\hat{Y}_i^* = \hat{Y}_i\) and \(ESS\) is unchanged. We saw in Exercise A1.6 that \(RSS\) is unchanged. Hence \(F\) is unchanged.

A2.12 The standard error of the coefficient of GDP. This is given by:

\[
\sqrt{\frac{\hat{\sigma}_u^2}{\sum (G_i^* - \bar{G})^2}}
\]

where \(\hat{\sigma}_u^2\), the estimator of the variance of the disturbance term, is \(\sum \hat{u}_i^2/(n - 2)\). We saw in Exercise 1.7 that \(\hat{u}_i^* = \hat{u}_i\) and so \(RSS\) is unchanged. Hence \(\hat{\sigma}_u^2 = \hat{\sigma}_u^2\).

Thus the new standard error is given by:

\[
\sqrt{\frac{\hat{\sigma}_u^2}{\sum (2G_i - 2\bar{G})^2}} = \frac{1}{\sqrt{2}} \sqrt{\frac{\hat{\sigma}_u^2}{\sum (G_i - \bar{G})^2}} = 0.005.
\]

\[
F = \frac{ESS}{(RSS/(n - 2))}
\]

where:

\[
ESS = \text{explained sum of squares} = \sum (\hat{Y}_i^* - \bar{Y})^2.
\]

Since \(\hat{u}_i^* = \hat{u}_i\), \(\hat{Y}_i^* = \hat{Y}_i\) and \(ESS\) is unchanged. Hence \(F\) is unchanged.

A2.13 One way of demonstrating that \(\hat{Y}_i^* = \hat{Y}_i - \bar{Y}\):

\[
\hat{Y}_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* = \hat{\beta}_2^* (X_i - \bar{X})
\]

\[
\hat{Y}_i - \bar{Y} = (\hat{\beta}_1 + \hat{\beta}_2 X_i) - \bar{Y} = (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}).
\]
Demonsration that the residuals are the same:
\[ \hat{u}_i^* = Y_i^* - \hat{Y}_i^* = \left( Y_i - \bar{Y} \right) - \left( \hat{Y}_i - \bar{Y} \right) = \hat{u}_i. \]

Demonstration that the OLS estimator of the variance of the disturbance term in (2) is equal to that in (1):
\[ \hat{\sigma}^2_u = \frac{\sum \hat{u}_i^2}{n - 2} = \frac{\sum \hat{u}_i^2}{n - 2} = \hat{\sigma}^2_u. \]

The standard error of the slope coefficient in (2) is equal to that in (1).
\[ \hat{\sigma}^2_{\hat{\beta}_2} = \frac{\hat{\sigma}^2_u}{\sum (X_i^* - \bar{X})^2} = \frac{\hat{\sigma}^2_u}{\sum X_i^2} = \frac{\hat{\sigma}^2_u}{\sum (X_i - \bar{X})^2} = \hat{\sigma}^2_{\hat{\beta}_2}. \]

Hence the standard errors are the same.

Demonstration that \( R^2 \) in (2) is equal to \( R^2 \) in (1):
\[ R^2 = \frac{\sum (Y_i^* - \bar{Y}^*)^2}{\sum (Y_i^* - \bar{Y})^2} = \frac{\sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = R^2. \]

The reason that specification (2) of the second researcher is incorrect is that the model does not include an intercept.

If the second researcher had fitted (3) instead of (2), this would not in fact have affected his estimator of \( \beta_2 \). Using (3), the researcher should have estimated \( \beta_2 \) as:
\[ \hat{\beta}_2^* = \frac{\sum X_i^*Y_i^*}{\sum X_i^2}. \]

However, Exercise 1.16 demonstrates that, effectively, he has done exactly this. Hence the estimator will be the same. It follows that dropping the unnecessary intercept would not have led to a gain in efficiency.

A2.14 We have:
\[ \hat{Y}_i^* = \hat{\beta}_2^* X_i^* = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} \hat{\beta}_2 \left( \frac{X_i - \bar{X} \hat{\beta}_2}{\hat{\sigma}_X} \right) = \frac{1}{\hat{\sigma}_Y} \hat{\beta}_2 (X_i - \bar{X}) \]

and:
\[ \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i = \bar{Y} + \hat{\beta}_2 (X_i - \bar{X}). \]

Hence:
\[ \hat{Y}_i^* = \frac{1}{\hat{\sigma}_Y} (\hat{Y}_i - \bar{Y}). \]
2. Properties of the regression coefficients and hypothesis testing

Also:

\[ \hat{u}_i^* = Y_i^* - \hat{Y}_i^* = \frac{1}{\hat{\sigma}_Y} (Y_i - \bar{Y}) - \frac{1}{\hat{\sigma}_Y} (\hat{Y}_i - \bar{Y}) = \frac{1}{\hat{\sigma}_Y} (Y_i - \hat{Y}_i) = \frac{1}{\hat{\sigma}_Y} \hat{u}_i \]

and:

\[ \text{s.e.}(\hat{\beta}_2^*) = \sqrt{\frac{1}{n-2} \sum \hat{u}_i^2} = \sqrt{\frac{1}{n-2} \sum \hat{u}_i^2} = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} \times \text{s.e.}(\hat{\beta}_2). \]

Given the expressions for \( \hat{\beta}_2^* \) and s.e.(\( \hat{\beta}_2^* \)), the \( t \) statistic for \( \hat{\beta}_2^* \) is the same as that for \( \hat{\beta}_2 \). Hence the \( F \) statistic will be the same and \( R^2 \) will be the same.

A2.15 The inclusion of the fifth observation does not cause the model to be misspecified or the regression model assumptions to be violated, so retaining it in the sample will not give rise to biased estimates. There would be no advantages in dropping it and there would be one major disadvantage. \( \sum (X_i - \bar{X})^2 \) would be greatly reduced and hence the variances of the coefficients would be increased, adversely affecting the precision of the estimates.

This said, in practice one would wish to check whether it is sensible to assume that the model relating \( Y \) to \( X \) for the other observations really does apply to the observation corresponding to \( X_5 \) as well. This question can be answered only by being familiar with the context and having some intuitive understanding of the relationship between \( Y \) and \( X \).
3.1 Overview

This chapter introduces regression models with more than one explanatory variable. Specific topics are treated with reference to a model with just two explanatory variables, but most of the concepts and results apply straightforwardly to more general models. The chapter begins by showing how the least squares principle is employed to derive the expressions for the regression coefficients and how the coefficients should be interpreted. It continues with a discussion of the precision of the regression coefficients and tests of hypotheses relating to them. Next comes multicollinearity, the problem of discriminating between the effects of individual explanatory variables when they are closely related. The chapter concludes with a discussion of $F$ tests of the joint explanatory power of the explanatory variables or subsets of them, and shows how a $t$ test can be thought of as a marginal $F$ test.

3.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to explain what is meant by:

- the principles behind the derivation of multiple regression coefficients (but you are not expected to learn the expressions for them or to be able to reproduce the mathematical proofs)
- how to interpret the regression coefficients
- the Frisch–Waugh–Lovell graphical representation of the relationship between the dependent variable and one explanatory variable, controlling for the influence of the other explanatory variables
- the properties of the multiple regression coefficients
- what factors determine the population variance of the regression coefficients
- what is meant by multicollinearity
- what measures may be appropriate for alleviating multicollinearity
- what is meant by a linear restriction
- the $F$ test of the joint explanatory power of the explanatory variables
3. Multiple regression analysis

- the $F$ test of the explanatory power of a group of explanatory variables
- why $t$ tests on the slope coefficients are equivalent to marginal $F$ tests.

You should know the expression for the population variance of a slope coefficient in a multiple regression model with two explanatory variables.

### 3.3 Additional exercises

A3.1 The output shows the result of regressing $FDHO$, expenditure on food consumed at home, on $EXP$, total household expenditure, and $SIZE$, number of persons in the household, using the $CES$ data set. Provide an interpretation of the regression coefficients and perform appropriate tests.

```plaintext
.reg FDHO EXP SIZE if FDHO>0
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs =</th>
<th>6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.1521e+09</td>
<td>2</td>
<td>576056293</td>
<td>$F(2, 6331) = 2257.59$</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>1.6154e+09</td>
<td>6331</td>
<td>255164.645</td>
<td>Prob &gt; $F$ = 0.0000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2.7676e+09</td>
<td>6333</td>
<td>437006.15</td>
<td>Adj R-squared = 0.4163</td>
<td></td>
</tr>
</tbody>
</table>

| FDHO | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|------|-------|-----------|---|------|----------------------|
| EXP  | .056366 | .0010435 | 54.02 | 0.000 | .0543204 .0584116 |
| SIZE | 115.1636 | 4.341912 | 26.52 | 0.000 | 106.652 123.6752 |
| _cons | 130.5997 | 13.53959 | 9.65 | 0.000 | 104.0575 157.1419 |

A3.2 Perform a regression parallel to that in Exercise A3.1 for your $CES$ category of expenditure, provide an interpretation of the regression coefficients and perform appropriate tests. Delete observations where expenditure on your category is zero.

A3.3 The output shows the result of regressing $FDHOPC$, expenditure on food consumed at home per capita, on $EXPPC$, total household expenditure per capita, and $SIZE$, number of persons in the household, using the $CES$ data set. Provide an interpretation of the regression coefficients and perform appropriate tests.

```plaintext
.reg FDHOPC EXPPC SIZE if FDHO>0
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs =</th>
<th>6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>202590496</td>
<td>2</td>
<td>101295248</td>
<td>$F(2, 6331) = 1572.95$</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>407705728</td>
<td>6331</td>
<td>64398.3143</td>
<td>R-squared = 0.3320</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>610296223</td>
<td>6333</td>
<td>96367.6336</td>
<td>Adj R-squared = 0.3317</td>
<td></td>
</tr>
</tbody>
</table>

Source: Christopher Dougherty, 2016. All rights reserved. Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
### Additional exercises

| FDHOPC | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|-------|---------------------|
| EXPPC  | 0.0480294 | 0.0010064 | 47.72 | 0.000 | 0.0460564 - 0.0500023 |
| SIZE   | -26.45917  | 2.253999 | -11.74 | 0.000 | -30.87777 - -22.04057  |
| _cons  | 283.2498  | 8.412603 | 33.67 | 0.000 | 266.7582 - 299.7413 |

**Exercise A3.4**

Perform a regression parallel to that in Exercise A3.3 for your CES category of expenditure. Provide an interpretation of the regression coefficients and perform appropriate tests.

**Exercise A3.5**

The output shows the result of regressing FDHOPC, expenditure on food consumed at home per capita, on EXPPC, total household expenditure per capita, and SIZEAM, SIZEAF, SIZEJM, SIZEJF, and SIZEIN, numbers of adult males, adult females, junior males, junior females, and infants, respectively, in the household, using the CES data set. Provide an interpretation of the regression coefficients and perform appropriate tests.

```stata
reg FDHOPC EXPPC SIZEAM SIZEAF SIZEJM SIZEJF SIZEIN if FDHO>0
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>202746894</td>
<td>6</td>
<td>33791149</td>
<td>F( 6, 6327) = 524.59</td>
</tr>
<tr>
<td>Residual</td>
<td>407549329</td>
<td>6327</td>
<td>64414.3084</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>610296223</td>
<td>6333</td>
<td>96367.6336</td>
<td>R-squared = 0.3322</td>
</tr>
</tbody>
</table>

| FDHOPC | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|-------|---------------------|
| EXPPC  | 0.0479717  | 0.0010087 | 47.56 | 0.000 | 0.0459943 - 0.0499491  |
| SIZEAM | -25.77747  | 4.757056  | -5.42 | 0.000 | -35.10291 - -16.45203  |
| SIZEAF | -32.38649  | 5.065782  | -6.39 | 0.000 | -42.31714 - -22.45584  |
| SIZEJM | -20.24693  | 5.731645  | -3.53 | 0.000 | -31.48290 - -9.01967    |
| SIZEJF | -26.66374  | 6.122262  | -4.36 | 0.000 | -38.66544 - -14.66203   |
| SIZEIN | -28.60470  | 11.75666  | -2.43 | 0.015 | -51.65174 - -5.557656    |
| _cons  | 287.5695   | 9.280372  | 30.99 | 0.000 | 269.37690 - 305.76221   |

**Exercise A3.6**

Perform a regression parallel to that in Exercise A3.5 for your CES category of expenditure. Provide an interpretation of the regression coefficients and perform appropriate tests.

**Exercise A3.7**

A researcher hypothesises that, for a typical enterprise, $V$, the logarithm of value added per worker, is related to $K$, the logarithm of capital per worker, and $S$, the logarithm of the average years of schooling of the workers, the relationship being:

$$V = \beta_1 + \beta_2 K + \beta_3 S + u$$

where $u$ is a disturbance term that satisfies the usual regression model assumptions. She fits the relationship (1) for a sample of 25 manufacturing enterprises, and (2) for a sample of 100 services enterprises. The table provides some data on the samples.
3. Multiple regression analysis

<table>
<thead>
<tr>
<th></th>
<th>(1) Manufacturing sample</th>
<th>(2) Services sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of enterprises</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Estimate of variance of $u$</td>
<td>0.16</td>
<td>0.64</td>
</tr>
<tr>
<td>Mean square deviation of $K$</td>
<td>4.00</td>
<td>16.00</td>
</tr>
<tr>
<td>Correlation between $K$ and $S$</td>
<td>0.60</td>
<td>0.60</td>
</tr>
</tbody>
</table>

The mean square deviation of $K$ is defined as $\frac{1}{n} \sum (K_i - \bar{K})^2$, where $n$ is the number of enterprises in the sample and $\bar{K}$ is the average value of $K$ in the sample.

The researcher finds that the standard error of the coefficient of $K$ is 0.050 for the manufacturing sample and 0.025 for the services sample. Explain the difference quantitatively, given the data in the table.

A3.8 A researcher is fitting earnings functions using a sample of data relating to individuals born in the same week in 1958. He decides to relate $Y$, gross hourly earnings in 2001, to $S$, years of schooling, and $PWE$, potential work experience, using the semilogarithmic specification:

$$\log Y = \beta_1 + \beta_2 S + \beta_3 PWE + u$$

where $u$ is a disturbance term assumed to satisfy the regression model assumptions. $PWE$ is defined as age – years of schooling – 5. Since the respondents were all aged 43 in 2001, this becomes:

$$PWE = 43 - S - 5 = 38 - S.$$ 

The researcher finds that it is impossible to fit the model as specified. Stata output for his regression is reproduced below:

```
. reg LGY S PWE

Source | SS       | df | MS        | Number of obs =  5660
-------+---------+----+-----------+------------------
Model  | 237.170265 | 1  | 237.170265 | F( 1, 5658) = 1232.62
Residual | 1088.66373 | 5658  | .192411405 | Prob > F = 0.0000
Total  | 1325.834  | 5659  | .234287682 | R-squared = 0.1789

------------------------------------------------------------------------------
LGY | Coef. Std. Err. t P>|t| [95% Conf. Interval]
--------+-----------------------------+-----------------------------+-----------------------------+-----------------------------+-----------------------------+
S | .1038011 .0029566 35.11 0.000 .0980051 .1095971
PWE | (dropped)                    |
_cons | .5000033 .0373785 13.38 0.000 .4267271 .5732795
------------------------------------------------------------------------------
```

Explain why the researcher was unable to fit his specification. Explain how the coefficient of $S$ might be interpreted.
3.4 Answers to the starred exercises in the textbook

3.5 Explain why the intercept in the regression of EARN on ES is equal to zero.

**Answer:**
The intercept is calculated as $\bar{EARN} - \hat{\beta}_2 \bar{ES}$. However, since the mean of the residuals from an OLS regression is zero, both $EARN$ and $ES$ are zero, and hence the intercept is zero.

3.6 Show that, in the general case, the mean of the residuals from a fitted OLS multiple regression is equal to zero, provided that an intercept is included in the specification. Note: This is an extension of one of the useful results in Section 1.5.

**Answer:**
If the model is:

$$Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k.$$  

For observation $i$ we have:

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}.$$

Hence:

$$\bar{\hat{u}} = \bar{Y} - \hat{\beta}_1 - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k$$

$$= \bar{Y} - \left[\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k\right] - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k = 0.$$

3.16 A researcher investigating the determinants of the demand for public transport in a certain city has the following data for 100 residents for the previous calendar year: expenditure on public transport, $E$, measured in dollars; number of days worked, $W$; and number of days not worked, $NW$. By definition $NW$ is equal to $365 - W$. He attempts to fit the following model:

$$E = \beta_1 + \beta_2 W + \beta_3 NW + u.$$

Explain why he is unable to fit this equation. (Give both intuitive and technical explanations.) How might he resolve the problem?

**Answer:**
There is exact multicollinearity since there is an exact linear relationship between $W$, $NW$ and the constant term. As a consequence it is not possible to tell whether variations in $E$ are attributable to variations in $W$ or variations in $NW$, or both. Noting that $NW_i - \bar{NW} = -(W_i - \bar{W})$, we have:

$$\hat{\beta}_2 = \frac{\sum (E_i - \bar{E})(W_i - \bar{W}) \sum (NW_i - \bar{NW})^2 - \sum (E_i - \bar{E})(NW_i - \bar{NW}) \sum (W_i - \bar{W})(NW_i - \bar{NW}) \sum (W_i - \bar{W}) (NW_i - \bar{NW})}{\sum (W_i - \bar{W})^2 \sum (NW_i - \bar{NW})^2 - \left(\sum (W_i - \bar{W})(NW_i - \bar{NW})\right)^2}$$

$$= \frac{\sum (E_i - \bar{E})(W_i - \bar{W}) \sum (W_i - \bar{W})^2 - \sum (E_i - \bar{E})(W_i - \bar{W}) \sum (W_i - \bar{W})(-W_i + \bar{W}) \sum (W_i - \bar{W})(-W_i + \bar{W})}{\sum (W_i - \bar{W})^2 \sum (W_i - \bar{W})^2 - \left(\sum (W_i - \bar{W})(-W_i + \bar{W})\right)^2}$$

$$= \frac{0}{0}.$$
3. Multiple regression analysis

One way of dealing with the problem would be to drop NW from the regression. The interpretation of $\hat{\beta}_2$ now is that it is an estimate of the extra expenditure on transport per day worked, compared with expenditure per day not worked.

3.21 The researcher in Exercise 3.16 decides to divide the number of days not worked into the number of days not worked because of illness, I, and the number of days not worked for other reasons, O. The mean value of I in the sample is 2.1 and the mean value of O is 120.2. He fits the regression (standard errors in parentheses):

$$\hat{E} = -9.6 + 2.10W + 0.45O \quad R^2 = 0.72$$

$$\begin{pmatrix} (8.3) & (1.98) & (1.77) \end{pmatrix}$$

Perform $t$ tests on the regression coefficients and an $F$ test on the goodness of fit of the equation. Explain why the $t$ tests and $F$ test have different outcomes.

**Answer:**

Although there is not an exact linear relationship between W and O, they must have a very high negative correlation because the mean value of I is so small. Hence one would expect the regression to be subject to multicollinearity, and this is confirmed by the results. The $t$ statistics for the coefficients of W and O are only 1.06 and 0.25, respectively, but the $F$ statistic:

$$F(2, 97) = \frac{0.72/2}{(1 - 0.72)/97} = 124.7$$

is greater than the critical value of $F$ at the 0.1 per cent level, 7.41.

3.5 **Answers to the additional exercises**

A3.1 The regression indicates that 5.6 cents out of the marginal expenditure dollar is spent on food consumed at home, and that expenditure on this category increases by $115 for each individual in the household, keeping total expenditure constant. Both of these effects are very highly significant. Just over 40 per cent of the variance in $FDHO$ is explained by $EXP$ and $SIZE$. The intercept has no plausible interpretation.

A3.2 With the exception of $LOCT$, all of the categories have positive coefficients for $EXP$, with high significance levels, but the $SIZE$ effect varies:

- Positive, significant at the 1 per cent level: $FDHO, TELE, CLOT, FOOT, GASO$.
- Positive, significant at the 5 per cent level: $LOCT$.
- Negative, significant at the 1 per cent level: $TEXT, FEES, READ$.
- Negative, significant at the 5 per cent level: $SHEL, EDUC$.
- Not significant: $FDAW, DOM, FURN, MAPP, SAPP, TRIP, HEAL, ENT, TOYS, TOB$.

At first sight it may seem surprising that $SIZE$ has a significant negative effect for some categories. The reason for this is that an increase in $SIZE$ means a reduction.
in expenditure per capita, if total household expenditure is kept constant, and thus \( SIZE \) has a (negative) income effect in addition to any direct effect. Effectively poorer, the larger household has to spend more on basics and less on luxuries. To determine the true direct effect, we need to eliminate the income effect, and that is the point of the re-specification of the model in the next exercise.

<table>
<thead>
<tr>
<th>( \hat{\beta}_2 )</th>
<th>s.e.(( \hat{\beta}_2 ))</th>
<th>( \hat{\beta}_3 )</th>
<th>s.e.(( \hat{\beta}_3 ))</th>
<th>( R^2 )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>0.238</td>
<td>0.0008</td>
<td>-8.09</td>
<td>4.19</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>0.0309</td>
<td>0.0010</td>
<td>16.39</td>
<td>4.50</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>0.0388</td>
<td>0.0026</td>
<td>52.34</td>
<td>14.06</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>0.1252</td>
<td>0.0090</td>
<td>-179.23</td>
<td>48.92</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>0.0121</td>
<td>0.0004</td>
<td>18.92</td>
<td>1.57</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>0.0538</td>
<td>0.0010</td>
<td>-20.72</td>
<td>4.47</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>0.0564</td>
<td>0.0010</td>
<td>115.16</td>
<td>4.34</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>0.0556</td>
<td>0.0005</td>
<td>3.24</td>
<td>2.05</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.0541</td>
<td>0.0071</td>
<td>-61.87</td>
<td>35.92</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>0.0347</td>
<td>0.0008</td>
<td>50.29</td>
<td>3.40</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>0.0580</td>
<td>0.0019</td>
<td>-9.96</td>
<td>8.60</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>0.1997</td>
<td>0.0027</td>
<td>-38.78</td>
<td>11.41</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>0.0198</td>
<td>0.0017</td>
<td>-9.01</td>
<td>8.99</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.0062</td>
<td>0.0011</td>
<td>14.61</td>
<td>4.72</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.0309</td>
<td>0.0050</td>
<td>44.48</td>
<td>23.94</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>0.0070</td>
<td>0.0002</td>
<td>-2.17</td>
<td>1.03</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>0.0049</td>
<td>0.0003</td>
<td>-1.06</td>
<td>1.58</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>0.0046</td>
<td>0.0008</td>
<td>-3.12</td>
<td>3.99</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>0.0150</td>
<td>0.0004</td>
<td>17.92</td>
<td>1.47</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.0041</td>
<td>0.0006</td>
<td>-0.71</td>
<td>2.90</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>0.0161</td>
<td>0.0016</td>
<td>6.79</td>
<td>6.24</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>0.0140</td>
<td>0.0010</td>
<td>12.19</td>
<td>4.88</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.0450</td>
<td>0.0045</td>
<td>37.48</td>
<td>31.21</td>
</tr>
</tbody>
</table>

A3.3 Another surprise, perhaps. The purpose of this specification is to test whether household size has an effect on expenditure per capita on food consumed at home, controlling for the income effect of variations in household size mentioned in the answer to Exercise A3.2. Expenditure per capita on food consumed at home increases by 4.8 cents out of the marginal dollar of total household expenditure per capita. Now \( SIZE \) has a very significant negative effect. Expenditure per capita on \( FDHO \) decreases by $26 per year for each extra person in the household, suggesting that larger households are more efficient than smaller ones with regard to expenditure on this category, the effect being highly significant. \( R^2 \) is lower than in Exercise A3.1, but a comparison is invalidated by the fact that the dependent variable is different.

A3.4 Nearly all of the categories have negative \( SIZE \) effects, the majority highly significant. One explanation of the negative effects could be economies of scale, but
3. Multiple regression analysis

this is not plausible in the case of some. Another might be family composition – larger families having more children. In the case of DOM, SIZE has a positive effect, significant at the 5 per cent level. Again, this might be attributable to larger families having more children and needing greater expenditure on childcare.

<table>
<thead>
<tr>
<th>EXP</th>
<th>SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>$\hat{\beta}_2$</td>
</tr>
<tr>
<td>ADM</td>
<td>2,815</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
</tr>
</tbody>
</table>

A3.5 The coefficients of the SIZE variables are fairly similar, suggesting that household composition is not important for this category of expenditure.

A3.6 The regression results for this specification are summarised in the table below. In the case of SHEL, the regression indicates that the SIZE effect is attributable to SIZEAM. To investigate this further, the regression was repeated: (1) restricting the sample to households with at least one adult male, and (2) restricting the sample to households with either no adult male or just 1 adult male. The first regression produces a negative effect for SIZEAM, but it is smaller than with the whole sample and not significant. In the second regression the coefficient of SIZEAM jumps dramatically, from $−424 to $−793, suggesting very strong economies of scale for this particular comparison.

As might be expected, the SIZE composition variables on the whole do not appear to have significant effects if the SIZE variable does not in Exercise A3.4. The
results for *TOB* are puzzling, in that the apparent economies of scale do not appear to be related to household composition.

<table>
<thead>
<tr>
<th>Category</th>
<th>ADM</th>
<th>CLOT</th>
<th>DOM</th>
<th>EDUC</th>
<th>ELEC</th>
<th>FDAW</th>
<th>FDHO</th>
<th>FOOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>0.0245</td>
<td>0.0309</td>
<td>0.0422</td>
<td>0.1191</td>
<td>0.0120</td>
<td>0.0531</td>
<td>0.0561</td>
<td>0.0056</td>
</tr>
<tr>
<td>SIZEAM</td>
<td>-37.17</td>
<td>12.84</td>
<td>-141.47</td>
<td>120.11</td>
<td>23.40</td>
<td>29.36</td>
<td>129.69</td>
<td>2.65</td>
</tr>
<tr>
<td>SIZEJM</td>
<td>1.33</td>
<td>17.11</td>
<td>114.68</td>
<td>-413.28</td>
<td>12.53</td>
<td>-24.45</td>
<td>126.94</td>
<td>1.23</td>
</tr>
<tr>
<td>SIZEJF</td>
<td>48.55</td>
<td>29.98</td>
<td>93.82</td>
<td>-287.35</td>
<td>8.93</td>
<td>-26.03</td>
<td>126.94</td>
<td>6.32</td>
</tr>
<tr>
<td>SIZEIN</td>
<td>-34.51</td>
<td>2.08</td>
<td>441.46</td>
<td>-123.20</td>
<td>4.05</td>
<td>-61.38</td>
<td>95.90</td>
<td>2.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>FURN</th>
<th>GASO</th>
<th>HEAL</th>
<th>HOUS</th>
<th>LIFE</th>
<th>LOCT</th>
<th>MAPP</th>
<th>PERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>0.0547</td>
<td>0.0341</td>
<td>0.0579</td>
<td>0.2022</td>
<td>0.0195</td>
<td>0.0061</td>
<td>0.0321</td>
<td>0.0071</td>
</tr>
<tr>
<td>SIZEAM</td>
<td>-119.30</td>
<td>90.70</td>
<td>3.01</td>
<td>-175.23</td>
<td>10.54</td>
<td>12.02</td>
<td>-13.99</td>
<td>-13.99</td>
</tr>
<tr>
<td>SIZEAF</td>
<td>-55.42</td>
<td>52.23</td>
<td>89.64</td>
<td>-111.39</td>
<td>25.43</td>
<td>19.16</td>
<td>-119.30</td>
<td>12.33</td>
</tr>
<tr>
<td>SIZEJM</td>
<td>-27.44</td>
<td>30.83</td>
<td>-62.83</td>
<td>52.32</td>
<td>-23.28</td>
<td>-6.41</td>
<td>131.15</td>
<td>-3.33</td>
</tr>
<tr>
<td>SIZEJF</td>
<td>-15.06</td>
<td>46.24</td>
<td>-57.94</td>
<td>35.6</td>
<td>249.2</td>
<td>480.2</td>
<td>753.6</td>
<td>28.5</td>
</tr>
</tbody>
</table>

| R²        | 0.110 | 0.310 | 0.181 | 0.475 | 0.109 | 0.077 | 0.116 | 0.228 |
| F         | 9.9 | 427.6 | 177.0 | 937.6 | 25.3 | 9.6 | 8.6 | 187.4 |
| n         | 487 | 5,710 | 4,802 | 6,223 | 1,253 | 692 | 399 | 3,817 |

<table>
<thead>
<tr>
<th>Category</th>
<th>READ</th>
<th>SAPP</th>
<th>TELE</th>
<th>TEXT</th>
<th>TOB</th>
<th>TOYS</th>
<th>TRIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>0.0499</td>
<td>0.0464</td>
<td>0.048</td>
<td>0.0400</td>
<td>0.0151</td>
<td>0.0148</td>
<td>0.0448</td>
</tr>
<tr>
<td>SIZEAM</td>
<td>-6.37</td>
<td>-1.64</td>
<td>29.33</td>
<td>7.42</td>
<td>30.92</td>
<td>-39.66</td>
<td>64.35</td>
</tr>
<tr>
<td>SIZEAF</td>
<td>1.69</td>
<td>8.95</td>
<td>35.59</td>
<td>2.58</td>
<td>22.09</td>
<td>1.30</td>
<td>4.87</td>
</tr>
<tr>
<td>SIZEJM</td>
<td>0.63</td>
<td>-13.21</td>
<td>6.38</td>
<td>-15.90</td>
<td>17.42</td>
<td>42.46</td>
<td>81.61</td>
</tr>
<tr>
<td>SIZEJF</td>
<td>4.73</td>
<td>1.17</td>
<td>12.74</td>
<td>-4.92</td>
<td>-45.12</td>
<td>19.34</td>
<td>102.45</td>
</tr>
<tr>
<td>R²</td>
<td>0.108</td>
<td>0.038</td>
<td>0.296</td>
<td>0.059</td>
<td>0.100</td>
<td>0.090</td>
<td>0.197</td>
</tr>
<tr>
<td>F</td>
<td>45.8</td>
<td>6.7</td>
<td>404.9</td>
<td>10.4</td>
<td>21.2</td>
<td>41.2</td>
<td>20.8</td>
</tr>
<tr>
<td>n</td>
<td>2,287</td>
<td>1,037</td>
<td>5,788</td>
<td>992</td>
<td>368</td>
<td>2,504</td>
<td>516</td>
</tr>
</tbody>
</table>
3. Multiple regression analysis

A3.7 The standard error is given by:

\[
s.e.(\hat{\beta}_2) = \hat{\sigma}_u \times \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{\text{MSD}(K)}} \times \frac{1}{\sqrt{1 - r_{K,S}^2}}.
\]

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>manufacturing sample</td>
<td>manufacturing sample</td>
</tr>
<tr>
<td>Number of enterprises</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Estimate of variance of u</td>
<td>0.16</td>
<td>0.64</td>
</tr>
<tr>
<td>Mean square deviation of K</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Correlation between K and S</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Standard errors</td>
<td></td>
<td>0.050</td>
</tr>
</tbody>
</table>

The table shows the four factors for the two sectors. Other things being equal, the larger number of enterprises and the greater MSD of \(K\) would separately cause the standard error of \(\hat{\beta}_2\) for the services sample to be half that in the manufacturing sample. However, the larger estimate of the variance of \(u\) would, taken in isolation, cause it to be double. The net effect, therefore, is that it is half.

A3.8 Exact multicollinearity. An extra year of schooling implies one fewer year of potential work experience. Thus the coefficient of schooling estimates the proportional increase in earnings associated with an additional year of schooling, taking account of the loss of a year of potential work experience.
Chapter 4
Transformations of variables

4.1 Overview

This chapter shows how least squares regression analysis can be extended to fit nonlinear models. Sometimes an apparently nonlinear model can be linearised by taking logarithms. $Y = \beta_1 X^{\beta_2}$ and $Y = \beta_1 e^{\beta_2 X}$ are examples. Because they can be fitted using linear regression analysis, they have proved very popular in the literature, there usually being little to be gained from using more sophisticated specifications. If you plot earnings on schooling, using the EAWE data set, or expenditure on a given category of expenditure on total household expenditure, using the CES data set, you will see that there is so much randomness in the data that one nonlinear specification is likely to be just as good as another, and indeed a linear specification may not be obviously inferior. Often the real reason for preferring a nonlinear specification to a linear one is that it makes more sense theoretically. The chapter shows how the least squares principle can be applied when the model cannot be linearised.

4.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain the difference between nonlinearity in parameters and nonlinearity in variables
- explain why nonlinearity in parameters is potentially a problem while nonlinearity in variables is not
- define an elasticity
- explain how to interpret an elasticity in simple terms
- perform basic manipulations with logarithms
- interpret the coefficients of semi-logarithmic and logarithmic regressions
- explain why the coefficients of semi-logarithmic and logarithmic regressions should not be interpreted using the method for regressions in natural units described in Chapter 1
- perform a RESET test of functional misspecification
4. Transformations of variables

- explain the role of the disturbance term in a nonlinear model
- explain how in principle a nonlinear model that cannot be linearised may be fitted
- perform a transformation for comparing the fits of models with linear and logarithmic dependent variables.

4.3 Further material

Box–Cox tests of functional specification

The theory behind the procedure for discriminating between a linear and a logarithmic specification of the dependent variable is explained in the Appendix to Chapter 10 of the text. However, the exposition there is fairly brief. An expanded version is offered here. It should be skipped on first reading because it makes use of material on maximum likelihood estimation. To keep the mathematics uncluttered, the theory will be described in the context of the simple regression model, where we are choosing between:

\[ Y = \beta_1 + \beta_2 X + u \]

and:

\[ \log Y = \beta_1 + \beta_2 X + u. \]

It generalises with no substantive changes to the multiple regression model.

The two models are actually special cases of the more general model:

\[ Y_\lambda = \frac{Y^\lambda - 1}{\lambda} = \beta_1 + \beta_2 X + u \]

with \( \lambda = 1 \) yielding the linear model (with an unimportant adjustment to the intercept) and \( \lambda = 0 \) yielding the logarithmic specification at the limit as \( \lambda \) tends to zero.

Assuming that \( u \) is iid (independently and identically distributed) \( N(0, \sigma^2) \), the density function for \( u_i \) is:

\[ f(u_i) = \frac{1}{\sigma \sqrt{2\pi}} e^{-u_i^2 / 2\sigma^2} \]

and hence the density function for \( Y_{\lambda i} \) is:

\[ f(Y_{\lambda i}) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2 / 2\sigma^2}. \]

From this we obtain the density function for \( Y_i \):

\[ f(Y_i) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(Y_i - \beta_1 - \beta_2 X_i)^2 / 2\sigma^2} \left| \frac{\partial Y_{\lambda i}}{\partial Y_i} \right| = \frac{1}{\sigma \sqrt{2\pi}} e^{-(Y_i - \beta_1 - \beta_2 X_i)^2 / 2\sigma^2} Y_i^{\lambda - 1}. \]

The factor \( \left| \frac{\partial Y_{\lambda i}}{\partial Y_i} \right| \) is the Jacobian for relating the density function of \( Y_{\lambda i} \) to that of \( Y_i \).

Hence the likelihood function for the parameters is:

\[ L(\beta_1, \beta_2, \sigma, \lambda) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n \prod_{i=1}^n e^{-(Y_i - \beta_1 - \beta_2 X_i)^2 / 2\sigma^2} \prod_{i=1}^n Y_i^{\lambda - 1}. \]
4.3. Further material

and the log-likelihood is:

$$\log L(\beta_1, \beta_2, \sigma, \lambda) = -\frac{n}{2} \log 2\pi \sigma^2 - \sum_{i=1}^{n} \frac{1}{2\sigma^2}(Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2 + \sum_{i=1}^{n} \log Y_i^{\lambda - 1}$$

$$= -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2 + (\lambda - 1) \sum_{i=1}^{n} \log Y_i.$$ 

From the first-order condition $\partial \log L / \partial \sigma = 0$, we have:

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2 = 0$$

giving:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2.$$ 

Substituting into the log-likelihood function, we obtain the concentrated log-likelihood:

$$\log L(\beta_1, \beta_2, \lambda) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \frac{1}{n} \sum_{i=1}^{n} (Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2 - \frac{n}{2} + (\lambda - 1) \sum_{i=1}^{n} \log Y_i.$$ 

The expression can be simplified (Zarembka, 1968) by working with $Y_i^*$ rather than $Y_i$, where $Y_i^*$ is $Y_i$ divided by $Y_{GM}$, the geometric mean of the $Y_i$ in the sample, for:

$$\sum_{i=1}^{n} \log Y_i^* = \sum_{i=1}^{n} \log(Y_i/Y_{GM}) = \sum_{i=1}^{n} (\log Y_i - \log Y_{GM})$$

$$= \sum_{i=1}^{n} \log Y_i - n \log Y_{GM} = \sum_{i=1}^{n} \log Y_i - n \log \left(\prod_{i=1}^{n} Y_i\right)^{1/n}$$

$$= \sum_{i=1}^{n} \log Y_i - \log \left(\prod_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \log Y_i - \sum_{i=1}^{n} \log Y_i = 0.$$ 

With this simplification, the log-likelihood is:

$$\log L(\beta_1, \beta_2, \lambda) = -\frac{n}{2} \left( \log 2\pi + \log \frac{1}{n} + 1 \right) - \frac{n}{2} \log \sum_{i=1}^{n} (Y_{\lambda i}^* - \beta_1 - \beta_2 X_i)^2$$

and it will be maximised when $\beta_1$, $\beta_2$ and $\lambda$ are chosen so as to minimise

$$\sum_{i=1}^{n} (Y_{\lambda i} - \beta_1 - \beta_2 X_i)^2$$, the residual sum of squares from a least squares regression of the scaled, transformed $Y$ on $X$. One simple procedure is to perform a grid search, scaling and transforming the data on $Y$ for a range of values of $\lambda$ and choosing the value that leads to the smallest residual sum of squares (Spitzer, 1982).

A null hypothesis $\lambda = \lambda_0$ can be tested using a likelihood ratio test in the usual way. Under the null hypothesis, the test statistic $2(\log L_\lambda - \log L_0)$ will have a chi-squared distribution with one degree of freedom, where $\log L_\lambda$ is the unconstrained log-likelihood and $L_0$ is the constrained one. Note that, in view of the preceding equation:

$$2(\log L_\lambda - \log L_0) = n(\log RSS_0 - \log RSS_\lambda)$$

71
4. Transformations of variables

where $RSS_0$ and $RSS_\lambda$ are the residual sums of squares from the constrained and unconstrained regressions with $Y^*$.

The most obvious tests are $\lambda = 0$ for the logarithmic specification and $\lambda = 1$ for the linear one. Note that it is not possible to test the two hypotheses directly against each other. As with all tests, one can only test whether a hypothesis is incompatible with the sample result. In this case we are testing whether the log-likelihood under the restriction is significantly smaller than the unrestricted log-likelihood. Thus, while it is possible that we may reject the linear but not the logarithmic, or vice versa, it is also possible that we may reject both or fail to reject both.

Example

The figure shows the residual sum of squares for values of $\lambda$ from $-1$ to $1$ for the wage equation example described in Section 4.2 in the text. The maximum likelihood estimate is 0.10, with $RSS = 130.3$. For the linear and logarithmic specifications, $RSS$ was 217.0 and 131.4, respectively, with likelihood ratio statistics $500(\log 217.0 - \log 130.3) = 255.0$ and $500(\log 131.4 - \log 130.3) = 420$. The logarithmic specification is clearly much to be preferred, but even it is rejected at the 5 per cent level, with $\chi^2(1) = 3.84$.

4.4 Additional exercises

A4.1 Is expenditure on your category per capita related to total expenditure per capita? An alternative model specification.

Define a new variable $LGCATPC$ as the logarithm of expenditure per capita on your category. Define a new variable $LGEXPPC$ as the logarithm of total household expenditure per capita. Regress $LGCATPC$ on $LGEXPPC$. Provide an interpretation of the coefficients, and perform appropriate statistical tests.

A4.2 Is expenditure on your category per capita related to household size as well as to total expenditure per capita? An alternative model specification.

Regress $LGCATPC$ on $LGEXPPC$ and $LGSIZE$. Provide an interpretation of the coefficients, and perform appropriate statistical tests.
A4.3 A researcher is considering two regression specifications:

\[ \log Y = \beta_1 + \beta_2 \log X + u \quad (1) \]

and:

\[ \log \frac{Y}{X} = \alpha_1 + \alpha_2 \log X + u \quad (2) \]

where \( u \) is a disturbance term.

Writing \( y = \log Y \), \( x = \log X \), and \( z = \log \frac{Y}{X} \), and using the same sample of \( n \) observations, the researcher fits the two specifications using OLS:

\[ \hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x \quad (3) \]

and:

\[ \hat{z} = \hat{\alpha}_1 + \hat{\alpha}_2 x \quad (4) \]

- Using the expressions for the OLS regression coefficients, demonstrate that \( \hat{\beta}_2 = \hat{\alpha}_2 + 1 \).
- Similarly, using the expressions for the OLS regression coefficients, demonstrate that \( \hat{\beta}_1 = \hat{\alpha}_1 \).
- Hence demonstrate that the relationship between the fitted values of \( y \), the fitted values of \( z \), and the actual values of \( x \), is \( \hat{y}_i - x_i = \hat{z}_i \).
- Hence show that the residuals for regression (3) are identical to those for (4).
- Hence show that the standard errors of \( \hat{\beta}_2 \) and \( \hat{\alpha}_2 \) are the same.
- Explain whether \( R^2 \) would be the same for the two regressions.

A4.4 A researcher has data on a measure of job performance, \( SKILL \), and years of work experience, \( EXP \), for a sample of individuals in the same occupation. Believing there to be diminishing returns to experience, the researcher proposes the model:

\[ SKILL = \beta_1 + \beta_2 \log(EXP) + \beta_3 \log(EXP^2) + u. \]

Comment on this specification.

A4.5 A researcher hypothesises that a variable \( Y \) is determined by a variable \( X \) and considers the following four alternative regression specifications, using cross-sectional data:

\[ Y = \beta_1 + \beta_2 X + u \quad (1) \]
\[ \log Y = \beta_1 + \beta_2 X + u \quad (2) \]
\[ Y = \beta_1 + \beta_2 \log X + u \quad (3) \]
\[ \log Y = \beta_1 + \beta_2 \log X + u. \quad (4) \]

Explain why a direct comparison of \( R^2 \), or of \( RSS \), in models (1) and (2) is illegitimate. What should be the strategy of the researcher for determining which of the four specifications has the best fit?
4. Transformations of variables

A4.6 Is a logarithmic specification preferable to a linear specification for an expenditure function?

Use your category of expenditure from the CES data set. Define \( \text{CATPCST} \) as \( \text{CATPC} \) scaled by its geometric mean and \( \text{LGCATST} \) as the logarithm of \( \text{CATPCST} \). Regress \( \text{CATPCST} \) on \( \text{EXPPC} \) and \( \text{SIZE} \) and regress \( \text{LGCATST} \) on \( \text{LGEXPPC} \) and \( \text{LGSIZE} \). Compare the RSS for these equations.

A4.7

```
. reg LGEARN S EXP ASVABC SASVABC
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>23.63683</td>
<td>4</td>
<td>5.90920575</td>
<td>F( 4, 495) = 22.68</td>
</tr>
<tr>
<td>Residual</td>
<td>128.9624</td>
<td>495</td>
<td>0.26053008</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>152.5992</td>
<td>499</td>
<td>.30581006</td>
<td>R-squared = 0.1549</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.1481</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = .51042</td>
</tr>
</tbody>
</table>

| LGEARN | Coef. Std. Err. t P>|t| [95% Conf. Interval] |
|--------|-------------------|--------|-------------------|-----------------|
| S      | 0.0764243         | 0.011679 | 6.54  | 0.000 | 0.0534603 | 0.0993883 |
| EXP    | 0.0400506         | 0.009648 | 4.15  | 0.000 | 0.0210948 | 0.0590065 |
| ASVABC | -0.2096325        | 0.1406659 | -1.49 | 0.137 | -.4860084 | 0.0667434 |
| SASVABC| 0.0188685         | 0.0093393 | 2.02  | 0.044 | 0.0005189 | 0.0372181 |
| _cons  | 1.386753          | 0.2109596 | 6.57  | 0.000 | .9722664 | 1.801241  |

The output above shows the result of regressing the logarithm of hourly earnings on years of schooling, years of work experience, \( \text{ASVABC} \), score, and \( \text{SASVABC} \), an interactive variable defined as the product of \( S \) and \( \text{ASVABC} \), using \( \text{EAWE Data Set 21} \). The mean values of \( S \), \( \text{EXP} \), and \( \text{ASVABC} \) in the sample were 14.9, 6.4, and 0.27, respectively. Give an interpretation of the regression output.

A4.8 Perform a RESET test of functional misspecification. Using your \( \text{EAWE data set} \), regress \( \text{WEIGHT11} \) on \( \text{HEIGHT} \). Save the fitted values as \( \text{YHAT} \) and define \( \text{YHATSQ} \) as its square. Add \( \text{YHATSQ} \) to the regression specification and test its coefficient.

4.5 Answers to the starred exercises in the textbook

4.8 Suppose that the logarithm of \( Y \) is regressed on the logarithm of \( X \), the fitted regression being:

\[
\log \hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 \log X.
\]

Suppose \( X^* = \mu X \), where \( \mu \) is a constant, and suppose that \( \log Y \) is regressed on \( \log X^* \). Determine how the regression coefficients are related to those of the original regression. Determine also how the \( t \) statistic for \( \hat{\beta}_2 \) and \( R^2 \) for the equation are related to those in the original regression.
4.5. Answers to the starred exercises in the textbook

**Answer:**

Nothing of substance is affected since the change amounts only to a fixed constant shift in the measurement of the explanatory variable.

Let the fitted regression be:

\[ \log \hat{Y} = \hat{\beta}_1^* + \hat{\beta}_2^* \log X^*. \]

Note that:

\[ \log X^*_i - \log X^* = \log \mu X_i - \frac{1}{n} \sum_{j=1}^{n} \log X_j^* \]

\[ = \log \mu X_i - \frac{1}{n} \sum_{j=1}^{n} \log \mu X_j \]

\[ = \log \mu \log X_i - \log \mu \sum_{j=1}^{n} \log X_j \]

\[ = \log X_i - \frac{1}{n} \sum_{j=1}^{n} \log X_j \]

\[ = \log X_i - \log X. \]

Hence \( \hat{\beta}_2^* = \hat{\beta}_2 \). To compute the standard error of \( \hat{\beta}_2^* \), we will also need \( \hat{\beta}_1^* \).

\[ \hat{\beta}_1^* = \log \mu - \hat{\beta}_2 \frac{1}{n} \sum_{j=1}^{n} (\log \mu + \log X_j) \]

\[ = \frac{\log \mu - \hat{\beta}_2 \log X}{\hat{\beta}_2} \frac{1}{\hat{\beta}_2^*} \]

Thus the residual \( \hat{u}_i^* \) is given by:

\[ \hat{u}_i^* = \log Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* \log X_i^* = \log Y_i - (\hat{\beta}_1 - \hat{\beta}_2 \log X) - \hat{\beta}_2 (\log X_i + \log \mu) = \hat{u}_i. \]

Hence the estimator of the variance of the disturbance term is unchanged and so the standard error of \( \hat{\beta}_2^* \) is the same as that for \( \hat{\beta}_2 \). As a consequence, the \( t \) statistic must be the same.

\[ R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (\log Y_i - \log \bar{Y})} = 1 - \frac{\sum \hat{u}_i^2}{\sum (\log Y_i - \log \bar{Y})} = R^2. \]

4.11 RSS was the same in Tables 4.6 and 4.8. Demonstrate that this was not a coincidence.

**Answer:**

This is a special case of the transformation in Exercise 4.7.
4. Transformations of variables

4.14

. gen LGHTSQ = ln(HEIGHTSQ)
. reg LGWT04 LGHEIGHT LGHTSQ

Source | SS df MS
-------------+------------------------------ Number of obs = 500
Model | 7.90843858 1 7.90843858 F( 1, 498) = 211.28
Residual | 18.6403163 498 .037430354             Prob > F = 0.0000
-------------+------------------------------ Adj R-squared = 0.2979
Total | 26.5487548 499 .053203918 Root MSE = .19347

------------------------------------------------------------------------------
LGWT04 | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-------------+----------------------------------------------------------------
LGHEIGHT | (dropped)
LGHTSQ | 1.053218 .0724577 14.54 0.000 .9108572 1.195578
_cons | -3.78834 .610925 -6.20 0.000 -4.988648 -2.588031
------------------------------------------------------------------------------

The output shows the results of regressing, \textit{LGWT04}, the logarithm of \textit{WEIGHT04}, on \textit{LGHEIGHT}, the logarithm of \textit{HEIGHT}, and \textit{LGHTSQ}, the logarithm of the square of \textit{HEIGHT}, using \textit{EAWE} Data Set 21. Explain the regression results, comparing them with those in Exercise 4.2.

Answer:

\textit{LGHTSQ} = 2 \textit{LGHEIGHT}, so the specification is subject to exact multicollinearity. In such a situation, Stata drops one of the variables responsible.

4.18

. nl (S = {beta1} + {beta2}/({beta3} + SIBLINGS)) if SIBLINGS>0
(\textit{obs} = 473)

Iteration 0: residual SS = 3502.041
Iteration 1: residual SS = 3500.884

....................
Iteration 14: residual SS = 3482.794

Source | SS df MS
-------------+------------------------------ Number of obs = 473
Model | 132.339291 2 66.1696453 R-squared = 0.0366
Residual | 3482.7939 470 7.41019979 Adj R-squared = 0.0325
-------------+------------------------------ Root MSE = 2.722168
Total | 3615.13319 472 7.65918049 Res. dev. = 2286.658

------------------------------------------------------------------------------
S | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-------------+----------------------------------------------------------------
/beta1 | 10.45811 5.371492 1.95 0.052 -.0970041 21.01322
/beta2 | 47.95198 125.3578 0.38 0.702 -198.3791 294.2831
/beta3 | 8.6994 15.10277 0.58 0.565 -20.97791 38.37671
------------------------------------------------------------------------------

Parameter betai taken as constant term in model & ANOVA table
The output uses *EAWE* Data Set 21 to fit the nonlinear model:

\[ S = \beta_1 + \frac{\beta_2}{\beta_3 + \text{SIBLINGS}} + u \]

where \( S \) is the years of schooling of the respondent and \( \text{SIBLINGS} \) is the number of brothers and sisters. The specification is an extension of that for Exercise 4.1, with the addition of the parameter \( \beta_3 \). Provide an interpretation of the regression results and compare it with that for Exercise 4.1.

**Answer:**

As in Exercise 4.1, the estimate of \( \beta_1 \) provides an estimate of the lower bound of schooling, 10.46 years, when the number of siblings is large. The other parameters do not have straightforward interpretations. The figure below represents the relationship. Comparing this figure with that for Exercise 4.1, it can be seen that it gives a very different picture of the adverse effect of additional siblings. The specification in Exercise 4.1 suggests that the adverse effect is particularly large for the first few siblings, and then attenuates. The revised specification indicates that the adverse effect is more evenly spread and is more enduring. However, the relationship has been fitted with imprecision since the estimates of \( \beta_2 \) and \( \beta_3 \) are not significant.

![Graph showing comparison between Exercises 4.1 and 4.18](image-url)
Transformations of variables

| LGFDHOPC | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|-------|-----|---------------------|
| LGEPPC   | .6092734 | .0088338  | 68.97 | 0.000 | .5919562 .6265905   |
| _cons    | .8988291 | .0703516  | 12.78 | 0.000 | .7609161 1.036742   |

The regression implies that the income elasticity of expenditure on food is 0.61 (supposing that total household expenditure can be taken as a proxy for permanent income). In addition to testing the null hypothesis that the elasticity is equal to zero, which is rejected at a very high significance level for all the categories, one might test whether it is different from 1, as a means of classifying the categories of expenditure as luxuries (elasticity $>1$) and necessities (elasticity $<1$).

The table gives the results for all the categories of expenditure.

Regression of LGCATPC on EXPPC

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\beta}_2$</th>
<th>s.e.$(\hat{\beta}_2)$</th>
<th>$t (\hat{\beta}_2 = 0)$</th>
<th>$t (\hat{\beta}_2 = 1)$</th>
<th>$R^2$</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>1.098</td>
<td>0.030</td>
<td>37.20</td>
<td>3.33</td>
<td>0.30</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>0.794</td>
<td>0.021</td>
<td>37.34</td>
<td>-9.69</td>
<td>0.23</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>0.812</td>
<td>0.049</td>
<td>16.54</td>
<td>-3.84</td>
<td>0.14</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>1.382</td>
<td>0.090</td>
<td>15.43</td>
<td>4.27</td>
<td>0.29</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>0.586</td>
<td>0.011</td>
<td>50.95</td>
<td>-36.05</td>
<td>0.30</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>0.947</td>
<td>0.015</td>
<td>64.68</td>
<td>-3.59</td>
<td>0.45</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>0.609</td>
<td>0.009</td>
<td>68.97</td>
<td>-44.23</td>
<td>0.42</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>0.608</td>
<td>0.027</td>
<td>22.11</td>
<td>-14.26</td>
<td>0.21</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.912</td>
<td>0.085</td>
<td>10.66</td>
<td>-1.03</td>
<td>0.19</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>0.677</td>
<td>0.012</td>
<td>56.92</td>
<td>-27.18</td>
<td>0.36</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>0.868</td>
<td>0.021</td>
<td>40.75</td>
<td>-6.22</td>
<td>0.25</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>1.033</td>
<td>0.014</td>
<td>73.34</td>
<td>2.34</td>
<td>0.46</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>0.607</td>
<td>0.047</td>
<td>13.00</td>
<td>-8.40</td>
<td>0.11</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.510</td>
<td>0.055</td>
<td>9.29</td>
<td>-8.92</td>
<td>0.11</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.817</td>
<td>0.033</td>
<td>9.87</td>
<td>-2.21</td>
<td>0.19</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>0.891</td>
<td>0.019</td>
<td>48.14</td>
<td>-5.88</td>
<td>0.38</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>0.909</td>
<td>0.032</td>
<td>28.46</td>
<td>-2.84</td>
<td>0.26</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>0.665</td>
<td>0.045</td>
<td>14.88</td>
<td>-7.49</td>
<td>0.17</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>0.710</td>
<td>0.012</td>
<td>58.30</td>
<td>-23.82</td>
<td>0.37</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.629</td>
<td>0.046</td>
<td>13.72</td>
<td>-8.09</td>
<td>0.16</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>0.721</td>
<td>0.035</td>
<td>20.39</td>
<td>-7.87</td>
<td>0.26</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>0.733</td>
<td>0.028</td>
<td>26.22</td>
<td>-9.57</td>
<td>0.21</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.723</td>
<td>0.077</td>
<td>9.43</td>
<td>-3.60</td>
<td>0.14</td>
</tr>
</tbody>
</table>

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
The income elasticity, 0.58, is now a little lower than before. The size elasticity is significantly negative, suggesting economies of scale and indicating that the model in the previous exercise was misspecified.

The specification is equivalent to that in Exercise 4.5 in the text. Writing the latter again as:

\[ LGCAT = \beta_1 + \beta_2 LGEXP + \beta_3 LGSIZE + u \]

we have:

\[ LGCAT - LGSIZE = \beta_1 + \beta_2(LGEXP - LGSIZE) + (\beta_3 + \beta_2 - 1)LGSIZE + u \]

and so:

\[ LGCATPC = \beta_1 + \beta_2 LGEXPPC + (\beta_3 + \beta_2 - 1)LGSIZE + u. \]

Note that the estimates of the income elasticity are identical to those in Exercise 4.5 in the text. This follows from the fact that the theoretical coefficient, \( \beta_2 \), has not been affected by the manipulation. The specification differs from that in Exercise A4.1 in that we have not dropped the \( LGSIZE \) term and so we are not imposing the restriction \( \beta_3 + \beta_2 - 1 = 0 \).
A4.3 A researcher is considering two regression specifications:

\[ \log Y = \beta_1 + \beta_2 \log X + u \]  
(1)

and:

\[ \log \frac{Y}{X} = \alpha_1 + \alpha_2 \log X + u \]  
(2)

where \( u \) is a disturbance term.

**Determine whether (2) is a reparameterised or a restricted version of (1).**

(2) may be rewritten:

\[ \log Y = \alpha_1 + (\alpha_2 + 1) \log X + u \]

so it is a reparameterised version of (1) with \( \beta_1 = \alpha_1 \) and \( \beta_2 = \alpha_2 + 1 \).

Writing \( y = \log Y \), \( x = \log X \), and \( z = \log \frac{Y}{X} \), and using the same sample of \( n \) observations, the researcher fits the two specifications using OLS:

\[ \hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x \]  
(3)

and:

\[ \hat{z} = \hat{\alpha}_1 + \hat{\alpha}_2 x. \]  
(4)
4.6. Answers to the additional exercises

Using the expressions for the OLS regression coefficients, demonstrate that \( \hat{\beta}_2 = \hat{\alpha}_2 + 1 \).

\[
\hat{\alpha}_2 = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})([y_i - x_i] - [\bar{y} - \bar{x}])}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} - \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \hat{\beta}_2 - 1.
\]

Similarly, using the expressions for the OLS regression coefficients, demonstrate that \( \hat{\beta}_1 = \hat{\alpha}_1 \).

\[
\hat{\alpha}_1 = \bar{z} - \hat{\alpha}_2 \bar{x} = (\bar{y} - \bar{x}) - \hat{\alpha}_2 \bar{x} = \bar{y} - (\hat{\alpha}_2 + 1) \bar{x} = \bar{y} - \hat{\beta}_2 \bar{x} = \hat{\beta}_1.
\]

Hence demonstrate that the relationship between the fitted values of \( y \), the fitted values of \( z \), and the actual values of \( x \), is \( \hat{y}_i - x_i = \hat{z}_i \).

\[
\hat{z}_i = \hat{\alpha}_1 + \hat{\alpha}_2 x_i = \hat{\beta}_1 + (\hat{\beta}_2 - 1)x_i = \hat{\beta}_1 + \hat{\beta}_2 x_i - x_i = \hat{y}_i - x_i.
\]

Hence show that the residuals for regression (3) are identical to those for (4).

Let \( \hat{\alpha}_i \) be the residual in (3) and \( \hat{\alpha}_i \) the residual in (4). Then:

\[
\hat{\alpha}_i = z_i - \hat{\alpha}_2 x_i = y_i - x_i - (\hat{y}_i - x_i) = y_i - \hat{y}_i = \hat{\alpha}_i.
\]

Hence show that the standard errors of \( \hat{\beta}_2 \) and \( \hat{\alpha}_2 \) are the same.

The standard error of \( \hat{\beta}_2 \) is:

\[
\text{s.e.}(\hat{\beta}_2) = \sqrt{\frac{\sum \hat{\alpha}_i^2/(n-2)}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum \hat{\alpha}_i^2/(n-2)}{\sum (x_i - \bar{x})^2}} = \text{s.e.}(\hat{\alpha}_2).
\]

Determine the relationship between the \( t \) statistic for \( \hat{\beta}_2 \) and the \( t \) statistic for \( \hat{\alpha}_2 \), and give an intuitive explanation for the relationship.

\[
t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{\text{s.e.}(\hat{\beta}_2)} = \frac{\hat{\alpha}_2 + 1}{\text{s.e.}(\hat{\alpha}_2)}.
\]

The \( t \) statistic for \( \hat{\beta}_2 \) is for the test of \( H_0 : \beta_2 = 0 \). Given the relationship, it is also for the test of \( H_0 : \alpha_2 = -1 \). The tests are equivalent since both of them reduce the model to \( \log(Y) \) depending only on an intercept and the disturbance term.

Explain whether \( R^2 \) would be the same for the two regressions.

\( R^2 \) will be different because it measures the proportion of the variance of the dependent variable explained by the regression, and the dependent variables are different.

A4.4 The proposed model:

\[
SKILL = \beta_1 + \beta_2 \log(EXP) + \beta_3 \log(EXP^2) + u
\]

cannot be fitted since:

\[
\log(EXP^2) = 2 \log(EXP)
\]

and the specification is therefore subject to exact multicollinearity.

81

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
4. Transformations of variables

A4.5 In (1) $R^2$ is the proportion of the variance of $Y$ explained by the regression. In (2) it is the proportion of the variance of log $Y$ explained by the regression. Thus, although related, they are not directly comparable. In (1) $RSS$ has dimension the squared units of $Y$. In (2) it has dimension the squared units of log $Y$. Typically it will be much lower in (2) because the logarithm of $Y$ tends to be much smaller than $Y$.

The specifications with the same dependent variable may be compared directly in terms of $RSS$ (or $R^2$) and hence two of the specifications may be eliminated immediately. The remaining two specifications should be compared after scaling, with $Y$ replaced by $Y^*$ where $Y^*$ is defined as $Y$ divided by the geometric mean of $Y$ in the sample. $RSS$ for the scaled regressions will then be comparable.

A4.6 The $RSS$ comparisons for all the categories of expenditure indicate that the logarithmic specification is overwhelmingly superior to the linear one. The differences are actually surprisingly large and suggest that some other factor may also be at work. One possibility is that the data contain many outliers, and these do more damage to the fit in linear than in logarithmic specifications. To see this, plot CATPC and EXPPC and compare with a plot of LGCATPC and LGEXPPC. (Strictly speaking, you should control for SIZE and LGSIZE using the Frisch–Waugh–Lovell method described in Chapter 3.)

The following Stata output gives the results of fitting the model for FDHO, assuming that both the dependent variable and the explanatory variables are subject to the Box–Cox transformation with the same value of $\lambda$. Iteration messages have been deleted. The maximum likelihood estimate of $\lambda$ is 0.10, so the logarithmic specification is a better approximation than the linear specification. The latter is very soundly rejected by the likelihood-ratio test.

```
. boxcox FDHO PC EXPPC SIZE if FDHO>0, model(lambda)
    Number of obs = 6334
    LR chi2(2) = 3592.55
    Log likelihood = -41551.328 Prob > chi2 = 0.000

------------------------------------------------------------------------------
  FDHO PC | Coef. Std. Err. z P>|z|  [95% Conf. Interval]
-------------+----------------------------------------------------------------
   /lambda  | 0.1019402  0.0117364  8.69 0.000   0.0789372   0.1249432
------------------------------------------------------------------------------

Estimates of scale-variant parameters
-------------------------------+---------------------------------------
     | Coef.
-------------------------------+---------------------------------------
Notrans |                        
   _cons | 2.292828
-------------------------------+---------------------------------------
Trans |                        
   EXPPC | 0.4608736
   SIZE | -0.1486856
-------------------------------+---------------------------------------
   /sigma | 0.9983288
```

82
4.6. Answers to the additional exercises

<table>
<thead>
<tr>
<th>Test HO:</th>
<th>Restricted log likelihood</th>
<th>LR statistic ch12</th>
<th>P-value Prob &gt; ch12</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda = -1</td>
<td>-50942.835</td>
<td>18783.01</td>
<td>0.000</td>
</tr>
<tr>
<td>lambda = 0</td>
<td>-41590.144</td>
<td>77.63</td>
<td>0.000</td>
</tr>
<tr>
<td>lambda = 1</td>
<td>-44053.749</td>
<td>5004.84</td>
<td>0.000</td>
</tr>
</tbody>
</table>

A4.7 Let the theoretical model for the regression be written:

\[ \text{LGEARN} = \beta_1 + \beta_2 S + \beta_3 \text{EXP} + \beta_4 \text{ASVABC} + \beta_5 \text{SA} + u. \]

The estimate of \( \beta_4 \) is negative, at first sight suggesting that cognitive ability has an adverse effect on earnings, contrary to common sense and previous results with wage equations of this kind. However, rewriting the model as:

\[ \text{LGEARN} = \beta_1 + \beta_2 S + \beta_3 \text{EXP} + (\beta_4 + \beta_5 S) \text{ASVABC} + u \]

it can be seen that, as a consequence of the inclusion of the interactive term, \( \beta_4 \) represents the effect of a marginal year of schooling for an individual with no schooling. Since no individual in the sample had fewer than 8 years of schooling, the perverse sign of the estimate illustrates only the danger of extrapolating outside the data range. It makes better sense to evaluate the implicit coefficient for an individual with the mean years of schooling, 14.9. This is 
\[ (-0.2096 + 0.0189 \times 14.9) = 0.072, \]

implying a much more plausible 7.2 per cent increase in earnings for each standard deviation increase in cognitive ability. The positive sign of the coefficient of \( \text{SA} \) suggests that schooling and cognitive ability have mutually reinforcing effects on earnings.

One way of avoiding nonsense parameter estimates is to measure the variables in question from their sample means. This has been done in the regression output below, where \( S1 \) and \( \text{ASVABC1} \) are schooling and \( \text{ASVABC} \) measured from their sample means and \( \text{SASVABC1} \) is their interaction. The coefficients of \( S \) and \( \text{ASVABC} \) now provide estimates of their effects when the other variable is equal to its sample mean.

```
reg LGEARN S1 EXP ASVABC1 SASVABC1
```

```
<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>23.6368304</td>
<td>4</td>
<td>5.90920759</td>
</tr>
<tr>
<td>Residual</td>
<td>128.962389</td>
<td>495</td>
<td>.260530079</td>
</tr>
<tr>
<td>Total</td>
<td>152.59922</td>
<td>499</td>
<td>.30581006</td>
</tr>
</tbody>
</table>

| LGEARN        | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|---------------|--------|-----------|-------|------|----------------------|
| S1            | .0815188 | .0116521 | 7.00  | 0.000 | .0586252 to 10.44125 |
| EXP           | .0400506 | .0096479 | 4.15  | 0.000 | .0210948 to .0590065 |
| ASVABC1       | .0715084 | .0298278 | 2.40  | 0.017 | .0129036 to 1.301132 |
| SASVABC1      | .0188685 | .0093393 | 2.02  | 0.044 | .0005819 to .0372181 |
| _cons         | 2.544783 | .0675566 | 37.67 | 0.000 | 2.41205 to 2.677516  |
```
4. Transformations of variables

A4.8 In the first part of the output, \( \text{WEIGHT11} \) is regressed on \( \text{HEIGHT} \), using \( \text{EAWE Data Set 21} \). The predict command saves the fitted values from the most recent regression, assigning them the variable name that follows the command, in this case \( \text{YHAT} \). \( \text{YHATSQ} \) is defined as the square of \( \text{YHAT} \), and this is added to the regression specification. Somewhat surprisingly, its coefficient is not significant. A logarithmic regression of \( \text{WEIGHT11} \) on \( \text{HEIGHT} \) yields an estimated elasticity of 2.05, significantly different from 1 at a high significance level. Multicollinearity is responsible for the failure to detect nonlinearity here. \( \text{YHAT} \) is very highly correlated with \( \text{HEIGHT} \).

```
.reg WEIGHT11 HEIGHT
----------------------------------------------------------------------------
 Source | SS   df  MS
---------+---------------------------------------------------------------------
 Model   | 236642.736 1 236642.736 F( 1, 498) = 139.97
 Residual| 841926.912 498 1690.61629 Prob > F = 0.0000
---------+---------------------------------------------------------------------
 Total   | 1078569.65 499 2161.46222 R-squared = 0.2194
          | Adj R-squared = 0.2178
          | Root MSE = 41.117
----------------------------------------------------------------------------

WEIGHT11 | Coef.  Std. Err.  t   P>|t|   [95% Conf. Interval]
---------+---------------------------------------------------------------------
 HEIGHT   | 5.369246  .4538259 11.83 0.000  4.477597 6.260895
 _cons    | -184.7802 30.8406  -5.99 0.000 -245.3739 -124.1865

.predict YHAT
.gen YHATSQ = YHAT*YHAT
.reg WEIGHT11 HEIGHT YHATSQ
----------------------------------------------------------------------------
 Source | SS   df  MS
---------+---------------------------------------------------------------------
 Model   | 237931.888 2 118965.944 F( 2, 497) = 70.33
 Residual| 840637.76 497 1691.42407 Prob > F = 0.0000
---------+---------------------------------------------------------------------
 Total   | 1078569.65 499 2161.46222 R-squared = 0.2206
          | Adj R-squared = 0.2175
          | Root MSE = 41.127
----------------------------------------------------------------------------

WEIGHT11 | Coef.  Std. Err.  t   P>|t|   [95% Conf. Interval]
---------+---------------------------------------------------------------------
 HEIGHT   | -.4995924 6.7377741 -0.07 0.941 -13.73756 12.73837
 YHATSQ   | .0030233  .003463  0.87 0.383  -.0037807 .0098273
 _cons    | 114.5523 344.2538  0.33 0.739 -561.8199 790.9244
```

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
Chapter 5
Dummy variables

5.1 Overview

This chapter explains the definition and use of a dummy variable, a device for allowing qualitative characteristics to be introduced into the regression specification. Although the intercept dummy may appear artificial and strange at first sight, and the slope dummy even more so, you will become comfortable with the use of dummy variables very quickly. The key is to keep in mind the graphical representation of the regression model.

5.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to explain:

- how the intercept and slope dummy variables are defined
- what impact they have on the regression specification
- how the choice of reference (omitted) category affects the interpretation of $t$ tests on the coefficients of dummy variables
- how a change of reference category would affect the regression results
- how to perform a Chow test
- when and why a Chow test is equivalent to a particular $F$ test of the joint explanatory power of a set of dummy variables.

5.3 Additional exercises

A5.1 In Exercise A1.4 the logarithm of earnings was regressed on height using EAWE Data Set 21 and, somewhat surprisingly, it was found that height had a highly significant positive effect. We have seen that the logarithm of earnings is more satisfactory than earnings as the dependent variable in a wage equation. Fitting the semilogarithmic specification, we obtain:
5. Dummy variables

`. reg LGEARN HEIGHT
----------------------------------------------------------------------------
Source | SS df MS Number of obs = 500
-----------+------------------------------ F( 1, 498) = 6.27
Model | 1.84965685 1 1.84965685 Prob > F = 0.0126
Residual | 146.79826 498 .294775622 R-squared = 0.0124
-----------+------------------------------ Adj R-squared = 0.0105
Total | 148.647917 499 .297891616 Root MSE = .54293
----------------------------------------------------------------------------
LGEARN | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----------+----------------------------------------------------------------
HEIGHT | .0148894 .005944 2.50 0.013 -.003211 .0265678
_cons | 1.746174 .4032472 4.33 0.000 .9538982 2.538449
----------------------------------------------------------------------------

The t statistic for HEIGHT is again significant, if only at the 5 per cent level. In Exercise A1.4 it was hypothesised that the effect might be attributable to males tending to have greater earnings than females and also tending to be taller. The output below shows the result of adding the dummy variable to the specification, to control for sex. Comment on the results.

`. reg LGEARN HEIGHT MALE
----------------------------------------------------------------------------
Source | SS df MS Number of obs = 500
-----------+------------------------------ F( 2, 497) = 4.20
Model | 2.47043329 2 1.23521664 Prob > F = 0.0155
Residual | 146.177483 497 .294119685 R-squared = 0.0166
-----------+------------------------------ Adj R-squared = 0.0127
Total | 148.647917 499 .297891616 Root MSE = .54233
----------------------------------------------------------------------------
LGEARN | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----------+----------------------------------------------------------------
HEIGHT | .0060845 .0084844 0.72 0.474 -.0105852 .0267541
MALE | .1007018 .0693157 1.45 0.147 -.0354862 .2368898
_cons | 2.292078 .5508559 4.16 0.000 1.209784 3.374371
----------------------------------------------------------------------------

A5.2 Does ethnicity have an effect on household expenditure?

The variable REFRACE in the CES data set is coded 1 if the reference individual in the household, usually the head of the household, is white and it is coded greater than 1 for other ethnicities. Define a dummy variable NONWHITE that is 0 if REFRACE is 1 and 1 if REFRACE is greater than 1. Regress LGCATPC on LGEXPPC, LGSIZE, and NONWHITE. Provide an interpretation of the coefficients, and perform appropriate statistical tests.

A5.3 Does education have an effect on household expenditure?

The variable REFEDUC in the CES data set provides information on the education of the reference individual in the household. Define dummy variables EDUCDO (high-school drop out or less), EDUCSC (some college), and EDUCBA (complete college or more) using the following rules:
5.3. Additional exercises

- \( \text{EDUCDO} = 1 \) if \( \text{REFEDUC} < 12 \), 0 otherwise
- \( \text{EDUCSC} = 1 \) if \( \text{REFEDUC} = 13 \) or 14, 0 otherwise
- \( \text{EDUCBA} = 1 \) if \( \text{REFEDUC} > 14 \), 0 otherwise.

Regress \( \text{LGCATPC} \) on \( \text{LGEXPPC}, \text{LGSIZE}, \text{EDUCDO}, \text{EDUCSC}, \) and \( \text{EDUCBA} \). Provide an interpretation of the coefficients, and perform appropriate statistical tests. Note that the reference (omitted) category for the dummy variables is high school graduate with no college \((\text{REFEDUC} = 12)\).

A5.4 Using the \( \text{CES} \) data set, evaluate whether the education dummies as a group have significant explanatory power for expenditure on your category of expenditure by comparing the residual sums of squares in the regressions in Exercises A4.2 and A5.3.

A5.5 Repeat Exercise A5.3 making \( \text{EDUCDO} \) the reference (omitted) category. Introduce a new dummy variable \( \text{EDUCHSD} \) for high school diploma, since this is no longer the omitted category:

- \( \text{EDUCHSD} = 1 \) if \( \text{REFEDUC} = 12 \), 0 otherwise.

Evaluate the impact on the interpretation of the coefficients and the statistical tests.

A5.6 A researcher has data on hourly earnings in dollars, \( \text{EARNINGS} \), years of schooling (highest grade completed), \( S \), and sector of employment, \( \text{GOV} \), for 1,355 male respondents in the National Longitudinal Survey of Youth 1979– for 2002. \( \text{GOV} \) is defined as a dummy variable equal to 0 if the respondent was working in the private sector and 1 if the respondent was working in the government sector. 91 per cent of the private sector workers and 95 per cent of the government sector workers had at least 12 years of schooling. The mean value of \( S \) was 13.5 for the private sector and 14.6 for the government sector. The researcher regresses \( \text{LGEARN} \), the natural logarithm of \( \text{EARNINGS} \):

- (1) on \( \text{GOV} \) alone
- (2) on \( \text{GOV} \) and \( S \)
- (3) on \( \text{GOV}, S, \) and \( \text{SGOV} \)

where the variable \( \text{SGOV} \) is defined to be the product of \( S \) and \( \text{GOV} \), with the results shown in the following table.

Standard errors are shown in parentheses and \( t \) statistics in square brackets. \( \text{RSS} = \) residual sum of squares.
5. Dummy variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOV</td>
<td>0.007</td>
<td>-0.121</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.038)</td>
<td>(0.193)</td>
</tr>
<tr>
<td></td>
<td>[0.16]</td>
<td>[-3.22]</td>
<td>[3.76]</td>
</tr>
<tr>
<td>S</td>
<td>—</td>
<td>0.116</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>[21.07]</td>
<td>[20.82]</td>
</tr>
<tr>
<td>SGOV</td>
<td>—</td>
<td>—</td>
<td>-0.059</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(0.013)</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>[-4.48]</td>
</tr>
<tr>
<td>constant</td>
<td>2.941 *</td>
<td>1.372</td>
<td>1.195</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.076)</td>
<td>(0.085)</td>
</tr>
<tr>
<td></td>
<td>[163.62]</td>
<td>[18.04]</td>
<td>[14.02]</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.000</td>
<td>0.247</td>
<td>0.258</td>
</tr>
<tr>
<td>RSS</td>
<td>487.7</td>
<td>367.2</td>
<td>361.8</td>
</tr>
</tbody>
</table>

- Explain verbally why the estimates of the coefficient of GOV are different in regressions (1) and (2).
- Explain the difference in the estimates of the coefficient of GOV in regressions (2) and (3).
- The correlation between GOV and SGOV was 0.977. Explain the variations in the standard error of the coefficient of GOV in the three regressions.

A5.7 A researcher has data on the average annual rate of growth of employment, $e$, and the average annual rate of growth of GDP, $x$, both measured as percentages, for a sample of 27 developing countries and 23 developed ones for the period 1985–1995. He defines a dummy variable $D$ that is equal to 1 for the developing countries and 0 for the others. Hypothesising that the impact of GDP growth on employment growth is lower in the developed countries than in the developing ones, he defines a slope dummy variable $xD$ as the product of $x$ and $D$ and fits the regression (standard errors in parentheses):

$$
\hat{e} = -1.45 + 0.19x + 0.78xD \quad R^2 = 0.61
$$

He also runs simple regressions of $e$ on $x$ for the whole sample, for the developed countries only, and for the developing countries only, with the following results:

- **Whole sample**
  $$
  \hat{e} = -0.56 + 0.24x \quad R^2 = 0.04
  $$
  $$(0.53) \quad (0.16) \quad RSS = 121.61$$

- **Developed countries**
  $$
  \hat{e} = -2.74 + 0.50x \quad R^2 = 0.35
  $$
  $$(0.58) \quad (0.15) \quad RSS = 18.63$$

- **Developing countries**
  $$
  \hat{e} = -0.85 + 0.78x \quad R^2 = 0.51
  $$
  $$(0.42) \quad (0.15) \quad RSS = 25.23$$

- Explain mathematically and graphically the role of the dummy variable $xD$ in this model.
5.3. Additional exercises

- The researcher could have included $D$ as well as $xD$ as an explanatory variable in the model. Explain mathematically and graphically how it would have affected the model.

- Suppose that the researcher had included $D$ as well as $xD$.
  - What would the coefficients of the regression have been?
  - What would the residual sum of squares have been?
  - What would the $t$ statistic for the coefficient of $D$ have been?

- Perform two tests of the researcher’s hypothesis. Explain why you would not test it with a $t$ test on the coefficient of $xD$ in regression (1).

A5.8 Does going to college have an effect on household expenditure?

Using the CES data set, define a dummy variable $COLLEGE$ that is 0 if $REFEDUC$ is less than 13 (no college education) and 1 if $REFEDUC$ is greater than 12 (partial or complete college education). Regress $LGCATPC$ on $LGEXPPC$ and $LGSIZE$: (1) for those respondents with $COLLEGE = 1$, (2) for those respondents with $COLLEGE = 0$, and (3) for the whole sample. Perform a Chow test.

A5.9 How does education impact on household expenditure?

In Exercise A5.8 you defined an intercept dummy $COLLEGE$ that allowed you to investigate whether going to college caused a shift in your expenditure function. Now define slope dummy variables that allow you to investigate whether going to college affects the coefficients of $LGEXPPC$ and $LGSIZE$. Define $LEXPCOL$ as the product of $LGEXPPC$ and $COLLEGE$, and define $LSIZECOL$ as the product of $LGSIZE$ and $COLLEGE$. Regress $LGATPC$ on $LGEXPPC$, $LGSIZE$, $COLLEGE$, $LEXPCOL$, and $LSIZECOL$. Provide an interpretation of the coefficients, and perform appropriate tests. Include a test of the joint explanatory power of the dummy variables by comparing RSS in this regression with that in Exercise A4.3. Verify that the outcome of this $F$ test is identical to that for the Chow test in Exercise A5.8.

A5.10 You are given the following data on 2,800 respondents in the National Longitudinal Survey of Youth 1979– with jobs in 2011:

- hourly earnings in the respondent’s main job at the time of the 2011 interview
- educational attainment (highest grade completed)
- mother’s and father’s educational attainment
- $ASVABC$ score
- sex
- ethnicity: black, Hispanic, or white, that is (not black nor Hispanic)
- whether the main job in 2011 was in the government sector or the private sector.

As a policy analyst, you are asked to investigate whether there is evidence of earnings discrimination, positive or negative, by sex or ethnicity in (1) the
5. Dummy variables

government sector, and (2) the private sector. Explain how you would do this, giving a mathematical representation of your regression specification(s).

You are also asked to investigate whether the incidence of earnings discrimination, if any, is significantly different in the two sectors. Explain how you would do this, giving a mathematical representation of your regression specification(s). In particular, discuss whether a Chow test would be useful for this purpose.

A5.11 A researcher has data from the National Longitudinal Survey of Youth 1997–2000 on hourly earnings, $Y$, years of schooling, $S$, and years of work experience, $EXP$, for a sample of 1,774 males and 1,468 females. She defines a dummy variable $MALE$ for being male, a slope dummy variable $SMALE$ as the product of $S$ and $MALE$, and another slope dummy variable $EXPMALE$ as the product of $EXP$ and $MALE$. She performs the following regressions (1) $\log Y$ on $S$ and $EXP$ for the entire sample, (2) $\log Y$ on $S$ and $EXP$ for males only, (3) $\log Y$ on $S$ and $EXP$ for females only, (4) $\log Y$ on $S$, $EXP$, and $MALE$ for the entire sample, and (5) $\log Y$ on $S$, $EXP$, $MALE$, $SMALE$, and $EXPMALE$ for the entire sample. The results are shown in the table, with standard errors in parentheses. $RSS$ is the residual sum of squares and $n$ is the number of observations.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>0.094</td>
<td>0.099</td>
<td>0.094</td>
<td>0.0967</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$EXP$</td>
<td>0.046</td>
<td>0.042</td>
<td>0.039</td>
<td>0.040</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$MALE$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.234</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.016)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>$SMALE$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>$EXPMALE$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>constant</td>
<td>5.165</td>
<td>5.283</td>
<td>5.166</td>
<td>5.111</td>
<td>5.166</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.083)</td>
<td>(0.068)</td>
<td>(0.052)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.319</td>
<td>0.277</td>
<td>0.363</td>
<td>0.359</td>
<td>0.359</td>
</tr>
<tr>
<td>$RSS$</td>
<td>714.6</td>
<td>411.0</td>
<td>261.6</td>
<td>672.8</td>
<td>672.5</td>
</tr>
<tr>
<td>$n$</td>
<td>3,242</td>
<td>1,774</td>
<td>1,468</td>
<td>3,242</td>
<td>3,242</td>
</tr>
</tbody>
</table>

The correlations between $MALE$ and $SMALE$, and $MALE$ and $EXPMALE$, were both 0.96. The correlation between $SMALE$ and $EXPMALE$ was 0.93.

- Give an interpretation of the coefficients of $S$ and $SMALE$ in regression (5).
- Give an interpretation of the coefficients of $MALE$ in regressions (4) and (5).
- The researcher hypothesises that the earnings function is different for males and females. Perform a test of this hypothesis using regression (4), and also using regressions (1) and (5).
- Explain the differences in the tests using regression (4) and using regressions (1) and (5).
5.3. Additional exercises

- At a seminar someone suggests that a Chow test could shed light on the researcher’s hypothesis. Is this correct?
- Explain which of (1), (4), and (5) would be your preferred specification.

A5.12 A researcher has data for the year 2000 from the National Longitudinal Survey of Youth 1997– on the following characteristics of the respondents: hourly earnings, \( EARNINGS \), measured in dollars; years of schooling, \( S \); years of work experience, \( EXP \); sex; and ethnicity (blacks, hispanics, and ‘whites’ (those not classified as black or hispanic)). She drops the hispanics from the sample, leaving 2,135 ‘whites’ and 273 blacks, and defines dummy variables \( MALE \) and \( BLACK \). \( MALE \) is defined to be 1 for males and 0 for females. \( BLACK \) is defined to be 1 for blacks and 0 for ‘whites’. She defines \( LGEARN \) to be the natural logarithm of \( EARNINGS \). She fits the following ordinary least squares regressions, each with \( LGEARN \) as the dependent variable:

- (1) Explanatory variables \( S, EXP, \) and \( MALE \), whole sample
- (2) Explanatory variables \( S, EXP, MALE, \) and \( BLACK \), whole sample
- (3) Explanatory variables \( S, EXP, \) and \( MALE \), ‘whites’ only
- (4) Explanatory variables \( S, EXP, \) and \( MALE \), blacks only.

She then defines interaction terms \( SB = S \times BLACK \), \( EB = EXP \times BLACK \), and \( MB = MALE \times BLACK \), and runs a fifth regression, still with \( LGEARN \) as the dependent variable:

- (5) Explanatory variables \( S, EXP, MALE, BLACK, SB, EB, MB \), whole sample.

The results are shown in the table. Unfortunately, some of those for Regression 4 are missing from the table. \( RSS \) = residual sum of squares. Standard errors are given in parentheses.
### 5. Dummy variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>whole sample</td>
<td>whole sample</td>
<td>'whites' only</td>
<td>blacks only</td>
<td>whole sample</td>
</tr>
<tr>
<td>S</td>
<td>0.124 (0.004)</td>
<td>0.121 (0.004)</td>
<td>0.122 (0.004)</td>
<td><strong>V</strong> 0.122 (0.004)</td>
<td></td>
</tr>
<tr>
<td>EXP</td>
<td>0.033 (0.002)</td>
<td>0.032 (0.002)</td>
<td>0.033 (0.003)</td>
<td><strong>W</strong> 0.033 (0.003)</td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>0.278 (0.020)</td>
<td>0.277 (0.020)</td>
<td>0.306 (0.021)</td>
<td><strong>X</strong> 0.306 (0.021)</td>
<td></td>
</tr>
<tr>
<td>BLACK</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.205 (0.032)</td>
<td></td>
</tr>
<tr>
<td>SB</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—0.009 (0.016)</td>
</tr>
<tr>
<td>EB</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—0.006 (0.007)</td>
</tr>
<tr>
<td>MB</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—0.280 (0.065)</td>
</tr>
<tr>
<td>constant</td>
<td>0.390 (0.075)</td>
<td>0.459 (0.076)</td>
<td>0.411 (0.084)</td>
<td><strong>Y</strong> 0.411 (0.082)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.335</td>
<td>0.341</td>
<td>0.332</td>
<td>0.321</td>
<td>0.347</td>
</tr>
<tr>
<td>RSS</td>
<td>610.0</td>
<td>605.1</td>
<td>555.7</td>
<td><strong>Z</strong> 600.0</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>2,408</td>
<td>2,408</td>
<td>2,135</td>
<td>273</td>
<td>2,408</td>
</tr>
</tbody>
</table>

- Calculate the missing coefficients $V$, $W$, $X$, and $Y$ in Regression 4 (just the coefficients, not the standard errors) and $Z$, the missing $RSS$, giving an explanation of your computations.

- Give an interpretation of the coefficient of $BLACK$ in Regression 2.

- Perform an $F$ test of the joint explanatory power of $BLACK$, $SB$, $EB$, and $MB$ in Regression 5.

- Explain whether it is possible to relate the $F$ test in part (c) to a Chow test based on Regressions 1, 3, and 4.

- Give an interpretation of the coefficients of $BLACK$ and $MB$ in Regression 5.

- Explain whether a simple $t$ test on the coefficient of $BLACK$ in Regression 2 is sufficient to show that the wage equations are different for blacks and ‘whites’.

**A5.13** As part of a workshop project, four students are investigating the effects of ethnicity and sex on earnings using data for the year 2002 in the National Longitudinal Survey of Youth 1979–. They all start with the same basic specification:

$$\log Y = \beta_1 + \beta_2 S + \beta_3 EXP + u$$

where $Y$ is hourly earnings, measured in dollars, $S$ is years of schooling completed, and $EXP$ is years of work experience. The sample contains 123 black males, 150 black females, 1,146 white males, and 1,127 white females. (All respondents were either black or white. The Hispanic subsample was dropped.) The output from fitting this basic specification is shown in column 1 of the table (standard errors in
5.3. Additional exercises

parentheses; \(RSS\) is residual sum of squares, \(n\) is the number of observations in the regression.

<table>
<thead>
<tr>
<th></th>
<th>Basic</th>
<th>Student C</th>
<th>Student D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>(S)</td>
<td>0.126</td>
<td>0.121</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>(EXP)</td>
<td>0.040</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>(MALE)</td>
<td>—</td>
<td>0.277</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.020)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>(BLACK)</td>
<td>—</td>
<td>—0.144</td>
<td>—0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.032)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>(MALEBLACK)</td>
<td>—</td>
<td>—</td>
<td>—0.290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.063)</td>
</tr>
<tr>
<td>constant</td>
<td>0.376</td>
<td>0.459</td>
<td>0.447</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.076)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.285</td>
<td>0.341</td>
<td>0.346</td>
</tr>
<tr>
<td>(RSS)</td>
<td>659</td>
<td>608</td>
<td>603</td>
</tr>
<tr>
<td>(n)</td>
<td>2,546</td>
<td>2,546</td>
<td>2,546</td>
</tr>
</tbody>
</table>

Student A divides the sample into the four ethnicity/sex categories. He chooses white females as the reference category and fits a regression that includes three dummy variables \(BM\), \(WM\), and \(BF\). \(BM\) is 1 for black males, 0 otherwise; \(WM\) is 1 for white males, 0 otherwise, and \(BF\) is 1 for black females, 0 otherwise.

Student B simply fits the basic specification separately for the four ethnicity/sex subsamples.

Student C defines dummy variables \(MALE\), equal to 1 for males and 0 for females, and \(BLACK\), equal to 1 for blacks and 0 for whites. She also defines an interactive dummy variable \(MALEBLACK\) as the product of \(MALE\) and \(BLACK\). She fits a regression adding \(MALE\) and \(BLACK\) to the basic specification, and a further regression adding \(MALEBLACK\) as well. The output from these regressions is shown in columns 2 and 3 in the table.

Student D divides the sample into males and females and performs the regression for both sexes separately, using the basic specification. The output is shown in columns 4a and 4b. She also divides the sample into whites and blacks, and again runs separate regressions using the basic specification. The output is shown in columns 5a and 5b.

Reconstruction of missing output.

Students A and B left their output on a bus on the way to the workshop. This is why it does not appear in the table.

- State what the missing output of Student A would have been, as far as this is can be done exactly, given the results of Students C and D. (Coefficients, standard errors, \(R^2\), \(RSS\).)
5. Dummy variables

- Explain why it is not possible to reconstruct any of the output of Student B.

Tests of hypotheses.

The approaches of the students allowed them to perform different tests, given the output shown in the table and the corresponding output for Students A and B. Explain the tests relating to the effects of sex and ethnicity that could be performed by each student, giving a clear indication of the null hypothesis in each case. (Remember, all of them started with the basic specification (1), before continuing with their individual regressions.) In the case of $F$ tests, state the test statistic in terms of its components.

- Student A (assuming he had found his output)
- Student B (assuming he had found his output)
- Student C
- Student D.

If you had been participating in the project and had had access to the data set, what regressions and tests would you have performed?

5.4 Answers to the starred exercises in the textbook

5.2 The Stata output for Data Set 21 shows the result of regressing weight in 2004, measured in pounds, on height, measured in inches, first with a linear specification, then with a logarithmic one, in both cases including a dummy variable $MALE$, defined as in Exercise 5.1. Give an interpretation of the coefficients and perform appropriate statistical tests. See Box 5.1 for a guide to the interpretation of dummy variable coefficients in logarithmic regressions.

```
. reg WEIGHT04 HEIGHT MALE

Source | SS        df | MS        Number of obs =  500
--------+-----------+-----------+-----------------------------
Model   | 215264.34  2 | 107632.17 | F(  2,  497) =  90.45
Residual| 591434.61  497| 1190.00927 Prob > F =  0.0000
--------+-----------+-----------+-----------------------------
Total   | 806698.95  499| 1616.63116 R-squared =  0.2668
        |            |           | Adj R-squared =  0.2668
--------+-----------+-----------+-----------------------------
WEIGHT04 | Coef.  Std. Err.  t   P>|t|  [95% Conf. Interval]
--------+----------------+---------+---------+-----------------+-------------+-------------------------+-------------------------+
HEIGHT  | 4.424345      .5213809 8.49  0.000 3.399662   5.448727
MALE    | 7.702828      4.225065 1.82  0.069 -.598363  16.00402
_cons   | -136.9713     33.9953 -4.03  0.000 -203.7635 -70.17904
--------+----------------+---------+---------+-----------------+-------------+-------------------------+-------------------------+
```

94

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
5.4. Answers to the starred exercises in the textbook

```
. reg LGWT04 LGHEIGHT MALE

Source | SS       df  MS
-------------+---------------------------------------------------------------------
Model       | 8.12184709     2  4.06092355 Prob > F = 0.0000
Residual    | 18.4269077     497 .037076273 R-squared = 0.3059
-------------+--------------------------------------------------
Total       | 26.5487548     499 .053203918 Root MSE = .19255
-------------+--------------------------------------------------

| Coef.       | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|-------------+-----------+-------+-----+-------------------------------|
| LGHEIGHT    | 1.7814    | .1978798 | 9.00 | 0.0000 | 1.392616 - 2.170185 |
| MALE        | .0566894  | .0236289 | 2.40 | 0.017  | .0102645 - .1031142 |
| _cons       | -2.44656  | .8261259 | -2.96| 0.003  | -4.06969 - .8234307 |
```

Answer:

The first regression indicates that weight increase by 4.4 pounds for each inch of stature and that males tend to weigh 7.7 pounds more than females, controlling for height, but the coefficient of MALE is not significant. The second regression indicates that the elasticity of weight with respect to height is 1.78, and that males weigh 5.7 per cent more than females, the latter effect now being significantly different from zero at the 5 per cent level.

The null hypothesis that the elasticity is zero is not worth testing, except perhaps in a negative sense, for if the result were not highly significant there would have to be something seriously wrong with the model specification. Two other hypotheses might be of greater interest: the elasticity being equal to 1, weight growing proportionally with height, and the elasticity being equal to 3, all dimensions increasing proportionally with height. The t statistics are 4.27 and −8.37, respectively, so both hypotheses are rejected.

5.5 Suppose that the relationship:

\[ Y_i = \beta_1 + \beta_2 X_i + u_i \]

is being fitted and that the value of \( X \) is missing for some observations. One way of handling the missing values problem is to drop those observations. Another is to set \( X = 0 \) for the missing observations and include a dummy variable \( D \) defined to be equal to 1 if \( X \) is missing, 0 otherwise. Demonstrate that the two methods must yield the same estimates of \( \beta_1 \) and \( \beta_2 \). Write down an expression for RSS using the second approach, decompose it into the RSS for observations with \( X \) present and RSS for observations with \( X \) missing, and determine how the resulting expression is related to RSS when the missing value observations are dropped.

Answer:

Let the fitted model, with \( D \) included, be:

\[ \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 D_i. \]
5. Dummy variables

If $X$ is missing for observations $m + 1$ to $n$, then:

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 D_i))^2$$

$$= \sum_{i=1}^{m} (Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 D_i))^2 + \sum_{i=m+1}^{n} (Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 D_i))^2$$

$$= \sum_{i=1}^{m} (Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i))^2 + \sum_{i=m+1}^{n} (Y_i - (\hat{\beta}_1 + \hat{\beta}_3))^2.$$

The normal equation for $\hat{\beta}_3$ will yield:

$$\hat{\beta}_3 = \hat{\beta}_1 - \bar{Y}_{\text{missing}}$$

where $\bar{Y}_{\text{missing}}$ is the mean value of $Y$ for those observations for which $X$ is missing. This relationship means that $\hat{\beta}_1$ and $\hat{\beta}_2$ may be chosen so as to minimise the first term in $RSS$. This, of course, is $RSS$ for the regression omitting the observations for which $X$ is missing, and hence $\hat{\beta}_1$ and $\hat{\beta}_2$ will be the same as for that regression.

5.7

```
. reg LGEARN EDUCPROF EDUCPHD EDUCMAST EDUCBA EDUCAA EDUCGED EDUCDO EXP MALE

Source | SS       df       MS           Number of obs = 500
--------+------------------ ------------- --------- ------------- -------------
Model    | 34.231898 8 4.27898724 Prob > F = 0.0000
Residual | 118.367322 491 .241073975 R-squared = 0.2243
-----------+------------------------------ Adj R-squared = 0.2117
Total     | 152.59922 499 .30581006 Root MSE = .49099
-----------+------------------------------
LGEARN    | Coef.  Std. Err.     t    P>|t|    [95% Conf. Interval]
-----------+------------------------+-------------------+---------------------------+--------------
EDUCPROF  | 1.233278 .1920661 6.42 0.000 .8559049 1.610651
EDUCPHD   | (dropped)                |                    |                          |
EDUCMAST  | .7442879 .0875306 8.50 0.000 .5723071 .9162686
EDUCBA    | .3144576 .0578615 5.43 0.000 .2007709 .4281443
EDUCAA    | .2076079 .084855 2.45 0.015 .0408843 .3743316
EDUCGED   | -.2000523 .0886594 -2.26 0.024 -.374251 -.0258537
EDUCDO    | -.2216305 .132202 -1.68 0.094 -.4813819 .038121
EXP       | .0261946 .0085959 3.05 0.002 .0093054 .0430839
MALE      | .1756002 .0445659 3.94 0.000 .0880369 .2631636
_cons     | 2.385391 .0804166 29.66 0.000 2.227388 2.543394
```

The Stata output shows the result of a semilogarithmic regression of earnings on highest educational qualification obtained, work experience, and the sex of the respondent, the educational qualifications being a professional degree, a PhD (no respondents in this sample), a Master’s degree, a Bachelor’s degree, an Associate of Arts degree, the GED certification, and no qualification (high school drop-out). The high school diploma was the reference category. Provide an interpretation of the coefficients and perform $t$ tests.
Answer:

The regression results indicate that those with professional degrees earn 123 per cent more than high school graduates, or 243 per cent more if calculated as $100(e^{1.233} - 1)$, the coefficient being significant at the 0.1 per cent level. There was no respondent with a PhD in this subsample. For the other qualifications the corresponding figures are:

- Master’s: 74.4, 110.4, 0.1 per cent.
- Bachelor’s: 31.4, 36.9, 0.1 per cent.
- Associate’s: 20.8, 23.1, 5 per cent.
- GED: −20.0, −18.1, 5 per cent.
- Drop-out: −22.2, −19.9, 5 per cent, using a one-sided test, as seems reasonable.

Males earn 17.6 per cent (19.2 per cent) more than females, and every year of work experience increases earnings by 2.6 per cent. The coefficient of those with a professional degree should be treated cautiously since there were only seven such individuals in the subsample (EAWE 21). For the other categories the numbers of observations were: Master’s 42; Bachelor’s 168; Associate’s 44; High school diploma 187; GED 37; and drop-out 15.

5.8 Given a hierarchical classification such as that of educational qualifications in Exercise 5.7, some researchers unthinkingly choose the bottom category as the omitted category. In the case of Exercise 5.7, this would be EDUCDO, the high school drop-outs. Explain why this procedure may be undesirable (and, in the case of Exercise 5.7, definitely would not be recommended).

Answer:

The use of drop-outs as the reference category would make the tests of the coefficients of the other categories of little interest. If one wishes to evaluate the earnings premium for a bachelor’s or associate’s degree, it is much more sensible to use high school diploma as the benchmark. There is also the consideration that the drop-out category is tiny and unrepresentative.

5.16 Column (1) of the table shows the result of regressing WEIGHT04 on HEIGHT, MALE, and ethnicity dummy variables, using EAWE Data Set 21. The omitted ethnicity category was ETHWHITE. Column (2) shows in abstract the result of the same regression, using ETHBLACK as the omitted ethnicity category instead of ETHWHITE. As far as this is possible, determine the numbers represented by the letters.
5. Dummy variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEIGHT</td>
<td>4.45 A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>(0.53)</td>
<td>(B)</td>
</tr>
<tr>
<td>MALE</td>
<td>7.68 C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>(4.26)</td>
<td>(D)</td>
</tr>
<tr>
<td>ETHBLACK</td>
<td>4.08</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(4.52)</td>
<td></td>
</tr>
<tr>
<td>ETHHISP</td>
<td>0.07 E</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>(4.90)</td>
<td>(F)</td>
</tr>
<tr>
<td>ETHWHITE</td>
<td>—</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(H)</td>
</tr>
<tr>
<td>constant</td>
<td>−139.41 I</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>(34.64)</td>
<td>(J)</td>
</tr>
<tr>
<td>R²</td>
<td>0.27 K</td>
<td>K</td>
</tr>
<tr>
<td>RSS</td>
<td>590,443</td>
<td>L</td>
</tr>
<tr>
<td>n</td>
<td>500 500</td>
<td></td>
</tr>
</tbody>
</table>

Answer:

The parts of the output unrelated to the dummy variables will not be affected, so A, B, C, D, K, and L are as in column (1). G = −4.08 and H = 4.52. E = 0.07 − 4.08 = −4.01. I = −139.41 + 4.08 = −135.33. F and J cannot be determined.

5.19 Is the effect of education on earnings different for members of a union? In the output below, COLLBARG is a dummy variable defined to be 1 for workers whose wages are determined by collective bargaining and 0 for the others. SBARG is a slope dummy variable defined as the product of S and COLLBARG. Provide an interpretation of the regression coefficients, comparing them with those in Exercise 5.10, and perform appropriate statistical tests.
5.4. Answers to the starred exercises in the textbook

```
. gen SBARG=S*COLLBARG
. reg LGEARN S EXP MALE COLLBARG SBARG
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>29.699</td>
<td>5</td>
<td>5.939</td>
<td>F( 5, 494) = 23.88</td>
</tr>
<tr>
<td>Residual</td>
<td>122.900</td>
<td>494</td>
<td>.249</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>152.599</td>
<td>499</td>
<td>.306</td>
<td>R-squared = 0.1946</td>
</tr>
</tbody>
</table>

| LGEARN | Coef.  | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------|--------|-----------|------|-----|----------------------|
| S      | .09376 | .010815   | 8.66 | 0.000 | .072426 -.1149241    |
| EXP    | .04230 | .0094148  | 4.49 | 0.000 | .0238037 .0607995    |
| MALE   | .17135 | .0453584  | 3.78 | 0.000 | .0822295 .2604679    |
| COLLBARG | .29828 | .357371   | 0.83 | 0.404 | -.4038769 1.000441   |
| SBARG  | -.00260 | .0226557  | -0.12| 0.908 | -.0471205 .0419064  |
| _cons  | 1.03478 | .2049246  | 5.05 | 0.000 | .6321502 1.437413    |

**Answer:**
In this specification, the coefficient of $S$ is an estimate of the effect of schooling on the earnings of those whose earnings are not subject to collective bargaining (henceforward, for short, unionised workers, though obviously the category includes some who do not actually belong to unions), and the coefficient of $SBARG$ is the extra effect in the case of those whose earnings are. One might have anticipated a negative coefficient, since seniority and skills are often thought to be more important than schooling for the earnings of union workers, but in fact there is no significant difference.

5.23 Column (1) of the table shows the result of regressing $HOURS$, hours worked per week, on $S$, $MALE$, and $MALES$ using $EAWE$ Data Set 21. $MALES$ is defined as the product of $MALE$ and $S$. Provide an interpretation of the coefficients.

Column (2) gives the output in abstract when $FEMALE$ is used instead of $MALE$ and $FEMALES$ instead of $MALES$. $FEMALES$ is the product of $FEMALE$ and $S$. As far as this is possible, determine the numbers represented by the letters.
5. Dummy variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>0.79</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(B)</td>
</tr>
<tr>
<td>$MALE$</td>
<td>14.00</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(4.99)</td>
<td></td>
</tr>
<tr>
<td>$FEMALE$</td>
<td>—</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>(D)</td>
<td></td>
</tr>
<tr>
<td>$MALES$</td>
<td>−0.69</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td></td>
</tr>
<tr>
<td>$FEMALES$</td>
<td>—</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>(F)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>25.56</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(3.71)</td>
<td>(H)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.05</td>
<td>I</td>
</tr>
<tr>
<td>$RSS$</td>
<td>49,384</td>
<td>J</td>
</tr>
<tr>
<td>$n$</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

Answer:
The coefficient of $MALE$ indicates that a male with no schooling works 14 hours longer than a similar female. The coefficient of $S$ indicates that a female works an extra 0.79 hours per year of schooling. For males, the corresponding figure would be 0.10 hours, taking account of the interactive effect.

$A = 0.79 - 0.69 = 0.10$. $C = -14.00$. $D = 4.99$. $E = 0.69$.

$G = 25.56 + 14.00 = 39.56$. $I$ and $J$ are not affected. $B$, $F$ and $H$ cannot be determined.

5.29 The first paragraph of Section 5.4 used the words ‘satisfactory’ and ‘better’. Such intuitive terms have no precise meaning in econometrics. What ideas were they trying to express?

Answer:
The Chow test is effectively an $F$ test of the joint explanatory power of a full set of dummy variables. If the joint explanatory power is significant, this implies that the model is misspecified if they are omitted. In this sense, it is ‘better’ to include them.

5.5 Answers to the additional exercises

A5.1 As was to be expected, the coefficient of $HEIGHT$ falls with the addition of $MALE$ to the specification and is no longer significant. However, the coefficient of $MALE$ is not significant, either. This is because $MALE$ and $HEIGHT$ are sufficiently correlated (correlation coefficient 0.71) to give rise to a problem of multicollinearity.
The regression indicates that, controlling for total household expenditure per capita and size of household, non-whites spend 2.0 per cent less per year than whites on food consumed at home. However, the effect is not significant. The coefficients of \( \text{LGEXPPC} \) and \( \text{LGSIZE} \) are not affected by the introduction of the dummy variable.

Summarising the effects for all the categories of expenditure, one finds:

- Positive, significant at the 1 per cent level: \( \text{HOUS, LOCT, PERS} \).
- Positive, significant at the 5 per cent level: \( \text{FOOT, TELE} \).
- Negative, significant at the 1 per cent level: \( \text{HEAL, TOB} \).
- Not significant: the rest.

Under the hypothesis that non-whites tend to live in urban areas, some of these effects may have more to do with residence than ethnicity – for example, the positive effect on \( \text{LOCT} \). The results for all the categories are shown in the table.
5. Dummy variables

<table>
<thead>
<tr>
<th>Dependent variable LGCATPC</th>
<th>LGEXPPC</th>
<th>LGSIZE</th>
<th>NONWHITE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>( \hat{\beta}_2 )</td>
<td>s.e.(( \hat{\beta}_2 ))</td>
</tr>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>1.078</td>
<td>0.033</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>0.843</td>
<td>0.024</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>0.927</td>
<td>0.055</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>1.231</td>
<td>0.101</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>0.475</td>
<td>0.012</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>0.879</td>
<td>0.016</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>0.583</td>
<td>0.010</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>0.404</td>
<td>0.031</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.826</td>
<td>0.104</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>0.676</td>
<td>0.013</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>0.773</td>
<td>0.023</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>1.001</td>
<td>0.016</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>0.470</td>
<td>0.050</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.418</td>
<td>0.061</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.725</td>
<td>0.094</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>0.834</td>
<td>0.020</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>0.760</td>
<td>0.034</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>0.465</td>
<td>0.049</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>0.642</td>
<td>0.013</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.384</td>
<td>0.049</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>0.552</td>
<td>0.037</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>0.639</td>
<td>0.031</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.691</td>
<td>0.084</td>
</tr>
</tbody>
</table>

A5.3

. reg LGFDHOPC LGEXPPC LGSIZE EDUCBA EDUCSC EDUCDO;

The dummies have been defined with high school graduate as the reference category. Their coefficients indicate a significant negative association between level
of education and expenditure on food consumed at home, controlling for expenditure per person and the size of the household. The finding does not shed light on the reason for the negative association. Possibly those with greater education tend to eat less. There is also a negative association between level of education and expenditure on tobacco.

<table>
<thead>
<tr>
<th>Category</th>
<th>ADM</th>
<th>CLOT</th>
<th>DOM</th>
<th>EDUC</th>
<th>ELEC</th>
<th>FDAW</th>
<th>FDHO</th>
<th>FOOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGEXPPC</td>
<td>1.049</td>
<td>0.832</td>
<td>0.040</td>
<td>1.132</td>
<td>0.541</td>
<td>0.882</td>
<td>0.627</td>
<td>0.307</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.026)</td>
<td>(0.058)</td>
<td>(0.107)</td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.010)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>LGSIZE</td>
<td>−0.060</td>
<td>0.141</td>
<td>0.386</td>
<td>−0.448</td>
<td>−0.334</td>
<td>−0.214</td>
<td>−0.066</td>
<td>−0.560</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.033)</td>
<td>(0.076)</td>
<td>(0.139)</td>
<td>(0.017)</td>
<td>(0.022)</td>
<td>(0.013)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>EDUCBA</td>
<td>0.239</td>
<td>0.072</td>
<td>0.187</td>
<td>0.601</td>
<td>−0.319</td>
<td>0.011</td>
<td>−0.164</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.047)</td>
<td>(0.113)</td>
<td>(0.214)</td>
<td>(0.024)</td>
<td>(0.031)</td>
<td>(0.019)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>EDUCSC</td>
<td>0.193</td>
<td>0.055</td>
<td>−0.035</td>
<td>0.320</td>
<td>−0.114</td>
<td>−0.014</td>
<td>−0.070</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.048)</td>
<td>(0.120)</td>
<td>(0.218)</td>
<td>(0.024)</td>
<td>(0.032)</td>
<td>(0.019)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>EDUCDO</td>
<td>0.000</td>
<td>0.035</td>
<td>0.075</td>
<td>0.133</td>
<td>0.055</td>
<td>0.065</td>
<td>0.102</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.062)</td>
<td>(0.163)</td>
<td>(0.320)</td>
<td>(0.031)</td>
<td>(0.044)</td>
<td>(0.025)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>R²</td>
<td>0.334</td>
<td>0.240</td>
<td>0.160</td>
<td>0.323</td>
<td>0.384</td>
<td>0.461</td>
<td>0.444</td>
<td>0.281</td>
</tr>
<tr>
<td>F</td>
<td>281.8</td>
<td>284.5</td>
<td>63.3</td>
<td>52.8</td>
<td>724.7</td>
<td>871.5</td>
<td>1,012.4</td>
<td>142.2</td>
</tr>
<tr>
<td>n</td>
<td>2,815</td>
<td>4,500</td>
<td>1,661</td>
<td>461</td>
<td>5,828</td>
<td>5,102</td>
<td>6,334</td>
<td>1,827</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>FURN</th>
<th>GASO</th>
<th>HEAL</th>
<th>HOUS</th>
<th>LIFE</th>
<th>LOCT</th>
<th>MAPP</th>
<th>PERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGEXPPC</td>
<td>0.875</td>
<td>0.719</td>
<td>0.822</td>
<td>0.960</td>
<td>0.468</td>
<td>0.464</td>
<td>0.728</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td>(0.014)</td>
<td>(0.024)</td>
<td>(0.053)</td>
<td>(0.017)</td>
<td>(0.067)</td>
<td>(0.100)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>LGSIZE</td>
<td>−0.228</td>
<td>0.015</td>
<td>−0.279</td>
<td>−0.155</td>
<td>−0.453</td>
<td>−0.394</td>
<td>−0.268</td>
<td>−0.213</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.018)</td>
<td>(0.031)</td>
<td>(0.021)</td>
<td>(0.066)</td>
<td>(0.086)</td>
<td>(0.124)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>EDUCBA</td>
<td>−0.345</td>
<td>−0.215</td>
<td>−0.222</td>
<td>0.190</td>
<td>0.045</td>
<td>−0.325</td>
<td>−0.058</td>
<td>−0.043</td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.026)</td>
<td>(0.044)</td>
<td>(0.031)</td>
<td>(0.087)</td>
<td>(0.143)</td>
<td>(0.171)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>EDUCSC</td>
<td>−0.363</td>
<td>−0.010</td>
<td>−0.152</td>
<td>0.127</td>
<td>−0.031</td>
<td>−0.404</td>
<td>−0.375</td>
<td>−0.002</td>
</tr>
<tr>
<td></td>
<td>(0.177)</td>
<td>(0.025)</td>
<td>(0.045)</td>
<td>(0.030)</td>
<td>(0.089)</td>
<td>(0.146)</td>
<td>(0.167)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>EDUCDO</td>
<td>0.071</td>
<td>−0.004</td>
<td>0.002</td>
<td>0.084</td>
<td>0.190</td>
<td>0.558</td>
<td>−0.150</td>
<td>−0.087</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.034)</td>
<td>(0.061)</td>
<td>(0.039)</td>
<td>(0.134)</td>
<td>(0.167)</td>
<td>(0.214)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>R²</td>
<td>0.206</td>
<td>0.373</td>
<td>0.276</td>
<td>0.471</td>
<td>0.156</td>
<td>0.154</td>
<td>0.219</td>
<td>0.388</td>
</tr>
<tr>
<td>F</td>
<td>24.9</td>
<td>679.8</td>
<td>366.1</td>
<td>1,105.8</td>
<td>46.0</td>
<td>25.0</td>
<td>22.1</td>
<td>483.4</td>
</tr>
<tr>
<td>n</td>
<td>487</td>
<td>5,710</td>
<td>4,802</td>
<td>6,223</td>
<td>1,253</td>
<td>692</td>
<td>399</td>
<td>3,817</td>
</tr>
</tbody>
</table>
A5.4 For *FDHO*, RSS was 1,988.4 without the education dummy variables and 1,946.0 with them. 3 degrees of freedom were consumed when adding them, and 6334 − 6 = 6328 degrees of freedom remained after they had been added. The $F$ statistic is, therefore:

$$F(3, 6328) = \frac{(1988.4 - 1946.0)/3}{1946.0/6328} = 45.98.$$  

The critical value of $F(3, 1000)$ at the 5 per cent level is 2.61. The critical value of $F(3, 6328)$ must be lower. Hence we reject the null hypothesis that the dummy variables have no explanatory power (that is, that all their coefficients are jointly equal to zero).
5.5 Answers to the additional exercises

<table>
<thead>
<tr>
<th>TEXT</th>
<th>992</th>
<th>1,032.9</th>
<th>1,031.8</th>
<th>0.36</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>873.4</td>
<td>813.5</td>
<td>28.18</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>2,828.3</td>
<td>2,826.7</td>
<td>0.48</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>792.8</td>
<td>790.6</td>
<td>0.48</td>
</tr>
</tbody>
</table>

The results for all the categories of expenditure have not been tabulated but are easily summarised:

- The analysis of variance in the upper half of the output is unaffected.
- The results for variables other than the dummy variables are unaffected.
- The results for `EDUCHSD` are identical to those for `EDUCDO` in the first regression, except for a change of sign in the coefficient, the $t$ statistic, and the limits of the confidence interval.
- The constant is equal to the old constant plus the coefficient of `EDUCDO` in the first regression.
- The coefficients of the other dummy variables are equal to their values in the first regression minus the coefficient of EDUCDO in the first regression.
- One substantive change is in the standard errors of `EDUCIC` and `EDUCCO`, caused by the fact that the comparisons are now between these categories and `EDUCDO`, not `EDUCHSD`.
- The other is that the $t$ statistics are for the new comparisons, not the old ones.

The results for all categories of expenditure have not been tabulated but are easily summarised:

- The analysis of variance in the upper half of the output is unaffected.
- The results for variables other than the dummy variables are unaffected.
- The results for `EDUCHSD` are identical to those for `EDUCDO` in the first regression, except for a change of sign in the coefficient, the $t$ statistic, and the limits of the confidence interval.
- The constant is equal to the old constant plus the coefficient of `EDUCDO` in the first regression.
- The coefficients of the other dummy variables are equal to their values in the first regression minus the coefficient of EDUCDO in the first regression.
- One substantive change is in the standard errors of `EDUCIC` and `EDUCCO`, caused by the fact that the comparisons are now between these categories and `EDUCDO`, not `EDUCHSD`.
- The other is that the $t$ statistics are for the new comparisons, not the old ones.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
5. Dummy variables

A5.6  **Explain verbally why the estimates of the coefficient of GOV are different in regressions (1) and (2).**

The second specification indicates that earnings are positively related to schooling and negatively related to working in the government sector. S has a significant coefficient in (2) and therefore ought to be in the model. If S is omitted from the specification the estimate of the coefficient of GOV will be biased upwards because schooling is positively correlated with working in the government sector. (We are told in the question that government workers on average have an extra year of schooling.) The bias is sufficiently strong to make the negative coefficient disappear.

**Explain the difference in the estimates of the coefficient of GOV in regressions (2) and (3).**

The coefficient of GOV in the third regression is effectively a linear function of S: 
\[ 0.726 - 0.059S \]. The coefficient of the GOV intercept dummy is therefore an estimate of the extra earnings of a government worker with no schooling. The premium disappears for S = 12 and becomes negative for higher values of S. The second regression does not take account of the variation of the coefficient of GOV with S and hence yields an average effect of GOV. The average effect was negative since only a small minority of government workers had fewer than 12 years of schooling.

*The correlation between GOV and SGOV was 0.977. Explain the variations in the standard error of the coefficient of GOV in the three regressions.*

The standard error in the first regression is meaningless given severe omitted variable bias. For comparing the standard errors in (2) and (3), it should be noted that the same problem in principle applies in (2), given that the coefficient of SGOV in (3) is highly significant. However, part of the reason for the huge increase must be the high correlation between GOV and SGOV.

A5.7  1. The dummy variable allows the slope coefficient to be different for developing and developed countries. From equation (1) one may derive the following relationships:

- developed countries: \( \hat{\epsilon} = -1.45 + 0.19x \)
- developing countries: \( \hat{\epsilon} = -1.45 + 0.19x + 0.78x = -1.45 + 0.97x. \)
5.5. Answers to the additional exercises

2. The inclusion of $D$ would allow the intercept to be different for the two types of country. If the model was written as:

$$ e = \beta_1 + \beta_2 x + \delta D + \lambda D x + u $$

the implicit relationships for the two types of country would be:

developed countries  
$$ e = \beta_1 + \beta_2 x + u $$

developing countries  
$$ e = \beta_1 + \beta_2 x + \delta + \lambda x + u $$

$$ = (\beta_1 + \delta) + (\beta_2 + \lambda)x + u. $$

3. When the specification includes both an intercept dummy and a slope dummy, the coefficients for the two categories will be the same as in the separate regressions (2) and (3). Hence the intercept and coefficient of $x$ will be the same as in the regression for the reference category, regression (3), and the coefficients of the dummies will be such that they modify the intercept and slope coefficient so that they are equal to their counterparts in regression (4):

$$ \hat{e} = -2.74 + 0.50x + 1.89D + 0.28xD. $$

Since the coefficients are the same, the overall fit for this regression will be the same as that for regressions (2) and (3). Hence $RSS = 18.63 + 25.23 = 43.86.$
5. Dummy variables

The $t$ statistic for the coefficient of $x$ will be the square root of the $F$ statistic for the test of the marginal explanatory power of $D$ when it is included in the equation. The $F$ statistic is:

$$F(1, 46) = \frac{(50.23 - 43.86)/1}{43.86/46} = 6.6808.$$ 

The $t$ statistic is therefore 2.58.

4. One method is to use a Chow test comparing $RSS$ for the pooled regression, regression (2), with the sum of $RSS$ regressions (3) and (4):

$$F(2, 46) = \frac{(121.61 - 43.86)/2}{43.86/46} = 40.8.$$ 

The critical value of $F(2, 40)$ at the 0.1 per cent significance level is 8.25. The critical value of $F(2, 46)$ must be lower. Hence the null hypothesis that the coefficients are the same for developed and developing countries is rejected.

We should also consider $t$ tests on the coefficients of $D$ and $xD$. We saw in (3) that the $t$ statistic for the coefficient of $D$ was 2.58, so we would reject the null hypothesis of no intercept shift at the 5 per cent level, and nearly at the 1 per cent level. We do not have enough information to derive the $t$ statistic for $xD$. We would not perform a $t$ test on the coefficient of $xD$ in regression (1) because that regression is clearly misspecified.

### Chow test

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$RSS$</th>
<th>$RSS$</th>
<th>$RSS$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>COLLEGE = 0</td>
<td>COLLEGE = 1</td>
<td></td>
</tr>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>3,945.2</td>
<td>789.5</td>
<td>3,129.9</td>
<td>6.15</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>5,766.1</td>
<td>1,837.9</td>
<td>3,913.8</td>
<td>3.77</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>4,062.5</td>
<td>1,048.5</td>
<td>2,984.0</td>
<td>4.10</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>1,380.1</td>
<td>278.0</td>
<td>1,087.0</td>
<td>2.05</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>2,636.3</td>
<td>962.6</td>
<td>1,594.6</td>
<td>60.02</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>3,369.1</td>
<td>1,114.8</td>
<td>2,251.7</td>
<td>1.32</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>1,988.4</td>
<td>751.9</td>
<td>1,205.3</td>
<td>33.63</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>1,373.5</td>
<td>513.1</td>
<td>858.5</td>
<td>0.82</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>913.9</td>
<td>238.7</td>
<td>662.1</td>
<td>2.32</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>2,879.3</td>
<td>1,043.2</td>
<td>1,811.7</td>
<td>16.27</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>6,062.5</td>
<td>2,211.7</td>
<td>3,796.6</td>
<td>14.42</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>4,825.6</td>
<td>2,234.6</td>
<td>2,565.6</td>
<td>10.55</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>1,559.2</td>
<td>424.0</td>
<td>1,119.6</td>
<td>4.20</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>1,075.1</td>
<td>283.3</td>
<td>769.3</td>
<td>4.88</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>576.8</td>
<td>205.6</td>
<td>367.5</td>
<td>0.84</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>3,002.2</td>
<td>918.5</td>
<td>2,081.1</td>
<td>1.10</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>2,892.1</td>
<td>752.6</td>
<td>2,129.1</td>
<td>2.75</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>1,148.9</td>
<td>342.9</td>
<td>802.1</td>
<td>1.18</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>3,055.1</td>
<td>1,132.8</td>
<td>1,903.2</td>
<td>12.10</td>
</tr>
</tbody>
</table>
5.5. Answers to the additional exercises

<table>
<thead>
<tr>
<th>TEXT</th>
<th>992</th>
<th>1,032.9</th>
<th>278.0</th>
<th>754.1</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>873.4</td>
<td>351.3</td>
<td>476.8</td>
<td>20.91</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>2,828.3</td>
<td>862.5</td>
<td>1,964.2</td>
<td>0.46</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>792.8</td>
<td>114.2</td>
<td>675.6</td>
<td>0.66</td>
</tr>
</tbody>
</table>

For FDHO, RSS for the logarithmic regression without college in Exercise A4.2 was 1,988.4. When the sample is split, RSS for COLLEGE = 0 is 751.9 and for COLLEGE = 1 is 1,205.3. Three degrees of freedom are consumed because the coefficients of LGEYPPPC and LGSIZE and the constant have to be estimated twice. The number of degrees of freedom remaining after splitting the sample is 6334 – 6 = 6328. Hence the F statistic is:

\[ F(3, 6328) = \frac{(1988.4 - (751.9 + 1205.3))/3}{(751.9 + 1205.3)/6328} = 33.63. \]

The critical value of F(3, 1000) at the 1 per cent level is 2.62 and so we reject the null hypothesis of no difference in the expenditure functions at that significance level. The results for all the categories are shown in the table.

A5.9

```plaintext
. gen LEXPCOL = LGEYPPPC*COLLEGE
. gen LSIZECOL = LGSIZE*COLLEGE
. reg LGFDHOPC LGEYPPPC LGSIZE COLLEGE LEXPCOL LSIZECOL
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F( 5, 6328) = 999.36</td>
</tr>
<tr>
<td>Model</td>
<td>1545.47231</td>
<td>5</td>
<td>309.094462</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>1957.1997</td>
<td>6328</td>
<td>.309291987</td>
<td>R-squared = 0.4412</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.4408</td>
</tr>
<tr>
<td>Total</td>
<td>3502.67201</td>
<td>6333</td>
<td>.553082585</td>
<td>Root MSE = .55614</td>
</tr>
</tbody>
</table>

| LGFDHOPC | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|-------|-----|------------------------|
| LGEYPPPC | .648295 | .0171599 | 37.78 | 0.000 | .6146559 - .6819342 |
| LGSIZE   | -.0559735 | .0216706 | -2.58 | 0.010 | -.0984552 -.0134917 |
| COLLEGE  | .3046012 | .1760486 | 1.73  | 0.084 | .0405137 .6497161 |
| LEXPCOL  | -.0558931 | .0211779 | -2.64 | 0.008 | -.104909 -.0134772 |
| LSIZECOL | -.0198021 | .0274525 | -0.72 | 0.471 | -.0736182 .034014 |
| _cons    | .7338499 | .1403321 | 5.23 | 0.000 | .4587514 1.008948 |

The example output is for FDHO. In Exercise A4.2, RSS was 1,988.4 for the same regression without the dummy variables. To perform the F test of the explanatory power of the intercept dummy variable and the two slope dummy variables as a group, we evaluate whether RSS for this regression is significantly lower. RSS has fallen from 1,988.4 to 1,957.2. 3 degrees of freedom are consumed by adding the dummy variables, and 6334 – 6 = 6328 degrees of freedom remain after adding the dummy variables. The F statistic is therefore:

\[ F(3, 6328) = \frac{(1988.4 - 1957.2)/3}{1957.2/6328} = 33.63. \]

This is highly significant. This F test is, of course, equivalent to the Chow test in the previous exercise. One possible explanation was offered there. The present
5. Dummy variables

Regression suggests another. The slope dummy variable \( LGEXPCOL \) has a significant negative coefficient, implying that the elasticity falls as income rises. This is plausible for a basic necessity such as food.

A5.10 (a) You should fit models such as:

\[
LGEARN = \beta_1 + \beta_2 S + \beta_3 ASVABC + \beta_4 MALE + \beta_5 ETHBLACK + \beta_6 ETHHISP + u
\]

separately for the private and government sectors. To investigate discrimination, for each sector \( t \) tests should be performed on the coefficients of \( MALE, ETHBLACK, \) and \( ETHHISP \) and an \( F \) test on the joint explanatory power of \( ETHBLACK \) and \( ETHHISP \).

(b) You should combine the earnings functions for the two sectors, while still allowing their parameters to differ, by fitting a model such as:

\[
LGEARN = \beta_1 + \beta_2 S + \beta_3 ASVABC + \beta_4 MALE + \beta_5 ETHBLACK + \beta_6 ETHHISP + \delta_1 GOV + \delta_2 GOVS + \delta_3 GOVASV + \delta_4 GOVMALE + \delta_5 GOVBLACK + \delta_6 GOVHISP + u
\]

where \( GOV \) is equal to 1 if the respondent works in the government sector and 0 otherwise, and \( GOVS, GOVASV, GOVMALE, GOVBLACK, \) and \( GOVHISP \) are slope dummy variables defined as the product of \( GOV \) and the respective variables. To investigate whether the level of discrimination is different in the two sectors, one should perform \( t \) tests on the coefficients of \( GOVMALE, GOVBLACK, \) and \( GOVHISP \) and an \( F \) test on the joint explanatory power of \( GOVBLACK \) and \( GOVHISP \).

A Chow test would not be appropriate because if it detected a significant difference in the earnings functions, this could be due to differences in the coefficients of \( S \) and \( ASVABC \) rather than the discrimination variables.

A5.11 Give an interpretation of the coefficients of \( S \) and \( SMALE \) in regression (5).

An extra year of schooling increases female earnings by 9.4 per cent. (Strictly, \( 100(e^{0.094} - 1) = 9.9 \) per cent.) For males, an extra year of schooling leads to an increase in earnings 0.5 per cent greater than for females, i.e. 9.9 per cent.

Give an interpretation of the coefficients of \( MALE \) in regressions (4) and (5).

(4): males earn 23.4 per cent more than females (controlling for other factors). (5): males with no schooling or work experience earn 11.7 per cent more than similar females.

The researcher hypothesises that the earnings function is different for males and females. Perform a test of this hypothesis using regression (4), and also using regressions (1) and (5).

Looking at regression (4), the coefficient of \( MALE \) is highly significant, indicating that the earnings functions are indeed different. Looking at regression (5), and comparing it with (1), the null hypothesis is that the coefficients of the male dummy variables in (5) are all equal to zero.

\[
F(3, 3236) = \frac{(714.6 - 672.5)/3}{672.5/3236} = 67.5.
\]
5.5. Answers to the additional exercises

The critical value of $F(3, 1000)$ at the 1 per cent level is 3.80. The corresponding critical value for $F(3, 3236)$ must be lower, so we reject the null hypothesis and conclude that the earnings functions are different.

**Explain the differences in the tests using regression (4) and using regressions (1) and (5).**

In regression (4) the coefficient of $MALE$ is highly significant. In regression (5) it is not. Likewise the coefficients of the slope dummies are not significant. This is (partly) due to the effect of multicollinearity. The male dummy variables are very highly correlated and as a consequence the standard error of the coefficient of $MALE$ is much larger than in regression (4). Nevertheless the $F$ test reveals that their joint explanatory power is highly significant.

**At a seminar someone suggests that a Chow test could shed light on the researcher’s hypothesis. Is this correct?**

Yes. Using regressions (1)–(3):

$$F(3, 3236) = \frac{(714.6 - (411.0 + 261.6))/3}{(411.0 + 261.6)/3236} = 67.4.$$  

The null hypothesis that the coefficients are the same for males and females is rejected at the 1 per cent level. The test is, of course, equivalent to the dummy variable test comparing (1) and (5).

**Explain which of (1), (4), and (5) would be your preferred specification.**

(4) seems best, given that the coefficients of $S$ and $EXP$ are fairly similar for males and females and that introducing the slope dummies causes multicollinearity. The $F$ statistic of their joint explanatory power is only 0.72, not significant at any significance level.

**Calculate the missing coefficients $V$, $W$, $X$, and $Y$ in Regression 4 (just the coefficients, not the standard errors) and $Z$, the missing RSS, giving an explanation of your computations.**

Since Regression 5 includes a complete set of black intercept and slope dummy variables, the basic coefficients will be the same as for a regression using the ‘whites’ only subsample and the coefficients modified by the dummies will give the counterparts for the blacks only subsample. Hence $V = 0.122 - 0.009 = 0.113$; $W = 0.033 - 0.006 = 0.027$; $X = 0.306 - 0.280 = 0.026$; and $Y = 0.411 + 0.205 = 0.616$. The residual sum of squares for Regression 5 will be equal to the sum of RSS for the ‘whites’ and blacks subsamples. Hence $Z = 600.0 - 555.7 = 44.3$.

**Give an interpretation of the coefficient of BLACK in Regression 2.**

It suggests that blacks earn 14.4 per cent less than whites, controlling for other characteristics.

**Perform an $F$ test of the joint explanatory power of BLACK, SB, EB, and MB in Regression 5.**

Write the model as:

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 EXP + \beta_4 MALE + \beta_5 BLACK + \beta_6 SB + \beta_7 EB + \beta_8 MB + u.$$
5. Dummy variables

The null hypothesis for the test is if \( H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0 \), and the alternative hypothesis is \( H_1 \): at least one coefficient different from 0. The \( F \) statistic is:

\[
F(4, 2400) = \frac{(610.0 - 600.0)/4}{600.0/2400} = \frac{2400}{240} = 10.0.
\]

This is significant at the 0.1 per cent level (critical value 4.65) and so the null hypothesis is rejected.

Explain whether it is possible to relate the \( F \) test in part (c) to a Chow test based on Regressions 1, 3, and 4.

The Chow test would be equivalent to the \( F \) test in this case.

Give an interpretation of the coefficients of \( BLACK \) and \( MB \) in Regression 5.

Re-write the model as:

\[
LGEARN = \beta_1 + \beta_2 S + \beta_3 EXP + \beta_4 MALE + (\beta_5 + \beta_6 S + \beta_7 EXP + \beta_8 MALE) BLACK + u.
\]

From this it follows that \( \beta_5 \) is the extra proportional earnings of a black, compared with a white, when \( S = EXP = MALE = 0 \). Thus the coefficient of \( BLACK \) indicates that a black female with no schooling or experience earns 20.5 per cent more than a similar white female. The interpretation of the coefficient of any interactive term requires care. Holding \( S = EXP = MALE = 0 \), the coefficients of \( MALE \) and \( BLACK \) indicate that black males will earn 30.6 + 20.5 = 51.1 per cent more than white females. The coefficient of \( MB \) modifies this estimate, reducing it by 28.0 per cent to 23.1 per cent.

Explain whether a simple \( t \) test on the coefficient of \( BLACK \) in Regression 2 is sufficient to show that the wage equations are different for blacks and whites.

Regression 2 is misspecified because it embodies the restriction that the effect of being black is the same for males and females, and that is contradicted by Regression 5. Hence any test is in principle invalid. However, the fact that the coefficient has a very high \( t \) statistic is suggestive that something associated with being black is affecting the wage equation.

A5.13 Reconstruction of missing output

Students A and B left their output on a bus on the way to the workshop. This is why it does not appear in the table.

State what the missing output of Student A would have been, as far as this can be done exactly, given the results of Students C and D. (Coefficients, standard errors, \( R^2 \), RSS.)

The output would be as for column (3) (coefficients, standard errors, \( R^2 \)), with the following changes:

- the row label \( MALE \) should be replaced with \( WM \)
- the row label \( BLACK \) should be replaced with \( BF \)
- the row label \( MALEBLACK \) should be replaced with \( BM \) and the coefficient for that row should be the sum of the coefficients in column (3): \( 0.308 - 0.011 - 0.290 = 0.007 \), and the standard error would not be known.
5.5. Answers to the additional exercises

Explain why it is not possible to reconstruct any of the output of Student B.

One could not predict the coefficients of either S or EXP in the four regressions performed by Student B. They will, except by coincidence, be different from any of the estimates of the other students because the coefficients for S and EXP in the other specifications are constrained in some way. As a consequence, one cannot predict exactly any part of the rest of the output, either.

Tests of hypotheses

- **Student A (assuming he had found his output)**
  
  Student A could perform tests of the differences in earnings between white males and white females, black males and white females, and black females and white females, through simple t tests on the coefficients of WM, BM, and BF.

  He could also test the null hypothesis that there are no sex/ethnicity differences with an F test, comparing RSS for his regression with that of the basic regression:
  
  $$F(3, 2540) = \frac{(922 - 603)/3}{603/2540}.$$  

  This would be compared with the critical value of F with 3 and 2,540 degrees of freedom at the significance level chosen and the null hypothesis of no sex/ethnicity effects would be rejected if the F statistic exceeded the critical value.

- **Student B (assuming he had found his output)**
  
  In the case of Student B, with four separate subsample regressions, candidates are expected say that no tests would be possible because no relevant standard errors would be available. We have covered Chow tests only for two categories. However, a four-category test could be performed, with:
  
  $$F(9, 2534) = \frac{(922 - X)/9}{X/2534},$$

  where RSS = 922 for the basic regression and X is the sum of RSS in the four separate regressions.

- **Student C**
  
  Student C could perform the same t tests and the same F test as Student A, with one difference: the t test of the difference between the earnings of black males and white females would not be available. Instead, the t statistic of MALEBLACK would allow a test of whether there is any interactive effect of being black and being male on earnings.

- **Student D**
  
  Student D could perform a Chow test to see if the wage equations of males and females differed:
  
  $$F(3, 2540) = \frac{(659 - (322 + 289))/3}{(322 + 289)/2540}.$$  

  RSS = 322 for males and 289 for females. This would be compared with the critical value of F with 3 and 2,540 degrees of freedom at the significance level.
5. Dummy variables

chosen and the null hypothesis of no sex/ethnicity effects would be rejected if
the $F$ statistic exceeded the critical value. She could also perform a
corresponding Chow test for blacks and whites:

$$F(3, 2540) = \frac{(659 - (609 + 44))/3}{(609 + 44)/2540}.$$

If you had been participating in the project and had had access to the data set, what
regressions and tests would you have performed?

The most obvious development would be to relax the sex/ethnicity restrictions on
the coefficients of $S$ and $EXP$ by including appropriate interaction terms. This
could be done by interacting these variables with the dummy variables defined by
Student A or those defined by Student C.
Chapter 6
Specification of regression variables

6.1 Overview

This chapter treats a variety of topics relating to the specification of the variables in a regression model. First there are the consequences for the regression coefficients, their standard errors, and $R^2$ of failing to include a relevant variable, and of including an irrelevant one. This leads to a discussion of the use of proxy variables to alleviate a problem of omitted variable bias. Next come $F$ and $t$ tests of the validity of a restriction, the use of which was advocated in Chapter 3 as a means of improving efficiency and perhaps mitigating a problem of multicollinearity. The chapter concludes by outlining the potential benefit to be derived from examining observations with large residuals after fitting a regression model.

6.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- derive the expression for the bias in an OLS estimator of a slope coefficient when the true model has two explanatory variables but the regression model has only one

- determine the likely direction of omitted variable bias, given data on the correlation between the explanatory variables

- explain the consequence of omitted variable bias for the standard errors of the coefficients and for $t$ tests and $F$ tests

- explain the consequences of including an irrelevant variable for the regression coefficients, their standard errors, and $t$ and $F$ tests

- explain how the regression results are affected by the substitution of a proxy variable for a missing explanatory variable

- perform an $F$ test of a restriction, stating the null hypothesis for the test

- perform a $t$ test of a restriction, stating the null hypothesis for the test.
6. Specification of regression variables

6.3 Additional exercises

A6.1 Does the omission of total household expenditure or household size give rise to omitted variable bias in your CES regressions?

Regress LGCATPC (1) on both LGEXPPC and LGSIZE, (2) on LGEXPPC only, and (3) on LGSIZE only. Assuming that (1) is the correct specification, analyse the likely direction of the bias in the estimate of the coefficient of LGEXPPC in (2) and that of LGSIZE in (3). Check whether the regression results are consistent with your analysis.

A6.2 A school has introduced an extra course of reading lessons for children starting school and a researcher is evaluating the impact of the course on the scores on a literacy test taken at the age of seven. In the first year of its implementation, those children whose surnames begin A–M are assigned to the extra course, while the rest have the normal curriculum. The researcher hypothesises that:

\[ Y = \beta_1 + \beta_2 D + \beta_3 A + u \]

where \( Y \) is the score on the literacy test, \( D \) is a dummy variable that is equal to 1 for those assigned to the extra course and 0 for the others, \( A \) is a measure of the cognitive ability of the child when starting school, and \( u \) is an iid (independently and identically distributed) disturbance term assumed to have a normal distribution. Unfortunately, the researcher has no data on \( A \). Using OLS (ordinary least squares), she fits the regression:

\[ \hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 D. \]

- Demonstrate that \( \hat{\beta}_2 \) is an unbiased estimator of \( \beta_2 \).
- A commentator says that the standard error of \( \hat{\beta}_2 \) will be invalid because an important variable, \( A \), has been omitted from the specification. The researcher replies that the standard error will remain valid if \( A \) can be assumed to have a normal distribution. Explain whether the commentator or the researcher is correct.
- Another commentator says that whether the distribution of \( A \) is normal or not makes no difference to the validity of the standard error. Evaluate this assertion.

A6.3 A researcher obtains data on household annual expenditure on books, \( B \), and annual household income, \( Y \), for 100 households. He hypothesises that \( B \) is related to \( Y \) and the average cognitive ability of adults in the household, \( IQ \), by the relationship:

\[ \log B = \beta_1 + \beta_2 \log Y + \beta_3 \log IQ + u \quad (A) \]

where \( u \) is a disturbance term that satisfies the regression model assumptions. He also considers the possibility that \( \log B \) may be determined by \( \log Y \) alone:

\[ \log B = \beta_1 + \beta_2 \log Y + u. \quad (B) \]
6.3. Additional exercises

He does not have data on IQ and decides to use average years of schooling of the adults in the household, S, as a proxy in specification (A). It may be assumed that Y and S are both nonstochastic. In the sample the correlation between log Y and log S is 0.86. He performs the following regressions: (1) log B on both log Y and log S, and (2) log B on log Y only, with the results shown in the table (standard errors in parentheses):

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log Y</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>log S</td>
<td>0.59</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>−6.89</td>
<td>−3.37</td>
</tr>
<tr>
<td></td>
<td>(2.28)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.29</td>
<td>0.27</td>
</tr>
</tbody>
</table>

- Assuming that (A) is the correct specification, explain, with a mathematical proof, whether you would expect the coefficient of log Y to be greater in regression (2).
- Assuming that (A) is the correct specification, describe the various benefits from using log S as a proxy for log IQ, as in regression (1), if log S is a good proxy.
- Explain whether the low value of $R^2$ in regression (1) implies that log S is not a good proxy.
- Assuming that (A) is the correct specification, provide an explanation of why the coefficients of log Y and log S in regression (1) are not significantly different from zero, using two-sided t tests.
- Discuss whether the researcher would be justified in using one-sided t tests in regression (1).
- Assuming that (B) is the correct specification, explain whether you would expect the coefficient of log Y to be lower in regression (1).
- Assuming that (B) is the correct specification, explain whether the standard errors in regression (1) are valid estimates.

A6.4 A researcher has the following data for the year 2012: $T$, annual total sales of cinema tickets per household, and $P$, the average price of a cinema ticket in the city. She believes that the true relationship is:

$$\log T = \beta_1 + \beta_2 \log P + \beta_3 \log Y + u$$

where $Y$ is average household income, but she lacks data on $Y$ and fits the regression (standard errors in parentheses):

$$\hat{\log T} = 13.74 + 0.17 \log P \quad R^2 = 0.01$$

$$(0.52) \quad (0.23)$$
6. Specification of regression variables

Explain analytically whether the slope coefficient is likely to be biased. You are
told that if the researcher had been able to obtain data on $Y$, her regression would
have been:

$$\log T = -1.63 - 0.48 \log P + 1.83 \log Y \quad R^2 = 0.44$$

(2.93) (0.21) (0.35)

You are also told that $Y$ and $P$ are positively correlated.

The researcher is not able to obtain data on $Y$ but, from local records, she is able
to obtain data on $H$, the average value of a house in each city, and she decides to
use it as a proxy for $Y$. She fits the following regression (standard errors in
parentheses):

$$\log T = -0.63 - 0.37 \log P + 1.69 \log H \quad R^2 = 0.36$$

(3.22) (0.22) (0.38)

Describe the theoretical benefits from using $H$ as a proxy for $Y$, discussing whether
they appear to have been obtained in this example.

A6.5 A researcher has data on years of schooling, $S$, weekly earnings in dollars, $W$, hours
worked per week, $H$, and hourly earnings, $E$ (computed as $W/H$) for a sample of
1,755 white males in the United States in the year 2000. She calculates $LW$, $LE$, and
$LH$ as the natural logarithms of $W$, $E$, and $H$, respectively, and fits the
following regressions, with the results shown in the table below (standard errors in
parentheses; $RSS =$ residual sum of squares):

- Column 1: a regression of $LE$ on $S$.
- Column 2: a regression of $LW$ on $S$ and $LH$.
- Column 3: a regression of $LE$ on $S$ and $LH$.

The correlation between $S$ and $LH$ is 0.06.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>FT</td>
<td>PT</td>
</tr>
<tr>
<td>Dependent variable</td>
<td>LE</td>
<td>LW</td>
<td>LE</td>
<td>LW</td>
<td>LW</td>
</tr>
<tr>
<td>$S$</td>
<td>0.099</td>
<td>0.098</td>
<td>0.098</td>
<td>0.101</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>$LH$</td>
<td>—</td>
<td>1.190</td>
<td>0.190</td>
<td>0.980</td>
<td>0.885</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.065)</td>
<td>(0.088)</td>
<td>(0.325)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.111</td>
<td>5.403</td>
<td>5.403</td>
<td>6.177</td>
<td>7.002</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td>(0.254)</td>
<td>(0.254)</td>
<td>(0.345)</td>
<td>(1.093)</td>
</tr>
<tr>
<td>$RSS$</td>
<td>741.5</td>
<td>737.9</td>
<td>737.9</td>
<td>626.1</td>
<td>100.1</td>
</tr>
<tr>
<td>Observations</td>
<td>1,755</td>
<td>1,755</td>
<td>1,755</td>
<td>1,669</td>
<td>86</td>
</tr>
</tbody>
</table>

- Explain why specification (1) is a restricted version of specification (2), stating
  and interpreting the restriction.
- Supposing the restriction to be valid, explain whether you expect the
  coefficient of $S$ and its standard error to differ, or be similar, in specifications
  (1) and (2).
6.3. Additional exercises

- Supposing the restriction to be invalid, how would you expect the coefficient of $S$ and its standard error to differ, or be similar, in specifications (1) and (2)?
- Perform an $F$ test of the restriction.
- Perform a $t$ test of the restriction.
- Explain whether the $F$ test and the $t$ test could lead to different conclusions.
- At a seminar, a commentator says that part-time workers tend to be paid worse than full-time workers and that their earnings functions are different. Defining full-time workers as those working at least 35 hours per week, the researcher divides the sample and fits the earnings functions for full-time workers (column 4) and part-time workers (column 5). Test whether the commentator’s assertion is correct.
- What are the implications of the commentator’s assertion for the test of the restriction?

A6.6 A researcher investigating whether government expenditure tends to crowd out investment has data on government recurrent expenditure, $G$, investment, $I$, and gross domestic product, $Y$, all measured in US$ billion, for 30 countries in 2012. She fits two regressions (standard errors in parentheses; $t$ statistics in square brackets; $RSS = \text{residual sum of squares}$).

1) A regression of $\log I$ on $\log G$ and $\log Y$:

\[
\hat{\log I} = -2.44 - 0.63 \log G + 1.60 \log Y \quad R^2 = 0.98 \quad (1)
\]

\[
(0.26) \quad (0.12) \quad (0.12) \quad RSS = 0.90
\]

\[
[9.42] \quad [-5.23] \quad [12.42]
\]

2) A regression of $\log(I/Y)$ on $\log(G/Y)$:

\[
\log \left( \frac{I}{Y} \right) = 2.65 - 0.63 \log \left( \frac{G}{Y} \right) \quad R^2 = 0.48 \quad (2)
\]

\[
(0.23) \quad (0.12) \quad RSS = 0.99
\]

\[
[11.58] \quad [-5.07]
\]

The correlation between $\log G$ and $\log Y$ in the sample is 0.98. The table gives some further basic data on $\log G$, $\log Y$, and $\log (G/Y)$.

<table>
<thead>
<tr>
<th></th>
<th>Sample mean</th>
<th>Mean square deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log G$</td>
<td>3.75</td>
<td>2.00</td>
</tr>
<tr>
<td>$\log Y$</td>
<td>5.57</td>
<td>1.95</td>
</tr>
<tr>
<td>$\log (G/Y)$</td>
<td>-1.81</td>
<td>0.08</td>
</tr>
</tbody>
</table>

- Explain why the second specification is a restricted version of the first. State the restriction.
- Perform a test of the restriction.
6. Specification of regression variables

- The researcher expected the standard error of the coefficient of \(\log(G/Y)\) in (2) to be smaller than the standard error of the coefficient of \(\log G\) in (1). Explain why she expected this.

- However, the standard error is the same, at least to two decimal places. Give an explanation.

- Show how the restriction could be tested using a \(t\) test in a reparameterised version of the specification for (1).

A6.7 Is expenditure per capita on your CES category related to total household expenditure per capita?

The model specified in Exercise A4.1 is a restricted version of that in Exercise 4.5 in the text. Perform an \(F\) test of the restriction. Also perform a \(t\) test of the restriction.

[Exercise 4.5: regress \(LGCAT\) on \(LGEXP\) and \(LGSIZE\); Exercise A4.1: regress \(LGCATPC\) on \(LGEXPPC\).]

A6.8 A researcher is considering two regression specifications:

\[
\log Y = \beta_1 + \beta_2 \log X + u \quad (1)
\]

and:

\[
\log \frac{Y}{X} = \alpha_1 + \alpha_2 \log X + u \quad (2)
\]

where \(u\) is a disturbance term. Determine whether (2) is a reparameterised or a restricted version of (1).

A6.9 Three researchers investigating the determinants of hourly earnings have the following data for a sample of 104 male workers in the United States in 2006: \(E\), hourly earnings in dollars; \(S\), years of schooling; \(NUM\), score on a test of numeracy; and \(VERB\), score on a test of literacy. The \(NUM\) and \(VERB\) tests are marked out of 100. The correlation between them is 0.81. Defining \(LGE\) to be the natural logarithm of \(E\), Researcher 1 fits the following regression (standard errors in parentheses; \(RSS\) = residual sum of squares):

\[
\hat{LGE} = 2.02 + 0.063S + 0.0044NUM + 0.0026VERB \quad RSS = 2,000
\]

\[(1.81) \quad (0.007) \quad (0.0011) \quad (0.0010)\]

Researcher 2 defines a new variable \(SCORE\) as the average of \(NUM\) and \(VERB\). She fits the regression:

\[
\hat{LGE} = 1.72 + 0.050S + 0.0068SCORE \quad RSS = 2,045
\]

\[(1.78) \quad (0.005) \quad (0.0010)\]

Researcher 3 fits the regression:

\[
\hat{LGE} = 2.02 + 0.063S + 0.0088SCORE - 0.0018VERB \quad RSS = 2,000
\]

\[(1.81) \quad (0.007) \quad (0.0022) \quad (0.0012)\]
6.3. Additional exercises

- Show that the specification of Researcher 2 is a restricted version of the specification of Researcher 1, stating the restriction.
- Perform an F test of the restriction.
- Show that the specification of Researcher 3 is a reparameterised version of the specification of Researcher 1 and hence perform a t test of the restriction in the specification of Researcher 2.
- Explain whether the F test and the t test could have led to different results.
- Perform a test of the hypothesis that the numeracy score has a greater effect on earnings than the literacy score.
- Compare the regression results of the three researchers.

A6.10 It is assumed that manufacturing output is subject to the production function:

\[ Q = AK^\alpha L^\beta \]  

(1)

where \( Q \) is output and \( K \) and \( L \) are capital and labour inputs. The cost of production is:

\[ C = \rho K + wL \]  

(2)

where \( \rho \) is the cost of capital and \( w \) is the wage rate. It can be shown that, if the cost is minimised, the wage bill \( wL \) will be given by the relationship:

\[ \log wL = \frac{1}{\alpha + \beta} \log Q + \frac{\alpha}{\alpha + \beta} \log \rho + \frac{\beta}{\alpha + \beta} \log w + \text{constant}. \]  

(3)

(Note: You are not expected to prove this.)

A researcher has annual data for 2002 for \( Q, K, L, \rho, \) and \( w \) (all monetary measures being converted into US dollars) for the manufacturing sectors of 30 industrialised countries and regresses log \( wL \) on log \( Q \), log \( \rho \), and log \( w \).

- Demonstrate that relationship (3) embodies a testable restriction and show how the model may be reformulated to take advantage of it.
- Explain how the restriction could be tested using an F test.
- Explain how the restriction could be tested using a t test.
- Explain the theoretical benefits of making use of a valid restriction. How could the researcher assess whether there are any benefits in practice, in this case?
- At a seminar, someone suggests that it is reasonable to hypothesise that manufacturing output is subject to constant returns to scale, so that \( \alpha + \beta = 1 \). Explain how the researcher could test this hypothesis (1) using an F test, (2) using a t test.

A6.11 A researcher hypothesises that the net annual growth of private sector purchases of government bonds, \( B \), is positively related to the nominal rate of interest on the bonds, \( I \), and negatively related to the rate of price inflation, \( P \):

\[ B = \beta_1 + \beta_2 I + \beta_3 P + u \]
6. Specification of regression variables

where \( u \) is a disturbance term. The researcher anticipates that \( \beta_2 > 0 \) and \( \beta_3 < 0 \). She also considers the possibility that \( B \) depends on the real rate of interest on the bonds, \( R \), where \( R = I - P \). Using a sample of observations for 40 countries, she regresses \( B \):

- (1) on \( I \) and \( P \)
- (2) on \( R \)
- (3) on \( I \)
- (4) on \( P \) and \( R \)

with the results shown in the corresponding columns of the table below (standard errors in parentheses; \( RSS \) is the residual sum of squares). The correlation coefficient for \( I \) and \( P \) was 0.97.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I )</td>
<td>2.17</td>
<td>—</td>
<td>0.69</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(1.04)</td>
<td>(0.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P )</td>
<td>—</td>
<td>1.37</td>
<td>—</td>
<td>2.17</td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td></td>
<td>(0.44)</td>
<td></td>
</tr>
<tr>
<td>( R )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td></td>
<td>(1.04)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-5.14</td>
<td>-3.15</td>
<td>-1.53</td>
<td>-5.14</td>
</tr>
<tr>
<td></td>
<td>(2.62)</td>
<td>(1.21)</td>
<td>(0.92)</td>
<td>(2.62)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.22</td>
<td>0.20</td>
<td>0.17</td>
<td>0.22</td>
</tr>
<tr>
<td>( RSS )</td>
<td>967.9</td>
<td>987.1</td>
<td>1,024.3</td>
<td>967.9</td>
</tr>
</tbody>
</table>

- Explain why the researcher was dissatisfied with the results of regression (1).
- Demonstrate that specification (2) may be considered to be a restricted version of specification (1).
- Perform an \( F \) test of the restriction, stating carefully your null hypothesis and conclusion.
- Perform a \( t \) test of the restriction.
- Demonstrate that specification (3) may also be considered to be a restricted version of specification (1).
- Perform both an \( F \) test and a \( t \) test of the restriction in specification (3), stating your conclusion in each case.
- At a seminar, someone suggests that specification (4) is also a restricted version of specification (1). Is this correct? If so, state the restriction.
- State, with an explanation, which would be your preferred specification.

A6.12 A researcher has a sample of 43 observations on a dependent variable, \( Y \), and two potential explanatory variables, \( X \) and \( Z \). He defines two further variables \( V \) and \( W \) as the sum of \( X \) and \( Z \) and the difference between them:

\[
V_i = X_i + Z_i
\]
\[
W_i = X_i - Z_i.
\]
He fits the following four regressions:

1. A regression of $Y$ on $X$ and $Z$.
3. A regression of $Y$ on $V$.

The table shows the regression results (standard errors in parentheses; $RSS =$ residual sum of squares; there was an intercept, not shown, in each regression). Unfortunately, a goat ate part of the regression output and some of the numbers are missing. These are indicated by letters.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z$</td>
<td>0.80</td>
<td></td>
<td></td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(I)</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V$</td>
<td>—</td>
<td>A</td>
<td>0.72</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(B)</td>
<td>(0.02)</td>
<td>(K)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W$</td>
<td>—</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RSS$</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each regression included an intercept (not shown).

Reconstruct each missing number if this is possible, giving a brief explanation. If it is not possible to reconstruct a number, give a brief explanation.

A6.13 In Exercise A6.7, a researcher proposes to test the restriction using variations in $R^2$ instead of variations in $RSS$. For food consumed at home, the unrestricted regression of $LGFDHO$ on $LGEXP$ and $LGSIZE$ had $R^2 = 0.4831$. For the regression of $LGFDHOPC$ on $LGEXPPC$, $R^2 = 0.4290$. Hence the researcher’s statistic is:

$$ F = \frac{(0.4831 - 0.4290)/1}{(1 - 0.4290)/6331} = 599.8. $$

Explain why this is different from the $F$ statistic reported for food consumed at home in the answer to Exercise A6.7.

6.4 Answers to the starred exercises in the textbook

6.4 The table gives the results of multiple and simple regressions of $LGFDHO$, the logarithm of annual household expenditure on food eaten at home, on $LGEXP$, the logarithm of total annual household expenditure, and $LGSIZE$, the logarithm of the number of persons in the household, using a sample of 6,334 households in the 2013 Consumer Expenditure Survey. The correlation coefficient for $LGEXP$ and $LGSIZE$ was 0.32. Explain the variations in the regression coefficients.
### 6. Specification of regression variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGEXP</td>
<td>0.58</td>
<td>0.67</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>LGSIZE</td>
<td>0.33</td>
<td>—</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>constant</td>
<td>1.16</td>
<td>0.70</td>
<td>6.04</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.48</td>
<td>0.43</td>
<td>0.19</td>
</tr>
</tbody>
</table>

**Answer:**

If the model is written as:

$$LGFDHO = \beta_1 + \beta_2 LGEXP + \beta_3 LGSIZE + u$$

the expected value of $\widehat{\beta}_2$ in the second regression is given by:

$$E(\widehat{\beta}_2) = \beta_2 + \beta_3 \frac{\sum (LGEXP_i - \overline{LGEXP})(LGSIZE_i - \overline{LGSIZE})}{\sum (LGEXP_i - \overline{LGEXP})^2}.$$  

We know that the covariance is positive because the correlation is positive, and it is reasonable to suppose that $\beta_3$ is also positive, especially given the highly significant positive estimate in the first regression, and so $\widehat{\beta}_2$ is biased upwards. This accounts for the large increase in its size in the second regression. In the third regression:

$$E(\widehat{\beta}_3) = \beta_3 + \beta_2 \frac{\sum (LGEXP_i - \overline{LGEXP})(LGSIZE_i - \overline{LGSIZE})}{\sum (LGSIZE_i - \overline{LGSIZE})^2}.$$  

$\beta_2$ is certainly positive, especially given the highly significant positive estimate in the first regression, and so $\widehat{\beta}_3$ is also biased upwards. As a consequence, the estimate in the third regression is greater than that in the first.

### 6.7 A researcher thinks that the level of activity in the shadow economy, $Y$, depends either positively on the level of the tax burden, $X$, or negatively on the level of government expenditure to discourage shadow economy activity, $Z$. $Y$ might also depend on both $X$ and $Z$. International cross-sectional data on $Y$, $X$, and $Z$, all measured in US$ million, are obtained for a sample of 30 industrialised countries and a second sample of 30 developing countries. The researcher regresses (1) log $Y$ on both log $X$ and log $Z$, (2) log $Y$ on log $X$ alone, and (3) log $Y$ on log $Z$ alone, for each sample, with the following results (standard errors in parentheses):

<table>
<thead>
<tr>
<th></th>
<th>Industrialised countries</th>
<th>Developing countries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>log $X$</td>
<td>0.699</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>(0.154)</td>
<td>(0.112)</td>
</tr>
<tr>
<td>log $Z$</td>
<td>-0.646</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.162)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-1.137</td>
<td>-1.065</td>
</tr>
<tr>
<td></td>
<td>(0.863)</td>
<td>(1.069)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.44</td>
<td>0.10</td>
</tr>
</tbody>
</table>
6.4. Answers to the starred exercises in the textbook

X was positively correlated with Z in both samples. Having carried out the appropriate statistical tests, write a short report advising the researcher how to interpret these results.

**Answer:**

One way to organise an answer to this exercise is, for each sample, to consider the evidence for and against each of the three specifications in turn. The t statistics for the slope coefficients are given in the following table. * indicates significance at the 5 per cent level, ** at the 1 per cent level, and *** at the 0.1 per cent level, using one-sided tests. (Justification for one-sided tests: one may rule out a negative coefficient for X and a positive one for Y.)

<table>
<thead>
<tr>
<th></th>
<th>Industrialised countries</th>
<th>Developing countries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>log X</td>
<td>4.54***</td>
<td>1.79*</td>
</tr>
<tr>
<td>log Z</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

**Industrialised countries:**

The first specification is clearly the only satisfactory one for this sample, given the t statistics. Writing the model as:

\[
\log Y = \beta_1 + \beta_2 \log X + \beta_3 \log Z + u
\]

in the second specification:

\[
E(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{\sum (\log X_i - \bar{\log X}) (\log Z_i - \bar{\log Z})}{\sum (\log X_i - \bar{\log X})^2}.
\]

Anticipating that \(\beta_3\) is negative, and knowing that X and Z are positively correlated, the bias term should be negative. The estimate of \(\hat{\beta}_2\) is indeed lower in the second specification. In the third specification:

\[
E(\hat{\beta}_3) = \beta_3 + \beta_2 \frac{\sum (\log X_i - \bar{\log X}) (\log Z_i - \bar{\log Z})}{\sum (\log Z_i - \bar{\log Z})^2}
\]

and the bias should be positive, assuming \(\beta_2\) is positive. \(\hat{\beta}_3\) is indeed less negative than in the first specification.

Note that the sum of the \(R^2\) statistics for the second and third specifications is less than \(R^2\) in the first. This is because the bias terms undermine the apparent explanatory power of X and Z in the second and third specifications. In the third specification, the bias term virtually neutralises the true effect and \(R^2\) is very low indeed.

**Developing countries:**

In principle the first specification is acceptable. The failure of the coefficient of Z to be significant might be due to a combination of a weak effect of Z and a relatively small sample.
6. Specification of regression variables

The second specification is also acceptable since the coefficient of \( Z \) and its \( t \)-statistic in the first specification are very low. Because the \( t \)-statistic of \( Z \) is low, \( R^2 \) is virtually unaffected when it is omitted.

The third specification is untenable because it cannot account for the highly significant coefficient of \( X \) in the first. The omitted variable bias is now so large that it overwhelms the negative effect of \( Z \) with the result that the estimated coefficient is positive.

6.11 A researcher has data on output per worker, \( Y \), and capital per worker, \( K \), both measured in thousands of dollars, for 50 firms in the textiles industry in 2012. She hypothesises that output per worker depends on capital per worker and perhaps also the technological sophistication of the firm, \( TECH \):

\[
Y = \beta_1 + \beta_2 K + \beta_3 TECH + u
\]

where \( u \) is a disturbance term. She is unable to measure \( TECH \) and decides to use expenditure per worker on research and development in 2012, \( R&D \), as a proxy for it. She fits the following regressions (standard errors in parentheses):

\[
\hat{Y} = 1.02 + 0.32K \quad R^2 = 0.749
\]

\[
(0.45) \quad (0.04)
\]

\[
\hat{Y} = 0.34 + 0.29K + 0.05R&D \quad R^2 = 0.750
\]

\[
(0.61) \quad (0.22) \quad (0.15)
\]

The correlation coefficient for \( K \) and \( R&D \) is 0.92. Discuss these regression results:

1. assuming that \( Y \) does depend on both \( K \) and \( TECH \)
2. assuming that \( Y \) depends only on \( K \).

**Answer:**

If \( Y \) depends on both \( K \) and \( TECH \), the first specification is subject to omitted variable bias, with the expected value of \( \hat{\beta}_2 \) being given by:

\[
E(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{\sum (K_i - \bar{K})(TECH_i - \bar{TECH})}{\sum (K_i - \bar{K})^2}.
\]

Since \( K \) and \( R&D \) have a high positive correlation, it is reasonable to assume that \( K \) and \( TECH \) are positively correlated. It is also reasonable to assume that \( \beta_3 \) is positive. Hence one would expect \( \hat{\beta}_2 \) to be biased upwards. It is indeed greater than in the second equation, but not by much. The second specification is clearly subject to multicollinearity, with the consequence that, although the estimated coefficients remain unbiased, they are erratic, this being reflected in large standard errors. The large variance of the estimate of the coefficient of \( K \) means that much of the difference between it and the estimate in the first specification is likely to be purely random, and this could account for the fact that the omitted variable bias appears to be so small.

If \( Y \) depends only on \( K \), the inclusion of \( R&D \) in the second specification gives rise to inefficiency. Since the standard errors in both equations remain valid, they can
be compared and it is evident that the loss of efficiency is severe. As expected in this case, the coefficient of $R&D$ is not significantly different from zero and the increase in $R^2$ in the second specification is minimal.

6.14 The first regression shows the result of regressing $LGFDHO$, the logarithm of annual household expenditure on food eaten at home, on $LGEXP$, the logarithm of total annual household expenditure, and $LGSIZE$, the logarithm of the number of persons in the household, using a sample of 6,334 households in the 2013 Consumer Expenditure Survey. In the second regression, $LGFDHOPC$, the logarithm of food expenditure per capita ($FDHO/$SIZE), is regressed on $LGEXPPC$, the logarithm of total expenditure per capita ($EXP/$SIZE). In the third regression $LGFDHOPC$ is regressed on $LGEXPPC$ and $LGSIZE$.

<table>
<thead>
<tr>
<th>.reg LGFDHO LGEXP LGSIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

| | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|---|-------|-----------|-------|--------|---------------------|
| LGEXP | .5842097 | .0097174 | 60.12 | 0.000 | .5651604 .6032591 |
| LGSIZE | .3343475 | .0127587 | 26.21 | 0.000 | .3093362 .3593589 |
| _cons | 1.158326 | .0820119 | 14.12 | 0.000 | .9975545 1.319097 |

<p>| .gen LGFDHOPC = ln(FDHO/SIZE) |
| .gen LGEXPPC = ln(EXP/SIZE) |</p>
<table>
<thead>
<tr>
<th>.reg LGFDHOPC LGEXPPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

| | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|---|-------|-----------|-------|--------|---------------------|
| LGEXPPC | .6092734 | .0088338 | 68.97 | 0.000 | .5919562 .6265905 |
| _cons | .8988292 | .0703516 | 12.78 | 0.000 | .7609162 1.036742 |
6. Specification of regression variables

```
. reg LGFDHOPC LGEXPPC LGSIZE
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1514.30723</td>
<td>2</td>
<td>757.153617</td>
<td>F( 2, 6331) = 2410.79</td>
</tr>
<tr>
<td>Residual</td>
<td>1988.36474</td>
<td>6331</td>
<td>.314068037</td>
<td>Prob&gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>3502.67197</td>
<td>6333</td>
<td>.553082579</td>
<td>R-squared = 0.4323</td>
</tr>
</tbody>
</table>

| LGFDHOPC | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|-------|-----|------------------------|
| LGEXPPC  | .5842097 | .0097174 | 60.12 | 0.000 | .5651604 .6032591 |
| LGSIZE   | -.0814427 | .0133333 | -6.11 | 0.000 | -.1075805 -.0553049 |
| _cons    | 1.158326 | .0820119 | 14.12 | 0.000 | .9975545 1.319097 |

1. Explain why the second model is a restricted version of the first, stating the restriction.

2. Perform an $F$ test of the restriction.

3. Perform a $t$ test of the restriction.

4. Summarise your conclusions from the analysis of the regression results.

**Answer:**

Write the first specification as:

$$LGFDHO = \beta_1 + \beta_2 \cdot LGEXP + \beta_3 \cdot LGSIZE + u.$$ 

Then the restriction implicit in the second specification is $\beta_3 = 1 - \beta_2$, for:

$$LGFDHO = \beta_1 + \beta_2 \cdot LGEXP + (1 - \beta_2) \cdot LGSIZE + u$$

$$LGFDHO - LGSIZE = \beta_1 + \beta_2 \cdot (LGEXP - LGSIZE) + u$$

$$\frac{FDHO}{SIZE} = \beta_1 + \beta_2 \log \frac{EXP}{SIZE} + u$$

$$LGFDHOPC = \beta_1 + \beta_2 \cdot LGEXPPC + u$$

the last equation being the second specification. The $F$ statistic for the null hypothesis $H_0: \beta_3 = 1 - \beta_2$ is:

$$F(1, 6331) = \frac{(2000.1 - 1988.4)/1}{1988.4/6331} = 37.3.$$  

The critical value of $F(1, 1000)$ at the 0.1 per cent level is 10.9, and hence the restriction is rejected at that significance level.

Alternatively, we could use the $t$ test approach. Under the null hypothesis that the restriction is valid, $\theta = 1 - \beta_2 - \beta_3 = 0$. Substituting for $\beta_3$, the unrestricted version may be rewritten:

$$LGFDHO = \beta_1 + \beta_2 \cdot LGEXP + (1 - \beta_2 - \theta) \cdot LGSIZE + u.$$
This may be rewritten:

\[
\log \frac{FDHO}{SIZE} = \beta_1 + \beta_2 \log \frac{EXP}{SIZE} - \theta \log SIZE + u
\]

that is:

\[
LGFDHOPC = \beta_1 + \beta_2 LGEXPPC - \theta LGSIZE + u.
\]

The \textit{t} statistic for the coefficient of \textit{LGSIZE} is \(-6.11\), so we reject the restriction at a very high significance level. Note that the \textit{t} statistic is the square root of the \textit{F} statistic and the critical value of \textit{t} at the 0.1 per cent level will be the square root of the critical value of \textit{F}.

### 6.5 Answers to the additional exercises

#### A6.1

The output below gives the results of a simple regression of \textit{LGFDHOPC} on \textit{LGSIZE}. See Exercise A4.1 for the simple regression of \textit{LGFDHOPC} on \textit{LGEXPPC} and Exercise A4.2 for the multiple regression of \textit{LGFDHOPC} on \textit{LGEXPPC} and \textit{LGSIZE}.

```
. reg LGFDHOPC LGSIZE
```

```
Source | SS       | df | MS                   Number of obs = 6334
--------+----------|----|----------------------|-----------------+-------------------+=+-------------------=
Model   | 379.128845  1 | 379.128845 | Prob > F = 0.0000
Residual| 3123.54316  6332 | 0.493294877 | R-squared = 0.1082
--------+-----------------+-------------------+=+-------------------=
Total   | 3502.67201  6333 | 0.553082585 | Root MSE = .70235
--------+-------------------+-------------------+=+-------------------=
LGFDHOPC | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
--------+----------|----------|-------|------|-----------------+-------------------=
  LGSIZE | -.4199282 | .0151473 | -27.72 | 0.000 | -.449622 to -.3902344 |
   _cons | 6.040547  | .0143586 | 420.69 | 0.000 | 6.012399 to 6.068695 |
```

If the true model is assumed to be:

\[
LGFDHOPC = \beta_1 + \beta_2 LGEXPPC + \beta_3 LGSIZE + u
\]

the expected value of \(\hat{\beta}_2\) in the simple regression of \textit{LGFDHOPC} on \textit{LGEXPPC} is given by:

\[
E(\hat{\beta}_2) = \beta_2 + \frac{\sum (LGEXPPC_i - LGEXPPC)(LGSIZE_i - LGSIZE)}{\sum (LGEXPPC_i - LGEXPPC)^2}
\]

We know that the numerator of the second factor in the bias term is negative because the correlation is negative:
6. Specification of regression variables

```
cor LGEXPPC LGSIZE
(obs=6334)

<table>
<thead>
<tr>
<th></th>
<th>LGEXPPC</th>
<th>LGSIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGEXPPC</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>LGSIZE</td>
<td>-0.4223</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```

It is reasonable to suppose that economies of scale will cause $\beta_3$ to be negative, and the highly significant negative estimate in the multiple regression provides empirical support, so $\hat{\beta}_3$ is biased upwards. This accounts for the increase in its size in the second regression. In the third regression:

$$E(\hat{\beta}_3) = \beta_3 + \beta_2 \frac{\sum (LGEXPPC_i - LGEXPPC) (LGSIZE_i - LGSIZE)}{\sum (LGSIZE_i - LGSIZE)^2}.$$ 

$\beta_2$ is certainly positive, especially given the highly significant positive estimate in the first regression, and so $\hat{\beta}_3$ is biased downwards. As a consequence, the estimate in the third regression is lower than that in the first.

Similar results are obtained for the other categories of expenditure. The correlation between LGEXPPC and LGSIZE varies because the missing observations are different for different categories.
A6.2 Demonstrate that $\hat{\beta}_2$ is an unbiased estimator of $\beta_2$.

$$\hat{\beta}_2 = \frac{\sum (D_i - \bar{D})(Y_i - \bar{Y})}{\sum (D_i - \bar{D})^2}$$

$$= \frac{\sum (D_i - \bar{D})((\beta_1 + \beta_2D_i + \beta_3A_i + u_i) - (\beta_1 + \beta_2\bar{D} + \beta_3\bar{A} + \bar{u}))}{\sum (D_i - \bar{D})^2}$$

$$= \beta_2 + \beta_3 \frac{\sum (D_i - \bar{D})(A_i - \bar{A})}{\sum (D_i - \bar{D})^2} + \frac{\sum (D_i - \bar{D})(u_i - \bar{u})}{\sum (D_i - \bar{D})^2}.$$

Hence:

$$\hat{\beta}_2 = \beta_2 + \beta_3 \sum d_i(A_i - \bar{A}) + \sum d_i(u_i - \bar{u})$$

where:

$$d_i = \frac{D_i - \bar{D}}{\sum(D_j - \bar{D})^2}.$$ 

Hence:

$$E(\hat{\beta}_2) = \beta_2 + \beta_3 \sum E(d_i(A_i - \bar{A})) + \sum E(d_i(u_i - \bar{u})).$$

Now, since the assignment to the course was random, $D$ is distributed independently of both $A$ and $u$, and hence:

$$E(d_i(A_i - \bar{A})) = E(d_i)E(A_i - \bar{A}) = 0$$

and:

$$E(d_i(u_i - \bar{u})) = E(d_i)E(u_i - \bar{u}) = 0.$$ 

Hence $\hat{\beta}_2$ is an unbiased estimator of $\beta_2$.

A commentator says that the standard error of $\hat{\beta}_2$ will be invalid because an important variable, $A$, has been omitted from the specification. The researcher replies that the standard error will remain valid if $A$ can be assumed to have a normal distribution. Explain whether the commentator or the researcher is correct.

The researcher is nearly correct. Given the random selection of the sample, $A$ will be distributed independently of $D$ and so it can be treated as part of the disturbance term and the standard error will remain valid. The requirement that $A$ have a normal distribution is too strong, since the expression for the standard error does not depend on it. However, if the standard error is to be used for $t$ tests, then it is important that the enlarged standard error should have a normal distribution, and this will be the case if an only if $A$ has a normal distribution (assuming that $u$ has one). If both $A$ and $u$ have normal distributions, a linear combination will also have one.

Another commentator says that whether the distribution of $A$ is normal or not makes no difference to the validity of the standard error. Evaluate this assertion.

The commentator is correct for the reasons just explained.
6. Specification of regression variables

A6.3 **Assuming that (A) is the correct specification, explain, with a mathematical proof, whether you would expect the coefficient of log $Y$ to be greater in regression (2).**

To simplify the algebra, throughout this answer log $B$, log $Y$, log $S$ and log $IQ$ will be written as $B$, $Y$, $S$ and $IQ$, it being understood that these are logarithms.

\[
\hat{\beta}_2 = \frac{\sum (B_i - \bar{B})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}
\]

\[
= \frac{\sum (\beta_1 + \beta_2 Y_i + \beta_3 IQ_i + u_i - \beta_1 - \beta_2 \bar{Y} - \beta_3 \bar{IQ} - \bar{u})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}
\]

\[
= \frac{\sum (\beta_2 Y_i - \beta_2 \bar{Y})(Y_i - \bar{Y}) + \sum (\beta_3 IQ_i - \beta_3 \bar{IQ})(Y_i - \bar{Y}) + \sum (u_i - \bar{u})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}
\]

\[
= \beta_2 + \frac{\sum (IQ_i - \bar{IQ})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} + \frac{\sum (u_i - \bar{u})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}.
\]

Hence:

\[
E(\hat{\beta}_2) = \beta_2 + \frac{\sum (IQ_i - \bar{IQ})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} + \frac{1}{\sum (Y_i - \bar{Y})^2} E\left(\sum (u_i - \bar{u})(Y_i - \bar{Y})\right)
\]

\[
= \beta_2 + \frac{\sum (IQ_i - \bar{IQ})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} + \frac{1}{\sum (Y_i - \bar{Y})^2} \sum E\left((u_i - \bar{u})(Y_i - \bar{Y})\right)
\]

\[
= \beta_2 + \frac{\sum (IQ_i - \bar{IQ})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} + \frac{1}{\sum (Y_i - \bar{Y})^2} \sum (Y_i - \bar{Y}) E(u_i - \bar{u})
\]

\[
= \beta_2 + \frac{\sum (IQ_i - \bar{IQ})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}.
\]

assuming that $Y$ and $IQ$ are nonstochastic. Thus $\hat{\beta}_2$ is biased, the direction of the bias depending on the signs of $\beta_3$ and $\sum (IQ_i - \bar{IQ})(Y_i - \bar{Y})$. We would expect the former to be positive and we expect the latter to be positive since we are told that the correlation between $S$ and $Y$ is positive and $S$ is a proxy for $IQ$. So we would expect an upward bias in regression (2).

**Assuming that (A) is the correct specification, describe the various benefits from using log $S$ as a proxy for log $IQ$, as in regression (1), if log $S$ is a good proxy.**

The use of $S$ as a proxy for $IQ$ will alleviate the problem of omitted variable bias. In particular, comparing the results of regression (1) with those that would have been obtained if $B$ had been regressed on $Y$ and $IQ$:
6.5. Answers to the additional exercises

- the coefficient of \( Y \) will be approximately the same
- its standard error will be approximately the same
- the \( t \) statistic for \( S \) will be approximately equal to that of \( IQ \)
- \( R^2 \) will be approximately the same.

*Explain whether the low value of \( R^2 \) in regression (1) implies that \( \log S \) is not a good proxy.*

Not necessarily. It could be that \( S \) is a poor proxy for \( IQ \), but it could also be that the original model had low explanatory power.

*Assuming that (A) is the correct specification, provide an explanation of why the coefficients of \( \log Y \) and \( \log S \) in regression (1) are not significantly different from zero, using two-sided \( t \) tests.*

The high correlation between \( Y \) and \( S \) has given rise to multicollinearity, the standard errors being so large that the coefficients are not significantly different from zero.

*Discuss whether the researcher would be justified in using one-sided \( t \) tests in regression (1).*

Yes. It is reasonable to suppose that expenditure on books should not be negatively influenced by either income or cognitive ability. (Note that one should not say that it is reasonable to suppose that expenditure on books is positively influenced by them. This rules out the null hypothesis.)

*Assuming that (B) is the correct specification, explain whether you would expect the coefficient of \( \log Y \) to be lower in regression (1).*

No. It would be randomly higher or lower, if \( S \) is an irrelevant variable.

*Assuming that (B) is the correct specification, explain whether the standard errors in regression (1) are valid estimates.*

Yes. The inclusion of an irrelevant variable in general does not invalidate the standard errors. It causes them to be larger than those in the correct specification.

**A6.4 Explain analytically whether the slope coefficient is likely to be biased.**

If the fitted model is:

\[
\hat{\log T} = \hat{\beta}_1 + \hat{\beta}_2 \log P
\]

then:

\[
\hat{\beta}_2 = \frac{\sum (\log P_i - \log \bar{P}) (\log T_i - \log \bar{T})}{\sum (\log P_i - \log \bar{P})^2}
\]

\[
= \frac{\sum (\log P_i - \log \bar{P}) (\beta_1 + \beta_2 \log P_i + \beta_2 \log Y_i + u_i - \beta_1 - \beta_2 \log \bar{P} - \beta_3 \log \bar{Y} - \bar{u})}{\sum (\log P_i - \log \bar{P})^2}
\]

\[
= \beta_2 + \beta_3 \frac{\sum (\log P_i - \log \bar{P}) (\log Y_i - \log \bar{Y})}{\sum (\log P_i - \log \bar{P})^2} + \frac{\sum (\log P_i - \log \bar{P}) (u_i - \bar{u})}{\sum (\log P_i - \log \bar{P})^2}.
\]
6. Specification of regression variables

Hence:

\[ E(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{\sum (\log P_i - \log \bar{P}) (\log Y_i - \log \bar{Y})}{\sum (\log P_i - \log \bar{P})^2} \]

provided that any random component of \( \log P \) is distributed independently of \( u \).

Since it is reasonable to assume \( \beta_3 > 0 \), and since we are told that \( Y \) and \( P \) are positively correlated, the bias will be upwards. This accounts for the nonsensical positive price elasticity in the fitted equation.

Describe the theoretical benefits from using \( H \) as a proxy for \( Y \), discussing whether they appear to have been obtained in this example.

Suppose that \( H \) is a perfect proxy for \( Y \):

\[ \log Y = \lambda + \mu \log H. \]

Then the relationship may be rewritten:

\[ \log T = \beta_1 + \beta_3 \lambda + \beta_2 \log P + \beta_3 \mu \log H + u. \]

The coefficient of \( \log P \) ought to be the same as in the true relationship. However in this example it is not the same. However it is of the right order of magnitude and much more plausible than the estimate in the first regression. The standard error of the coefficient ought to be the same as in the true relationship, and this is the case.

The coefficient of \( \log H \) will be an estimate of \( \beta_3 \mu \), and since \( \mu \) is unknown, \( \beta_3 \) is not identified. However, if it can be assumed that the average household income in a city is proportional to average house values, it could be asserted that \( \mu \) is equal to 1, in which case the coefficient of \( \log H \) will be a direct estimate of \( \beta_3 \) after all.

The coefficient of \( \log H \) is indeed quite close to that of \( \log Y \). The \( t \) statistic for the coefficient of \( \log H \) ought to be the same as that for \( \log Y \), and this is approximately true, being a little lower. \( R^2 \) ought to be the same, but it is somewhat lower, suggesting that \( H \) appears to have been a good proxy, but not a perfect one.

A6.5 Explain why specification (1) is a restricted version of specification (2), stating and interpreting the restriction.

First note that, since \( E = W/H \), \( LE = \log(W/H) = LW - LH \).

Write specification (2) as:

\[ LW = \beta_1 + \beta_2 S + \beta_3 LH + u. \]

If one imposes the restriction \( \beta_3 = 1 \), the model becomes specification (1):

\[ LW - LH = \beta_1 + \beta_2 S + u. \]

The restriction implies that weekly earnings are proportional to hours worked, controlling for schooling.

Supposing the restriction to be valid, explain whether you expect the coefficient of \( S \) and its standard error to differ, or be similar, in specifications (1) and (2).

If the restriction is valid, the coefficient of \( S \) should be similar in the restricted specification (1) and the unrestricted specification (2). Both estimates will be
unbiased, but that in specification (1) will be more efficient. The gain in efficiency in specification (1) should be reflected in a smaller standard error. However, the gain will be small, given the low correlation.

Supposing the restriction to be invalid, how would you expect the coefficient of $S$ and its standard error to differ, or be similar, in specifications (1) and (2)?

The estimate of the coefficient of $S$ would be biased. The standard error in specification (1) would be invalid and so a comparison with the standard error in specification (2) would be illegitimate.

Perform an $F$ test of the restriction.

The null and alternative hypotheses are $H_0 : \beta_3 = 1$ and $H_1 : \beta_3 \neq 1$.

$$F(1, 1752) = \frac{(741.5 - 737.9)/1}{737.9/1752} = 8.5.$$  

The critical value of $F(1, 1000)$ at the 1 per cent level is 6.66. The critical value of $F(1, 1752)$ must be lower. Thus we reject the restriction at the 1 per cent level. (The critical value at the 0.1 per cent level is about 10.8.)

Perform a $t$ test of the restriction.

The restriction is so simple that it can be tested with no reparameterisation: a simple $t$ test on the coefficient of $LH$ in specification (2), $H_0 : \beta_3 = 1$.

Alternatively, mechanically following the standard procedure, we rewrite the restriction as $\beta_3 - 1 = 0$. The reparameterisation will be:

$$\theta = \beta_3 - 1$$

and so:

$$\beta_3 = \theta + 1.$$  

Substituting this into the unrestricted specification, the latter may be rewritten:

$$LW = \beta_1 + \beta_2 S + (\theta + 1) LH + u.$$  

Hence:

$$LW - LH = \beta_1 + \beta_2 S + \theta LH + u.$$  

This is regression specification (3) and the restriction may be tested with a $t$ test on the coefficient of $LH$, the null hypothesis being $H_0 : \theta = \beta_3 - 1 = 0$. The $t$ statistic is 2.92, which is significant at the 1 per cent level, implying that the restriction should be rejected.

Explain whether the $F$ test and the $t$ test could lead to different conclusions.

The tests must lead to the same conclusion since the $F$ statistic is the square of the $t$ statistic and the critical value of $F$ is the square of the critical value of $t$.

At a seminar, a commentator says that part-time workers tend to be paid worse than full-time workers and that their earnings functions are different. Defining full-time workers as those working at least 35 hours per week, the researcher divides the sample and fits the earnings functions for full-time workers (column 4) and part-time workers (column 5). Test whether the commentators assertion is correct.
6. Specification of regression variables

The appropriate test is a Chow test. The test statistic under the null hypothesis of no difference in the earnings functions is:

\[ F(3, 1749) = \frac{(737.9 - 626.1 - 100.1)/3}{(626.1 + 100.1)/1749} = 9.39. \]

The critical value of \( F(3, 1000) \) at the 0.1 per cent level is 5.46. Hence we reject the null hypothesis and conclude that the commentator is correct.

What are the implications of the commentator's assertion for the test of the restriction?

The elasticity of \( LH \) is now not significantly different from 1 for either full-time or part-time workers, so the restriction is no longer rejected.

A6.6 Explain why the second specification is a restricted version of the first. State the restriction.

Write the second equation as:

\[ \log \frac{I}{Y} = \beta_1 + \beta_2 \log \left( \frac{G}{Y} \right) + u. \]

It may be re-written as:

\[ \log I = \beta_1 + \beta_2 \log G + (1 - \beta_2) \log Y + u. \]

This is a special case of the specification of the first equation:

\[ \log I = \beta_1 + \beta_2 \log G + \beta_3 \log Y + u \]

with the restriction \( \beta_3 = 1 - \beta_2 \).

Perform a test of the restriction.

The null hypothesis is \( H_0 : \beta_2 + \beta_3 = 1. \) The test statistic is:

\[ F(1, 27) = \frac{(0.99 - 0.90)/1}{0.90/27} = 2.7. \]

The critical value of \( F(1, 27) \) is 4.21 at the 5 per cent level. Hence we do not reject the null hypothesis that the restriction is valid.

The researcher expected the standard error of the coefficient of \( \log (G/Y) \) in (2) to be smaller than the standard error of the coefficient of \( \log G \) in (1). Explain why she expected this.

The imposition of the restriction, if valid, should lead to a gain in efficiency and this should be reflected in lower standard errors.

However the standard error is the same, at least to two decimal places. Give an explanation.

The standard errors of the coefficients of \( G \) in (1) and \( G/Y \) in (2) are given by:

\[ \sqrt{\frac{\hat{\sigma}_u^2}{n\text{MSD}(G)}} \times \frac{1}{1 - r_{G,Y}^2} \quad \text{and} \quad \sqrt{\frac{\hat{\sigma}_u^2}{n\text{MSD}(G/Y)}}. \]
6.5. Answers to the additional exercises

respectively, where \( \hat{\sigma}^2_u \) is an estimate of the variance of the disturbance term, \( n \) is the number of observations, MSD is the mean square deviation in the sample, and \( r_{G,Y} \) is the sample correlation coefficient of \( G \) and \( Y \). \( n \) is the same for both standard errors and \( \hat{\sigma}^2_u \) will be very similar. We are told that \( r_{G,Y} = 0.98 \), so its square is 0.96 and the second factor in the expression for the standard error of \( G \) is \((1/0.04) = 25\). Hence, other things being equal, the standard error of \( G/Y \) should be much lower than that of \( G \). However the table shows that the MSD of \( G/Y \) is only 1/25 as great as that of \( G \). This just about exactly negates the gain in efficiency attributable to the elimination of the correlation between \( G \) and \( Y \).

Show how the restriction could be tested using a t test in a reparameterised version of the specification for (1).

Define \( \theta = \beta_2 + \beta_3 - 1 \), so that the restriction may be written \( \theta = 0 \). Then \( \beta_3 = \theta - \beta_2 + 1 \). Use this to substitute for \( \beta_3 \) in the unrestricted model:

\[
\log I = \beta_1 + \beta_2 \log G + \beta_3 \log Y + u
\]

\[
= \beta_1 + \beta_2 \log G + (\theta - \beta_2 + 1) \log Y + u.
\]

Then:

\[
\log I - \log Y = \beta_1 + \beta_2 (\log G - \log Y) + \theta \log Y + u
\]

and:

\[
\log \left( \frac{I}{Y} \right) = \beta_1 + \beta_2 \left( \frac{G}{Y} \right) + \theta \log Y + u.
\]

Hence the restriction may be tested by a t test of the coefficient of \( \log Y \) in a regression using this specification.

A6.7 This is a generalisation of the example with \( FDHO \) in Exercise 6.14 in the text. The reason for the discrepancy in the number of observations is not known. Possibly it used an earlier version of the data set.
6. Specification of regression variables

```
. reg LGFDHOPC LGEXPPC
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1502.58932</td>
<td>1</td>
<td>1502.58932</td>
<td>F( 1, 6332) = 4757.00</td>
</tr>
<tr>
<td>Residual</td>
<td>2000.08269</td>
<td>6332</td>
<td>.315869029</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>3502.67201</td>
<td>6333</td>
<td>.553082585</td>
<td>Adj R-squared = 0.4289</td>
</tr>
</tbody>
</table>

```
LGFDHOPC | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----------+----------------------------------------------------------------
LGEXPPC   | .6092734 .0088338 68.97 0.000 .5919562 .6265905
_cons     | .8988291 .0703516 12.78 0.000 .7609161 1.036742
```

Write the first specification as:

\[ LGFDHO = \beta_1 + \beta_2 LGEXP + \beta_3 LGSIZE + u. \]

Then the restriction implicit in the second specification is \( \beta_3 = 1 - \beta_2 \), for then:

\[ LGFDHO = \beta_1 + \beta_2 LGEXP + (1 - \beta_2)LGSIZE + u \]
\[ LGFDHO - LGSIZE = \beta_1 + \beta_2 (LGEXP - LGSIZE) + u \]
\[ \log \frac{FDHO}{SIZE} = \beta_1 + \beta_2 \log \frac{EXP}{SIZE} + u \]
\[ LGFDHOPC = \beta_1 + \beta_2 LGEXPPC + u \]

the last equation being the second specification. The \( F \) statistic for the null hypothesis \( H_0 : \beta_3 = 1 - \beta_2 \) is:

\[ F(1, 6331) = \frac{(2000.1 - 1988.4)/1}{1998.4/6331} = 37.25. \]

The critical value of \( F(1, 1000) \) at the 0.1 per cent level is 10.9, and hence the restriction is rejected at that significance level. This is not a surprising result, given that the estimates of \( \beta_2 \) and \( \beta_3 \) in the unrestricted specification were 0.58 and 0.33, respectively, their sum being well short of 1, as implied by the restriction.

Summarising the results of the test for all the categories, we have:

- Restriction rejected at the 1 per cent level: FDHO, FDAW, HOUS, TELE, FURN, MAPP, SAPP, CLOT, HEAL, ENT, FEES, READ, TOB.
- Restriction rejected at the 5 per cent level: TRIP, LOCT.
- Restriction not rejected at the 5 per cent level: DOM, TEXT, FOOT, GASO, TOYS, EDUC.
6.5. Answers to the additional exercises

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>RSS restricted</th>
<th>RSS unrestricted</th>
<th>F</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>3,947.5</td>
<td>3,945.2</td>
<td>1.6</td>
<td>−1.26</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>5,792.0</td>
<td>5,766.1</td>
<td>20.2</td>
<td>4.50</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>4,138.0</td>
<td>4,062.5</td>
<td>30.8</td>
<td>5.55</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>1,404.6</td>
<td>1,373.5</td>
<td>176.4</td>
<td>−13.28</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>2,842.9</td>
<td>2,636.3</td>
<td>456.4</td>
<td>−21.36</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>3,430.9</td>
<td>3,369.1</td>
<td>93.6</td>
<td>−9.68</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>2,000.1</td>
<td>1,988.4</td>
<td>37.2</td>
<td>−6.11</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>1,506.4</td>
<td>1,373.5</td>
<td>136.4</td>
<td>−13.28</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>920.0</td>
<td>913.9</td>
<td>3.2</td>
<td>−1.80</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>2,879.4</td>
<td>2,879.3</td>
<td>0.0</td>
<td>−0.20</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>6,183.4</td>
<td>6,062.5</td>
<td>95.7</td>
<td>−9.79</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>4,859.4</td>
<td>4,825.6</td>
<td>43.6</td>
<td>−6.60</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>1,622.7</td>
<td>1,559.2</td>
<td>50.9</td>
<td>−7.13</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>1,108.1</td>
<td>1,075.1</td>
<td>21.1</td>
<td>−4.60</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>583.5</td>
<td>576.8</td>
<td>4.6</td>
<td>−2.14</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>3,049.1</td>
<td>3,002.2</td>
<td>59.6</td>
<td>−7.72</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>3,038.1</td>
<td>2,892.1</td>
<td>115.3</td>
<td>−10.74</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>1,239.6</td>
<td>1,148.9</td>
<td>81.6</td>
<td>−9.03</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>3,133.1</td>
<td>3,055.1</td>
<td>147.6</td>
<td>−12.15</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>1,150.5</td>
<td>1,032.9</td>
<td>112.6</td>
<td>−10.61</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>956.3</td>
<td>873.4</td>
<td>109.4</td>
<td>−10.46</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>2,885.4</td>
<td>2,828.3</td>
<td>50.5</td>
<td>−7.11</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>795.4</td>
<td>792.8</td>
<td>1.7</td>
<td>−1.30</td>
</tr>
</tbody>
</table>

For the $t$ test, we first rewrite the restriction as $\beta_2 + \beta_3 - 1 = 0$. The test statistic is therefore $\theta = \beta_2 + \beta_3 - 1$. This allows us to write $\beta_3 = \theta - \beta_2 + 1$. Substituting for $\beta_3$, the unrestricted version becomes:

$$LGFDHO = \beta_1 + \beta_2LGEXP + (\theta - \beta_2 + 1)LGSIZE + u.$$  

Hence the unrestricted version may be rewritten:

$$LGFDHO - LGSIZE = \beta_1 + \beta_2(LGEXP - LGSIZE) + \theta LGSIZE + u$$

that is:

$$LGFDHOPC = \beta_1 + \beta_2LGEXPPC + \theta LGSIZE + u.$$  

We use a $t$ test to see if the coefficient of $LGSIZE$ is significantly different from 0. If it is not, we can drop the $LGSIZE$ term and we conclude that the restricted specification is an adequate representation of the data. If it is, we have to stay with the unrestricted specification.

From the output for the third regression, we see that $t$ is $−6.11$ and hence the null hypothesis $H_0 : \beta_2 + \beta_3 - 1 = 0$ is rejected (critical value of $t$ at the 0.1 per cent level is 3.29). Note that the $t$ statistic is the square root of the $F$ statistic and the critical value of $t$ at the 0.1 per cent level is the square root of the critical value of $F$. The results for the other categories are likewise identical to those for the $F$ test.
### Specification of regression variables

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6334</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1514.30728</td>
<td>2</td>
<td>757.15364</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>1988.36473</td>
<td>6331</td>
<td>.314068035</td>
<td>R-squared = 0.4323</td>
</tr>
<tr>
<td>Total</td>
<td>3502.67201</td>
<td>6333</td>
<td>.553082585</td>
<td>Root MSE = 0.56042</td>
</tr>
</tbody>
</table>

| Coef.    | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-----------|-------|------|----------------------|
| LGFDHOPC | .5842097  | .0097174 | 60.12 | 0.000 | .5651604 .6032591 |
| LGEXPPC  | -.0814427 | .0133333 | -6.11 | 0.000 | -.1075806 -.0553049 |
| _cons    | 1.158326  | .0820119 | 14.12 | 0.000 | .9975545 1.319097 |

A6.8 (2) may be rewritten:

\[ \log Y = \alpha_1 + (\alpha_2 + 1) \log X + u \]

so it is a reparameterised version of (1) with \( \beta_1 = \alpha_1 \) and \( \beta_2 = \alpha_2 + 1 \).

A6.9 *Show that the specification of Researcher 2 is a restricted version of the specification of Researcher 1, stating the restriction.*

Let the model be written:

\[ LGE = \beta_1 + \beta_2 S + \beta_3 NUM + \beta_4 VERB + u. \]

The restriction is \( \beta_4 = \beta_3 \) since \( NUM \) and \( VERB \) are given equal weights in the construction of \( SCORE \). Using the restriction, the model can be rewritten

\[ LGE = \beta_1 + \beta_2 S + \beta_3 (NUM + VERB) + u \]

\[ = \beta_1 + \beta_2 S + 2\beta_3 SCORE + u. \]

*Perform an F test of the restriction.*

The null and alternative hypotheses are \( H_0: \beta_4 = \beta_3 \) and \( H_1: \beta_4 \neq \beta_3 \). The \( F \) statistic is:

\[ F(1, 100) = \frac{(2045 - 2000)/1}{2000/100} = 2.25. \]

The critical value of \( F(1, 100) \) is 3.94 at the 5 per cent level. Hence we do not reject the restriction at the 5 per cent level.

*Show that the specification of Researcher 3 is a reparameterised version of the specification of Researcher 1 and hence perform a t test of the restriction in the specification of Researcher 2.*

The restriction may be rewritten \( \beta_4 - \beta_3 = 0 \). The test statistic is therefore \( \theta = \beta_4 - \beta_3 \). Hence \( \beta_4 = \theta + \beta_3 \). Substituting for \( \beta_4 \) in the unrestricted model, one has:

\[ LGE = \beta_1 + \beta_2 S + \beta_3 NUM + (\theta + \beta_3) VERB + u \]

\[ = \beta_1 + \beta_2 S + \beta_3 (NUM + VERB) + \theta VERB + u \]

\[ = \beta_1 + \beta_2 S + 2\beta_3 SCORE + \theta VERB + u. \]
This is the specification of Researcher 3. To test the hypothesis that the restriction is valid, we perform a \(t\) test on the coefficient of \(VERB\). The \(t\) statistic is \(-1.5\), so we do not reject the restriction at the 5 per cent level.

**Explain whether the \(F\) test in (b) and the \(t\) test in (c) could have led to different results.**

No, the \(F\) test and the \(t\) test must give the same result because the \(F\) statistic must be the square of the \(t\) statistic and the critical value of \(F\) must be the square of the critical value of \(t\) for any given significance level. Note that this assumes a two-sided \(t\) test. If one is in a position to perform a one-sided test, the \(t\) test would be more powerful.

**Perform a test of the hypothesis that the numeracy score has a greater effect on earnings than the literacy score.**

One should perform a one-sided \(t\) test on the coefficient of \(VERB\) in regression 3 with the null hypothesis \(H_0: \theta = 0\) and the alternative hypothesis \(H_1: \theta < 0\). The null hypothesis is not rejected and hence one concludes that there is no significant difference.

**Compare the regression results of the three researchers.**

The regression results of Researchers 1 and 3 are equivalent, the only difference being that the coefficient of \(VERB\) provides a direct estimate of \(\beta_4\) in the specification of Researcher 1 and \((\beta_4 - \beta_3)\) in the specification of Researcher 3. Assuming the restriction is valid, there is a large gain in efficiency in the estimation of \(\beta_3\) in specification (2) because its standard error is effectively 0.0005, as opposed to 0.0011 in specifications (1) and (3).

**A6.10 Demonstrate that relationship (3) embodies a testable restriction and show how the model may be reformulated to take advantage of it.**

The coefficients of \(\log \rho\) and \(\log w\) sum to 1. Hence the model should be reformulated as:

\[
\log L = \frac{1}{\alpha + \beta} \log Q + \frac{\alpha}{\alpha + \beta} \log \frac{\rho}{w} \tag{4}
\]

(plus a disturbance term).

**Explain how the restriction could be tested using an \(F\) test.**

Let \(RSS_U\) and \(RSS_R\) be the residual sums of squares from the unrestricted and restricted regressions. To test the null hypothesis that the coefficients of \(\log \rho\) and \(\log w\) sum to 1, one should calculate the \(F\) statistic:

\[
F(1, 27) = \frac{(RSS_R - RSS_U)/1}{RSS_U/27}
\]

and compare it with the critical values of \(F(1, 27)\).

**Explain how the restriction could be tested using a \(t\) test.**

Alternatively, writing (3) as an unrestricted model:

\[
\log wL = \gamma_1 \log Q + \gamma_2 \log \rho + \gamma_3 \log w + u \tag{5}
\]
6. Specification of regression variables

the restriction is $\gamma_2 + \gamma_3 - 1 = 0$. Define $\theta = \gamma_2 + \gamma_3 - 1$. Then $\gamma_3 = \theta - \gamma_2 + 1$ and the unrestricted model may be rewritten as:

$$\log wL = \gamma_1 \log Q + \gamma_2 \log \rho + (\theta - \gamma_2 + 1) \log w + u.$$

Hence:

$$\log wL - \log w = \gamma_1 \log Q + \gamma_2 (\log \rho - \log w) + \theta \log w + u.$$

Hence:

$$\log L = \gamma_1 \log Q + \gamma_2 \log \frac{\rho}{w} + \theta \log w + u.$$

Thus one should regress $\log L$ on $\log Q$, $\log (\rho/w)$, and $\log w$ and perform a $t$ test on the coefficient of $\log w$.

Explain the theoretical benefits of making use of a valid restriction. How could the researcher assess whether there are any benefits in practice, in this case?

The main theoretical benefit of making use of a valid restriction is that one obtains more efficient estimates of the coefficients. The use of a restriction would eliminate the problem of duplicate estimates of the same parameter. Reduced standard errors should provide evidence of the gain in efficiency.

At a seminar, someone suggests that it is reasonable to hypothesise that manufacturing output is subject to constant returns to scale, so that $\alpha + \beta = 1$. Explain how the researcher could test this hypothesis (1) using an F test, (2) using a $t$ test.

Under the assumption of constant returns to scale, the model becomes:

$$\log \frac{L}{Q} = \alpha \log \frac{\rho}{w}.$$  \hfill (6)

One could test the hypothesis by computing the $F$ statistic:

$$F(1, 28) = \frac{(RSS_R - RSS_U)/1}{RSS_U/28}$$

where $RSS_U$ and $RSS_R$ are for the specifications in (4) and (6) respectively.

Alternatively, one could perform a simple $t$ test of the hypothesis that the coefficient of $\log Q$ in (4) is equal to 1.

A6.11 Explain why the researcher was dissatisfied with the results of regression (1).

The high correlation between $I$ and $P$ has given rise to a problem of multicollinearity. The standard errors are relatively large and the $t$ statistics low.

Demonstrate that specification (2) may be considered to be a restricted version of specification (1).

The restriction is $\beta_3 = -\beta_2$. Imposing it, we have:

$$B = \beta_1 + \beta_2 I + \beta_3 P + u$$

$$= \beta_1 + \beta_2 I - \beta_2 P + u$$

$$= \beta_1 + \beta_2 R + u.$$
Perform an F test of the restriction, stating carefully your null hypothesis and conclusion.

The null hypothesis is $H_0 : \beta_3 = -\beta_2$. The test statistic is:

$$F(1, 37) = \frac{(987.1 - 967.9)/1}{967.9/37} = 0.73.$$  

The null hypothesis is not rejected at any significance level since $F < 1$.

Perform a t test of the restriction

The unrestricted specification may be rewritten:

$$B = \beta_1 + \beta_2 I + \beta_3 P + u = \beta_1 + \beta_2 (P + R) + \beta_3 P + u = \beta_1 + (\beta_2 + \beta_3)P + \beta_2 R + u.$$  

Thus a t test on the coefficient of $P$ in this specification is a test of the restriction. The null hypothesis is not rejected, given that the $t$ statistic is 0.86. Of course, the $F$ statistic is the square of the $t$ statistic and the tests are equivalent.

Demonstrate that specification (3) may also be considered to be a restricted version of specification (1)

The restriction is $\beta_3 = 0$.

Perform both an $F$ test and a $t$ test of the restriction in specification (3), stating your conclusion in each case.

$$F(1, 37) = \frac{(1024.3 - 967.9)/1}{967.9/37} = 2.16.$$  

The critical value of $F(1, 37)$ at 5 per cent is approximately 4.08, so the null hypothesis that $P$ does not influence $B$ is not rejected. Of course, with $t = -1.47$, the $t$ test, which is equivalent, leads to the same conclusion.

At a seminar, someone suggests that specification (4) is also a restricted version of specification (1). Is this correct? If so, state the restriction.

No, it is not correct. As shown above, it is an alternative form of the unrestricted specification.

State, with an explanation, which would be your preferred specification.

None of the specifications has been rejected. The second should be preferred because it should be more efficient than the unrestricted specification. The much lower standard error of the slope coefficient provides supportive evidence. The third specification should be eliminated on the grounds that price inflation ought to be a determinant.

A6.12 Write the original model:

$$Y = \beta_1 + \beta_2 X + \beta_3 Z + u.$$  

Then, with:

$$X = 0.5(V + W), \quad Z = 0.5(V - W)$$
6. Specification of regression variables

the other specifications are:

\[ Y = \beta_1 + 0.5(\beta_2 + \beta_3)V + 0.5(\beta_2 - \beta_3)W + u \]  \hspace{1cm} (2)

\[ Y = \beta_1 + \beta_2 V + u \]  \hspace{1cm} (3)

with the implicit restriction \( \beta_3 = \beta_2 \), and, using \( X = V - Z \):

\[ Y = \beta_1 + \beta_2 V + (\beta_3 - \beta_2)Z + u. \]  \hspace{1cm} (4)

(2) and (4) are reparameterisations of (1), so the measures of fit are unchanged: \( E = L = 0.60 \), \( F = M = 200 \).

Given the relationships among the parameters, \( A = 0.70 \), \( C = -0.10 \), \( J = 0.60 \), \( H = 0.20 \).

The standard errors \( B \) and \( D \) cannot be reconstructed because the standard errors of \( \hat{\beta}_2 \) and \( \hat{\beta}_3 \) cannot be used (on their own) to construct standard errors of linear combinations (a loose explanation is acceptable because we have hardly touched on covariances between estimators).

\( K = 0.04 \) since \( J = \) coefficient of \( X \) in specification (1).

The \( F \) statistic for the restriction \( \beta_3 = \beta_2 \) implicit in specification (3) is:

\[ F(1, 40) = \frac{(220 - 200)/1}{200/40} = 4.0. \]

In terms of \( R^2 \) it would be:

\[ F(1, 40) = \frac{(0.60 - G)/1}{0.40/40}. \]

Hence \( G = 0.56 \).

A two-sided \( t \) test on the coefficient of \( Z \) in specification (4) provides an equivalent test of the restriction. The \( t \) statistic must therefore be \( \sqrt{4.0} = 2.0 \) and so \( I = 0.10 \).

[Note: One may also compute \( G \) using the \( t \) statistic for the coefficient of \( V \) in specification (3):

\[ G \] \hspace{1cm} \frac{(1 - G)/41}{t^2}. \]

Yet another was of computing \( G \) is as follows. Since \( R^2 \) in specification (1) is 0.60, \( TSS \) must be 500, using:

\[ R^2 = 1 - \frac{RSS}{TSS}. \]

\( TSS \) is the same in specification (3). Hence one obtains \( G = 0.56 \).]

A6.13 \( F \) statistics should always be computed using \( RSS \), not \( R^2 \). Often the \( R^2 \) version is equivalent, but often it is not, and this is a case in point. The reason is very simple: the dependent variables in the two specifications are different, and so the \( R^2 \) for the specifications are not comparable. The \( RSS \) are comparable because:

\[ LGFDHOPC - LGFDHOPC = (LGFDHO - LGSIZE) - (\hat{LGFDHO} - LGSIZE) \]

\[ = LGFDHO - \hat{LGFDHO}. \]
Chapter 7

Heteroskedasticity

7.1 Overview

This chapter begins with a general discussion of homoskedasticity and heteroskedasticity: the meanings of the terms, the reasons why the distribution of a disturbance term may be subject to heteroskedasticity, and the consequences of the problem for OLS estimators. It continues by presenting several tests for heteroskedasticity and methods of alleviating the problem. It shows how apparent heteroskedasticity may be caused by model misspecification. It concludes with a description of the use of heteroskedasticity-consistent standard errors.

7.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain the concepts of homoskedasticity and heteroskedasticity
- describe how the problem of heteroskedasticity may arise
- explain the consequences of heteroskedasticity for OLS estimators, their standard errors, and $t$ and $F$ tests
- perform the Goldfeld–Quandt test for heteroskedasticity
- perform the White test for heteroskedasticity
- explain how the problem of heteroskedasticity may be alleviated
- explain why a mathematical misspecification of the regression model may give rise to a problem of apparent heteroskedasticity
- explain the use of heteroskedasticity-consistent standard errors.

7.3 Additional exercises

A7.1 Is the disturbance term in your CES expenditure function heteroskedastic?

Sort the data by $EXPPC$. Excluding observations for which $EXPPC$ is zero, regress $CATPC$ on $EXPPC$ and $SIZE$ (a) for the first three-eighths of the non-zero
7. Heteroskedasticity

observations, and (b) for the last three-eighths. Perform a Goldfeld–Quandt test to test for heteroskedasticity in the EXPPC dimension. Repeat using LGCATPC as the dependent variable and regressing it on LGEXPPC and LGSIZE.


A7.3 The observations for the occupational schools (see Chapter 5 in the text) in the figure suggest that a simple linear regression of cost on number of students, restricted to the subsample of these schools, would be subject to heteroskedasticity. Download the data set from the Online Resource Centre and use a Goldfeld–Quandt test to investigate whether this is the case. If the relationship is heteroskedastic, what could be done to alleviate the problem?

A7.4 A researcher hypothesises that larger economies should be more self-sufficient than smaller ones and that \( M/G \), the ratio of imports, \( M \), to gross domestic product, \( G \), should be negatively related to \( G \):

\[
\frac{M}{G} = \beta_1 + \beta_2 G + u
\]

with \( \beta_2 < 0 \). Using data for a sample of 42 countries, with \( M \) and \( G \) both measured in US$ billion, he fits the regression (standard errors in parentheses):

\[
\begin{align*}
\frac{\hat{M}}{G} &= 0.37 - 0.000086G \quad R^2 = 0.12 \\
&\quad (0.03) (0.000036)
\end{align*}
\]

He plots a scatter diagram, reproduced as Figure 7.1, and notices that the ratio \( M/G \) tends to have relatively high variance when \( G \) is small. He also plots a scatter diagram for \( M \) and \( G \), reproduced as Figure 7.2. Defining \( GSQ \) as the square of \( G \), he regresses \( M \) on \( G \) and \( GSQ \):

\[
\begin{align*}
\hat{M} &= 7.27 + 0.30G - 0.000049GSQ \quad R^2 = 0.86 \\
&\quad (10.77) (0.03) (0.000009)
\end{align*}
\]
Finally, he plots a scatter diagram for log $M$ and log $G$, reproduced as Figure 7.3, and regresses log $M$ on log $G$:

$$\hat{\log M} = -0.14 + 0.80 \log G \quad R^2 = 0.78 \quad (3)$$

$$(0.37) \quad (0.07)$$

Having sorted the data by $G$, he tests for heteroskedasticity by regressing specifications (1) – (3) first for the 16 countries with smallest $G$, and then for the 16 countries with the greatest $G$. $RSS_1$ and $RSS_2$, the residual sums of squares for these regressions, are summarised in the following table.

<table>
<thead>
<tr>
<th>Specification</th>
<th>$RSS_1$</th>
<th>$RSS_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.53</td>
<td>0.21</td>
</tr>
<tr>
<td>(2)</td>
<td>3178</td>
<td>71404</td>
</tr>
<tr>
<td>(3)</td>
<td>3.45</td>
<td>3.60</td>
</tr>
</tbody>
</table>

**Figure 7.1:** Scatter diagram of $M/G$ against $G$.

**Figure 7.2:** Scatter diagram of $M$ against $G$. 
7. Heteroskedasticity

Figure 7.3: Scatter diagram of log $M$ against log $G$.

- Discuss whether (1) appears to be an acceptable specification, given the data in the table and Figure 7.1.
- Explain what the researcher hoped to achieve by running regression (2).
- Discuss whether (2) appears to be an acceptable specification, given the data in the table and Figure 7.2.
- Explain what the researcher hoped to achieve by running regression (3).
- Discuss whether (3) appears to be an acceptable specification, given the data in the table and Figure 7.3.
- What are your conclusions concerning the researcher’s hypothesis?

A7.5 A researcher has data on the number of children attending, $N$, and annual recurrent expenditure, $EXP$, measured in US$, for 50 nursery schools in a US city for 2006 and hypothesises that the cost function is of the quadratic form:

$$EXP = \beta_1 + \beta_2N + \beta_3 NSQ + u$$

where $NSQ$ is the square of $N$, anticipating that economies of scale will cause $\beta_3$ to be negative. He fits the following equation:

$$\hat{EXP} = 17999 + 1060N - 1.29 NSQ$$

$R^2 = 0.74$ (1)

(12908) (133) (0.30)

Suspecting that the regression was subject to heteroskedasticity, the researcher runs the regression twice more, first with the 19 schools with lowest enrolments, then with the 19 schools with the highest enrolments. The residual sums of squares in the two regressions are 8.0 million and 64.0 million, respectively.

The researcher defines a new variable, $EXPN$, expenditure per student, as $EXPN = EXP/N$, and fits the equation:

$$\hat{EXPN} = 1080 - 1.25N + 16114 NREC$$

$R^2 = 0.65$ (2)

(90) (0.25) (6000)
7.3. Additional exercises

where $NREC = 1/N$. He again runs regressions with the 19 smallest schools and the 19 largest schools and the residual sums of squares are 900,000 and 600,000.

- Perform a Goldfeld–Quandt test for heteroskedasticity on both of the regression specifications.
- Explain why the researcher ran the second regression.
- $R^2$ is lower in regression (2) than in regression (1). Does this mean that regression (1) is preferable?

**A7.6** This is a continuation of Exercise A6.5.

- When the researcher presents her results at a seminar, one of the participants says that, since $I$ and $G$ have been divided by $Y$, (2) is less likely to be subject to heteroskedasticity than (1). Evaluate this suggestion.

**A7.7** A researcher has data on annual household expenditure on food, $F$, and total annual household expenditure, $E$, both measured in dollars, for 400 households in the United States for 2010. The scatter plot for the data is shown as Figure 7.4. The basic model of the researcher is:

$$F = \beta_1 + \beta_2 E + u \quad (1)$$

where $u$ is a disturbance term. The researcher suspects heteroskedasticity and performs a Goldfeld–Quandt test and a White test. For the Goldfeld–Quandt test, she sorts the data by size of $E$ and fits the model for the subsample with the 150 smallest values of $E$ and for the subsample with the 150 largest values. The residual sums of squares ($RSS$) for these regressions are shown in column (1) of the table. She also fits the regression for the entire sample, saves the residuals, and then fits an auxiliary regression of the squared residuals on $E$ and its square. $R^2$ for this regression is also shown in column (1) in the table. She performs parallel tests of heteroskedasticity for two alternative models:

$$\frac{F}{A} = \beta_1 \frac{1}{A} + \beta_2 \frac{E}{A} + \upsilon \quad (2)$$

$$\log F = \beta_1 + \beta_2 \log E + w. \quad (3)$$

$A$ is household size in terms of equivalent adults, giving each adult a weight of 1 and each child a weight of 0.7. The scatter plot for $F/A$ and $E/A$ is shown as Figure 7.5, and that for $\log F$ and $\log E$ as Figure 7.6. The data for the heteroskedasticity tests for models (2) and (3) are shown in columns (2) and (3) of the table.

<table>
<thead>
<tr>
<th>Specification</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goldfeld–Quandt test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RSS$ smallest 150</td>
<td>200 million</td>
<td>40 million</td>
<td>20.0</td>
</tr>
<tr>
<td>$RSS$ largest 150</td>
<td>820 million</td>
<td>240 million</td>
<td>21.0</td>
</tr>
<tr>
<td><strong>White test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ from auxiliary regression</td>
<td>0.160</td>
<td>0.140</td>
<td>0.001</td>
</tr>
</tbody>
</table>

- Perform the Goldfeld–Quandt test for each model and state your conclusions.
7. Heteroskedasticity

- Explain why the researcher thought that model (2) might be an improvement on model (1).
- Explain why the researcher thought that model (3) might be an improvement on model (1).
- When models (2) and (3) are tested for heteroskedasticity using the White test, auxiliary regressions must be fitted. State the specification of this auxiliary regression for model (2).
- Perform the White test for the three models.
- Explain whether the results of the tests seem reasonable, given the scatter plots of the data.

![Figure 7.4: Scatter diagram of household expenditure on food against total household expenditure.](image)

**Figure 7.4:** Scatter diagram of household expenditure on food against total household expenditure.

![Figure 7.5: Scatter diagram of household expenditure on food per equivalent adult against total household expenditure per equivalent adult.](image)

**Figure 7.5:** Scatter diagram of household expenditure on food per equivalent adult against total household expenditure per equivalent adult.
A7.8 Explain what is correct, mistaken, confused or in need of further explanation in the following statements relating to heteroskedasticity in a regression model:

- ‘Heteroskedasticity occurs when the disturbance term in a regression model is correlated with one of the explanatory variables.’

- ‘In the presence of heteroskedasticity ordinary least squares (OLS) is an inefficient estimation technique and this causes $t$ tests and $F$ tests to be invalid.’

- ‘OLS remains unbiased but it is inconsistent.

- ‘Heteroskedasticity can be detected with a Chow test.’

- ‘Alternatively one can compare the residuals from a regression using half of the observations with those from a regression using the other half and see if there is a significant difference. The test statistic is the same as for the Chow test.’

- ‘One way of eliminating the problem is to make use of a restriction involving the variable correlated with the disturbance term.’

- ‘If you can find another variable related to the one responsible for the heteroskedasticity, you can use it as a proxy and this should eliminate the problem.’

- ‘Sometimes apparent heteroskedasticity can be caused by a mathematical misspecification of the regression model. This can happen, for example, if the dependent variable ought to be logarithmic, but a linear regression is run.’
7. Heteroskedasticity

7.4 Answers to the starred exercises in the textbook

7.5 The following regressions were fitted using the Shanghai school cost data introduced in Section 6.1 (standard errors in parentheses):

\[
\hat{\text{COST}} = 24000 + 339N \quad R^2 = 0.39
\]
\[
(27000) \quad (50)
\]

\[
\hat{\text{COST}} = 51000 - 4000OCC + 152N + 284NOCC \quad R^2 = 0.68.
\]
\[
(31000) \quad (41000) \quad (60) \quad (76)
\]

where \(\text{COST}\) is the annual cost of running a school, \(N\) is the number of students, \(OCC\) is a dummy variable defined to be 0 for regular schools and 1 for occupational schools, and \(NOCC\) is a slope dummy variable defined as the product of \(N\) and \(OCC\). There are 74 schools in the sample. With the data sorted by \(N\), the regressions are fitted again for the 26 smallest and 26 largest schools, the residual sums of squares being as shown in the table.

<table>
<thead>
<tr>
<th>Regression</th>
<th>26 smallest</th>
<th>26 largest</th>
</tr>
</thead>
<tbody>
<tr>
<td>First regression</td>
<td>7.8 \times 10^{10}</td>
<td>54.4 \times 10^{10}</td>
</tr>
<tr>
<td>Second regression</td>
<td>6.7 \times 10^{10}</td>
<td>13.8 \times 10^{10}</td>
</tr>
</tbody>
</table>

Perform a Goldfeld–Quandt test for heteroskedasticity for the two models and, with reference to Figure 6.5, explain why the problem of heteroskedasticity is less severe in the second model.

Answer:

For both regressions \(RSS\) will be denoted \(RSS_1\) for the 26 smallest schools and \(RSS_2\) for the 26 largest schools. In the first regression, 
\[
\frac{RSS_2}{RSS_1} = \frac{(54.4 \times 10^{10})/(7.8 \times 10^{10})} = 6.97.
\]
There are 24 degrees of freedom in each subsample (26 observations, 2 parameters estimated). The critical value of \(F(24, 24)\) is approximately 3.7 at the 0.1 per cent level, and so we reject the null hypothesis of homoskedasticity at that level. In the second regression, 
\[
\frac{RSS_2}{RSS_1} = \frac{(13.8 \times 10^{10})/(6.7 \times 10^{10})} = 2.06.
\]
There are 22 degrees of freedom in each subsample (26 observations, 4 parameters estimated). The critical value of \(F(22, 22)\) is 2.05 at the 5 per cent level, and so we (just) do not reject the null hypothesis of homoskedasticity at that significance level.

Why is the problem of heteroskedasticity less severe in the second regression? The figure in Exercise A7.2 reveals that the cost function is much steeper for the occupational schools than for the regular schools, reflecting their higher marginal cost. As a consequence the two sets of observations diverge as the number of students increases and the scatter is bound to appear heteroskedastic, irrespective of whether the disturbance term is truly heteroskedastic or not. The first regression takes no account of this and the Goldfeld–Quandt test therefore indicates significant heteroskedasticity. In the second regression the problem of apparent heteroskedasticity does not arise because the intercept and slope dummy variables allow separate implicit regression lines for the two types of school.
Looking closely at the diagram, the observations for the occupational schools exhibit a classic pattern of true heteroskedasticity, and this would be confirmed by a Goldfeld–Quandt test confined to the subsample of those schools (see Exercise A7.2). However the observations for the regular schools appear to be homoskedastic and this accounts for the fact that we did not (quite) reject the null hypothesis of homoskedasticity for the combined sample.

7.6 The file educ.dta on the website contains international cross-sectional data on aggregate expenditure on education, EDUC, gross domestic product, GDP, and population, POP, for a sample of 38 countries in 1997. EDUC and GDP are measured in US$ million and POP is measured in thousands. Download the data set, plot a scatter diagram of EDUC on GDP, and comment on whether the data set appears to be subject to heteroskedasticity. Sort the data set by GDP and perform a Goldfeld–Quandt test for heteroskedasticity, running regressions using the subsamples of 14 countries with the smallest and greatest GDP.

Answer:
The figure plots expenditure on education, EDUC, and gross domestic product, GDP, for the 38 countries in the sample, measured in $ billion rather than $ million. The observations exhibit heteroskedasticity. Sorting them by GDP and regressing EDUC on GDP for the subsamples of 14 countries with smallest and greatest GDP, the residual sums of squares for the first and second subsamples, denoted RSS$_1$ and RSS$_2$, respectively, are 1,660,000 and 63,113,000, respectively. Hence:

$$F(12, 12) = \frac{\text{RSS}_2}{\text{RSS}_1} = \frac{63113000}{1660000} = 38.02.$$  

The critical value of $F(12, 12)$ at the 0.1 per cent level is 7.00, and so we reject the null hypothesis of homoskedasticity.

![Figure 7.7: Expenditure on education and GDP ($ billion).](image)

7.9 Repeat Exercise 7.6, using the Goldfeld–Quandt test to investigate whether scaling by population or by GDP, or whether running the regression in logarithmic form,
7. Heteroskedasticity

would eliminate the heteroskedasticity. Compare the results of regressions using the entire sample and the alternative specifications.

**Answer:**

Dividing through by population, \( POP \), the model becomes:

\[
\frac{EDUC}{POP} = \beta_1 \frac{1}{POP} + \beta_2 \frac{GDP}{POP} + \frac{u}{POP}
\]

with expenditure on education per capita, denoted \( EDUC_{POP} \), hypothesised to be a function of gross domestic product per capita, \( GDP_{POP} \), and the reciprocal of population, \( POP_{PREC} \), with no intercept. Sorting the sample by \( GDP_{POP} \) and running the regression for the subsamples of 14 countries with smallest and largest \( GDP_{POP} \), \( RSS_1 = 0.006788 \) and \( RSS_2 = 1.415516 \). Now:

\[
F(12, 12) = \frac{RSS_2}{RSS_1} = \frac{1.415516}{0.006788} = 208.5.
\]

Thus the model is still subject to heteroskedasticity at the 0.1 per cent level. This is evident in Figure 7.8.

![Figure 7.8: Expenditure on education per capita and GDP per capita ($ per capita).](image)

Dividing through instead by \( GDP \), the model becomes:

\[
\frac{EDUC}{GDP} = \beta_1 \frac{1}{GDP} + \beta_2 + \frac{u}{GDP}
\]

with expenditure on education as a share of gross domestic product, denoted \( EDUC_{GDP} \), hypothesised to be a simple function of the reciprocal of gross domestic product, \( GDP_{PREC} \), with no intercept. Sorting the sample by \( GDP_{PREC} \) and running the regression for the subsamples of 14 countries with smallest and largest \( GDP_{PREC} \), \( RSS_1 = 0.00413 \) and \( RSS_2 = 0.00238 \). Since \( RSS_2 \) is less than \( RSS_1 \), we test for heteroskedasticity under the hypothesis that the standard deviation of the disturbance term is inversely related to \( GDP_{PREC} \):

\[
F(12, 12) = \frac{RSS_1}{RSS_2} = \frac{0.00413}{0.00238} = 1.74.
\]
7.4. Answers to the starred exercises in the textbook

Figure 7.9: Expenditure on education as a proportion of GDP and the reciprocal of GDP (measured in $ billion).

The critical value of $F(12, 12)$ at the 5 per cent level is 2.69, so we do not reject the null hypothesis of homoskedasticity. Could one tell this from Figure 7.9? It is a little difficult to say.

Finally, we will consider a logarithmic specification. If the true relationship is logarithmic, and homoskedastic, it would not be surprising that the linear model appeared heteroskedastic. Sorting the sample by GDP, $RSS_1$ and $RSS_2$ are 2.733 and 3.438 for the subsamples of 14 countries with smallest and greatest GDP. The $F$ statistic is:

$$F(12, 12) = \frac{RSS_1}{RSS_2} = \frac{3.438}{2.733} = 1.26.$$  

Thus again we would not reject the null hypothesis of homoskedasticity.

Figure 7.10: Expenditure on education and GDP, logarithmic.
7. Heteroskedasticity

The third and fourth models both appear to be free from heteroskedasticity. How do we choose between them? We will examine the regression results, shown for the two models with the full sample:

```
. reg EDUCGDP GDPREC

Source | SS      df    MS
---------+------------------ Number of obs = 38
Model    | .001348142 1  .001348142
Residual | .008643037 36  .000240084
---------+------------------------------ Adj R-squared = 0.1109
Total    | .009991179 37  .000270032 Root MSE = .01549

------------------------------------------------------------------------------
EDUCGDP | Coef.   Std. Err.  t    P>|t|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
  GDPREC  | -234.0823  98.78309   -2.370  0.023   -434.4236   -33.74086
   _cons  |   .0484593   .0036696   13.205  0.000    .0410169    .0559016

------------------------------------------------------------------------------

. reg LGEE LGGDP

Source | SS      df    MS
---------+------------------ Number of obs = 38
Model    | 51.9905508 1  51.9905508
Residual | 7.6023197 36  .211175547
---------+------------------------------ Adj R-squared = 0.8689
Total    | 59.5928705 37  1.61061812 Root MSE = .45954

------------------------------------------------------------------------------
LGEE     | Coef.   Std. Err.  t    P>|t|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
  LGGDP   |   1.160594   .0739673   15.691  0.000     1.010582    1.310607
   _cons  |  -5.025204   .8152239   -6.164  0.000   -6.678554   -3.371853

In equation form, the first regression is:

\[
\frac{EDUC}{GDP} = 0.048 - \frac{234.1}{GDP} \quad R^2 = 0.13
\]

Multiplying through by \( GDP \), it may be rewritten:

\[
EDUC = -234.1 + 0.048GDP.
\]

It implies that expenditure on education accounts for 4.8 per cent of gross domestic product at the margin. The constant does not have any sensible interpretation. We will compare this with the output from an OLS regression that makes no attempt to eliminate heteroskedasticity:
7.4. Answers to the starred exercises in the textbook

```
. reg EDUC GDP

Source | SS     df MS
--------+---------+---------+----------
Model   | 1.0571e+09 1 1.0571e+09
Residual| 74645819.2 36 2073494.98
--------+---------+---------+----------
Total   | 1.1317e+09 37 3058911.0

Number of obs = 38
F( 1,  36) = 509.80
Prob > F    = 0.0000
R-squared   = 0.9340
Adj R-squared = 0.9322
Root MSE    = 1440.0

EDUC    | Coef.  Std. Err. t P>|t| [95% Conf. Interval]
--------+----------------------+---------+---------+-----------------------------+-----------------------------+---------+---------+
GDP     | .0480656  .0021288 22.579 0.000  .0437482   .052383
_cons   | -160.4669 311.699  -0.515 0.610 -792.6219   471.688
--------+----------------------+---------+---------+-----------------------------+-----------------------------+---------+---------+
```

The slope coefficient, 0.048, is identical to three decimal places. This is not entirely a surprise, since heteroskedasticity does not give rise to bias and so there should be no systematic difference between the estimate from an OLS regression and that from a specification that eliminates heteroskedasticity. Of course, it is a surprise that the estimates are so close. Generally there would be some random difference, and of course the OLS estimate would tend to be less accurate. In this case, the main difference is in the estimated standard error. That for the OLS regression is actually smaller than that for the regression of EDUCGDP on GDPREC, but it is misleading. It is incorrectly calculated and we know that, since OLS is inefficient, the true standard error for the OLS estimate is actually larger.

The logarithmic regression in equation form is:

\[
\log \text{EDUC} = -5.03 + 1.16 \log \text{GDP} \quad R^2 = 0.87
\]

\[
(0.82) \quad (0.07)
\]

implying that the elasticity of expenditure on education with regard to gross domestic product is 1.16. In substance the interpretations of the models are similar, since both imply that the proportion of GDP allocated to education increases slowly with GDP, but the elasticity specification seems a little more informative and probably serves as a better starting point for further exploration. For example, it would be natural to add the logarithm of population to see if population had an independent effect.

7.10 It was reported above that the heteroskedasticity-consistent estimate of the standard error of the coefficient of GDP in equation (7.18) was 0.18. Explain why the corresponding standard error in equation (7.20) ought to be lower and comment on the fact that it is not.

**Answer:**

(7.20), unlike (7.18) appears to be free from heteroskedasticity and therefore should provide more efficient estimates of the coefficients, reflected in lower standard errors when computed correctly. However the sample may be too small for the heteroskedasticity-consistent estimator to be a good guide.

7.11 A health economist plans to evaluate whether screening patients on arrival or spending extra money on cleaning is more effective in reducing the incidence of
infections by the MRSA bacterium in hospitals. She hypothesises the following model:

\[ MRSA_i = \beta_1 + \beta_2 S_i + \beta_3 C_i + u_i \]

where, in hospital \( i \), MRSA is the number of infections per thousand patients, \( S \) is expenditure per patient on screening, and \( C \) is expenditure per patient on cleaning. \( u_i \) is a disturbance term that satisfies the usual regression model assumptions. In particular, \( u_i \) is drawn from a distribution with mean zero and constant variance \( \sigma^2 \). The researcher would like to fit the relationship using a sample of hospitals. Unfortunately, data for individual hospitals are not available. Instead she has to use regional data to fit:

\[ \overline{MRSA}_j = \beta_1 + \beta_2 \overline{S}_j + \beta_3 \overline{C}_j + \overline{u}_j \]

where \( \overline{MRSA}_j, \overline{S}_j, \overline{C}_j \), and \( \overline{u}_j \) are the averages of \( MRSA, S, C, \) and \( u \) for the hospitals in region \( j \). There were different numbers of hospitals in the regions, there being \( n_j \) hospitals in region \( j \).

Show that the variance of \( \overline{u}_j \) is equal to \( \sigma^2 / n_j \) and that an OLS regression using the grouped regional data to fit the relationship will be subject to heteroskedasticity.

Assuming that the researcher knows the value of \( n_j \) for each region, explain how she could re-specify the regression model to make it homoskedastic. State the revised specification and demonstrate mathematically that it is homoskedastic. Give an intuitive explanation of why the revised specification should tend to produce improved estimates of the parameters.

**Answer:**

\[
\text{var}(\overline{u}_j) = \text{var} \left( \frac{1}{n_j} \sum_{k=1}^{n_j} u_{jk} \right) = \left( \frac{1}{n_j} \right)^2 \text{var} \left( \sum_{k=1}^{n_j} u_{jk} \right) = \left( \frac{1}{n_j} \right)^2 n_j \text{var}(u_{jk})
\]

since the covariance terms are all 0. Hence:

\[
\text{var}(\overline{u}_j) = \left( \frac{1}{n_j} \right)^2 n_j \sigma^2 = \frac{\sigma^2}{n_j}.
\]

To eliminate the heteroskedasticity, multiply observation \( j \) by \( \sqrt{n_j} \). The regression becomes:

\[
\sqrt{n_j} \overline{MRSA}_j = \beta_1 \sqrt{n_j} + \beta_2 \sqrt{n_j} \overline{S}_j + \beta_3 \sqrt{n_j} \overline{C}_j + \sqrt{n_j} \overline{u}_j.
\]

The variance of the disturbance term is now:

\[
\text{var} \left( \sqrt{n_j} \overline{u}_j \right) = \left( \sqrt{n_j} \right)^2 \text{var}(\overline{u}_j) = n_j \frac{\sigma^2}{n_j} = \sigma^2
\]

and is thus the same for all observations.

From the expression for \( \text{var}(\overline{u}_j) \), we see that, the larger the group, the more reliable should be its observation (the closer its observation should tend to be to the population relationship). The scaling gives greater weight to the more reliable observations and the resulting estimators should be more efficient.
A7.1 The first step is to drop the zero-observations from the data set and sort it by EXPPC. The $F$ statistic is then computed as:

$$F(n_2 - k, n_1 - k) = \frac{RSS_2/(n_2 - k)}{RSS_1/(n_1 - k)}$$

where $n_1$ and $n_2$ are the number of available observations and $k$ is the number of parameters in the regression specification.

```
drop if FDHO == 0  
(0 observations deleted)  
gen EXPPC = EXP/SIZE  
sort EXPPC  
gen LGEXPPC = ln(EXPPC)  
gen LGSIZE = ln(SIZE)  
gen FDHOPC = FDHO/SIZE  
gen LGFDHOPC = ln(FDHOPC)  
reg FDHOPC EXPPC SIZE in 1/2375  

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 2375</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7382348.18</td>
<td>2</td>
<td>3691174.09</td>
<td>F( 2, 2372) = 278.36</td>
</tr>
<tr>
<td>Residual</td>
<td>31453534.1</td>
<td>2372</td>
<td>13260.3432</td>
<td>R-squared = 0.1901</td>
</tr>
<tr>
<td>Total</td>
<td>38835882.2</td>
<td>2374</td>
<td>16358.8383</td>
<td>Adj R-squared = 0.1894</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 2375</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>40643447.8</td>
<td>2</td>
<td>20321723.9</td>
<td>F( 2, 2372) = 170.94</td>
</tr>
<tr>
<td>Residual</td>
<td>281980931</td>
<td>2372</td>
<td>118878.976</td>
<td>R-squared = 0.1260</td>
</tr>
<tr>
<td>Total</td>
<td>322624379</td>
<td>2374</td>
<td>135899.064</td>
<td>Root MSE = 344.79</td>
</tr>
</tbody>
</table>

| FDHOPC | Coef. | Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|-----|-----|----------------------|
| EXPPC  | .1107869 | .0051862 | 21.36 | 0.000 | .1006169 .1209569   |
| SIZE   | -4.462209 | 1.438899 | -3.10 | 0.002 | -7.283838 -1.640579 |
| _cons  | 85.38055 | 9.590628 | 8.90 | 0.000 | 66.57366 104.1874 |

| FDHOPC | Coef. | Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|-----|-----|----------------------|
| EXPPC  | .0286606 | .0019716 | 14.54 | 0.000 | .0247944 .0325268   |
| SIZE   | -54.33452 | 7.047302 | -7.71 | 0.000 | -68.15403 -40.51501 |
| _cons  | 508.6148 | 22.37631 | 22.73 | 0.000 | 464.7356 552.4939 |
7. Heteroskedasticity

```
. reg LGFDHOPC LGEXPPC LGSIZE in 1/2375
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 2375</th>
</tr>
</thead>
<tbody>
<tr>
<td>F(    2, 2372) = 369.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>207.241064</td>
<td>2</td>
<td>103.620532</td>
<td>369.49</td>
</tr>
<tr>
<td>Residual</td>
<td>665.204785</td>
<td>2372</td>
<td>.280440866</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>872.445849</td>
<td>2374</td>
<td>.367500357</td>
<td>0.2375</td>
</tr>
</tbody>
</table>

```
. reg LGFDHOPC LGEXPPC LGSIZE in 3960/6334
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 2375</th>
</tr>
</thead>
<tbody>
<tr>
<td>F(    2, 2372) = 138.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>94.0495475</td>
<td>2</td>
<td>47.0247737</td>
<td>138.91</td>
</tr>
<tr>
<td>Residual</td>
<td>802.969196</td>
<td>2372</td>
<td>.338519887</td>
<td>0.1048</td>
</tr>
<tr>
<td>Total</td>
<td>897.018744</td>
<td>2374</td>
<td>.377851198</td>
<td>0.1041</td>
</tr>
</tbody>
</table>

```
LGFDHOPC  | Coef.  | Std. Err. | t     | P>|t| | [95% Conf. Interval]
---------|--------|-----------|-------|-----|-----------------------------|
LGEXPPC  | .6510802 | .0265608 | 24.51 | 0.000 | .5989953 - .703165          |
LGSIZE   | -.0567001 | .0198997 | -2.85 | 0.004 | -.0957727 - .0176775       |
_cons    | .6450249 | .1965331 | 3.28  | 0.001 | .2596305 - 1.030419        |
```

```
. reg LGFDHOPC LGEXPPC LGSIZE in 1/2375
```

The $F$ statistic for the linear specification is:

$$F(2372, 2372) = \frac{281980931/2372}{31453534/2372} = 8.97.$$ 

This is significant at the 0.1 per cent level. The corresponding $F$ statistic for the logarithmic specification is 1.21. The critical value of $F(200, 200)$ at the 5 per cent level is 1.26. The critical value for $F(2372, 2372)$ must be lower, so the null hypothesis of homoskedasticity is probably rejected at that level. However, the problem has evidently been largely eliminated.

The logarithmic specification in general appears to be much less heteroskedastic than the linear one and for some categories the null hypothesis of homoskedasticity would not be rejected. Note that for a few of these $RSS_2 < RSS_1$ for the logarithmic specification.
### 7.5. Answers to the additional exercises

The table shows the construction of the White test statistics for the linear and logarithmic specifications for each category of expenditure. The regressors in the auxiliary regression were expenditure per capita and its square, size and its square, and the product of expenditure per capita and size. Hence there were five degrees of freedom for the chi-squared test. The critical values are 11.1 and 15.1 at the 5 per cent and 1 per cent levels. Thus there is strong evidence of heteroskedasticity for all of the categories in the linear specification. There is also evidence for some categories in the logarithmic specification. It is possible that the White test, being more general, is finding evidence of heteroskedasticity not detected by the Goldfeld–Quandt test.

<table>
<thead>
<tr>
<th>Category</th>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( RSS_1 \times 10^{-6} )</th>
<th>( RSS_2 \times 10^{-6} )</th>
<th>( F )</th>
<th>( RSS_1 )</th>
<th>( RSS_2 )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>1,056</td>
<td>1,056</td>
<td>1.95</td>
<td>62.93</td>
<td>32.30</td>
<td>1,324.96</td>
<td>1,593.31</td>
<td>1.20</td>
</tr>
<tr>
<td>CLOT</td>
<td>1,688</td>
<td>1,688</td>
<td>7.17</td>
<td>316.80</td>
<td>44.17</td>
<td>2,107.28</td>
<td>2,196.79</td>
<td>1.04</td>
</tr>
<tr>
<td>DOM</td>
<td>623</td>
<td>623</td>
<td>7.23</td>
<td>238.90</td>
<td>33.05</td>
<td>1,571.19</td>
<td>1,505.92</td>
<td>1.04*</td>
</tr>
<tr>
<td>EDUC</td>
<td>210</td>
<td>210</td>
<td>11.70</td>
<td>376.01</td>
<td>32.15</td>
<td>495.12</td>
<td>507.07</td>
<td>1.02</td>
</tr>
<tr>
<td>ELEC</td>
<td>2,186</td>
<td>2,186</td>
<td>7.55</td>
<td>33.34</td>
<td>4.41</td>
<td>1,034.70</td>
<td>923.18</td>
<td>1.12*</td>
</tr>
<tr>
<td>FDAW</td>
<td>1,913</td>
<td>1,913</td>
<td>9.00</td>
<td>278.13</td>
<td>30.89</td>
<td>1,136.09</td>
<td>1,361.12</td>
<td>1.20</td>
</tr>
<tr>
<td>FDHO</td>
<td>2,375</td>
<td>2,375</td>
<td>31.45</td>
<td>281.98</td>
<td>8.97</td>
<td>665.20</td>
<td>802.97</td>
<td>1.21</td>
</tr>
<tr>
<td>FOOT</td>
<td>685</td>
<td>685</td>
<td>0.55</td>
<td>5.74</td>
<td>10.37</td>
<td>513.08</td>
<td>514.24</td>
<td>1.00</td>
</tr>
<tr>
<td>FURN</td>
<td>183</td>
<td>183</td>
<td>7.17</td>
<td>258.26</td>
<td>36.00</td>
<td>322.50</td>
<td>368.42</td>
<td>1.14</td>
</tr>
<tr>
<td>GASO</td>
<td>2,141</td>
<td>2,141</td>
<td>11.06</td>
<td>159.54</td>
<td>14.43</td>
<td>921.26</td>
<td>1,245.55</td>
<td>1.35</td>
</tr>
<tr>
<td>HEAL</td>
<td>1,801</td>
<td>1,801</td>
<td>32.91</td>
<td>876.72</td>
<td>26.64</td>
<td>2,233.73</td>
<td>2,192.92</td>
<td>1.02*</td>
</tr>
<tr>
<td>HOUS</td>
<td>2,334</td>
<td>2,334</td>
<td>105.48</td>
<td>3,031.19</td>
<td>28.74</td>
<td>2,129.27</td>
<td>1,475.02</td>
<td>1.44*</td>
</tr>
<tr>
<td>LIFE</td>
<td>470</td>
<td>470</td>
<td>2.85</td>
<td>48.37</td>
<td>16.95</td>
<td>503.19</td>
<td>667.14</td>
<td>1.33</td>
</tr>
<tr>
<td>LOCT</td>
<td>260</td>
<td>260</td>
<td>0.58</td>
<td>5.32</td>
<td>9.13</td>
<td>366.16</td>
<td>409.90</td>
<td>1.12</td>
</tr>
<tr>
<td>MAPP</td>
<td>150</td>
<td>150</td>
<td>2.85</td>
<td>37.01</td>
<td>12.96</td>
<td>211.71</td>
<td>243.18</td>
<td>1.15</td>
</tr>
<tr>
<td>PERS</td>
<td>1,431</td>
<td>1,431</td>
<td>0.47</td>
<td>9.01</td>
<td>19.34</td>
<td>1,045.70</td>
<td>1,204.31</td>
<td>1.15</td>
</tr>
<tr>
<td>READ</td>
<td>858</td>
<td>858</td>
<td>0.36</td>
<td>4.95</td>
<td>13.69</td>
<td>1,076.35</td>
<td>1,085.38</td>
<td>1.01</td>
</tr>
<tr>
<td>SAPP</td>
<td>389</td>
<td>389</td>
<td>0.56</td>
<td>10.68</td>
<td>19.04</td>
<td>396.41</td>
<td>433.37</td>
<td>1.09</td>
</tr>
<tr>
<td>TELE</td>
<td>2,171</td>
<td>2,171</td>
<td>3.27</td>
<td>26.80</td>
<td>8.19</td>
<td>1,133.43</td>
<td>1,123.46</td>
<td>1.01*</td>
</tr>
<tr>
<td>TEXT</td>
<td>372</td>
<td>372</td>
<td>0.57</td>
<td>2.05</td>
<td>3.61</td>
<td>410.29</td>
<td>393.80</td>
<td>1.04*</td>
</tr>
<tr>
<td>TOB</td>
<td>433</td>
<td>433</td>
<td>1.56</td>
<td>27.81</td>
<td>17.84</td>
<td>312.71</td>
<td>338.28</td>
<td>1.08</td>
</tr>
<tr>
<td>TOYS</td>
<td>939</td>
<td>939</td>
<td>6.83</td>
<td>87.65</td>
<td>12.83</td>
<td>1,079.76</td>
<td>1,064.92</td>
<td>1.01*</td>
</tr>
<tr>
<td>TRIP</td>
<td>194</td>
<td>194</td>
<td>9.62</td>
<td>77.65</td>
<td>8.07</td>
<td>300.70</td>
<td>335.75</td>
<td>1.12</td>
</tr>
</tbody>
</table>

* indicates \( RSS_2 < RSS_1 \)
7. Heteroskedasticity

<table>
<thead>
<tr>
<th>White tests</th>
<th>Linear</th>
<th>Logarithmic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>ADM</td>
<td>2,815</td>
<td>0.1710</td>
</tr>
<tr>
<td>CLOT</td>
<td>4,500</td>
<td>0.0180</td>
</tr>
<tr>
<td>DOM</td>
<td>1,661</td>
<td>0.0191</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>0.1432</td>
</tr>
<tr>
<td>ELEC</td>
<td>5,828</td>
<td>0.0487</td>
</tr>
<tr>
<td>FDAW</td>
<td>5,102</td>
<td>0.1072</td>
</tr>
<tr>
<td>FDHO</td>
<td>6,334</td>
<td>0.1143</td>
</tr>
<tr>
<td>FOOT</td>
<td>1,827</td>
<td>0.0191</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.3287</td>
</tr>
<tr>
<td>GASO</td>
<td>5,710</td>
<td>0.0575</td>
</tr>
<tr>
<td>HEAL</td>
<td>4,802</td>
<td>0.0608</td>
</tr>
<tr>
<td>HOUS</td>
<td>6,223</td>
<td>0.2002</td>
</tr>
<tr>
<td>LIFE</td>
<td>1,253</td>
<td>0.0535</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.0388</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.0882</td>
</tr>
<tr>
<td>PERS</td>
<td>3,817</td>
<td>0.0607</td>
</tr>
<tr>
<td>READ</td>
<td>2,287</td>
<td>0.0158</td>
</tr>
<tr>
<td>SAPP</td>
<td>1,037</td>
<td>0.0221</td>
</tr>
<tr>
<td>TELE</td>
<td>5,788</td>
<td>0.0724</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.0183</td>
</tr>
<tr>
<td>TOB</td>
<td>1,155</td>
<td>0.0235</td>
</tr>
<tr>
<td>TOYS</td>
<td>2,504</td>
<td>0.0347</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.0571</td>
</tr>
</tbody>
</table>

A7.3 Having sorted by $N$, the number of students, $RSS_1$ and $RSS_2$ are $2.02 \times 10^{10}$ and $22.59 \times 10^{10}$, respectively, for the subsamples of the 13 smallest and largest schools. The $F$ statistic is 11.18. The critical value of $F(11, 11)$ at the 0.1 per cent level must be a little below 8.75, the critical value for $F(10, 10)$, and so the null hypothesis of homoskedasticity is rejected at that significance level.

One possible way of alleviating the heteroskedasticity is by scaling through by the number of students. The dependent variable now becomes the unit cost per student year, and this is likely to be more uniform than total recurrent cost. Scaling through by $N$, and regressing $UNITCOST$, defined as $COST$ divided by $N$, on $NREC$, the reciprocal of $N$, having first sorted by $NREC$, $RSS_1$ and $RSS_2$ are now 349,000 and 504,000. The $F$ statistic is therefore 1.44, and this is not significant even at the 5 per cent level since the critical value must be a little above 2.69, the critical value for $F(12, 12)$. The regression output for this specification using the full sample is shown.

```plaintext
. reg UNITCOST NREC

Source | SS    | df | MS
--------|-------|----|----
Model   | 27010.3792 | 1  | 27010.3792
Residual| 1164624.44  | 32 | 36394.5138
--------+--------|----|----
Total   | 1191634.82 | 33 | 36110.1461

Number of obs = 34
F( 1,   32) = 0.74
Prob > F = 0.3954
R-squared = 0.0227
Adj R-squared = -0.0079
Root MSE = 190.77
```
7.5. Answers to the additional exercises

In equation form, the regression is:

\[
\frac{\hat{COST}}{N} = 524.8 + 10976 \frac{1}{N} \quad R^2 = 0.03
\]

\[
(53.9) \quad (12741)
\]

Multiplying through by \(N\), it may be rewritten:

\[
\hat{COST} = 10976 + 524.8N.
\]

The estimate of the marginal cost is somewhat higher than the estimate of 436 obtained using OLS in Section 5.3 of the text.

A second possible way of alleviating the heteroskedasticity is to hypothesise that the true relationship is logarithmic, in which case the use of an inappropriate linear specification would give rise to apparent heteroskedasticity. Scaling through by \(N\), and regressing \(LGCOST\), the (natural) logarithm of \(COST\), on \(LGN\), the logarithm of \(N\), \(RSS_1\) and \(RSS_2\) are 2.16 and 1.58. The \(F\) statistic is therefore 1.37, and again this is not significant even at the 5 per cent level. The regression output for this specification using the full sample is shown.

```
. reg LGCOST LGN
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs =</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>14.7086057</td>
<td>1</td>
<td>14.7086057</td>
<td>F( 1, 32) =</td>
<td>100.98</td>
</tr>
<tr>
<td>Residual</td>
<td>4.66084501</td>
<td>32</td>
<td>.145651406</td>
<td>Prob &gt; F =</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>19.3694507</td>
<td>33</td>
<td>.58695305</td>
<td>R-squared =</td>
<td>0.7594</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared =</td>
<td>0.7519</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE =</td>
<td>.38164</td>
</tr>
</tbody>
</table>

| LGCOST | Coef. | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------|------|-----------|------|------|-------------------|
| LGN    | .909126 | .0904681 | 10.049 | 0.000 | .7248858 - 1.093404 |
| _cons  | 6.808312 | .5435035 | 12.527 | 0.000 | 5.701232 - 7.915393 |

The estimate of the elasticity of cost with respect to number of students, 0.91, is less than 1 and thus suggests that the schools are subject to economies of scale. However, we are not able to reject the null hypothesis that the elasticity is equal to 1 and thus that costs are proportional to numbers, the \(t\) statistic for the null hypothesis being too low:

\[
t = \frac{0.909 - 1.000}{0.091} = -1.00.
\]
7. Heteroskedasticity

A7.4 Discuss whether (1) appears to be an acceptable specification, given the data in the table and Figure 7.1.

Using the Goldfeld–Quandt test to test specification (1) for heteroskedasticity assuming that the standard deviation of \( u \) is inversely proportional to \( G \), we have:

\[
F(14, 14) = \frac{0.53}{0.21} = 2.52.
\]

The critical value of \( F(14, 14) \) at the 5 per cent level is 2.48, so we just reject the null hypothesis of homoskedasticity at that level. Figure 7.1 does strongly suggest heteroskedasticity. Thus (1) does not appear to be an acceptable specification.

Explain what the researcher hoped to achieve by running regression (2).

If it is true that the standard deviation of \( u \) is inversely proportional to \( G \), the heteroskedasticity could be eliminated by multiplying through by \( G \). This is the motivation for the second specification. An intercept that in principle does not exist has been added, thereby changing the model specification slightly.

Discuss whether (2) appears to be an acceptable specification, given the data in the table and Figure 7.2.

\[
F(13, 13) = \frac{71404}{3178} = 22.47.
\]

The critical value of \( F(13, 13) \) at the 0.1 per cent level is about 6.4, so the null hypothesis of homoskedasticity is rejected. Figure 7.2 confirms the heteroskedasticity.

Explain what the researcher hoped to achieve by running regression (3).

Heteroskedasticity can appear to be present in a regression in natural units if the true relationship is logarithmic. The disturbance term in a logarithmic regression is effectively increasing or decreasing the value of the dependent variable by random proportions. Its effect in absolute terms will therefore tend to be greater, the larger the value of \( G \). The researcher is checking to see if this is the reason for the heteroskedasticity in the second specification.

Discuss whether (3) appears to be an acceptable specification, given the data in the table and Figure 7.3.

Obviously there is no problem with the Goldfeld–Quandt test, since:

\[
F(14, 14) = \frac{3.60}{3.45} = 1.04.
\]

Figure 7.3 looks free from heteroskedasticity.

What are your conclusions concerning the researcher’s hypothesis?

Evidence in support of the hypothesis is provided by (3) where, with:

\[
t = \frac{0.80 - 1}{0.07} = -2.86
\]

the elasticity is significantly lower than 1. Figures 7.1 and 7.2 also strongly suggest that on balance larger economies have lower import ratios than smaller ones.
A7.5 Perform a Goldfeld–Quandt test for heteroskedasticity on both of the regression specifications.

The $F$ statistics for the G–Q test for the two specifications are:

$$F(16, 16) = \frac{64}{16}/\frac{8}{16} = 8.0 \quad \text{and} \quad F(16, 16) = \frac{900}{16}/\frac{600}{16} = 1.5.$$  

The critical value of $F(16, 16)$ is 2.33 at the 5 per cent level and 5.20 at the 0.1 per cent level. Hence one would reject the null hypothesis of homoskedasticity at the 0.1 per cent level for regression 1 and one would not reject it even at the 5 per cent level for regression 2.

Explain why the researcher ran the second regression.

He hypothesised that the standard deviation of the disturbance term in observation $i$ was proportional to $N_i$: $\sigma_i = \lambda N_i$ for some $\lambda$. If this is the case, dividing through by $N_i$ makes the specification homoskedastic, since:

$$\text{var}\left(\frac{u_i}{N_i}\right) = \frac{1}{N^2}\text{var}(u_i) = \frac{1}{N^2}(\lambda N_i)^2 = \lambda^2$$

and is therefore the same for all $i$.

$R^2$ is lower in regression (2) than in regression (1). Does this mean that regression (1) is preferable?

$R^2$ is not comparable because the dependent variable is different in the two regressions. Regression (2) is to be preferred since it is free from heteroskedasticity and therefore ought to tend to yield more precise estimates of the coefficients with valid standard errors.

A7.6 When the researcher presents her results at a seminar, one of the participants says that, since $I$ and $G$ have been divided by $Y$, (2) is less likely to be subject to heteroskedasticity than (1). Evaluate this suggestion.

If the restriction is valid, imposing it will have no implications for the disturbance term and so it could not lead to any mitigation of a potential problem of heteroskedasticity. [If there were heteroskedasticity, and if the specification were linear, scaling through by a variable proportional in observation $i$ to the standard deviation of $u_i$ in observation $i$ would lead to the elimination of heteroskedasticity. The present specification is logarithmic and dividing $I$ and $G$ by $Y$ does not affect the disturbance term.]

A7.7 Perform the Goldfeld–Quandt test for each model and state your conclusions.

The ratios are 4.1, 6.0, and 1.05. In each case we should look for the critical value of $F(148, 148)$. The critical values of $F(150, 150)$ at the 5 per cent, 1 per cent, and 0.1 per cent levels are 1.31, 1.46, and 1.66, respectively. Hence we reject the null hypothesis of homoskedasticity at the 0.1 per cent level (1 per cent is OK) for models (1) and (2). We do not reject it even at the 5 per cent level for model (3).
7. Heteroskedasticity

Explain why the researcher thought that model (2) might be an improvement on model (1).

If the assumption that the standard deviation of the disturbance term is proportional to household size, scaling through by \( A \) should eliminate the heteroskedasticity, since:

\[
E(v^2) = E\left(\frac{u^2}{A^2}\right) = \frac{1}{A^2}E(u^2) = \lambda^2
\]

if the standard deviation of \( u = \lambda A \).

Explain why the researcher thought that model (3) might be an improvement on model (1).

It is possible that the (apparent) heteroskedasticity is attributable to mathematical misspecification. If the true model is logarithmic, a homoskedastic disturbance term would appear to have a heteroskedastic effect if the regression is performed in the original units.

When models (2) and (3) are tested for heteroskedasticity using the White test, auxiliary regressions must be fitted. State the specification of this auxiliary regression for model (2).

The dependent variable is the squared residuals from the model regression. The explanatory variables are the reciprocal of \( A \) and its square, \( E/A \) and its square, and the product of the reciprocal of \( A \) and \( E/A \). (No constant.)

Perform the White test for the three models.

\( nR^2 \) is 64.0, 56.0, and 0.4 for the three models. Under the null hypothesis of homoskedasticity, this statistic has a chi-squared distribution with degrees of freedom equal to the number of terms on the right side of the regression, minus one. This is two for models (1) and (3). The critical value of chi-squared with two degrees of freedom is 5.99, 9.21, and 13.82 at the 5, 1, and 0.1 per cent levels. Hence \( H_0 \) is rejected at the 0.1 per cent level for model (1), and not rejected even at the 5 per cent level for model (3). In the case of model (2), there are five terms on the right side of the regression. The critical value of chi-squared with four degrees of freedom is 18.47 at the 0.1 per cent level. Hence \( H_0 \) is rejected at that level.

Explain whether the results of the tests seem reasonable, given the scatter plots of the data.

Absolutely. In Figures 7.1 and 7.2, the variances of the dispersions of the dependent variable clearly increase with the size of the explanatory variable. In Figure 7.3, the dispersion is much more even.

A7.8 ‘Heteroskedasticity occurs when the disturbance term in a regression model is correlated with one of the explanatory variables.’

This is false. Heteroskedasticity occurs when the variance of the disturbance term is not the same for all observations.
In the presence of heteroskedasticity ordinary least squares (OLS) is an inefficient estimation technique and this causes $t$ tests and $F$ tests to be invalid.

It is true that OLS is inefficient and that the $t$ and $F$ tests are invalid, but ‘and this causes’ is wrong.

OLS remains unbiased but it is inconsistent.

It is true that OLS is unbiased, but false that it is inconsistent.

Heteroskedasticity can be detected with a Chow test.

This is false.

Alternatively one can compare the residuals from a regression using half of the observations with those from a regression using the other half and see if there is a significant difference. The test statistic is the same as for the Chow test.

The first sentence is basically correct with the following changes and clarifications: one is assuming that the standard deviation of the disturbance term is proportional to one of the explanatory variables; the sample should first be sorted according to the size of the explanatory variable; rather than split the sample in half, it would be better to compare the first three-eighths (or one third) of the observations with the last three-eighths (or one third); ‘comparing the residuals’ is too vague: the $F$ statistic is $F(n' - k, n' - k) = \frac{RSS_2}{RSS_1}$ assuming $n'$ observations and $k$ parameters in each subsample regression, and placing the larger $RSS$ over the smaller.

The second sentence is false.

One way of eliminating the problem is to make use of a restriction involving the variable correlated with the disturbance term.

This is nonsense.

If you can find another variable related to the one responsible for the heteroskedasticity, you can use it as a proxy and this should eliminate the problem.

This is more nonsense.

Sometimes apparent heteroskedasticity can be caused by a mathematical misspecification of the regression model. This can happen, for example, if the dependent variable ought to be logarithmic, but a linear regression is run.

True. A homoskedastic disturbance term in a logarithmic regression, which is responsible for proportional changes in the dependent variable, may appear to be heteroskedastic in a linear regression because the absolute changes in the dependent variable will be proportional to its size.
7. Heteroskedasticity
Chapter 8
Stochastic regressors and measurement errors

8.1 Overview

Until this point it has been assumed that the only random element in a regression model is the disturbance term. This chapter extends the analysis to the case where the variables themselves have random components. The initial analysis shows that in general OLS estimators retain their desirable properties. A random component attributable to measurement error, the subject of the rest of the chapter, is however another matter. While measurement error in the dependent variable merely inflates the variances of the regression coefficients, measurement error in the explanatory variables causes OLS estimates of the coefficients to be biased and invalidates standard errors, t tests, and F tests. The analysis is illustrated with reference to the Friedman permanent income hypothesis, the most celebrated application of measurement error analysis in the economic literature. The chapter then introduces instrumental variables (IV) estimation and gives an example of its use to fit the Friedman model. The chapter concludes with a description of the Durbin–Wu–Hausman test for investigating whether measurement errors are serious enough to warrant using IV instead of OLS.

8.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain the conditions under which OLS estimators remain unbiased when the variables in the regression model possess random components
- derive the large-sample expression for the bias in the slope coefficient in a simple regression model with measurement error in the explanatory variable
- demonstrate, within the context of the same model, that measurement error in the dependent variable does not cause the regression coefficients to be biased but does increase their standard errors
- describe the Friedman permanent income hypothesis and explain why OLS estimates of a conventional consumption function will be biased if it is correct
- explain what is meant by an instrumental variables estimator and state the conditions required for its use
8. Stochastic regressors and measurement errors

- demonstrate that the IV estimator of the slope coefficient in a simple regression model is consistent, provided that the conditions required for its use are satisfied
- explain the factors responsible for the population variance of the IV estimator of the slope coefficient in a simple regression model
- perform the Durbin–Wu–Hausman test in the context of measurement error.

8.3 Additional exercises

A8.1 A researcher believes that a variable \( Y \) is determined by the simple regression model:

\[
Y = \beta_1 + \beta_2 X + u.
\]

She thinks that \( X \) is not distributed independently of \( u \) but thinks that another variable, \( Z \), would be a suitable instrument. The instrumental estimator of the intercept, \( \hat{\beta}^{IV}_1 \), is given by:

\[
\hat{\beta}^{IV}_1 = \bar{Y} - \hat{\beta}^{IV}_2 \bar{X}
\]

where \( \hat{\beta}^{IV}_2 \) is the IV estimator of the slope coefficient. [Exercise 8.12 in the textbook asks for a proof that \( \hat{\beta}^{IV}_1 \) is a consistent estimator of \( \beta_1 \).]

Explain, with a brief mathematical proof, why \( \hat{\beta}^{OLS}_1 \), the ordinary least squares estimator of \( \beta_1 \), would be inconsistent, if the researcher is correct in believing that \( X \) is not distributed independently of \( u \).

The researcher has only 20 observations in her sample. Does the fact that \( \hat{\beta}^{IV}_1 \) is consistent guarantee that it has desirable small-sample properties? If not, explain how the researcher might investigate the small-sample properties.

A8.2 Suppose that the researcher in Exercise A8.1 is wrong and \( X \) is in fact distributed independently of \( u \). Explain the consequences of using \( \hat{\beta}^{IV}_1 \) instead of \( \hat{\beta}^{OLS}_1 \) to estimate \( \beta_1 \).

**Note:** The population variance of \( \hat{\beta}^{IV}_1 \) is given by:

\[
\sigma^2_{\hat{\beta}^{IV}_1} = \left( 1 + \frac{\mu^2_X}{\sigma^2_X} \times \frac{1}{r^2_{XZ}} \right) \frac{\sigma^2_u}{n}
\]

where \( \mu_X \) is the population mean of \( X \), \( \sigma^2_X \) is its population variance, \( r_{XZ} \) is the correlation between \( X \) and \( Z \), and \( \sigma^2_u \) is the population variance of the disturbance term, \( u \). For comparison, the population variance of the OLS estimator is:

\[
\sigma^2_{\hat{\beta}^{OLS}_1} = \left( 1 + \frac{\mu^2_X}{\sigma^2_X} \right) \frac{\sigma^2_u}{n}
\]

when the model is correctly specified and the regression model assumptions are satisfied.
A8.3 A researcher investigating the incidence of teenage knife crime has the following data for each of 35 cities for 2008:

- $K =$ number of knife crimes per 1,000 population in 2008
- $N =$ number of teenagers per 1,000 population living in social deprivation in 2008.

The researcher hypothesises that the relationship between $K$ and $N$ is given by:

$$K = \beta_1 + \beta_2 N + u \quad (1)$$

where $u$ is a disturbance term that satisfies the usual regression model assumptions. However, knife crime tends to be under-reported, with the degree of under-reporting worst in the most heavily afflicted boroughs, so that:

$$R = K + w \quad (2)$$

where $R =$ number of reported knife crimes per 1,000 population in 2008 and $w$ is a random variable with $E(w) < 0$ and $\text{cov}(w, K) < 0$. $w$ may be assumed to be distributed independently of $u$. Note that $\text{cov}(w, K) < 0$ implies $\text{cov}(w, N) < 0$. Derive analytically the sign of the bias in the estimator of $\beta_2$ if the researcher regresses $R$ on $N$ using ordinary least squares.

A8.4 Suppose that in the model:

$$Y = \beta_1 + \beta_2 X + u$$

where the disturbance term $u$ satisfies the regression model assumptions, the variable $X$ is subject to measurement error, being underestimated by a fixed amount $\alpha$ in all observations.

- Discuss whether it is true that the ordinary least squares estimator of $\beta_2$ will be biased downwards by an amount proportional to both $\alpha$ and $\beta_2$.
- Discuss whether it is true that the fitted values of $Y$ from the regression will be reduced by an amount $\alpha\beta_2$.
- Discuss whether it is true that $R^2$ will be reduced by an amount proportional to $\alpha$.

A8.5 A researcher believes that the rate of migration from Country B to Country A, $M_t$, measured in thousands of persons per year, is a linear function of the relative average wage, $RW_t$, defined as the average wage in Country A divided by the average wage in Country B, both measured in terms of the currency of Country A:

$$M_t = \beta_1 + \beta_2 RW_t + u_t \quad (1)$$

$u_t$ is a disturbance term that satisfies the regression model assumptions. However, Country B is a developing country with limited resources for statistical surveys and the wage data for that country, derived from a small sample of social security records, are widely considered to be unrepresentative, with a tendency to overstate the true average wage because those working in the informal sector are excluded. As a consequence the measured relative wage, $MRW_t$, is given by

$$MRW_t = RW_t + w_t \quad (2)$$
where \( w_t \) is a random quantity with expected value less than 0. It may be assumed to be distributed independently of \( u_t \) and \( RW_t \).

The researcher also has data on relative GDP per capita, \( RGDP_t \), defined as the ratio of GDP per capita in countries A and B, respectively, both measured in terms of the currency of Country A. He has annual observations on \( M_t \), \( MRW_t \), and \( RGDP_t \) for a 30-year period. The correlation between \( MRW_t \) and \( RGDP_t \) in the sample period is 0.8. Analyse mathematically the consequences for the estimates of the intercept and the slope coefficient, the standard errors and the \( t \) statistics, if the migration equation (1) is fitted:

- using ordinary least squares with \( MRW_t \) as the explanatory variable.
- using OLS, with \( RGDP_t \) as a proxy for \( RW_t \).
- using instrumental variables, with \( RGDP_t \) as an instrument for \( MRW_t \).

A8.6 Suppose that in Exercise A8.5 \( RGDP_t \) is subject to the same kind of measurement error as \( RW_t \), and that as a consequence there is an exact linear relationship between \( RGDP_t \) and \( MRW_t \). Demonstrate mathematically how this would affect the IV estimator of \( \beta_2 \) in part (3) of Exercise A8.5 and give a verbal explanation of your result.

### 8.4 Answers to the starred exercises in the textbook

8.5 A variable \( Q \) is determined by the model:

\[
Q = \beta_1 + \beta_2 X + v
\]

where \( X \) is a variable and \( v \) is a disturbance term that satisfies the regression model assumptions. The dependent variable is subject to measurement error and is measured as \( Y \) where:

\[
Y = Q + r
\]

and \( r \) is the measurement error, distributed independently of \( v \). Describe analytically the consequences of using OLS to fit this model if:

1. The expected value of \( r \) is not equal to zero (but \( r \) is distributed independently of \( Q \)).
2. \( r \) is not distributed independently of \( Q \) (but its expected value is zero).

**Answer:**

Substituting for \( Q \), the model may be rewritten:

\[
Y = \beta_1 + \beta_2 X + v + r = \beta_1 + \beta_2 X + u
\]

where \( u = v + r \). Then:

\[
\hat{\beta}_2 = \beta_2 + \frac{\left( X_i - \bar{X} \right) (u_i - \bar{u})}{\sum \left( X_i - \bar{X} \right)^2} = \beta_2 + \frac{\sum \left( X_i - \bar{X} \right) (v_i - \bar{v}) + \sum \left( X_i - \bar{X} \right) (r_i - \bar{r})}{\sum \left( X_i - \bar{X} \right)^2}
\]
8.4. Answers to the starred exercises in the textbook

and:

\[ E(\hat{\beta}_2) = E \left( \hat{\beta}_2 + \frac{\sum (X_i - \bar{X})(v_i - \bar{v}) + \sum (X_i - \bar{X})(r_i - \bar{r})}{\sum (X_i - \bar{X})^2} \right) \]

\[ = \beta_2 + \frac{1}{\sum (X_i - \bar{X})^2} \left( \sum (X_i - \bar{X})(v_i - \bar{v}) + \sum (X_i - \bar{X})(r_i - \bar{r}) \right) \]

\[ = \beta_2 + \frac{1}{\sum (X_i - \bar{X})^2} \left( \sum (X_i - \bar{X}) E(v_i - \bar{v}) + \sum (X_i - \bar{X}) E(r_i - \bar{r}) \right) \]

\[ = \beta_2 \]

provided that \( X \) is nonstochastic. (If \( X \) is stochastic, the proof that the expected value of the error term is zero is parallel to that in Section 8.2 of the text.) Thus \( \hat{\beta}_2 \) remains an unbiased estimator of \( \beta_2 \).

However, the estimator of the intercept is affected if \( E(r) \) is not zero.

\[ \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \beta_1 + \beta_2 \bar{X} + \bar{u} - \hat{\beta}_2 \bar{X} = \beta_1 + \beta_2 \bar{X} + \bar{v} + \bar{r} - \hat{\beta}_2 \bar{X}. \]

Hence:

\[ E(\hat{\beta}_1) = \beta_1 + \beta_2 \bar{X} + E(\bar{v}) + E(\bar{r}) - E(\hat{\beta}_2 \bar{X}) \]

\[ = \beta_1 + \beta_2 \bar{X} + E(\bar{v}) + E(\bar{r}) - \bar{X}E(\hat{\beta}_2) \]

\[ = \beta_1 + E(\bar{r}). \]

Thus the intercept is biased if \( E(r) \) is not equal to zero, for then \( E(\bar{r}) \) is not equal to zero.

If \( r \) is not distributed independently of \( Q \), the situation is a little bit more complicated. For it to be distributed independently of \( Q \), it must be distributed independently of both \( X \) and \( v \), since these are the determinants of \( Q \). Thus if it is not distributed independently of \( Q \), one of these two conditions must be violated. We will consider each in turn.

(a) \( r \) not distributed independently of \( X \). We now have:

\[ \text{plim} \hat{\beta}_2 = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum (X_i - \bar{X})(v_i - \bar{v}) + \text{plim} \frac{1}{n} \sum (X_i - \bar{X})(r_i - \bar{r})}{\text{plim} \frac{1}{n} \sum (X_i - \bar{X})^2} \]

\[ = \beta_2 + \frac{\sigma_{Xr}}{\sigma_X^2}. \]

Since \( \sigma_{Xr} \neq 0 \), \( \hat{\beta}_2 \) is an inconsistent estimator of \( \beta_2 \). It follows that \( \hat{\beta}_1 \) will also be an inconsistent estimator of \( \beta_1 \):

\[ \hat{\beta}_1 = \beta_1 + \beta_2 \bar{X} + \bar{v} + \bar{r} - \hat{\beta}_2 \bar{X}. \]
8. Stochastic regressors and measurement errors

Hence:

\[
\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \bar{X} + \text{plim } \bar{v} + \text{plim } \bar{r} - \bar{X} \text{plim } \hat{\beta}_2
\]

\[
= \beta_1 + \bar{X}(\beta_2 - \text{plim } \hat{\beta}_2)
\]

and this is different from \( \beta_1 \) if \( \text{plim } \hat{\beta}_2 \) is not equal to \( \beta_2 \).

(b) \( r \) is not distributed independently of \( v \). This condition is not required in the proof of the unbiasedness of either \( \hat{\beta}_1 \) or \( \hat{\beta}_2 \) and so both remain unbiased.

8.6 A variable \( Y \) is determined by the model:

\[
Y = \beta_1 + \beta_2 Z + v
\]

where \( Z \) is a variable and \( v \) is a disturbance term that satisfies the regression model conditions. The explanatory variable is subject to measurement error and is measured as \( X \) where:

\[
X = Z + w
\]

and \( w \) is the measurement error, distributed independently of \( v \). Describe analytically the consequences of using OLS to fit this model if:

(1) the expected value of \( w \) is not equal to zero (but \( w \) is distributed independently of \( Z \))

(2) \( w \) is not distributed independently of \( Z \) (but its expected value is zero).

**Answer:**

Substituting for \( Z \), we have:

\[
Y = \beta_1 + \beta_2 (X - w) + v = \beta_1 + \beta_2 X + u
\]

where \( u = v - \beta_2 w \).

\[
\hat{\beta}_2 = \beta_2 + \frac{\sum (X_i - \bar{X}) (u_i - \bar{u})}{\sum (X_i - \bar{X})^2}.
\]

It is not possible to obtain a closed-form expression for the expectation of the error term since both its numerator and its denominator depend on \( w \). Instead we take plims, having first divided the numerator and the denominator of the error term by \( n \) so that they have limits:

\[
\text{plim } \hat{\beta}_2 = \beta_2 + \frac{\text{plim } \frac{1}{n} \sum (X_i - \bar{X}) (u_i - \bar{u})}{\text{plim } \frac{1}{n} \sum (X_i - \bar{X})^2}
\]

\[
= \beta_2 + \frac{\text{cov}(X, u)}{\text{var}(X)} = \beta_2 + \frac{\text{cov}([Z + w], [v - \beta_2 w])}{\text{var}(X)}
\]

\[
= \beta_2 + \frac{\text{cov}(Z, v) - \beta_2 \text{cov}(Z, w) + \text{cov}(w, v) - \beta_2 \text{cov}(w, w)}{\text{var}(X)}.
\]
8.4. Answers to the starred exercises in the textbook

If \( E(w) \) is not equal to zero, \( \hat{\beta}_2 \) is not affected. The first three terms in the numerator are zero and:

\[
\text{plim} \, \hat{\beta}_2 = \beta_2 + \frac{-\beta_2 \sigma_w^2}{\sigma_X^2}
\]

so \( \hat{\beta}_2 \) remains inconsistent as in the standard case. If \( w \) is not distributed independently of \( Z \), then the second term in the numerator is not 0. \( \hat{\beta}_2 \) remains inconsistent, but the expression is now:

\[
\text{plim} \, \hat{\beta}_2 = \beta_2 + \frac{-\beta_2 (\sigma_{Zw} + \sigma_w^2)}{\sigma_X^2}.
\]

The OLS estimator of the intercept is affected in both cases, but like the slope coefficient, it was inconsistent anyway.

\[
\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \beta_1 + \beta_2 \bar{X} + \bar{u} - \hat{\beta}_2 \bar{X} = \beta_1 + \beta_2 \bar{X} + \bar{v} - \beta_2 \bar{w} - \beta_2 \bar{X}.
\]

Hence:

\[
\text{plim} \, \hat{\beta}_1 = \beta_1 + (\beta_2 - \text{plim} \, \hat{\beta}_2) \bar{X} + \text{plim} \, \bar{v} - \beta_2 \text{plim} \, \bar{w}.
\]

In the standard case this would reduce to:

\[
\text{plim} \, \hat{\beta}_1 = \beta_1 + (\beta_2 - \text{plim} \, \hat{\beta}_2) \bar{X} = \beta_1 + \beta_2 \frac{\sigma_w^2}{\sigma_X^2} \bar{X}.
\]

If \( w \) has expected value \( \mu_w \), not equal to zero:

\[
\text{plim} \, \hat{\beta}_1 = \beta_1 + \beta_2 \left( \frac{\sigma_w^2}{\sigma_X^2} \bar{X} - \mu_w \right).
\]

If \( w \) is not distributed independently of \( Z \):

\[
\text{plim} \, \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\sigma_{Zw} + \sigma_w^2}{\sigma_X^2} \bar{X}.
\]

8.10 A researcher investigating the shadow economy using international crosssectional data for 25 countries hypothesises that consumer expenditure on shadow goods and services, \( Q \), is related to total consumer expenditure, \( Z \), by the relationship:

\[
Q = \beta_1 + \beta_2 Z + v
\]

where \( v \) is a disturbance term that satisfies the regression model assumptions. \( Q \) is part of \( Z \) and any error in the estimation of \( Q \) affects the estimate of \( Z \) by the same amount. Hence:

\[
Y_i = Q_i + w_i
\]

and:

\[
X_i = Z_i + w_i
\]

where \( Y_i \) is the estimated value of \( Q_i \), \( X_i \) is the estimated value of \( Z_i \), and \( w_i \) is the measurement error affecting both variables in observation \( i \). It is assumed that the expected value of \( w \) is 0 and that \( v \) and \( w \) are distributed independently of \( Z \) and of each other.
8. Stochastic regressors and measurement errors

1. Derive an expression for the large-sample bias in the estimate of $\beta_2$ when OLS is used to regress $Y$ on $X$, and determine its sign if this is possible. [Note: The standard expression for measurement error bias is not valid in this case.]

2. In a Monte Carlo experiment based on the model above, the true relationship between $Q$ and $Z$ is:

$$Q = 2.0 + 0.2Z.$$ 

A sample of 25 observations is generated using the integers 1, 2,..., 25 as data for $Z$. The variance of $Z$ is 52.0. A normally distributed random variable with mean 0 and variance 25 is used to generate the values of the measurement error in the dependent and explanatory variables. The results with 10 samples are summarised in the table below. Comment on the results, stating whether or not they support your theoretical analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\hat{\beta}_1$</th>
<th>s.e.($\hat{\beta}_1$)</th>
<th>$\hat{\beta}_2$</th>
<th>s.e.($\hat{\beta}_2$)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.85</td>
<td>1.09</td>
<td>0.42</td>
<td>0.07</td>
<td>0.61</td>
</tr>
<tr>
<td>2</td>
<td>-0.37</td>
<td>1.45</td>
<td>0.36</td>
<td>0.10</td>
<td>0.36</td>
</tr>
<tr>
<td>3</td>
<td>-2.85</td>
<td>0.88</td>
<td>0.49</td>
<td>0.06</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>-2.21</td>
<td>1.59</td>
<td>0.54</td>
<td>0.10</td>
<td>0.57</td>
</tr>
<tr>
<td>5</td>
<td>-1.08</td>
<td>1.43</td>
<td>0.47</td>
<td>0.09</td>
<td>0.55</td>
</tr>
<tr>
<td>6</td>
<td>-1.32</td>
<td>1.39</td>
<td>0.51</td>
<td>0.08</td>
<td>0.64</td>
</tr>
<tr>
<td>7</td>
<td>-3.12</td>
<td>1.12</td>
<td>0.54</td>
<td>0.07</td>
<td>0.71</td>
</tr>
<tr>
<td>8</td>
<td>-0.64</td>
<td>0.95</td>
<td>0.45</td>
<td>0.06</td>
<td>0.74</td>
</tr>
<tr>
<td>9</td>
<td>0.57</td>
<td>0.89</td>
<td>0.38</td>
<td>0.05</td>
<td>0.69</td>
</tr>
<tr>
<td>10</td>
<td>-0.54</td>
<td>1.26</td>
<td>0.40</td>
<td>0.08</td>
<td>0.50</td>
</tr>
</tbody>
</table>

3. The figure below plots the points $(Q, Z)$, represented as circles, and $(Y, X)$, represented as solid markers, for the first sample, with each $(Q, Z)$ point linked to the corresponding $(Y, X)$ point. Comment on this graph, given your answers to parts 1 and 2.

![Graph](image_url)

**Answer:**

1. Substituting for $Q$ and $Z$ in the first equation:

$$(Y - w) = \beta_1 + \beta_2(X - w) + v.$$
8.4. Answers to the starred exercises in the textbook

Hence:

\[ Y = \beta_1 + \beta_2 X + v + (1 - \beta_2)w \]

\[ = \beta_1 + \beta_2 X + u \]

where \( u = v + (1 - \beta_2)w \). So:

\[ \hat{\beta}_2 = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2} \]

It is not possible to obtain a closed-form expression for the expectation of the error term since both its numerator and its denominator depend on \( w \). Instead we take plims, having first divided the numerator and the denominator of the error term by \( n \) so that they have limits:

\[
\text{plim} \hat{\beta}_2 = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum (X_i - \bar{X})(u_i - \bar{u})}{\text{plim} \frac{1}{n} \sum (X_i - \bar{X})^2} = \beta_2 + \frac{\text{cov}(X, u)}{\text{var}(u)} = \beta_2 + \frac{\text{cov}(Z + w, [v + (1 - \beta_2)w])}{\text{var}(X)}
\]

\[ = \beta_2 + \frac{\text{cov}(Z, v) + (1 - \beta_2)\text{cov}(Z, w) + \text{cov}(w, v) + (1 - \beta_2)\text{cov}(w, w)}{\text{var}(X)} \]

Since \( v \) and \( w \) are distributed independently of \( Z \) and of each other, \( \text{cov}(Z, v) = \text{cov}(Z, w) = \text{cov}(w, v) = 0 \), and so:

\[ \text{plim} \hat{\beta}_2 = \beta_2 + (1 - \beta_2)\frac{\sigma_w^2}{\sigma_X^2} \]

\( \beta_2 \) clearly should be positive and less than 1, so the bias is positive.

2. \( \sigma_X^2 = \sigma_Z^2 + \sigma_w^2 \), given that \( w \) is distributed independently of \( Z \), and hence \( \sigma_X^2 = 52 + 25 = 77 \). Thus:

\[ \text{plim} \hat{\beta}_2 = 0.2 + \frac{(1 - 0.2) \times 25}{77} = 0.46 \]

The estimates of the slope coefficient do indeed appear to be distributed around this number.

As a consequence of the slope coefficient being overestimated, the intercept is underestimated, negative estimates being obtained in each case despite the fact that the true value is positive. The standard errors are invalid, given the severe problem of measurement error.

3. The diagram shows how the measurement error causes the observations to be displaced along 45° lines. Hence the slope of the regression line will be a compromise between the true slope, \( \beta_2 \) and 1. More specifically, \( \text{plim} \hat{\beta}_2 \) is a
8. Stochastic regressors and measurement errors

weight average of $\beta_2$ and 1, the weights being proportional to the variances of $Z$ and $w$:

\[
\text{plim } \hat{\beta}_2 = \beta_1 + (1 - \beta_2) \frac{\sigma^2_w}{\sigma^2_Z + \sigma^2_w} = \frac{\sigma^2_Z}{\sigma^2_Z + \sigma^2_w} \beta_2 + \frac{\sigma^2_w}{\sigma^2_Z + \sigma^2_w}.
\]

8.16 It is possible that the ASVABC test score is a poor measure of the kind of ability relevant for earnings. Accordingly, perform an OLS regression of the logarithm of hourly earnings on $S$, $EXP$, $ASVABC$, $MALE$, $ETHBLACK$, and $ETHHISP$ using your EAWE data set and an IV regression using $SM$, $SF$, and $SIBLINGS$ as instruments for $ASVABC$. Perform a Durbin–Wu–Hausman test to evaluate whether $ASVABC$ appears to be subject to measurement error.

**Answer:**

Contrary to expectations, the coefficient of $ASVABC$ is lower in the IV regression. It is 0.048 in the OLS regression and $-0.094$ in the IV regression. The chi-squared statistic, 1.21, is low. One might therefore conclude that there is no serious measurement error and the change in the coefficient is random. Another possibility is that the instruments are too weak. $ASVABC$ is not highly correlated with any of the instruments and the standard error of the coefficient rises from 0.028 in the OLS regression to 0.132 in the IV regression.

```
. ivreg LGEARN S EXP MALE ETHBLACK ETHHISP (ASVABC=SM SF SIBLINGS)
```

**Instrumental variables (2SLS) regression**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>27.63</td>
<td>6</td>
<td>4.60</td>
<td>F(  6, 493) = 22.29</td>
</tr>
<tr>
<td>Residual</td>
<td>121.50</td>
<td>493</td>
<td>.25</td>
<td>R-squared = 0.1853</td>
</tr>
<tr>
<td>Total</td>
<td>149.13</td>
<td>499</td>
<td>.29</td>
<td>Root MSE = .49644</td>
</tr>
</tbody>
</table>

| LGEARN | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|-------|-------|---------------------|
| ASVABC | -.0938253 | .1319694 | -0.71 | 0.477 | -.3531172 .1654666 |
| S      | .1203265  | .0251596  | 4.78  | 0.000 | .0708931 .1697599  |
| EXP    | .0444094  | .0092246  | 4.81  | 0.000 | .026285 .0625338   |
| MALE   | .1909863  | .0456252  | 4.19  | 0.000 | .1013424 .2806302  |
| ETHBLACK| -.1678914 | .1355897 | -1.24 | 0.216 | -.4342963 .0985136 |
| ETHHISP| .0756980  | .0828383  | 0.91  | 0.361 | -.0870617 .2384576 |
| _cons  | .6503199  | .3570741  | 1.82  | 0.069 | -.0512548 1.351895  |

Instrumented: ASVABC

Instruments: S EXP MALE ETHBLACK ETHHISP SM SF SIBLINGS
8.4. Answers to the starred exercises in the textbook

. estimates store IV1

. reg LGEARN S EXP ASVABC MALE ETHBLACK ETHHISP

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F(  6, 493) = 23.81</td>
</tr>
<tr>
<td>Model</td>
<td>33.509496</td>
<td>6</td>
<td>5.58492493</td>
<td>Prob&gt; F = 0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>115.623489</td>
<td>493</td>
<td>.234530403</td>
<td>R-squared = 0.2247</td>
</tr>
<tr>
<td>Total</td>
<td>149.133038</td>
<td>499</td>
<td>.298863804</td>
<td>Root MSE = 0.48428</td>
</tr>
</tbody>
</table>

| LGEARN | Coef. | Std. Err. | t   | P>|t|    | [95% Conf. Interval] |
|--------|-------|-----------|-----|--------|----------------------|
| S      | .0953713 | .0106101 | 8.99 | 0.000  | .0745246 .1162179    |
| EXP    | .043139  | .0089279 | 4.83 | 0.000  | .0255976 .0606805    |
| ASVABC | .0477892 | .0282877 | 1.69 | 0.092  | -.00779 .1033685     |
| MALE   | .1954406 | .0443323 | 4.41 | 0.000  | .1083371 .2825441    |
| ETHBLACK | -.0448382 | .074738  | -0.60 | 0.549  | -.1916824 .102006    |
| ETHHISP | .1226463 | .0692577 | 1.77 | 0.077  | -.0134303 .258723    |
| _cons  | .9766376 | .1938648 | 5.04 | 0.000  | .5957345 1.357541     |

. estimates store OLS1

. hausman IV1 OLS1, constant

---- Coefficients ----
     | (b) | (B) | (b-B) | sqrt(diag(V_b-V_B)) |
-----|-----|-----|-------|----------------------|
| IV1 | .0938253 | .0477892 | -.1416145 | .1289021 |
| OLS1| .1203265 | .0953713 | .0249552 | .022813  |
| ASVABC | -.0444094 | .043139  | .0012704 | .0023208 |
| S    | .1909863 | .1954406 | -.0044543 | .0107847 |
| EXP  | -.1678914 | -.0448382 | -.1230532 | .1131318 |
| MALE | .075698  | .1226463 | -.0469484 | .0454484 |
| ETHBLACK | .6503199 | .9766376 | -.3263177 | .2998639 |

b = consistent under Ho and Ha; obtained from ivreg
B = inconsistent under Ha, efficient under Ho; obtained from regress
Test: Ho: difference in coefficients not systematic
chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 1.21
Prob>chi2 = 0.9908

. cor ASVABC SM SF SIBLINGS
(obs=500)

<table>
<thead>
<tr>
<th></th>
<th>ASVABC</th>
<th>SM</th>
<th>SF</th>
<th>SIBLINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASVABC</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM</td>
<td>0.3426</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>0.3613</td>
<td>0.5622</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>SIBLINGS</td>
<td>-0.2360</td>
<td>-0.3038</td>
<td>-0.2516</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
8. Stochastic regressors and measurement errors

8.17 What is the difference between an instrumental variable and a proxy variable (as described in Section 6.4)? When would you use one and when would you use the other?

**Answer:**

An instrumental variable estimator is used when one has data on an explanatory variable in the regression model but OLS would give inconsistent estimates because the explanatory variable is not distributed independently of the disturbance term. The instrumental variable partially replaces the original explanatory variable in the estimator and the estimator is consistent.

A proxy variable is used when one has no data on an explanatory variable in a regression model. The proxy variable is used as a straight substitute for the original variable. The interpretation of the regression coefficients will depend on the relationship between the proxy and the original variable, and the properties of the other estimators in the model and the tests and diagnostic statistics will depend on the degree of correlation between the proxy and the original variable.

### 8.5 Answers to the additional exercises

**A8.1**

\[
\hat{\beta}_1^{\text{OLS}} = \bar{Y} - \hat{\beta}_2^{\text{OLS}} \bar{X}
\]

\[= \beta_1 + \beta_2 \bar{X} + \bar{u} - \hat{\beta}_2^{\text{OLS}} \bar{X}.\]

Therefore:

\[\text{plim} \hat{\beta}_1^{\text{OLS}} = \beta_1 - (\text{plim} \hat{\beta}_2^{\text{OLS}} - \beta_2) \text{plim} \bar{X} \neq \beta_1.\]

However:

\[
\hat{\beta}_1^{\text{IV}} = \bar{Y} - \hat{\beta}_2^{\text{IV}} \bar{X}
\]

\[= \beta_1 + \beta_2 \bar{X} + \bar{u} - \hat{\beta}_2^{\text{IV}} \bar{X}
\]

\[= \beta_1 - (\hat{\beta}_2^{\text{IV}} - \beta_2) \bar{X} + \bar{u}.\]

Therefore:

\[\text{plim} \hat{\beta}_1^{\text{IV}} = \beta_1 - (\text{plim} \hat{\beta}_2^{\text{IV}} - \beta_2) \text{plim} \bar{X}
\]

\[= \beta_1.\]

Consistency does not guarantee desirable small-sample properties. The latter could be investigated with a Monte Carlo experiment.

**A8.2** Both estimators will be consistent (actually, unbiased) but the IV estimator will be less efficient than the OLS estimator, as can be seen from a comparison of the expressions for the population variances.
A8.3 The regression model is:

\[ R = \beta_1 + \beta_2 N + u + w. \]

Hence:

\[ \hat{\beta}_2^{\text{OLS}} = \beta_2 + \frac{\sum (N_i - \bar{N}) (u_i + w_i - \bar{u} - \bar{w})}{\sum (N_i - \bar{N})^2}. \]

It is not possible to obtain a closed-form expression for the expectation since \( N \) and \( w \) are correlated. Hence, instead, we investigate the plim:

\[
\text{plim} \hat{\beta}_2^{\text{OLS}} = \beta_2 + \text{plim} \frac{\frac{1}{n} \sum (N_i - \bar{N}) (u_i + w_i - \bar{u} - \bar{w})}{\frac{1}{n} \sum (N_i - \bar{N})^2} = \beta_2 + \frac{\text{cov}(N, u) + \text{cov}(N, w)}{\text{var}(N)} < \beta_2
\]

since \( \text{cov}(N, u) = 0 \) and \( \text{cov}(N, w) < 0 \).

A8.4 Discuss whether it is true that the ordinary least squares estimator of \( \beta_2 \) will be biased downwards by an amount proportional to both \( \alpha \) and \( \beta_2 \).

It is not true. Let the measured \( X \) be \( X' \), where \( X' = X - \alpha \). Then:

\[ \hat{\beta}_2^{\text{OLS}} = \frac{\sum (X'_i - X') (Y' - \bar{Y})}{\sum (X'_i - X')^2} = \frac{\sum (X_i - \alpha - [X - \alpha]) (Y_i - \bar{Y})}{\sum (X_i - \alpha - [X - \alpha])^2} = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}. \]

Thus the measurement error has no effect on the estimate of the slope coefficient.

Discuss whether it is true that the fitted values of \( Y \) from the regression will be reduced by an amount proportional to \( \alpha \beta_2 \).

The estimator of the intercept will be \( \bar{Y} - \hat{\beta}_2 \bar{X} = \bar{Y} - \hat{\beta}_2 (\bar{X} - \alpha) \). Hence the fitted value in observation \( i \) will be:

\[ \bar{Y} - \hat{\beta}_2 (\bar{X} - \alpha) + \hat{\beta}_2 X'_i = \bar{Y} - \hat{\beta}_2 (\bar{X} - \alpha) + \hat{\beta}_2 (X_i - \alpha) = \bar{Y} - \hat{\beta}_2 \bar{X} + \hat{\beta}_2 X_i \]

which is what it would be in the absence of the measurement error.

Discuss whether it is true that \( R^2 \) will be reduced by an amount proportional to \( \alpha \).

Since \( R^2 \) is the variance of the fitted values of \( Y \) divided by the variance of the actual values, it will be unaffected.

A8.5 Using ordinary least squares with \( MRW_i \) as the explanatory variable.

\[
\text{plim} \hat{\beta}_2^{\text{OLS}} = \beta_2 - \beta_2 \frac{\sigma_w^2}{\sigma_{Rw}^2 + \sigma_w^2} = \beta_2 \frac{\sigma_{Rw}^2}{\sigma_{Rw}^2 + \sigma_w^2}
\]

(standard theory). Hence the bias is towards zero.

\[
\hat{\beta}_1^{\text{OLS}} = \bar{M} - \hat{\beta}_2^{\text{OLS}} \bar{MRW}
\]

\[
= \beta_1 + \beta_2 \bar{R W} + \bar{u} - \hat{\beta}_2^{\text{OLS}} (\bar{R W} + \bar{w})
\]

\[
= \beta_1 + (\beta_2 - \hat{\beta}_2^{\text{OLS}}) \bar{R W} + \bar{u} - \hat{\beta}_2^{\text{OLS}} \bar{w}
\]
8. Stochastic regressors and measurement errors

and so:

$$\text{plim } \hat{\beta}_{1}^{\text{OLS}} = \beta_1 + \beta_2 \frac{\sigma_w^2}{\sigma_{RW}^2 + \sigma_w^2} \bar{R}W - \beta_2 \frac{\sigma_{RW}^2}{\sigma_w^2 + \sigma_{RW}^2} \mu_w$$

where $\mu_w$ is the population mean of $w$. The first component of the bias will be positive and the second negative, given that $\mu_w$ is negative. It is not possible without further information to predict the direction of the bias. The standard errors and $t$ statistics will be invalidated if there is substantial measurement error in MRW.

Using OLS, with RGDP$_t$ as a proxy for RW.

Suppose $RW = \alpha_1 + \alpha_2 RGDP$. Then the migration equation may be rewritten:

$$M_t = \beta_1 + \beta_2 (\alpha_1 + \alpha_2 RGDP_t) + u_t = (\beta_1 + \alpha_1 \beta_2) + \alpha_2 \beta_2 RGDP_t + u_t.$$

In general it would not be possible to derive estimates of either $\beta_1$ or $\beta_2$. Likewise one has no information on the standard errors of either $\hat{\beta}_1$ or $\hat{\beta}_2$. Nevertheless the $t$ statistic for the slope coefficient would be approximately equal to the $t$ statistic in a regression of $M$ on $RW$, if the proxy is a good one. $R^2$ will be approximately the same as it would have been in a regression of $M$ on $RW$, if the proxy is a good one. One might hypothesise that RGDP might be approximately equal to $RW$, in which case $\alpha_1 = 0$ and $\alpha_2 = 1$ and one can effectively fit the original model.

Using instrumental variables, with RGDP$_t$ as an instrument for MRW$_t$.

The IV estimator of $\beta_2$ is consistent:

$$\hat{\beta}_2^{\text{IV}} = \frac{\sum (M_i - \bar{M}) (RGDP_i - \bar{RGDP})}{\sum (MRW_i - \bar{MRW}) (RGDP_i - \bar{RGDP})} = \beta_2 + \frac{\sum (u_i - \beta_2 w_i - \bar{u} + \beta_2 \bar{w}) (RGDP_i - \bar{RGDP})}{\sum (MRW_i - \bar{MRW}) (RGDP_i - \bar{RGDP})}.$$

Hence plim $\hat{\beta}_2^{\text{IV}} = \beta_2$ if $u$ and $w$ are distributed independently of RGDP. Likewise the IV estimator of $\hat{\beta}_1$ is consistent:

$$\hat{\beta}_1^{\text{IV}} = \bar{M} - \hat{\beta}_2^{\text{IV}} \bar{MRW} = \beta_1 + \beta_2 \bar{RW} + \bar{u} - \hat{\beta}_2^{\text{IV}} \bar{RW} - \hat{\beta}_2^{\text{IV}} \bar{w}.$$

Hence:

$$\text{plim } \hat{\beta}_1^{\text{IV}} = \beta_1 + \beta_2 \bar{RW} + \text{plim } \bar{u} - \text{plim } \hat{\beta}_2^{\text{IV}} \bar{RW} - \text{plim } \hat{\beta}_2^{\text{IV}} \text{plim } \bar{w} = \beta_1$$

since plim $\hat{\beta}_2^{\text{IV}} = \beta_2$ and plim $u = \text{plim } w = 0$. The standard errors will be higher, and hence $t$ statistics lower, than they would have been if it had been possible to run the original regression using OLS.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
A8.6 Suppose $RGDP = \theta + \phi MRW$. Then:

$$\hat{\beta}_2^{IV} = \frac{\sum (M_i - \bar{M})(RGDP_i - \bar{RGDP})}{\sum (MRW_i - \bar{MRW})(RGDP_i - \bar{RGDP})}$$

$$= \frac{\sum (M_i - \bar{M})(\phi MRW_i - \phi \bar{MRW})}{\sum (MRW_i - \bar{MRW})(\phi MRW_i - \phi \bar{MRW})}$$

$$= \hat{\beta}_2^{OLS}.$$ 

The instrument is no longer valid because it is correlated with the measurement error.
8. Stochastic regressors and measurement errors
Chapter 9
Simultaneous equations estimation

9.1 Overview

Until this point the analysis has been confined to the fitting of a single regression equation on its own. In practice, most economic relationships interact with others in a system of simultaneous equations, and when this is the case the application of ordinary least squares (OLS) to a single relationship in isolation yields biased estimates. Having defined what is meant by an endogenous variable, an exogenous variable, a structural equation, and a reduced form equation, the first objective of this chapter is to demonstrate this. The second is to show how it may be possible to use instrumental variables (IV) estimation, with exogenous variables acting as instruments for endogenous ones, to obtain consistent estimates of the coefficients of a relationship. The conditions for exact identification, underidentification, and overidentification are discussed. In the case of overidentification, it is shown how two-stage least squares can be used to obtain estimates that are more efficient than those obtained with simple IV estimation. The chapter concludes with a discussion of the problem of unobserved heterogeneity and the use of the Durbin–Wu–Hausman test in the context of simultaneous equations estimation.

9.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain what is meant by:
  - an endogenous variable
  - an exogenous variable
  - a structural equation
  - a reduced form equation

- explain why the application of OLS to a single equation in isolation is likely to yield inconsistent estimates of the coefficients if the equation is part of a simultaneous equations model

- derive an expression for the large-sample bias in the slope coefficient when OLS is used to fit a simple regression equation in a simultaneous equations model
9. Simultaneous equations estimation

- explain how consistent estimates of the coefficients of an equation in a simultaneous equations model might in principle be obtained using instrumental variables
- explain what is meant by exact identification, underidentification, and overidentification
- explain the principles underlying the use of two-stage least squares, and the reason why it is more efficient than simple IV estimation
- explain what is meant by the problem of unobserved heterogeneity

9.3 Further material

Good governance and economic development

In development economics it has long been observed that there is a positive association between economic performance, \( Y \), and good governance, \( R \), especially in developing countries. However, quantification of the relationship is made problematic by the fact that it is unlikely that causality is unidirectional. While good governance may contribute to economic performance, better performing countries may also develop better institutions. Hence in its simplest form one has a simultaneous equations mode:

\[
Y = \beta_1 + \beta_2 R + u \tag{1}
\]
\[
R = \alpha_1 + \alpha_2 Y + v \tag{2}
\]

where \( u \) and \( v \) are disturbance terms. Assuming that the latter are distributed independently, an OLS regression of the first equation will lead to an upwards biased estimate of \( \beta_2 \), at least in large samples. The proof is left as an exercise (Exercise A9.10). Thus to fit the first equation, one needs an instrument for \( R \). Obviously a better-specified model would have additional explanatory variables in both equations, but there is a problem. In general any variable that influences \( R \) is also likely to influence \( Y \) and is therefore unavailable as an instrument.

In a study of 64 ex-colonial countries that is surely destined to become a classic, ‘The colonial origins of comparative development: an empirical investigation’, *American Economic Review* 91(5): 1369–1401, December 2001, Acemoglu, Johnson, and Robinson (henceforward AJR) argue that settler mortality rates provide a suitable instrument. Put simply, the thesis is that where mortality rates were low, European colonisers founded neo-European settlements with European institutions and good governance. Such settlements eventually prospered. Examples are the United States, Canada, Australia, and New Zealand. Where mortality rates were high, on account of malaria, yellow fever and other diseases for which Europeans had little or no immunity, settlements were not viable. In such countries the main objective of the coloniser was economic exploitation, especially of mineral wealth. Institutional development was not a consideration. Post-independence regimes have often been as predatory as their
predecessors, indigenous rulers taking the place of the former colonisers. Think of the Belgian Congo, first exploited by King Leopold and more recently by Mobutu.

The study is valuable as an example of IV estimation in that it places minimal technical demands on the reader. There is nothing that would not be easily comprehensible to students in an introductory econometrics course that covers IV. Nevertheless, it gives careful attention to the important technical issues. In particular, it discusses at length the validity of the exclusion restriction. To use mortality as an instrument for $R$ in the first equation, one must be sure that it is not a determinant of $Y$ in its own right, either directly or indirectly (other than through $R$).

The conclusion of the study is surprising. According to theory (see Exercise A9.10), the OLS estimate of $\beta_2$ will be biased upwards by the endogeneity of $R$. The objective of the study was to demonstrate that the estimate remains positive and significant even when the upward bias has been removed by using IV. However, the IV estimate turns out to be higher than the OLS estimate. In fact it is nearly twice as large. AJR suggest that this is attributable to measurement error in the measurement of $R$. This would cause the OLS estimate to be biased downwards, and the bias would be removed (asymptotically) by the use of IV. AJR conclude that the downward bias in the OLS estimate caused by measurement error is greater than the upward bias caused by endogeneity.

### 9.4 Additional exercises

**A9.1** In a certain agricultural country, aggregate consumption, $C$, is simply equal to 2,000 plus a random quantity $z$ that depends upon the weather:

$$ C = 2000 + z. $$

$z$ has mean zero and standard deviation 100. Aggregate investment, $I$, is subject to a four-year trade cycle, starting at 200, rising to 300 at the top of the cycle, and falling to 200 in the next year and to 100 at the bottom of the cycle, rising to 200 again the year after that, and so on. Aggregate income, $Y$, is the sum of $C$ and $I$:

$$ Y = C + I. $$

Data on $C$ and $I$, and hence $Y$, are given in the table. $z$ was generated by taking normally distributed random numbers with mean zero and unit standard deviation and multiplying them by 100.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$C$</th>
<th>$I$</th>
<th>$Y$</th>
<th>$t$</th>
<th>$C$</th>
<th>$I$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,813</td>
<td>200</td>
<td>2,013</td>
<td>11</td>
<td>1,981</td>
<td>200</td>
<td>2,181</td>
</tr>
<tr>
<td>2</td>
<td>1,893</td>
<td>300</td>
<td>2,193</td>
<td>12</td>
<td>2,211</td>
<td>100</td>
<td>2,311</td>
</tr>
<tr>
<td>3</td>
<td>2,119</td>
<td>200</td>
<td>2,319</td>
<td>13</td>
<td>2,127</td>
<td>200</td>
<td>2,327</td>
</tr>
<tr>
<td>4</td>
<td>1,967</td>
<td>100</td>
<td>2,067</td>
<td>14</td>
<td>1,953</td>
<td>300</td>
<td>2,253</td>
</tr>
<tr>
<td>5</td>
<td>1,997</td>
<td>200</td>
<td>2,197</td>
<td>15</td>
<td>2,141</td>
<td>200</td>
<td>2,341</td>
</tr>
<tr>
<td>6</td>
<td>2,050</td>
<td>300</td>
<td>2,350</td>
<td>16</td>
<td>1,836</td>
<td>100</td>
<td>1,936</td>
</tr>
<tr>
<td>7</td>
<td>2,035</td>
<td>200</td>
<td>2,235</td>
<td>17</td>
<td>2,103</td>
<td>200</td>
<td>2,303</td>
</tr>
<tr>
<td>8</td>
<td>2,088</td>
<td>100</td>
<td>2,188</td>
<td>18</td>
<td>2,058</td>
<td>300</td>
<td>2,358</td>
</tr>
<tr>
<td>9</td>
<td>2,023</td>
<td>200</td>
<td>2,223</td>
<td>19</td>
<td>2,119</td>
<td>200</td>
<td>2,319</td>
</tr>
<tr>
<td>10</td>
<td>2,144</td>
<td>300</td>
<td>2,444</td>
<td>20</td>
<td>2,032</td>
<td>100</td>
<td>2,132</td>
</tr>
</tbody>
</table>
9. Simultaneous equations estimation

An orthodox economist regresses $C$ on $Y$, using the data in the table, and obtains (standard errors in parentheses):

$$\hat{C} = 512 + 0.68Y \quad R^2 = 0.67$$

(252) (0.11) $F = 36.49$

Explain why this result was obtained, despite the fact that $C$ does not depend on $Y$ at all. In particular, comment on the $t$ and $F$ statistics.

A9.2 A small macroeconomic model of a closed economy consists of a consumption function, an investment function, and an income identity:

$$C_t = \beta_1 + \beta_2Y_t + u_t$$
$$I_t = \alpha_1 + \alpha_2r_t + v_t$$
$$Y_t = C_t + I_t + G_t$$

where $C_t$ is aggregate consumer expenditure in year $t$, $I_t$ is aggregate investment, $G_t$ is aggregate current public expenditure, $Y_t$ is aggregate output, and $r_t$ is the rate of interest. State which variables in the model are endogenous and exogenous, and explain how you would fit the equations, if you could.

A9.3 The model is now expanded to include a demand for money equation and an equilibrium condition for the money market:

$$\begin{align*}
M^d_t &= \delta_1 + \delta_2Y_t + \delta_3r_t + w_t \\
M^d_t &= M_t
\end{align*}$$

where $M^d_t$ is the demand for money in year $t$ and $M_t$ is the supply of money, assumed exogenous. State which variables are endogenous and exogenous in the expanded model and explain how you would fit the equations, including those in Exercise A9.2, if you could.

A9.4 Table 9.2 reports a simulation comparing OLS and IV parameter estimates and standard errors for 10 samples. The reported $R^2$ (not shown in that table) for the OLS and IV regressions are shown in the table below.

<table>
<thead>
<tr>
<th>Sample</th>
<th>OLS $R^2$</th>
<th>IV $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.59</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>0.69</td>
<td>0.52</td>
</tr>
<tr>
<td>3</td>
<td>0.78</td>
<td>0.73</td>
</tr>
<tr>
<td>4</td>
<td>0.61</td>
<td>0.37</td>
</tr>
<tr>
<td>5</td>
<td>0.40</td>
<td>0.06</td>
</tr>
<tr>
<td>6</td>
<td>0.72</td>
<td>0.57</td>
</tr>
<tr>
<td>7</td>
<td>0.60</td>
<td>0.33</td>
</tr>
<tr>
<td>8</td>
<td>0.58</td>
<td>0.44</td>
</tr>
<tr>
<td>9</td>
<td>0.69</td>
<td>0.43</td>
</tr>
<tr>
<td>10</td>
<td>0.39</td>
<td>0.13</td>
</tr>
</tbody>
</table>
9.4. Additional exercises

We know that, for large samples, the IV estimator is preferable to the OLS estimator because it is consistent, while the OLS estimator is inconsistent. However, do the smaller OLS standard errors in Table 9.2 and the larger OLS values of $R^2$ in the present table indicate that OLS is actually preferable for small samples ($n = 20$ in the simulation)?

A9.5 A researcher investigating the relationship between aggregate wages, $W$, aggregate profits, $P$, and aggregate income, $Y$, postulates the following model:

$$W = \beta_1 + \beta_2 Y + u$$  \hspace{1cm} (1)

$$P = \alpha_1 + \alpha_2 Y + \alpha_3 K + v$$  \hspace{1cm} (2)

$$Y = W + P$$  \hspace{1cm} (3)

where $K$ is aggregate stock of capital and $u$ and $v$ are disturbance terms that satisfy the usual regression model assumptions and may be assumed to be distributed independently of each other. The third equation is an identity, all forms of income being classified either as wages or as profits. The researcher intends to fit the model using data from a sample of industrialised countries, with the variables measured on a per capita basis in a common currency. $K$ may be assumed to be exogenous.

- Explain why ordinary least squares (OLS) would yield inconsistent estimates if it were used to fit (1) and derive the large-sample bias in the slope coefficient.
- Explain what can be inferred about the finite-sample properties of OLS if used to fit (1).
- Demonstrate mathematically how one might obtain a consistent estimate of $\beta_2$ in (1).
- Explain why (2) is not identified (underidentified).
- Explain whether (3) is identified.
- At a seminar, one of the participants asserts that it is possible to obtain an estimate of $\alpha_2$ even though equation (2) is underidentified. Any change in income that is not a change in wages must be a change in profits, by definition, and so one can estimate $\alpha_2$ as $(1 - \hat{\beta}_2)$, where $\hat{\beta}_2$ is the consistent estimate of $\beta_2$ found in the third part of this question. The researcher does not think that this is right but is confused and says that he will look into it after the seminar. What should he have said?

A9.6 A researcher has data on $e$, the annual average rate of growth of employment, $x$ the annual average rate of growth of output, and $p$, the annual average rate of growth of productivity, for a sample of 25 countries, the average rates being calculated for the period 1995–2005 and expressed as percentages. The researcher hypothesises that the variables are related by the following model:

$$e = \beta_1 + \beta_2 x + u$$  \hspace{1cm} (1)

$$x = e + p.$$  \hspace{1cm} (2)

The second equation is an identity because $p$ is defined as the difference between $x$ and $e$. The researcher believes that $p$ is exogenous. The correlation coefficient for $x$ and $p$ is 0.79.
9. Simultaneous equations estimation

- Explain why the OLS estimator of $\beta_2$ would be inconsistent, if the researcher’s model is correctly specified. Derive analytically the large-sample bias, and state whether it is possible to determine its sign.

- Explain how the researcher might use $p$ to construct an IV estimator of $\beta_2$, that is consistent if $p$ is exogenous. Demonstrate analytically that the estimator is consistent.

- The OLS and IV regressions are summarised below (standard errors in parentheses). Comment on them, making use of your answers to the first two parts of this question.

  \[
  \text{OLS} \quad \hat{e} = -0.52 + 0.48x \quad (3)
  \]

  \[
  \text{IV} \quad \hat{e} = 0.37 + 0.17x \quad (4)
  \]

- A second researcher hypothesises that both $x$ and $p$ are exogenous and that equation (2) should be written:

  \[ e = x - p. \]

  On the assumption that this is correct, explain why the slope coefficients in (3) and (4) are both biased and determine the direction of the bias in each case.

- Explain what would be the result of fitting (5), regressing $e$ on $x$ and $p$.

A9.7 A researcher has data from the World Bank World Development Report 2000 on $F$, average fertility (average number of children born to each woman during her life), $M$, under-five mortality (number of children, per 100, dying before reaching the age of 5), and $S$, average years of female schooling, for a sample of 54 countries. She hypothesises that fertility is inversely related to schooling and positively related to mortality, and that mortality is inversely related to schooling:

\[
F = \beta_1 + \beta_2 S + \beta_3 M + u \quad (1)
\]

\[
M = \alpha_1 + \alpha_2 S + v \quad (2)
\]

where $u$ and $v$ are disturbance terms that may be assumed to be distributed independently of each other. $S$ may be assumed to be exogenous.

- Derive the reduced form equations for $F$ and $M$.

- Explain what would be the most appropriate method to fit equation (1).

- Explain what would be the most appropriate method to fit equation (2).

The researcher decides to fit (1) using ordinary least squares, and she decides also to perform a simple regression of $F$ on $S$, again using ordinary least squares, with the following results (standard errors in parentheses):

\[
\hat{F} = 4.08 - 0.17S + 0.015M \quad R^2 = 0.83 \quad (3)
\]

\[
\hat{F} = 6.99 - 0.36S \quad R^2 = 0.71 \quad (4)
\]
9.4. Additional exercises

- Explain why the coefficient of $S$ differs in the two equations.
- Explain whether one may validly perform $t$ tests on the coefficients of (4).

At a seminar someone hypothesises that female schooling may be negatively influenced by fertility, especially in the poorer developing countries in the sample, and this would affect (4). To investigate this, the researcher adds the following equation to the model:

$$ S = \delta_1 + \delta_2 F + \delta_3 G + w \quad (5) $$

where $G$ is GNP per capita and $w$ is a disturbance term. She regresses $F$ on $S$ (1) instrumenting for $S$ with $G$ (column (b) in the output below), and (2) using ordinary least squares, as in equation (4) (column (B) in the output below). The correlation between $S$ and $G$ was 0.70. She performs a Durbin–Wu–Hausman test to compare the coefficients.

<table>
<thead>
<tr>
<th>---- Coefficients ----</th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>-.2965323</td>
<td>-.3637397</td>
<td>.0672074</td>
<td>.0347484</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>6.162605</td>
<td>6.992907</td>
<td>-.8303019</td>
<td>.4194891</td>
<td></td>
</tr>
</tbody>
</table>

Test: Ho: difference in coefficients not systematic

$$ \text{chi2}(1) = (b-B)'[(V_{b-V_B})^{(-1)}](b-B) $$

$$ = 3.31 $$

$$ \text{Prob>chi2} = 0.1158 $$

- Discuss whether $G$ is likely to be a valid instrument.
- What should the researcher’s conclusions be with regard to the test?

A9.8 Aggregate demand $Q_D$ for a certain commodity is determined by its price, $P$, aggregate income, $Y$, and population, $POP$:

$$ Q_D = \beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + u_D $$

and aggregate supply is given by:

$$ Q_S = \alpha_1 + \alpha_2 P + u_S $$

where $u_D$ and $u_S$ are independently distributed disturbance terms.

- Demonstrate that the estimator of $\alpha_2$ will be inconsistent if ordinary least squares (OLS) is used to fit the supply equation, showing that the large-sample bias is likely to be negative.
- Demonstrate that a consistent estimator of $\alpha_2$ will be obtained if the supply equation is fitted using instrumental variables (IV), using $Y$ as an instrument.

The model is used for a Monte Carlo experiment, with $\alpha_2$ set equal to 0.2 and suitable values chosen for the other parameters. The table shows the estimates of
9. Simultaneous equations estimation

\( \alpha_2 \) obtained in 10 samples using OLS, using IV with \( Y \) as an instrument, using IV with \( POP \) as an instrument, and using two-stage least squares (TSLS) with \( Y \) and \( POP \). s.e. is standard error. The correlation between \( P \) and \( Y \) averaged 0.50 across the samples. The correlation between \( P \) and \( POP \) averaged 0.63 across the samples. Discuss the results obtained.

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>IV with ( Y )</th>
<th>IV with ( POP )</th>
<th>TSLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>coef.</td>
<td></td>
<td>coef.</td>
<td>coef.</td>
<td>coef.</td>
</tr>
<tr>
<td>s.e.</td>
<td></td>
<td>s.e.</td>
<td>s.e.</td>
<td>s.e.</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.22</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
<td>0.24</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.11</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>0.18</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>0.16</td>
<td>0.20</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
<td>0.27</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.09</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>0.14</td>
<td>0.24</td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.08</td>
<td>0.05</td>
<td>0.20</td>
</tr>
<tr>
<td>7</td>
<td>0.20</td>
<td>0.22</td>
<td>0.26</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.05</td>
<td>0.25</td>
<td>0.03</td>
</tr>
<tr>
<td>8</td>
<td>0.15</td>
<td>0.21</td>
<td>0.24</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.06</td>
<td>0.23</td>
<td>0.04</td>
</tr>
<tr>
<td>9</td>
<td>0.11</td>
<td>0.17</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.05</td>
<td>0.15</td>
<td>0.03</td>
</tr>
<tr>
<td>10</td>
<td>0.17</td>
<td>0.16</td>
<td>0.24</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.05</td>
<td>0.20</td>
<td>0.03</td>
</tr>
</tbody>
</table>

A9.9 A researcher has the following data for a sample of 1,000 manufacturing enterprises on the following variables, each measured as an annual average for the period 2001–2005: \( G \), average annual percentage rate of growth of sales; \( R \), expenditure on research and development; and \( A \), expenditure on advertising. \( R \) and \( A \) are measured as a proportion of sales revenue. He hypothesises the following model:

\[
G = \beta_1 + \beta_2 R + \beta_3 A + u_G \quad (1)
\]

\[
R = \alpha_1 + \alpha_2 G + u_R \quad (2)
\]

where \( u_G \) and \( u_R \) are disturbance terms distributed independently of each other.

A second researcher believes that expenditure on quality control, \( Q \), measured as a proportion of sales revenue, also influences the growth of sales, and hence that the first equation should be written:

\[
G = \beta_1 + \beta_2 R + \beta_3 A + \beta_4 Q + u_G. \quad (1^*)
\]

\( A \) and \( Q \) may be assumed to be exogenous variables.

- Derive the reduced form equation for \( G \) for the first researcher.
- Explain why ordinary least squares (OLS) would be an inconsistent estimator of the parameters of equation (2).
- The first researcher uses instrumental variables (IV) to estimate \( \alpha_2 \) in (2). Explain the procedure and demonstrate that the IV estimator of \( \alpha_2 \) is consistent.
- The second researcher uses two stage least squares (TSLS) to estimate \( \alpha_2 \) in (2). Explain the procedure and demonstrate that the TSLS estimator is consistent.
9.4. Additional exercises

- Explain why the TSLS estimator used by the second researcher ought to produce ‘better’ results than the IV estimator used by the first researcher, if the growth equation is given by (1*). Be specific about what you mean by ‘better’.

- Suppose that the first researcher is correct and the growth equation is actually given by (1), not (1*). Compare the properties of the two estimators in this case.

- Suppose that the second researcher is correct and the model is given by (1*) and (2), but A is not exogenous after all. Suppose that A is influenced by G:

  \[ A = \gamma_1 + \gamma_2 G + u_A \]  

  (3)

  where \( u_A \) is a disturbance term distributed independently of \( u_G \) and \( u_R \). How would this affect the properties of the IV estimator of \( \alpha_2 \) used by the first researcher?

A9.10 A researcher has data for 100 workers in a large organisation on hourly earnings, EARNINGS, skill level of the worker, SKILL, and a measure of the intelligence of the worker, IQ. She hypothesises that LGEARN, the natural logarithm of EARNINGS, depends on SKILL, and that SKILL depends on IQ.

  \[
  \begin{align*}
  \text{LGEARN} &= \beta_1 + \beta_2 \text{SKILL} + u \\
  \text{SKILL} &= \alpha_1 + \alpha_2 \text{IQ} + v
  \end{align*}
  \]  

  (1)

  (2)

  where \( u \) and \( v \) are disturbance terms. The researcher is not sure whether \( u \) and \( v \) are distributed independently of each other.

- State, with a brief explanation, whether each variable is endogenous or exogenous, and derive the reduced form equations for the endogenous variables.

- Explain why the researcher could use ordinary least squares (OLS) to fit equation (1) if \( u \) and \( v \) are distributed independently of each other.

- Show that the OLS estimator of \( \beta_2 \) is inconsistent if \( u \) and \( v \) are positively correlated and determine the direction of the large-sample bias.

- Demonstrate mathematically how the researcher could use instrumental variables (IV) estimation to obtain a consistent estimate of \( \beta_2 \).

- Explain the advantages and disadvantages of using IV, rather than OLS, to estimate \( \beta_2 \), given that the researcher is not sure whether \( u \) and \( v \) are distributed independently of each other.

- Describe in general terms a test that might help the researcher decide whether to use OLS or IV. What are the limitations of the test?

- Explain whether it is possible for the researcher to fit equation (2) and obtain consistent estimates.
9. Simultaneous equations estimation

A9.11 This exercise relates to the Further material section.

In general in an introductory econometrics course, issues and problems are treated separately, one at a time. In practice in empirical work, it is common for multiple problems to be encountered simultaneously. When this is the case, the one-at-a-time analysis may no longer be valid. In the case of the AJR study, both endogeneity and measurement error seem to be issues. This exercise looks at both together, within the context of that model.

Let $S$ be the correct good governance variable and let $R$ be the measured variable, with measurement error $w$. Thus the model may be written:

\[
Y = \beta_1 + \beta_2 S + u \\
S = \alpha_1 + \alpha_2 Y + v \\
R = S + w.
\]

It may be assumed that $w$ has zero expectation and constant variance $\sigma^2_w$ across observations, and that it is distributed independently of $S$ and the disturbance terms in the equations in the model. Investigate the likely direction of the bias in the OLS estimator of $\beta_2$ in large samples.

9.5 Answers to the starred exercises in the textbook

9.1 A simple macroeconomic model consists of a consumption function and an income identity:

\[
C = \beta_1 + \beta_2 Y + u \\
Y = C + I
\]

where $C$ is aggregate consumption, $I$ is aggregate investment, $Y$ is aggregate income, and $u$ is a disturbance term. On the assumption that $I$ is exogenous, derive the reduced form equations for $C$ and $Y$.

**Answer:**

Substituting for $Y$ in the first equation:

\[C = \beta_1 + \beta_2(C + I) + u.\]

Hence:

\[C = \frac{\beta_1}{1 - \beta_2} + \frac{\beta_2 I}{1 - \beta_2} + \frac{u}{1 - \beta_2}\]

and:

\[Y = C + I = \frac{\beta_1}{1 - \beta_2} + \frac{I}{1 - \beta_2} + \frac{u}{1 - \beta_2}.\]

9.2 It is common to write an earnings function with the logarithm of the hourly wage as the dependent variable and characteristics such as years of schooling, cognitive ability, years of work experience, etc as the explanatory variables. Explain whether
such an equation should be regarded as a reduced form equation or a structural equation.

**Answer:**

In the conventional model of the labour market, the wage rate and the quantity of labour employed are both endogenous variables jointly determined by the interaction of demand and supply. According to this model, the wage equation is a reduced form equation.

9.3 In the simple macroeconomic model:

\[
C = \beta_1 + \beta_2 Y + u
\]

\[
Y = C + I
\]

described in Exercise 9.1, demonstrate that OLS would yield inconsistent results if used to fit the consumption function, and investigate the direction of the bias in the slope coefficient.

**Answer:**

The first step in the analysis of the OLS slope coefficient is to break it down into the true value and error component in the usual way:

\[
\hat{\beta}_{OLS}^2 = \frac{\sum (Y_i - \bar{Y})(C_i - \bar{C})}{\sum (Y_i - \bar{Y})^2} = \beta_2 + \frac{\sum (Y_i - \bar{Y})(u_i - \bar{u})}{\sum (Y_i - \bar{Y})^2}.
\]

From the reduced form equation in Exercise 9.1 we see that \(Y\) depends on \(u\) and hence we will not be able to obtain a closed-form expression for the expectation of the error term. Instead we take plims, having first divided the numerator and the denominator of the error term by \(n\) so that they will possess limits as \(n\) goes to infinity.

\[
\text{plim} \hat{\beta}_{OLS}^2 = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum (Y_i - \bar{Y})(u_i - \bar{u})}{\text{plim} \frac{1}{n} \sum (Y_i - \bar{Y})^2} = \beta_2 + \frac{\text{cov}(Y, u)}{\text{var}(Y)}.
\]

We next substitute for \(Y\) since it is an endogenous variable. We have two choices: we could substitute from the structural equation, or we could substitute from the reduced form. If we substituted from the structural equation, in this case the income identity, we would introduce another endogenous variable, \(C\), and we would find ourselves going round in circles. So we must choose the reduced form.

\[
\text{plim} \hat{\beta}_{OLS}^2 = \beta_2 + \frac{\text{cov} \left( \frac{\beta_1}{1-\beta_2} + \frac{I}{1-\beta_2} + \frac{u}{1-\beta_2} \right, u)}{\text{var} \left( \frac{\beta_1}{1-\beta_2} + \frac{I}{1-\beta_2} + \frac{u}{1-\beta_2} \right)}
\]

\[
= \beta_2 + \frac{1}{1-\beta_2} \left( \text{cov}(I, u) + \text{cov}(u, u) \right)
\]

\[
= \beta_2 + \frac{1}{1-\beta_2} \frac{\text{cov}(I, u) + \text{var}(u)}{\text{var}(I) + \text{var}(u) + 2\text{cov}(I, u)}.
\]
9. Simultaneous equations estimation

On the assumption that \( I \) is exogenous, it is distributed independently of \( u \) and \( \text{cov}(I,u) = 0 \). So:

\[
\text{plim} \hat{\beta}_2^{\text{OLS}} = \beta_2 + (1 - \beta_2) \frac{\sigma_u^2}{\sigma_I^2 + \sigma_u^2}
\]

since the sample variances tend to the population variances as the sample becomes large. Since the variances are positive, the sign of the bias depends on the sign of \((1 - \beta_2)\). It is reasonable to assume that the marginal propensity to consume is positive and less than 1, in which case this term will be positive and the large-sample bias in \( \hat{\beta}_2^{\text{OLS}} \) will be upwards.

The OLS estimate of the intercept is also inconsistent:

\[
\hat{\beta}_1^{\text{OLS}} = \bar{C} - \hat{\beta}_2^{\text{OLS}} \bar{Y} = \beta_1 + \beta_2 \bar{Y} + \bar{u} - \hat{\beta}_2^{\text{OLS}} \bar{Y}.
\]

Hence:

\[
\text{plim} \hat{\beta}_1^{\text{OLS}} = \beta_1 + (\beta_2 - \text{plim} \hat{\beta}_2^{\text{OLS}}) \text{plim} \bar{Y} = \beta_1 - (1 - \beta_2) \frac{\sigma_u^2}{\sigma_I^2 + \sigma_u^2} \text{plim} \bar{Y}.
\]

This is evidently biased downwards, as one might expect, given that the slope coefficient was biased upwards.

9.6 The table gives consumption per capita, \( C \), gross fixed capital formation per capita, \( I \), and gross domestic product per capita, \( Y \), all measured in US$, for 33 countries in 1998. The output from an OLS regression of \( C \) on \( Y \), and an IV regression using \( I \) as an instrument for \( Y \), are shown. Comment on the differences in the results.

<table>
<thead>
<tr>
<th></th>
<th>( C )</th>
<th>( I )</th>
<th>( Y )</th>
<th></th>
<th>( C )</th>
<th>( I )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>15,024</td>
<td>4,749</td>
<td>19,461</td>
<td>South Korea</td>
<td>4,596</td>
<td>1,448</td>
<td>6,829</td>
</tr>
<tr>
<td>Austria</td>
<td>19,813</td>
<td>6,787</td>
<td>26,104</td>
<td>Luxembourg</td>
<td>26,400</td>
<td>9,767</td>
<td>42,650</td>
</tr>
<tr>
<td>Belgium</td>
<td>18,367</td>
<td>5,174</td>
<td>24,522</td>
<td>Malaysia</td>
<td>1,683</td>
<td>873</td>
<td>3,268</td>
</tr>
<tr>
<td>Canada</td>
<td>15,786</td>
<td>4,017</td>
<td>20,855</td>
<td>Mexico</td>
<td>3,359</td>
<td>1,056</td>
<td>4,328</td>
</tr>
<tr>
<td>China–PR</td>
<td>446</td>
<td>293</td>
<td>768</td>
<td>Netherlands</td>
<td>17,558</td>
<td>4,865</td>
<td>24,086</td>
</tr>
<tr>
<td>China–HK</td>
<td>17,067</td>
<td>7,262</td>
<td>24,452</td>
<td>New Zealand</td>
<td>11,236</td>
<td>2,658</td>
<td>13,992</td>
</tr>
<tr>
<td>Denmark</td>
<td>25,199</td>
<td>6,947</td>
<td>32,769</td>
<td>Norway</td>
<td>23,415</td>
<td>9,221</td>
<td>32,933</td>
</tr>
<tr>
<td>Finland</td>
<td>17,991</td>
<td>4,741</td>
<td>24,952</td>
<td>Pakistan</td>
<td>389</td>
<td>79</td>
<td>463</td>
</tr>
<tr>
<td>France</td>
<td>19,178</td>
<td>4,622</td>
<td>24,587</td>
<td>Philippines</td>
<td>760</td>
<td>176</td>
<td>868</td>
</tr>
<tr>
<td>Germany</td>
<td>20,058</td>
<td>5,716</td>
<td>26,219</td>
<td>Portugal</td>
<td>8,579</td>
<td>2,644</td>
<td>9,976</td>
</tr>
<tr>
<td>Greece</td>
<td>9,991</td>
<td>2,460</td>
<td>11,551</td>
<td>Spain</td>
<td>11,255</td>
<td>3,415</td>
<td>14,052</td>
</tr>
<tr>
<td>Iceland</td>
<td>25,294</td>
<td>6,706</td>
<td>30,622</td>
<td>Sweden</td>
<td>20,687</td>
<td>4,487</td>
<td>26,866</td>
</tr>
<tr>
<td>India</td>
<td>291</td>
<td>84</td>
<td>385</td>
<td>Switzerland</td>
<td>27,648</td>
<td>7,815</td>
<td>36,864</td>
</tr>
<tr>
<td>Indonesia</td>
<td>351</td>
<td>216</td>
<td>613</td>
<td>Thailand</td>
<td>1,226</td>
<td>479</td>
<td>1,997</td>
</tr>
<tr>
<td>Ireland</td>
<td>13,045</td>
<td>4,791</td>
<td>20,132</td>
<td>UK</td>
<td>19,743</td>
<td>4,316</td>
<td>23,844</td>
</tr>
<tr>
<td>Italy</td>
<td>16,134</td>
<td>4,075</td>
<td>20,580</td>
<td>USA</td>
<td>26,387</td>
<td>6,540</td>
<td>32,377</td>
</tr>
<tr>
<td>Japan</td>
<td>21,478</td>
<td>7,923</td>
<td>30,124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Christopher Dougherty, 2016. All rights reserved. Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
9.5. Answers to the starred exercises in the textbook

. reg C Y

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2.5686e+09</td>
<td>1</td>
<td>2.5686e+09</td>
<td>F( 1, 31) = 1331.29</td>
</tr>
<tr>
<td>Residual</td>
<td>59810749.2</td>
<td>31</td>
<td>1929379.01</td>
<td>Prob&gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2.6284e+09</td>
<td>32</td>
<td>82136829.4</td>
<td>R-squared = 0.9772</td>
</tr>
</tbody>
</table>

| C | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|---|-------|-----------|---|------|------------------|
| Y | 0.7303066 | 0.0200156 | 36.49 | 0.000 | 0.6894845 - 0.7711287 |
| _cons | 379.4871 | 443.6764 | 0.86 | 0.399 | -525.397 - 1284.371 |

. ivregress 2sls C (Y=I)

Instrumental variables (2SLS) regression

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2.5686e+09</td>
<td>1</td>
<td>2.5686e+09</td>
<td>Wald chi2(1) = 1269.09</td>
</tr>
<tr>
<td>Residual</td>
<td>59810749.2</td>
<td>31</td>
<td>1929379.01</td>
<td>Prob&gt; chi2 = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2.6284e+09</td>
<td>32</td>
<td>82136829.4</td>
<td>R-squared = 0.9770</td>
</tr>
</tbody>
</table>

| C | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|---|-------|-----------|---|------|------------------|
| Y | 0.7183909 | 0.0201658 | 35.62 | 0.000 | 0.6788667 - 0.7579151 |
| _cons | 600.946 | 442.7386 | 1.36 | 0.175 | -266.8057 - 1468.698 |

Instrumented: Y
Instruments: I

Answer:

Assuming the simple macroeconomic model:

\[ C = \beta_1 + \beta_2 Y + u \]
\[ Y = C + I \]

where \( C \) is consumption per capita, \( I \) is investment per capita, and \( Y \) is income per capita, and \( I \) is assumed exogenous, the OLS estimator of the marginal propensity to consume will be biased upwards. As was shown in Exercise 9.3:

\[ \text{plim}\ \hat{\beta}_2^{OLS} = \beta_2 + (1 - \beta_2) \frac{\sigma_u^2}{\sigma_I^2 + \sigma_u^2}. \]

Hence the IV estimate should be expected to be lower, but only by a small amount, given the data. With \( \hat{\beta}_2 \) estimated at 0.72, \( (1 - \hat{\beta}_2) \) is 0.28. \( \sigma_u^2 \) is estimated at 1.95 million and \( \sigma_I^2 \) is 7.74 million. Hence, on the basis of these estimates, the bias should be about 0.06. The actual difference in the OLS and IV estimates is smaller still. However, the actual difference would depend on the purely random sampling error as well as the bias, and it is possible that in this case the sampling error happens to have offset the bias to some extent.
9. Simultaneous equations estimation

9.11 Consider the price inflation/wage inflation model given by equations (9.1) and (9.2):

\[ p = \beta_1 + \beta_2 w + u_p \]
\[ w = \alpha_1 + \alpha_2 p + \alpha_3 U + u_w. \]

We have seen that the first equation is exactly identified, \( U \) being used as an instrument for \( w \). Suppose that TSLS is applied to this model, despite the fact that it is exactly identified, rather than overidentified. How will the results differ?

**Answer:**

If we fit the reduced form, we obtain a fitted equation:

\[ \hat{w} = h_1 + h_2 U. \]

The TSLS estimator is then given by

\[
\hat{\beta}_2^{TSLS} = \frac{\sum (\hat{w}_i - \hat{w})(p_i - \bar{p})}{\sum (\hat{w}_i - \bar{w})(w_i - \bar{w})} = \frac{\sum (h_1 + h_2 U_i - h_1 - h_2 \bar{U})(p_i - \bar{p})}{\sum (h_1 + h_2 U_i - h_1 - h_2 \bar{U})(w_i - \bar{w})}
\]

\[
= \frac{\sum h_2 (U_i - \bar{U})(p_i - \bar{p})}{\sum h_2 (U_i - \bar{U})(w_i - \bar{w})} = \hat{\beta}_2^{IV}
\]

where \( \hat{\beta}_2^{IV} \) is the IV estimator using \( U \). Hence the estimator is exactly the same.

[Note: This is a special case of Exercise 8.18.]

9.15 Suppose the first equation in the model in Box 9.2 is fitted, with \( Q \) used as an instrument for \( Y \). Describe the likely properties of the estimator of \( \alpha_2 \).

**Answer:**

The first equation in Box 9.2 is:

\[ X = \alpha_1 + \alpha_2 Y + u \]

The reduced form equation for \( Y \) is:

\[ Y = \frac{1}{1 - \alpha_2 \beta_2} (\beta_1 + \alpha_1 \beta_2 + \beta_2 u + v). \]

\( Q \) is not a valid instrument for \( Y \) because it is not a determinant of \( Y \).

Mathematically, it can be shown that:

\[ \text{plim} \hat{\alpha}_2^{IV} = \alpha_2 + \frac{\text{cov}(Q, u)}{\text{cov}(Q, Y)}. \]

The numerator of the second term is zero, but so is its denominator and therefore the expression is undefined.
9.6 Answers to the additional exercises

A9.1 The positive coefficient of $Y_t$ in the regression is attributable wholly to simultaneous equations bias. The three figures show this graphically.

The first diagram shows what the time series for $C_t$, $I_t$, and $Y_t$ would look like if there were no random component of consumption. The series for $C_t$ is constant at 2,000. That for $I_t$ is a wave form, and that for $Y_t$ is the same wave form shifted upward by 2,000. The second diagram shows the effect of adding the random component to consumption. $Y_t$ still has a wave form, but there is a clear correlation between it and $C_t$. 
In the third diagram, \( C_t \) is plotted against \( Y_t \), with and without the random component. The three large circles represent the data when there is no random component. One circle represents the five data points \([C = 2,000, Y = 2,100]\); the middle circle represents the ten data points \([C = 2,000, Y = 2,200]\); and the other circle represents the five data points \([C = 2,000, Y = 2,300]\). A regression line based on these three points would be horizontal (the dashed line). The solid circles represent the 20 data points when the random component is affecting \( C_t \) and \( Y_t \), and the solid line is the regression line for these points. Note that these 20 data points fall into three groups: five which lie on a 45 degree line through the left large circle, 10 which lie on the 45 degree line through the middle circle (actually, you can only see nine), and five on the 45 degree line through the right circle.

If OLS is used to fit the equation:

\[
\hat{\beta}_{OLS}^2 = \frac{\sum (Y_i - \bar{Y})(C_i - \bar{C})}{\sum (Y_i - \bar{Y})^2} = \frac{\sum (Y_i - \bar{Y})([2000 + z_i] - [2000 + \hat{z}])}{\sum (Y_i - \bar{Y})^2} = \frac{\sum (Y_i - \bar{Y})(z_i - \hat{z})}{\sum (Y_i - \bar{Y})^2}.
\]

Note that at this stage we have broken down the slope coefficient into its true value plus an error term. The true value does not appear explicitly because it is zero, so we only have the error term. We cannot take expectations because both the numerator and the denominator are functions of \( z \):

\[
Y = C + I = 2000 + I + z.
\]

\( z \) is a component of \( C \) and hence of \( Y \). As a second-best procedure, we investigate the large-sample properties of the estimator by taking \( \text{plims} \). We must first divide the numerator and denominator by \( n \) so that they tend to finite limits:

\[
\text{plim} \hat{\beta}_{OLS}^2 = \frac{\text{plim} \frac{1}{n} \sum (Y_i - \bar{Y})(z_i - \hat{z})}{\text{plim} \frac{1}{n} \sum (Y_i - \bar{Y})^2} = \frac{\text{cov}(Y, z)}{\text{var}(Y)}.
\]

Substituting for \( Y \) from its reduced form equation:

\[
\text{plim} \hat{\beta}_{OLS}^2 = \frac{\text{cov}([2000 + I + z], z)}{\text{var}(2000 + I + z)} = \frac{\text{cov}(I, z) + \text{var}(z)}{\text{var}(I) + \text{var}(z) + 2\text{cov}(I, z)} = \frac{\sigma_z^2}{\sigma_I^2 + \sigma_z^2}.
\]
cov(I, z) = 0 because I is distributed independently of z. σ^2_z is equal to 10,000 (since we are told that σ_z is equal to 100). Over a four-year cycle, the mean value of I is 200 and hence its population variance is given by:

\[ \sigma^2_I = \frac{1}{4} [0 + 100^2 + 0 + (-100)^2] = 5000. \]

Hence:

\[ \text{plim } \beta_{OLS}^2 = \frac{10000}{15000} = 0.67. \]

The actual coefficient in the 20-observation sample, 0.68, is very close to this (probably atypically close for such a model).

The estimator of the intercept, whose true value is 2,000, is biased downwards because \( \beta_{OLS}^2 \) is biased upwards. The standard errors of the coefficients are invalid because the regression model assumption B.7 is violated, and hence t tests would be invalid.

By virtue of the fact that \( Y = C + I \), C is being regressed against a variable which is largely composed of itself. Hence the high \( R^2 \) is inevitable, despite the fact that there is no behavioural relationship between C and Y. Mathematically, \( R^2 \) is equal to the square of the sample correlation between the actual and fitted values of C. Since the fitted values of C are a linear function of the values of Y, \( R^2 \) is equal to the square of the sample correlation between C and Y. The population correlation coefficient is given by

\[ \rho_{C,Y} = \frac{\text{cov}(C,Y)}{\sqrt{\text{var}(C)\text{var}(Y)}} = \frac{\text{cov}([2000 + z],[2000 + I + z])}{\sqrt{\text{var}([2000 + z])\text{var}([2000 + I + z])}} = \frac{\text{var}(z)}{\sqrt{\text{var}(z)\text{var}(I + z)}} = \frac{\sigma^2_z}{\sqrt{\sigma^2_z (\sigma^2_I + \sigma^2_z)}}. \]

Hence in large samples:

\[ R^2 = \frac{10000^2}{10000^2 + 5000^2} = 0.67. \]

\( R^2 \) in the regression is exactly equal to this, the closeness probably being something of a coincidence.

Since regression model assumption B.7 is violated, the F statistic cannot be used to perform an F test of goodness of fit.

A9.2 \( C_t, I_t, \) and \( Y_t \) are endogenous, the first two being the dependent variables of the behavioural relationships and the third being defined by an identity. \( G_t \) and \( r_t \) are exogenous.

Either \( I_t \) or \( r_t \) could be used as an instrument for \( Y_t \) in the consumption function. If it can be assumed that \( u_t \) and \( v_t \) are distributed independently, it can also be regarded as exogenous as far as the determination of \( C_t \) and \( Y_t \) are concerned. It would be preferable to \( r_t \) since it is more highly correlated with \( Y_t \). One’s first thought, then, would be to use TSLS, with the first stage fitting the equation:

\[ Y_t = \frac{\beta_1}{1 - \beta_2} + I_t \frac{1}{1 - \beta_2} + G_t \frac{1}{1 - \beta_2} + u_t \frac{1}{1 - \beta_2}. \]
9. Simultaneous equations estimation

Note, however, that the equation implies the restriction that the coefficients of $I_t$ and $G_t$ are equal. Hence all one has to do is to define a variable:

$$Z_t = I_t + G_t$$

and use $Z_t$ as an instrument for $Y_t$ in the consumption function.

The investment function would be fitted using OLS since $r_t$ is exogenous. The income identity does not need to be fitted.

A9.3 $M_t^d$ is endogenous because it is determined by the second of the two new relationships. The addition of the first of these relationships makes $r_t$ endogenous. To see this, substituting for $C_t$ and $I_t$ in the income identity, using the consumption function and the investment function, one obtains:

$$Y_t = \frac{(\alpha_1 + \beta_1) + \alpha_2 r_t + u_t + v_t}{1 - \beta_2}.$$  

This is usually known as the IS curve. Substituting for $M_t^d$ in the first of the two new relationships, using the second, one has:

$$M_t = \delta_1 + \delta_2 Y_t + \delta_3 r_t + w_t.$$  

This is usually known as the LM curve. The equilibrium values of both $Y_t$ and $r_t$ are determined by the intersection of these two curves and hence $r_t$ is endogenous as well as $Y_t$. $G_t$ remains exogenous, as before, and $M_t$ is also exogenous.

The consumption and investment functions are overidentified and one would use TSLS to fit them, the exogenous variables being government expenditure and the supply of money. The demand for money equation is exactly identified, two of the explanatory variables, $r_t$ and $Y_t$, being endogenous, and the two exogenous variables being available to act as instruments for them.

A9.4 The OLS standard errors are invalid so a comparison is illegitimate. They are not of any great interest anyway because the OLS estimator is biased. Figure 9.3 in the text shows that the variance of the OLS estimator is smaller than that of the IV estimator, but, using a criterion such as the mean square error, there is no doubt that the IV estimator should be preferred. The comment about $R^2$ is irrelevant. OLS has a better fit but we have had to abandon the least squares principle because it yields inconsistent estimates.

A9.5 Explain why ordinary least squares (OLS) would yield inconsistent estimates if it were used to fit (1) and derive the large-sample bias in the slope coefficient.

At some point we will need the reduced form equation for $Y$. Substituting into the third equation from the first two, and rearranging, it is:

$$Y = \frac{1}{1 - \alpha_2 - \beta_2} (\alpha_1 + \beta_1 + \alpha_3 K + u + v).$$

Since $Y$ depends on $u$, the assumption that the disturbance term be distributed independently of the regressors is violated in (1).

$$\hat{\beta}_2^{OLS} = \frac{\sum (Y_i - \bar{Y})(W_i - \bar{W})}{\sum (Y_i - \bar{Y})^2} = \beta_2 + \frac{\sum (Y_i - \bar{Y})(u_i - \bar{u})}{\sum (Y_i - \bar{Y})^2}.$$
after substituting for \( W \) from (1) and simplifying. We are not able to obtain a closed-form expression for the expectation of the error term because \( u \) influences both its numerator and denominator, directly and by virtue of being a component of \( Y \), as seen in the reduced form. Dividing both the numerator and denominator by \( n \), and noting that:

\[
\operatorname{plim} \frac{1}{n} \sum \left( Y_i - \bar{Y} \right)^2 = \operatorname{var}(Y)
\]
as a consequence of a law of large numbers, and that it can also be shown that:

\[
\operatorname{plim} \frac{1}{n} \sum \left( Y_i - \bar{Y} \right) (u_i - \bar{u}) = \operatorname{cov}(Y, u)
\]
we can write

\[
\operatorname{plim} \hat{\beta}_2^{OLS} = \beta_2 + \frac{\operatorname{plim} \frac{1}{n} \sum \left( Y_i - \bar{Y} \right) (u_i - \bar{u})}{\operatorname{plim} \frac{1}{n} \sum \left( Y_i - \bar{Y} \right)^2} = \beta_2 + \frac{\operatorname{cov}(Y, u)}{\operatorname{var}(Y)}.
\]

Now:

\[
\operatorname{cov}(Y, u) = \operatorname{cov}
\left( \frac{1}{1 - \alpha_2 - \beta_2} \left( \alpha_1 + \beta_1 + \alpha_3 K + u + v \right), u \right)
\]

\[
= \frac{1}{1 - \alpha_2 - \beta_2} \left( \alpha_3 \operatorname{cov}(K, u) + \operatorname{var}(u) + \operatorname{cov}(v, u) \right)
\]

the covariance of \( u \) with the constants being zero. Since \( K \) is exogenous, \( \operatorname{cov}(K, u) = 0 \). We are told that \( u \) and \( v \) are distributed independently of each other, and so \( \operatorname{cov}(u, v) = 0 \). Hence:

\[
\operatorname{plim} \hat{\beta}_2^{OLS} = \beta_2 + \frac{1}{1 - \alpha_2 - \beta_2} \frac{\sigma_u^2}{\operatorname{plim} \operatorname{var}(Y)}.
\]

From the reduced form equation for \( Y \) it is evident that \((1 - \alpha_2 - \beta_2) > 0\), and so the large-sample bias will be positive.

**Explain what can be inferred about the finite-sample properties of OLS if used to fit (1).**

It is not possible for an estimator that is unbiased in a finite sample to develop a bias if the sample size increases. Therefore, since the estimator is biased in large samples, it must also be biased in finite ones. The plim may well be a guide to the mean of the estimator in a finite sample, but this is not guaranteed and it is unlikely to be exactly equal to the mean.

**Demonstrate mathematically how one might obtain a consistent estimate of \( \beta_2 \) in (1).**

Use \( K \) as an instrument for \( Y \):

\[
\hat{\beta}_2^{IV} = \frac{\sum \left( K_i - \bar{K} \right) \left( W_i - \bar{W} \right)}{\sum \left( K_i - \bar{K} \right) \left( Y_i - \bar{Y} \right)} = \beta_2 + \frac{\sum \left( K_i - \bar{K} \right) (u_i - \bar{u})}{\sum \left( K_i - \bar{K} \right) \left( Y_i - \bar{Y} \right)}
\]
9. Simultaneous equations estimation

after substituting for $W$ from (1) and simplifying. We are not able to obtain a closed-form expression for the expectation of the error term because $u$ influences both its numerator and denominator, directly and by virtue of being a component of $Y$, as seen in the reduced form. Dividing both the numerator and denominator by $n$, and noting that it can be shown that:

$$\text{plim} \frac{1}{n} \sum (K_i - \bar{K}) (u_i - \bar{u}) = \text{cov}(K, u) = 0$$

since $K$ is exogenous, and that:

$$\text{plim} \frac{1}{n} \sum (K_i - \bar{K}) (Y_i - \bar{Y}) = \text{cov}(K, Y)$$

we can write:

$$\text{plim} \hat{\beta}_2^W = \beta_2 + \frac{\text{cov}(K, u)}{\text{cov}(K, Y)} = \beta_2.$$

$\text{cov}(K, Y)$ is non-zero since the reduced form equation for $Y$ reveals that $K$ is a determinant of $Y$. Hence the instrumental variable estimator is consistent.

Explain why (2) is not identified (underidentified).

(2) is underidentified because the endogenous variable $Y$ is a regressor and there is no valid instrument to use with it. The only potential instrument is the exogenous variable $K$ and it is already a regressor in its own right.

Explain whether (3) is identified.

(3) is an identity so the issue of identification does not arise.

At a seminar, one of the participants asserts that it is possible to obtain an estimate of $\alpha_2$ even though equation (2) is underidentified. Any change in income that is not a change in wages must be a change in profits, by definition, and so one can estimate $\alpha_2$ as $(1 - \hat{\beta}_2)$, where $\hat{\beta}_2$ is the consistent estimate of $\beta_2$ found in the third part of this question. The researcher does not think that this is right but is confused and says that he will look into it after the seminar. What should he have said?

The argument would be valid if $Y$ were exogenous, in which case one could characterise $\beta_2$ and $\alpha_2$ as being the effects of $Y$ on $W$ and $P$, holding other variables constant. But $Y$ is endogenous, and so the coefficients represent only part of an adjustment process. $Y$ cannot change autonomously, only in response to variations in $K$, $u$, or $v$.

The reduced form equations for $W$ and $P$ are:

$$W = \beta_1 + \frac{\beta_2}{1 - \alpha_2 - \beta_2}(\alpha_1 + \beta_1 + \alpha_3K + u + v) + u$$

$$= \frac{1}{1 - \alpha_2 - \beta_2}(\beta_1 + \alpha_1\beta_2 - \alpha_2\beta_1 + \alpha_3\beta_2K + (1 - \alpha_2)u + \beta_2v)$$

$$P = \alpha_1 + \frac{\alpha_2}{1 - \alpha_2 - \beta_2}(\alpha_1 + \beta_1 + \alpha_3K + u + v) + \alpha_3K + v$$

$$= \frac{1}{1 - \alpha_2 - \beta_2}(\alpha_1 - \alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3(1 - \beta_2)K + \alpha_2u + (1 - \beta_2)v).$$
Thus, for example, a change in \( K \) will lead to changes in \( W \) and \( P \) in the proportions \( \beta_2 : (1 - \beta_2) \), not \( \beta_2 : \alpha_2 \). The same is true of changes caused by a variation in \( v \). For a variation in \( u \), the proportions would be \( (1 - \alpha_2) : \alpha_2 \).

9.6 Answers to the additional exercises

A9.6 Explain why the OLS estimator of \( \beta_2 \) would be inconsistent, if the researcher’s model is correctly specified. Derive analytically the large sample bias, and state whether it is possible to determine its sign.

The reduced form equation for \( x \) is:

\[
 x = \frac{\beta_1 + p + u}{1 - \beta_2}.
\]

Thus:

\[
\hat{\beta}_{2 \text{OLS}} = \frac{\sum (x_i - \bar{x})(e_i - \bar{e})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(\beta_1 + \beta_2 x_i + u_i - \beta_1 - \beta_2 \bar{x} - \bar{u})}{\sum (x_i - \bar{x})^2} = \beta_2 + \frac{\sum (x_i - \bar{x})(u_i - \bar{u})}{\sum (x_i - \bar{x})^2}.
\]

It is not possible to obtain a closed-form expression for the expectation of the estimator because the error term is a nonlinear function of \( u \). Instead we investigate whether the estimator is consistent, first dividing the numerator and the denominator of the error term by \( n \) so that they tend to limits as the sample size becomes large.

\[
\text{plim} \hat{\beta}_{2 \text{OLS}} = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum \left( \frac{1}{1-\beta_2} \left[ \beta_1 + p_i + u_i - \beta_1 - \bar{p} - \bar{u} \right] \right) (u_i - \bar{u})}{\text{plim} \frac{1}{n} \sum (x_i - \bar{x})^2}
\]

\[
= \beta_2 + \frac{1}{1 - \beta_2} \text{plim} \frac{1}{n} \sum (p_i - \bar{p})(u_i - \bar{u}) + \text{plim} \frac{1}{n} \sum (u_i - \bar{u})^2
\]

\[
= \beta_2 + \frac{1}{1 - \beta_2} \frac{\text{cov}(p, u) + \text{var}(u)}{\text{var}(x)} = \beta_2 + \frac{1}{1 - \beta_2} \frac{\sigma_u^2}{\sigma_x^2}
\]

since \( \text{cov}(p, u) = 0 \), \( p \) being exogenous. It is reasonable to assume that employment grows less rapidly than output, and hence \( \beta_2 \), and so \( (1 - \beta_2) \), are less than 1. The bias is therefore likely to be positive.

Explain how the researcher might use \( p \) to construct an IV estimator of \( \beta_2 \) that is consistent if \( p \) is exogenous. Demonstrate analytically that the estimator is consistent.

\( p \) is available as an instrument, being exogenous, and therefore independent of \( u \), being correlated with \( x \), and not being in the equation in its own right.

\[
\hat{\beta}_{2 \text{IV}} = \frac{\sum (p_i - \bar{p})(e_i - \bar{e})}{\sum (p_i - \bar{p})(x_i - \bar{x})} = \frac{\sum (p_i - \bar{p})(\beta_1 + \beta_2 x_i + u_i - \beta_1 - \beta_2 \bar{x} - \bar{u})}{\sum (p_i - \bar{p})(x_i - \bar{x})} = \beta_2 + \frac{\sum (p_i - \bar{p})(u_i - \bar{u})}{\sum (p_i - \bar{p})(x_i - \bar{x})}.
\]
9. Simultaneous equations estimation

Hence, dividing the numerator and the denominator of the error term by $n$ so that they tend to limits as the sample size becomes large,

$$\text{plim} \hat{\beta}_{2,\text{IV}} = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum (p_i - \bar{p})(u_i - \bar{u})}{\text{plim} \frac{1}{n} \sum (p_i - \bar{p})(x_i - \bar{x})} = \beta_2 + \frac{\text{cov}(p, u)}{\text{cov}(p, x)} = \beta_2$$

since $\text{cov}(p, u) = 0$, $p$ being exogenous, and $\text{cov}(p, x) \neq 0$, $x$ being determined partly by $p$.  

The OLS and IV regressions are summarised below (standard errors in parentheses). Comment on them, making use of your answers to the first two parts of this question.

OLS\[ \hat{e} = -0.52 + 0.48x \]  \hspace{0.5cm} (3) \hspace{0.5cm} (0.27) \hspace{0.5cm} (0.08)  

IV\[ \hat{e} = 0.37 + 0.17x \]  \hspace{0.5cm} (4) \hspace{0.5cm} (0.42) \hspace{0.5cm} (0.14)

The IV estimate of the slope coefficient is lower than the OLS estimate, as expected. The standard errors are not comparable because the OLS ones are invalid.

A second researcher hypothesises that both $x$ and $p$ are exogenous and that equation (2) should be written:

$$e = x - p. \hspace{0.5cm} (5)$$

On the assumption that this is correct, explain why the slope coefficients in (3) and (4) are both biased and determine the direction of the bias in each case.

If (5) is correct, (3) is a misspecification that omits $p$ and includes a redundant intercept. From the identity, the true values of the coefficients of $x$ and $p$ are 1 and $-1$, respectively. For (3):

$$E(\hat{\beta}_{2,\text{OLS}}) = 1 - 1 \times \frac{\sum (x_i - \bar{x})(p_i - \bar{p})}{\sum (x_i - \bar{x})^2}.$$  

$x$ and $p$ are positively correlated, so the bias will be downwards.

For (4):

$$\hat{\beta}_{2,\text{IV}} = \frac{\sum (p_i - \bar{p})(e_i - \bar{e})}{\sum (p_i - \bar{p})(x_i - \bar{x})} = \frac{\sum (p_i - \bar{p})[(x_i - p_i) - (\bar{x} - \bar{p})]}{\sum (p_i - \bar{p})(x_i - \bar{x})}$$

$$= 1 - \frac{\sum (p_i - \bar{p})^2}{\sum (p_i - \bar{p})(x_i - \bar{x})} = 1 - \frac{1}{n} \frac{\sum (p_i - \bar{p})^2}{\sum (p_i - \bar{p})(x_i - \bar{x})}.$$

Hence:

$$\text{plim} \hat{\beta}_{2,\text{IV}} = 1 - \frac{\text{var}(p)}{\text{cov}(x, p)}$$

and so again the bias is downwards.

Explain what would be the result of fitting (5), regressing $e$ on $x$ and $p$.

One would obtain a perfect fit with the coefficient of $x$ equal to 1, the coefficient of $p$ equal to $-1$, and $R^2 = 1$. 

© Christopher Dougherty, 2016. All rights reserved.  
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
A9.7 Derive the reduced form equations for $F$ and $M$.

(2) is the reduced form equation for $M$. Substituting for $M$ in (1), we have:

$$F = (\beta_1 + \alpha_1 \beta_3) + (\beta_2 + \alpha_2 \beta_3)S + u + \beta_3 v.$$ 

Explain what would be the most appropriate method to fit equation (1).

Since $M$ does not depend on $u$, OLS may be used to fit (1).

Explain what would be the most appropriate method to fit equation (2).

There are no endogenous explanatory variables in (2), so again OLS may be used.

Explain why the coefficient of $S$ differs in the two equations.

In (3), the coefficient is an estimate of the direct effect of $S$ on fertility, controlling for $M$. In (4), the reduced form equation, it is an estimate of the total effect, taking account of the indirect effect via $M$ (female education reduces mortality, and a reduction in mortality leads to a reduction in fertility).

Explain whether one may validly perform $t$ tests on the coefficients of (4).

It is legitimate to use OLS to fit (4), so the $t$ tests are valid.

Discuss whether $G$ is likely to be a valid instrument.

$G$ should be a valid instrument since it is highly correlated with $S$, it may reasonably be considered to be exogenous and therefore uncorrelated with the disturbance term in (4), and it does not appear in the equation in its own right (though perhaps it should).

What should the researchers conclusions be with regard to the test?

With 1 degree of freedom as indicated by the output, the critical value of chi-squared at the 5 per cent significance level is 3.84. Therefore we do not reject the null hypothesis of no significant difference between the estimates of the coefficients and conclude that there is no need to instrument for $S$. (4) should be preferred because OLS is more efficient than IV, when both are consistent.

A9.8 Demonstrate that the estimate of $\alpha_2$ will be inconsistent if ordinary least squares (OLS) is used to fit the supply equation, showing that the large-sample bias is likely to be negative.

The reduced form equation for $P$ is:

$$P = \frac{1}{\alpha_2 - \beta_2} (\beta_1 - \alpha_1 + \beta_3 Y + \beta_4 POP + u_D - u_S).$$

The OLS estimator of $\alpha_2$ is:

$$\hat{\alpha}_2^{OLS} = \frac{\sum (P_i - \bar{P}) (Q_i - \bar{Q})}{\sum (P_i - \bar{P})^2} = \frac{\sum (P_i - \bar{P}) (\alpha_1 + \alpha_2 P_i + u_{Si} - \alpha_1 - \alpha_2 \bar{P} - \bar{u}_S)}{\sum (P_i - \bar{P})^2} = \frac{\alpha_2 + \sum (P_i - \bar{P}) (u_{Si} - \bar{u}_S)}{\sum (P_i - \bar{P})^2}.$$
9. Simultaneous equations estimation

We cannot take expectations because $u_s$ is a determinant of both the numerator and the denominator of the error term, in view of the reduced form equation for $P$. Instead, we take probability limits, after first dividing the numerator and the denominator of the error term by $n$ to ensure that limits exist.

$$\text{plim } \alpha_2^{\text{OLS}} = \alpha_2 + \frac{\text{plim } \frac{1}{n} \sum \left( P_i - \bar{P} \right) (u_{Si} - \bar{u}_s)}{\text{plim } \frac{1}{n} \sum \left( P_i - \bar{P} \right)^2} = \alpha_2 + \frac{\text{cov}(P, u_s)}{\text{var}(P)}.$$

Substituting from the reduced form equation for $P$:

$$\text{plim } \alpha_2^{\text{OLS}} = \alpha_2 + \frac{\text{cov} \left( \frac{1}{\alpha_2 - \beta_2} (\beta_1 - \alpha_1 + \beta_3 Y + \beta_4 \text{POP} + u_D - u_s), u_s \right)}{\text{var}(P)} = \alpha_2 - \frac{1}{\alpha_2 - \beta_2} \frac{\sigma^2_{u_s}}{\sigma^2_P}$$

assuming that $Y$ and $\text{POP}$ are exogenous and so $\text{cov}(u_s, Y) = \text{cov}(u_s, \text{POP}) = 0$. We are told that $u_s$ and $u_D$ are distributed independently, so $\text{cov}(u_s, u_D) = 0$.

Since it is reasonable to suppose that $\alpha_2$ is positive and $\beta_2$ is negative, the large-sample bias will be negative.

**Demonstrate that a consistent estimate of $\alpha_2$ will be obtained if the supply equation is fitted using instrumental variables (IV), using $Y$ as an instrument.**

$$\hat{\alpha}_2^{\text{IV}} = \frac{\sum \left( Y_i - \bar{Y} \right) (Q_i - \bar{Q})}{\sum \left( Y_i - \bar{Y} \right) \left( P_i - \bar{P} \right)} = \frac{\sum \left( Y_i - \bar{Y} \right) \left( \alpha_1 + \alpha_2 P_i + u_{Si} - \alpha_1 - \alpha_2 \bar{P} - \bar{u}_s \right)}{\sum \left( Y_i - \bar{Y} \right) \left( P_i - \bar{P} \right)}$$

$$= \alpha_2 + \frac{\sum \left( Y_i - \bar{Y} \right) (u_{Si} - \bar{u}_s)}{\sum \left( Y_i - \bar{Y} \right) \left( P_i - \bar{P} \right)}.$$

We cannot take expectations because $u_s$ is a determinant of both the numerator and the denominator of the error term, in view of the reduced form equation for $P$. Instead, we take probability limits, after first dividing the numerator and the denominator of the error term by $n$ to ensure that limits exist.

$$\text{plim } \hat{\alpha}_2^{\text{IV}} = \alpha_2 + \frac{\text{plim } \frac{1}{n} \sum \left( Y_i - \bar{Y} \right) (u_{Si} - \bar{u}_s)}{\text{plim } \frac{1}{n} \sum \left( Y_i - \bar{Y} \right) \left( P_i - \bar{P} \right)} = \alpha_2 + \frac{\text{cov}(Y, u)}{\text{cov}(Y, P)} = \alpha_2$$

since $\text{cov}(Y, u_s) = 0$ and $\text{cov}(P, Y) \neq 0$, $Y$ being a determinant of $P$.

**The model is used for a Monte Carlo experiment ... Discuss the results obtained.**

- The OLS estimates are clearly biased downwards.
- The IV and TSLS estimates appear to be distributed around the true value, although one would need a much larger number of samples to be sure of this.
- The IV estimates with POP appear to be slightly closer to the true value than those with $Y$, as should be expected given the higher correlation, and the TSLS estimates appear to be slightly closer than either, again as should be expected.
9.6. Answers to the additional exercises

• The OLS standard errors should be ignored. The standard errors for the IV regressions using \( \text{POP} \) tend to be smaller than those using \( Y \), reflecting the fact that \( \text{POP} \) is a better instrument. Those for the TSLS regressions are smallest of all, reflecting its greater efficiency.

A9.9 Derive the reduced form equation for \( G \) for the first researcher.

\[
G = \frac{1}{1 - \alpha_2 \beta_2} (\beta_1 + \alpha_1 \beta_2 + \beta_3 A + u_G + \beta_2 u_R).
\]

**Explain why ordinary least squares (OLS) would be an inconsistent estimator of the parameters of equation (2).**

The reduced form equation for \( G \) demonstrates that \( G \) is not distributed independently of the disturbance term \( u_R \), a requirement for the consistency of OLS when fitting (2).

**The first researcher uses instrumental variables (IV) to estimate \( \alpha_2 \) in (2). Explain the procedure and demonstrate that the IV estimator of \( \alpha_2 \) is consistent.**

The first researcher would use \( A \) as an instrument for \( G \). It is exogenous, so independent of \( u_R \); correlated with \( G \); and not in the equation in its own right. The estimator of the slope coefficient is:

\[
\hat{\alpha} \text{IV}_2 = \frac{\sum (A_i - \bar{A})(R_i - \bar{R})}{\sum (A_i - \bar{A})(G_i - \bar{G})} = \frac{\sum (A_i - \bar{A}) [\alpha_1 + \alpha_2 G_i + u_{Ri}] - [\alpha_1 + \alpha_2 \bar{G} = \bar{u}]}{\sum (A_i - \bar{A})(G_i - \bar{G})}.
\]

Hence:

\[
\operatorname{plim} \hat{\alpha}_2 \text{IV} = \alpha_2 + \operatorname{plim} \frac{\frac{1}{n} \sum (A_i - \bar{A})(u_{Ri} - \bar{u}_R)}{\frac{1}{n} \sum (A_i - \bar{A})(G_i - \bar{G})} = \alpha_2 + \frac{\operatorname{cov}(A, u_R)}{\operatorname{cov}(A, G)} = \alpha_2
\]

since \( \operatorname{cov}(A, u_R) = 0 \), \( A \) being exogenous, and \( \operatorname{cov}(A, G) \neq 0 \), \( A \) being a determinant of \( G \).

**The second researcher uses two stage least squares (TSLS) to estimate \( \alpha_2 \) in (2). Explain the procedure and demonstrate that the TSLS estimator is consistent.**

The reduced form equation for \( G \) for the second researcher is:

\[
G = \frac{1}{1 - \alpha_2 \beta_2} (\beta_1 + \alpha_1 \beta_2 + \beta_3 A + \beta_4 Q + u_G + \beta_2 u_R).
\]

It is fitted using TSLS. The fitted values of \( G \) are used as the instrument:

\[
\hat{\alpha} \text{TSLS}_2 = \frac{\sum (\hat{G}_i - \bar{G})(R_i - \bar{R})}{\sum (\hat{G}_i - \bar{G})(G_i - \bar{G})}.
\]
9. Simultaneous equations estimation

Following the same method as in the third part of the question:

\[
\text{plim } \hat{\alpha}_2^{\text{TLS}} = \alpha_2 + \frac{\text{cov}(\hat{G}, u_R)}{\text{cov}(\hat{G}, G)} = \alpha_2
\]

\(\text{cov}(\hat{G}, u_R)\) because \(\hat{G}\) is a linear combination of the exogenous variables, and \(\text{cov}(\hat{G}, G) \neq 0\).

**Explain why the TSLS estimator used by the second researcher ought to produce ‘better’ results than the IV estimator used by the first researcher, if the growth equation is given by (1*). Be specific about what you mean by ‘better’.**

The TSLS estimator of \(\alpha_2\) should have a smaller variance. The variance of an IV estimator is inversely proportional to the square of the correlation of \(G\) and the instrument. \(\hat{G}\) is the linear combination of \(A\) and \(Q\) that has the highest correlation. It will therefore, in general, have a lower variance than the IV estimator using \(A\).

**Suppose that the first researcher is correct and the growth equation is actually given by (1), not (1*). Compare the properties of the two estimators in this case.**

If the first researcher is correct, \(A\) is the optimal instrument because it will be more highly correlated with \(G\) (in the population) than the TSLS combination of \(A\) and \(Q\) and it will therefore be more efficient.

**Suppose that the second researcher is correct and the model is given by (1*) and (2), but \(A\) is not exogenous after all. Suppose that \(A\) is influenced by \(G\):**

\[
A = \gamma_1 + \gamma_2 G + u_A
\]

where \(u_A\) is a disturbance term distributed independently of \(u_G\) and \(u_R\). How would this affect the properties of the IV estimator of \(\alpha_2\) used by the first researcher?

\(\text{cov}(A, u_R)\) would not be equal to 0 and so the estimator would be inconsistent.

**A9.10 State, with a brief explanation, whether each variable is endogenous or exogenous, and derive the reduced form equations for the endogenous variables.**

In this model \(LGEARN\) and \(SKILL\) are endogenous. \(IQ\) is exogenous. The reduced form equation for \(LGEARN\) is:

\[
LGEARN = \beta_1 + \alpha_1 \beta_2 + \alpha_2 \beta_2 IQ + u + \beta_2 v.
\]

The reduced form equation for \(SKILL\) is the structural equation.

**Explain why the researcher could use ordinary least squares (OLS) to fit equation (1) if \(u\) and \(v\) are distributed independently of each other.**

\(SKILL\) is not determined either directly or indirectly by \(u\). Thus in equation (1) there is no violation of the requirement that the regressor be distributed independently of the disturbance term.
Show that the OLS estimator of $\beta_2$ is inconsistent if $u$ and $v$ are positively correlated and determine the direction of the large-sample bias.

Writing $L$ for $LGEARN$, $S$ for $SKILL$:

$$\hat{\beta}_{OLS}^2 = \frac{\sum (S_i - \bar{S})(L_i - \bar{L})}{\sum (S_i - \bar{S})^2} = \frac{\sum (S_i - \bar{S})[(\beta_1 + \beta_2 S_i + u_i) - (\beta_1 + \beta_2 S + \bar{u})]}{\sum (S_i - \bar{S})^2}$$

$$= \beta_2 + \frac{\sum (S_i - \bar{S})(u_i - \bar{u})}{\sum (S_i - \bar{S})^2}.$$ 

We cannot obtain a closed-form expression for the expectation of the error term since $S$ depends on $v$ and $v$ is correlated with $u$. Hence instead we take plims, dividing the numerator and the denominator by $n$ to ensure that the limits exist:

$$\text{plim } \hat{\beta}_{OLS}^2 = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum (S_i - \bar{S})(u_i - \bar{u})}{\text{plim} \frac{1}{n} \sum (S_i - \bar{S})^2} = \beta_2 + \frac{\text{cov}(S, u)}{\text{var}(S)}.$$ 

Now:

$$\text{cov}(S, u) = \text{cov}([\alpha_1 + \alpha_2 IQ + v], u) = \text{cov}(v, u)$$

since $\alpha_1$ is a constant and $IQ$ is exogenous. Hence the numerator of the error term is positive in large samples. The denominator, being a variance, is also positive. So the large-sample bias is positive.

**Demonstrate mathematically how the researcher could use instrumental variables (IV) estimation to obtain a consistent estimate of $\beta_2$.**

The researcher could use $IQ$ as an instrument for $SKILL$:

$$\hat{\beta}_{IV}^2 = \frac{\sum (I_i - \bar{I})(L_i - \bar{L})}{\sum (I_i - \bar{I})(S_i - \bar{S})} = \frac{\sum (I_i - \bar{I})[(\beta_1 + \beta_2 S_i + u_i) - (\beta_1 + \beta_2 S + \bar{u})]}{\sum (I_i - \bar{I})(S_i - \bar{S})}$$

$$= \beta_2 + \frac{\sum (I_i - \bar{I})(u_i - \bar{u})}{\sum (I_i - \bar{I})(S_i - \bar{S})}.$$ 

We cannot obtain a closed-form expression for the expectation of the error term since $S$ depends on $v$ and $v$ is correlated with $u$. Hence instead we take plims, dividing the numerator and the denominator by $n$ to ensure that the limits exist:

$$\text{plim } \hat{\beta}_{IV}^2 = \beta_2 + \frac{\text{plim} \frac{1}{n} \sum (I_i - \bar{I})(u_i - \bar{u})}{\text{plim} \frac{1}{n} \sum (I_i - \bar{I})(S_i - \bar{S})} = \beta_2 + \frac{\text{cov}(I, u)}{\text{cov}(I, s)}.$$ 

The numerator of the error term is zero because $I$ is exogenous. The denominator is not zero because $S$ is determined by $I$. Hence the IV estimator is consistent.
9. Simultaneous equations estimation

Explain the advantages and disadvantages of using IV, rather than OLS, to estimate \( \beta_2 \), given that the researcher is not sure whether \( u \) and \( v \) are distributed independently of each other.

The advantage of IV is that, being consistent, there will be no bias in large samples and hence one may hope that there is no serious bias in a finite sample. One disadvantage is that there is a loss of efficiency if \( u \) and \( v \) are independent. Even if they are not independent, the IV estimator may be inferior to the OLS estimator using some criterion such as the mean square error that allows a trade-off between the bias of an estimator and its variance.

Describe in general terms a test that might help the researcher decide whether to use OLS or IV. What are the limitations of the test?

Durbin–Wu–Hausman test. Also known as Hausman test. The test statistic is a chi-squared statistic based on the differences of all the coefficients in the regression. The null hypothesis is that \( \text{SKILL} \) is distributed independently of \( u \) and the differences in the coefficients are random. If the test statistic exceeds its critical value, given the significance level of the test, we reject the null hypothesis and conclude that we ought to use IV rather than OLS. The main limitation is lack of power if the instrument is weak.

Explain whether it is possible for the researcher to fit equation (2) and obtain consistent estimates.

There is no reason why the equation should not be fitted using OLS.

A9.11 Substituting for \( Y \) from the first equation into the second, and re-arranging, we have the reduced form equation for \( S \):

\[
S = \frac{\alpha_1 + \alpha_2 \beta_1 + v + \alpha_2 u}{1 - \alpha_2 \beta_2}.
\]

Substituting from the third equation into the first, we have:

\[
Y = \beta_1 + \beta_2 R + u - \beta_2 w.
\]

If this equation is fitted using OLS, we have:

\[
\text{plim } \hat{\beta}_2^{\text{OLS}} = \beta_2 + \frac{\text{cov}(R, [u - \beta_2 w])}{\text{var}(R)} = \beta_2 + \frac{\text{cov}([S + w], [u - \beta_2 w])}{\text{var}(S + w)}
\]

\[
= \beta_2 + \frac{\alpha_2 \gamma \sigma_u^2 - \beta_2 \sigma_w^2}{\sigma_S^2 + \sigma_w^2} = \beta_2 + \frac{\alpha_2 \gamma \sigma_u^2 - \beta_2 \sigma_w^2}{\gamma^2 (\sigma_S^2 + \alpha_2 \sigma_u^2) + \sigma_w^2}
\]

where:

\[
\gamma = \frac{1}{1 - \alpha_2 \beta_2}.
\]

The denominator of the bias term is positive. Hence the bias will be positive if (the component attributable to simultaneity) is greater than (the component attributable to measurement error), and negative if it is smaller.
Chapter 10
Binary choice and limited dependent variable models, and maximum likelihood estimation

10.1 Overview

The first part of this chapter describes the linear probability model, logit analysis, and probit analysis, three techniques for fitting regression models where the dependent variable is a qualitative characteristic. Next it discusses tobit analysis, a censored regression model fitted using a combination of linear regression analysis and probit analysis. This leads to sample selection models and heckman analysis. The second part of the chapter introduces maximum likelihood estimation, the method used to fit all of these models except the linear probability model.

10.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- describe the linear probability model and explain its defects
- describe logit analysis, giving the mathematical specification
- describe probit analysis, including the mathematical specification
- calculate marginal effects in logit and probit analysis
- explain why OLS yields biased estimates when applied to a sample with censored observations, even when the censored observations are deleted
- explain the problem of sample selection bias and describe how the heckman procedure may provide a solution to it (in general terms, without mathematical detail)
- explain the principle underlying maximum likelihood estimation
- apply maximum likelihood estimation from first principles in simple models.
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

10.3 Further material

Limiting distributions and the properties of maximum likelihood estimators

Provided that weak regularity conditions involving the differentiability of the likelihood function are satisfied, maximum likelihood (ML) estimators have the following attractive properties in large samples:

1. They are consistent.
2. They are asymptotically normally distributed.
3. They are asymptotically efficient.

The meaning of the first property is familiar. It implies that the probability density function of the estimator collapses to a spike at the true value. This being the case, what can the other assertions mean? If the distribution becomes degenerate as the sample size becomes very large, how can it be described as having a normal distribution? And how can it be described as being efficient, when its variance, and the variance of any other consistent estimator, tend to zero?

To discuss the last two properties, we consider what is known as the limiting distribution of an estimator. This is the distribution of the estimator when the divergence between it and its population mean is multiplied by $\sqrt{n}$. If we do this, the distribution of a typical estimator remains nondegenerate as $n$ becomes large, and this enables us to say meaningful things about its shape and to make comparisons with the distributions of other estimators (also multiplied by $\sqrt{n}$).

To put this mathematically, suppose that there is one parameter of interest, $\theta$, and that $\hat{\theta}$ is its ML estimator. Then (2) says that:

$$\sqrt{n} (\hat{\theta} - \theta) \sim N(0, \sigma^2)$$

for some variance $\sigma^2$. (3) says that, given any other consistent estimator $\tilde{\theta}$, $\sqrt{\tilde{\theta} - \theta}$ cannot have a smaller variance.

Test procedures for maximum likelihood estimation

This section on ML tests contains material that is a little advanced for an introductory econometrics course. It is provided because likelihood ratio tests are encountered in the sections on binary choice models and because a brief introduction may be of help to those who proceed to a more advanced course.

There are three main approaches to testing hypotheses in maximum likelihood estimation: likelihood ratio (LR) tests, Wald tests, and Lagrange multiplier (LM) tests. Since the theory behind Lagrange multiplier tests is relatively complex, the present discussion will be confined to the first two types. We will start by assuming that the probability density function of a random variable $X$ is a known function of a single unknown parameter $\theta$ and that the likelihood function for $\theta$ given a sample of $n$ observations on $X$, $L(\theta | X_1, \ldots, X_n)$, satisfies weak regularity conditions involving its
differentiability. In particular, we assume that $\theta$ is determined by the first-order condition $dL/d\theta = 0$. (This rules out estimators such as that in Exercise A10.7) The null hypothesis is $H_0 : \theta = \theta_0$, the alternative hypothesis is $H_1 : \theta \neq \theta_0$, and the maximum likelihood estimate of $\theta$ is $\hat{\theta}$.

### Likelihood ratio tests

A likelihood ratio test compares the value of the likelihood function at $\theta = \hat{\theta}$ with its value at $\theta = \theta_0$. In view of the definition of $\hat{\theta}$, $L(\hat{\theta}) \geq L(\theta_0)$ for all $\theta_0$. However, if the null hypothesis is true, the ratio $L(\hat{\theta})/L(\theta_0)$ should not be significantly greater than 1. As a consequence, the logarithm of the ratio:

$$
\log \left( \frac{L(\hat{\theta})}{L(\theta_0)} \right) = \log L(\hat{\theta}) - \log L(\theta_0)
$$

should not be significantly different from zero. In that it involves a comparison of the measures of goodness of fit for unrestricted and restricted versions of the model, the LR test is similar to an $F$ test.

Under the null hypothesis, it can be shown that in large samples the test statistic:

$$
LR = 2 \left( \log L(\hat{\theta}) - \log L(\theta_0) \right)
$$

has a chi-squared distribution with one degree of freedom. If there are multiple parameters of interest, and multiple restrictions, the number of degrees of freedom is equal to the number of restrictions.

### Examples

We will return to the example in Section 10.6 in the textbook, where we have a normally-distributed random variable $X$ with unknown population mean $\mu$ and known standard deviation equal to 1. Given a sample of $n$ observations, the likelihood function is:

$$
L(\hat{\mu} | X_1, \ldots, X_n) = \left( \frac{1}{\sqrt{2\pi}e^{(X_1-\mu)^2/2}} \right) \times \cdots \times \left( \frac{1}{\sqrt{2\pi}e^{(X_n-\mu)^2/2}} \right).
$$

The log-likelihood is:

$$
\log L(\hat{\mu} | X_1, \ldots, X_n) = n \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} \sum (X_i - \hat{\mu})^2
$$

and the unrestricted ML estimate is $\hat{\mu} = \bar{X}$. The LR statistic for the null hypothesis $H_0 : \mu = \mu_0$ is therefore:

$$
LR = 2 \left( n \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} \sum (X_i - \bar{X})^2 \right) - \left( n \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} (X_i - \mu_0)^2 \right)
$$

$$
= \sum (X_i - \mu_0)^2 - \sum (X_i - \bar{X})^2 = n(\bar{X} - \mu_0)^2.
$$

If we relaxed the assumption $\sigma = 1$, the unrestricted likelihood function is:

$$
L(\hat{\mu}, \hat{\sigma} | X_1, \ldots, X_n) = \left( \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}(X_i-\hat{\mu})^2/\hat{\sigma}^2} \right) \times \cdots \times \left( \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}(X_n-\hat{\mu})^2/\hat{\sigma}^2} \right)
$$
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

and the log-likelihood is:

\[
\log L(\hat{\mu}, \hat{\sigma} \mid X_1, \ldots, X_n) = n \log \left(\frac{1}{\sqrt{2\pi}}\right) - n \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} \sum (X_i - \hat{\mu})^2.
\]

The first-order condition obtained by differentiating by \(\sigma\) is:

\[
\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum (X_i - \mu)^2 = 0
\]

from which we obtain:

\[
\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \hat{\mu})^2.
\]

Substituting back into the log-likelihood function, the latter now becomes a function of \(\mu\) only (and is known as the concentrated log-likelihood function or, sometimes, the profile log-likelihood function):

\[
\log L(\mu \mid X_1, \ldots, X_n) = n \log \left(\frac{1}{\sqrt{2\pi}}\right) - n \log \left(\frac{1}{n} \sum (X_i - \mu)^2\right)^{1/2} - \frac{n}{2}.
\]

As before, the ML estimator of \(\mu\) is \(\bar{X}\). Hence the LR statistic is:

\[
LR = 2 \left( n \log \left(\frac{1}{\sqrt{2\pi}}\right) - n \log \left(\frac{1}{n} \sum (X_i - \bar{X})^2\right)^{1/2} - \frac{n}{2} \right) - \left( n \log \left(\frac{1}{\sqrt{2\pi}}\right) - n \log \left(\frac{1}{n} \sum (X_i - \mu_0)^2\right)^{1/2} - \frac{n}{2} \right) \\
= n \left( \log \sum (X_i - \mu_0)^2 - \log \sum (X_i - \bar{X})^2 \right).
\]

It is worth noting that this is closely related to the \(F\) statistic obtained when one fits the least squares model:

\[X_i = \mu + u_i.\]

The least squares estimator of \(\mu\) is \(\bar{X}\) and \(RSS = \sum (X_i - \bar{X})^2\).

If one imposes the restriction \(\mu = \mu_0\), we have \(RSS_R = \sum (X_i - \mu_0)^2\) and the \(F\) statistic:

\[F(1, n - 1) = \frac{\sum (X_i - \mu_0)^2 - \sum (X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2 / (n - 1)}.\]

Returning to the LR statistic, we have:

\[
LR = n \log \frac{\sum (X_i - \mu_0)^2}{\sum (X_i - \bar{X})^2} = n \log \left(1 + \frac{\sum (X_i - \mu_0)^2 - \sum (X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right) \\
\approx n \frac{\sum (X_i - \mu_0)^2 - \sum (X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2} = \frac{n}{n - 1} F \approx F.
\]

Note that we have used the approximation \(\log(1 + a) = a\) which is valid when \(a\) is small enough for higher powers to be neglected.
10.3. Further material

**Wald tests**

Wald tests are based on the same principle as t tests in that they evaluate whether the discrepancy between the maximum likelihood estimate \( \theta \) and the hypothetical value \( \theta_0 \) is significant, taking account of the variance in the estimate. The test statistic for the null hypothesis \( H_0 : \hat{\theta} - \theta_0 = 0 \) is:

\[
\frac{(\hat{\theta} - \theta_0)^2}{\hat{\sigma}_\theta^2}
\]

where \( \hat{\sigma}_\theta^2 \) is the estimate of the variance of \( \theta \) evaluated at the maximum likelihood value. \( \hat{\sigma}_\theta^2 \) can be estimated in various ways that are asymptotically equivalent if the likelihood function has been specified correctly. A common estimator is that obtained as minus the inverse of the second differential of the log-likelihood function evaluated at the maximum likelihood estimate. Under the null hypothesis that the restriction is valid, the test statistic has a chi-squared distribution with one degree of freedom. When there are multiple restrictions, the test statistic becomes more complex and the number of degrees of freedom is equal to the number of restrictions.

**Examples**

We will use the same examples as for the LR test, first, assuming that \( \sigma = 1 \) and then assuming that it has to be estimated along with \( \mu \). In the first case the log-likelihood function is:

\[
\log L(\mu | X_1, \ldots, X_n) = n \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} \sum (X_i - \mu)^2.
\]

The first differential is \( \sum (X_i - \mu) \) and the second is \( -n \), so the estimate of the variance is \( 1/n \). The Wald test statistic is therefore \( n (\hat{X} - \mu_0)^2 \).

In the second example, where \( \sigma \) was unknown, the concentrated log-likelihood function is:

\[
\log L(\mu | X_1, \ldots, X_n) = n \log \left( \frac{1}{\sqrt{2\pi}} \right) - n \log \left( \frac{1}{n} \sum (X_i - \mu)^2 \right)^{1/2} - \frac{n}{2}
\]

\[
= n \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{n}{2} \log \frac{1}{n} - \frac{n}{2} \log \left( \sum (X_i - \mu)^2 \right) - \frac{n}{2}.
\]

The first derivative with respect to \( \mu \) is:

\[
\frac{d \log L}{d \mu} = n \frac{\sum (X_i - \mu)}{\sum (X_i - \mu)^2}.
\]

The second derivative is:

\[
\frac{d^2 \log L}{d \mu^2} = n \left( \frac{-n (\sum (X_i - \mu)^2) - (\sum (X_i - \mu)) (-2 \sum (X_i - \mu))}{[\sum (X_i - \mu)^2]^2} \right).
\]

Evaluated at the ML estimator \( \hat{\mu} = \bar{X} \), \( \sum (X_i - \mu) = 0 \) and hence:

\[
\frac{d^2 \log L}{d \mu^2} = -\frac{n^2}{\sum (X_i - \mu)^2}
\]
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

giving an estimated variance \( \hat{\sigma}^2 / n \), given:

\[
\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2.
\]

Hence the Wald test statistic is:

\[
\left( \frac{\bar{X} - \mu_0}{\hat{\sigma}^2 / n} \right)^2.
\]

Under the null hypothesis, this is distributed as a chi-squared statistic with one degree of freedom.

When there is just one restriction, as in the present case, the Wald statistic is the square of the corresponding asymptotic t statistic (asymptotic because the variance has been estimated asymptotically). The chi-squared test and the t test are equivalent, given that, when there is one degree of freedom, the critical value of the chi-squared statistic for any significance level is the square of the critical value of the normal distribution.

**LR test of restrictions in a regression model**

Given the regression model:

\[
Y_i = \beta_1 + \sum_{j=2}^{k} \beta_j X_{ij} + u_i
\]

with \( u \) assumed to be iid \( N(0, \sigma^2) \), the log-likelihood function for the parameters is:

\[
\log L(\beta_1, \ldots, \beta_k, \sigma \mid Y_i, X_i, i = 1, \ldots, n) = n \log \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{1}{2\sigma^2} \sum \left( Y_i - \beta_1 - \sum_{j=2}^{k} \beta_j X_{ij} \right)^2.
\]

This is a straightforward generalisation of the expression for a simple regression derived in Section 10.6 in the textbook. Hence

\[
\log L(\beta_1, \ldots, \beta_k, \sigma \mid Y_i, X_i, i = 1, \ldots, n) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} Z
\]

where:

\[
Z = \sum \left( Y_i - \beta_1 - \sum_{j=2}^{k} \beta_j X_{ij} \right)^2.
\]

The estimates of the \( \beta \) parameters affect only \( Z \). To maximise the log-likelihood, they should be chosen so as to minimise \( Z \), and of course this is exactly what one is doing when one is fitting a least squares regression. Hence \( Z = RSS \). It remains to determine the ML estimate of \( \sigma \). Taking the partial differential with respect to \( \sigma \), we obtain one of the first-order conditions for a maximum:

\[
\frac{\partial \log L(\beta_1, \ldots, \beta_k, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} RSS = 0.
\]

From this we obtain:

\[
\hat{\sigma}^2 = \frac{RSS}{n}.
\]
Hence the ML estimator is the sum of the squares of the residuals divided by \( n \). This is different from the least squares estimator, which is the sum of the squares of the residuals divided by \( n - k \), but the difference disappears as the sample size becomes large. Substituting for \( \hat{\sigma}^2 \) in the log-likelihood function, we obtain the concentrated likelihood function:

\[
\log L(\beta_1, \ldots, \beta_k | Y_i, X_i, i = 1, \ldots, n) = -n \log \left( \frac{RSS}{n} \right)^{1/2} - \frac{n}{2} \log 2\pi - \frac{1}{2} \frac{RSS}{n}
\]

\[
= -\frac{n}{2} \log RSS - \frac{n}{2} \log 2\pi - \frac{n}{2}
\]

\[
= -\frac{n}{2} (\log RSS + 1 + \log 2\pi - \log n).
\]

We will re-write this as:

\[
\log L_U = -\frac{n}{2} (\log RSS_U + 1 + \log 2\pi - \log n)
\]

the subscript U emphasising that this is the unrestricted log-likelihood. If we now impose a restriction on the parameters and maximise the loglikelihood function subject to the restriction, it will be:

\[
\log L_R = -\frac{n}{2} (\log RSS_R + 1 + \log 2\pi - \log n)
\]

where \( RSS_R \geq RSS_U \) and hence \( \log L_R \leq \log L_U \). The LR statistic for a test of the restriction is therefore:

\[
2(\log L_U - L_R) = n(\log RSS_R - \log RSS_U) = n \log \frac{RSS_R}{RSS_U}.
\]

It is distributed as a chi-squared statistic with one degree of freedom under the null hypothesis that the restriction is valid. If there is more than one restriction, the test statistic is the same but the number of degrees of freedom under the null hypothesis that all the restrictions are valid is equal to the number of restrictions.

An example of its use is the common factor test in Section 12.3 in the text. As with all maximum likelihood tests, it is valid only for large samples. Thus for testing linear restrictions we should prefer the \( F \) test approach because it is valid for finite samples.

### 10.4 Additional exercises

A10.1 **What factors affect the decision to make a purchase of your category of expenditure in the CES data set?**

Define a new variable \( CATBUY \) that is equal to 1 if the household makes any purchase of your category and 0 if it makes no purchase at all. Regress \( CATBUY \) on \( EXPPC, SIZE, REFAGE, \) and \( COLLEGE \) (as defined in Exercise A5.6) using: (1) the linear probability model, (2) the logit model, and (3) the probit model. Calculate the marginal effects at the mean of \( EXPPC, SIZE, REFAGE, \) and \( COLLEGE \) for the logit and probit models and compare them with the coefficients of the linear probability model.
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

A10.2 Logit analysis was used to relate the event of a respondent working (WORKING, defined to be 1 if the respondent was working, and 0 otherwise) to the respondent’s educational attainment (S, defined as the highest grade completed) using 1994 data from the National Longitudinal Survey of Youth 1979–. In this year the respondents were aged 29–36 and a substantial number of females had given up work to raise a family. The analysis was undertaken for females and males separately, with the output shown below (first females, then males, with iteration messages deleted):

```
.logit WORKING S if MALE==0
Logit Estimates
Number of obs = 2726
chi2(1) = 70.42
Prob > chi2 = 0.0000
Log Likelihood = -1586.5519 Pseudo R2 = 0.0217

| WORKING | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------+-------+-----------+-------+-------+---------------------------------|
| S       | 0.151  | 0.019     | 8.12  | 0.000 | 0.115                           |
| _cons   | -1.05  | 0.24      | -4.29 | 0.000 | -1.54                           |
```

```
.logit WORKING S if MALE==1
Logit Estimates
Number of obs = 2573
chi2(1) = 75.03
Prob > chi2 = 0.0000
Log Likelihood = -802.6542 Pseudo R2 = 0.0446

| WORKING | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------+-------+-----------+-------+-------+---------------------------------|
| S       | 0.25   | 0.03      | 8.16  | 0.000 | 0.19                           |
| _cons   | -0.97  | 0.38      | -2.57 | 0.010 | -1.71                           |
```

95 per cent of the respondents had S in the range 9–18 years and the mean value of S was 13.3 and 13.2 years for females and males, respectively.

From the logit analysis, the marginal effect of S on the probability of working at the mean was estimated to be 0.030 and 0.020 for females and males, respectively. Ordinary least squares regressions of WORKING on S yielded slope coefficients of 0.029 and 0.020 for females and males, respectively.

As can be seen from the second figure below, the marginal effect of educational attainment was lower for males than for females over most of the range S ≥ 9. Discuss the plausibility of this finding.

As can also be seen from the second figure, the marginal effect of educational attainment decreases with educational attainment for both males and females over the range S ≥ 9. Discuss the plausibility of this finding.

Compare the estimates of the marginal effect of educational attainment using logit analysis with those obtained using ordinary least squares.
10.4. Additional exercises

A10.3 A researcher has data on weight, height, and schooling for 540 respondents in the National Longitudinal Survey of Youth 1979– for the year 2002. Using the data on weight and height, he computes the body mass index for each individual. If the body mass index is 30 or greater, the individual is defined to be obese. He defines a binary variable, \( OBESE \), that is equal to 1 for the 164 obese individuals and 0 for the other 376. He wishes to investigate whether obesity is related to schooling and fits an ordinary least squares (OLS) regression of \( OBESE \) on \( S \), years of schooling, with the following result (\( t \) statistics in parentheses):

\[
\hat{OBESE} = 0.595 - 0.021 S \\
(5.30) 
\]

This is described as the linear probability model (LPM). He also fits the logit
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

model:

\[ F(Z) = \frac{1}{1 + e^{-Z}} \]

where \( F(Z) \) is the probability of being obese and \( Z = \beta_1 + \beta_2 S \), with the following result (again, \( t \) statistics in parentheses):

\[ \hat{Z} = 0.588 - 0.105S \quad (1.07) \quad (2.60) \]

The figure below shows the probability of being obese and the marginal effect of schooling as a function of \( S \), given the logit regression. Most (492 out of 540) of the individuals in the sample had 12 to 18 years of schooling.

![Scatter diagram of probability of being obese against years of schooling.](image)

**Figure 10.3:** Scatter diagram of probability of being obese against years of schooling.

- Discuss whether the relationships indicated by the probability and marginal effect curves appear to be plausible.
- Add the probability function and the marginal effect function for the LPM to the diagram. Explain why you drew them the way you did.
- The logit model is considered to have several advantages over the LPM. Explain what these advantages are. Evaluate the importance of the advantages of the logit model in this particular case.
- The LPM is fitted using OLS. Explain how, instead, it might be fitted using maximum likelihood estimation:
  - Write down the probability of being obese for any obese individual, given \( S_i \) for that individual, and write down the probability of not being obese for any non-obese individual, again given \( S_i \) for that individual.
  - Write down the likelihood function for this sample of 164 obese individuals and 376 non-obese individuals.
  - Explain how one would use this function to estimate the parameters. [Note: You are not expected to attempt to derive the estimators of the parameters.]
10.4. Additional exercises

- Explain whether your maximum likelihood estimators will be the same or different from those obtained using least squares.

A10.4 A researcher interested in the relationship between parenting, age and schooling has data for the year 2000 for a sample of 1,167 married males and 870 married females aged 35 to 42 in the National Longitudinal Survey of Youth 1979-. In particular, she is interested in how the presence of young children in the household is related to the age and education of the respondent. She defines \( \text{CHILDL6} \) to be 1 if there is a child less than 6 years old in the household and 0 otherwise and regresses it on \( \text{AGE} \), age, and \( S \), years of schooling, for males and females separately using probit analysis. Defining the probability of having a child less than 6 in the household to be \( p = F(Z) \) where:

\[
Z = \beta_1 + \beta_2 \text{AGE} + \beta_3 S
\]

she obtains the results shown in the table below (asymptotic standard errors in parentheses).

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{AGE} )</td>
<td>-0.137</td>
<td>-0.154</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>( S )</td>
<td>0.132</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>constant</td>
<td>0.194</td>
<td>0.547</td>
</tr>
<tr>
<td></td>
<td>(0.358)</td>
<td>(0.492)</td>
</tr>
<tr>
<td>( \bar{Z} )</td>
<td>-0.399</td>
<td>-0.874</td>
</tr>
<tr>
<td>( f(\bar{Z}) )</td>
<td>0.368</td>
<td>0.272</td>
</tr>
</tbody>
</table>

For males and females separately, she calculates:

\[
\bar{Z} = \bar{\beta}_1 + \bar{\beta}_2 \bar{\text{AGE}} + \bar{\beta}_3 \bar{S}
\]

where \( \bar{\text{AGE}} \) and \( \bar{S} \) are the mean values of \( \text{AGE} \) and \( S \) and \( \bar{\beta}_1, \bar{\beta}_2, \) and \( \bar{\beta}_3 \) are the probit coefficients in the corresponding regression, and she further calculates:

\[
f(\bar{Z}) = \frac{1}{\sqrt{2\pi}} e^{-\bar{Z}^2/2}
\]

where \( f(Z) = dF/dZ \). The values of \( \bar{Z} \) and \( f(\bar{Z}) \) are shown in the table.

- Explain how one may derive the marginal effects of the explanatory variables on the probability of having a child less than 6 in the household, and calculate for both males and females the marginal effects at the means of \( \text{AGE} \) and \( S \).
- Explain whether the signs of the marginal effects are plausible. Explain whether you would expect the marginal effect of schooling to be higher for males or for females.
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

- At a seminar someone asks the researcher whether the marginal effect of \( S \) is significantly different for males and females. The researcher does not know how to test whether the difference is significant and asks you for advice. What would you say?

A10.5 A health economist investigating the relationship between smoking, schooling, and age, defines a dummy variable \( D \) to be equal to 1 for smokers and 0 for nonsmokers. She hypothesises that the effects of schooling and age are not independent of each other and defines an interactive term schooling*age. She includes this as an explanatory variable in the probit regression. Explain how this would affect the estimation of the marginal effects of schooling and age.

A10.6 A researcher has data on the following variables for 5,061 respondents in the National Longitudinal Survey of Youth 1979–:

- \( MARRIED \), marital status in 1994, defined to be 1 if the respondent was married with spouse present and 0 otherwise;
- \( MALE \), defined to be 1 if the respondent was male and 0 if female;
- \( AGE \) in 1994 (the range being 29–37);
- \( S \), years of schooling, defined as highest grade completed, and
- \( ASVABC \), score on a test of cognitive ability, scaled so as to have mean 50 and standard deviation 10.

She uses probit analysis to regress \( MARRIED \) on the other variables, with the output shown:

```
. probit MARRIED MALE AGE S ASVABC
Probit estimates
Number of obs = 5061
LR chi2(4) = 229.78
Prob > chi2 = 0.0000
Log likelihood = -3286.1289 Pseudo R2 = 0.0338
------------------------------------------------------------------------------
 MARRIED | Coef. Std. Err. z P>|z| [95% Conf. Interval]
-------------+----------------------------------------------------------------
 MALE | -.1215281 .036332 -3.34 0.001 -.1927375 -.0503188
 AGE | .028571 .0081632 3.50 0.000 .0125715 .0445705
 S | -.0017465 .00919 -0.19 0.849 -.0197587 .0162656
 ASVABC | .0252911 .0022895 11.05 0.000 .0208038 .0297784
 _cons | -1.816455 .2798724 -6.49 0.000 -2.364995 -1.267916
------------------------------------------------------------------------------
```

The means of the explanatory variables, and their marginal effects evaluated at the means, are shown in the table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Marginal effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>0.4841</td>
<td>-0.0467</td>
</tr>
<tr>
<td>AGE</td>
<td>32.52</td>
<td>0.0110</td>
</tr>
<tr>
<td>S</td>
<td>13.31</td>
<td>-0.0007</td>
</tr>
<tr>
<td>ASVABC</td>
<td>48.94</td>
<td>0.0097</td>
</tr>
</tbody>
</table>
10.5. Answers to the starred exercises in the textbook

- Discuss the conclusions one may reach, given the probit output and the table, commenting on their plausibility.
- The researcher considers including $CHILD$, a dummy variable defined to be 1 if the respondent had children, and 0 otherwise, as an explanatory variable. When she does this, its $z$-statistic is 33.65 and its marginal effect 0.5685. Discuss these findings.

10.7 Suppose that the time, $t$, required to complete a certain process has probability density function:

$$f(t) = \alpha e^{-\alpha(t-\beta)} \text{ with } t > \beta > 0$$

and you have a sample of $n$ observations with times $T_1, \ldots, T_n$. Determine the maximum likelihood estimate of $\alpha$, assuming that $\beta$ is known.

A10.8 In Exercise 10.14 in the text, an event could occur with probability $p$. Given that the event occurred $m$ times in a sample of $n$ observations, the exercise required demonstrating that $m/n$ was the ML estimator of $p$. Derive the LR statistic for the null hypothesis $p = p_0$. If $m = 40$ and $n = 100$, test the null hypothesis $p = 0.5$.

A10.9 For the variable in Exercise A10.8, derive the Wald statistic and test the null hypothesis $p = 0.5$.

### 10.5 Answers to the starred exercises in the textbook

10.1 [This exercise does not have a star in the text, but an answer to it is needed for comparison with the answer to Exercise 10.3.]

The output shows the result of an investigation of how the probability of a respondent obtaining a bachelor’s degree from a four-year college is related to the score on ASVABC, using EAWE Data Set 21. $BACH$ is a dummy variable equal to 1 for those with bachelor’s degrees (years of schooling at least 16) and 0 otherwise. ASVABC is a measure of cognitive ability, scaled so that in the population it has mean 0 and standard deviation 1. Provide an interpretation of the coefficients. Explain why OLS is not a satisfactory estimation method for this kind of model.

```
. reg BACH ASVABC
  Source | SS df MS Number of obs = 500
----------+------------------------------ F( 1, 498) = 123.14
Model | 24.7674233 1 24.7674233 Prob > F = 0.0000
Residual | 100.160577 498 .201125656 R-squared = 0.1983
----------+------------------------------ Adj R-squared = 0.1966
Total | 124.928 499 .250356713 Root MSE = .44847

BACH | Coef. Std. Err. t P>|t| [95% Conf. Interval]
----------+------------------------------------
ASVABC | .2479312 .0223421 11.10 0.000 .2040348 .2918277
_cons | .4206845 .0209535 20.08 0.000 .3795163 .4618526
```
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

Answer:

The slope coefficient indicates that the probability of earning a bachelor’s degree rises by 25 per cent for every additional unit of the ASVABC score. ASVABC is scaled so that one unit is one standard deviation and it has mean zero. While this may be realistic for a range of values of ASVABC, it is not for very low ones. Very few of those with scores in the low end of the spectrum earned bachelors degrees and variations in the ASVABC score would be unlikely to have an effect on the probability. The intercept literally indicates that an individual with average score would have a 42 per cent probability of earning a bachelor’s degree.

However, the linear probability model predicts nonsense negative probabilities for all those with scores less of −1.70 or less. It also suffers from the problem that the standard errors and t and F tests are invalid because the disturbance term does not have a normal distribution. Its distribution is not even continuous, consisting of only two possible values for each value of ASVABC.

10.3 The output shows the results of fitting a logit regression for BACH, as defined in Exercise 10.1, with the iteration messages deleted. 48.8 per cent of the respondents earned bachelor’s degrees.

```
.logit BACH ASVABC
```

| Coef.     | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|-----------|-----------|------|-------|---------------------|
| ASVABC    | 1.240198  | .1377998 | 9.00  | 0.000 | .9701151 1.51028 |
| _cons     | -.4077999 | .1088093 | -3.75 | 0.000 | -.6210623 -.1945375 |

The diagram shows the probability of earning a bachelor’s degree as a function of ASVABC. It also shows the marginal effect function.

- With reference to the diagram, discuss the variation of the marginal effect of the ASVABC score implicit in the logit regression.
- Sketch the probability and marginal effect diagrams for the OLS regression in Exercise 10.1 and compare them with those for the logit regression.

Answer:

ASVABC is scaled so that it has a mean of zero. From the curve for the cumulative probability in the figure it can be seen that, for a respondent with mean score, the probability of graduating from college is about 40 per cent. For those one standard deviation above the mean, it is nearly 70 per cent. For those one standard deviation below, a little lower than 20 percent. Looking at the curve for the marginal probability, it can be seen that the marginal effect is greatest for those of average cognitive ability, and still quite high a standard deviation either way. For those two standard deviations above the mean, the marginal effect is low because most are going to college anyway. For those two standard deviations below, the effect is gain low, for the opposite reason.
10.5. Answers to the starred exercises in the textbook

Figure 10.4: Scatter diagram of cumulative and marginal effects against ASVABC.

For the linear probability model in Exercise 10.1, the counterpart to the cumulative probability curve in the figure is a straight line using the regression result. In this example, the predictions of the linear probability model do not differ much from those of the logit model over the central range of the data. Its deficiencies become visible only at the extremes. The OLS counterpart to the marginal probability curve is a horizontal straight line at 0.25, showing that the marginal effect is somewhat underestimated in the central range and overestimated elsewhere.

Figure 10.5: Scatter diagram of cumulative and marginal effects against ASVABC.

10.7 The following probit regression, with iteration messages deleted, was fitted using 2,108 observations on females in the National Longitudinal Survey of Youth using the LFP2011 data set described in Exercise 10.2. The respondents were aged 27 to 31 and many of them were raising young families.
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

```
.probit WORKING S AGE CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP if MALE==0
```

Probit regression
---
Number of obs = 2108
LR chi2(7) = 170.55
Prob > chi2 = 0.0000
Log likelihood = -972.89229 Pseudo R2 = 0.0806
---

```
|    | Coef. | Std. Err. | z    | P>|z|    | [95% Conf. Interval] |
|----|-------|-----------|------|--------|----------------------------|
| S  | .1046085 | .0127118  | 8.23 | 0.000  | .0796939 - .1295232       |
| AGE| -.0029273 | .0237761  | -0.12| 0.902  | -.0495277 - .043673       |
| CHILDL06| -.4490263 | .08128 | -5.52 | 0.000  | -.6083322 - -.2897204    |
| CHILDL16| -.3055774 | .1060307 | -2.88 | 0.004  | -.5133938 - -.097761    |
| MARRIED| -.1286145 | .0724189 | -1.78 | 0.076  | -.2705529 - .0133239    |
| ETHBLACK| -.1070784 | .0861386 | -1.24 | 0.214  | -.2759069 - .0617502   |
| ETHHISP| .0364241  | .0987625  | 0.37 | 0.712  | -.1571468 - .229995     |
| _cons| -.1885982 | .7046397 | -0.27 | 0.789  | -1.569667 - 1.19247     |
```

**WORKING** is a binary variable equal to 1 if the respondent was working in 2011, 0 otherwise. **CHILDL06** is a dummy variable equal to 1 if there was a child aged less than 6 in the household, 0 otherwise. **CHILDL16** is a dummy variable equal to 1 if there was a child aged less than 16, but no child less than 6, in the household, 0 otherwise. **MARRIED** is equal to 1 if the respondent was married with spouse present, 0 otherwise. The remaining variables are as described in Appendix B. The mean values of the variables are given in the output from the sum command:

```
.sum WORKING S AGE CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP if MALE==0
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORKING</td>
<td>2108</td>
<td>.7988615</td>
<td>.4009465</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>2108</td>
<td>14.32922</td>
<td>2.882736</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>AGE</td>
<td>2108</td>
<td>28.99336</td>
<td>1.386405</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td>CHILDL06</td>
<td>2108</td>
<td>.4407021</td>
<td>.4965891</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CHILDL16</td>
<td>2108</td>
<td>.1465844</td>
<td>.3537751</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MARRIED</td>
<td>2108</td>
<td>.420778</td>
<td>.4938011</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ETHBLACK</td>
<td>2108</td>
<td>.1783681</td>
<td>.3829132</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ETHHISP</td>
<td>2108</td>
<td>.1233397</td>
<td>.3289047</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Calculate the marginal effects and discuss whether they are plausible.

**Answer:**

The marginal effects are calculated in the table below. As might be expected, having a child aged less than 6 has a large adverse effect, very highly significant. Schooling also has a very significant effect, more educated mothers making use of their investment by tending to stay in the labour force. Age has a significant negative effect, the reason for which is not obvious (the respondents were aged 29–36 in 1994). Being black also has an adverse effect, the reason for which is likewise not obvious. (The **WORKING** variable is defined to be 1 if the individual has recorded hourly earnings of at least $3. If the definition is tightened to including also the requirement that the employment status is employed, the latter effect is smaller, but still significant at the 5 per cent level.)
10.5. Answers to the starred exercises in the textbook

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>$\beta_2$</th>
<th>Mean $\times \beta_2$</th>
<th>$f(Z)$</th>
<th>$\beta_2 \times f(Z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>14.3292</td>
<td>0.1046</td>
<td>1.4990</td>
<td>0.2627</td>
<td>0.0275</td>
</tr>
<tr>
<td>AGE</td>
<td>28.9934</td>
<td>-0.0029</td>
<td>-0.00849</td>
<td>0.2627</td>
<td>-0.0008</td>
</tr>
<tr>
<td>CHILD06</td>
<td>0.4407</td>
<td>-0.4490</td>
<td>-0.1979</td>
<td>0.2627</td>
<td>-0.1180</td>
</tr>
<tr>
<td>CHILDL16</td>
<td>0.1466</td>
<td>-0.3056</td>
<td>-0.0448</td>
<td>0.2627</td>
<td>-0.0803</td>
</tr>
<tr>
<td>MARRIED</td>
<td>0.4208</td>
<td>-0.1286</td>
<td>-0.0541</td>
<td>0.2627</td>
<td>-0.0338</td>
</tr>
<tr>
<td>ETHBLACK</td>
<td>0.1784</td>
<td>-0.1071</td>
<td>-0.0191</td>
<td>0.2627</td>
<td>-0.0281</td>
</tr>
<tr>
<td>ETHHISP</td>
<td>0.1233</td>
<td>0.0045</td>
<td>0.2627</td>
<td>0.0096</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>1.0000</td>
<td>-0.1886</td>
<td>-0.1886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>0.9141</td>
<td></td>
</tr>
</tbody>
</table>

10.12 Show that the tobit model may be regarded as a special case of a selection bias model.

**Answer:**

The selection bias model may be written:

$$B_i^* = \delta_1 + \sum_{j=2}^{m} \delta_j Q_{ji} + \varepsilon_i$$

$$Y_i^* = \beta_1 \sum_{j=2}^{k} \beta_j X_{ji} + u_i$$

$$Y_i = Y_i^* \quad \text{for} \quad B_i^* > 0$$

$$Y_i \text{ is not observed for} \quad B_i^* \leq 0$$

where the $Q$ variables determine selection. The tobit model is the special case where the $Q$ variables are identical to the $X$ variables and $B^*$ is the same as $Y^*$.

10.14 An event is hypothesised to occur with probability $p$. In a sample of $n$ observations, it occurred $m$ times. Demonstrate that the maximum likelihood estimator of $p$ is $m/n$.

**Answer:**

In each observation where the event did occur, the probability was $p$. In each observation where it did not occur, the probability was $(1 - p)$. Since there were $m$ of the former and $n - m$ of the latter, the joint probability was $p^m(1 - p)^{n-m}$.

Reinterpreting this as a function of $p$, given $m$ and $n$, the log-likelihood function for $p$ is:

$$\log L(p) = m \log p + (n - m) \log(1 - p).$$

Differentiating with respect to $p$, we obtain the first-order condition for a minimum:

$$\frac{d \log L(p)}{dp} = \frac{m}{p} - \frac{n - m}{1 - p} = 0.$$  

This yields $p = m/n$. We should check that the second differential is negative and that we have therefore found a maximum. The second differential is:

$$\frac{d^2 \log L(p)}{dp^2} = -\frac{m}{p^2} - \frac{n - m}{(1 - p)^2}.$$
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

Evaluated at $p = m/n$:
\[ \frac{d^2 \log L(p)}{dp^2} = -\frac{n^2}{m} - \frac{n - m}{(1 - \frac{m}{n})^2} = -n^2 \left( \frac{1}{m} + \frac{1}{n - m} \right). \]

This is negative, so we have indeed chosen the value of $p$ that maximises the probability of the outcome.

10.18 Returning to the example of the random variable $X$ with unknown mean $\mu$ and variance $\sigma^2$, the log-likelihood for a sample of $n$ observations was given by equation (10.36):
\[ \log L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + \frac{1}{\sigma^2} \left( -\frac{1}{2} (X_1 - \mu)^2 - \cdots - \frac{1}{2} (X_n - \mu)^2 \right). \]

The first-order condition for $\mu$ produced the ML estimator of $\mu$ and the first order condition for $\sigma$ then yielded the ML estimator for $\sigma$. Often, the variance is treated as the primary dispersion parameter, rather than the standard deviation. Show that such a treatment yields the same results in the present case. Treat $\sigma^2$ as a parameter, differentiate log $L$ with respect to it, and solve.

**Answer:**
\[ \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{\sigma^4} \left( -\frac{1}{2} (X_1 - \mu)^2 - \cdots - \frac{1}{2} (X_n - \mu)^2 \right). \]

Hence:
\[ \hat{\sigma}^2 = \frac{1}{n} \left( (X_1 - \mu)^2 + \cdots + (X_n - \mu)^2 \right) \]
as before. The ML estimator of $\mu$ is $\bar{X}$ as before.

10.19 In Exercise 10.7, log $L_0$ is $-1058.17$. Compute the pseudo-$R^2$ and confirm that it is equal to that reported in the output.

**Answer:**
As defined in equation (10.48):
\[ \text{pseudo-}R^2 = 1 - \frac{\log L}{\log L_0} = 1 - \frac{-972.8923}{-1058.17} = 0.0806 \]
as appears in the output.

10.20 In Exercise 10.7, compute the likelihood ratio statistic $2(\log L - \log L_0)$, confirm that it is equal to that reported in the output, and perform the likelihood ratio test.

**Answer:**
The likelihood ratio statistic is $2(-972.89 + 1058.17) = 170.56$, which is that reported in the output, apart from rounding error in the last digit. Under the null hypothesis that the coefficients of the explanatory variables are all jointly equal to 0, this is distributed as a chi-squared statistic with degrees of freedom equal to the number of explanatory variables, in this case 7. The critical value of chi-squared at the 0.1 per cent significance level with 7 degrees of freedom is 24.32, and so we reject the null hypothesis at that level.
### 10.6 Answers to the additional exercises

**A10.1** In the case of *FDHO* there were no non-purchasing households and so it was not possible to undertake the analysis.

The results for the logit analysis and the probit analysis were very similar. The linear probability model also yielded similar results for most of the commodities, the coefficients being similar to the logit and probit marginal effects and the *t* statistics being of the same order of magnitude as the *z* statistics for the logit and probit.

Most of the effects seem plausible with simple explanations. The total expenditure of the household and the size of the household were both highly significant factors in the decision to make a purchase for most categories of expenditure. The main exception, *TOB*, was instead influenced (negatively: survival bias?) by the age of the reference individual and, unsurprisingly, by his or her education.

---

**Linear probability model, dependent variable *CATBUY***

<table>
<thead>
<tr>
<th></th>
<th>EXPPC×10^{−4}</th>
<th>SIZE×10^{−2}</th>
<th>REFAGE×10^{−2}</th>
<th>COLLEGE</th>
<th>Cases with probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2.815</td>
<td>0.38</td>
<td>20.41</td>
<td>4.00</td>
<td>9.54</td>
</tr>
<tr>
<td>CLOT</td>
<td>4.500</td>
<td>0.33</td>
<td>18.74</td>
<td>5.38</td>
<td>13.61</td>
</tr>
<tr>
<td>DOM</td>
<td>1.661</td>
<td>0.30</td>
<td>17.37</td>
<td>4.18</td>
<td>10.78</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>0.13</td>
<td>11.83</td>
<td>3.13</td>
<td>12.38</td>
</tr>
<tr>
<td>ELEC</td>
<td>5.828</td>
<td>0.08</td>
<td>7.33</td>
<td>2.71</td>
<td>11.09</td>
</tr>
<tr>
<td>FDAW</td>
<td>5.102</td>
<td>0.23</td>
<td>14.57</td>
<td>2.23</td>
<td>6.41</td>
</tr>
<tr>
<td>FDHO*</td>
<td>6.334</td>
<td>0.28</td>
<td>15.83</td>
<td>5.93</td>
<td>14.81</td>
</tr>
<tr>
<td>FOOT</td>
<td>1.827</td>
<td>0.28</td>
<td>15.83</td>
<td>5.93</td>
<td>14.81</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.14</td>
<td>13.47</td>
<td>1.65</td>
<td>6.87</td>
</tr>
<tr>
<td>GASO</td>
<td>5.710</td>
<td>0.09</td>
<td>7.70</td>
<td>3.23</td>
<td>12.07</td>
</tr>
<tr>
<td>HEAL</td>
<td>4.802</td>
<td>0.21</td>
<td>12.82</td>
<td>3.18</td>
<td>8.77</td>
</tr>
<tr>
<td>HOUS</td>
<td>6.223</td>
<td>0.03</td>
<td>5.24</td>
<td>0.52</td>
<td>4.36</td>
</tr>
<tr>
<td>LIFE</td>
<td>1.253</td>
<td>0.35</td>
<td>15.82</td>
<td>3.91</td>
<td>11.02</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.04</td>
<td>3.42</td>
<td>−0.23</td>
<td>−0.80</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.10</td>
<td>10.34</td>
<td>1.59</td>
<td>7.23</td>
</tr>
<tr>
<td>PERS</td>
<td>3.817</td>
<td>0.30</td>
<td>15.56</td>
<td>4.55</td>
<td>10.53</td>
</tr>
<tr>
<td>READ</td>
<td>2.287</td>
<td>0.25</td>
<td>13.48</td>
<td>2.52</td>
<td>5.98</td>
</tr>
<tr>
<td>SAPP</td>
<td>1.037</td>
<td>0.20</td>
<td>13.80</td>
<td>2.86</td>
<td>8.61</td>
</tr>
<tr>
<td>TELE</td>
<td>5.788</td>
<td>0.07</td>
<td>6.29</td>
<td>3.52</td>
<td>14.09</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.19</td>
<td>13.25</td>
<td>2.45</td>
<td>7.50</td>
</tr>
<tr>
<td>TOB</td>
<td>1.155</td>
<td>−0.01</td>
<td>−0.54</td>
<td>0.24</td>
<td>0.69</td>
</tr>
<tr>
<td>TOYS</td>
<td>2.504</td>
<td>0.24</td>
<td>12.14</td>
<td>6.26</td>
<td>14.36</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>0.23</td>
<td>21.63</td>
<td>0.93</td>
<td>3.88</td>
</tr>
</tbody>
</table>

*FDHO* had no observations with zero expenditure.
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

**Logit model, dependent variable **\( \text{CATBUY} \)**

<table>
<thead>
<tr>
<th></th>
<th>( EXPPC \times 10^4 )</th>
<th>( SIZE \times 10^{-2} )</th>
<th>( REFAGE \times 10^{-2} )</th>
<th>( \text{COLLEGE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2.815</td>
<td>2.06</td>
<td>18.34</td>
<td>20.02</td>
</tr>
<tr>
<td>CLOT</td>
<td>4.500</td>
<td>2.51</td>
<td>17.22</td>
<td>32.00</td>
</tr>
<tr>
<td>DOM</td>
<td>1.661</td>
<td>1.50</td>
<td>15.28</td>
<td>22.50</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>1.38</td>
<td>11.60</td>
<td>35.93</td>
</tr>
<tr>
<td>ELEC</td>
<td>5.828</td>
<td>1.63</td>
<td>7.28</td>
<td>44.17</td>
</tr>
<tr>
<td>FDAW</td>
<td>5.102</td>
<td>2.71</td>
<td>14.40</td>
<td>17.42</td>
</tr>
<tr>
<td>FDHO</td>
<td>6.334</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOOT</td>
<td>1.827</td>
<td>1.39</td>
<td>14.69</td>
<td>29.17</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>1.43</td>
<td>12.00</td>
<td>21.16</td>
</tr>
<tr>
<td>GASO</td>
<td>5.710</td>
<td>1.50</td>
<td>7.50</td>
<td>47.81</td>
</tr>
<tr>
<td>HEAL</td>
<td>4.802</td>
<td>2.29</td>
<td>13.58</td>
<td>21.11</td>
</tr>
<tr>
<td>HOUS</td>
<td>6.223</td>
<td>4.31</td>
<td>5.78</td>
<td>37.81</td>
</tr>
<tr>
<td>LIFE</td>
<td>1.253</td>
<td>1.38</td>
<td>13.94</td>
<td>24.61</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.41</td>
<td>3.50</td>
<td>173.58</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>1.21</td>
<td>9.65</td>
<td>23.27</td>
</tr>
<tr>
<td>PERS</td>
<td>3.817</td>
<td>1.78</td>
<td>15.07</td>
<td>21.91</td>
</tr>
<tr>
<td>READ</td>
<td>2.287</td>
<td>1.18</td>
<td>12.35</td>
<td>11.97</td>
</tr>
<tr>
<td>SAPP</td>
<td>1.037</td>
<td>1.24</td>
<td>12.47</td>
<td>19.99</td>
</tr>
<tr>
<td>TELE</td>
<td>5.788</td>
<td>1.24</td>
<td>6.20</td>
<td>51.87</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>1.20</td>
<td>11.97</td>
<td>17.77</td>
</tr>
<tr>
<td>TOB</td>
<td>1.155</td>
<td>0.07</td>
<td>0.64</td>
<td>1.28</td>
</tr>
<tr>
<td>TOYS</td>
<td>2.504</td>
<td>1.04</td>
<td>11.53</td>
<td>27.08</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>1.92</td>
<td>15.76</td>
<td>9.60</td>
</tr>
</tbody>
</table>

**Probit model, dependent variable **\( \text{CATBUY} \)**

<table>
<thead>
<tr>
<th></th>
<th>( EXPPC \times 10^4 )</th>
<th>( SIZE \times 10^{-2} )</th>
<th>( REFAGE \times 10^{-2} )</th>
<th>( \text{COLLEGE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>2.815</td>
<td>1.17</td>
<td>19.26</td>
<td>11.97</td>
</tr>
<tr>
<td>CLOT</td>
<td>4.500</td>
<td>1.34</td>
<td>18.00</td>
<td>18.37</td>
</tr>
<tr>
<td>DOM</td>
<td>1.661</td>
<td>0.89</td>
<td>15.77</td>
<td>13.35</td>
</tr>
<tr>
<td>EDUC</td>
<td>561</td>
<td>0.78</td>
<td>11.88</td>
<td>19.78</td>
</tr>
<tr>
<td>ELEC</td>
<td>5.828</td>
<td>0.71</td>
<td>7.18</td>
<td>19.93</td>
</tr>
<tr>
<td>FDAW</td>
<td>5.102</td>
<td>1.37</td>
<td>14.87</td>
<td>9.53</td>
</tr>
<tr>
<td>FDHO</td>
<td>6.334</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOOT</td>
<td>1.827</td>
<td>0.82</td>
<td>15.39</td>
<td>17.60</td>
</tr>
<tr>
<td>FURN</td>
<td>487</td>
<td>0.80</td>
<td>12.45</td>
<td>11.37</td>
</tr>
<tr>
<td>GASO</td>
<td>5.710</td>
<td>0.61</td>
<td>7.37</td>
<td>21.79</td>
</tr>
<tr>
<td>HEAL</td>
<td>4.802</td>
<td>1.18</td>
<td>13.94</td>
<td>11.97</td>
</tr>
<tr>
<td>HOUS</td>
<td>6.223</td>
<td>1.33</td>
<td>5.76</td>
<td>14.17</td>
</tr>
<tr>
<td>LIFE</td>
<td>1.253</td>
<td>0.81</td>
<td>14.78</td>
<td>14.40</td>
</tr>
<tr>
<td>LOCT</td>
<td>692</td>
<td>0.21</td>
<td>3.30</td>
<td>0.80</td>
</tr>
<tr>
<td>MAPP</td>
<td>399</td>
<td>0.67</td>
<td>9.94</td>
<td>12.10</td>
</tr>
<tr>
<td>PERS</td>
<td>3.817</td>
<td>0.97</td>
<td>15.47</td>
<td>12.93</td>
</tr>
<tr>
<td>READ</td>
<td>2.287</td>
<td>0.70</td>
<td>12.74</td>
<td>7.14</td>
</tr>
<tr>
<td>SAPP</td>
<td>1.037</td>
<td>0.73</td>
<td>12.95</td>
<td>11.49</td>
</tr>
<tr>
<td>TELE</td>
<td>5.788</td>
<td>0.55</td>
<td>6.11</td>
<td>24.85</td>
</tr>
<tr>
<td>TEXT</td>
<td>992</td>
<td>0.71</td>
<td>12.53</td>
<td>10.21</td>
</tr>
<tr>
<td>TOB</td>
<td>1.155</td>
<td>0.05</td>
<td>0.79</td>
<td>0.84</td>
</tr>
<tr>
<td>TOYS</td>
<td>2.504</td>
<td>0.62</td>
<td>11.91</td>
<td>16.57</td>
</tr>
<tr>
<td>TRIP</td>
<td>516</td>
<td>1.06</td>
<td>16.91</td>
<td>4.84</td>
</tr>
</tbody>
</table>
10.6. Answers to the additional exercises

<table>
<thead>
<tr>
<th>Marginal effects, linear probability model, logit and probit</th>
<th>EXPPC4 \times 10^{-4}</th>
<th>SIZE \times 10^{-2}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPM</td>
<td>logit</td>
</tr>
<tr>
<td>ADM</td>
<td>0.38</td>
<td>0.51</td>
</tr>
<tr>
<td>CLOT</td>
<td>0.33</td>
<td>0.48</td>
</tr>
<tr>
<td>DOM</td>
<td>0.30</td>
<td>0.28</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>ELEC</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>FAW</td>
<td>0.23</td>
<td>0.36</td>
</tr>
<tr>
<td>FOOT</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>FURN</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>GASO</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>HEAL</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td>HOUS</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>LIFE</td>
<td>0.35</td>
<td>0.21</td>
</tr>
<tr>
<td>LOCT</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>MAPP</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>PERS</td>
<td>0.30</td>
<td>0.42</td>
</tr>
<tr>
<td>READ</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>SAPP</td>
<td>0.20</td>
<td>0.16</td>
</tr>
<tr>
<td>TELE</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>TEXT</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>TOB</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>TOYS</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>TRIP</td>
<td>0.23</td>
<td>0.11</td>
</tr>
</tbody>
</table>
### Marginal effects, linear probability model, logit and probit

<table>
<thead>
<tr>
<th></th>
<th>REFAGE $\times 10^{-2}$</th>
<th>COLLEGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPM logit probit</td>
<td>LPM logit probit</td>
</tr>
<tr>
<td>ADM</td>
<td>-0.34 -0.42 -0.40</td>
<td>0.22 0.24 0.24</td>
</tr>
<tr>
<td>CLOT</td>
<td>-0.35 -0.33 -0.34</td>
<td>0.05 0.04 0.04</td>
</tr>
<tr>
<td>DOM</td>
<td>0.16 0.17 0.17</td>
<td>0.09 0.10 0.10</td>
</tr>
<tr>
<td>EDUC</td>
<td>-0.12 -0.14 -0.15</td>
<td>0.05 0.05 0.05</td>
</tr>
<tr>
<td>ELEC</td>
<td>0.16 0.13 0.13</td>
<td>0.02 0.01 0.01</td>
</tr>
<tr>
<td>FDAW</td>
<td>-0.27 -0.24 -0.26</td>
<td>0.11 0.08 0.09</td>
</tr>
<tr>
<td>FDHO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOOT</td>
<td>-0.22 -0.25 -0.25</td>
<td>0.01 0.02 0.02</td>
</tr>
<tr>
<td>FURN</td>
<td>-0.07 -0.08 -0.08</td>
<td>0.01 0.02 0.02</td>
</tr>
<tr>
<td>GASO</td>
<td>-0.00 0.01 0.01</td>
<td>0.07 0.05 0.06</td>
</tr>
<tr>
<td>HEAL</td>
<td>0.82 0.80 0.85</td>
<td>0.11 0.09 0.09</td>
</tr>
<tr>
<td>HOUS</td>
<td>0.04 0.02 0.03</td>
<td>0.01 0.00 0.01</td>
</tr>
<tr>
<td>LIFE</td>
<td>0.19 0.19 0.20</td>
<td>0.04 0.04 0.04</td>
</tr>
<tr>
<td>LOCT</td>
<td>-0.15 -0.15 -0.15</td>
<td>0.00 0.00 0.00</td>
</tr>
<tr>
<td>MAPP</td>
<td>-0.00 0.00 0.00</td>
<td>-0.01 -0.01 -0.01</td>
</tr>
<tr>
<td>PERS</td>
<td>0.29 0.31 0.31</td>
<td>0.12 0.11 0.12</td>
</tr>
<tr>
<td>READ</td>
<td>0.37 0.40 0.40</td>
<td>0.16 0.18 0.17</td>
</tr>
<tr>
<td>SAPP</td>
<td>-0.03 -0.04 -0.04</td>
<td>0.03 0.04 0.04</td>
</tr>
<tr>
<td>TELE</td>
<td>0.31 0.23 0.25</td>
<td>0.01 0.01 0.01</td>
</tr>
<tr>
<td>TEXT</td>
<td>-0.03 -0.04 -0.04</td>
<td>0.04 0.04 0.04</td>
</tr>
<tr>
<td>TOB</td>
<td>-0.17 -0.17 -0.17</td>
<td>-0.10 -0.09 -0.09</td>
</tr>
<tr>
<td>TOYS</td>
<td>-0.13 -0.14 -0.14</td>
<td>0.06 0.06 0.06</td>
</tr>
<tr>
<td>TRIP</td>
<td>-0.03 -0.03 -0.03</td>
<td>0.03 0.04 0.04</td>
</tr>
</tbody>
</table>

A10.2 The finding that the marginal effect of educational attainment was lower for males than for females over most of the range $S \geq 9$ is plausible because the probability of working is much closer to 1 for males than for females for $S \geq 9$, and hence the possible sensitivity of the participation rate to $S$ is smaller.

The explanation of the finding that the marginal effect of educational attainment decreases with educational attainment for both males and females over the range $S \geq 9$ is similar. For both sexes, the greater is $S$, the greater is the participation rate, and hence the smaller is the scope for it being increased by further education.

The OLS estimates of the marginal effect of educational attainment are given by the slope coefficients and they are very similar to the logit estimates at the mean, the reason being that most of the observations on $S$ are confined to the middle part of the sigmoid curve where it is relatively linear.

A10.3 Discuss whether the relationships indicated by the probability and marginal effect curves appear to be plausible.

The probability curve indicates an inverse relationship between schooling and the probability of being obese. This seems entirely plausible. The more educated tend to have healthier lifestyles, including eating habits. Over the relevant range, the marginal effect falls a little in absolute terms (is less negative) as schooling
10.6. Answers to the additional exercises

increases. This is in keeping with the idea that further schooling may have less
effect on the highly educated than on the less educated (but the difference is not
large).

Add the probability function and the marginal effect function for the LPM to the
diagram. Explain why you drew them the way you did.

![Figure 10.6: Scatter diagram of probability of being obese and marginal effect against
years of schooling.]

The estimated probability function for the LPM is just the regression equation and
the marginal effect is the coefficient of $S$. They are shown as the dashed lines in the
diagram.

The logit model is considered to have several advantages over the LPM. Explain
what these advantages are. Evaluate the importance of the advantages of the logit
model in this particular case.

The disadvantages of the LPM are (1) that it can give nonsense fitted values
(predicted probabilities greater than 1 or less than 0); (2) the disturbance term in
observation $i$ must be equal to either $-1 - F(Z_i)$ (if the dependent variable is equal
to 1) or $-F(Z_i)$ (if the dependent variable is equal to 0) and so it violates the usual
assumption that the disturbance term is normally distributed, although this may
not matter asymptotically; (3) the disturbance term will be heteroskedastic
because $Z_i$ is different for different observations; (4) the LPM implicitly assumes
that the marginal effect of each explanatory variable is constant over its entire
range, which is often intuitively unappealing.

In this case, nonsense predictions are clearly not an issue. The assumption of a
constant marginal effect does not seem to be a problem either, given the
approximate linearity of the logit $F(Z)$.

The LPM is fitted using OLS. Explain how, instead, it might be fitted using
maximum likelihood estimation:

Write down the probability of being obese for any obese individual, given $S_i$ for that
individual, and write down the probability of not being obese for any non-obese
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

individual, again given $S_i$ for that individual.

Obese: $p_i^O = \beta_1 + \beta_2 S_i$; not obese: $p_i^{NO} = 1 - \beta_1 - \beta_2 S_i$.

Write down the likelihood function for this sample of 164 obese individuals and 376 non-obese individuals.

$$L(\beta_1, \beta_2 \mid \text{data}) = \prod_{\text{OBESE}} p_i^O \prod_{\text{NOT OBSE}} p_i^{NO} = \prod_{\text{OBESE}} (\beta_1 + \beta_2 S_i) \prod_{\text{NOT OBSE}} (1 - \beta_1 - \beta_2 S_i).$$

Explain how one would use this function to estimate the parameters. [Note: You are not expected to attempt to derive the estimators of the parameters.]

You would use some algorithm to find the values of $\beta_1$ and $\beta_2$ that maximise the function.

Explain whether your maximum likelihood estimators will be the same or different from those obtained using least squares.

Least squares involves finding the extremum of a completely different expression and will therefore lead to different estimators.

10.4 Explain how one may derive the marginal effects of the explanatory variables on the probability of having a child less than 6 in the household, and calculate for both males and females the marginal effects at the means of AGE and $S$.

Since $p$ is a function of $Z$, and $Z$ is a linear function of the $X$ variables, the marginal effect of $X_j$ is:

$$\frac{\partial p}{\partial X_j} = \frac{dp}{dZ} \frac{\partial Z}{\partial X_j} = \frac{dp}{dZ}\beta_j$$

where $\beta_j$ is the coefficient of $X_j$ in the expression for $Z$. In the case of probit analysis, $p = F(Z)$ is the cumulative standardised normal distribution. Hence $dp/dZ$ is just the standardised normal distribution.

For males, this is 0.368 when evaluated at the means. Hence the marginal effect of $AGE$ is $0.368 \times -0.137 = -0.050$ and that of $S$ is $0.368 \times 0.132 = 0.049$. For females the corresponding figures are $0.272 \times -0.154 = -0.042$ and $0.272 \times 0.094 = 0.026$, respectively. So for every extra year of age, the probability is reduced by 5.0 per cent for males and 4.2 per cent for females. For every extra year of schooling, the probability increases by 4.9 per cent for males and 2.6 per cent for females.

Explain whether the signs of the marginal effects are plausible. Explain whether you would expect the marginal effect of schooling to be higher for males or for females.

Yes. Given that the cohort is aged 35–42, the respondents have passed the age at which most adults start families, and the older they are, the less likely they are to have small children in the household. At the same time, the more educated the respondent, the more likely he or she is to have started having a family relatively late, so the positive effect of schooling is also plausible. However, given the age of the cohort, it is likely to be weaker for females than for males, given that most females intending to have families will have started them by this time, irrespective of their education.
At a seminar someone asks the researcher whether the marginal effect of $S$ is significantly different for males and females. The researcher does not know how to test whether the difference is significant and asks you for advice. What would you say?

Fit a probit regression for the combined sample, adding a male intercept dummy and male slope dummies for $AGE$ and $S$. Test the coefficient of the slope dummy for $S$.

10.5 The $Z$ function will be of the form:

$$Z = \beta_1 + \beta_2 A + \beta_3 S + \beta_4 AS$$

so the marginal effects are:

$$\frac{\partial p}{\partial A} = \frac{dp}{dZ} \frac{\partial Z}{\partial A} = f(Z)(\beta_2 + \beta_4 S)$$

and:

$$\frac{\partial p}{\partial S} = \frac{dp}{dZ} \frac{\partial Z}{\partial S} = f(Z)(\beta_3 + \beta_4 A).$$

Both factors depend on the values of $A$ and/or $S$, but the marginal effects could be evaluated for a representative individual using the mean values of $A$ and $S$ in the sample.

A10.6 Discuss the conclusions one may reach, given the probit output and the table, commenting on their plausibility.

Being male has a small but highly significant negative effect. This is plausible because males tend to marry later than females and the cohort is still relatively young.

Age has a highly significant positive effect, again plausible because older people are more likely to have married than younger people.

Schooling has no apparent effect at all. It is not obvious whether this is plausible.

Cognitive ability has a highly significant positive effect. Again, it is not obvious whether this is plausible.

The researcher considers including $CHILD$, a dummy variable defined to be 1 if the respondent had children, and 0 otherwise, as an explanatory variable. When she does this, its $z$-statistic is 33.65 and its marginal effect 0.5685. Discuss these findings.

Obviously one would expect a high positive correlation between being married and having children and this would account for the huge and highly significant coefficient. However getting married and having children are often a joint decision, and accordingly it is simplistic to suppose that one characteristic is a determinant of the other. The finding should not be taken at face value.

A10.7 Determine the maximum likelihood estimate of $\alpha$, assuming that $\beta$ is known.

The log-likelihood function is:

$$\log L(\alpha \mid \beta, T_1, \ldots, T_n) = n \log \alpha - \alpha \sum (T_i - \beta).$$
10. Binary choice and limited dependent variable models, and maximum likelihood estimation

Setting the first derivative with respect to $\alpha$ equal to zero, we have:

$$\frac{n}{\hat{\alpha}} - \sum (T_i - \beta) = 0$$

and hence:

$$\hat{\alpha} = \frac{1}{\sum T_i - \beta}.$$

The second derivative is $-n/\hat{\alpha}^2$, which is negative, confirming we have maximised the loglikelihood function.

A10.8 From the solution to Exercise 10.14, the log-likelihood function for $p$ is:

$$\log L(p) = m \log p + (n - m) \log(1 - p).$$

Thus the LR statistic is:

$$LR = 2 \left( m \log \left( \frac{m}{n} \right) + (n - m) \log \left( \frac{1}{1 - p_0} \right) \right).$$

If $m = 40$ and $n = 100$, the LR statistic for $H_0 : p = 0.5$ is:

$$LR = 2 \left( 40 \log \left( \frac{0.4}{0.5} \right) + 60 \log \left( \frac{0.6}{0.5} \right) \right) = 4.03.$$

We would reject the null hypothesis at the 5 per cent level (critical value of chi-squared with one degree of freedom 3.84) but not at the 1 per cent level (critical value 6.64).

A10.9 The first derivative of the log-likelihood function is:

$$\frac{d \log L(p)}{dp} = \frac{m}{p} - \frac{n - m}{1 - p} = 0$$

and the second differential is:

$$\frac{d \log L(p)}{dp^2} = -\frac{m}{p^2} - \frac{n - m}{(1 - p)^2}.$$

Evaluated at $p = m/n$:

$$\frac{d \log L(p)}{dp^2} = -\frac{n^2}{m} \cdot \frac{n - m}{(1 - \frac{m}{n})^2} = -n^2 \left( \frac{1}{m} + \frac{1}{n - m} \right) = -\frac{n^3}{m(n - m)}.$$

The variance of the ML estimate is given by:

$$\left( -\frac{d \log L(p)}{dp^2} \right)^{-1} = \left( \frac{n^3}{m(n - m)} \right)^{-1} = \frac{m(n - m)}{n^3}.$$

The Wald statistic is therefore:

$$\left( \frac{m}{n} - p_0 \right)^2 = \left( \frac{m}{n} - p_0 \right)^2.$$

Given the data, this is equal to:

$$\frac{(0.4 - 0.5)^2}{100 \times 0.4 \times 0.6} = 4.17.$$

Under the null hypothesis this has a chi-squared distribution with one degree of freedom, and so the conclusion is the same as in Exercise A.8.
Chapter 11
Models using time series data

11.1 Overview

This chapter introduces the application of regression analysis to time series data, beginning with static models and then proceeding to dynamic models with lagged variables used as explanatory variables. It is shown that multicollinearity is likely to be a problem in models with unrestricted lag structures and that this provides an incentive to use a parsimonious lag structure, such as the Koyck distribution. Two models using the Koyck distribution, the adaptive expectations model and the partial adjustment model, are described, together with well-known applications to aggregate consumption theory, Friedman’s permanent income hypothesis in the case of the former and Brown’s habit persistence consumption function in the case of the latter. The chapter concludes with a discussion of prediction and stability tests in time series models.

11.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain why multicollinearity is a common problem in time series models, especially dynamic ones with lagged explanatory variables
- describe the properties of a model with a lagged dependent variable (ADL(1,0) model)
- describe the assumptions underlying the adaptive expectations and partial adjustment models
- explain the properties of OLS estimators of the parameters of ADL(1,0) models
- explain how predetermined variables may be used as instruments in the fitting of models using time series data
- explain in general terms the objectives of time series analysts and those constructing VAR models.
11. Models using time series data

11.3 Additional exercises

A11.1 The output below shows the result of linear and logarithmic regressions of expenditure on food on income, relative price, and population (measured in thousands) using the Demand Functions data set, together with the correlations among the variables. Provide an interpretation of the regression coefficients and perform appropriate statistical tests.

---

**Dependent Variable: FOOD**

Method: Least Squares

Sample: 1959 2003

Included observations: 45

<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>-19.49285</td>
<td>88.86914</td>
<td>-0.219343</td>
<td>0.8275</td>
</tr>
<tr>
<td>DPI</td>
<td>0.031713</td>
<td>0.010658</td>
<td>2.975401</td>
<td>0.0049</td>
</tr>
<tr>
<td>PREFOOD</td>
<td>0.403356</td>
<td>0.365133</td>
<td>1.104681</td>
<td>0.2757</td>
</tr>
<tr>
<td>POP</td>
<td>0.001140</td>
<td>0.000563</td>
<td>2.024017</td>
<td>0.0495</td>
</tr>
</tbody>
</table>

R-squared: 0.988529  Mean dependent var: 422.0374
Adjusted R-squared: 0.987690  S.D. dependent var: 91.58053
S.E. of regression: 10.16104  Akaike info criter: 7.559685
Sum squared resid: 4233.113  Schwarz criterion: 7.720278
Log likelihood: -166.0929  F-statistic: 1177.745
Durbin-Watson stat: 0.404076  Prob(F-statistic): 0.000000

---

**Dependent Variable: LGFOOD**

Method: Least Squares

Sample: 1959 2003

Included observations: 45

<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>5.293654</td>
<td>2.762757</td>
<td>1.916077</td>
<td>0.0623</td>
</tr>
<tr>
<td>LGDPI</td>
<td>0.589239</td>
<td>0.080158</td>
<td>7.351014</td>
<td>0.0000</td>
</tr>
<tr>
<td>LGPRFOOD</td>
<td>-0.122598</td>
<td>0.084355</td>
<td>-1.453361</td>
<td>0.1537</td>
</tr>
<tr>
<td>LGPOP</td>
<td>-0.289219</td>
<td>0.258762</td>
<td>-1.117706</td>
<td>0.2702</td>
</tr>
</tbody>
</table>

R-squared: 0.992245  Mean dependent var: 6.021331
Adjusted R-squared: 0.991678  S.D. dependent var: 0.222787
S.E. of regression: 0.020324  Akaike info criter: -4.869317
Sum squared resid: 0.016936  Schwarz criterion: -4.708725
Log likelihood: 113.5596  F-statistic: 1748.637
Durbin-Watson stat: 0.488502  Prob(F-statistic): 0.000000

---

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
11.3. Additional exercises

### Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>LGFOOD</th>
<th>LGDPI</th>
<th>LGPRFOOD</th>
<th>LGPOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGFOOD</td>
<td>1.000</td>
<td>0.996</td>
<td>-0.613</td>
<td>0.991</td>
</tr>
<tr>
<td>LGDPI</td>
<td>0.996</td>
<td>1.000</td>
<td>-0.605</td>
<td>0.995</td>
</tr>
<tr>
<td>LGPRFOOD</td>
<td>-0.613</td>
<td>-0.605</td>
<td>1.000</td>
<td>-0.641</td>
</tr>
<tr>
<td>LGPOP</td>
<td>0.991</td>
<td>0.995</td>
<td>-0.641</td>
<td>1.000</td>
</tr>
</tbody>
</table>

A11.2 Perform regressions parallel to those in Exercise A11.1 using your category of expenditure and provide an interpretation of the coefficients.

A11.3 The output shows the result of a logarithmic regression of expenditure on food per capita, on income per capita, both measured in US$ million, and the relative price index for food. Provide an interpretation of the coefficients, demonstrate that the specification is a restricted version of the logarithmic regression in Exercise A11.1, and perform an $F$ test of the restriction.

```markdown
Dependent Variable: LGFOODPC
Method: Least Squares
Sample: 1959 2003
Included observations: 45

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-5.425877</td>
<td>0.353655</td>
<td>-15.34231</td>
</tr>
<tr>
<td>LGDPIPC</td>
<td>0.280229</td>
<td>0.014641</td>
<td>19.14024</td>
</tr>
<tr>
<td>LGPRFOODPC</td>
<td>0.052952</td>
<td>0.082588</td>
<td>0.641160</td>
</tr>
</tbody>
</table>

R-squared: 0.92748
Mean dependent var: -6.321984
```

A11.4 Perform a regression parallel to that in Exercise A11.3 using your category of expenditure. Provide an interpretation of the coefficients, and perform an $F$ test of the restriction.

A11.5 The output shows the result of a logarithmic regression of expenditure on food per capita, on income per capita, the relative price index for food, and population. Provide an interpretation of the coefficients, demonstrate that the specification is equivalent to that for the logarithmic regression in Exercise A11.1, and use it to perform a $t$ test of the restriction in Exercise A11.3.

```markdown
Dependent Variable: LGFOODPC
Method: Least Squares
Sample: 1959 2003
```

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
11. Models using time series data

Included observations: 45

<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>5.293654</td>
<td>2.762757</td>
<td>1.916077</td>
<td>0.0623</td>
</tr>
<tr>
<td>LGDPI</td>
<td>0.589239</td>
<td>0.080158</td>
<td>7.351014</td>
<td>0.0000</td>
</tr>
<tr>
<td>LGPRFOOD</td>
<td>-0.122598</td>
<td>0.084355</td>
<td>-1.453361</td>
<td>0.1537</td>
</tr>
<tr>
<td>LGPOP</td>
<td>-0.699980</td>
<td>0.179299</td>
<td>-3.903973</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

R-squared 0.947037  Mean dependent var -6.321984
Adjusted R-squared 0.943161  S.D. dependent var 0.085249
S.E. of regression 0.020324  Akaike info criter -4.869317
Sum squared resid 0.016936  Schwarz criterion -4.708725
Log likelihood 113.5596  F-statistic 244.3727
Durbin-Watson stat 0.488502  Prob(F-statistic) 0.000000

A11.6 Perform a regression parallel to that in Exercise A11.5 using your category of expenditure, and perform a $t$ test of the restriction implicit in the specification in Exercise A11.4.

A11.7 In Exercise 11.9 you fitted the model:

$$LGCAT = \beta_1 + \beta_2 LGDPI + \beta_3 LGDPI(-1) + \beta_4 LGPRCAT + \beta_5 LGPRCAT(-1) + u$$

where $CAT$ stands for your category of expenditure.

- Show that $(\beta_2 + \beta_3)$ and $(\beta_4 + \beta_5)$ are theoretically the long-run (equilibrium) income and price elasticities.
- Reparameterise the model and fit it to obtain direct estimates of these long-run elasticities and their standard errors.
- Confirm that the estimates are equal to the sum of the individual short-run elasticities found in Exercise 11.9.
- Compare the standard errors with those found in Exercise 11.9 and state your conclusions.

A11.8 In a certain bond market, the demand for bonds, $B_t$, in period $t$ is negatively related to the expected interest rate, $i^e_{t+1}$, in period $t + 1$:

$$B_t = \beta_1 + \beta_2 i^e_{t+1} + u_t$$  \hspace{1cm} (1)

where $u_t$ is a disturbance term not subject to autocorrelation. The expected interest rate is determined by an adaptive expectations process:

$$i^e_{t+1} - i^e_t = \lambda(i_t - i^e_t)$$  \hspace{1cm} (2)

where $i_t$ is the actual rate of interest in period $t$. A researcher uses the following model to fit the relationship:

$$B_t = \gamma_1 + \gamma_2 i_t + \gamma_3 B_{t-1} + v_t$$  \hspace{1cm} (3)

where $v_t$ is a disturbance term.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
11.3. Additional exercises

- Show how this model may be derived from the demand function and the adaptive expectations process.
- Explain why inconsistent estimates of the parameters will be obtained if equation (3) is fitted using ordinary least squares (OLS). (A mathematical proof is not required. Do not attempt to derive expressions for the bias.)
- Describe a method for fitting the model that would yield consistent estimates.
- Suppose that $u_t$ was subject to the first-order autoregressive process:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where $\varepsilon_t$ is not subject to autocorrelation. How would this affect your answer to the second part of this question?
- Suppose that the true relationship was actually:

$$B_t = \beta_1 + \beta_2 i_t + u_t \quad (1*)$$

with $u_t$ not subject to autocorrelation, and the model is fitted by regressing $B_t$ on $i_t$ and $B_{t-1}$, as in equation (3), using OLS. How would this affect the regression results?
- How plausible do you think an adaptive expectations process is for modelling expectations in a bond market?

A11.9 The output shows the result of a logarithmic regression of expenditure on food on income, relative price, population, and lagged expenditure on food using the Demand Functions data set. Provide an interpretation of the regression coefficients, paying attention to both short-run and long-run dynamics, and perform appropriate statistical tests.

---------------------------------------------------------------------------
Dependent Variable: LGFOOD
Method: Least Squares
Sample(adjusted): 1960 2003
Included observations: 44 after adjusting endpoints
---------------------------------------------------------------------------
Variable       Coefficient  Std. Error  t-Statistic  Prob.  
---------------------------------------------------------------------------
  C             1.487645    2.072156    0.717921    0.4771 
  LGDPI         0.143829    0.090334    1.592194    0.1194 
  LGPRFOOD     -0.095749    0.061118   -1.566613    0.1253 
  LGPOP        -0.046515    0.189453   -0.245524    0.8073 
  LGFOOD(-1)   0.727290    0.113831    6.389195    0.0000 
---------------------------------------------------------------------------
R-squared       0.995886    Mean dependent var 6.030691 
Adjusted R-squared 0.995464  S.D. dependent var 0.216227 
S.E. of regression 0.04564  Akaike info criter-5.754397 
Sum squared resid 0.048272  Schwarz criterion -5.311188 
Log likelihood 126.3066  F-statistic 2359.938 
Durbin-Watson stat 1.103102  Prob(F-statistic) 0.000000 
---------------------------------------------------------------------------

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
11. Models using time series data

**A11.10** Perform a regression parallel to that in Exercise A11.9 using your category of expenditure. Provide an interpretation of the coefficients, and perform appropriate statistical tests.

**A11.11** In his classic study *Distributed Lags and Investment Analysis* (1954), Koyck investigated the relationship between investment in railcars and the volume of freight carried on the US railroads using data for the period 1884–1939. Assuming that the desired stock of railcars in year $t$ depended on the volume of freight in year $t-1$ and year $t-2$ and a time trend, and assuming that investment in railcars was subject to a partial adjustment process, he fitted the following regression equation using OLS (standard errors and constant term not reported):

$$\hat{I}_t = 0.077F_{t-1} + 0.017F_{t-2} - 0.0033t - 0.110K_{t-1} \quad R^2 = 0.85$$

where $I_t = K_t - K_{t-1}$ is investment in railcars in year $t$ (thousands), $K_t$ is the stock of railcars at the end of year $t$ (thousands), and $F_t$ is the volume of freight handled in year $t$ (ton-miles).

Provide an interpretation of the equation and describe the dynamic process implied by it. (Note: It is best to substitute $K_t - K_{t-1}$ for $I_t$ in the regression and treat it as a dynamic relationship determining $K_t$.)

**A11.12** Two researchers agree that a model consists of the following relationships:

\[
\begin{align*}
Y_t & = \alpha_1 + \alpha_2 X_t + u_t \quad (1) \\
X_t & = \beta_1 + \beta_2 Y_{t-1} + v_t \quad (2) \\
Z_t & = \gamma_1 + \gamma_2 Y_t + \gamma_3 X_t + \gamma_4 Q_t + w_t \quad (3)
\end{align*}
\]

where $u_t$, $v_t$, and $w_t$, are disturbance terms that are drawn from fixed distributions with zero mean. It may be assumed that they are distributed independently of $Q_t$ and of each other and that they are not subject to autocorrelation. All the parameters may be assumed to be positive and it may be assumed that $\alpha_2 \beta_2 < 1$.

- One researcher asserts that consistent estimates will be obtained if (2) is fitted using OLS and (1) is fitted using IV, with $Y_{t-1}$ as an instrument for $X_t$. Determine whether this is true.
- The other researcher asserts that consistent estimates will be obtained if both (1) and (2) are fitted using OLS, and that the estimate of $\beta_2$ will be more efficient than that obtained using IV. Determine whether this is true.
11.4  Answers to the starred exercises in the textbook

11.6

<table>
<thead>
<tr>
<th>Year</th>
<th>Y</th>
<th>K</th>
<th>L</th>
<th>Year</th>
<th>Y</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1899</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>1911</td>
<td>153</td>
<td>216</td>
<td>145</td>
</tr>
<tr>
<td>1900</td>
<td>101</td>
<td>107</td>
<td>105</td>
<td>1912</td>
<td>177</td>
<td>226</td>
<td>152</td>
</tr>
<tr>
<td>1901</td>
<td>112</td>
<td>114</td>
<td>110</td>
<td>1913</td>
<td>184</td>
<td>236</td>
<td>154</td>
</tr>
<tr>
<td>1902</td>
<td>122</td>
<td>122</td>
<td>118</td>
<td>1914</td>
<td>169</td>
<td>244</td>
<td>149</td>
</tr>
<tr>
<td>1903</td>
<td>124</td>
<td>131</td>
<td>123</td>
<td>1915</td>
<td>189</td>
<td>266</td>
<td>154</td>
</tr>
<tr>
<td>1904</td>
<td>122</td>
<td>138</td>
<td>116</td>
<td>1916</td>
<td>225</td>
<td>298</td>
<td>182</td>
</tr>
<tr>
<td>1905</td>
<td>143</td>
<td>149</td>
<td>125</td>
<td>1917</td>
<td>227</td>
<td>335</td>
<td>196</td>
</tr>
<tr>
<td>1906</td>
<td>152</td>
<td>163</td>
<td>133</td>
<td>1918</td>
<td>223</td>
<td>366</td>
<td>200</td>
</tr>
<tr>
<td>1907</td>
<td>151</td>
<td>176</td>
<td>138</td>
<td>1919</td>
<td>218</td>
<td>387</td>
<td>193</td>
</tr>
<tr>
<td>1908</td>
<td>126</td>
<td>185</td>
<td>121</td>
<td>1920</td>
<td>231</td>
<td>407</td>
<td>193</td>
</tr>
<tr>
<td>1909</td>
<td>155</td>
<td>198</td>
<td>140</td>
<td>1921</td>
<td>179</td>
<td>417</td>
<td>147</td>
</tr>
<tr>
<td>1910</td>
<td>159</td>
<td>208</td>
<td>144</td>
<td>1922</td>
<td>240</td>
<td>431</td>
<td>161</td>
</tr>
</tbody>
</table>

Source: Cobb and Douglas (1928)

The table gives the data used by Cobb and Douglas (1928) to fit the original Cobb–Douglas production function:

\[ Y_t = \beta_1 K_t^{\beta_2} L_t^{\beta_3} v_t \]

\( Y_t, K_t, \) and \( L_t, \) being index number series for real output, real capital input, and real labour input, respectively, for the manufacturing sector of the United States for the period 1899–1922 (1899 = 100). The model was linearised by taking logarithms of both sides and the following regressions was run (standard errors in parentheses):

\[
\hat{\log} Y = -0.18 + 0.23 \log K + 0.81 \log L \\
R^2 = 0.96 \\
(0.43) (0.06) (0.15)
\]

Provide an interpretation of the regression coefficients.

Answer:

The elasticities of output with respect to capital and labour are 0.23 and 0.81, respectively, both coefficients being significantly different from zero at very high significance levels. The fact that the sum of the elasticities is close to one suggests that there may be constant returns to scale. Regressing output per worker on capital per worker, one has:

\[
\hat{\log} \frac{Y}{L} = 0.01 + 0.25 \log \frac{K}{L} \\
R^2 = 0.63 \\
(0.02) (0.04)
\]

The smaller standard error of the slope coefficient suggests a gain in efficiency. Fitting a reparameterised version of the unrestricted model:

\[
\hat{\log} \frac{Y}{L} = -0.18 + 0.23 \log \frac{K}{L} + 0.04 \log L \\
R^2 = 0.64 \\
(0.43) (0.06) (0.09)
\]

we find that the restriction is not rejected.
11. Models using time series data

11.7 The Cobb–Douglas model in Exercise 11.6 makes no allowance for the possibility that output may be increasing as a consequence of technical progress, independently of \( K \) and \( L \). Technical progress is difficult to quantify and a common way of allowing for it in a model is to include an exponential time trend:

\[
Y_t = \beta_1 K_t^{\beta_2} L_t^{\beta_3} e^{\rho t v_t}
\]

where \( \rho \) is the rate of technical progress and \( t \) is a time trend defined to be 1 in the first year, 2 in the second, etc. The correlations between log \( K \), log \( L \) and \( t \) are shown in the table. Comment on the regression results.

\[
\hat{\log Y} = 2.81 - 0.53 \log K + 0.91 \log L + 0.047t \quad R^2 = 0.97
\]

\[
(1.38) \quad (0.34) \quad (0.14) \quad (0.021)
\]

<table>
<thead>
<tr>
<th></th>
<th>LGK</th>
<th>LGL</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGK</td>
<td>1.000000</td>
<td>0.909562</td>
<td>0.996834</td>
</tr>
<tr>
<td>LGL</td>
<td>0.909562</td>
<td>1.000000</td>
<td>0.896344</td>
</tr>
<tr>
<td>TIME</td>
<td>0.996834</td>
<td>0.896344</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Answer:
The elasticity of output with respect to labour is higher than before, now implausibly high given that, under constant returns to scale, it should measure the share of wages in output. The elasticity with respect to capital is negative and nonsensical. The coefficient of time indicates an annual exponential growth rate of 4.7 per cent, holding \( K \) and \( L \) constant. This is unrealistically high for the period in question. The implausibility of the results, especially those relating to capital and time (correlation 0.997), may be attributed to multicollinearity.

11.16 Demonstrate that the dynamic process (11.18) implies the long-run relationship given by (11.15).

Answer:
Equations (11.15) and (11.18) are:

\[
\hat{Y} = \frac{\beta_1}{1 - \beta_3} + \frac{\beta_2}{1 - \beta_3} \hat{X}
\]

\[
Y_t = \beta_1 (1 + \beta_3 + \beta_3^2 + \cdots) + \beta_2 X_t + \beta_2 \beta_3 X_{t-1} + \beta_2 \beta_3^2 X_{t-2} + \cdots + u_t + \beta_3 u_{t-1} + \beta_3^2 u_{t-2} + \cdots
\]

Putting \( X = \hat{X} \) for all \( X \) in (11.18), and ignoring the disturbance terms, the long-run relationship between \( Y \) and \( X \) is given by:

\[
\hat{Y} = \beta_1 (1 + \beta_3 + \beta_3^2 + \cdots) + \beta_2 \hat{X} + \beta_2 \beta_3 \hat{X} + \beta_2 \beta_3^2 \hat{X} + \cdots
\]

\[
= \frac{\beta_1}{1 - \beta_3} + (1 + \beta_3 + \beta_3^2 + \cdots) \beta_2 \hat{X}
\]

\[
= \frac{\beta_1}{1 - \beta_3} + \frac{\beta_2}{1 - \beta_3} \hat{X}.
\]
11.4. Answers to the starred exercises in the textbook

11.17 The compound disturbance term in the adaptive expectations model (11.37) does potentially give rise to a problem that will be discussed in Chapter 12 when we come to the topic of autocorrelation. It can be sidestepped by representing the model in the alternative form.

\[ Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 \lambda(1 - \lambda)X_{t-1} + \cdots + \beta_2 \lambda(1 - \lambda)^s X_{t-s} + \beta_2(1 - \lambda)^{s+1} X_{t-s} + u_t. \]

Show how this form might be obtained, and discuss how it might be fitted.

**Answer:**

We start by reprising equations (11.31) – (11.34) in the text. We assume that the dependent variable \( Y_t \) is related to \( X_{t+1}^e \), the value of \( X \) anticipated in the next time period:

\[ Y_t = \gamma_1 + \gamma_2 X_{t+1}^e + u_t. \]  \hspace{1cm} (11.31)

To make the model operational, we hypothesise that expectations are updated in response to the discrepancy between what had been anticipated for the current time period, \( X_{t+1}^e \), and the actual outcome, \( X_t \):

\[ X_{t+1}^e - X_t^e = \lambda(X_t - X_t^e) \]  \hspace{1cm} (11.32)

where \( \lambda \) may be interpreted as a speed of adjustment. We can rewrite this as (11.33):

\[ X_{t+1}^e = \lambda X_t + (1 - \lambda)X_t^e. \]  \hspace{1cm} (11.33)

Hence we obtain (11.34):

\[ Y_t = \gamma_1 + \gamma_2 \lambda X_t + \gamma_2(1 - \lambda)X_t^e + u_t. \]  \hspace{1cm} (11.34)

This includes the unobservable \( X_t^e \) on the right side. However, lagging (11.33), we have:

\[ X_t^e = \lambda X_{t-1} + (1 - \lambda)X_{t-1}^e. \]

Hence:

\[ Y_t = \gamma_1 + \gamma_2 \lambda X_t + \gamma_2(1 - \lambda)X_{t-1} + \gamma_2(1 - \lambda)^2 X_{t-1}^e + u_t. \]

This includes the unobservable \( X_{t-1}^e \) on the right side. However, continuing to lag and substitute, we have:

\[ Y_t = \gamma_1 + \gamma_2 \lambda X_t + \gamma_2 \lambda(1 - \lambda) X_{t-1} + \cdots + \gamma_2 \lambda(1 - \lambda)^s X_{t-s} + \gamma_2(1 - \lambda)^{s+1} X_{t-s}^e + u_t. \]

Provided that \( s \) is large enough for \( \gamma_2(1 - \lambda)^{s+1} \) to be very small, this may be fitted, omitting the unobservable final term, with negligible omitted variable bias.

We would fit it with a nonlinear regression technique that respected the constraints implicit in the theoretical structure of the coefficients.

11.19 The output below shows the result of fitting the model:

\[ LGFOOD = \beta_1 + \beta_2 LGDPI + \beta_2(1 - \lambda) LGDPI(-1) + \beta_2 \lambda(1 - \lambda)^2 LGDPI(-2) + \beta_2 \lambda(1 - \lambda)^3 LGDPI(-3) + \beta_3 LGPRFOOD + u \]

using the data on expenditure on food in the Demand Functions data set. \( LGFOOD \) and \( LGPRFOOD \) are the logarithms of expenditure on food and the
11. Models using time series data

relative price index series for food. C(1), C(2), C(3), and C(4) are estimates of \( \beta_1 \), \( \beta_2 \), \( \lambda \) and \( \beta_3 \), respectively. Explain how the regression equation could be interpreted as an adaptive expectations model and discuss the dynamics implicit in it, both short-run and long-run. Should the specification have included further lagged values of \( LGDPI \)?

<table>
<thead>
<tr>
<th>Dependent Variable: LGFOOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method: Least Squares</td>
</tr>
<tr>
<td>Sample(adjusted): 1962-2003</td>
</tr>
<tr>
<td>Included observations: 42 after adjusting endpoints</td>
</tr>
<tr>
<td>Convergence achieved after 25 iterations</td>
</tr>
</tbody>
</table>

\[
LGFOOD = C(1) + C(2) \cdot LGDPI + C(2) \cdot C(3) \cdot (1-C(3)) \cdot LGDPI(-1) + C(2) \cdot (C(3) \cdot (1-C(3)) \cdot LGDPI(-2) + C(2) \cdot C(3) \cdot (1-C(3)) \cdot LGDPI(-3) + C(4) \cdot LGPRFOOD
\]

<table>
<thead>
<tr>
<th>Coefficient Std. Error t-Statistic Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>C(1) 2.339513 0.468550 4.993091 0.0000</td>
</tr>
<tr>
<td>C(2) 0.496425 0.012264 40.47818 0.0000</td>
</tr>
<tr>
<td>C(3) 0.915046 0.442851 2.066264 0.0457</td>
</tr>
<tr>
<td>C(4) -0.089681 0.083250 -1.077247 0.2882</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-squared</th>
<th>Mean dependent var</th>
<th>6.049936</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted R-squared</td>
<td>S.D. dependent var</td>
<td>0.201706</td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>Akaike info criter-4.765636</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>Schwarz criterion</td>
<td>-4.600143</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>Durbin-Watson stat</td>
<td>0.449978</td>
</tr>
</tbody>
</table>

Answer:

Suppose that the model is:

\[
LGFOOD_t = \gamma_1 + \gamma_2 \cdot LGDPI_{t+1}^e + \gamma_3 \cdot LGPRFOOD_t + u_t
\]

where \( LGDPI_{t+1}^e \) is expected \( LGDPI \) at time \( t+1 \), and that expectations for income are subject to the adaptive expectations process:

\[
LGDPI_{t+1}^e - LGDPI_t^e = \lambda (LGDPI_t - LGDPI_t^e).
\]

The adaptive expectations process may be rewritten:

\[
LGDPI_{t+1}^e = \lambda LGDPI_t + (1-\lambda) LGDPI_t^e.
\]

Lagging this equation one period and substituting, one has:

\[
LGDPI_{t+1}^e = \lambda LGDPI_t + \lambda(1-\lambda) LGDPI_{t-1} + (1-\lambda)^2 LGDPI_{t-1}^e.
\]

Lagging a second time and substituting, one has:

\[
LGDPI_{t+1}^e = \lambda LGDPI_t + \lambda(1-\lambda) LGDPI_{t-1} + \lambda(1-\lambda)^2 LGDPI_{t-2} + (1-\lambda)^3 LGDPI_{t-2}^e.
\]

Lagging a third time and substituting, one has:

\[
LGDPI_{t+1}^e = \lambda LGDPI_t + \lambda(1-\lambda) LGDPI_{t-1} + \lambda(1-\lambda)^2 LGDPI_{t-2} + \lambda(1-\lambda)^3 LGDPI_{t-3} + (1-\lambda)^4 LGDPI_{t-3}^e.
\]

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
11.4. Answers to the starred exercises in the textbook

Substituting this into the model, dropping the final unobservable term, one has the regression specification as stated in the question.

The estimates imply that the short-run income elasticity is 0.50. The speed of adjustment of expectations is 0.92. Hence the long-run income elasticity is $0.50/0.92 = 0.54$. The price side of the model has been assumed to be static. The estimate of the price elasticity is $-0.09$. The coefficient of the dropped unobservable term is $\gamma(1-\lambda)^4$. Given our estimates of $\gamma$ and $\lambda$, its estimate is 0.0003. Hence we are justified in neglecting it.

11.22 A researcher is fitting the following supply and demand model for a certain commodity, using a sample of time series observations:

\[
Q_{dt} = \beta_1 + \beta_2 P_t + u_{dt}
\]
\[
Q_{st} = \alpha_1 + \alpha_2 P_t + u_{st}
\]

where $Q_{dt}$ is the amount demanded at time $t$, $Q_{st}$ is the amount supplied, $P_t$ is the market clearing price, and $u_{dt}$ and $u_{st}$ are disturbance terms that are not necessarily independent of each other. It may be assumed that the market clears and so $Q_{dt} = Q_{st}$.

- What can be said about the identification of (a) the demand equation, (b) the supply equation?
- What difference would it make if supply at time $t$ was determined instead by price at time $t - 1$? That is:

\[
Q_{st} = \alpha_1 + \alpha_2 P_{t-1} + u_{st}.
\]

- What difference would it make if it could be assumed that $u_{dt}$ is distributed independently of $u_{st}$?

**Answer:**

The reduced form equation for $P_t$ is:

\[
P_t = \frac{1}{\alpha_2 - \beta_2}(\beta_1 - \alpha_1 + u_{dt} - u_{st}).
\]

$P_t$ is not independent of the disturbance term in either equation and so OLS would yield inconsistent estimates. There is no instrument available, so both equations are underidentified.

Provided that $u_{dt}$ is not subject to autocorrelation, $P_{t-1}$ could be used as an instrument in the demand equation. Provided that $u_{st}$ is not subject to autocorrelation, OLS could be used to fit the second equation. It makes no difference whether or not $u_{dt}$ is distributed independently of $u_{st}$.

The first equation could, alternatively, be fitted using OLS, with the variables switched. From the second equation, $P_{t-1}$ determines $Q_t$, and then, given $Q_t$, the demand equation determines $P_t$:

\[
P_t = \frac{1}{\beta_2}(Q_t - \beta_1 - u_{dt}).
\]

The reciprocal of the slope coefficient provides a consistent estimator of $\beta_2$. 

© Christopher Dougherty, 2016. All rights reserved. Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
11. Models using time series data

11.24 Consider the following simple macroeconomic model:

\[ C_t = \beta_1 + \beta_2 Y_t + u_{C_t} \]
\[ I_t = \alpha_1 + \alpha_2 (Y_t - Y_{t-1}) + u_{I_t} \]
\[ Y_t = C_t + I_t \]

where \( C_t, I_t, \) and \( Y_t \) are aggregate consumption, investment, and income and \( u_{C_t} \) and \( u_{I_t} \) are disturbance terms. The first relationship is a conventional consumption function. The second relates investment to the change of output from the previous year. (This is known as an ‘accelerator’ model.) The third is an income identity. What can be said about the identification of the relationships in the model?

**Answer:**

The restriction on the coefficients of \( Y_t \) and \( Y_{t-1} \) in the investment equation complicates matters. A simple way of handling it is to define:

\[ \Delta Y_t = Y_t - Y_{t-1} \]

and to rewrite the investment equation as:

\[ I_t = \alpha_1 + \alpha_2 \Delta Y_t + u_{I_t}. \]

We now have four endogenous variables and four equations, and one exogenous variable. The consumption and investment equations are exactly identified. We would fit them using \( Y_{t-1} \) as an instrument for \( Y_t \) and \( \Delta Y_t \), respectively. The other two equations are identities and do not need to be fitted.

11.5 **Answers to the additional exercises**

A11.1 The linear regression indicates that expenditure on food increases by $0.032 billion for every extra $ billion of disposable personal income (in other words, by 3.2 cents out of the marginal dollar), that it increases by $0.403 billion for every point increase in the price index, and that it increases by $0.001 billion for every additional thousand population. The income coefficient is significant at the 1 per cent level (ignoring problems to be discussed in Chapter 12). The positive price coefficient makes no sense (remember that the dependent variable is measured in real terms). The intercept has no plausible interpretation.

The logarithmic regression indicates that the income elasticity is 0.59 and highly significant, and the price elasticity is \(-0.12\), not significant. The negative elasticity for population is not plausible. One would expect expenditure on food to increase in line with population, controlling for other factors, and hence, as a first approximation, the elasticity should be equal to 1. However, an increase in population, keeping income constant, would lead to a reduction in income per capita and hence to a negative income effect. Given that the income elasticity is less than 1, one would still expect a positive elasticity overall for population. At least the estimate is not significantly different from zero. In view of the high correlation, 0.995, between \( LGDPI \) and \( LGPOP \), the negative estimate may well be a result of multicollinearity.
11.5. Answers to the additional exercises

A11.2

<table>
<thead>
<tr>
<th>OLS logarithmic regressions</th>
<th>LGDPI</th>
<th>LGP</th>
<th>LGPOP</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cof.</strong></td>
<td><strong>s.e.</strong></td>
<td><strong>cof.</strong></td>
<td><strong>s.e.</strong></td>
<td><strong>cof.</strong></td>
</tr>
<tr>
<td>ADM</td>
<td>−1.43</td>
<td>0.20</td>
<td>−0.28</td>
<td>0.10</td>
</tr>
<tr>
<td>BOOK</td>
<td>−0.29</td>
<td>0.28</td>
<td>−1.18</td>
<td>0.21</td>
</tr>
<tr>
<td>BUSI</td>
<td>0.36</td>
<td>0.19</td>
<td>−0.11</td>
<td>0.27</td>
</tr>
<tr>
<td>CLOT</td>
<td>0.71</td>
<td>0.10</td>
<td>−0.70</td>
<td>0.05</td>
</tr>
<tr>
<td>DENT</td>
<td>1.23</td>
<td>0.14</td>
<td>−0.95</td>
<td>0.09</td>
</tr>
<tr>
<td>DOC</td>
<td>0.97</td>
<td>0.14</td>
<td>0.26</td>
<td>0.13</td>
</tr>
<tr>
<td>FLOW</td>
<td>0.46</td>
<td>0.32</td>
<td>0.16</td>
<td>0.33</td>
</tr>
<tr>
<td>FOOD</td>
<td>0.59</td>
<td>0.08</td>
<td>−0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>FURN</td>
<td>0.36</td>
<td>0.28</td>
<td>−0.48</td>
<td>0.26</td>
</tr>
<tr>
<td>GAS</td>
<td>1.27</td>
<td>0.24</td>
<td>−0.24</td>
<td>0.06</td>
</tr>
<tr>
<td>GASO</td>
<td>1.46</td>
<td>0.16</td>
<td>−0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>HOUS</td>
<td>0.91</td>
<td>0.08</td>
<td>−0.54</td>
<td>0.06</td>
</tr>
<tr>
<td>LEGL</td>
<td>1.17</td>
<td>0.16</td>
<td>−0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>MAGS</td>
<td>1.05</td>
<td>0.22</td>
<td>−0.73</td>
<td>0.44</td>
</tr>
<tr>
<td>MASS</td>
<td>−1.92</td>
<td>0.22</td>
<td>−0.57</td>
<td>0.14</td>
</tr>
<tr>
<td>OPHT</td>
<td>0.30</td>
<td>0.45</td>
<td>0.28</td>
<td>0.59</td>
</tr>
<tr>
<td>RELG</td>
<td>0.56</td>
<td>0.13</td>
<td>−0.99</td>
<td>0.23</td>
</tr>
<tr>
<td>TELE</td>
<td>0.91</td>
<td>0.13</td>
<td>−0.61</td>
<td>0.11</td>
</tr>
<tr>
<td>TOB</td>
<td>0.54</td>
<td>0.17</td>
<td>−0.42</td>
<td>0.04</td>
</tr>
<tr>
<td>TOYS</td>
<td>0.59</td>
<td>0.10</td>
<td>−0.54</td>
<td>0.06</td>
</tr>
</tbody>
</table>

The price elasticities mostly lie in the range 0 to −1, as they should, and therefore seem plausible. However the very high correlation between income and population, 0.995, has given rise to a problem of multicollinearity and as a consequence the estimates of their elasticities are very erratic. Some of the income elasticities look plausible, but that may be pure chance, for many are unrealistically high, or negative when obviously they should be positive. The population elasticities are even less convincing.

<table>
<thead>
<tr>
<th>Correlations between prices, income and population</th>
<th>LGP, LGDPI</th>
<th>LGP, LGPOP</th>
<th>LGP, LGDPI</th>
<th>LGP, LGPOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>0.61</td>
<td>0.61</td>
<td>GASO</td>
<td>0.05</td>
</tr>
<tr>
<td>BOOK</td>
<td>0.88</td>
<td>0.87</td>
<td>HOUS</td>
<td>0.49</td>
</tr>
<tr>
<td>BUSI</td>
<td>0.98</td>
<td>0.97</td>
<td>LEGL</td>
<td>0.99</td>
</tr>
<tr>
<td>CLOT</td>
<td>−0.94</td>
<td>−0.96</td>
<td>MAGS</td>
<td>0.99</td>
</tr>
<tr>
<td>DENT</td>
<td>0.94</td>
<td>0.96</td>
<td>MASS</td>
<td>0.90</td>
</tr>
<tr>
<td>DOC</td>
<td>0.98</td>
<td>0.98</td>
<td>OPHT</td>
<td>−0.68</td>
</tr>
<tr>
<td>FLOW</td>
<td>−0.93</td>
<td>−0.95</td>
<td>RELG</td>
<td>0.92</td>
</tr>
<tr>
<td>FOOD</td>
<td>−0.60</td>
<td>−0.64</td>
<td>TELE</td>
<td>−0.98</td>
</tr>
<tr>
<td>FURN</td>
<td>−0.95</td>
<td>−0.97</td>
<td>TOB</td>
<td>0.83</td>
</tr>
<tr>
<td>GAS</td>
<td>0.77</td>
<td>0.76</td>
<td>TOYS</td>
<td>−0.97</td>
</tr>
</tbody>
</table>
11. Models using time series data

A11.3 The regression indicates that the income elasticity is 0.40 and the price elasticity 0.21, the former very highly significant, the latter significant at the 1 per cent level using a one-sided test. If the specification is:

$$\log \frac{FOOD}{POP} = \beta_1 + \beta_2 \log \frac{DPI}{POP} + \beta_3 \log PRELFOOD + u$$

it may be rewritten:

$$\log FOOD = \beta_1 + \beta_2 \log DPI + \beta_3 \log PRELFOOD + (1 - \beta_2) \log POP + u.$$ 

This is a restricted form of the specification in Exercise A11.2:

$$\log FOOD = \beta_1 + \beta_2 \log DPI + \beta_3 \log PRELFOOD + \beta_4 \log POP + u$$

with $\beta_4 = 1 - \beta_2$. We can test the restriction by comparing $RSS$ for the two regressions:

$$F(1, 41) = \frac{(0.023232 - 0.016936)/1}{0.016936/41} = 15.24.$$

The critical value of $F(1, 40)$ at the 0.1 per cent level is 12.61. The critical value for $F(1, 41)$ must be slightly lower. Thus we reject the restriction. Since the restricted version is misspecified, our interpretation of the coefficients of this regression and the $t$ tests are invalidated.

A11.4 Given that the critical values of $F(1, 41)$ at the 5 and 1 per cent levels are 4.08 and 7.31 respectively, the results of the $F$ test may be summarised as follows:

- Restriction not rejected: CLOT, DENT, DOC, FURN, HOUS.
- Restriction rejected at the 5 per cent level: MAGS.
- Restriction rejected at the 1 per cent level: ADM, BOOK, BUSI, FLOW, FOOD, GAS, GASO, LEGL, MASS, OPHT, RELG, TELE, TOB, TOYS.

However, for reasons that will become apparent in the next chapter, these findings must be regarded as provisional.
## A11.5 Answers to the additional exercises

### Tests of a restriction

<table>
<thead>
<tr>
<th>Category</th>
<th>$RSS_U$</th>
<th>$RSS_R$</th>
<th>$F$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>0.125375</td>
<td>0.480709</td>
<td>116.20</td>
<td>10.78</td>
</tr>
<tr>
<td>BOOK</td>
<td>0.223664</td>
<td>0.461853</td>
<td>43.66</td>
<td>6.61</td>
</tr>
<tr>
<td>BUSI</td>
<td>0.084516</td>
<td>0.167580</td>
<td>40.30</td>
<td>6.35</td>
</tr>
<tr>
<td>CLOT</td>
<td>0.021326</td>
<td>0.021454</td>
<td>0.25</td>
<td>-0.50</td>
</tr>
<tr>
<td>DENT</td>
<td>0.033275</td>
<td>0.034481</td>
<td>1.49</td>
<td>1.22</td>
</tr>
<tr>
<td>DOC</td>
<td>0.068759</td>
<td>0.069726</td>
<td>0.58</td>
<td>-0.76</td>
</tr>
<tr>
<td>FLOW</td>
<td>0.220256</td>
<td>0.262910</td>
<td>7.94</td>
<td>2.82</td>
</tr>
<tr>
<td>FOOD</td>
<td>0.016936</td>
<td>0.023232</td>
<td>15.24</td>
<td>-3.90</td>
</tr>
<tr>
<td>FURN</td>
<td>0.157153</td>
<td>0.162677</td>
<td>1.44</td>
<td>1.20</td>
</tr>
<tr>
<td>GAS</td>
<td>0.185578</td>
<td>0.300890</td>
<td>25.48</td>
<td>-5.05</td>
</tr>
<tr>
<td>GASSO</td>
<td>0.078334</td>
<td>0.139278</td>
<td>31.90</td>
<td>-5.65</td>
</tr>
<tr>
<td>HOUSE</td>
<td>0.011270</td>
<td>0.012106</td>
<td>3.04</td>
<td>1.74</td>
</tr>
<tr>
<td>LEGL</td>
<td>0.082628</td>
<td>0.102698</td>
<td>9.96</td>
<td>-3.16</td>
</tr>
<tr>
<td>MAGS</td>
<td>0.096620</td>
<td>0.106906</td>
<td>4.36</td>
<td>-2.09</td>
</tr>
<tr>
<td>MASS</td>
<td>0.143775</td>
<td>0.330813</td>
<td>53.34</td>
<td>7.30</td>
</tr>
<tr>
<td>OPHT</td>
<td>0.663413</td>
<td>0.822672</td>
<td>9.84</td>
<td>3.14</td>
</tr>
<tr>
<td>RELG</td>
<td>0.053785</td>
<td>0.135532</td>
<td>62.32</td>
<td>7.89</td>
</tr>
<tr>
<td>TELE</td>
<td>0.054519</td>
<td>0.080728</td>
<td>19.71</td>
<td>4.44</td>
</tr>
<tr>
<td>TOB</td>
<td>0.062452</td>
<td>0.087652</td>
<td>16.54</td>
<td>-4.07</td>
</tr>
<tr>
<td>TOYS</td>
<td>0.031269</td>
<td>0.071656</td>
<td>52.96</td>
<td>7.28</td>
</tr>
</tbody>
</table>

### A11.5 If the specification is:

\[
\log \frac{FOOD}{POP} = \beta_1 + \beta_2 \log \frac{DPI}{POP} + \beta_3 \log PREGIDO + \gamma_1 POP + u
\]

it may be rewritten:

\[
\log FOOD = \beta_1 + \beta_2 \log DPI + \beta_3 \log PREGIDO + (1 - \beta_2 + \gamma_1) \log POP + u.
\]

This is equivalent to the specification in Exercise A11.1:

\[
\log FOOD = \beta_1 + \beta_2 \log DPI + \beta_3 \log PREGIDO + \beta_4 \log POP + u
\]

with \( \beta_4 = 1 - \beta_2 + \gamma_1 \). Note that this is not a restriction. (1) – (3) are just different ways of writing the unrestricted model.

A t test of \( H_0 : \gamma_1 = 0 \) is equivalent to a t test of \( H_0 : \beta_4 = 1 - \beta_2 \), that is, that the restriction in Exercise A11.3 is valid. The t statistic for LGPOP in the regression is -3.90, and hence again we reject the restriction. Note that the test is equivalent to the F test. -3.90 is the square root of 15.24, the F statistic, and it can be shown that the critical value of t is the square root of the critical value of F.

### A11.6 The t statistics for all the categories of expenditure are supplied in the table in the answer to Exercise A11.4. Of course they are equal to the square root of the F statistic, and their critical values are the square roots of the critical values of F, so the conclusions are identical and, like those of the F test, should be treated as provisional.
A11.7 Show that $\beta_2 + \beta_3$ and $(\beta_4 + \beta_5)$ are theoretically the long-run (equilibrium) income and price elasticities.

In equilibrium, $\overline{LGCAT} = \overline{LGDPI}$, $LGDPI = LGDPI(-1) = \overline{LGDPI}$ and $LGPRCAT = LGPRCAT(-1) = \overline{LGPRCAT}$. Hence, ignoring the transient effect of the disturbance term:

$$LGCAT = \beta_1 + \beta_2 \overline{LGDPI} + \beta_3 \overline{LGDPI} + \beta_4 \overline{LGPRCAT} + \beta_5 \overline{LGPRCAT} = \beta_1 + (\beta_2 + \beta_3) \overline{LGDPI} + (\beta_4 + \beta_5) \overline{LGPRCAT}.$$

Thus the long-run equilibrium income and price elasticities are $\theta = \beta_2 + \beta_3$ and $\phi = \beta_4 + \beta_5$, respectively.

Reparameterise the model and fit it to obtain direct estimates of these long-run elasticities and their standard errors.

We will reparameterise the model to obtain direct estimates of $\theta$ and $\phi$ and their standard errors. Write $\beta_3 = \theta - \beta_2$ and $\phi = \beta_4 + \beta_5$ and substitute for $\beta_3$ and $\beta_5$ in the model. We obtain:

$$LGCAT = \beta_1 + \beta_2 \overline{LGDPI} + (\theta - \beta_2) \overline{LGDPI}(-1) + \beta_4 \overline{LGPRCAT}(-1) + u$$
$$= \beta_1 + \beta_2 \overline{LGDPI} + \theta \overline{LGDPI}(-1) + \beta_4 \overline{LGPRCAT}(-1) + u$$
$$+ \beta_4 (LGPRCAT - LGPRCAT(-1)) + \phi LGPRCAT(-1) + u$$
$$= \beta_1 + \beta_2 DLGDPI + \theta LGDPI(-1) + \beta_4 DLGPRCAT + \phi LGPRCAT(-1) + u$$

where $DLGDPI = LGDPI - LGDPI(-1)$ and $DLGPRCAT = LGPRCAT - LGPRCAT(-1)$.

The output for $HOUS$ is shown below. $DLGPRCAT$ has been abbreviated as $DLGP$.

<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>0.020785</td>
<td>0.144497</td>
<td>0.143844</td>
<td>0.8864</td>
</tr>
<tr>
<td>DLGDPI</td>
<td>0.329571</td>
<td>0.150397</td>
<td>2.191340</td>
<td>0.0345</td>
</tr>
<tr>
<td>LGDPI(-1)</td>
<td>1.013147</td>
<td>0.006815</td>
<td>148.6735</td>
<td>0.0000</td>
</tr>
<tr>
<td>DLGP</td>
<td>-0.088813</td>
<td>0.165651</td>
<td>-0.536144</td>
<td>0.5949</td>
</tr>
<tr>
<td>LGPRHOUS(-1)</td>
<td>-0.447176</td>
<td>0.035927</td>
<td>-12.44689</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Confirm that the estimates are equal to the sum of the individual short-run elasticities found in Exercise 11.9.

The estimates of the long-run income and price elasticities are 1.01 and $-0.45$, respectively. The output below is for the model in its original form, where the

Dependent Variable: LGHOUS
Method: Least Squares
Sample(adjusted): 1960 2003
Included observations: 44 after adjusting endpoints

R-squared          0.999039  Mean dependent var 6.379059
Adjusted R-squared 0.998940  S.D. dependent var 0.421861
S.E. of regression 148.6735
Akaike info criter-5.631127
Schwarz criterion -5.428379
Log likelihood     128.8848  F-statistic 10131.80
Prob(F-statistic)  0.000000
coefficients are all short-run elasticities. It may be seen that, for both income and price, the sum of the estimates of the short-run elasticities is indeed equal to the estimate of the long-run elasticity in the reparameterised specification.

<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>0.020785</td>
<td>0.144497</td>
<td>0.143844</td>
<td>0.8864</td>
</tr>
<tr>
<td>LGDPI</td>
<td>0.329571</td>
<td>0.150397</td>
<td>2.191340</td>
<td>0.0345</td>
</tr>
<tr>
<td>LGDPI(-1)</td>
<td>0.683575</td>
<td>0.147111</td>
<td>4.646648</td>
<td>0.0000</td>
</tr>
<tr>
<td>LGPRHOU5</td>
<td>-0.088813</td>
<td>0.165651</td>
<td>-0.536144</td>
<td>0.5949</td>
</tr>
<tr>
<td>LGPRHOU5(-1)</td>
<td>-0.358363</td>
<td>0.165782</td>
<td>-2.161660</td>
<td>0.0368</td>
</tr>
</tbody>
</table>

R-squared 0.999039 Mean dependent var 6.379059
Adjusted R-squared 0.998940 S.D. dependent var 0.421861
S.E. of regression 0.013735 Akaike info criter -5.631127
Sum squared resid 0.007357 Schwarz criterion -5.428379
Log likelihood 128.8848 F-statistic 10131.80
Durbin-Watson stat 0.536957 Prob(F-statistic) 0.000000

Compare the standard errors with those found in Exercise 11.9 and state your conclusions.

The standard errors of the long-run elasticities in the reparameterised version are much smaller than those of the short-run elasticities in the original specification, and the t statistics accordingly much greater. Our conclusion is that it is possible to obtain relatively precise estimates of the long-run impact of income and price, even though multicollinearity prevents us from deriving precise short-run estimates.

A11.8 Show how this model may be derived from the demand function and the adaptive expectations process.

The adaptive expectations process may be rewritten:

\[ i_{t+1}^e = \lambda i_t + (1 - \lambda) i_t^e. \]

Substituting this into (1), one obtains:

\[ B_t = \beta_1 + \beta_2 i_t + \beta_2 (1 - \lambda) i_t^e + u_t. \]

We note that if we lag (1) by one time period:

\[ B_{t-1} = \beta_1 + \beta_2 i_t^e + u_{t-1}. \]

Hence:

\[ \beta_2 i_t^e = B_{t-1} - \beta_1 - u_{t-1}. \]

Substituting this into the second equation above, one has:

\[ B_t = \beta_1 \lambda + \beta_2 \lambda i_t + (1 - \lambda) B_{t-1} + u_t - (1 - \lambda) u_{t-1}. \]
11. Models using time series data

This is equation (3) in the question, with $\gamma_1 = \beta_1 \lambda$, $\gamma_2 = \beta_2 \lambda$, $\gamma_3 = 1 - \lambda$, and $v_t = u_t - (1 - \lambda)u_{t-1}$.

**Explain why inconsistent estimates of the parameters will be obtained if equation (3) is fitted using ordinary least squares (OLS).** (A mathematical proof is not required. Do not attempt to derive expressions for the bias.)

In equation (3), the regressor $B_{t-1}$ is partly determined by $u_{t-1}$. The disturbance term $v_t$ also has a component $u_{t-1}$. Hence the requirement that the regressors and the disturbance term be distributed independently of each other is violated. The violation will lead to inconsistent estimates because the regressor and the disturbance term are contemporaneously correlated.

**Describe a method for fitting the model that would yield consistent estimates.**

If the first equation in this exercise is true for time period $t + 1$, it is true for time period $t$:

$$i_t^* = \lambda i_{t-1} + (1 - \lambda)i_{t-1}^*.$$

Substituting into the second equation in (a), we now have:

$$B_t = \beta_1 + \beta_2 \lambda i_t + \beta_2 \lambda(1 - \lambda)i_{t-1} + (1 - \lambda)^2i_{t-1}^* + u_t.$$

Continuing to lag and substitute, we have:

$$B_t = \beta_1 + \beta_2 \lambda i_t + \beta_2 \lambda(1 - \lambda)i_{t-1} + \cdots + \beta_2 \lambda(1 - \lambda)^s i_{t-s+1} + u_t.$$

For $s$ large enough, $(1 - \lambda)^s$ will be so small that we can drop the unobservable term $i_{t-s+1}^*$ with negligible omitted variable bias. The disturbance term is distributed independently of the regressors and hence we obtain consistent estimates of the parameters. The model should be fitted using a nonlinear estimation technique that takes account of the restrictions implicit in the specification.

**Suppose that $u_t$ were subject to the first-order autoregressive process:**

$$u_t = \rho u_{t-1} + \varepsilon_t$$

**where $\varepsilon_t$ is not subject to autocorrelation. How would this affect your answer to the second part of this question?**

$v_t$ is now given by:

$$v_t = u_t - (1 - \lambda)u_{t-1} = \rho u_{t-1} + \varepsilon_t - (1 - \lambda)u_{t-1} = \varepsilon_t - (1 - \rho - \lambda)u_{t-1}.$$

Since $\rho$ and $\lambda$ may reasonably be assumed to lie between 0 and 1, it is possible that their sum is approximately equal to 1, in which case $v_t$ is approximately equal to the innovation $\varepsilon_t$. If this is the case, there would be no violation of the regression assumption described in the second part of this question and one could use OLS to fit (3) after all.

**Suppose that the true relationship was actually:**

$$B_t = \beta_1 + \beta_2 i_t + u_t \quad (1*)$$

**with $u_t$ not subject to autocorrelation, and the model is fitted by regressing $B_t$ on $i_t$ and $B_{t-1}$, as in equation (3), using OLS. How would this affect the regression results?**
The estimators of the coefficients will be inefficient in that $B_{t-1}$ is a redundant variable. The inclusion of $B_{t-1}$ will also give rise to finite sample bias that would disappear in large samples.

How plausible do you think an adaptive expectations process is for modelling expectations in a bond market?

The adaptive expectations model is implausible since the expectations process would change as soon as those traders taking advantage of their knowledge of it started earning profits.

A11.9 The regression indicates that the short-run income, price, and population elasticities for expenditure on food are 0.14, −0.10, and −0.05, respectively, and that the speed of adjustment is $(1 - 0.73) = 0.27$. Dividing by 0.27, the long-run elasticities are 0.52, −0.37, and −0.19, respectively. The income and price elasticities seem plausible. The negative population elasticity makes no sense, but it is small and insignificant. The estimates of the short-run income and price elasticities are likewise not significant, but this is not surprising given that the point estimates are so small.

A11.10 The table gives the result of the specification with a lagged dependent variable for all the categories of expenditure.

<table>
<thead>
<tr>
<th></th>
<th>LGDPI</th>
<th>LGP</th>
<th>LGPOP</th>
<th>LGCAT(−1)</th>
<th>Long-run effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>−0.38</td>
<td>0.18</td>
<td>−0.10</td>
<td>0.06</td>
<td>−0.18</td>
</tr>
<tr>
<td>BOOK</td>
<td>−0.36</td>
<td>0.20</td>
<td>−0.21</td>
<td>0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>BUSI</td>
<td>0.10</td>
<td>0.13</td>
<td>0.03</td>
<td>0.18</td>
<td>0.11</td>
</tr>
<tr>
<td>CLOT</td>
<td>0.44</td>
<td>0.10</td>
<td>−0.40</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>DENT</td>
<td>0.71</td>
<td>0.18</td>
<td>−0.46</td>
<td>0.16</td>
<td>1.34</td>
</tr>
<tr>
<td>DOC</td>
<td>0.23</td>
<td>0.14</td>
<td>−0.11</td>
<td>0.10</td>
<td>1.04</td>
</tr>
<tr>
<td>FLOW</td>
<td>0.20</td>
<td>0.24</td>
<td>−0.31</td>
<td>0.27</td>
<td>0.81</td>
</tr>
<tr>
<td>FOOD</td>
<td>0.14</td>
<td>0.09</td>
<td>−0.10</td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>FURN</td>
<td>0.07</td>
<td>0.22</td>
<td>−0.07</td>
<td>0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>GAS</td>
<td>0.10</td>
<td>0.17</td>
<td>−0.06</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>GASO</td>
<td>0.32</td>
<td>0.11</td>
<td>−0.10</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>HOUS</td>
<td>0.30</td>
<td>0.05</td>
<td>−0.09</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>LEGL</td>
<td>0.40</td>
<td>0.14</td>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>MAGS</td>
<td>0.57</td>
<td>0.21</td>
<td>−0.48</td>
<td>0.37</td>
<td>0.12</td>
</tr>
<tr>
<td>MASS</td>
<td>−0.28</td>
<td>0.29</td>
<td>−0.23</td>
<td>0.11</td>
<td>−1.14</td>
</tr>
<tr>
<td>OPHT</td>
<td>0.30</td>
<td>0.24</td>
<td>−0.28</td>
<td>0.33</td>
<td>2.48</td>
</tr>
<tr>
<td>RELG</td>
<td>0.34</td>
<td>0.09</td>
<td>−0.71</td>
<td>0.17</td>
<td>0.09</td>
</tr>
<tr>
<td>TELE</td>
<td>0.15</td>
<td>0.14</td>
<td>0.00</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>TOB</td>
<td>0.12</td>
<td>0.14</td>
<td>−0.12</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>TOYS</td>
<td>0.31</td>
<td>0.11</td>
<td>−0.27</td>
<td>0.08</td>
<td>0.12</td>
</tr>
</tbody>
</table>
11. Models using time series data

A11.11 In his classic study Distributed Lags and Investment Analysis (1954), Koyck investigated the relationship between investment in railcars and the volume of freight carried on the US railroads using data for the period 1884–1939. Assuming that the desired stock of railcars in year \( t \) depended on the volume of freight in year \( t - 1 \) and year \( t - 2 \) and a time trend, and assuming that investment in railcars was subject to a partial adjustment process, he fitted the following regression equation using OLS (standard errors and constant term not reported):

\[
\hat{I}_t = 0.077F_{t-1} + 0.017F_{t-2} - 0.0033t - 0.110K_{t-1} \quad R^2 = 0.85
\]

where \( I_t = K_t - K_{t-1} \) is investment in railcars in year \( t \) (thousands), \( K_t \) is the stock of railcars at the end of year \( t \) (thousands), and \( F_t \) is the volume of freight handled in year \( t \) (ton-miles).

Provide an interpretation of the equation and describe the dynamic process implied by it. (Note: It is best to substitute \( K_t - K_{t-1} \) for \( I_t \) in the regression and treat it as a dynamic relationship determining \( K_t \)).

Given the information in the question, the model may be written:

\[
K_t^* = \beta_1 + \beta_2F_{t-1} + \beta_3F_{t-2} + \beta_4t + u_t
\]

\[
K_t - K_{t-1} = I_t = \lambda(K_t^* - K_{t-1}).
\]

Hence:

\[
I_t = \lambda\beta_1 + \lambda\beta_2F_{t-1} + \lambda\beta_3F_{t-2} + \lambda\beta_4t - \lambda K_{t-1} + \lambda u_t.
\]

From the fitted equation:

\[
\hat{\lambda} = 0.110
\]

\[
\hat{\beta}_2 = \frac{0.077}{0.110} = 0.70
\]

\[
\hat{\beta}_3 = \frac{0.017}{0.110} = 0.15
\]

\[
\hat{\beta}_4 = \frac{-0.0033}{0.110} = -0.030.
\]

Hence the short-run effect of an increase of 1 million ton-miles of freight is to increase investment in railcars by 7,000 one year later and 1,500 two years later. It does not make much sense to talk of a short-run effect of a time trend.

In the long-run equilibrium, neglecting the effects of the disturbance term, \( K_t \) and \( K_t^* \) are both equal to the equilibrium value \( \bar{K} \) and \( F_{t-1} \) and \( F_{t-2} \) are both equal to their equilibrium value \( \bar{F} \). Hence, using the first equation:

\[
\bar{K} = \beta_1 + (\beta_2 + \beta_3)\bar{F} + \beta_4t.
\]

Thus an increase of one million ton-miles of freight will increase the stock of railcars by 940 and the time trend will be responsible for a secular decline of 33 railcars per year.
One researcher asserts that consistent estimates will be obtained if (2) is fitted using OLS and (1) is fitted using IV, with $Y_{t-1}$ as an instrument for $X_t$. Determine whether this is true.

(2) may indeed be fitted using OLS. Strictly speaking, there may be an element of bias in finite samples because of noncontemporaneous correlation between $u_t$ and future values of $Y_{t-1}$.

We could indeed use $Y_{t-1}$ as an instrument for $X_t$ in (1) because $Y_{t-1}$ is a determinant of $X_t$ but is not (contemporaneously) correlated with $u_t$.

The other researcher asserts that consistent estimates will be obtained if both (1) and (2) are fitted using OLS, and that the estimate of $\beta_2$ will be more efficient than that obtained using IV. Determine whether this is true.

This assertion is also correct. $X_t$ is not correlated with $u_t$, and OLS estimators are more efficient than IV estimators when both are consistent. Strictly speaking, there may be an element of bias in finite samples because of noncontemporaneous correlation between $u_t$ and future values of $X_t$. 
11. Models using time series data
Chapter 12

Properties of regression models with time series data

12.1 Overview

This chapter begins with a statement of the regression model assumptions for regressions using time series data, paying particular attention to the assumption that the disturbance term in any time period be distributed independently of the regressors in all time periods. There follows a general discussion of autocorrelation: the meaning of the term, the reasons why the disturbance term may be subject to it, and the consequences of it for OLS estimators. The chapter continues by presenting the Durbin–Watson test for AR(1) autocorrelation and showing how the problem may be eliminated. Next it is shown why OLS yields inconsistent estimates when the disturbance term is subject to autocorrelation and the regression model includes a lagged dependent variable as an explanatory variable. Then the chapter shows how the restrictions implicit in the AR(1) specification may be tested using the common factor test, and this leads to a more general discussion of how apparent autocorrelation may be caused by model misspecification. This in turn leads to a general discussion of the issues involved in model selection and, in particular, to the general-to-specific methodology.

12.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain the concept of autocorrelation and the difference between positive and negative autocorrelation
- describe how the problem of autocorrelation may arise
- describe the consequences of autocorrelation for OLS estimators, their standard errors, and t and F tests, and how the consequences change if the model includes a lagged dependent variable
- perform the Breusch–Godfrey and Durbin–Watson d tests for autocorrelation
- explain how the problem of AR(1) autocorrelation may be eliminated
- describe the restrictions implicit in the AR(1) specification
12. Properties of regression models with time series data

- perform the common factor test
- explain how apparent autocorrelation may arise as a consequence of the omission of an important variable or the mathematical misspecification of the regression model
- demonstrate that the static, AR(1), and ADL(1,0) specifications are special cases of the ADL(1,1) model
- explain the principles of the general-to-specific approach to model selection and the defects of the specific-to-general approach.

12.3 Additional exercises

A12.1 The output shows the result of a logarithmic regression of expenditure on food on income, relative price, and population, using an AR(1) specification. Compare the results with those in Exercise A11.1.

```
Dependent Variable: LGFOOD
Method: Least Squares
Sample(adjusted): 1960 2003
Included observations: 44 after adjusting endpoints
Convergence achieved after 14 iterations

Variable      Coefficient  Std. Error  t-Statistic  Prob.
-------------  -----------  -----------  -----------  ----
    C          2.945983    3.943913    0.746969    0.4596
 LGDPI        0.469216    0.118230    3.968687    0.0003
 LGPRFOOD    -0.361862    0.122069   -2.964413    0.0052
 LGPOP        0.072193    0.379563    0.190200    0.8501
  AR(1)       0.880631    0.092512    9.519085    0.0000

R-squared 0.996695  Mean dependent var 6.030691
Adjusted R-squared 0.996356  S.D. dependent var 0.216227
S.E. of regression 0.013053  Akaike info criter-5.732970
Sum squared resid 0.006645  Schwarz criterion -5.530221
Log likelihood 131.1253  F-statistic 2940.208
Durbin--Watson stat 1.556480  Prob(F-statistic) 0.000000

Inverted AR Roots .88
```

A12.2 Perform Breusch–Godfrey and Durbin–Watson tests for autocorrelation for the logarithmic regression in Exercise A11.2. If you reject the null hypothesis of no autocorrelation, run the regression again using an AR(1) specification, and compare the results with those in Exercise A11.2.

A12.3 Perform an OLS ADL(1,1) logarithmic regression of expenditure on your category on current income, price, and population and lagged expenditure, income, price, and population. Use the results to perform a common factor test of the validity of the AR(1) specification in Exercise A12.2.
12.3. Additional exercises

A12.4 A researcher has annual data on \( \textit{LIFE} \), aggregate consumer expenditure on life insurance, \( \textit{DPI} \), aggregate disposable personal income, and \( \textit{PRELLIFE} \), a price index for the cost of life insurance relative to general inflation, for the United States for the period 1959–1994. \( \textit{LIFE} \) and \( \textit{DPI} \) are measured in US$ billion. \( \textit{PRELLIFE} \) is an index number series with 1992 = 100. She defines \( \textit{LGLIFE} \), \( \textit{LGDPI} \), and \( \textit{LGPRLIFE} \) as the natural logarithms of \( \textit{LIFE} \), \( \textit{DPI} \), and \( \textit{PRELLIFE} \), respectively. She fits the regressions shown in columns (1) – (4) of the table, each with \( \textit{LGLIFE} \) as the dependent variable. (Standard errors in parentheses; OLS = ordinary least squares; AR(1) is a specification appropriate when the disturbance term follows a first-order autoregressive process; \( B-G \) is the Breusch–Godfrey test statistic for AR(1) autocorrelation; \( d \) = Durbin–Watson \( d \) statistic; \( \hat{\rho} \) is the estimate of the autoregressive parameter in a first-order autoregressive process.)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \textit{LGDPI} )</td>
<td>1.37</td>
<td>1.41</td>
<td>0.42</td>
<td>0.28</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.25)</td>
<td>(0.60)</td>
<td>(0.17)</td>
<td></td>
</tr>
<tr>
<td>( \textit{LGPRLIFE} )</td>
<td>−0.67</td>
<td>−0.78</td>
<td>−0.59</td>
<td>−0.26</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.50)</td>
<td>(0.51)</td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td>( \textit{LGLIFE}(−1) )</td>
<td>—</td>
<td>—</td>
<td>0.82</td>
<td>0.79</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>( \textit{LGDPI}(−1) )</td>
<td>—</td>
<td>—</td>
<td>−0.15</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \textit{LGPRLIFE}(−1) )</td>
<td>—</td>
<td>—</td>
<td>0.38</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>−4.39</td>
<td>−4.20</td>
<td>−0.50</td>
<td>−0.51</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(0.88)</td>
<td>(1.69)</td>
<td>(0.72)</td>
<td>(0.70)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.958</td>
<td>0.985</td>
<td>0.986</td>
<td>0.986</td>
<td>0.984</td>
</tr>
<tr>
<td>( RSS )</td>
<td>0.2417</td>
<td>0.0799</td>
<td>0.0719</td>
<td>0.0732</td>
<td>0.0843</td>
</tr>
<tr>
<td>( B-G )</td>
<td>23.48</td>
<td>—</td>
<td>0.61</td>
<td>0.34</td>
<td>0.10</td>
</tr>
<tr>
<td>( d )</td>
<td>0.36</td>
<td>1.85</td>
<td>2.02</td>
<td>1.92</td>
<td>2.05</td>
</tr>
<tr>
<td>( \hat{\rho} )</td>
<td>—</td>
<td>0.82</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.11)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Discuss whether specification (1) is an adequate representation of the data.
- Discuss whether specification (3) is an adequate representation of the data.
- Discuss whether specification (2) is an adequate representation of the data.
- Discuss whether specification (4) is an adequate representation of the data.
- If you were presenting these results at a seminar, what would you say were your conclusions concerning the most appropriate of specifications (1) – (4)?
- At the seminar a commentator points out that in specification (4) neither \( \textit{LGDPI} \) nor \( \textit{LGPRLIFE} \) have significant coefficients and so these variables should be dropped. As it happens, the researcher has considered this specification, and the results are shown as specification (5) in the table. What would be your answer to the commentator?
A researcher has annual data on the yearly rate of change of the consumer price index, \( p \), and the yearly rate of change of the nominal money supply, \( m \), for a certain country for the 51-year period 1958–2008. He fits the following regressions, each with \( p \) as the dependent variable. The first four regressions are fitted using OLS. The fifth is fitted using a specification appropriate when the disturbance term is assumed to follow an AR(1) process. \( p(-1) \) indicates \( p \) lagged one year. \( m(-1) \), \( m(-2) \), and \( m(-3) \) indicate \( m \) lagged 1, 2, and 3 years, respectively.

1. explanatory variable \( m \).
2. explanatory variables \( m \), \( m(-1) \), \( m(-2) \), and \( m(-3) \).
3. explanatory variables \( m \), \( p(-1) \), and \( m(-1) \).
4. explanatory variables \( m \) and \( p(-1) \).
5. explanatory variable \( m \).

The results are shown in the table. Standard errors are shown in parentheses. RSS is the residual sum of squares. \( B – G \) is the Breusch–Godfrey test statistic for AR(1) autocorrelation. \( d \) is the Durbin–Watson \( d \) statistic.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>AR(1)</td>
</tr>
<tr>
<td>( m )</td>
<td>0.95</td>
<td>0.50</td>
<td>0.40</td>
<td>0.18</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.30)</td>
<td>(0.12)</td>
<td>(0.09)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>( m(-1) )</td>
<td>—</td>
<td>0.30</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(-2) )</td>
<td>—</td>
<td>—</td>
<td>−0.15</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(-3) )</td>
<td>—</td>
<td>0.30</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p(-1) )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(0.20)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>constant</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>RSS</td>
<td>0.0200</td>
<td>0.0150</td>
<td>0.0100</td>
<td>0.0120</td>
<td>0.0105</td>
</tr>
<tr>
<td>( B – G )</td>
<td>35.1</td>
<td>27.4</td>
<td>0.39</td>
<td>0.26</td>
<td>0.57</td>
</tr>
<tr>
<td>( d )</td>
<td>0.10</td>
<td>0.21</td>
<td>2.00</td>
<td>2.00</td>
<td>1.90</td>
</tr>
</tbody>
</table>

- Looking at all five regressions together, evaluate the adequacy of:
  - specification 1.
  - specification 2.
  - specification 3.
  - specification 4.
- Explain why specification 5 is a restricted version of one of the other specifications, stating the restriction, and explaining the objective of the manipulations that lead to specification 5.
- Perform a test of the restriction embodied in specification 5.
- Explain which would be your preferred specification.
A12.6 Derive the short-run (current year) and long-run (equilibrium) effect of \( m \) on \( p \) for each of the five specifications in Exercise A12.5, using the estimated coefficients.

A12.7 A researcher has annual data on aggregate consumer expenditure on taxis, \( TAXI \), and aggregate disposable personal income, \( DPI \), both measured in $ billion at 2000 constant prices, and a relative price index for taxis, \( P \), equal to 100 in 2000, for the United States for the period 1981–2005.

Defining \( LGTAXI \), \( LGDPI \), and \( LGP \) as the natural logarithms of \( TAXI \), \( DPI \), and \( P \), respectively, he fits regressions (1) – (4) shown in the table. OLS = ordinary least squares; AR(1) indicates that the equation was fitted using a specification appropriate for first-order autoregressive autocorrelation; \( \hat{\rho} \) is an estimate of the parameter in the AR(1) process; \( B–G \) is the Breusch–Godfrey statistic for AR(1) autocorrelation; \( d \) is the Durbin–Watson \( d \) statistic; standard errors are given in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( LGDPI )</td>
<td>2.06</td>
<td>1.28</td>
<td>2.28</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.84)</td>
<td>(0.05)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>( LGP )</td>
<td>—</td>
<td>—</td>
<td>-0.99</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.09)</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-12.75</td>
<td>-7.45</td>
<td>-9.58</td>
<td>-9.45</td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td>(5.89)</td>
<td>(0.40)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>( \hat{\rho} )</td>
<td>—</td>
<td>0.88</td>
<td>—</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( B–G )</td>
<td>17.84</td>
<td>—</td>
<td>1.47</td>
<td>—</td>
</tr>
<tr>
<td>( d )</td>
<td>0.31</td>
<td>1.40</td>
<td>1.46</td>
<td>1.88</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 12.1 shows the actual values of \( LGTAXI \) and the fitted values from regression (1). Figure 12.2 shows the residuals from regression (1) and the values of \( LGP \).

- Evaluate regression (1).
- Evaluate regression (2). Explain mathematically what assumptions were being made by the researcher when he used the AR(1) specification and why he hoped the results would be better than those obtained with regression (1).
- Evaluate regression (3).
- Evaluate regression (4). In particular, discuss the possible reasons for the differences in the standard errors in regressions (3) and (4).
- At a seminar one of the participants says that the researcher should consider adding lagged values of \( LGTAXI \), \( LGDPI \), and \( LGP \) to the specification. What would be your view?
A researcher has annual data on $I$, investment as a percentage of gross domestic product, and $r$, the real long-term rate of interest for a certain economy for the period 1981–2010. He regresses $I$ on $r$, (1) using ordinary least squares (OLS), (2) using an estimator appropriate for AR(1) residual autocorrelation, and (3) using OLS but adding $I(-1)$ and $r(-1)$ ($I$ and $r$ lagged one time period) as explanatory variables. The results are shown in columns (1), (2), and (3) of the table below. The residuals from regression (1) are shown in Figure 12.3.

He then obtains annual data on $g$, the rate of growth of gross domestic product of the economy, for the same period, and repeats the regressions, adding $g$ (and, where appropriate, $g(-1)$) to the specifications as an explanatory variable. The results are shown in columns (4), (5), and (6) of the table. $r$ and $g$ are measured as per cent per year. The data for $g$ are plotted in the figure.
Figure 12.3: Residuals from regression (1).

<table>
<thead>
<tr>
<th></th>
<th>OLS (1)</th>
<th>AR(1) (2)</th>
<th>OLS (3)</th>
<th>OLS (4)</th>
<th>AR(1) (5)</th>
<th>OLS (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>-0.87</td>
<td>-0.83</td>
<td>-0.87</td>
<td>-1.81</td>
<td>-1.88</td>
<td>-1.71</td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td>(1.05)</td>
<td>(1.08)</td>
<td>(0.49)</td>
<td>(0.50)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>$I(-1)$</td>
<td>—</td>
<td>—</td>
<td>0.37</td>
<td>—</td>
<td>—</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.16)</td>
<td></td>
<td></td>
<td>(0.18)</td>
</tr>
<tr>
<td>$r(-1)$</td>
<td>—</td>
<td>—</td>
<td>0.64</td>
<td>—</td>
<td>—</td>
<td>-0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.08)</td>
<td></td>
<td></td>
<td>(0.64)</td>
</tr>
<tr>
<td>$g$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.61</td>
<td>1.61</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.17)</td>
<td>(0.18)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>$g(-1)$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>—</td>
<td>0.37</td>
<td>—</td>
<td>—</td>
<td>-0.16</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.20)</td>
</tr>
<tr>
<td></td>
<td>(3.64)</td>
<td>(3.90)</td>
<td>(4.48)</td>
<td>(1.77)</td>
<td>(1.64)</td>
<td>(2.69)</td>
</tr>
<tr>
<td>$B-G$</td>
<td>4.42</td>
<td>—</td>
<td>4.24</td>
<td>0.70</td>
<td>—</td>
<td>0.98</td>
</tr>
<tr>
<td>$d$</td>
<td>0.99</td>
<td>1.36</td>
<td>1.33</td>
<td>2.30</td>
<td>2.05</td>
<td>2.09</td>
</tr>
<tr>
<td>RSS</td>
<td>120.5</td>
<td>103.9</td>
<td>103.5</td>
<td>27.4</td>
<td>26.8</td>
<td>23.5</td>
</tr>
</tbody>
</table>

Note: standard errors are given in parentheses. $\hat{\rho}$ is the estimate of the autocorrelation parameter in the AR(1) specification. $B-G$ is the Breusch–Godfrey statistic for AR(1) autocorrelation. $d$ is the Durbin–Watson $d$ statistic.

- Explain why the researcher was not satisfied with regression (1).
- Evaluate regression (2). Explain why the coefficients of $I(-1)$ and $r(-1)$ are not reported, despite the fact that they are part of the regression specification.
- Evaluate regression (3).
12. Properties of regression models with time series data

- Evaluate regression (4).
- Evaluate regression (5).
- Evaluate regression (6).
- Summarise your conclusions concerning the evaluation of the different regressions. Explain whether an examination of the figure supports your conclusions.

A12.9 In Exercise A11.5 you performed a test of a restriction. The result of this test will have been invalidated if you found that the specification was subject to autocorrelation. How should the test be performed, assuming the correct specification is ADL(1,1)?

A12.10 Given data on a univariate process:

\[ Y_t = \beta_1 + \beta_2 y_{t-1} + u_t \]

where \(|\beta_2| < 1\) and \(u_t\) is iid, the usual OLS estimators will be consistent but subject to finite-sample bias. How should the model be fitted if \(u_t\) is subject to an AR(1) process?

A12.11 Explain what is correct, incorrect, confused or incomplete in the following statements, giving a brief explanation if not correct.

- The disturbance term in a regression model is said to be autocorrelated if its values in a sample of observations are not distributed independently of each other.
- When the disturbance term is subject to autocorrelation, the ordinary least squares estimators are inefficient and inconsistent, but they are not biased, and the t tests are invalid.
- It is a common problem in time series models because it always occurs when the dependent variable is correlated with its previous values.
- If this is the case, it could be eliminated by including the lagged value of the dependent variable as an explanatory variable.
- However, if the model is correctly specified and the disturbance term satisfies the regression model assumptions, adding the lagged value of the dependent variable as an explanatory variable will have the opposite effect and cause the disturbance term to be autocorrelated.
- A second way of dealing with the problem of autocorrelation is to use an instrumental variable.
- If the autocorrelation is of the AR(1) type, randomising the order of the observations will cause the Breusch–Godfrey statistic to be near zero, and the Durbin–Watson statistic to be near 2, thereby eliminating the problem.
12.4 Answers to the starred exercises in the textbook

12.7 Prove that $\sigma_u^2$ is related to $\sigma_e^2$ as shown in (12.31), and show that weighting the first observation by $\sqrt{1 - \rho^2}$ eliminates the heteroskedasticity.

Answer:

(12.31) is:

$$\sigma_u^2 = \frac{1}{1 - \rho^2} \sigma_e^2$$

and it assumes the first order AR(1) process (12.26): $u_t = \rho u_{t-1} + \varepsilon_t$. From the AR(1) process, neglecting transitory effects, $\sigma_{u_t} = \sigma_{u_{t-1}} = \sigma_u$ and so:

$$\sigma_u^2 = \rho^2 \sigma_u^2 + \sigma_e^2 = \frac{1}{1 - \rho^2} \sigma_e^2.$$  
(Note that the covariance between $u_{t-1}$ and $\varepsilon_t$ is zero.) If the first observation is weighted by $\sqrt{1 - \rho^2}$, the variance of the disturbance term will be:

$$\left(\sqrt{1 - \rho^2}\right)^2 \sigma_u^2 = (1 - \rho^2) \frac{1}{1 - \rho^2} \sigma_e^2 = \sigma_e^2$$

and it will therefore be the same as in the other observations in the sample.

12.10 The table gives the results of three logarithmic regressions using the Cobb–Douglas data for $Y_t$, $K_t$, and $L_t$, index number series for real output, real capital input, and real labor input, respectively, for the manufacturing sector of the United States for the period 1899–1922, reproduced in Exercise 11.6 (method of estimation as indicated; standard errors in parentheses; $d =$ Durbin–Watson $d$ statistic; $B–G =$ Breusch–Godfrey test statistic for first-order autocorrelation):

<table>
<thead>
<tr>
<th></th>
<th>1: OLS</th>
<th>2: AR(1)</th>
<th>3: OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $K$</td>
<td>0.23</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>log $L$</td>
<td>0.81</td>
<td>0.86</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>log $Y(-1)$</td>
<td>—</td>
<td>—</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log $K(-1)$</td>
<td>—</td>
<td>—</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log $L(-1)$</td>
<td>—</td>
<td>—</td>
<td>-1.01</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-0.18</td>
<td>-0.35</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.51)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>—</td>
<td>0.19</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.96</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>RSS</td>
<td>0.0710</td>
<td>0.0697</td>
<td>0.0259</td>
</tr>
<tr>
<td>$d$</td>
<td>1.52</td>
<td>1.54</td>
<td>1.46</td>
</tr>
<tr>
<td>$B–G$</td>
<td>0.36</td>
<td>—</td>
<td>1.54</td>
</tr>
</tbody>
</table>
12. Properties of regression models with time series data

The first regression is that performed by Cobb and Douglas. The second fits the same specification, allowing for AR(1) autocorrelation. The third specification uses OLS with lagged variables. Evaluate the three regression specifications.

Answer:

For the first specification, the Breusch–Godfrey LM test for autocorrelation yields statistics of 0.36 (first order) and 1.39 (second order), both satisfactory. For the Durbin–Watson test, \( d_L \) and \( d_U \) are 1.19 and 1.55 at the 5 per cent level and 0.96 and 1.30 at the 1 per cent level, with 24 observations and two explanatory variables. Hence the specification appears more or less satisfactory. Fitting the model with an AR(1) specification makes very little difference, the estimate of \( \rho \) being low. However, when we fit the general ADL(1,1) model, neither of the first two specifications appears to be an acceptable simplification. The \( F \) statistic for dropping all the lagged variables is:

\[
F(3, 18) = \frac{(0.0710 - 0.0259)/3}{0.0259/18} = 10.45.
\]

The critical value of \( F(3, 18) \) at the 0.1 per cent level is 8.49. The common factor test statistic is:

\[
23 \log \frac{0.0657}{0.0259} = 22.77
\]

and the critical value of chi-squared with two degrees of freedom is 13.82 at the 0.1 per cent level. The Breusch–Godfrey statistic for first-order autocorrelation is 1.54.

We come to the conclusion that Cobb and Douglas, who actually fitted a restricted version of the first specification, imposing constant returns to scale, were a little fortunate to obtain the plausible results they did.

12.11 Derive the final equation in Box 12.2 from the first two equations in the box. What assumptions need to be made when fitting the model?

Answer:

This exercise overlaps Exercise 11.17. The first two equations in the box are:

\[
Y_t = \beta_1 + \beta_2 X_{t+1}^e + u_t
\]

\[
X_{t+1}^e - X_t^e = \lambda (X_t - X_t^e).
\]

We can rewrite the second equation as:

\[
X_{t+1}^e = \lambda X_t + (1 - \lambda) X_t^e.
\]

Substituting this into the first equation, we have:

\[
Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 (1 - \lambda) X_t^e + u_t.
\]

This includes the unobservable \( X_t^e \) on the right side. However, lagging the second equation, we have:

\[
X_t^e = \lambda X_{t-1} + (1 - \lambda) X_{t-1}^e.
\]

Hence:

\[
Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 \lambda (1 - \lambda) X_{t-1} + \beta_2 (1 - \lambda)^2 X_{t-1}^e + u_t.
\]
12.4. Answers to the starred exercises in the textbook

This includes the unobservable $X_{t-1}$ on the right side. However, continuing to lag and substitute, we have:

$$Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 \lambda (1 - \lambda) X_{t-1} + \cdots + \beta_2 \lambda (1 - \lambda)^s X_{t-s} + \beta_2 (1 - \lambda)^{s+1} X_{t-s} + u_t.$$ 

Provided that $s$ is large enough for $\beta_2 (1 - \lambda)^{s+1}$ to be very small, this may be fitted, omitting the unobservable final term, with negligible omitted variable bias. We would fit it with a nonlinear regression technique that respected the constraints implicit in the theoretical structure of the coefficients. The disturbance term is unaffected by the manipulations. Hence it is sufficient to assume that it is well-behaved in the original specification.

12.14 Using the 50 observations on two variables $Y$ and $X$ shown in the diagram below, an investigator runs the following five regressions (estimation method as indicated; standard errors in parentheses; all variables as logarithms in the logarithmic regressions; $d = \text{Durbin–Watson} d$ statistic; $B–G = \text{Breusch–Godfrey} \text{ test statistic}$):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Logarithmic</td>
<td>Linear</td>
<td>Logarithmic</td>
<td>Linear</td>
</tr>
<tr>
<td>$X$</td>
<td>0.16</td>
<td>0.03</td>
<td>2.39</td>
<td>2.39</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.05)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>$Y(-1)$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>$-0.11$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.15)</td>
</tr>
<tr>
<td>$X(-1)$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>$1.30$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.75)</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>—</td>
<td>1.16</td>
<td>—</td>
<td>$-0.14$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td></td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>$-21.88$</td>
<td>$-2.52$</td>
<td>$-11.00$</td>
<td>$-10.99$</td>
<td>$-12.15$</td>
</tr>
<tr>
<td></td>
<td>(3.17)</td>
<td>(8.03)</td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(1.67)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.858</td>
<td>0.974</td>
<td>0.993</td>
<td>0.993</td>
<td>0.993</td>
</tr>
<tr>
<td>$RSS$</td>
<td>7663</td>
<td>1366</td>
<td>1.011</td>
<td>0.993</td>
<td>0.946</td>
</tr>
<tr>
<td>$d$</td>
<td>0.26</td>
<td>2.75</td>
<td>2.17</td>
<td>1.86</td>
<td>21.95</td>
</tr>
<tr>
<td>$B–G$</td>
<td>39.54</td>
<td>—</td>
<td>0.85</td>
<td>—</td>
<td>1.03</td>
</tr>
</tbody>
</table>

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
12. Properties of regression models with time series data

Discuss each of the five regressions, explaining which is your preferred specification.

**Answer:**

The scatter diagram reveals that the relationship is nonlinear. If it is fitted with a linear regression, the residuals must be positive for the largest and smallest values of \( X \) and negative for the middle ones. As a consequence it is no surprise to find a high Breusch–Godfrey statistic, above 10.83, the critical value of \( \chi^2(1) \) at the 0.1% level, and a low Durbin–Watson statistic, below 1.32, the critical value at the 1 per cent level. Equally it is no surprise to find that an AR(1) specification does not yield satisfactory results, the Durbin–Watson statistic now indicating negative autocorrelation.

By contrast the logarithmic specification appears entirely satisfactory, with a Breusch–Godfrey statistic of 0.85 and a Durbin–Watson statistic of 1.82 (\( d_U \) is 1.59 at the 5 per cent level). Comparing it with the ADL(1,1) specification, the \( F \) statistic for dropping the lagged variables is:

\[
F(2, 46) = \frac{(1.084 - 1.020)/2}{1.020/46} = 1.44.
\]

The critical value of \( F(2, 40) \) at the 5 per cent level is 3.23. Hence we conclude that specification (3) is an acceptable simplification. Specifications (4) and (5) are inefficient, and this accounts for their larger standard errors.

12.15 Using the data on food in the Demand Functions data set, the following regressions were run, each with the logarithm of food as the dependent variable: (1) an OLS regression on a time trend \( T \) defined to be 1 in 1959, 2 in 1960, etc., (2) an AR(1) regression using the same specification, and (3) an OLS regression on \( T \) and the logarithm of food lagged one time period, with the results shown in the table (standard errors in parentheses).

<table>
<thead>
<tr>
<th></th>
<th>1: OLS</th>
<th>2: AR(1)</th>
<th>3: OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>0.0181</td>
<td>0.0166</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0021)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>( \text{LGFOOD}(-1) )</td>
<td>—</td>
<td>—</td>
<td>0.8551</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0886)</td>
</tr>
<tr>
<td>constant</td>
<td>5.7768</td>
<td>5.8163</td>
<td>0.8571</td>
</tr>
<tr>
<td></td>
<td>(0.0106)</td>
<td>(0.0586)</td>
<td>(0.5101)</td>
</tr>
<tr>
<td>( \hat{\rho} )</td>
<td>—</td>
<td>0.8551</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0886)</td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.9750</td>
<td>0.9931</td>
<td>0.9931</td>
</tr>
<tr>
<td>( RSS )</td>
<td>0.0327</td>
<td>0.0081</td>
<td>0.0081</td>
</tr>
<tr>
<td>( d )</td>
<td>0.2752</td>
<td>1.3328</td>
<td>1.3328</td>
</tr>
<tr>
<td>( h )</td>
<td>—</td>
<td>—</td>
<td>2.32</td>
</tr>
</tbody>
</table>

Discuss why each regression specification appears to be unsatisfactory. Explain why it was not possible to perform a common factor test.
Answer:
The Durbin–Watson statistic in regression (1) is very low, suggesting AR(1) autocorrelation. However, it remains below 1.40, \( d_L \) for a 5 per cent significance test with one explanatory variable and 35 observations, in the AR(1) specification in regression (2). The reason of course is that the model is very poorly specified, with two obvious major variables, income and price, excluded.

With regard to the impossibility of performing a common factor test, suppose that the original model is written:

\[
LGFOOD_t = \beta_1 + \beta_2T + u_t.
\]

Lagging the model and multiplying through by \( \rho \), we have:

\[
\rho LGFOOD_{t-1} = \beta_1 \rho + \beta_2 \rho (T - 1) + \rho u_{t-1}.
\]

Subtracting and rearranging, we obtain the AR(1) specification:

\[
LGFOOD_t = \beta_1 (1 - \rho) + \rho LGFOOD_{t-1} + \beta_2 T - \beta_2 \rho (T - 1) + u_t - \rho u_{t-1}
\]

\[
= \beta_1 (1 - \rho) + \beta_2 \rho + \rho LGFOOD_{t-1} + \beta_2 (1 - \rho) T + \epsilon_t.
\]

However, this specification does not include any restrictions. The coefficient of \( LGFOOD_{t-1} \) provides an estimate of \( \rho \). The coefficient of \( T \) then provides an estimate of \( \beta_2 \). Finally, given these estimates, the intercept provides an estimate of \( \beta_1 \). The AR(1) and ADL(1,1) specifications are equivalent in this model, the reason being that the variable \((T - 1)\) is merged into \( T \) and the intercept.

12.5 Answers to the additional exercises

A12.1 The Durbin–Watson statistic in the OLS regression is 0.49, causing us to reject the null hypothesis of no autocorrelation at the 1 per cent level. The Breusch–Godfrey statistic (not shown) is 25.12, also causing the null hypothesis of no autocorrelation to be rejected at a high significance level. Apart from a more satisfactory Durbin–Watson statistic, the results for the AR(1) specification are similar to those of the OLS one. The income and price elasticities are a little larger. The estimate of the population elasticity, negative in the OLS regression, is now effectively zero, suggesting that the direct effect of population on expenditure on food is offset by a negative income effect. The standard errors are larger than those for the OLS regression, but the latter are invalidated by the autocorrelation and therefore should not be taken at face value.

A12.2 All of the regressions exhibit strong evidence of positive autocorrelation. The Breusch–Godfrey test statistic for AR(1) autocorrelation is above the critical value of 10.82 (critical value of chi-squared with one degree of freedom at the 0.1% significance level) and the Durbin–Watson \( d \) statistic is below 1.20 (\( d_L \), 1 per cent level, 45 observations, \( k = 4 \)). The Durbin–Watson statistics for the AR(1) specification are generally much more healthy than those for the OLS one, being scattered around 2.
12. Properties of regression models with time series data

Breusch–Godfrey and Durbin–Watson statistics, logarithmic OLS regression including population

<table>
<thead>
<tr>
<th></th>
<th>B–G</th>
<th>d</th>
<th>B–G</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>19.37</td>
<td>0.683</td>
<td>GASO</td>
<td>36.21</td>
</tr>
<tr>
<td>BOOK</td>
<td>25.85</td>
<td>0.484</td>
<td>HOUSE</td>
<td>23.88</td>
</tr>
<tr>
<td>BUSI</td>
<td>24.31</td>
<td>0.507</td>
<td>LEGL</td>
<td>24.30</td>
</tr>
<tr>
<td>CLOT</td>
<td>18.47</td>
<td>0.706</td>
<td>MAGS</td>
<td>19.27</td>
</tr>
<tr>
<td>DENT</td>
<td>14.02</td>
<td>0.862</td>
<td>MASS</td>
<td>21.97</td>
</tr>
<tr>
<td>DOC</td>
<td>12.47</td>
<td>0.984</td>
<td>OPHT</td>
<td>31.64</td>
</tr>
<tr>
<td>FLOW</td>
<td>24.13</td>
<td>0.535</td>
<td>RELG</td>
<td>26.30</td>
</tr>
<tr>
<td>FOOD</td>
<td>24.95</td>
<td>0.489</td>
<td>TELE</td>
<td>30.08</td>
</tr>
<tr>
<td>FURN</td>
<td>22.92</td>
<td>0.563</td>
<td>TOB</td>
<td>27.84</td>
</tr>
<tr>
<td>GAS</td>
<td>23.41</td>
<td>0.569</td>
<td>TOYS</td>
<td>20.04</td>
</tr>
</tbody>
</table>

Since autocorrelation does not give rise to bias, one would not expect to see systematic changes in the point estimates of the coefficients. However, since multicollinearity is to some extent a problem for most categories, the coefficients do exhibit greater volatility than is usual when comparing OLS and AR(1) results. Fortunately, most of the major changes seem to be for the better. In particular, some implausibly high income elasticities are lower. Likewise, the population elasticities are a little less erratic, but most are still implausible, with large standard errors that reflect the continuing underlying problem of multicollinearity.

<table>
<thead>
<tr>
<th></th>
<th>AR(1) logarithmic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LGDPI</td>
</tr>
<tr>
<td>coef.</td>
<td>s.e.</td>
</tr>
<tr>
<td>ADM</td>
<td>-0.34</td>
</tr>
<tr>
<td>BOOK</td>
<td>0.46</td>
</tr>
<tr>
<td>BUSI</td>
<td>0.43</td>
</tr>
<tr>
<td>CLOT</td>
<td>1.07</td>
</tr>
<tr>
<td>DENT</td>
<td>1.14</td>
</tr>
<tr>
<td>DOC</td>
<td>0.85</td>
</tr>
<tr>
<td>FLOW</td>
<td>0.71</td>
</tr>
<tr>
<td>FOOD</td>
<td>0.47</td>
</tr>
<tr>
<td>FURN</td>
<td>1.73</td>
</tr>
<tr>
<td>GAS</td>
<td>-0.02</td>
</tr>
<tr>
<td>GASO</td>
<td>0.75</td>
</tr>
<tr>
<td>HOUS</td>
<td>0.27</td>
</tr>
<tr>
<td>LEGL</td>
<td>0.89</td>
</tr>
<tr>
<td>MAGS</td>
<td>0.98</td>
</tr>
<tr>
<td>MASS</td>
<td>0.06</td>
</tr>
<tr>
<td>OPHT</td>
<td>1.99</td>
</tr>
<tr>
<td>RELG</td>
<td>0.86</td>
</tr>
<tr>
<td>TELE</td>
<td>0.70</td>
</tr>
<tr>
<td>TOB</td>
<td>0.38</td>
</tr>
<tr>
<td>TOYS</td>
<td>0.89</td>
</tr>
</tbody>
</table>
12.5. Answers to the additional exercises

A12.3 The table gives the residual sum of squares for the unrestricted ADL(1,1) specification and that for the restricted AR(1) one, the fourth column giving the chi-squared statistic for the common factor test.

Before performing the common factor test, one should check that the ADL(1,1) specification is itself free from autocorrelation using the Breusch–Godfrey test. The fifth column gives the $B–G$ statistic for AR(1) autocorrelation. All but one of the statistics are below the critical value at the 5 per cent level, 3.84. The exception is that for $LEGL$. It should be remembered that the Breusch–Godfrey test is a large-sample test and in this application, with only 44 observations, the sample is rather small.

<table>
<thead>
<tr>
<th>Common factor test and tests of autocorrelation for ADL(1,1) model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>ADM</td>
</tr>
<tr>
<td>BOOK</td>
</tr>
<tr>
<td>BUSI</td>
</tr>
<tr>
<td>CLOT</td>
</tr>
<tr>
<td>DENT</td>
</tr>
<tr>
<td>DOC</td>
</tr>
<tr>
<td>FLOW</td>
</tr>
<tr>
<td>FOOD</td>
</tr>
<tr>
<td>FURN</td>
</tr>
<tr>
<td>GAS</td>
</tr>
<tr>
<td>GASO</td>
</tr>
<tr>
<td>HOUS</td>
</tr>
<tr>
<td>LEGL</td>
</tr>
<tr>
<td>MAGS</td>
</tr>
<tr>
<td>MASS</td>
</tr>
<tr>
<td>OPHT</td>
</tr>
<tr>
<td>RELG</td>
</tr>
<tr>
<td>TELE</td>
</tr>
<tr>
<td>TOB</td>
</tr>
<tr>
<td>TOYS</td>
</tr>
</tbody>
</table>

For the common factor test, the critical values of chi-squared are 7.81 and 11.34 at the 5 and 1 per cent levels, respectively, with 3 degrees of freedom. Summarising the results, we find:

- AR(1) specification not rejected: $BUSI$, $DENT$, $DOC$, $FLOW$, $FURN$, $MAGS$, $MASS$, $OPHT$, $RELG$, $TOB$, $TOYS$.
- AR(1) specification rejected at 5 per cent level: $BOOK$, $CLOT$, $FOOD$, $GAS$, $TELE$.
- AR(1) specification rejected at 1 per cent level: $ADM$, $GASO$, $HOUS$, $LEGL$.

A12.4 Discuss whether specification (1) is an adequate representation of the data.

The Breusch–Godfrey statistic is well in excess of the critical value at the 0.1 per cent significance level, 10.83. Likewise, the Durbin–Watson statistic is far below
12. Properties of regression models with time series data

1.15, \(d_L\) at the 1 per cent level with two explanatory variables and 36 observations. There is therefore strong evidence of either severe AR(1) autocorrelation or some serious misspecification.

**Discuss whether specification (3) is an adequate representation of the data.**

The only item that we can check is whether it is free from autocorrelation. The Breusch–Godfrey statistic is well under 3.84, the critical value at the 5 per cent significance level, and so there is no longer evidence of autocorrelation or misspecification.

**Discuss whether specification (2) is an adequate representation of the data.**

Let the original model be written:

\[
LGLIFE = \beta_1 + \beta_2 LGDPI + \beta_3 LGDPRLIFE + u
\]

\[
u_t = \rho u_{t-1} + \varepsilon_t.
\]

The AR(1) specification is then:

\[
LGLIFE = \beta_1 (1 - \rho) + \rho LGLIFE(-1) + \beta_2 LGDPI - \beta_2 \rho LGDPI(-1)
\]

\[+ \beta_3 LGDPRLIFE - \beta_3 \rho LGPRLIFE(-1) + \varepsilon_t.
\]

This is a restricted version of the ADL(1,1) model because it incorporates nonlinear restrictions on the coefficients of \(LGDPI(-1)\) and \(LGPRLIFE(-1)\). In the ADL(1,1) specification, minus the product of the coefficients of \(LGLIFE(-1)\) and \(LGDPI\) is \(-0.82 \times 0.42 = -0.34\). The coefficient of \(LGDPI(-1)\) is smaller than this, but then its standard error is large. Minus the product of the coefficients of \(LGLIFE(-1)\) and \(LGPRLIFE\) is \(-0.82 \times -0.59 = 0.48\). The coefficient of \(LGPRLIFE(-1)\) is fairly close, bearing in mind that its standard error is also large. The coefficient of \(LGLIFE(-1)\) is exactly equal to the estimate of \(\rho\) in the AR(1) specification.

The common factor test statistic is:

\[
35 \log \frac{0.799}{0.719} = 3.69.
\]

The null hypothesis is that the two restrictions are valid. Under the null hypothesis, the test statistic has a chi-squared distribution with 2 degrees of freedom. Its critical value at the 5 per cent level is 5.99. Hence we do not reject the restrictions and the AR(1) specification therefore does appear to be acceptable.

**Discuss whether specification (4) is an adequate representation of the data.**

We note that \(LGLDPI(-1)\) and \(LGPRLIFE(-1)\) do not have significant \(t\) statistics, but since they are being dropped simultaneously, we should perform an \(F\) test of their joint explanatory power:

\[
F(2, 29) = \frac{(0.732 - 0.719)/2}{0.719/29} = 0.26.
\]

Since this is less than 1, it is not significant at any significance level and so we do not reject the null hypothesis that the coefficients of \(LGLDPI(-1)\) and
LGPRLIFE$(-1)$ are both 0. Hence it does appear that we can drop these variables. We should also check for autocorrelation. The Breusch–Godfrey statistic indicates that there is no problem.

*If you were presenting these results at a seminar, what would you say were your conclusions concerning the most appropriate of specifications (1) – (4)?*

There is no need to mention (1). (3) is not a candidate because we have found acceptable simplifications that are likely to yield more efficient parameter estimates, and this is reflected in the larger standard errors compared with (2) and (4). We cannot discriminate between (2) and (4).

*At the seminar a commentator points out that in specification (4) neither LGDPI nor LGPRLIFE have significant coefficients and so these variables should be dropped. As it happens, the researcher has considered this specification, and the results are shown as specification (5) in the table. What would be your answer to the commentator?*

Comparing (3) and (5):

$$F(4, 29) = \frac{(0.843 - 0.719)/4}{0.719/29} = 1.25.$$  

The critical value of $F(4, 29)$ at the 5 per cent level is 2.70, so it would appear that the joint explanatory power of the 4 income and price variables is not significant. However, it does not seem sensible to drop current income and current price from the model. The reason that they have so little explanatory power is that the short-run effects are small, life insurance being subject to long-term contracts and thus a good example of a category of expenditure with a large amount of inertia. The fact that income in the AR(1) specification has a highly significant coefficient is concrete evidence that it should not be dropped.

A12.5  *Looking at all five regressions together, evaluate the adequacy of:*

- *specification 1.  
- *specification 2.  
- *specification 3.  
- *specification 4.  
- Specification 1 has a very high Breusch–Godfrey statistic and a very low Durbin–Watson statistic. There is evidence of either severe autocorrelation or model misspecification.  
- Specification 2 also has a very high Breusch–Godfrey statistic and a very low Durbin–Watson statistic. Further, there is evidence of multicollinearity: large standard errors (although comparisons are very dubious given low DW), and implausible coefficients.  
- Specification 3 seems acceptable. In particular, there is no evidence of autocorrelation since the Breusch–Godfrey statistic is low.  
- Specification 4: dropping $m(-1)$ may be expected to cause omitted variable bias since the $t$ statistic for its coefficient was $-3.0$ in specification 3.*
12. Properties of regression models with time series data

(Equivalently, the $F$ statistic is:

$$F(1, 46) = \frac{(0.0120 - 0.0100)/1}{0.0100/46} = 0.2 \times 46 = 9.2$$

the square of the $t$ statistic and similarly significant.)

*Explain why specification 5 is a restricted version of one of the other specifications, stating the restriction, and explaining the objective of the manipulations that lead to specification 5.*

Write the original model and AR(1) process:

$$p_t = \beta_1 + \beta_2 m_t + u_t$$
$$u_y = \rho u_{t-1} + \epsilon_t.$$

Then fitting:

$$p_t = \beta_1 (1 - \rho) + \rho p_{t-1} + \beta_2 m_t - \beta_2 \rho m_{t-1} + \epsilon_t$$

removes the autocorrelation. This is a restricted version of specification 3, with restriction that the coefficient of $m_{t-1}$ is equal to minus the product of the coefficients of $m_t$ and $p_{t-1}$.

*Perform a test of the restriction embodied in specification 5.*

Comparing specifications 3 and 5, the common factor test statistic is:

$$n \log \left( \frac{RSS_R}{RSS_U} \right) = 50 \log \left( \frac{0.0105}{0.0100} \right) = 50 \log 1.05 \cong 50 \times 0.05 = 2.5.$$  

Under the null hypothesis that the restriction implicit in the specification is valid, the test statistic is distributed as chi-squared with one degree of freedom. The critical value at the 5 per cent significance level is 3.84, so we do not reject the restriction. Accordingly, specification 5 appears to be an adequate representation of the data.

*Explain which would be your preferred specification.*

Specifications (3) and (5) both appear to be adequate representations of the data. (5) should yield more efficient estimators of the parameters because, exploiting an apparently-valid restriction, it is less susceptible to multicollinearity, and this appears to be confirmed by the lower standard errors.

A12.6 The models are:

1. $p_t = \beta_1 + \beta_2 m_t + u_t$
2. $p_t = \beta_1 + \beta_2 m_t + \beta_3 m_{t-1} + \beta_4 m_{t-2} + \beta_5 m_{t-3} + u_t$
3. $p_t = \beta_1 + \beta_2 m_t + \beta_3 m_{t-1} + \beta_6 p_{t-1} + u_t$
4. $p_t = \beta_1 + \beta_2 m_t + \beta_6 p_{t-1} + u_t$
5. $p_t = \beta_1 (1 - \beta_6) + \beta_6 p_{t-1} + \beta_2 m_t - \beta_2 \beta_6 m_{t-1} + \epsilon_t$ (writing $\rho = \beta_6$).
Hence we obtain the following estimates of $\partial p_t/\partial m_t$:
1. $0.95$
2. $0.50$
3. $0.40$
4. $0.18$
5. $0.90$.

Putting $p$ and $m$ equal to equilibrium values, and ignoring the disturbance term, we have:
1. $\tilde{p} = \beta_1 + \beta_2 \bar{m}$
2. $\tilde{p} = \beta_1 + (\beta_2 + \beta_3 + \beta_4) \bar{m}$
3. $\tilde{p} = \frac{1}{1-\beta_6} (\beta_1 + (\beta_2 + \beta_3) \bar{m})$
4. $\tilde{p} = \frac{1}{1-\beta_6} (\beta_1 + \beta_2 \bar{m})$
5. $\tilde{p} = \beta_1 + \beta_2 \bar{m}$.

Hence we obtain the following estimates of $d\tilde{p}/d\bar{m}$:
1. $0.95$
2. $0.95$
3. $1.00$
4. $0.90$
5. $0.90$.

A12.7 Evaluate regression (1).

Regression (1) has a very high Breusch–Godfrey statistic and a very low Durbin–Watson statistic. The null hypothesis of no autocorrelation is rejected at the 1 per cent level for both tests. Alternatively, the test statistics might indicate some misspecification problem.

Evaluate regression (2). Explain mathematically what assumptions were being made by the researcher when he used the AR(1) specification and why he hoped the results would be better than those obtained with regression (1).

Regression (2) has been run on the assumption that the disturbance term follows an AR(1) process:

$$u_t = \rho u_{t-1} + \epsilon_t.$$  

On the assumption that the regression model should be:

$$LGTAXI_t = \beta_1 + \beta_2 LGDPI_t + u_t,$$

the autocorrelation can be eliminated in the following way: lag the regression model by one time period and multiply through by $\rho$:

$$\rho LGTAXI_{t-1} = \beta_1 \rho + \beta_2 \rho LGDPI_{t-1} + \rho u_{t-1}.$$ 

Subtract this from the regression model:

$$LGTAXI_t - \rho LGTAXI_{t-1} = \beta_1 (1 - \rho) + \beta_2 LGDPI_t - \beta_2 \rho LGDPI_{t-1} + u_t - \rho u_{t-1}.$$
12. Properties of regression models with time series data

Hence one obtains a specification free from autocorrelation:

\[ LGTXI_t = \beta_1 (1 - \rho) + \rho LGTXI_{t-1} + \beta_2 LGDPI_t - \beta_2 \rho LGDPI_{t-1} + \varepsilon_t. \]

The Durbin–Watson statistic is still low, suggesting that fitting the AR(1) specification was an inappropriate response to the problem.

Evaluate regression (3).

In regression (3) the Breusch–Godfrey statistic suggests that, for this specification, there is not a problem of autocorrelation (the Durbin–Watson statistic is indecisive). This suggests that the apparent autocorrelation in the regression (1) is in fact attributable to the omission of the price variable.

This is corroborated by the diagrams, which show that large negative residuals occurred when the price rose and positive ones when it fell. The effect is especially obvious in the final years of the sample period.

Evaluate regression (4). In particular, discuss the possible reasons for the differences in the standard errors in regressions (3) and (4).

In regression (4), the Durbin–Watson statistic does not indicate a problem of autocorrelation. Overall, there is little to choose between regressions (3) and (4). It is possible that there was some autocorrelation in regression (3) and that it has been rectified by using AR(1) in regression (4). It is also possible that autocorrelation was not actually a problem in regression (3). Regressions (3) and (4) yield similar estimates of the income and price elasticities and in both cases the elasticities are significantly different from zero at a high significance level. If regression (4) is the correct specification, the lower standard errors in regression (3) should be disregarded because they are invalid. If regression (3) is the correct specification, AR(1) estimation will yield inefficient estimates; which could account for the higher standard errors in regression (4).

At a seminar one of the participants says that the researcher should consider adding lagged values of LGTXI, LGDPI, and LGP to the specification. What would be your view?

Specifications (2) and (4) already contain the lagged values, with restrictions on the coefficients of LGDPI(-1) and LGP(-1).

A12.8 Explain why the researcher was not satisfied with regression (1).

The researcher was not satisfied with the results of regression (1) because the Breusch–Godfrey statistic was 4.42, above the critical value at the 5 per cent level, 3.84, and because the Durbin–Watson \( d \) statistic was only 0.99. The critical value of \( d \) with one explanatory variable and 30 observations is 1.35. Thus there is evidence that the specification may be subject to autocorrelation.

Evaluate regression (2). Explain why the coefficients of \( I(-1) \) and \( r(-1) \) are not reported, despite the fact that they are part of the regression specification.

Specification (2) is equally unsatisfactory. The fact that the Durbin–Watson statistic has remained low is an indication that the reason for the low \( d \) in (1) was not an AR(1) disturbance term. RSS is very high compared with those in specifications (4) – (6). The coefficient of \( I(-1) \) is not reported as such because it
12.5. Answers to the additional exercises

is the estimate $\hat{\rho}$. The coefficient of $r(-1)$ is not reported because it is constrained to be minus the product of $\hat{\rho}$ and the coefficient of $I$.

**Evaluate regression (3).**

Specification (3) is the unrestricted ADL(1,1) model of which the previous AR(1) model was a restricted version and it suffers from the same problems. There is still evidence of positive autocorrelation, since the Breusch–Godfrey statistic, 4.24, is high and $RSS$ is still much higher than in the three remaining specifications.

**Evaluate regression (4).**

Specification (4) seems fine. The null hypothesis of no autocorrelation is not rejected by either the Breusch–Godfrey statistic or the Durbin–Watson statistic. The coefficients are significant and have the expected signs.

**Evaluate regression (5).**

The AR(1) specification (5) does not add anything because there was no evidence of autocorrelation in (4). The estimate of $\rho$ is not significantly different from zero.

**Evaluate regression (6).**

Specification (6) does not add anything either. $t$ tests on the coefficients of the lagged variables indicate that they are individually not significantly different from zero. Likewise the joint hypothesis that their coefficients are all equal to zero is not rejected by an $F$ test comparing $RSS$ in (4) and (6):

$$F(3, 23) = \frac{(27.4 - 23.5)/3}{23.5/23} = 1.27.$$

The critical value of $F(3, 23)$ at the 5 per cent level is 3.03. [There is no point in comparing (5) and (6) using a common factor test, but for the record the test statistic is:

$$n \log_e \frac{RSS_R}{RSS_U} = 29 \log_e \frac{26.8}{23.5} = 3.81.$$

The critical value of chi-squared with 2 degrees of freedom at the 5 per cent level is 5.99.]

**Summarise your conclusions concerning the evaluation of the different regressions. Explain whether an examination of the figure supports your conclusions.**

The overall conclusion is that the static model (4) is an acceptable representation of the data and the apparent autocorrelation in specifications (1) – (3) is attributable to the omission of $g$. Figure 12.3 shows very clearly that the residuals in specification (1) follow the same pattern as $g$, confirming that the apparent autocorrelation in the residuals is in fact attributable to the omission of $g$ from the specification.

A12.9  **In Exercise A11.5 you performed a test of a restriction. The result of this test will have been invalidated if you found that the specification was subject to autocorrelation. How should the test be performed, assuming the correct specification is ADL(1,1)?**

© Christopher Dougherty, 2016. All rights reserved.

Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
Properties of regression models with time series data

If the ADL(1,1) model is written:
\[ \log \text{CAT} = \beta_1 + \beta_2 \log \text{DPI} + \beta_3 \log P + \beta_4 \log \text{POP} + \beta_5 \log \text{CAT}_{-1} + \beta_6 \log \text{DPI}_{-1} + \beta_7 \log P_{-1} + \beta_8 \log \text{POP}_{-1} + u \]

the restricted version with expenditure per capita a function of income per capita is:
\[ \log \frac{\text{CAT}}{\text{POP}} = \beta_1 + \beta_2 \log \frac{\text{DPI}}{\text{POP}} + \beta_3 \log P + \beta_5 \log \frac{\text{CAT}_{-1}}{\text{POP}_{-1}} + \beta_6 \log \frac{\text{DPI}_{-1}}{\text{POP}_{-1}} + \beta_7 \log P_{-1} + u. \]

Comparing the two equations, we see that the restrictions are \( \beta_4 = 1 - \beta_2 \) and \( \beta_8 = -\beta_5 - \beta_6 \). The usual \( F \) statistic should be constructed and compared with the critical values of \( F(2,28) \).

A12.10 Let the AR(1) process be written:
\[ u_t = \rho u_{t-1} + \varepsilon_t. \]

As the specification stands, OLS would yield inconsistent estimates because both the explanatory variable and the disturbance term depend on \( u_{t-1} \). Applying the standard procedure, multiplying the lagged relationship by \( \rho \) and subtracting, one has:
\[ Y_t - \rho Y_{t-1} = \beta_1 (1 - \rho) + \beta_2 Y_{t-1} - \beta_2 \rho Y_{t-1} + u_t - \rho u_{t-1}. \]

Hence:
\[ Y_t = \beta_1 (1 - \rho) + (\beta_2 + \rho) Y_{t-1} - \beta_2 \rho Y_{t-2} + \varepsilon_t. \]

It follows that the model should be fitted as a second-order, rather than as a first-order, process. There are no restrictions on the coefficients. OLS estimators will be consistent, but subject to finite-sample bias.

A12.11 Explain what is correct, incorrect, confused or incomplete in the following statements, giving a brief explanation if not correct.

- The disturbance term in a regression model is said to be autocorrelated if its values in a sample of observations are not distributed independently of each other.
  Correct.
- When the disturbance term is subject to autocorrelation, the ordinary least squares estimators are inefficient ...
  Correct.
- ...and inconsistent...
  Incorrect, unless there is a lagged dependent variable.
- ...but they are not biased...
  Correct, unless there is a lagged dependent variable.
12.5. Answers to the additional exercises

- ...and the t tests are invalid.
  Correct.

- It is a common problem in time series models because it always occurs when the dependent variable is correlated with its previous values.
  Incorrect.

- If this is the case, it could be eliminated by including the lagged value of the dependent variable as an explanatory variable.
  In general, incorrect. However, a model requiring a lagged dependent variable could appear to exhibit autocorrelation if the lagged dependent variable were omitted, and including it could eliminate the apparent problem.

- However, if the model is correctly specified and the disturbance term satisfies the regression model assumptions, adding the lagged value of the dependent variable as an explanatory variable will have the opposite effect and cause the disturbance term to be autocorrelated.
  Nonsense.

- A second way of dealing with the problem of autocorrelation is to use an instrumental variable.
  More nonsense.

- If the autocorrelation is of the AR(1) type, randomising the order of the observations will cause the Durbin–Watson statistic to be near 2...
  Correct.

- ...thereby eliminating the problem.
  Incorrect. The problem will have been disguised, not rectified.
12. Properties of regression models with time series data
Chapter 13
Introduction to nonstationary time series

13.1 Overview

This chapter begins by defining the concepts of stationarity and nonstationarity as applied to univariate time series and, in the case of nonstationary series, the concepts of difference-stationarity and trend-stationarity. It next describes the consequences of nonstationarity for models fitted using nonstationary time-series data and gives an account of the Granger–Newbold Monte Carlo experiment with random walks. Next the two main methods of detecting nonstationarity in time series are described, the graphical approach using correlograms and the more formal approach using Augmented Dickey–Fuller unit root tests. This leads to the topic of cointegration. The chapter concludes with a discussion of methods for fitting models using nonstationary time series: detrending, differencing, and error-correction models.

13.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain what is meant by stationarity and nonstationarity.
- explain what is meant by a random walk and a random walk with drift
- derive the condition for the stationarity of an AR(1) process
- explain what is meant by an integrated process and its order of integration
- explain why Granger and Newbold obtained the results that they did
- explain what is depicted by a correlogram
- perform an Augmented Dickey–Fuller unit root test to test a time series for nonstationarity
- test whether a set of time series are cointegrated
- construct an error-correction model and describe its advantages over detrending and differencing.
13. Introduction to nonstationary time series

13.3 Further material

Addition to Section 13.6 Cointegration

Section 13.6 contains the following paragraph on page 507:

In the case of a cointegrating relationship, least squares estimators can be shown to be superconsistent (Stock, 1987). An important consequence is that OLS may be used to fit a cointegrating relationship, even if it belongs to a system of simultaneous relationships, for any simultaneous equations bias tends to zero asymptotically.

This cries out for an illustrative simulation, so here is one. Consider the model:

\[ Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + \varepsilon_{Yt} \]
\[ X_t = \alpha_1 + \alpha_2 Y_t + \varepsilon_{Xt} \]
\[ Z_t = \rho Z_{t-1} + \varepsilon_{Zt} \]

where \( Y_t \) and \( X_t \) are endogenous variables, \( Z_t \) is exogenous, and \( \varepsilon_{Yt} \), \( \varepsilon_{Xt} \), and \( \varepsilon_{Zt} \) are iid \( N(0, 1) \) disturbance terms. We expect OLS estimators to be inconsistent if used to fit either of the first two equations. However, if \( \rho = 1 \), \( Z \) is nonstationary, and \( X \) and \( Y \) will also be nonstationary. So, if we fit the second equation, for example, the OLS estimator of \( \alpha_2 \) will be superconsistent. This is illustrated by a simulation where the first two equations are:

\[ Y_t = 1.0 + 0.8X_t + 0.5Z_t + \varepsilon_{Yt} \]
\[ X_t = 2.0 + 0.4Y_t + \varepsilon_{Xt} \]

The distributions in the right of the figure below (dashed lines) are for the case \( \rho = 0.5 \). \( Z \) is stationary, and so are \( Y \) and \( X \). You will have no difficulty in demonstrating that \( \lim \hat{\alpha}_2^{OLS} = 0.68 \). The distributions to the left of the figure (solid lines) are for \( \rho = 1 \), and you can see that in this case the estimator is consistent. But is it superconsistent? The variance seems to be decreasing relatively slowly, not fast, especially for small sample sizes. The explanation is that the superconsistency becomes apparent only for very large sample sizes, as shown in the second figure.
13.4 Additional exercises

A13.1 The Figure 13.1 plots the logarithm of the US population for the period 1959–2003. It is obviously nonstationary. Discuss whether it is more likely to be difference-stationary or trend-stationary.

A13.2 Figure 13.2 plots the first difference of the logarithm of the US population for the period 1959–2003. Explain why the vertical axis measures the proportional growth rate. Comment on whether the series appears to be stationary or nonstationary.

A13.3 The regression output below shows the results of ADF unit root tests on the logarithm of the US population, and its difference, for the period 1959–2003. Comment on the results and state whether they confirm or contradict your conclusions in Exercise A13.2.
13. Introduction to nonstationary time series

Figure 13.2: Logarithm of the US population, first difference.

Augmented Dickey--Fuller Unit Root Test on LGPOP

Null Hypothesis: LGPOP has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Fixed)

<table>
<thead>
<tr>
<th>t-Statistic</th>
<th>Prob.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Dickey--Fuller test statistic</td>
<td>-2.030967</td>
</tr>
<tr>
<td>Test critical values 1% level</td>
<td>-4.186481</td>
</tr>
<tr>
<td>5% level</td>
<td>-3.518090</td>
</tr>
<tr>
<td>10% level</td>
<td>-3.189732</td>
</tr>
</tbody>
</table>


Augmented Dickey--Fuller Test Equation
Dependent Variable: D(LGPOP)
Method: Least Squares
Sample(adjusted): 1961 2003
Included observations: 43 after adjusting endpoints

<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>LGPOP(-1)</td>
<td>-0.047182</td>
<td>0.023231</td>
<td>-2.030967</td>
<td>0.0491</td>
</tr>
<tr>
<td>D(LGPOP(-1))</td>
<td>0.687772</td>
<td>0.058979</td>
<td>11.66139</td>
<td>0.0000</td>
</tr>
<tr>
<td>C</td>
<td>0.574028</td>
<td>0.281358</td>
<td>2.040209</td>
<td>0.0481</td>
</tr>
<tr>
<td>@TREND(1959)</td>
<td>0.000507</td>
<td>0.000246</td>
<td>2.060295</td>
<td>0.0461</td>
</tr>
</tbody>
</table>

R-squared 0.839263  Mean dependent var 0.011080
Adjusted R-squared 0.826898  S.D. dependent var 0.001804
S.E. of regression 0.000750  Akaike info criter-11.46327
Sum squared resid 2.20E-05  Schwarz criterion -11.29944
Log likelihood 250.4603  F-statistic 67.87724
Durbin-Watson stat 1.164933  Prob(F-statistic) 0.000000
13.4. Additional exercises

Augmented Dickey--Fuller Unit Root Test on DLGPOP

Null Hypothesis: DLGPOP has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Fixed)

<table>
<thead>
<tr>
<th>t-Statistic</th>
<th>Prob.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Dickey--Fuller test statistic</td>
<td>-2.513668</td>
</tr>
</tbody>
</table>

Test critical values
1% level | -4.192337 |
5% level | -3.520787 |
10% level | -3.191277 |


Augmented Dickey--Fuller Test Equation
Dependent Variable: D(DLGPOP)
Method: Least Squares
Sample(adjusted): 1962 2003
Included observations: 42 after adjusting endpoints

<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>DLGPOP(-1)</td>
<td>-0.161563</td>
<td>0.064274</td>
<td>-2.513668</td>
<td>0.0163</td>
</tr>
<tr>
<td>D(DLGPOP(-1))</td>
<td>0.294717</td>
<td>0.117766</td>
<td>2.502573</td>
<td>0.0167</td>
</tr>
<tr>
<td>C</td>
<td>0.001714</td>
<td>0.000796</td>
<td>2.152327</td>
<td>0.0378</td>
</tr>
<tr>
<td>@TREND(1959)</td>
<td>-1.32E-07</td>
<td>9.72E-06</td>
<td>-0.013543</td>
<td>0.9893</td>
</tr>
</tbody>
</table>

R-squared | 0.320511 |
Mean dependent var | -0.000156 |
Adjusted R-squared | 0.266867 |
S.D. dependent var | 0.000827 |
Akaike info criter | 11.57806 |
Schwarz criterion | 11.41257 |
F-statistic | 5.974780 |
Prob(F-statistic) | 0.001932 |

A13.4 A researcher believes that a time series is generated by the process:

\[ X_t = \rho X_{t-1} + \varepsilon_t \]

where \( \varepsilon_t \) is a white noise series generated randomly from a normal distribution with mean zero, constant variance, and no autocorrelation. Explain why the null hypothesis for a test of nonstationarity is that the series is nonstationary, rather than stationary.

A13.5 A researcher correctly believes that a time series is generated by the process:

\[ X_t = \rho X_{t-1} + \varepsilon_t \]

where \( \varepsilon_t \) is a white noise series generated randomly from a normal distribution with mean zero, constant variance, and no autocorrelation. Unknown to the researcher, the true value of \( \rho \) is 0.7. The researcher uses a unit root test to test the series for nonstationarity. The output is shown. Discuss the result of the test.
13. Introduction to nonstationary time series

**Augmented Dickey–Fuller Unit Root Test on X**

============================================================
ADF Test Statistic -2.528841 1% Critical Value*-3.6289
5% Critical Value -2.9472
10% Critical Value -2.6118

============================================================

*MacKinnon critical values for rejection of hypothesis of a unit root.*

**Augmented Dickey–Fuller Test Equation**
Dependent Variable: D(X)
Method: Least Squares
Sample(adjusted): 2 36
Included observations: 35 after adjusting endpoints

<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>X(-1)</td>
<td>-0.379661</td>
<td>0.150132</td>
<td>-2.528841</td>
<td>0.0164</td>
</tr>
<tr>
<td>C</td>
<td>0.222066</td>
<td>0.203435</td>
<td>1.091580</td>
<td>0.2829</td>
</tr>
</tbody>
</table>

R-squared 0.162331  Mean dependent var-0.052372
Adjusted R-squared 0.136947  S.D. dependent var 1.095782
S.E. of regression 1.017988  Akaike info criteri2.928979
Sum squared resid 34.19792  Schwarz criterion 3.017856
Log likelihood -49.25714  F-statistic 6.395035
Durbin-Watson stat 1.965388  Prob(F-statistic) 0.016406

A13.6 Test of cointegration. Perform a logarithmic regression of expenditure on your commodity on income, relative price, and population. Save the residuals and test them for stationarity. (Note: the critical values in the regression output do not apply to tests of cointegration. For the correct critical values, see the text.)

A13.7 A variable $Y_t$ is generated by the autoregressive process:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \varepsilon_t$$

where $\beta_2 = 1$ and $\varepsilon_t$ satisfies the regression model assumptions. A second variable $Z_t$ is generated as the lagged value of $Y_t$:

$$Z_t = Y_{t-1}.$$ 

Show that $Y$ and $Z$ are nonstationary processes. Show that nevertheless they are cointegrated.

A13.8 $X_t$ and $Z_t$ are independent I(1) (integrated of order 1) time series. $W_t$ is a stationary time series. $Y_t$ is generated as the sum of $X_t$, $Z_t$, and $W_t$. Not knowing this, a researcher regresses $Y_t$ on $X_t$ and $Z_t$. Explain whether he would find a cointegrating relationship.
A13.9 Two random walks $RA_t$ and $RB_t$, and two stationary processes $SA_t$ and $SB_t$ are generated by the following processes:

$$
RA_t = RA_{t-1} + \varepsilon_{1t} \\
RB_t = RB_{t-1} + \varepsilon_{2t} \\
SA_t = \rho_A SA_{t-1} + \varepsilon_{3t}, \quad 0 < \rho_A < 1 \\
SB_t = \rho_B SB_{t-1} + \varepsilon_{4t}, \quad 0 < \rho_B < 1
$$

where $\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t},$ and $\varepsilon_{4t}$, are iid $N(0, 1)$ (independently and identically distributed from a normal distribution with mean 0 and variance 1).

- Two series $XA_t$ and $XB_t$ are generated as:
  $$
  XA_t = RA_t + SA_t \\
  XB_t = RB_t + SB_t.
  $$

Explain whether it is possible for $XA_t$ and $XB_t$ to be stationary.

Explain whether it is possible for them to be cointegrated.

- Two series $YA_t$ and $YB_t$ are generated as:
  $$
  YA_t = RA_t + SA_t \\
  YB_t = RA_t + SB_t.
  $$

Explain whether it is possible for $YA_t$ and $YB_t$ to be cointegrated.

- Two series $ZA_t$ and $ZB_t$ are generated as:
  $$
  ZA_t = RA_t + RB_t + SA_t \\
  ZB_t = RA_t - RB_t + SB_t.
  $$

Explain whether it is possible for $ZA_t$ and $ZB_t$ to be stationary.

Explain whether it is possible for them to be cointegrated.

13.5 Answers to the starred exercises in the textbook

13.1 Demonstrate that the MA(1) process:

$$
X_t = \varepsilon_t + \alpha_2 \varepsilon_{t-1}
$$

is stationary. Does the result generalise to higher-order MA processes?

Answer:

The expected value of $X_t$ is zero and therefore independent of time:

$$
E(X_t) = E(\varepsilon_t + \alpha_2 \varepsilon_{t-1}) = E(\varepsilon_t) + \alpha_2 E(\varepsilon_{t-1}) = 0 + 0 = 0.
$$
13. Introduction to nonstationary time series

Since \( \varepsilon_t \) and \( \varepsilon_{t-1} \) are uncorrelated:
\[
\sigma^{2}_{X_t} = \sigma^{2}_{\varepsilon_t} + \alpha^{2}_{2}\sigma^{2}_{\varepsilon_{t-1}}
\]
and this is independent of time. Finally, because:
\[
X_{t-1} = \varepsilon_{t-1} + \alpha_{2}\varepsilon_{t-2},
\]
the population covariance of \( X_t \) and \( X_{t-1} \) is given by:
\[
\sigma_{X_t, X_{t-1}} = \alpha_{2}\sigma^{2}_{\varepsilon}.
\]

This is fixed and independent of time. The population covariance between \( X_t \) and \( X_{t-s} \) is zero for all \( s > 1 \) since then \( X_t \) and \( X_{t-1} \) have no elements in common. Thus the third condition for stationarity is also satisfied.

All MA processes are stationary, the general proof being a simple extension of that for the MA(1) case.

13.2 A stationary AR(1) process:
\[
X_t = \beta_1 + \beta_2 X_{t-1} + \varepsilon_t
\]
with \( |\beta_2| < 1 \), has initial value \( X_0 \), where \( X_0 \) is defined as:
\[
X_0 = \frac{\beta_1}{1 - \beta_2} + \sqrt{\frac{1}{1 - \beta^2_2}}\varepsilon_0.
\]

Demonstrate that \( X_0 \) is a random draw from the ensemble distribution for \( X \).

Answer:

Lagging and substituting, it was shown, equation (13.12), that:
\[
X_t = \beta_2 X_0 + \beta_1 \frac{1 - \beta^t_2}{1 - \beta_2} + \beta_2^{t-1}\varepsilon_1 + \cdots + \beta_2^{2}\varepsilon_{t-2} + \beta_2\varepsilon_{t-1} + \varepsilon_t.
\]

With the stochastic definition of \( X_0 \), we now have:
\[
X_t = \beta_2^t \left( \frac{\beta_1}{1 - \beta_2} + \sqrt{\frac{1}{1 - \beta^2_2}}\varepsilon_0 \right) + \beta_1 \frac{1 - \beta^t_2}{1 - \beta_2} + \beta_2^{t-1}\varepsilon_1 + \cdots + \beta_2^{2}\varepsilon_{t-2} + \beta_2\varepsilon_{t-1} + \varepsilon_t
\]

\[
= \frac{\beta_1}{1 - \beta_2} + \beta_2^t \sqrt{\frac{1}{1 - \beta^2_2}}\varepsilon_0 + \beta_2^{t-1}\varepsilon_1 + \cdots + \beta_2^{2}\varepsilon_{t-2} + \beta_2\varepsilon_{t-1} + \varepsilon_t.
\]

Hence:
\[
E(X_t) = \frac{\beta_1}{1 - \beta_2}
\]

and:
\[
\text{var}(X_t) = \text{var} \left( \beta_2^t \sqrt{\frac{1}{1 - \beta^2_2}}\varepsilon_0 + \beta_2^{t-1}\varepsilon_1 + \cdots + \beta_2^{2}\varepsilon_{t-2} + \beta_2\varepsilon_{t-1} + \varepsilon_t \right)
\]

\[
= \frac{\beta_2^t\sigma^2_{\varepsilon} + (\beta_2^{t-2} + \cdots + \beta_2^t + \beta_2^2 + 1)\sigma^2_{\varepsilon}}{1 - \beta^2_2}
\]

\[
= \frac{\beta_2^t\sigma^2_{\varepsilon} + 1 - \beta^2_2\sigma^2_{\varepsilon}}{1 - \beta^2_2} = \frac{\sigma^2_{\varepsilon}}{1 - \beta^2_2}.
\]
13.5. Answers to the starred exercises in the textbook

Given the generating process for \( X_0 \), one has:

\[
E(X_0) = \frac{\beta_1}{1 - \beta_2} \quad \text{and} \quad \text{var}(X_0) = \frac{\sigma^2_\varepsilon}{1 - \beta_2^2}.
\]

Hence \( X_0 \) is a random draw from the ensemble distribution. Implicitly it has been assumed that the distributions of \( \varepsilon \) and \( X_0 \) are both normal. This should have been stated explicitly.

13.4 Suppose that \( Y_t \) is determined by the process:

\[
Y_t = Y_{t-1} + \varepsilon_t + \lambda \varepsilon_{t-1}
\]

where \( \varepsilon_t \) is iid. Show that the process for \( Y_t \) is nonstationary unless \( \lambda \) takes a certain value.

**Answer:**

Lagging and substituting back to time 0:

\[
Y_t = Y_0 + \sum_{s=1}^{t} \varepsilon_t + \lambda \sum_{s=0}^{t-1} \varepsilon_t = Y_0 + (1 + \lambda) \sum_{s=1}^{t-1} \varepsilon_t + \varepsilon_t + \lambda \varepsilon_0.
\]

The expectation of \( Y_t \), taken at time 0, is \( Y_0 \) and independent of time. The variance of \( Y_t \) is \( ((t-1)(1+\lambda)^2 + 1 + \lambda^2) \sigma^2_\varepsilon \). The process is nonstationary because the variance is dependent on time, unless \( \lambda = -1 \), in which case the process is stationary. It reduces to:

\[
Y_t = Y_0 + \varepsilon_t - \varepsilon_0.
\]

The covariance between \( Y_t \) and \( Y_{t-s} \) is zero for all \( s \) greater than 0 if \( \varepsilon_0 \) is taken as predetermined. It is equal to the variance of \( \varepsilon \) if \( \varepsilon_0 \) is treated as random. Either way, it is independent of time.

13.11 Suppose that a series is generated as:

\[
X_t = \beta_2 X_{t-1} + \varepsilon_t
\]

with \( \beta_2 \) equal to \( 1 - \delta \), where \( \delta \) is small. Demonstrate that, if \( \delta \) is small enough that terms involving \( \delta^2 \) may be neglected, the variance may be approximated as:

\[
\sigma^2_{X_t} = (1 - (2t - 2)\delta + \cdots + (1 - 2\delta) + 1) \sigma^2_\varepsilon
\]

\[
= (1 - (t - 1)\delta) t \sigma^2_\varepsilon
\]

and draw your conclusions concerning the properties of the time series.

**Answer:**

\[
X_t = \beta_2^t X_0 + \beta_2^{t-1} \varepsilon_1 + \cdots + \varepsilon_t.
\]

Hence:

\[
\sigma^2_{X_t} = (\beta_2^{2t-2} + \cdots + \beta_2^2 + 1) \sigma^2_\varepsilon
\]

\[
= ((1 - \delta)^{2t-2} + \cdots + (1 - \delta)^2 + 1) \sigma^2_\varepsilon
\]

\[
= ((1 - (2t - 2)\delta) + \cdots + (1 - 2\delta) + 1) \sigma^2_\varepsilon
\]

293
13. Introduction to nonstationary time series

assuming that \( \delta \) is so small that terms involving \( \delta^2 \) may be neglected. (Note that the expansion of \((1 + x)^n\) is \( (1 + nx + \frac{n(n-1)x^2}{2!} + \cdots) \) and if \( x \) is so small that terms involving \( x^2 \) and higher powers of \( x \) may be neglected, the expansion reduces to \((1 + nx)\).) Thus:

\[
\sigma^2_{X_t} = (t - 2\delta(t - 1 + \cdots + 1)) \sigma^2_{\varepsilon}
= (t - \delta t(t - 1)) \sigma^2_{\varepsilon}
= (1 - (t - 1)\delta) t \sigma^2_{\varepsilon}.
\]

It follows that, for finite \( t \), the variance is a function of \( t \) and hence that the series exhibits nonstationary behavior for finite \( t \), even though it is stationary.

13.15 Demonstrate that, for Case (e), \( Y_t \) is determined by:

\[
Y_t = t \beta_1 + \frac{t(t + 1)}{2} \delta + Y_0 + \sum_{s=1}^{t} \varepsilon_s.
\]

This implies that the process is a convex quadratic function of time, implausible empirically.

**Answer:**

The simplest proof is a proof by induction. Suppose that the expression is valid for time \( t \). Then \( Y_{t+1} \) is given by:

\[
Y_t = \beta_1 + Y_t + \delta(t + 1) + \varepsilon_{t+1}
= \beta_1 + \left( t \beta_1 + \frac{t(t + 1)}{2} \delta + Y_0 + \sum_{s=1}^{t} \varepsilon_s \right) + \delta(t + 1) + \varepsilon_{t+1}
= (t + 1)\beta_1 + \frac{(t + 1)(t + 2)}{2} \delta + Y_0 + \sum_{s=1}^{t+1} \varepsilon_s
\]

and so it is valid for time \( t + 1 \). But it is true for time 1. So it is valid for all \( t \geq 1 \).

13.17 Demonstrate that the OLS estimator of \( \delta \) in the model:

\[
Y_t = \beta_1 + \delta t + \varepsilon_t, \quad t = 1, \ldots, T
\]

is hyperconsistent. Show also that it is unbiased in finite samples, despite the fact that \( Y_t \) is nonstationary.

**Answer:**

Let \( \hat{\delta} \) be the OLS estimator of \( \delta \). Following the analysis in Chapter 2, \( \hat{\delta} \) may be decomposed as:

\[
\hat{\delta} = \delta + \sum_{t=1}^{T} a_t u_t
\]

where:

\[
a_t = \frac{t - 0.5T}{\sum_{s=1}^{T} (s - 0.5T)^2}.
\]
Since $a_t$ is deterministic:

$$E(\hat{\delta}) = \delta + \sum_{t=1}^{T} a_t E(u_t) = \delta$$

and the estimator is unbiased. The variance of $\hat{\delta}$, conditional on $T$, is:

$$\sigma^2_{\hat{\delta}} = \frac{\sigma^2_e}{\sum_{t=1}^{T} (t - 0.5(T + 1))^2}.$$ 

Now:

$$\sum_{t=1}^{T} \left( t - \frac{1}{2}(T + 1) \right)^2 = \sum_{t=1}^{T} t^2 - (T + 1) \sum_{t=1}^{T} t + \frac{1}{4} T(T + 1)^2$$

$$= \frac{1}{6} T(T + 1)(2T + 1) - \frac{1}{2} T(T + 1)^2 + \frac{1}{4} T(T + 1)^2$$

$$= \frac{T + 1}{12} (4T^2 + 2T - 6T^2 - 6T + 3T^2 + 3T)$$

$$= \frac{T^3 - T}{12}.$$ 

Thus the variance is (asymptotically) inversely proportional to $T^3$ and the estimator is hyperconsistent.

### 13.6 Answers to the additional exercises

A13.1 The population series exhibits steady growth and is therefore obviously nonstationary. The growth is partly due to an excess of births over deaths and partly due to immigration. The question is whether variations in these factors are likely to be offsetting in the sense that a relatively large birth/death excess one year is somehow automatically counterbalanced by a relatively small one in a subsequent year, or that a relatively large rate of immigration one year stimulates a reaction that leads to a relatively small one later. Such compensating mechanisms do not seem to exist, so trend stationarity may be ruled out. Population is a very good example of an integrated series with the effects of shocks being permanently incorporated in its level.

A13.2 It is difficult to come to any firm conclusion regarding this series. At first sight it looks like a random walk. On closer inspection, you will notice that after an initial decline in the first few years, the series appears to be stationary, with a high degree of correlation. The series is too short to allow one to discriminate between the two possibilities.

A13.3 As expected, given that the series is evidently nonstationary, the coefficient of $LGPOP(-1), -0.05$, is close to zero and not significant. When we difference the
13. Introduction to nonstationary time series

series, the coefficient of $DLGPOP(-1)$ is $-0.16$ and not significant, even at the 5 per cent level. One possibility, which does not seem plausible, is that the population series is $I(2)$. It is more likely that it is $I(1)$, the first difference being stationary but highly autocorrelated.

A13.4 If the process is nonstationary, $\rho = 1$. If it is stationary, it could lie anywhere in the range $-1 < \rho < 1$. We must have a specific value for the null hypothesis. Hence we are forced to use nonstationarity as the null hypothesis, despite the inconvenience of having to compute alternative critical values of $t$.

A13.5 The model has been rewritten:

$$X_t - X_{t-1} = (\rho - 1)X_{t-1} + \varepsilon_t$$

so that the coefficient of $X_{t-1}$ is zero under the null hypothesis of nonstationarity. We see that the null hypothesis is not rejected at any significance level, despite the fact that we know that the series is stationary. However, the estimate of the coefficient of $X_{t-1}$, $-0.38$, is not particularly close to zero. It implies an estimate of 0.67 for $\rho$, close to the actual value. This is a common outcome. Unit root tests generally have low power, making it generally difficult or impossible to discriminate between nonstationary processes and highly autocorrelated stationary processes.

A13.6 Where the hypothetical cointegrating relationship has a constant but no trend, as in the present case, the critical values of $t$ are $-3.34$ and $-3.90$ at the 5 and 1 per cent levels, respectively (Davidson and MacKinnon, 1993). Hence the test indicates that we have a cointegrating relationship only for $DENT$ and then only at the 5 per cent level. However, one knows in advance that the residuals are likely to be highly autocorrelated. Many of the coefficients are greater than 0.2 in absolute terms and perfectly compatible with a hypothesis of highly autocorrelated stationarity.

<table>
<thead>
<tr>
<th>Test of cointegration</th>
<th>$\beta_2$ s.e.</th>
<th>$t$</th>
<th>$\beta_2$ s.e.</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ADM$</td>
<td>$-0.09$ 0.06</td>
<td>$-1.69$</td>
<td>$GASO$</td>
<td>$-0.08$ 0.05</td>
</tr>
<tr>
<td>$BOOK$</td>
<td>$-0.17$ 0.08</td>
<td>$-2.24$</td>
<td>$HOUS$</td>
<td>$-0.31$ 0.12</td>
</tr>
<tr>
<td>$BUSI$</td>
<td>$-0.23$ 0.09</td>
<td>$-2.40$</td>
<td>$LEGL$</td>
<td>$-0.26$ 0.10</td>
</tr>
<tr>
<td>$CLOT$</td>
<td>$-0.41$ 0.13</td>
<td>$-3.17$</td>
<td>$MAGS$</td>
<td>$-0.39$ 0.13</td>
</tr>
<tr>
<td>$DENT$</td>
<td>$-0.51$ 0.15</td>
<td>$-3.51$</td>
<td>$MASS$</td>
<td>$-0.07$ 0.05</td>
</tr>
<tr>
<td>$DOC$</td>
<td>$-0.35$ 0.12</td>
<td>$-2.99$</td>
<td>$OPHT$</td>
<td>$-0.14$ 0.08</td>
</tr>
<tr>
<td>$FLOW$</td>
<td>$-0.22$ 0.10</td>
<td>$-2.14$</td>
<td>$RELG$</td>
<td>$-0.17$ 0.07</td>
</tr>
<tr>
<td>$FOOD$</td>
<td>$-0.29$ 0.11</td>
<td>$-2.61$</td>
<td>$TELE$</td>
<td>$-0.22$ 0.09</td>
</tr>
<tr>
<td>$FURN$</td>
<td>$-0.32$ 0.10</td>
<td>$-3.29$</td>
<td>$TOB$</td>
<td>$-0.16$ 0.10</td>
</tr>
<tr>
<td>$GAS$</td>
<td>$-0.24$ 0.09</td>
<td>$-2.79$</td>
<td>$TOYS$</td>
<td>$-0.17$ 0.09</td>
</tr>
</tbody>
</table>

A13.7 The expected value of $Y_t$ is $\beta_1 t + Y_0$, and thus it is not independent of $t$, one of the conditions for stationarity. Similarly for $Z_t$. However:

$$Y_t - \beta_1 - \beta_2 Z_t = \varepsilon_t$$

and is therefore $I(0)$.
13.6. Answers to the additional exercises

A13.8

\[ Y_t - X_t - Z_t = W_t. \]

Since \( W_t \) is stationary, the left side of the equation is a cointegrating relationship.

A13.9 Two series \( XA_t \) and \( XB_t \) are generated as:

\[
XA_t = RA_t + SA_t \\
XB_t = RB_t + SB_t.
\]

Explain whether it is possible for \( XA_t \) and \( XB_t \) to be stationary.
Explain whether it is possible for them to be cointegrated.

A combination of a nonstationary process and a stationary one is nonstationary. Hence both \( X_A \) and \( X_B \) are nonstationary.

Since the nonstationary components of \( X_A \) and \( X_B \) are unrelated, there is no linear combination that is stationary, and so the series are not cointegrated.

Two series \( YA_t \) and \( YB_t \) are generated as

\[
YA_t = RA_t + SA_t \\
YB_t = RA_t + SB_t.
\]

Explain whether it is possible for \( YA_t \) and \( YB_t \) to be cointegrated.

\[ YA_t - YB_t = SA_t - SB_t. \]

This is a cointegrating relationship for \( YA_t \) and \( YB_t \) since \( SA_t - SB_t \) is stationary.

Two series \( ZA_t \) and \( ZB_t \) are generated as

\[
ZA_t = RA_t + RB_t + SA_t \\
ZB_t = RA_t - RB_t + SB_t.
\]

Explain whether it is possible for \( ZA_t \) and \( ZB_t \) to be stationary.

No linear combination of \( RA_t \) and \( RB_t \) can be stationary since they are independent random walks, and so \( ZA_t \) and \( ZB_t \) are both nonstationary.

Explain whether it is possible for them to be cointegrated.

No linear combination of \( ZA_t \) and \( ZB_t \) can eliminate both \( RA_t \) and \( RB_t \), so there is no cointegrating relationship.
13. Introduction to nonstationary time series
Chapter 14
Introduction to panel data

14.1 Overview

Increasingly, researchers are now using panel data where possible in preference to cross-sectional data. One major reason is that dynamics may be explored with panel data in a way that is seldom possible with crosssectional data. Another is that panel data offer the possibility of a solution to the pervasive problem of omitted variable bias. A further reason is that panel data sets often contain very large numbers of observations and the quality of the data is high. This chapter describes fixed effects regression and random effects regression, alternative techniques that exploit the structure of panel data.

14.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to:

- explain the differences between panel data, cross-sectional data, and time series data
- explain the benefits that can be obtained using panel data
- explain the differences between OLS pooled regressions, fixed effects regressions, and random effects regressions
- explain the potential advantages of the fixed effects model over pooled OLS
- explain the differences between the within-groups, first differences, and least squares dummy variables variants of the fixed effects model
- explain the assumptions required for the use of the random effects model
- explain the advantages of the random effects model over the fixed effects model when the assumptions are valid
- explain how to use a Durbin–Wu–Hausman test to determine whether the random effects model may be used instead of the fixed effects model.
14. Introduction to panel data

14.3 Additional exercises

A14.1 The NLSY2000 data set contains the following data for a sample of 2,427 males and 2,392 females for the years 1980–2000: years of work experience, \( EXP \), years of schooling, \( S \), and age, \( AGE \). A researcher investigating the impact of schooling on willingness to work regresses \( EXP \) on \( S \), including potential work experience, \( PWE \), as a control. \( PWE \) was defined as:

\[
PWE = AGE - S - 5.
\]

The following regressions were performed for males and females separately:

1. an ordinary least squares (OLS) regression pooling the observations
2. a within-groups fixed effects regression
3. a random effects regression.

The results of these regressions are shown in the table below. Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>FE</td>
<td>RE</td>
<td>OLS</td>
</tr>
<tr>
<td>( S )</td>
<td>0.78 (0.01)</td>
<td>0.65 (0.01)</td>
<td>0.72 (0.01)</td>
<td>0.89 (0.01)</td>
</tr>
<tr>
<td>( PWE )</td>
<td>0.83 (0.003)</td>
<td>0.94 (0.001)</td>
<td>0.94 (0.001)</td>
<td>0.74 (0.004)</td>
</tr>
<tr>
<td>constant</td>
<td>−10.16 (0.09)</td>
<td>dropped</td>
<td>−10.56 (0.14)</td>
<td>−11.11 (0.12)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.79 —</td>
<td>—</td>
<td>0.71 —</td>
<td>—</td>
</tr>
<tr>
<td>( n )</td>
<td>24,057</td>
<td>24,057</td>
<td>24,057</td>
<td>18,758</td>
</tr>
<tr>
<td>DHW ( \chi^2(2) )</td>
<td>10.76</td>
<td></td>
<td>1.43</td>
<td></td>
</tr>
</tbody>
</table>

- Explain why the researcher included \( PWE \) as a control.
- Evaluate the results of the Durbin–Wu–Hausman tests.
- For males and females separately, explain the differences in the coefficients of \( S \) in the OLS and FE regressions.
- For males and females separately, explain the differences in the coefficients of \( PWE \) in the OLS and FE regressions.

A14.2 Using the NLSY2000 data set, a researcher fits OLS and fixed effects regressions of the logarithm of hourly wages on schooling, years of work experience, \( EXP \), \( ASVABC \) score, and dummies \( MALE \), \( ETHBLACK \), and \( ETHHISP \) for being male, black, or hispanic. Schooling was split into years of high school, \( SH \), and years of college, \( SC \). The results are shown in the table below, with standard errors placed in parentheses.
If an individual reported being in high school or college, the observation for that individual for that year was deleted from the sample. As a consequence, the observations for most individuals in the sample begin when the formal education of that individual has been completed. However, a small minority of individuals, having apparently completed their formal education and having taken employment, subsequently resumed their formal education, either to complete high school with a general educational development (GED) degree equivalent to the high school diploma, or to complete one or more years of college.

- Discuss the differences in the estimates of the coefficient of \( SH \).
- Discuss the differences in the estimates of the coefficient of \( SC \).

A researcher has data on \( G \), the average annual rate of growth of GDP 2001–2005, and \( S \), the average years of schooling of the workforce in 2005, for 28 European Union countries. She believes that \( G \) depends on \( S \) and on \( E \), the level of entrepreneurship in the country, and a disturbance term \( u \):

\[
G = \beta_1 + \beta_2 S + \beta_3 E + u. \tag{1}
\]

\( u \) may be assumed to satisfy the usual regression model assumptions. Unfortunately the researcher does not have data on \( E \).

- Explain intuitively and mathematically the consequences of performing a simple regression of \( G \) on \( S \). For this purpose \( S \) and \( E \) may be treated as nonstochastic variables.

The researcher does some more research and obtains data on \( G^* \), the average annual rate of growth of GDP 1996–2000, and \( S^* \), the average years of schooling of the workforce in 2000, for the same countries. She thinks that she
14. Introduction to panel data

can deal with the unobservable variable problem by regressing \( \Delta G \), the change in \( G \), on \( \Delta S \), the change in \( S \), where:

\[
\begin{align*}
\Delta G &= G - G^* \\
\Delta S &= S - S^*
\end{align*}
\]

assuming that \( E \) would be much the same for each country in the two periods. She fits the equation:

\[
\Delta G = \delta_1 + \delta_2 \Delta S + w \tag{2}
\]

where \( w \) is a disturbance term that satisfies the usual regression model assumptions.

• Compare the properties of the estimators of the coefficient of \( S \) in (1) and of the coefficient of \( \Delta S \) in (2).

• Explain why in principle you would expect the estimate of \( \delta_1 \) in (2) not to be significant. Suppose that nevertheless the researcher finds that the coefficient is significant. Give two possible explanations.

Random effects regressions have potential advantages over fixed effect regressions.

• Could the researcher have used a random effects regression in the present case?

A14.4 A researcher has the following data for 3,763 respondents in the National Longitudinal Survey of Youth 1979–2000: hourly earnings in dollars in 1994 and 2000, years of schooling as recorded in 1994 and 2000, and years of work experience as recorded in 1994 and 2000. The respondents were aged 14–21 in 1979, so they were aged 29–36 in 1994 and 35–42 in 2000. 371 of the respondents had increased their formal schooling between 1994 and 2000, 210 by one year, 101 by two years, 47 by three years, and 13 by more than three years, mostly at college level in non-degree courses. The researcher performs the following regressions:

1. the logarithm of hourly earnings in 1994 on schooling and work experience in 1994
2. the logarithm of hourly earnings in 2000 on schooling and work experience in 2000
3. the change in the logarithm of hourly earnings from 1994 to 2000 on the changes in schooling and work experience in that interval.

The results are shown in columns (1) – (3) in the table (t statistics in parentheses), and are presented at a seminar.
### 14.3. Additional exercises

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Schooling</td>
<td>0.114 (30.16)</td>
<td>0.116 (28.99)</td>
<td>—</td>
<td>0.108 (24.53)</td>
<td>—</td>
</tr>
<tr>
<td>Experience</td>
<td>0.052 (18.81)</td>
<td>0.038 (14.59)</td>
<td>—</td>
<td>0.037 (14.10)</td>
<td>—</td>
</tr>
<tr>
<td>Cognitive ability score</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.004 (4.79)</td>
<td>—</td>
</tr>
<tr>
<td>Male</td>
<td>0.214 (12.03)</td>
<td>0.229 (11.77)</td>
<td>—</td>
<td>0.230 (11.88)</td>
<td>—</td>
</tr>
<tr>
<td>Black</td>
<td>-0.149 (-5.23)</td>
<td>-0.199 (-6.44)</td>
<td>—</td>
<td>-0.167 (-5.29)</td>
<td>—</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.039 (1.11)</td>
<td>0.053 (1.38)</td>
<td>—</td>
<td>0.071 (1.84)</td>
<td>—</td>
</tr>
<tr>
<td>Change in schooling</td>
<td>—</td>
<td>—</td>
<td>0.090 (5.00)</td>
<td>—</td>
<td>-0.006 (-0.16)</td>
</tr>
<tr>
<td>Change in experience</td>
<td>—</td>
<td>—</td>
<td>0.024 (2.75)</td>
<td>—</td>
<td>0.003 (0.15)</td>
</tr>
<tr>
<td>constant</td>
<td>4.899 (74.59)</td>
<td>5.023 (65.02)</td>
<td>0.102 (2.13)</td>
<td>4.966 (63.69)</td>
<td>0.389 (3.05)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.265</td>
<td>0.243</td>
<td>0.007</td>
<td>0.248</td>
<td>0.0002</td>
</tr>
<tr>
<td>$n$</td>
<td>3,763</td>
<td>3,763</td>
<td>3,763</td>
<td>3,763</td>
<td>371</td>
</tr>
</tbody>
</table>

- The researcher is unable to explain why the coefficient of the change in schooling in regression (3) is so much lower than the schooling coefficients in (1) and (2). Someone says that it is because he has left out relevant variables such as cognitive ability, region of residence, etc, and the coefficients in (1) and (2) are therefore biased. Someone else says that cannot be the explanation because these variables are also omitted from regression (3). Explain what would be your view.

- He runs regressions (1) and (2) again, adding a measure of cognitive ability. The results for the 2000 regression are shown in column (4). The results for 1994 were very similar. Discuss possible reasons for the fact that the estimate of the schooling coefficient differs from those in (2) and (3).

- Someone says that the researcher should not have included a constant in regression (3). Explain why she made this remark and assess whether it is valid.

- Someone else at the seminar says that the reason for the relatively low coefficient of schooling in regression (3) is that it mostly represented non-degree schooling. Hence one would not expect to find the same relationship between schooling and earnings as for the regular pre-employment schooling of young people. Explain in general terms what investigation the researcher should undertake in response to this suggestion.

- Another person suggests that the small minority of individuals who went back to school or college in their thirties might have characteristics different from...
14. Introduction to panel data

those of the individuals who did not, and that this could account for a
different coefficient. Explain in general verbal terms what investigation the
researcher should undertake in response to this suggestion.

- Finally, another person says that it might be a good idea to look at the
relationship between earnings and schooling for the subsample who went back
to school or college, restricting the analysis to these 371 individuals. The
researcher responds by running the regression for that group alone. The result
is shown in column (5) in the table. The researcher also plots a scatter
diagram, reproduced below, showing the change in the logarithm of earnings
and the change in schooling. For those with one extra year of schooling, the
mean change in log earnings was 0.40. For those with two extra years, 0.37.
For those with three extra years, 0.47. What conclusions might be drawn from
the regression results?

A14.5 In the discussion of the DWH test, it was stated that the test compares the
coefficients of those variables not dropped in the FE regression. Explain why the
constant is not included in the comparison.

14.4 Answer to the starred exercise in the textbook

14.9 The *NLSY2000* data set contains the following data for a sample of 2,427 males
and 2,392 females for the years 1980–2000: weight in pounds, years of schooling,
age, marital status in the form of a dummy variable *MARRIED* defined to be 1 if
the respondent was married, 0 if single, and height in inches. Hypothesizing that
weight is influenced by schooling, age, marital status, and height, the following
regressions were performed for males and females separately:

1. an ordinary least squares (OLS) regression pooling the observations
2. a within-groups fixed effects regression
3. a random effects regression.
The results of these regressions are shown in the table. Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Males OLS</th>
<th>Males FE</th>
<th>Males RE</th>
<th>Females OLS</th>
<th>Females FE</th>
<th>Females RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of schooling</td>
<td>-0.98  (0.09)</td>
<td>-0.02  (0.23)</td>
<td>-0.45  (0.16)</td>
<td>-1.95  (0.12)</td>
<td>-0.60  (0.27)</td>
<td>-1.25  (0.18)</td>
</tr>
<tr>
<td>Age</td>
<td>1.61  (0.04)</td>
<td>1.64  (0.02)</td>
<td>1.65  (0.02)</td>
<td>2.03  (0.05)</td>
<td>1.66  (0.03)</td>
<td>1.72  (0.03)</td>
</tr>
<tr>
<td>Married</td>
<td>3.70  (0.48)</td>
<td>2.92  (0.33)</td>
<td>3.00  (0.32)</td>
<td>-8.27  (0.59)</td>
<td>3.08  (0.46)</td>
<td>1.98  (0.44)</td>
</tr>
<tr>
<td>Height</td>
<td>5.07  (0.08)</td>
<td>dropped</td>
<td>4.95  (0.18)</td>
<td>3.48  (0.10)</td>
<td>dropped</td>
<td>3.38  (0.21)</td>
</tr>
<tr>
<td>constant</td>
<td>-209.52 (5.39)</td>
<td>dropped</td>
<td>-209.81 (12.88)</td>
<td>-105.90  (6.62)</td>
<td>dropped</td>
<td>-107.61 (13.43)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.27  —</td>
<td>—</td>
<td>0.17  —</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$n$</td>
<td>17,299 17,299</td>
<td>17,299 17,299</td>
<td>17,299 13,160</td>
<td>13,160 13,160</td>
<td>13,160 13,160</td>
<td></td>
</tr>
</tbody>
</table>

DWH $\chi^2(3)$ | 7.22 | 92.94

Explain why height is excluded from the FE regression.

Evaluate, for males and females separately, whether the fixed effects or random effects model should be preferred.

For males and females separately, compare the estimates of the coefficients in the OLS and FE models and attempt to explain the differences.

Explain in principle how one might test whether individual-specific fixed effects jointly have significant explanatory power, if the number of individuals is small. Explain why the test is not practical in this case.

**Answer:**

Height is constant over observations. Hence, for each individual:

$$\overline{\text{HEIGHT}}_{it} - \text{HEIGHT}_i = 0$$

for all $t$, where $\overline{\text{HEIGHT}}_i$ is the mean height for individual $i$ for the observations for that individual. Hence height has to be dropped from the regression model.

The critical value of chi-squared, with three degrees of freedom, is 7.82 at the 5 percent level and 16.27 at the 0.1 percent level. Hence there is a possibility that the random effects model may be appropriate for males, but it is definitely not appropriate for females.

**Males**

The OLS regression suggests that schooling has a small (one pound less per year of schooling) but highly significant negative effect on weight. The fixed effects regression eliminates the effect, indicating that an unobserved effect is responsible: males with unobserved qualities that have a positive effect on educational attainment, controlling for other measured variables, have lower weight as a consequence of the same unobserved qualities. We cannot compare estimates of the effect of height since it is dropped from the FE regression. The effect of age is the same in the two regressions. There is a small but highly significant positive effect of being married, the OLS estimate possibly being inflated by an unobserved effect.
14. Introduction to panel data

Females

The main, and very striking, difference is in the marriage coefficient. The OLS regression suggests that marriage reduces weight by eight pounds, a remarkable amount. The FE regression suggests the opposite, that marriage leads to an increase in weight that is similar to that for males. The clear implication is that women who weigh less are relatively successful in the marriage market, but once they are married they put on weight.

For schooling the story is much the same as for males, except that the OLS coefficient is much larger and the coefficient remains significant at the 5 percent level in the FE regression. The effect of age appears to be exaggerated in the OLS regression, for reasons that are not obvious.

One might test whether individual-specific fixed effects jointly have significant explanatory power by performing a LSDV regression, eliminating the intercept in the model and adding a dummy variable for each individual. One would compare RSS for this regression with that for the regression without the dummy variables, using a standard $F$ test. In the present case it is not a practical proposition because there are more than 17,000 males and 13,000 females.

14.5 Answers to the additional exercises

A14.1 Explain why the researcher included PWE as a control.

Clearly actual work experience is positively influenced by $PWE$. Omitting it would cause the coefficient of $S$ to be biased downwards since $PWE$ and $S$ are negatively correlated.

Evaluate the results of the Durbin–Wu–Hausman tests

With two degrees of freedom, the critical value of chi-squared is 5.99 at the 5 percent level and 9.21 at the 1 percent level. Thus the random effects model is rejected for males but seemingly not for females.

For males and females separately, explain the differences in the coefficients of $S$ in the OLS and FE regressions.

For both sexes the OLS estimate is greater than the FE estimate. One possible reason is that some unobserved characteristics, for example drive, are positively correlated with both acquiring schooling, and seeking and gaining employment.

For males and females separately, explain the differences in the coefficients of PWE in the OLS and FE regressions.

Since $S$ and $PWE$ are negatively correlated, these same unobserved characteristics would cause the OLS estimate of the coefficient of $PWE$ to be biased downwards.

A14.2 First, note that the DWH statistic is significant at the 5 per cent level (critical value 7.82) but not at the 1 per cent level (critical value 11.35).

The coefficients of $SH$ and $SC$ in the OLS regression is an estimate of the impact of variations in years of high school and years of college among all the individuals in the sample. Most individuals in fact completed high school and so had $SH = 12$. 

© Christopher Dougherty, 2016. All rights reserved. Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
However, a small minority did not and this variation made possible the estimation of the \(SH\) coefficient. The majority of the remainder did not complete any years of college and therefore had \(SC = 0\), but a substantial minority did have a partial or complete college education, some even pursuing postgraduate studies, and this variation made possible the estimation of the \(SC\) coefficient.

Most individuals completed their formal education before entering employment. For them, \(SH_t = \bar{SH}_i\) for all \(t\) and hence \(SH_t - \bar{SH}_i = 0\) for all \(t\). As a consequence, the observations for such individuals provide no variation in the \(SH\) variable. Likewise they provide no variation in the \(SC\) variable. If all observations pertained to such individuals, schooling would be washed out in the FE regression along with other unchanging characteristics such as sex, ethnicity, and \(ASVABC\) score. The schooling coefficients in the FE regression therefore relate to those individuals who returned to formal education after a break in which they found employment.

The fact that these individuals account for a relatively small proportion of the observations in the data set has an adverse effect on the precision of the FE estimates of the coefficients of \(SH\) and \(SC\). This is reflected in standard errors that are much larger than those obtained in the OLS pooled regression.

*Discuss the differences in the estimates of the coefficient of \(SH\).*

Most of the variation in \(SH\) in the FE regressions come from individuals earning the GED degree. This degree provides an opportunity for high school drop-outs to make good their shortfall by taking courses and passing the examinations required for this diploma. These courses may be civilian or military adult education classes, but very often they are programmes offered to those in jail. In principle the GED should be equivalent to the high school diploma, but there is some evidence that standards are sometimes lower. The results in the table appear to corroborate this view. The OLS regression indicates that a year of high school raises earnings by 2.6 per cent, with the coefficient being highly significant, whereas the FE coefficient indicates that the effect is only 0.5 per cent and not significant.

*Discuss the differences in the estimates of the coefficient of \(SC\).*

Some of the variation in \(SC\) in the FE regressions comes from individuals entering employment for a year or two after finishing high school and then going to college, resuming their formal education. However, most comes from individuals returning to college for a year or two after having been employment for a number of years. A typical example is a high school graduate who has settled down in an occupation and who has then decided to upgrade his or her professional skills by taking a two-year associate of arts degree. Similarly one encounters college graduates who upgrade to masters level after having worked for some time. One would expect such students to be especially well motivated – they are often undertaking studies that are relevant to an established career, and they are often bearing high opportunity costs from loss of earnings while studying – and accordingly one might expect the payoff in terms of increased earnings to be relatively high. This seems to be borne out in a comparison of the OLS and FE estimates of the coefficient of \(SC\), though the difference is not dramatic.

On the surface, this exercise appeared to be about how one might use FE to eliminate the bias in OLS pooled regression caused by unobserved effects. Has the analysis been successful in this respect? Absolutely not. In particular, the apparent
14. Introduction to panel data

...conclusion that high school education has virtually no effect on earnings should not be taken at face value. The reason is that the issue of biases attributable to unobserved effects has been overtaken by the much more important issue of the difference in the interpretation of the $SH$ and $SC$ coefficients discussed in the exercise. This illustrates a basic point in econometrics: understanding the context of the data is often just as important as being proficient at technical analysis.

A14.3 Explain intuitively and mathematically the consequences of performing a simple regression of $G$ on $S$. For this purpose $S$ and $E$ may be treated as nonstochastic variables.

If one fits the regression:

$$\hat{G} = \hat{\beta}_1 + \hat{\beta}_2 S$$

then

$$\hat{\beta}_2 = \frac{\sum (S_i - \bar{S})(G_i - \bar{G})}{\sum (S_i - \bar{S})^2}$$

$$= \frac{\sum (S_i - \bar{S})((\beta_1 + \beta_2 S_i + \beta_3 E_i + u_i) - (\beta_1 + \beta_2 \bar{S} + \beta_2 \bar{E} + \bar{u}))}{\sum (S_i - \bar{S})^2}$$

$$= \beta_2 + \beta_3 \frac{\sum (S_i \bar{S})(E_i - \bar{E})}{\sum (S_i - \bar{S})^2} + \frac{\sum (S_i - \bar{S})(u_i - \bar{u})}{\sum (S_i - \bar{S})^2}.$$

Taking expectations, and making use of the invitation to treat $S$ and $E$ as nonstochastic:

$$E(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{\sum (S_i \bar{S})(E_i - \bar{E})}{\sum (S_i - \bar{S})^2} + \frac{\sum (S_i - \bar{S})E(u_i - \bar{u})}{\sum (S_i - \bar{S})^2}$$

$$= \beta_2 + \beta_3 \frac{\sum (S_i - \bar{S})(E_i - \bar{E})}{\sum (S_i - \bar{S})^2}.$$

Hence the estimator is biased unless $S$ and $E$ happen to be uncorrelated in the sample. As a consequence, the standard errors will be invalid.

Compare the properties of the estimators of the coefficient of $S$ in (1) and of the coefficient of $\Delta S$ in (2).

Given (1), the differenced model should have been:

$$\Delta G = \delta_2 \Delta S + w$$

where $w = u - u^*$. The estimator of the coefficient of $\Delta S$ in (2) should be unbiased, while that of $S$ in (1) will be subject to omitted variable bias. However:
14.5. Answers to the additional exercises

- it is possible that the bias in (1) may be small. This would be the case if $E$ were a relatively unimportant determinant of $G$ or if its correlation with $S$ were low.

- it is possible that the variance in $\Delta S$ is smaller than that of $S$. This would be the case if $S$ were changing slowly in each country, or if the rate of change of $S$ were similar in each country.

Thus there may be a trade-off between bias and variance and it is possible that the estimator of $\beta_2$ using specification (1) could actually be superior according to some criterion such as the mean square error. It should be noted that the inclusion of $\delta_1$ in (2) will make the estimation of $\delta_2$ even less efficient.

Explain why in principle you would expect the estimate of $\delta_1$ in (2) not to be significant. Suppose that nevertheless the researcher finds that the coefficient is significant. Give two possible explanations.

If specification (1) is correct, there should be no intercept in (2) and for this reason the estimate of the intercept should not be significantly different from zero. If it is significant, this could have occurred as a matter of Type I error. Alternatively, it might indicate a shift in the relationship between the two time periods. Suppose that (1) should have included a dummy variable set equal to 0 in the first time period and 1 in the second. $\hat{\delta}_1$ would then be an estimate of its coefficient.

Could the researcher have used a random effects regression in the present case?

Random effects requires the sample to be drawn randomly from a population and for unobserved effects to be uncorrelated with the regressors. The first condition is not satisfied here, so random effects would be inappropriate.

A14.4 The researcher is unable to explain why the coefficient of the change in schooling in regression (3) is so much lower than the schooling coefficients in (1) and (2).

Someone says that it is because he has left out relevant variables such as cognitive ability, region of residence, etc, and the coefficients in (1) and (2) are therefore biased. Someone else says that cannot be the explanation because these variables are also omitted from regression (3). Explain what would be your view.

Suppose that the true model is:

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 EXP + \beta_4 ASVABC + \beta_5 MALE$$

$$+ \beta_6 ETHBLACK + \beta_7 ETHHISP + \beta_8 X_8 + u$$

where $X_8$ is some further fixed characteristic of the respondent. ASVABC and $X_8$ are absent from regressions (1) and (2) and so those regressions will be subject to omitted variable bias. In particular, since ASVABC is likely to be positively correlated with $S$, and to have a positive coefficient, its omission will tend to bias the coefficient of $S$ upwards.

However, if the specification is valid for both 1994 and 2000 and unchanged, one can eliminate the omitted variable bias by taking first differences as in regression (3):

$$\Delta LGEARN = \beta_2 \Delta S + \beta_3 \Delta EXP + \Delta u.$$  

By fitting this specification one should obtain unbiased estimates of the coefficients of schooling and experience, and the former should therefore be smaller than in (1)
and (2). Note that all the fixed characteristics have been washed out. The suggestion that ASVABC should have been included in (3) is therefore incorrect. Note that (3) should not have included an intercept. This is discussed later in the question.

He runs regressions (1) and (2) again, adding a measure of cognitive ability. The results for the 2000 regression are shown in column (4). The results for 1994 were very similar. Discuss possible reasons for the fact that the estimate of the schooling coefficient differs from those in (2) and (3).

The estimate of the coefficient of $S$ differs from that in (2) because the omitted variable bias attributable to the omission of ASVABC in that specification has now been corrected. However it is still biased if $X_8$ (representing other omitted characteristics) is a determinant of earnings and is correlated with $S$. This partial rectification of the omitted variable problem accounts for the fact that the coefficient of $S$ in (4) lies between those in (2) and (3).

Someone says that the researcher should not have included a constant in regression (3). Explain why she made this remark and assess whether it is valid.

Given the specification in (1) and (2), there should have been no intercept in the first differences specification (3). One would therefore expect the estimate of the intercept to be somewhere near zero in the sense of not being significantly different from it. Nevertheless, it is significantly different at the 5 percent level. However, suppose that the relationship shifted between 1994 and 2000, and that the shift could be represented by a dummy variable $D$ equal to zero in 1994 and 1 in 2000, with coefficient $\delta$. Then (3) should have an intercept $\delta$. Its estimate, 0.102, suggests that earnings grew by 10 percent from 1994 to 2000, holding other factors constant. This seems entirely reasonable, perhaps even a little low.

Alternatively, the apparently significant $t$ statistic might have arisen as a matter of Type I error.

Someone else at the seminar says that the reason for the relatively low coefficient of schooling in regression (3) is that it mostly represented non-degree schooling. Hence one would not expect to find the same relationship between schooling and earnings as for the regular preemployment schooling of young people. Explain in general verbal terms what investigation the researcher should undertake in response to this suggestion.

Divide $S$ into two variables, schooling as of 1994 and extra schooling as of 2000, with separate coefficients. Then use a standard $F$ test (or $t$ test) of a restriction to test whether the coefficients are significantly different.

Another person suggests that the small minority of individuals who went back to school or college in their thirties might have characteristics different from those of the individuals who did not, and that this could account for a different coefficient. Explain in general verbal terms what investigation the researcher should undertake in response to this suggestion.

The issue is sample selection bias and an appropriate procedure would be that proposed by Heckman. One would use probit analysis with an appropriate set of determinants to model the decision to return to school between 1994 and 2000, and a regression model to explain variations in the logarithm of earnings of those
respondents who do return to school, linking the two models by allowing their disturbance terms to be correlated. One would test whether the estimate of this correlation is significantly different from zero.

Finally, another person says that it might be a good idea to look at the relationship between earnings and schooling for the subsample who went back to school or college, restricting the analysis to these 371 individuals. The researcher responds by running the regression for that group alone. The result is shown in column (5) in the table. The researcher also plots a scatter diagram, reproduced below, showing the change in the logarithm of earnings and the change in schooling. For those with one extra year of schooling, the mean change in log earnings was 0.40. For those with two extra years, 0.37. For those with three extra years, 0.47. What conclusions might be drawn from the regression results?

The schooling coefficient is effectively zero! [These are real data, incidentally.] The scatter diagram shows why. Irrespective of whether the respondent had one, two, or three years of extra schooling, the gain is about the same, on average. (These are the only categories with large numbers of observations, given the information at the beginning of the question, confirmed by the scatter diagram.) So the results indicate that the fact of going back to school, rather than the duration of the schooling, is the relevant determinant of the change in earnings. The intercept indicates that this subsample on average increased their earnings between 1994 and 2000 by 38.9 percent. (As a first approximation. The actual proportion would be better estimated as $e^{0.389} - 1 = 0.476.$) This figure is confirmed by the diagram, and it would appear to be much greater than the effect of regular schooling. One explanation could be sample selection bias, as already discussed. A more likely possibility is that the respondents were presented with opportunities to increase their earnings substantially if they undertook certain types of formal course, and they took advantage of these opportunities.

A14.5 In a random effects regression, the interpretation of an intercept is not affected by the estimation technique. In a fixed effects regression, the intercept is washed out. Hence there is no basis for a comparison. In general, the model is fitted without an intercept. The only case where an intercept should be included is in first-differences fixed effects estimation of a model containing a deterministic trend. For example, suppose one is fitting the model:

$$Y_{it} = \beta_1 + \beta_2 X_{it} + \delta t + u_{it}.$$  

For individual $i$ in the previous time period, one has:

$$Y_{i,t-1} = \beta_1 + \beta_2 X_{i,t-1} + \delta(t - 1) + u_{i,t-1}.$$  

Subtracting, one obtains:

$$Y_{it} - Y_{i,t-1} = \beta_2(X_{it} - X_{i,t-1}) + \delta + u_{it} - u_{i,t-1}.$$  

The model now does have an intercept, but its meaning is different from that in the original specification. It now provides an estimate of $\delta$, not $\beta_1$.  

© Christopher Dougherty, 2016. All rights reserved.  
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.
14. Introduction to panel data
Chapter 15
Regression analysis with linear algebra primer

15.1 Overview

This primer is intended to provide a mathematical bridge to a master’s level course that uses linear algebra for students who have taken an undergraduate econometrics course that does not. Why should we make the mathematical shift? The most immediate reason is the huge double benefit of allowing us to generalise the core results to models with many explanatory variables while simultaneously permitting a great simplification of the mathematics. This alone justifies the investment in time – probably not more than ten hours – required to acquire the necessary understanding of basic linear algebra.

In fact, one could very well put the question the other way. Why do introductory econometrics courses not make this investment and use linear algebra from the start? Why do they (almost) invariably use ordinary algebra, leaving students to make the switch when they take a second course?

The answer to this is that the overriding objective of an introductory econometrics course must be to encourage the development of a solid intuitive understanding of the material and it is easier to do this with familiar, everyday algebra than with linear algebra, which for many students initially seems alien and abstract. An introductory course should ensure that at all times students understand the purpose and value of what they are doing. This is far more important than proofs and for this purpose it is usually sufficient to consider models with one, or at most two, explanatory variables. Even in the relatively advanced material, where we are forced to consider asymptotics because we cannot obtain finite-sample results, the lower-level mathematics holds its own. This is especially obvious when we come to consider finite-sample properties of estimators when only asymptotic results are available mathematically. We invariably use a simple model for a simulation, not one that requires a knowledge of linear algebra.

These comments apply even when it comes to proofs. It is usually helpful to see a proof in miniature where one can easily see exactly what is involved. It is then usually sufficient to know that in principle it generalises, without there being any great urgency to see a general proof. Of course, the linear algebra version of the proof will be general and often simpler, but it will be less intuitively accessible and so it is useful to have seen a miniature proof first. Proofs of the unbiasedness of the regression coefficients under appropriate assumptions are obvious examples.

At all costs, one wishes to avoid the study of econometrics becoming an extended exercise in abstract mathematics, most of which practitioners will never use again. They will use regression applications and as long as they understand what is happening in
15. Regression analysis with linear algebra primer

principle, the actual mechanics are of little interest.

This primer is not intended as an exposition of linear algebra as such. It assumes that a
basic knowledge of linear algebra, for which there are many excellent introductory
textbooks, has already been acquired. For the most part, it is sufficient that you should
know the rules for multiplying two matrices together and for deriving the inverse of a
square matrix, and that you should understand the consequences of a square matrix
having a zero determinant.

15.2 Notation

Matrices and vectors will be written bold, upright, matrices upper case, for example \( A \),
and vectors lower case, for example \( b \). The transpose of a matrix will be denoted by a
prime, so that the transpose of \( A \) is \( A' \), and the inverse of a matrix will be denoted by a
superscript \(-1\), so that the inverse of \( A \) is \( A^{-1} \).

15.3 Test exercises

Answers to all of the exercises in this primer will be found at its end. If you are unable
to answer the following exercises, you need to spend more time learning basic matrix
algebra before reading this primer. The rules in Exercises 3–5 will be used frequently
without further explanation.

1. Demonstrate that the inverse of the inverse of a matrix is the original matrix.

2. Demonstrate that if a (square) matrix possesses an inverse, the inverse is unique.

3. Demonstrate that, if \( A = BC \), \( A' = C'B' \).

4. Demonstrate that, if \( A = BC \), \( A^{-1} = C^{-1}B^{-1} \), provided that \( B^{-1} \) and \( C^{-1} \) exist.

5. Demonstrate that \( [A']^{-1} = [A^{-1}]' \).

15.4 The multiple regression model

The most obvious benefit from switching to linear algebra is convenience. It permits an
elegant simplification and generalisation of much of the mathematical analysis
associated with regression analysis. We will consider the general multiple regression
model:

\[
Y_i = \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i
\]

where the second subscript identifies the variable and the first the observation. In the
textbook, as far as the fourth edition, the subscripts were in the opposite order. The
reason for the change of notation here, which will be adopted in the next edition of the
textbook, is that it is more compatible with a linear algebra treatment.

314
Equation (1) is a row relating to observation \( i \) in a sample of \( n \) observations. The entire layout would be:

\[
\begin{bmatrix}
Y_1 \\
\vdots \\
Y_i \\
\vdots \\
Y_n
\end{bmatrix} =
\begin{bmatrix}
\beta_1 X_{11} + \cdots + \beta_j X_{ij} + \cdots + \beta_k X_{1k} \\
\vdots \\
\beta_1 X_{i1} + \cdots + \beta_j X_{ij} + \cdots + \beta_k X_{ik} \\
\vdots \\
\beta_1 X_{n1} + \cdots + \beta_j X_{nj} + \cdots + \beta_k X_{nk}
\end{bmatrix}
+ 
\begin{bmatrix}
u_1 \\
\vdots \\
u_i \\
\vdots \\
u_n
\end{bmatrix}.
\]

This, of course, may be written in linear algebra form as:

\[ y = X\beta + u \]  \hspace{1cm} (2)

where:

\[
y =
\begin{bmatrix}
Y_1 \\
\vdots \\
Y_i \\
\vdots \\
Y_n
\end{bmatrix}, \\
X =
\begin{bmatrix}
X_{11} \cdots X_{1j} \cdots X_{1k} \\
\vdots \\
X_{i1} \cdots X_{ij} \cdots X_{ik} \\
\vdots \\
X_{n1} \cdots X_{nj} \cdots X_{nk}
\end{bmatrix}, \\
\beta =
\begin{bmatrix}
\beta_1 \\
\vdots \\
\beta_i \\
\vdots \\
\beta_n
\end{bmatrix}, \text{ and } u =
\begin{bmatrix}
u_1 \\
\vdots \\
u_i \\
\vdots \\
u_n
\end{bmatrix}
\]

with the first subscript of \( X_{ij} \) relating to the row and the second to the column, as is conventional with matrix notation. This was the reason for the change in the order of the subscripts in equation (1).

Frequently, it is convenient to think of the matrix \( X \) as consisting of a set of column vectors:

\[ X = [x_1 \cdots x_j \cdots x_k] \]

where:

\[ x_j =
\begin{bmatrix}
X_{1j} \\
\vdots \\
X_{ij} \\
\vdots \\
X_{nj}
\end{bmatrix}. \]

\( x_j \) is the set of observations relating to explanatory variable \( j \). It is written lower case, bold, not italic because it is a vector.

\subsection*{15.5 The intercept in a regression model}

As described above, there is no special intercept term in the model. If, as is usually the case, one is needed, it is accommodated within the matrix framework by including an \( X \) variable, typically placed as the first, with value equal to 1 in all observations:

\[ x_1 =
\begin{bmatrix}
1 \\
\vdots \\
1 \\
\vdots \\
1
\end{bmatrix}. \]
15. Regression analysis with linear algebra primer

The coefficient of this unit vector is the intercept in the regression model. If it is included, and located as the first column, the $X$ matrix becomes:

\[
X = \begin{bmatrix}
1 & X_{12} & \cdots & X_{1j} & \cdots & X_{1k} \\
\vdots & \vdots & & \vdots & & \vdots \\
1 & X_{i2} & \cdots & X_{ij} & \cdots & X_{ik} \\
\vdots & \vdots & & \vdots & & \vdots \\
1 & X_{n2} & \cdots & X_{nj} & \cdots & X_{nk}
\end{bmatrix} = [1 \ x_2 \ \cdots \ x_j \ \cdots \ x_k].
\]

15.6 The OLS regression coefficients

Using the matrix and vector notation, we may write the fitted equation:

\[
\hat{Y}_i = \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}
\]
as:

\[
\hat{y} = X\hat{\beta}
\]
with obvious definitions of $\hat{y}$ and $\hat{\beta}$. Then we may define the vector of residuals as:

\[
\hat{u} = y - \hat{y} = y - X\hat{\beta}
\]

and the residual sum of squares as:

\[
RSS = \hat{u}'\hat{u} = (y - X\hat{\beta})'(y - X\hat{\beta})
\]

\[
= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}
\]

\[
= y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}
\]

($y'X\hat{\beta} = \hat{\beta}'X'y$ since it is a scalar.) The next step is to obtain the normal equations:

\[
\frac{\partial RSS}{\partial \hat{\beta}_j} = 0
\]
for $j = 1, \ldots, k$ and solve them (if we can) to obtain the least squares coefficients. Using linear algebra, the normal equations can be written:

\[
X'X\hat{\beta} - X'y = 0.
\]
The derivation is straightforward but tedious and has been consigned to Appendix A. $X'X$ is a square matrix with $k$ rows and columns. If assumption A.2 is satisfied (that it is not possible to write one $X$ variable as a linear combination of the others), $X'X$ has an inverse and we obtain the OLS estimator of the coefficients:

\[
\hat{\beta} = [X'X]^{-1}X'y. \quad (3)
\]

Exercises

6. If $Y = \beta_1 + \beta_2 X + u$, obtain the OLS estimators of $\beta_1$ and $\beta_2$ using (3).

7. If $Y = \beta_2 + u$, obtain the OLS estimator of $\beta_2$ using (3).

8. If $Y = \beta_1 + u$, obtain the OLS estimator of $\beta_1$ using (3).
15.7 Unbiasedness of the OLS regression coefficients

Substituting for \( y \) from (2) into (3), we have:

\[
\hat{\beta} = [X'X]^{-1}X'(X\beta + u)
\]
\[
= [X'X]^{-1}X'\beta + [X'X]^{-1}X'u
\]
\[
= \beta + [X'X]^{-1}X'u.
\]

Hence each element of \( \hat{\beta} \) is equal to the corresponding value of \( \beta \) plus a linear combination of the values of the disturbance term in the sample. Next:

\[
E(\hat{\beta} | X) = \beta + E([X'X]^{-1}X'u | X).
\]

To proceed further, we need to be specific about the data generation process (DGP) for \( X \) and the assumptions concerning \( u \) and \( X \). In Model A, we have no DGP for \( X \): the data are simply taken as given. When we describe the properties of the regression estimators, we are either talking about the potential properties, before the sample has been drawn, or about the distributions that we would expect in repeated samples using those given data on \( X \). If we make the assumption \( E(u | X) = 0 \), then:

\[
E(\hat{\beta} | X) = \beta + E([X'X]^{-1}X'u | X) = \beta
\]

and so \( \hat{\beta} \) is an unbiased estimator of \( \beta \). It should be stressed that unbiasedness in Model A, along with all other properties of the regression estimators, are conditional on the actual given data for \( X \).

In Model B, we allow \( X \) to be drawn from a fixed joint distribution of the explanatory variables. The appropriate assumption for the disturbance term is that it is distributed independently of \( X \) and hence its conditional distribution is no different from its absolute distribution: \( E(u | X) = E(u) \) for all \( X \). We also assume \( E(u) = 0 \). The independence of the distributions of \( X \) and \( u \) allows us to write:

\[
E(\hat{\beta} | X) = \beta + E([X'X]^{-1}X'u | X)
\]
\[
= \beta + E([X'X]^{-1}X')E(u)
\]
\[
= \beta.
\]

15.8 The variance-covariance matrix of the OLS regression coefficients

We define the variance-covariance matrix of the disturbance term to be the matrix whose element in row \( i \) and column \( j \) is the population covariance of \( u_i \) and \( u_j \). By assumption A.4, the covariance of \( u_i \) and \( u_j \) is constant and equal to \( \sigma^2_u \) if \( j = i \) and by assumption A.5 it is equal to zero if \( j \neq i \). Thus the variance-covariance matrix is:
15. Regression analysis with linear algebra primer

\[
\begin{bmatrix}
\sigma_u^2 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & \sigma_u^2 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \sigma_u^2 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \sigma_u^2 & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & \sigma_u^2 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \sigma_u^2 \\
\end{bmatrix}
\]

that is, a matrix whose diagonal elements are all equal to \( \sigma_u^2 \) and whose off-diagonal elements are all zero. It may more conveniently be written \( \mathbf{I}_n \sigma_u^2 \) where \( \mathbf{I}_n \) is the identity matrix of order \( n \).

Similarly, we define the variance-covariance matrix of the regression coefficients to be the matrix whose element in row \( i \) and column \( j \) is the population covariance of \( \hat{\beta}_i \) and \( \hat{\beta}_j \):

\[
\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = E\left( (\hat{\beta}_i - E(\hat{\beta}_i))(\hat{\beta}_j - E(\hat{\beta}_j)) \right) = E\left( (\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j) \right).
\]

The diagonal elements are of course the variances of the individual regression coefficients. We denote this matrix \( \text{var}(\hat{\beta}) \). If we are using the framework of Model A, everything will be conditional on the actual given data for \( \mathbf{X} \), so we should refer to \( \text{var}(\hat{\beta} \mid \mathbf{X}) \) rather than \( \text{var}(\hat{\beta}) \). Then:

\[
\text{var}(\hat{\beta} \mid \mathbf{X}) = E((\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))' \mid \mathbf{X})
\]

\[
= E((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \mid \mathbf{X})
\]

\[
= E\left( ([\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{u})([\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{u})' \mid \mathbf{X} \right)
\]

\[
= E\left( [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1} \mid \mathbf{X} \right)
\]

\[
= [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}' E(\mathbf{u}\mathbf{u}' \mid \mathbf{X})[\mathbf{X}'\mathbf{X}]^{-1}
\]

\[
= [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}' \mathbf{I}_n \sigma_u^2 \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}
\]

\[
= [\mathbf{X}'\mathbf{X}]^{-1} \sigma_u^2.
\]

If we are using Model B, we can obtain the unconditional variance of \( \mathbf{b} \) using the standard decomposition of a variance in a joint distribution:

\[
\text{var}(\hat{\beta}) = E\left[ \text{var}(\hat{\beta} \mid \mathbf{X}) \right] + \text{var}\left[ E(\hat{\beta} \mid \mathbf{X}) \right].
\]

Now \( E(\hat{\beta} \mid \mathbf{X}) = \beta \) for all \( \mathbf{X} \), so \( \text{var}[E(\hat{\beta} \mid \mathbf{X})] = \text{var}(\beta) = 0 \) since \( \beta \) is a constant vector, so:

\[
\text{var}(\hat{\beta}) = E\left( [\mathbf{X}'\mathbf{X}]^{-1} \sigma_u^2 \right) = \sigma_u^2 E\left( [\mathbf{X}'\mathbf{X}]^{-1} \right)
\]

the expectation being taken over the distribution of \( \mathbf{X} \).

To estimate \( \text{var}(\hat{\beta}) \), we need to estimate \( \sigma_u^2 \). An unbiased estimator is provided by \( \hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k) \). For a proof, see Appendix B.
15.9 The Gauss–Markov theorem

We will demonstrate that the OLS estimators are the minimum variance unbiased estimators that are linear in \( y \). For simplicity, we will do this within the framework of Model A, with the analysis conditional on the given data for \( X \). The analysis generalises straightforwardly to Model B, where the explanatory variables are stochastic but drawn from fixed distributions.

Consider the general estimator in this class:

\[
\hat{\beta}^* = Ay
\]

where \( A \) is a \( k \) by \( n \) matrix. Let:

\[
C = A - [X'X]^{-1}X'.
\]

Then:

\[
\hat{\beta}^* = ([X'X]^{-1}X' + C) y
\]

\[
= ([X'X]^{-1}X' + C) (X\beta + u)
\]

\[
= \beta + CX\beta + [X'X]^{-1}X' u + Cu.
\]

Unbiasedness requires:

\[
CX = 0_k
\]

where \( 0 \) is a \( k \) by \( k \) matrix consisting entirely of zeros. Then, with \( E(\hat{\beta}^*) = \beta \), the variance-covariance matrix of \( \hat{\beta}^* \) is given by:

\[
E \left[ (\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)' \right] = E \left[ ([X'X]^{-1}X' + C) uu' ([X'X]^{-1}X' + C)' \right]
\]

\[
= ([X'X]^{-1}X' + C) I_n \sigma_u^2 ([X'X]^{-1}X' + C)'
\]

\[
= ([X'X]^{-1}X' + C) ([X'X]^{-1}X' + C)' \sigma_u^2
\]

\[
= ([X'X]^{-1} + CC')' \sigma_u^2.
\]

Now diagonal element \( i \) of \( CC' \) is the inner product of row \( i \) of \( C \) and column \( i \) of \( C' \). These are the same, so it is given by:

\[
\sum_{s=1}^{k} c_{ik}^2
\]

which is positive unless \( c_{is} = 0 \) for all \( s \). Hence minimising the variances of the estimators of all of the elements of \( \beta \) requires \( C = 0 \). This implies that OLS provides the minimum variance unbiased estimator.

15.10 Consistency of the OLS regression coefficients

Since:

\[
\hat{\beta} = \beta + [X'X]^{-1}X' u
\]
15. Regression analysis with linear algebra primer

the probability limit of $\hat{\beta}$ is given by:

$$\text{plim } \hat{\beta} = \beta + \text{plim } [X'X]^{-1}X'u$$

$$= \beta + \text{plim } \left( \frac{1}{n} X'X \right)^{-1} \frac{1}{n} X'u\right).$$

Now, if we are working with cross-sectional data with the explanatory variables drawn from fixed (joint) distributions, it can be shown that:

$$\text{plim } \left[ \frac{1}{n} X'X \right]^{-1}$$

has a limiting matrix and that:

$$\text{plim } \frac{1}{n} X'u = 0.$$

Hence we can decompose:

$$\text{plim } \left( \frac{1}{n} X'X \right)^{-1} \frac{1}{n} X'u\right) = \text{plim } \left[ \frac{1}{n} X'X \right]^{-1} \text{plim } \frac{1}{n} X'u = 0$$

and so $\text{plim } \hat{\beta} = \beta$. Note that this is only an outline of the proof. For a proper proof and a generalisation to less restrictive assumptions, see Greene pp.64–65.

15.11 Frisch–Waugh–Lovell theorem

We will precede the discussion of the Frisch–Waugh–Lovell (FWL) theorem by introducing the residual-maker matrix. We have seen that, when we fit:

$$y = X\beta + u$$

using OLS, the residuals are given by:

$$\hat{u} = y - \hat{y} = y - X\hat{\beta}.$$ 

Substituting for $\hat{\beta}$, we have:

$$\hat{u} = y - X[X'X]^{-1}X'y$$

$$= \left[ I - X[X'X]^{-1}X' \right] y$$

$$= My$$

where:

$$M = I - X[X'X]^{-1}X'.$$

$M$ is known as the ‘residual-maker’ matrix because it converts the values of $y$ into the residuals of $y$ when regressed on $X$. Note that $M$ is symmetric, because $M' = M$, and idempotent, meaning that $MM = M$.

Now suppose that we divide the $k$ variables comprising $X$ into two subsets, the first $s$ and the last $k - s$. (For the present purposes, it makes no difference whether there is or
15.11. Frisch–Waugh–Lovell theorem

is not an intercept in the model, and if there is one, whether the vector of ones responsible for it is in the first or second subset.) We will partition \( X \) as:

\[
X = [X_1 \ X_2]
\]

where \( X + 1 \) comprises the first \( s \) columns and \( X_2 \) comprises the last \( k - s \), and we will partition \( \beta \) similarly, so that the theoretical model may be written:

\[
y = [X_1 \ X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u.
\]

The FWL theorem states that the OLS estimates of the coefficients in \( \beta_1 \) are the same as those that would be obtained by the following procedure: regress \( y \) on the variables in \( X_2 \) and save the residuals as \( \hat{u}_y \). Regress each of the variables in \( X_1 \) on \( X_2 \) and save the matrix of residuals as \( \hat{u}_{X_1} \). If we regress \( \hat{u}_y \) on \( \hat{u}_{X_1} \), we will obtain the same estimates of the coefficients of \( \beta_1 \) as we did in the straightforward multiple regression. (Why we might want to do this is another matter. We will come to this later.) Applying the preceding discussion relating to the residual-maker, we have:

\[
\hat{u}_y = M_2 y
\]

where:

\[
M_2 = I - X_2[X_2'X_2]^{-1}X_2'
\]

and:

\[
\hat{u}_{X_1} = M_2 X_1.
\]

Let the vector of coefficients obtained when we regress \( \hat{u}_y \) on \( \hat{u}_{X_1} \) be denoted \( \hat{\beta}_1^* \). Then:

\[
\hat{\beta}_1^* = [\hat{u}_{X_1} \hat{u}_{X_1}]^{-1}\hat{u}_{X_1} \hat{u}_y = [X_1'M_2'M_2X_1]^{-1}X_1'M_2'M_2y = [X_1'M_2X_1]^{-1}X_1M_2y.
\]

(Remember that \( M_2 \) is symmetric and idempotent.) Now we will derive an expression for \( \hat{\beta}_1 \) from the orthodox multiple regression of \( y \) on \( X \). For this purpose, it is easiest to start with the normal equations:

\[
X'X\hat{\beta} - X'y = 0.
\]

We partition \( \hat{\beta} \) as \( \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \), \( X' \) is \( \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \), and we have the following:

\[
X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix},
\]

\[
X'X\hat{\beta} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 \\ X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 \end{bmatrix}.
\]

\[
X'y = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}.
\]
15. Regression analysis with linear algebra primer

Hence, splitting the normal equations into their upper and lower components, we have:

\[ X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 - X_1'y = 0 \]

and:

\[ X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 - X_2'y = 0. \]

From the second we obtain:

\[ X_2'X_2\hat{\beta}_2 = X_2'y - X_2'X_1\hat{\beta}_1 \]

and so:

\[ \hat{\beta}_2 = [X_2'X_2]^{-1}[X_2'y - X_2'X_1\hat{\beta}_1]. \]

Substituting for \( \hat{\beta}_2 \) in the first normal equation:

\[ X_1'X_1\hat{\beta}_1 + X_1'X_2[X_2'X_2]^{-1}[X_2'y - X_2'X_1\hat{\beta}_1] - X_1'y = 0. \]

Hence:

\[ X_1'X_1\hat{\beta}_1 - X_1'X_2[X_2'X_2]^{-1}X_2'X_1\hat{\beta}_1 = X_1'y - X_1'X_2[X_2'X_2]^{-1}X_2'y \]

and so:

\[ X_1'[I - X_2[X_2'X_2]^{-1}X_2']X_1\hat{\beta}_1 = X_1'[I - X_2[X_2'X_2]^{-1}X_2']y. \]

Hence:

\[ X_1'M_2X_1\hat{\beta}_1 = X_1'M_2y \]

and:

\[ \hat{\beta}_1 = [X_1'M_2X_1]^{-1}X_1'M_2y = \hat{\beta}_1^*. \]

Why should we be interested in this result? The original purpose remains instructive. In early days, econometricians working with time series data, especially macroeconomic data, were concerned to avoid the problem of spurious regressions. If two variables both possessed a time trend, it was very likely that ‘significant’ results would be obtained when one was regressed on the other, even if there were no genuine relationship between them. To avoid this, it became the custom to detrend the variables before using them by regressing each on a time trend and then working with the residuals from these regressions. Frisch and Waugh (1933) pointed out that this was an unnecessarily laborious procedure. The same results would be obtained using the original data, if a time trend was added as an explanatory variable.

Generalising, and this was the contribution of Lovell, we can infer that, in a multiple regression model, the estimator of the coefficient of any one variable is not influenced by any of the other variables, irrespective of whether they are or are not correlated with the variable in question. The result is so general and basic that it should be understood by all students of econometrics. Of course, it fits neatly with the fact that the multiple regression coefficients are unbiased, irrespective of any correlations among the variables.

A second reason for being interested in the result is that it allows one to depict graphically the relationship between the observations on the dependent variable and those on any single explanatory variable, controlling for the influence of all the other explanatory variables. This is described in the textbook in Section 3.2.
Exercise

9. Using the FKL theorem, demonstrate that, if a multiple regression model contains an intercept, the same slope coefficients could be obtained by subtracting the means of all of the variables from the data for them and then regressing the model omitting an intercept.

15.12 Exact multicollinearity

We will assume, as is to be expected, that \(k\), the number of explanatory variables (including the unit vector, if there is one), is less than \(n\), the number of observations. If the explanatory variables are independent, the \(X\) matrix will have rank \(k\) and likewise \(X'X\) will have rank \(k\) and will possess an inverse. However, if one or more linear relationships exist among the explanatory variables, the model will be subject to exact multicollinearity. The rank of \(X\), and hence of \(X'X\), will then be less than \(k\) and \(X'X\) will not possess an inverse.

Suppose we write \(X\) as a set of column vectors \(x_j\), each corresponding to the observations on one of the explanatory variables:

\[
X = [x_1 \ldots x_j \ldots x_k]
\]

where:

\[
x_j = \begin{bmatrix}
    x_{1j} \\
    \vdots \\
    x_{ij} \\
    \vdots \\
    x_{nj}
\end{bmatrix}.
\]

Then:

\[
X' = \begin{bmatrix}
    x_1' \\
    \vdots \\
    x_j' \\
    \vdots \\
    x_k'
\end{bmatrix}
\]

and the normal equations:

\[
X'X\hat{\beta} - X'y = 0
\]

may be written:

\[
\begin{bmatrix}
    x_1'X\hat{\beta} \\
    \vdots \\
    x_j'X\hat{\beta} \\
    \vdots \\
    x_k'X\hat{\beta}
\end{bmatrix} - \begin{bmatrix}
    x_1'y \\
    \vdots \\
    x_j'y \\
    \vdots \\
    x_k'y
\end{bmatrix} = 0.
\]
15. Regression analysis with linear algebra primer

Now suppose that one of the explanatory variables, say the last, can be written as a linear combination of the others:

\[ x_k = \sum_{i=1}^{k-1} \lambda_i x_i. \]

Then the last of the normal equations is that linear combination of the other \( k - 1 \). Hence it is redundant, and we are left with a set of \( k - 1 \) equations for determining the \( k \) unknown regression coefficients. The problem is not that there is no solution. It is the opposite: there are too many possible solutions, in fact an infinite number. One coefficient could be chosen arbitrarily, and then the normal equations would provide a solution for the other \( k - 1 \). Some regression applications deal with this situation by dropping one of the variables from the regression specification, effectively assigning a value of zero to its coefficient.

Exact multicollinearity is unusual because it mostly occurs as a consequence of a logical error in the specification of the regression model. The classic example is the dummy variable trap. This occurs when a set of dummy variables \( D_j, j = 1, \ldots, s \) are defined for a qualitative characteristic that has \( s \) categories. If all \( s \) dummy variables are included in the specification, in observation \( i \) we will have:

\[ \sum_{j=1}^{s} D_{ij} = 1 \]

since one of the dummy variables must be equal to 1 and the rest are all zero. But this is the (unchanging) value of the unit vector. Hence the sum of the dummy variables is equal to the unit vector. As a consequence, if the unit vector and all of the dummy variables are simultaneously included in the specification, there will be exact multicollinearity. The solution is to drop one of the dummy variables, making it the reference category, or, alternatively, to drop the intercept (and hence unit vector), effectively making the dummy variable coefficient for each category the intercept for that category. As explained in the textbook, it is illogical to wish to include a complete set of dummy variables as well as the intercept, for then no interpretation can be given to the dummy variable coefficients.

15.13 Estimation of a linear combination of regression coefficients

Suppose that one wishes to estimate a linear combination of the regression parameters:

\[ \sum_{j=1}^{k} \lambda_j \beta_j. \]

In matrix notation, we may write this as \( \lambda' \beta \) where:

\[ \lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_j \\ \vdots \\ \lambda_k \end{bmatrix}. \]
The corresponding linear combination of the regression coefficients, $\lambda' \hat{\beta}$, provides an unbiased estimator of $\lambda' \beta$. However, we will often be interested also in its standard error, and this is not quite so straightforward. We obtain it via the variance:

$$\text{var}(\lambda' \hat{\beta}) = E[(\lambda' \hat{\beta} - E(\lambda' \hat{\beta}))^2] = E[(\lambda' \hat{\beta} - \lambda' \beta)^2].$$

Since $(\lambda' \hat{\beta} - \lambda' \beta)$ is a scalar, it is equal to its own transpose, and so $(\lambda' \hat{\beta} - \lambda' \beta)^2$ may be written:

$$\text{var}(\lambda' \hat{\beta}) = \lambda'E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \lambda = \lambda'[X'X]^{-1}\lambda \sigma_u^2.$$

The square root of this expression provides the standard error of $\lambda' \hat{\beta}$ after we have replaced $\sigma_u^2$ by its estimator $\hat{u}' \hat{u}/(n - k)$ in the usual way.

### 15.14 Testing linear restrictions

An obvious application of the foregoing is its use in testing a linear restriction. Suppose that one has a hypothetical restriction:

$$\sum_{j=1}^{k} \lambda_j \beta_j = \lambda_0.$$

We can perform a $t$ test of the restriction using the $t$ statistic:

$$t = \frac{\lambda' \hat{\beta} - \lambda_0}{\text{s.e.}(\lambda' \hat{\beta})}$$

where the standard error is obtained via the variance-covariance matrix as just described. Alternatively, we could reparameterise the regression specification so that one of the coefficients is $\lambda' \beta$. In practice, this is often more convenient since it avoids having to work with the variance-covariance matrix. If there are multiple restrictions that should be tested simultaneously, the appropriate procedure is an $F$ test comparing $RSS$ for the unrestricted and fully restricted models.

### 15.15 Weighted least squares and heteroskedasticity

Suppose that the regression model:

$$y = X\beta + u$$

satisfies the usual regression model assumptions and suppose that we premultiply the elements of the model by the $n$ by $n$ matrix $A$ whose diagonal elements are $A_{ii}$,
15. Regression analysis with linear algebra primer

\( i = 1, \ldots, n, \) and whose off-diagonal elements are all zero:

\[
A = \begin{bmatrix}
A_{11} & \cdots & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & A_{ii} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \cdots & A_{nn}
\end{bmatrix}.
\]

The model becomes:

\[ Ay = AX\beta + Au. \]

If we fit it using least squares, the point estimates of the coefficients are given by:

\[
\hat{\beta}_{WLS} = \left[ X'AX \right]^{-1}X'Ay
\]

(WLS standing for weighted least squares). This is unbiased but heteroskedastic
because the disturbance term in observation \( i \) is \( A_{ii}u_i \) and has variance \( A_{ii}^2\sigma_u^2 \).

Now suppose that the disturbance term in the original model was heteroskedastic, with
variance \( \sigma_u^2 \) in observation \( i \). If we define the matrix \( A \) so that the diagonal elements are determined by:

\[
A_{ii} = \frac{1}{\sqrt{\sigma_u^2}}
\]

the corresponding variance in the weighted regression will be 1 for all observations and
the WLS model will be homoskedastic. The WLS estimator is then:

\[
\hat{\beta}_{WLS} = [X'C]^{-1}X'Cy
\]

where:

\[
C = A'A = \begin{bmatrix}
\frac{1}{\sigma_u^2} & \cdots & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \frac{1}{\sigma_u^2} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \cdots & \frac{1}{\sigma_u^2}
\end{bmatrix}.
\]

The variance-covariance matrix of the WLS coefficients, conditional on the data for \( X \), is:

\[
\text{var}(\hat{\beta}_{WLS}) = E \left[ (\hat{\beta}_{WLS} - E(\hat{\beta}_{WLS}))(\hat{\beta}_{WLS} - E(\hat{\beta}_{WLS}))' \right]
\]

\[
= E \left[ (\hat{\beta}_{WLS} - \beta)(\hat{\beta}_{WLS} - \beta)' \right]
\]

\[
= E \left[ \left([X'A'AX]^{-1}X'A'Au\right)\left([X'A'AX]^{-1}X'A'Au\right)' \right]
\]

\[
= \left([X'A'AX]^{-1}X'A'Au'u'A'AX[X'A'AX]^{-1} \right]
\]

\[
= \left([X'A'AX]^{-1}X'A'AE(u'u')A'AX[X'A'AX]^{-1} \right]
\]

\[
= \left([X'A'AX]^{-1}X'A'AX[X'A'AX]^{-1} \right]
\]

\[
= \left([X'C]^{-1}\sigma_u^2 \right]
\]

\[
= \left([X'C]^{-1}\sigma_u^2 \right]
\]

326
since $A$ has been defined so that:

$$AE(uu')A' = I.$$ 

Of course, in practice we seldom know $\sigma^2_u$, but if it is appropriate to hypothesise that the standard deviation is proportional to some measurable variable $Z_i$, then the WLS regression will be homoskedastic if we define $A$ to have diagonal element $i$ equal to the reciprocal of $Z_i$.

### 15.16 IV estimators and TSLS

Suppose that we wish to fit the model:

$$y = X\beta + u$$

where one or more of the explanatory variables is not distributed independently of the disturbance term. For convenience, we will describe such variables as ‘endogenous’, irrespective of the reason for the violation of the independence requirement. Given a sufficient number of suitable instruments, we may consider using the IV estimator:

$$\hat{\beta}^{IV} = [WX]^{-1}W'y (4)$$

where $W$ is the matrix of instruments. In general $W$ will be a mixture of (1) those original explanatory variables that are distributed independently of the disturbance term (these are then described as acting as instruments for themselves), and (2) new variables that are correlated with the endogenous variables but distributed independently of the disturbance term. If we substitute for $y$:

$$\hat{\beta}^{IV} = [WX]^{-1}W'(X\beta + u) = \beta + [WX]^{-1}W'u.$$ 

We cannot obtain a closed-form expression for the expectation of the error term, so instead we take plims:

$$\text{plim} \, \hat{\beta}^{IV} = \beta + \text{plim} \left( \frac{1}{n}WX\right)^{-1}\frac{1}{n}W'u.$$ 

Now if we are using cross-sectional data, it is usually reasonable to suppose that:

$$\text{plim} \left( \frac{1}{n}WX\right)^{-1} \quad \text{and} \quad \text{plim} \left( \frac{1}{n}W'u\right)$$

both exist, in which case we can decompose the plim of the error term:

$$\text{plim} \, \hat{\beta}^{IV} = \beta + \text{plim} \left( \frac{1}{n}WX\right)^{-1} \text{plim} \left( \frac{1}{n}W'u\right).$$ 

Further, if the matrix of instruments has been correctly chosen, it can be shown that:

$$\text{plim} \left( \frac{1}{n}W'u\right) = 0$$
15. Regression analysis with linear algebra primer

and hence the IV estimator is consistent.

It is not possible to derive a closed-form expression for the variance of the IV estimator in finite samples. The best we can do is to invoke a central limit theorem that gives the limiting distribution asymptotically and work backwards from that, as an approximation, for finite samples. A central limit theorem can be used to establish that:

\[ \sqrt{n} (\hat{\beta}^{IV} - \beta) \xrightarrow{d} N \left( 0, \sigma^2_u \text{plim} \left[ \frac{1}{n} W'X \right]^{-1} \text{plim} \left[ \frac{1}{n} W'W \right] \text{plim} \left[ \frac{1}{n} X'W \right]^{-1} \right) . \]

From this, we may infer, that as an approximation, for sufficiently large samples:

\[ \hat{\beta}^{IV} \sim N \left( \beta, \frac{\sigma^2_u}{n} \text{plim} \left[ \frac{1}{n} W'X \right]^{-1} \text{plim} \left[ \frac{1}{n} W'W \right] \text{plim} \left[ \frac{1}{n} X'W \right]^{-1} \right) . \] (5)

We have implicitly assumed so far that \( W \) has the same dimensions as \( X \) and hence that \( W'X \) is a square \( k \) by \( k \) matrix. However, the model may be overidentified, with the number of columns of \( W \) exceeding \( k \). In that case, the appropriate procedure is two-stage least squares. One regresses each of the variables in \( X \) on \( W \) and saves the fitted values. The matrix of fitted values is then used as the instrument matrix in place of \( W \).

Exercises

10. Using (4) and (5), demonstrate that, for the simple regression model:

\[ Y_i = \beta_1 = \beta_2 X_i + u_i \]

with \( Z \) acting as an instrument for \( X \) (and the unit vector acting as an instrument for itself):

\[ \hat{\beta}_1^{IV} = \bar{Y} - \hat{\beta}_2^{IV} \bar{X} \]

\[ \hat{\beta}_2^{IV} = \frac{\sum (Z_i - \bar{Z}) (Y_i - \bar{Y})}{\sum (Z_i - \bar{Z}) (X_i - \bar{X})} \]

and, as an approximation:

\[ \text{var} (\hat{\beta}_2^{IV}) = \frac{\sigma^2_u}{\sum (X_i - \bar{X})^2} \times \frac{1}{r_{XZ}^2} \]

where \( Z \) is the instrument for \( X \) and \( r_{XZ} \) is the correlation between \( X \) and \( Z \).

11. Demonstrate that any variable acting as an instrument for itself is unaffected by the first stage of two-stage least squares.

12. Demonstrate that TSLS is equivalent to IV if the equation is exactly identified.
15.17 Generalised least squares

The final topic in this introductory primer is generalised least squares and its application to autocorrelation (autocorrelated disturbance terms). One of the basic regression model assumptions is that the disturbance terms in the observations in a sample are distributed identically and independently of each other. If this is the case, the variance-covariance matrix of the disturbance terms is the identity matrix of order \( n \), multiplied by \( \sigma^2_u \). We have encountered one type of violation, heteroskedasticity, where the values of the disturbance term are independent but not identical. The consequence was that the off-diagonal elements of the variance-covariance matrix remained zero, but the diagonal elements differed. Mathematically, autocorrelation is complementary. It occurs when the values of the disturbance term are not independent and as a consequence some, or all, of the off-diagonal elements are non-zero. It is usual in initial treatments to retain the assumption of identical distributions, so that the diagonal elements of the variance-covariance matrix are the same. Of course, in principle one could have both types of violation at the same time.

In abstract, it is conventional to denote the variance-covariance matrix of the disturbance term \( \Omega \sigma^2_u \), where \( \Omega \) is the Greek upper case omega, writing the model:

\[
y = X\beta \quad \text{with} \quad E(uu') = \Omega \sigma^2_u.
\] (6)

If the values of the disturbance term are iid, \( \Omega = I \). If they are not iid, OLS is in general inefficient and the standard errors are estimated incorrectly. Then, it is desirable to transform the model so that the transformed disturbance terms are iid. One possible way of doing this is to multiply through by some suitably chosen matrix \( P \), fitting:

\[
Py = PX\beta + Pu
\]

Choosing \( P \) so that \( E(Puu'P') = I\alpha \) where \( \alpha \) is some scalar. The solution for heteroskedasticity was a simple example of this type. We had:

\[
\Omega = \begin{bmatrix}
\sigma^2_u & \cdots & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \sigma^2_u & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \cdots & \sigma^2_u
\end{bmatrix}
\]

and the appropriate choice of \( P \) was:

\[
P = \begin{bmatrix}
\sqrt{\frac{1}{\sigma^2_u_1}} & \cdots & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \sqrt{\frac{1}{\sigma^2_u_i}} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \cdots & \sqrt{\frac{1}{\sigma^2_u_n}}
\end{bmatrix}
\]

In the case of heteroskedasticity, the choice of \( P \) is obvious, provided, of course, that one knows the values of the diagonal elements of \( \Omega \). The more general theory requires an understanding of eigenvalues and eigenvectors that will be assumed. \( \Omega \) is a
15. Regression analysis with linear algebra primer

A symmetric matrix since \( \text{cov}(u_i, u_j) \) is the same as \( \text{cov}(u_j, u_i) \). Hence all its eigenvalues are real. Let \( \Lambda \) be the diagonal matrix with the eigenvalues as the diagonal elements. Then there exists a matrix of eigenvectors, \( C \), such that:

\[
C' \Omega C = \Lambda. \tag{7}
\]

\( C \) has the properties that \( CC' = I \) and \( C' = C^{-1} \). Since \( \Lambda \) is a diagonal matrix, if its eigenvalues are all positive (which means that it is what is known as a ‘positive definite’ matrix), it can be factored as \( \Lambda = \Lambda^{1/2} \Lambda^{1/2} \) where \( \Lambda^{1/2} \) is a diagonal matrix whose diagonal elements are the square roots of the eigenvalues. It follows that the inverse of \( \Lambda \) can be factored as \( \Lambda^{-1} = \Lambda^{-1/2} \Lambda^{-1/2} \). Then, in view of (7):

\[
\Lambda^{-1/2} [C' \Omega C] \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda^{1/2} \Lambda^{1/2} \Lambda^{-1/2} = I. \tag{8}
\]

This, if we define \( P = \Lambda^{-1/2} C' \), (8) becomes:

\[
P \Omega P' = I.
\]

As a consequence, if we premultiply (6) through by \( P \), we have:

\[
Py = PX\beta + Pu
\]

or:

\[
y^* = X'*\beta + u^*
\]

where \( y^* = Py \), \( X^* = PX \), and \( u^* = Pu \), and \( E(u'u'^t) = \sigma_u^2 I \). An OLS regression of \( y^* \) on \( X^* \) will therefore satisfy the usual regression model assumptions and the estimator of \( \beta \) will have the usual properties. Of course, the approach usually requires the estimation of \( \Omega \), \( \Omega \) being positive definite, and there being no problems in extracting the eigenvalues and determining the eigenvectors.

**Exercise**

13. Suppose that the disturbance term in a simple regression model (with an intercept) is subject to AR(1) autocorrelation with \( |\rho| < 1 \), and suppose that the sample consists of just two observations. Determine the variance-covariance matrix of the disturbance term, find its eigenvalues, and determine its eigenvectors. Hence determine \( P \) and state the transformed model. Verify that the disturbance term in the transformed model is iid.

### 15.18 Appendix A: Derivation of the normal equations

We have seen that \( RSS \) is given by:

\[
RSS = y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}. \tag{A.1}
\]

The normal equations are:

\[
\frac{\partial RSS}{\partial \beta_j} = 0 \tag{A.2}
\]

330
for $j = 1, \ldots, k$. We will show that they can be written:

$$X'X\hat{\beta} - X'y = 0.$$ 

The proof is mathematically unchallenging but tedious because one has to keep careful track of the dimensions of all of the elements in the equations. As far as I know, it is of no intrinsic interest and once one has seen it there should never be any reason to look at it again.

First note that the term $y'y$ in (A.1) is not a function of any of the $b_j$ and disappears in (A.2). Accordingly we will restrict our attention to the other two terms on the right side of (A.1). Suppose that we write the $X$ matrix as a set of column vectors:

$$X = [x_1 \cdots x_j \cdots x_k] \quad \text{(A.3)}$$

where:

$$x_j = \begin{bmatrix} X_{1j} \\ \vdots \\ X_{nj} \end{bmatrix}.$$ 

Then:

$$y'X\hat{\beta} = [y'x_1 \cdots y'x_j \cdots y'x_k] \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_j \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = [y'x_1\hat{\beta}_1 + \cdots + y'x_j\hat{\beta}_j + \cdots + y'x_k\hat{\beta}_k].$$

Hence:

$$\frac{\partial y'X\hat{\beta}}{\partial \hat{\beta}_j} = y'x_j.$$ 

We now consider the $\hat{\beta}'X'X\hat{\beta}$ term. Using (A.3):

$$\hat{\beta}'X'X\hat{\beta} = [x_1\hat{\beta}_1 + \cdots + x_j\hat{\beta}_j + \cdots + x_k\hat{\beta}_k]'[x_1\hat{\beta}_1 + \cdots + x_j\hat{\beta}_j + \cdots + x_k\hat{\beta}_k]$$

$$= \sum_{p=1}^{k} \sum_{q=1}^{k} \hat{\beta}_p \hat{\beta}_q x'_p x_q.$$ 

The subset of terms including $\hat{\beta}_j$ is:

$$\sum_{q=1}^{k} \hat{\beta}_j \hat{\beta}_q x'_j x_q + \sum_{p=1}^{k} \hat{\beta}_p \hat{\beta}_j x'_p x_j.$$ 

Hence:

$$\frac{\partial \hat{\beta}'X'X\hat{\beta}}{\partial \hat{\beta}_j} = \sum_{q=1}^{k} \hat{\beta}_q x'_j x_q + \sum_{p=1}^{k} \hat{\beta}_p x'_p x_j = 2 \sum_{p=1}^{k} \hat{\beta}_p x'_p x_j.$$
15. Regression analysis with linear algebra primer

Putting these results together:

\[
\frac{\partial \text{RSS}}{\partial \hat{\beta}_j} = \frac{\partial [y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}]}{\partial \hat{\beta}_j} = -2y'x_j + 2 \sum_{p=1}^{k} \hat{\beta}_p x'_p x_j.
\]

Hence the normal equation \( \frac{\partial \text{RSS}}{\partial \hat{\beta}_j} = 0 \) is:

\[
\sum_{p=1}^{k} \hat{\beta}_p x'_p x_j = x'_j y.
\]

(Note that \( x'_j x_j = x'_j x_j \) and \( y'x_j = x'_j y \) since they are scalars.) Hence:

\[
x'_j \left[ \sum_{p=1}^{k} \hat{\beta}_p x_p \right] = x'_j y.
\]

Hence:

\[
x'_j X\hat{\beta} = x'_j y
\]

since:

\[
X\hat{\beta} = [x_1 \cdots x_p \cdots x_k] \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \sum_{p=1}^{k} x_p \hat{\beta}_p.
\]

Hence, stacking the \( k \) normal equations:

\[
\begin{bmatrix} x_1'X\hat{\beta} \\ \vdots \\ x_j'X\hat{\beta} \\ \vdots \\ x_k'X\hat{\beta} \end{bmatrix} = \begin{bmatrix} x_1'y \\ \vdots \\ x_j'y \\ \vdots \\ x_k'y \end{bmatrix}.
\]

Hence:

\[
\begin{bmatrix} x'_1 \\ \vdots \\ x'_j \\ \vdots \\ x'_k \end{bmatrix} X\hat{\beta} = \begin{bmatrix} x'_1 \\ \vdots \\ x'_j \\ \vdots \\ x'_k \end{bmatrix} y.
\]

Hence:

\[
X'X\hat{\beta} = X'y.
\]

15.19 Appendix B: Demonstration that \( \hat{u}'\hat{u}/(n - k) \) is an unbiased estimator of \( \sigma^2_u \)

This classic proof is both elegant, in that it is much shorter than any proof not using matrix algebra, and curious, in that it uses the trace of a matrix, a feature that I have
never seen used for any other purpose. The trace of a matrix, defined for square matrices only, is the sum of its diagonal elements. We will first need to demonstrate that, for any two conformable matrices whose product is square:

\[ tr(AB) = tr(BA). \]

Let \( A \) have \( n \) rows and \( m \) columns, and let \( B \) have \( m \) rows and \( n \) columns. Diagonal element \( i \) of \( AB \) is:

\[ \sum_{p=1}^{m} a_{ip} b_{pi}. \]

Hence:

\[ tr(AB) = \sum_{i=1}^{n} \left( \sum_{p=1}^{m} a_{ip} b_{pi} \right). \]

Similarly, diagonal element \( i \) of \( BA \) is:

\[ \sum_{p=1}^{n} b_{ip} a_{pi}. \]

Hence:

\[ tr(BA) = \sum_{i=1}^{m} \left( \sum_{p=1}^{n} b_{ip} a_{pi} \right). \]

What we call the symbols used to index the summations makes no difference. Re-writing \( p \) as \( i \) and \( i \) as \( p \), and noting that the order of the summation makes no difference, we have \( tr(BA) = tr(AB) \).

We also need to note that:

\[ tr(A + B) = tr(A) + tr(B) \]

where \( A \) and \( B \) are square matrices of the same dimension. This follows immediately from the way that we sum conformable matrices.

By definition:

\[ \hat{u} = y - \hat{y} = y - X\hat{\beta}. \]

Using:

\[ \hat{\beta} = [X'X]^{-1}X'y \]

we have:

\[ \hat{u} = y - X[X'X]^{-1}X'y \]

\[ = X\beta + u - X[X'X]^{-1}X'(X\beta + u) \]

\[ = I_n u - X[X'X]^{-1}X'u \]

\[ = Mu \]

where \( I_n \) is an identity matrix of dimension \( n \) and:

\[ M = I_n - X[X'X]^{-1}X'. \]
15. Regression analysis with linear algebra primer

Hence:

\[ \hat{u} \hat{u} = u'M'Mu. \]

Now \( M \) is symmetric and idempotent: \( M' = M \) and \( MM = M \). Hence:

\[ \hat{u} \hat{u} = u'Mu \]

\( \hat{u} \hat{u} \) is a scalar, and so the expectation of \( \hat{u} \hat{u} \) and the expectation of the trace of \( \hat{u} \hat{u} \) are the same. So:

\[ E(\hat{u} \hat{u}) = E(tr(\hat{u} \hat{u})) = E(tr(u'Mu)) = E(tr(Muu')) = tr(E(Muu')). \]

The penultimate line uses \( tr(AB) = tr(BA) \). The last line uses the fact that the expectation of the sum of the diagonal elements of a matrix is equal to the sum of their individual expectations. Assuming that \( X \), and hence \( M \), is nonstochastic:

\[ E(\hat{u} \hat{u}) = tr(ME(uu')) = tr(MI_n \sigma_u^2) = \sigma_u^2 tr(M) = \sigma_u^2 (tr(I_n) - tr(X[X'X]^{-1}X')) = \sigma_u^2 (tr(I_n) - tr(X[X'X]^{-1}X')). \]

The last step uses \( tr(A + B) = tr(A) + tr(B) \). The trace of an identity matrix is equal to its dimension. Hence:

\[ E(\hat{u} \hat{u}) = \sigma_u^2 (n - tr(X[X'X]^{-1}X')) = \sigma_u^2 (n - tr(X[X'X]^{-1})) = \sigma_u^2 (n - tr(I_k)) = \sigma_u^2 (n - k). \]

Hence \( \hat{u} \hat{u} / (n - k) \) is an unbiased estimator of \( \sigma_u^2 \).

15.20 Appendix C: Answers to the exercises

1. Given any square matrix \( C \), another matrix \( D \) is said to be its inverse if and only if \( CD = DC = I \). Thus, if \( B \) is the inverse of \( A \), \( AB = BA = I \). Now focus on the matrix \( B \). Since \( BA = AB = I \), \( A \) is its inverse. Hence the inverse of an inverse is the original matrix.

2. Suppose that two different matrices \( B \) and \( C \) both satisfied the conditions for being the inverse of \( A \). Then \( BA = I \) and \( AC = I \). Consider the matrix \( BAC \). Using \( BA = I \), \( BAC = C \). However, using \( AC = I \), \( BAC = B \). Hence \( B = C \) and it is not possible for \( A \) to have two separate inverses.

3. \( A_{ij} \), and hence \( A'_{ji} \), is the inner product of row \( i \) of \( B \) and column \( j \) of \( C \). If one writes \( D = C'B \), \( D_{ji} \) is the inner product of row \( j \) of \( C' \) and column \( i \) of \( B' \), that is, column \( j \) of \( C \) and row \( i \) of \( B \). Hence \( D_{ji} = A_{ij} \), so \( D = A' \) and \( C'B' = (BC)' \).

4. Let \( D \) be the inverse of \( A \). Then \( D \) must satisfy \( AD = DA = I \). Now \( A = BC \), so \( D \) must satisfy \( BCD = DBC = I \). \( C^{-1}B^{-1} \) satisfies both of these conditions, since \( BCC^{-1}B^{-1} = BIB^{-1} = I \) and \( C^{-1}B^{-1}BC = C^{-1}IC = I \). Hence \( C^{-1}B^{-1} \) is the inverse of \( BC \) (assuming that \( B^{-1} \) and \( C^{-1} \) exist).

334
5. Let $B = A^{-1}$. Then $BA = AB = I$. Hence, using the result from Exercise 3, $A'B' = B'A' = I' = I$. Hence $B'$ is the inverse of $A'$. In other words, $[A^{-1}]' = [A']^{-1}$.

6. The relationship $Y = \beta_1 + \beta_2 X + u$ may be written in linear algebra form as $y = X\beta + u$ where $X = [1 \ x]$ and 1 is the unit vector and:

$$x = \begin{bmatrix}
X_1 \\
\vdots \\
X_i \\
\vdots \\
X_n
\end{bmatrix}.$$

Then:

$$X'X = \begin{bmatrix} 1' & x' \end{bmatrix} \begin{bmatrix} 1 & x \end{bmatrix} = \begin{bmatrix} 1'1 & 1'x \\ x'1 & x'x \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}.$$

The determinant of $X'X$ is:

$$n \sum X_i^2 - \left( \sum X_i \right)^2 = n \sum X_i^2 - n^2 \bar{X}^2.$$

Hence:

$$[X'X]^{-1} = \frac{1}{n \sum X_i^2 - n^2 \bar{X}^2} \begin{bmatrix} \sum X_i^2 & -n \bar{X} \\ -n \bar{X} & n \end{bmatrix}.$$

We also have:

$$X'y = \begin{bmatrix} 1'y \\ x'y \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}.$$

So:

$$\hat{\beta} = [X'X]X'y$$

$$= \frac{1}{n \sum X_i^2 - n^2 \bar{X}^2} \begin{bmatrix} \sum X_i^2 & -n \bar{X} \\ -n \bar{X} & n \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$= \frac{1}{n \sum X_i^2 - n^2 \bar{X}^2} \begin{bmatrix} \bar{Y} \sum X_i^2 - n \bar{X} \sum X_i Y_i \\ \sum (X_i - \bar{X}) (Y_i - \bar{Y}) \end{bmatrix}.$$

Thus:

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

and:

$$\hat{\beta}_1 = \frac{\bar{Y} \sum X_i^2 - n \bar{X} \sum X_i Y_i}{\sum (X_i - \bar{X})^2}.$$
15. Regression analysis with linear algebra primer

\( \hat{\beta}_1 \) may be written in its more usual form as follows:

\[
b_1 = \frac{\sum X_i^2 - n\bar{X}^2 + \bar{Y}n\bar{X}^2 - \bar{X} \sum X_iY_i}{\sum (X_i - \bar{X})^2}
\]

\[
= \frac{\bar{Y} \left( \sum (X_i - \bar{X})^2 \right) - \bar{X} \left( \sum X_iY_i - n\bar{Y} \right)}{\sum (X_i - \bar{X})^2}
\]

\[
= \bar{Y} - \frac{\bar{X} \left( \sum (X_i - \bar{X}) (Y_i - \bar{Y}) \right)}{\sum (X_i - \bar{X})^2}
\]

\[
= \bar{Y} - \hat{\beta}_2 \bar{X}.
\]

7. If \( Y = \beta_2 X + u \), \( y = X \beta + u \) where:

\[
X = x = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{bmatrix}.
\]

Then:

\[
X'X = x'x = \sum X_i^2.
\]

The inverse of \( X'X \) is \( 1/\sum X_i^2 \). In this model, \( X'y = x'y = \sum X_iY_i \). So:

\[
\hat{\beta} = [X'X]^{-1}X'y = \frac{\sum X_iY_i}{\sum X_i^2}.
\]

8. If \( Y = \beta_1 + u \), \( y = X \beta + u \) where \( X = 1 \), the unit vector. Then \( X'X = 1'1 = n \) and its inverse is \( 1/n \).

\[
X'y = 1'y = \sum Y_i = n\bar{Y}.
\]

So:

\[
\hat{\beta} = [X'X]^{-1}X'y = \frac{1}{n}n\bar{Y} = \bar{Y}.
\]

9. We will start with \( Y \). If we regress it on the intercept, we are regressing it on \( 1 \), the unit vector, and, as we saw in Exercise 8, the coefficient is \( \bar{Y} \). Hence the residual in observation \( i \) is \( Y_i - \bar{Y} \). The same is true for each of the \( X \) variables when regressed on the intercept. So when we come to regress the residuals of \( Y \) on the residuals of the \( X \) variables, we are in fact using the demeaned data for \( Y \) and the demeaned data for the \( X \) variables.

10. The general form of the IV estimator is \( \hat{\beta}^{IV} = [W'X]^{-1}W'y \). In the case of the simple regression model, with \( Z \) acting as an instrument for \( X \) and the unit vector acting as an instrument for itself, \( W = [1 \ z'] \) and \( X = [1 \ x] \). Thus:

\[
W'X = \begin{bmatrix} 1' \\ z' \end{bmatrix} [1 \ x] = \begin{bmatrix} 1'1 & 1'x \\ z'1 & z'x \end{bmatrix} = \begin{bmatrix} \sum X_i \\ \sum Z_iZ_iX_i \end{bmatrix}.
\]
The determinant of $W'X$ is:

$$n \sum Z_i X_i - \left( \sum Z_i \right) \left( \sum X_i \right) = n \sum Z_i X_i - n^2 \bar{Z} \bar{X}.$$ 

Hence:

$$[W'X]^{-1} = \frac{1}{n \sum Z_i X_i - n^2 \bar{Z} \bar{X}} \begin{bmatrix} \sum Z_i X_i & -n \bar{X} \\ -n \bar{Z} & n \end{bmatrix}.$$

We also have:

$$W'y = \begin{bmatrix} 1' & z' \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Z_i Y_i \end{bmatrix}.$$

So:

$$\hat{\beta}^{IV} = [W'X]^{-1} W'y$$

$$= \frac{1}{n \sum Z_i X_i - n^2 \bar{Z} \bar{X}} \begin{bmatrix} \sum Z_i X_i & -n \bar{X} \\ -n \bar{Z} & n \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum Z_i Y_i \end{bmatrix}$$

$$= \frac{1}{n \sum Z_i X_i - n^2 \bar{Z} \bar{X}} \begin{bmatrix} \bar{Y} \sum Z_i X_i - \bar{X} \sum Z_i Y_i \\ \sum \left( Z_i - \bar{Z} \right) \left( Y_i - \bar{Y} \right) \end{bmatrix}.$$ 

Thus:

$$\hat{\beta}^{IV}_2 = \frac{\sum \left( Z_i - \bar{Z} \right) \left( Y_i - \bar{Y} \right)}{\sum \left( Z_i - \bar{Z} \right) \left( X_i - \bar{X} \right)}$$

and:

$$\hat{\beta}^{IV}_1 = \frac{\bar{Y} \sum Z_i X_i - \bar{X} \sum Z_i Y_i}{\sum \left( Z_i - \bar{Z} \right) \left( X_i - \bar{X} \right)}.$$

$\hat{\beta}^{IV}_1$ may be written in its more usual form as follows:

$$\hat{\beta}^{IV}_1 = \frac{\bar{Y} \left( \sum Z_i X_i - n \bar{Z} \bar{X} \right) + \bar{Y} n \bar{Z} \bar{X} - \bar{X} \sum Z_i Y_i}{\sum \left( Z_i - \bar{Z} \right) \left( X_i - \bar{X} \right)}$$

$$= \frac{\bar{Y} \left( \sum \left( Z_i - \bar{Z} \right) \left( X_i - \bar{X} \right) \right) - \bar{X} \left( \sum Z_i Y_i - n \bar{Z} \bar{X} \right)}{\sum \left( Z_i - \bar{Z} \right) \left( X_i - \bar{X} \right)}$$

$$= \bar{Y} - \frac{\bar{X} \left( \sum \left( Z_i - \bar{Z} \right) \left( Y_i - \bar{Y} \right) \right)}{\sum \left( Z_i - \bar{Z} \right) \left( X_i - \bar{X} \right)}$$

$$= \bar{Y} - \hat{\beta}^{IV}_2 \bar{X}.$$ 

11. By definition, if one of the variables in $X$ is acting as an instrument for itself, it is included in the $W$ matrix. If it is regressed on $W$, a perfect fit is obtained by
15. Regression analysis with linear algebra primer

assigning its column in \( W \) a coefficient of 1 and assigning zero values to all the other coefficients. Hence its fitted values are the same as its original values and it is not affected by the first stage of Two-Stage Least Squares.

12. If the variables in \( X \) are regressed on \( W \) and the matrix of fitted values of \( X \) saved:

\[
\hat{X} = W[W'W]^{-1}W'X.
\]

If \( \hat{X} \) is used as the matrix of instruments:

\[
\hat{\beta}_{\text{TLS}} = (\hat{X}'X)^{-1}\hat{X}'y = [X'W[W'W]^{-1}W'X][X'W[W'W]^{-1}W'y] = [W'X]^{-1}W'W[X'W][X'W]^{-1}W'y = [W'X]^{-1}W'y = \hat{\beta}_{\text{IV}}.
\]

Note that, in going from the second line to the third, we have used \([ABC]^{-1} = C^{-1}B^{-1}A^{-1}\), and we have exploited the fact that \( W'X \) is square and possesses an inverse.

13. The variance-covariance matrix of \( u \) is:

\[
\begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix}
\]

and hence the characteristic equation for the eigenvalues is:

\[(1 - \lambda)^2 - \rho^2 = 0.\]

The eigenvalues are therefore \( 1 - \rho \) and \( 1 + \rho \). Since we are told \(|\rho| < 1\), the matrix is positive definite.

Let:

\[
c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.
\]

If \( \lambda = 1 - \rho \), the matrix \( A - \lambda I \) is given by:

\[
A - \lambda I = \begin{bmatrix} \rho & \rho \\ \rho & \rho \end{bmatrix}
\]

and hence the equation:

\[
[A - \lambda I]c = 0
\]

yields:

\[
\rho c_1 + \rho c_2 = 0.
\]

Hence, also imposing the normalisation:

\[
c'c = c_1^2 + c_2^2 = 1
\]
we have $c_1 = 1/\sqrt{2}$ and $c_2 = -1/\sqrt{2}$, or vice versa. If $\lambda = 1 + \rho$:

$$A - \lambda I = \begin{bmatrix} -\rho & \rho \\ \rho & -\rho \end{bmatrix}$$

and hence $[A - \lambda I]c = 0$ yields:

$$-\rho c_1 + \rho c_2 = 0.$$ 

Hence, also imposing the normalisation:

$$c'c = c_1^2 + c_2^2 = 1$$

we have $c_1 = c_2 = 1/\sqrt{2}$. Thus:

$$C = \begin{bmatrix} \sqrt{1-\rho} & \sqrt{1+\rho} \\ \sqrt{1-\rho} & \sqrt{1+\rho} \end{bmatrix}$$

and:

$$P = \Lambda^{-1/2}C' = \begin{bmatrix} \sqrt{1-\rho} & 0 \\ 0 & \sqrt{1+\rho} \end{bmatrix} \begin{bmatrix} \sqrt{1-\rho} & -\sqrt{1-\rho} \\ \sqrt{1-\rho} & \sqrt{1-\rho} \end{bmatrix} = \begin{bmatrix} 1 & -\sqrt{1-\rho} \\ \sqrt{1-\rho} & 1 \end{bmatrix}.$$ 

It may then be verified that $P\Omega P' = I$:

$$= \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$ 

The transformed model has:

$$y^* = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{1-\rho}(y_1 - y_2) \\ \sqrt{1+\rho}(y_1 + y_2) \end{bmatrix}$$

and parallel transformations for the $X$ variables and $u$. Given that:

$$u^* = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{1-\rho}(u_1 - u_2) \\ \sqrt{1+\rho}(u_1 + u_2) \end{bmatrix}$$
Regression analysis with linear algebra primer

none of its elements is the white noise $\varepsilon$ in the AR(1) process, but nevertheless its elements are iid.

$$\text{var}(u^*_1) = \frac{1}{2} \frac{1}{1 - \rho} \left( \text{var}(u_1) + \text{var}(u_2) - 2\text{cov}(u_1, u_2) \right)$$

$$= \frac{1}{2} \frac{1}{1 - \rho} \left( \sigma^2_u + \sigma^2_u - 2\rho\sigma^2_u \right) = \sigma^2_u$$

$$\text{var}(u^*_2) = \frac{1}{2} \frac{1}{1 + \rho} \left( \text{var}(u_1) + \text{var}(u_2) + 2\text{cov}(u_1, u_2) \right)$$

$$= \frac{1}{2} \frac{1}{1 + \rho} \left( \sigma^2_u + \sigma^2_u + 2\rho\sigma^2_u \right) = \sigma^2_u$$

$$\text{cov}(u^*_1, u^*_2) = \frac{1}{2} \frac{1}{\sqrt{1 - \rho^2}} \text{cov} \left( (u_1 - u_2), (u_1 + u_2) \right)$$

$$= \frac{1}{2} \frac{1}{\sqrt{1 - \rho^2}} \left( \text{var}(u_1) + \text{cov}(u_1, u_2) - \text{cov}(u_2, u_1) - \text{var}(u_2) \right) = 0.$$

Hence $E(u^*u^{*\prime}) = I\sigma^2_u$. Of course, this was the objective of the $P$ transformation.
Appendix A

Syllabus for the EC2020 Elements of econometrics examination

This syllabus is intended to provide an explicit list of all the mathematical formulae and proofs that you are expected to know for the EC2020 Elements of Econometrics examination. You are warned that the examination is intended to be an opportunity for you to display your understanding of the material, rather than of your ability to reproduce standard items.

A.1 Review: Random variables and sampling theory


*Formulae and proofs:* This chapter is concerned with statistics, not econometrics, and is not examinable. However, you are expected to know the results in this chapter and to be able to use them.

A.2 Chapter 1 Simple regression analysis


*Formulae and proofs:* You are expected to know, and be able to derive, the expressions for the regression coefficients in a simple regression model, including variations where either the intercept or the slope coefficient may be assumed to be zero. You are expected to know the definition of $R^2$ and how it is related to the residual sum of squares. You are expected to know the relationship between $R^2$ and the correlation between the actual and fitted values of the dependent variable, but not to be able to prove it.
A.3 Chapter 2 Properties of the regression coefficients

Types of data and regression model. Assumptions for Model A. Regression coefficients as random variables. Unbiasedness of the regression coefficients. Precision of the regression coefficients. Gauss–Markov theorem. $t$ test of a hypothesis relating to a regression coefficient. Type I error and Type II error. Confidence intervals. One-sided tests. $F$ test of goodness of fit.

*Formulae and proofs:* You are expected to know the regression model assumptions for Model A. You are expected to know, though not be able to prove, that, in the case of a simple regression model, an $F$ test on the goodness of fit is equivalent to a two-sided $t$ test on the slope coefficient. You are expected to know how to make a theoretical decomposition of an estimator and hence how to investigate whether or not it is biased. In particular, you are expected to be able to show that the OLS estimator of the slope coefficient in a simple regression model can be decomposed into the true value plus a weighted linear combination of the values of the disturbance term in the sample. You are expected to be able to derive the expression for the variance of the slope coefficient in a simple regression model. You are expected to know how to estimate the variance of the disturbance term, given the residuals, but you are not expected to be able to derive the expression. You are expected to understand the Gauss–Markov theorem, but you are not expected to be able to prove it.

A.4 Chapter 3 Multiple regression analysis


*Formulae and proofs:* You are expected to know how, in principle, the multiple regression coefficients are derived, but you do not have to remember the expressions, nor do you have to be able to derive them mathematically. You are expected to know, but not to be able to derive, the expressions for the population variance of a slope coefficient and its standard error in a model with two explanatory variables. You are expected to be able to perform $F$ tests on the goodness of fit of the model as a whole and for the improvement in fit when a group of explanatory variables is added to the model. You are expected to be able to demonstrate the properties of predictions within the context of the classical linear regression model. In particular, you are expected to be able to demonstrate that the expected value of the prediction error is 0, if the model is correctly specified and the regression model assumptions are satisfied. You are not expected to know the population variance of the prediction error.
A.5 Chapter 4 Transformation of variables


*Formulae and proofs:* You are expected to know how to perform a Box–Cox transformation for comparing the goodness of fit of alternative versions of a model with $Y$ and $\log Y$ as the dependent variable.

A.6 Chapter 5 Dummy variables


*Formulae and proofs:* You are expected to be able to perform a Chow test and a test of the explanatory power of a group of dummy variables, and to understand the relationship between them.

A.7 Chapter 6 Specification of regression variables


*Formulae and proofs:* You are expected to be able to derive the expression for omitted variable bias when the true model has two explanatory variables and the fitted model omits one of them. You are expected to know how to perform an $F$ test on the validity of a linear restriction, given appropriate data on the residual sum of squares. You are expected to understand the logic behind the $t$ test of a linear restriction and to be able to reparameterise a regression specification to perform such a test in a simple context. You are expected to be able to perform $F$ tests of multiple linear restrictions.

A.8 Chapter 7 Heteroskedasticity


*Formulae and proofs:* You are expected to know how to perform the Goldfeld–Quandt and White tests for heteroskedasticity.
A. Syllabus for the EC2020 Elements of econometrics examination

A.9 Chapter 8 Stochastic regressors and measurement errors

Stochastic regressors. Assumptions for models with stochastic regressors. Finite sample and asymptotic properties of the regression coefficients in models with stochastic regressors. Measurement error and its consequences. Friedman’s Permanent Income Hypothesis. Instrumental variables (IV). Asymptotic properties of IV estimators, including the asymptotic limiting distribution of \( \sqrt{n}(\hat{\beta}_2^{IV} - \beta_2) \). \( \hat{\beta}_2^{IV} \) is the IV estimator of \( \beta_2 \) in a simple regression model. Use of simulation to investigate the finite-sample properties of estimators when only asymptotic properties can be determined analytically. Application of the Durbin–Wu–Hausman test.

 формула и доказательства: You are expected to be able to demonstrate that, in a simple regression model, the OLS estimator of the slope coefficient is inconsistent when there is measurement error in the explanatory variable. You should know the expression for the bias and be able to derive it. You should be able to explain the consequences of measurement error in the dependent variable. You should know the expression for an instrumental variable estimator of the slope coefficient in a simple regression model and be able to demonstrate that it yields consistent estimates, provided that certain assumptions are satisfied. You should also know the expression for the asymptotic population variance of an instrumental variable estimator in a simple regression model and to understand why it provides only an approximation for finite samples. You are not expected to know the formula for the Durbin–Wu–Hausman test.

A.10 Chapter 9 Simultaneous equations estimation


 формула и доказательства: You are expected to be able to derive an expression for simultaneous equations bias in a simple regression equation and to be able to demonstrate the consistency of an IV estimator in a simple regression equation. You are expected to be able to explain in general terms why TSLS is used in overidentified models.

A.11 Chapter 10 Binary choice models and maximum likelihood estimation


 формула и доказательства: You are expected to know the expression for the probability of an
event occurring in the logit model, and to know the expressions for the marginal functions in the logit and probit models. You would not be expected to calculate marginal effects in an examination, but you should be able to explain how they are calculated and to comment on calculations of them. You are expected to be able to derive a maximum likelihood estimator in a simple example. In more complex examples, you would only be expected to explain how the estimates are obtained, in principle. You are expected to be able to perform, from first principles, likelihood ratio tests in a simple context.

A.12 Chapter 11 Models using time series data

Static demand functions fitted using aggregate time series data. Lagged variables and naive attempts to model dynamics. Autoregressive distributed lag (ADL) models with applications in the form of the partial adjustment and adaptive expectations models. Error correction models. Asymptotic properties of OLS estimators of ADL models, including asymptotic limiting distributions. Use of simulation to investigate the finite sample properties of parameter estimators for the ADL(1,0) model. Use of predetermined variables as instruments in simultaneous equations models using time series data. (Section 11.7 of the textbook, Alternative dynamic representations..., is not in the syllabus.)

Formulae and proofs: You are expected to be able to analyze the short-run and long-run dynamics inherent in ADL(1,0) models in general and the adaptive expectations and partial adjustment models in particular. You are expected to be able to explain why the OLS estimators of the parameters of ADL(1,0) models are subject to finite-sample bias and, within the context of the model \( Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t \) to be able to demonstrate that they are consistent.

A.13 Chapter 12 Autocorrelation


Formulae and proofs: You are expected to know how to perform the tests for autocorrelation mentioned above and to know how to perform a common factor test. You are expected to be able to explain why the properties of estimators obtained by fitting the AR(1) nonlinear regression specification are not necessarily superior to those obtained using OLS.
A. Syllabus for the EC2020 Elements of Econometrics examination

A.14 Chapter 13 Introduction to nonstationary processes


Formulae and proofs: You are expected to be able to determine whether a simple random process is stationary or nonstationary. You would not be expected to perform a unit root test in an examination, but you are expected to understand the test and to be able to comment on the results of such a test.

A.15 Chapter 14 Introduction to panel data models

Definition of panel data set (longitudinal data set). Pooled OLS model. Definition of, and consequences of, unobserved heterogeneity. Withingroups fixed effects model. First differences fixed-effects model. Least squares dummy variable model. Calculation of degrees of freedom in fixed effects models. Random effects model, with assumption required for the use of this model. \( F \) test for discriminating between fixed effects and pooled OLS as the appropriate specification. Durbin–Wu–Hausman test for discriminating between fixed effects and random effects models as the appropriate specification.

Formulae and proofs: You are expected to be able to demonstrate mathematically how the within-groups and first differences versions of the fixed effects approach eliminate unobserved heterogeneity, and to be able to explain how the least squares dummy variable model provides an alternative solution. You are expected to be able to explain mathematically why the random effects model is subject to a form of autocorrelation.