

# Contents

<i>List of Figures</i>	xxi
<i>List of Tables</i>	xxiii
<i>Preface</i>	xxvii
<i>Acknowledgements</i>	xxx
<b>Part I Statistical Background and Basic Data Handling</b>	<b>1</b>
<b>1 Fundamental Concepts</b>	<b>3</b>
Introduction	4
A simple example	4
A statistical framework	6
Properties of the sampling distribution of the mean	7
Hypothesis testing and the central limit theorem	8
Central limit theorem	10
Conclusion	13
<b>2 The Structure of Economic Data and Basic Data Handling</b>	<b>14</b>
Learning objectives	14
The structure of economic data	15
Cross-sectional data	15
Time series data	15
Panel data	16
Basic data handling	17
Looking at raw data	17
Graphical analysis	17
Summary statistics	19
Questions	25
<b>Part II The Classical Linear Regression Model</b>	<b>27</b>
<b>3 Simple Regression</b>	<b>29</b>
Learning objectives	29
Introduction to regression: the classical linear regression model (CLRM)	30
Why do we do regressions?	30

x *Contents*

The classical linear regression model	30
The Ordinary Least Squares (OLS) method of estimation	32
Alternative expressions for $\hat{\beta}$	34
The assumptions of the CLRM	35
General	35
The assumptions	36
Violations of the assumptions	37
Properties of the OLS estimators	38
Linearity	38
Unbiasedness	39
Efficiency and BLUEness	40
Consistency	42
The overall goodness of fit	43
Problems associated with $R^2$	44
Hypothesis testing and confidence intervals	45
Testing the significance of the OLS coefficients	46
Confidence intervals	47
How to estimate a simple regression in EViews and Stata	48
Simple regression in EViews	48
Simple regression in Stata	48
Reading the Stata simple regression results output	49
Reading the EViews simple regression results output	49
Presentation of regression results	50
Economic theory applications	50
Application 1: the demand function	50
Application 2: the production function	51
Application 3: Okun's law	52
Application 4: the Keynesian consumption function	52
Computer example: the Keynesian consumption function	53
Solution	53
Questions and exercises	58
<b>4 Multiple Regression</b>	<b>62</b>
Learning objectives	62
Introduction	64
Derivation of multiple regression coefficients	64
The three-variable model	64
The $k$ -variables case	65
Derivation of the coefficients with matrix algebra	66
The structure of the $X'X$ and $X'Y$ matrices	67
The assumptions of the multiple regression model	68
The variance–covariance matrix of the errors	69
Properties of multiple regression model OLS estimators	69
Linearity	69
Unbiasedness	70
Consistency	70
BLUEness	70
$R^2$ and adjusted $R^2$	72

General criteria for model selection	73
Multiple regression estimation in EViews and Stata	74
Multiple regression in EViews	74
Multiple regression in Stata	74
Reading the EViews multiple regression results output	75
Hypothesis testing	75
Testing individual coefficients	75
Testing linear restrictions	75
The $F$ -form of the likelihood ratio test	77
Testing the joint significance of the $X$ s	78
$F$ -test for overall significance in EViews	78
Adding or deleting explanatory variables	79
Omitted and redundant variables test in EViews	79
How to perform the Wald test in EViews	80
The $t$ -test (a special case of the Wald procedure)	80
The Lagrange multiplier (LM) test	81
The LM test in EViews	82
Computer example: Wald, omitted and redundant variables tests	82
A Wald test of coefficient restrictions	83
A redundant variable test	83
An omitted variable test	84
Computer example: commands for Stata	84
Financial econometrics application: the capital asset pricing model in action	87
A few theoretical remarks regarding the CAPM	87
The empirical application of the CAPM	89
EViews programming and the CAPM application	90
Advanced EViews programming and the CAPM application	96
Questions and exercises	97
<b>Part III Violating the Assumptions of the CLRM</b>	<b>101</b>
<b>5 Multicollinearity</b>	<b>103</b>
Learning objectives	103
Introduction	104
Perfect multicollinearity	104
Consequences of perfect multicollinearity	105
Imperfect multicollinearity	106
Consequences of imperfect multicollinearity	107
Detecting problematic multicollinearity	109
Simple correlation coefficient	109
$R^2$ from auxiliary regressions	109
Computer examples	110
Example 1: induced multicollinearity	110
Example 2: with the use of real economic data	112
Questions and exercises	115
Questions	115

xii *Contents*

<b>6</b>	<b>Heteroskedasticity</b>	<b>119</b>
	Learning objectives	119
	Introduction: what is heteroskedasticity?	120
	Consequences of heteroskedasticity	
	for OLS estimators	122
	A general approach	122
	A mathematical approach	123
	Detecting heteroskedasticity	126
	The informal way	126
	The Breusch–Pagan LM test	127
	The Glesjer LM test	130
	The Harvey–Godfrey LM test	132
	The Park LM test	133
	Criticism of the LM tests	135
	The Goldfeld–Quandt test	135
	White’s test	137
	Computer example: heteroskedasticity tests	139
	The Breusch–Pagan test	140
	The Glesjer test	142
	The Harvey–Godfrey test	142
	The Park test	143
	The Goldfeld–Quandt test	144
	White’s test	146
	Commands for the computer example in Stata	146
	Engle’s ARCH test*	148
	Computer example of the ARCH-LM test	149
	Resolving heteroskedasticity	150
	Generalized (or weighted) least squares	150
	Computer example: resolving heteroskedasticity	152
	Questions and exercises	155
<b>7</b>	<b>Autocorrelation</b>	<b>159</b>
	Learning objectives	159
	Introduction: what is autocorrelation?	160
	What causes autocorrelation?	160
	First- and higher-order autocorrelation	161
	Consequences of autocorrelation	
	for the OLS estimators	162
	A general approach	162
	A more mathematical approach	163
	Detecting autocorrelation	165
	The graphical method	165
	Example: detecting autocorrelation using the graphical method	165
	The Durbin–Watson test	167
	Computer example of the DW test	169
	The Breusch–Godfrey LM test for serial correlation	170
	Computer example of the Breusch–Godfrey test	171
	Durbin’s <i>h</i> -test in the presence of lagged dependent variables	173
	Computer example of Durbin’s <i>h</i> -test	174

Resolving autocorrelation	175
When $\rho$ is known	176
Computer example of the generalized differencing approach	176
When $\rho$ is unknown	178
Computer example of the iterative procedure	179
Resolving autocorrelation in Stata	181
Questions and exercises	181
Appendix	182
<b>8 Misspecification: Wrong Regressors, Measurement Errors and Wrong Functional Forms</b>	<b>184</b>
Learning objectives	184
Introduction	185
Omitting influential or including non-influential explanatory variables	185
Consequences of omitting influential variables	185
Including a non-influential variable	186
Omission and inclusion of relevant and irrelevant variables at the same time	187
The plug-in solution in the omitted variable bias	187
Various functional forms	189
Introduction	189
Linear-log functional form	189
Reciprocal functional form	190
Polynomial functional form	190
Functional form including interaction terms	191
Log-linear functional form	192
The double-log functional form	192
The Box–Cox transformation	193
Measurement errors	194
Measurement error in the dependent variable	195
Measurement error in the explanatory variable	195
Tests for misspecification	197
Normality of residuals	197
The Ramsey RESET test for general misspecification	199
Tests for non-nested models	201
Computer example: the Box–Cox transformation in EViews	203
Approaches in choosing an appropriate model	206
The traditional view: average economic regression	206
The Hendry ‘general to specific approach’	207
Questions and exercises	208
<b>Part IV Topics in Econometrics</b>	<b>211</b>
<b>9 Dummy Variables</b>	<b>213</b>
Learning objectives	213
Introduction: the nature of qualitative information	214
The use of dummy variables	214

## xiv Contents

Constant dummy variables	214
Slope dummy variables	216
The combined effect of intercept and slope dummies	218
Computer example of the use of dummy variables	219
Using a constant dummy	220
Using a slope dummy	220
Using both dummies together	221
Special cases of the use of dummy variables	222
Using dummy variables with multiple categories	222
Using more than one dummy variable	224
Using seasonal dummy variables	225
Computer example of dummy variables with multiple categories	226
Financial econometrics application: the January effect in emerging stock markets	228
Tests for structural stability	231
The dummy variable approach	231
The Chow test for structural stability	231
Financial econometrics application: the day-of-the-week effect in action	232
How to create daily dummies in Stata	234
Questions and exercises	235
<b>10 Dynamic Econometric Models</b>	<b>239</b>
Learning objectives	239
Introduction	240
Distributed lag models	240
The Koyck transformation	241
The Almon transformation	243
Other models of lag structures	244
Autoregressive models	244
The partial adjustment model	244
Computer example of the partial adjustment model	245
The adaptive expectations model	247
Tests of autocorrelation in autoregressive models	249
Exercises	249
<b>11 Simultaneous Equation Models</b>	<b>251</b>
Learning objectives	251
Introduction: basic definitions	252
Consequences of ignoring simultaneity	253
The identification problem	253
Basic definitions	253
Conditions for identification	254
Example of the identification procedure	255
A second example: the macroeconomic model of a closed economy	255
Estimation of simultaneous equation models	256
Estimation of an exactly identified equation: the ILS method	257

Estimation of an over-identified equation: the TSLS method	257
Computer example: the IS–LM model	258
Estimation of simultaneous equations in Stata	261
Exercises	262
<b>12 Limited Dependent Variable Regression Models</b>	<b>265</b>
Learning objectives	265
Introduction	266
The linear probability model	266
Problems with the linear probability model	267
$\hat{D}_i$ is not bounded by the (0,1) range	267
Non-normality and heteroskedasticity of the disturbances	268
The coefficient of determination as a measure of overall fit	268
The logit model	269
A general approach	269
Interpretation of the estimates in logit models	270
Goodness of fit	271
A more mathematical approach	272
The probit model	274
A general approach	274
A more mathematical approach	275
Multinomial and ordered logit and probit models	276
Multinomial logit and probit models	277
Ordered logit and probit models	277
The Tobit model	278
Computer example: probit and logit models in EViews and Stata	278
Logit and probit models in EViews	278
Logit and probit models in Stata	281
Exercises	282
<b>Part V Time Series Econometrics</b>	<b>285</b>
<b>13 ARIMA Models and the Box–Jenkins Methodology</b>	<b>287</b>
Learning objectives	287
An introduction to time series econometrics	288
ARIMA models	288
Stationarity	289
Autoregressive time series models	289
The AR(1) model	289
The AR( $p$ ) model	291
Properties of the AR models	293
Moving average models	294
The MA(1) model	294
The MA( $q$ ) model	294
Invertibility in MA models	295
Properties of the MA models	296
ARMA models	297
Integrated processes and the ARIMA models	297

xvi *Contents*

An integrated series	297
Example of an ARIMA model	298
Box–Jenkins model selection	298
Identification	299
Estimation	300
Diagnostic checking	300
The Box–Jenkins approach step by step	301
Computer example: the Box–Jenkins approach	301
The Box–Jenkins approach in EViews	301
The Box–Jenkins approach in Stata	305
Questions and exercises	307
<b>14 Modelling the Variance: ARCH–GARCH Models</b>	<b>309</b>
Learning objectives	309
Introduction	310
The ARCH model	311
The ARCH(1) model	312
The ARCH( $q$ ) model	312
Testing for ARCH effects	313
Estimation of ARCH models by iteration	313
Estimating ARCH models in EViews	314
A more mathematical approach	318
The GARCH model	321
The GARCH( $p, q$ ) model	321
The GARCH(1, 1) model as an infinite ARCH process	321
Estimating GARCH models in EViews	322
Alternative specifications	323
The GARCH in mean, or GARCH-M, model	324
Estimating GARCH-M models in EViews	325
The threshold GARCH (TGARCH) model	328
Estimating TGARCH models in EViews	328
The exponential GARCH (EGARCH) model	329
Estimating EGARCH models in EViews	330
Adding explanatory variables in the mean equation	331
Adding explanatory variables in the variance equation	331
Estimating ARCH/GARCH-type models in Stata	332
Advanced EViews programming for the estimation of GARCH-type models	334
Application: a GARCH model of UK GDP and the effect of socio-political instability	338
Questions and exercises	342
<b>15 Vector Autoregressive (VAR) Models and Causality Tests</b>	<b>346</b>
Learning objectives	346
Vector autoregressive (VAR) models	347
The VAR model	347
Pros and cons of the VAR models	348
Causality tests	349
The Granger causality test	349



The Sims causality test	351
Financial econometrics application: financial development and economic growth – what is the causal relationship?	351
Estimating VAR models and causality tests in EViews and Stata	354
Estimating VAR models in EViews	354
Estimating VAR models in Stata	357
Exercises	359
<b>16 Non-Stationarity and Unit-Root Tests</b>	<b>362</b>
Learning objectives	362
Introduction	363
Unit roots and spurious regressions	363
What is a unit root?	363
Spurious regressions	366
Explanation of the spurious regression problem	368
Testing for unit roots	370
Testing for the order of integration	370
The simple Dickey–Fuller (DF) test for unit roots	370
The augmented Dickey–Fuller (ADF) test for unit roots	372
The Phillips–Perron (PP) test	372
Unit-root tests in EViews and Stata	374
Performing unit-root tests in EViews	374
Performing unit-root tests in Stata	376
Application: unit-root tests on various macroeconomic variables	377
Financial econometrics application: unit-root tests for the financial development and economic growth case	379
Questions and exercises	381
<b>17 Cointegration and Error-Correction Models</b>	<b>383</b>
Learning objectives	383
Introduction: what is cointegration?	384
Cointegration: a general approach	384
Cointegration: a more mathematical approach	385
Cointegration and the error-correction mechanism (ECM): a general approach	386
The problem	386
Cointegration (again)	387
The error-correction model (ECM)	387
Advantages of the ECM	387
Cointegration and the error-correction mechanism: a more mathematical approach	388
A simple model for only one lagged term of $X$ and $Y$	388
A more general model for large numbers of lagged terms	390
Testing for cointegration	392
Cointegration in single equations: the Engle–Granger approach	392
Drawbacks of the EG approach	394
The EG approach in EViews and Stata	395
Cointegration in multiple equations and the Johansen approach	396
Advantages of the multiple-equation approach	397
The Johansen approach (again)	397

xviii *Contents*

The steps of the Johansen approach in practice	398
The Johansen approach in EViews and Stata	403
Financial econometrics application: cointegration tests for the financial development and economic growth case	408
Monetization ratio	409
Turnover ratio	412
Claims and currency ratios	412
A model with more than one financial development proxy variable	414
Questions and exercises	416
<b>18 Identification in Standard and Cointegrated Systems</b>	<b>418</b>
Learning objectives	418
Introduction	419
Identification in the standard case	419
The order condition	421
The rank condition	422
Identification in cointegrated systems	422
A worked example	424
Computer example of identification	426
Conclusion	428
Questions and exercises	429
<b>19 Solving Models</b>	<b>430</b>
Learning objectives	430
Introduction	431
Solution procedures	431
Model add factors	433
Simulation and impulse responses	434
Stochastic model analysis	435
Setting up a model in EViews	437
Conclusion	440
Exercises	441
<b>20 Time-Varying Coefficient Models: A New Way of Estimating Bias-Free Parameters</b>	<b>442</b>
Learning objectives	442
Introduction	443
TVC estimation	444
Theorem 1	445
Coefficient drivers	446
Assumption 1 (auxiliary information)	446
Assumption 2	446
Choosing coefficient drivers	447
First requirement: selecting the complete driver set	447
Second requirement: splitting the driver set	448
Financial econometrics application: rating agencies' decisions and the sovereign bond spread between Greece and Germany	451
Conclusion	456
Questions and exercises	456

<b>Part VI Panel Data Econometrics</b>	<b>457</b>
<b>21 Traditional Panel Data Models</b>	<b>459</b>
Learning objectives	459
Introduction: the advantages of panel data	460
The linear panel data model	461
Different methods of estimation	461
The common constant method	461
The fixed effects method	462
The random effects method	463
The Hausman test	464
Computer examples with panel data	465
Inserting panel data in EViews	465
Estimating a panel data regression in EViews	469
The Hausman test in EViews	470
Estimating a panel data regression in Stata	473
The Hausman test in Stata	474
<b>22 Dynamic Heterogeneous Panels</b>	<b>475</b>
Learning objectives	475
Introduction	476
Bias in dynamic panels	476
Bias in the simple OLS estimator	476
Bias in the fixed effects model	477
Bias in the random effects model	477
Solutions to the bias problem (caused by the dynamic nature of the panel)	477
Bias of heterogeneous slope parameters	478
Solutions to heterogeneity bias: alternative methods of estimation	479
The mean group (MG) estimator	479
The pooled mean group (PMG) estimator	480
Application: the effects of uncertainty in economic growth and investment	482
Evidence from traditional panel data estimation	482
Mean group and pooled mean group estimates	483
<b>23 Non-Stationary Panels</b>	<b>485</b>
Learning objectives	485
Introduction	486
Panel unit-root tests	486
The Levin and Lin (LL) test	487
The Im, Pesaran and Shin (IPS) test	488
The Maddala and Wu (MW) test	489
Computer examples of panel unit-root tests	489
Panel cointegration tests	491
Introduction	491
The Kao test	492
The McCoskey and Kao test	493
The Pedroni tests	494
The Larsson <i>et al.</i> test	495
Computer examples of panel cointegration tests	496

<b>Part VII Using Econometric Software</b>	<b>501</b>
<b>24 Practicalities of Using EViews and Stata</b>	<b>503</b>
About EViews	504
Starting up with EViews	504
Creating a workfile and importing data	506
Copying and pasting data	506
Verifying and saving the data	507
Examining the data	507
Commands, operators and functions	508
About Stata	509
Starting up with Stata	509
The Stata menu and buttons	510
Creating a file when importing data	511
Copying/pasting data	512
Cross-sectional and time series data in Stata	512
First way – time series data with no time variable	512
Second way – time series data with time variable	513
Time series – daily frequency	513
Time series – monthly frequency	514
All frequencies	515
Saving data	515
Basic commands in Stata	516
Understanding command syntax in Stata	517
<i>Appendix: Statistical Tables</i>	<b>519</b>
<i>Bibliography</i>	<b>525</b>
<i>Index</i>	<b>531</b>

# 1 Fundamental Concepts

## CHAPTER CONTENTS

Introduction	4
A simple example	4
A statistical framework	6
Properties of the sampling distribution of the mean	7
Hypothesis testing and the central limit theorem	8
Conclusion	13

## 4 Statistical background and basic data handling

### Introduction

This chapter outlines some of the fundamental concepts that lie behind much of the rest of this book, including the ideas of a population distribution and a sampling distribution, the importance of random sampling, the law of large numbers and the central limit theorem. It then goes on to show how these ideas underpin the conventional approach to testing hypotheses and constructing confidence intervals.

Econometrics has a number of roles in terms of forecasting and analysing real data and problems. At the core of these roles, however, is the desire to pin down the magnitudes of effects and test their significance. Economic theory often points to the direction of a causal relationship (if income rises we may expect consumption to rise), but theory rarely suggests an exact magnitude. Yet, in a policy or business context, having a clear idea of the magnitude of an effect may be extremely important, and this is the realm of econometrics.

The aim of this chapter is to clarify some basic definitions and ideas in order to give the student an intuitive understanding of these underlying concepts. The account given here will therefore deliberately be less formal than much of the material later in the book.

### A simple example

Consider a very simple example to illustrate the idea we are putting forward here. Table 1.1 shows the average age at death for both men and women in the 15 European countries that made up the European Union (EU) before its enlargement.

Simply looking at these figures makes it fairly obvious that women can expect to live longer than men in each of these countries, and if we take the average across all

**Table 1.1** Average age at death for the EU15 countries (2002)

	<i>Women</i>	<i>Men</i>
Austria	81.2	75.4
Belgium	81.4	75.1
Denmark	79.2	74.5
Finland	81.5	74.6
France	83.0	75.5
Germany	80.8	74.8
Greece	80.7	75.4
Ireland	78.5	73.0
Italy	82.9	76.7
Luxembourg	81.3	74.9
Netherlands	80.6	75.5
Portugal	79.4	72.4
Spain	82.9	75.6
Sweden	82.1	77.5
UK	79.7	75.0
beginlwmath26pt] Mean	81.0	75.1
Standard deviation	1.3886616	1.2391241

countries we can clearly see that again, on a Europe-wide basis, women tend to live longer than men. However, there is quite considerable variation between the countries, and it might be reasonable to ask whether in general, in the world population, we would expect women to live longer than men.

A natural way to approach this would be to look at the difference in the mean life expectancy between men and women for the whole of Europe and to ask whether this is significantly different from zero. This involves a number of fundamental steps: first the difference in average life expectancy has to be estimated, then a measure of its uncertainty must be constructed, and finally the hypothesis that the difference is zero needs to be tested.

Table 1.1 gives the average (or mean) life expectancy for men and women for the EU as a whole, simply defined as:

$$\bar{Y}_w = \frac{1}{15} \sum_{i=1}^{15} Y_{wi} \quad \bar{Y}_m = \frac{1}{15} \sum_{i=1}^{15} Y_{mi} \quad (1.1)$$

where  $\bar{Y}_w$  is the EU average life expectancy for women and  $\bar{Y}_m$  is the EU average life expectancy for men. A natural estimate of the difference between the two means is  $(\bar{Y}_w - \bar{Y}_m)$ . Table 1.1 also gives the average dispersion for each of these means, defined as the standard deviation, which is given by:

$$S.D._j = \sqrt{\sum_{i=1}^{15} (Y_{ji} - \bar{Y}_j)^2} \quad j = w, m \quad (1.2)$$

As we have an estimate of the difference and an estimate of the uncertainty of our measures, we can now construct a formal hypothesis test. The test for the difference between two means is:

$$t = \frac{\bar{Y}_w - \bar{Y}_m}{\sqrt{\frac{s_w^2}{15} + \frac{s_m^2}{15}}} = \frac{81 - 75.1}{\sqrt{\frac{1.389^2}{15} + \frac{1.24^2}{15}}} = 12.27 \quad (1.3)$$

The  $t$ -statistic of 12.27 is greater than 1.96, which means that there is less than a 5% chance of finding a  $t$ -statistic of 12.27 purely by chance when the true difference is zero. Hence we can conclude that there is a significant difference between the life expectancies of men and women.

Although this appears very intuitive and simple, there are some underlying subtleties, and these are the subject of this chapter. The questions to be explored are: what theoretical framework justifies all this? Why is the difference in means a good estimate of extra length of life for women? Is this a good estimate for the world as a whole? What is the measure of uncertainty captured by the standard deviation, and what does it really mean? In essence, what is the underlying theoretical framework that justifies what happened?

## A statistical framework

The statistical framework that underlies the approach above rests on a number of key concepts, the first of which is the population. We assume that there is a population of events or entities that we are interested in. This population is assumed to be infinitely large and comprises all the outcomes that concern us. The data in Table 1.1 are for the EU15 countries for the year 2002. If we were interested only in this one year for this one set of countries, then there would be no statistical question to be asked. According to the data, women lived longer than men in that year in that area. That is simply a fact. But the population is much larger; it comprises all men and women in all periods, and to make an inference about this population we need some statistical framework. It might, for example, just be chance that women lived longer than men in that one year. How can we determine this?

The next important concepts are random variables and the population distribution. A random variable is simply a measurement of any event that occurs in an uncertain way. So, for example, the age at which a person dies is uncertain, and therefore the age of an individual at death is a random variable. Once a person dies, the age at death ceases to be a random variable and simply becomes an observation or a number. The population distribution defines the probability of a certain event happening; for example, the population distribution would define the probability of a man dying before he is 60 ( $\Pr(Y_m < 60)$ ). The population distribution has various moments that define its shape. The first two moments are the mean (sometimes called the expected value,  $E(Y_m) = \mu_{Y_m}$ , or the average) and the variance ( $E(Y_m - \mu_{Y_m})^2$ , which is the square of the standard deviation and is often defined as  $\sigma_{Y_m}^2$ ).

The moments described above are sometimes referred to as the unconditional moments; that is to say, they apply to the whole population distribution. But we can also condition the distribution and the moments on a particular piece of information. To make this clear, consider the life expectancy of a man living in the UK. Table 1.1 tells us that this is 75 years. What, then, is the life expectancy of a man living in the UK who is already 80? Clearly not 75! An unconditional moment is the moment for the complete distribution under consideration; a conditional moment is the moment for those members of the population who fulfil some condition, in this case being 80. We can consider a conditional mean  $E(Y_m | Y_{im} = 80)$ , in this case the mean of men aged 80, or conditional higher moments such as the conditional variance, which will be the subject of a later chapter. This is another way of thinking of subgroups of the population: we could think of the population as consisting of all people, or we could think of the distribution of the population of men and women separately. What we would like to know about is the distribution of the population we are interested in, that is, the mean of the life expectancy of all men and all women. If we could measure this, again there would be no statistical issue to address; we would simply know whether, on average, women live longer than men. Unfortunately, typically we can only ever have direct measures on a sample drawn from the population, and we have to use this sample to draw some inference about the population.

If the sample obeys some basic properties we can proceed to construct a method of deriving inference. The first key idea is that of random sampling: the individuals who make up our sample should be drawn at random from the population. The life expectancy of a man is a random variable; that is to say, the age at death of any individual is uncertain. Once we have observed the age at death and the



observation becomes part of our sample it ceases to be a random variable. The data set then comprises a set of individual observations, each of which has been drawn at random from the population. So our sample of ages at death for men becomes  $Y_m = (Y_{1m}, Y_{2m}, \dots, Y_{nm})$ . The idea of random sampling has some strong implications: because any two individuals are drawn at random from the population they should be **independent** of each other; that is to say, knowing the age at death of one man tells us nothing about the age at death of the other man. Also, as both individuals have been drawn from the same population, they should have an **identical distribution**. So, based on the assumption of random sampling, we can assert that each of the observations in our sample should have an independent and identical distribution; this is often expressed as iid.

We are now in a position to begin to construct a statistical framework. We want to make some inference about a population distribution from which only a sample has been observed. How can we know whether the method we choose to analyse the sample is a good one or not? The answer to this question lies in another concept, called the sampling distribution. If we draw a sample from our population, let's suppose we have a method for analysing that sample. It could be anything; for example, take the odd-numbered observations and sum them and divide by 20. This will give us an estimate. If we had another sample this would give us another estimate, and if we kept drawing samples this would give us a whole sequence of estimates based on this technique. We could then look at the distribution of all these estimates, and this would be the sampling distribution of this particular technique. Suppose the estimation procedure produces an estimate of the population mean which we call  $\tilde{Y}_m$ , then the sampling distribution will have a mean and a variance  $E(\tilde{Y}_m)$  and  $E(\tilde{Y}_m - E(\tilde{Y}_m))^2$ ; in essence, the sampling distribution of a particular technique tells us most of what we need to know about the technique. A good estimator will generally have the property of **unbiasedness**, which implies that its mean value is equal to the population feature we want to estimate. That is,  $E(\tilde{Y}_m) = \eta$ , where  $\eta$  is the feature of the population we wish to measure. In the case of unbiasedness, even in a small sample we expect the estimator to get the right answer on average. A slightly weaker requirement is **consistency**; here we only expect the estimator to get the answer correct if we have an infinitely large sample,  $\lim_{n \rightarrow \infty} E(\tilde{Y}_m) = \eta$ . A good estimator will be either unbiased or consistent, but there may be more than one possible procedure which has this property. In this case we can choose between a number of estimators on the basis of **efficiency**; this is simply given by the variance of the sampling distribution. Suppose we have another estimation technique, which gives rise to  $\tilde{\tilde{Y}}$ , which is also unbiased; then we would prefer  $\tilde{Y}$  to this procedure if  $\text{var}(\tilde{Y}) < \text{var}(\tilde{\tilde{Y}})$ . This simply means that, on average, both techniques get the answer right, but the errors made by the first technique are, on average, smaller.

## Properties of the sampling distribution of the mean

In the example above, based on Table 1.1, we calculated the mean life expectancy of men and women. Why is this a good idea? The answer lies in the sampling distribution of the mean as an estimate of the population mean. The mean of the sampling distribution of the mean is given by:

## 8 Statistical background and basic data handling

$$E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n} \sum_{i=1}^n E(Y_i) = \frac{1}{n} \sum_{i=1}^n \mu_Y = \mu_Y \quad (1.4)$$

So the expected value of the mean of a sample is equal to the population mean, and hence the mean of a sample is an unbiased estimate of the mean of the population distribution. The mean thus fulfils our first criterion for being a good estimator. But what about the variance of the mean?

$$\begin{aligned} \text{var}(\bar{Y}) &= E(\bar{Y} - \mu_Y)^2 = E\left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Y_i - \mu_Y)(Y_j - \mu_Y)\right) \\ \text{beginlwmath28pt]} &= \frac{1}{n^2} \left( \sum_{i=1}^n \text{var}(Y_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \text{cov}(Y_i, Y_j) \right) = \frac{\sigma_Y^2}{n} \end{aligned} \quad (1.5)$$

So the variance of the mean around the true population mean is related to the sample size that is used to construct the mean and the variance of the population distribution. As the sample size increases, the variance in the population shrinks, which is quite intuitive, as a large sample gives rise to a better estimate of the population mean. If the true population distribution has a smaller mean the sampling distribution will also have a smaller mean. Again, this is very intuitive; if everyone died at exactly the same age the population variance would be zero, and any sample we drew from the population would have a mean exactly the same as the true population mean.

## Hypothesis testing and the central limit theorem

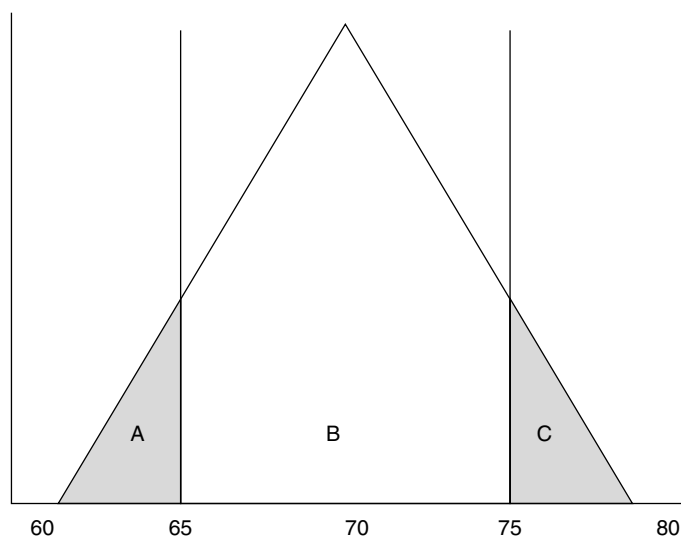
It would seem that the mean fulfils our two criteria for being a good estimate of the population as a whole: it is unbiased and its efficiency increases with the sample size. However, before we can begin to test a hypothesis about this mean, we need some idea of the shape of the whole sampling distribution. Unfortunately, while we have derived a simple expression for the mean and the variance, it is not in general possible to derive the shape of the complete sampling distribution. A hypothesis test proceeds by making an assumption about the truth; we call this the null hypothesis, often referred to as  $H$ . We then set up a specific alternative hypothesis, typically called  $H_a$ . The test consists of calculating the probability that the observed value of the statistic could have arisen purely by chance, assuming that the null hypothesis is true. Suppose that our null hypothesis is that the true population mean for age at death for men is 70,  $H: E(\bar{Y}_m) = 70$ . Having observed a mean of 75.1, we might then test the alternative that the mean is greater than 70. We would do this by calculating the probability that 75.1 could arise purely by chance when the true value of the population mean is 70. With a continuous distribution the probability of any exact point coming up is zero, so strictly what we are calculating is the probability of drawing any value for the mean that is greater than 75.1. We can then compare this probability with a predetermined value, which we call the **significance level** of the test. If the probability is less than the significance level, we reject the null hypothesis in favour of the alternative. In traditional statistics the

significance level is usually set at 1%, 5% or 10%. If we were using a 5% significance level and we found that the probability of observing a mean greater than 75.1 was 0.01, as  $0.01 < 0.05$  we would reject the hypothesis that the true value of the population mean is 70 against the alternative that it is greater than 70.

The alternative hypothesis can typically be specified in two ways, which give rise to either a one-sided test or a two-sided test. The example above is a one-sided test, as the alternative was that the age at death was greater than 70, but we could equally have tested the possibility that the true mean was either greater or less than 70, in which case we would have been conducting a two-sided test. In the case of a two-sided test we would be calculating the probability that a value either greater than 75.1 or less than  $70 - (75.1 - 70) = 64.9$  could occur by chance. Clearly this probability would be higher than in the one-sided test.

Figure 1.1 shows the basic idea of hypothesis testing. It illustrates a possible sampling distribution for the mean life expectancy of men under the null hypothesis that the population mean is 70. It is an unlikely shape, being effectively a triangle, but we will discuss this later; for the moment, simply assume that this is the shape of the distribution. By definition, the complete area under the triangle sums to 1. This simply means that with probability 1 (certainty) the mean will lie between 62 and 78 and that it is centred on 70. We actually observe a mean of 75.1, and if we wish to test the hypothesis that the true mean is 70 against the alternative that it is greater than 70 (a one-sided test) we calculate the probability of observing a value of 75.1 or greater. This is given by area C in the figure. If we wished to conduct the two-sided test, that the alternative is either greater than 75.1 or less than 64.9, we would calculate the sum of areas A and C, which is clearly greater than C. If we adopted a 5% critical value and if  $C < 0.05$ , we would reject the null on a one-sided test. If  $C + A < 0.05$ , we would reject the null at a 5% level on the two-sided test.

As noted above, while we have calculated the mean and the variance of the sampling distribution in the case of the mean, it is not generally possible to calculate the



**Figure 1.1** A possible distribution for life expectancy

## 10 Statistical background and basic data handling

shape of the complete distribution. However, there is a remarkable theorem which does generally allow us to do this as the sample size grows large. This is the central limit theorem.

### Central limit theorem

If a set of data is iid with  $n$  observations,  $(Y_1, Y_2, \dots, Y_n)$ , and with a finite variance then as  $n$  goes to infinity the distribution of  $\bar{Y}$  becomes normal. So as long as  $n$  is reasonably large we can think of the distribution of the mean as being approximately normal.

This is a remarkable result; what it says is that, regardless of the form of the population distribution, the sampling distribution will be normal as long as it is based on a large enough sample. To take an extreme example, suppose we think of a lottery which pays out one winning ticket for every 100 tickets sold. If the prize for a winning ticket is \$100 and the cost of each ticket is \$1, then, on average, we would expect to earn \$1 per ticket bought. But the population distribution would look very strange; 99 out of every 100 tickets would have a return of zero and one ticket would have a return of \$100. If we tried to graph the distribution of returns it would have a huge spike at zero and a small spike at \$100 and no observations anywhere else. But, as long as we draw a reasonably large sample, when we calculate the mean return over the sample it will be centred on \$1 with a normal distribution around 1.

The importance of the central limit theorem is that it allows us to know what the sampling distribution of the mean should look like as long as the mean is based on a reasonably large sample. So we can now replace the arbitrary triangular distribution in Figure 1.1 with a much more reasonable one, the normal distribution.

A final small piece of our statistical framework is the **law of large numbers**. This simply states that if a sample  $(Y_1, Y_2, \dots, Y_n)$  is IID with a finite variance then  $\bar{Y}$  is a consistent estimator of  $\mu$ , the true population mean. This can be formally stated as  $\Pr(|\bar{Y} - \mu| < \varepsilon) \rightarrow 1$  as  $n \rightarrow \infty$ , meaning that the probability that the absolute difference between the mean estimate and the true population mean will be less than a small positive number tends to one as the sample size tends to infinity. This can be proved straightforwardly, since, as we have seen, the variance of the sampling distribution of the mean is inversely proportional to  $n$ ; hence as  $n$  goes to infinity the variance of the sampling distribution goes to zero and the mean is forced to the true population mean.

We can now summarize:  $\bar{Y}$  is an unbiased and consistent estimate of the true population mean  $\mu$ ; it is approximately distributed as a normal distribution with a variance which is inversely proportional to  $n$ ; this may be expressed as  $N(\mu, \sigma^2/n)$ . So if we subtract the population mean from  $\bar{Y}$  and divide by its standard deviation we will create a variable which has a mean of zero and a unit variance. This is called standardizing the variable.

$$\frac{\bar{Y} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1) \quad (1.6)$$

One small problem with this formula, however, is that it involves  $\sigma^2$ . This is the population variance, which is unknown, and we need to derive an estimate of it. We may estimate the population variance by:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (1.7)$$

Here we divide by  $n-1$  because we effectively lose one observation when we estimate the mean. Consider what happens when we have a sample of one. The estimate of the mean would be identical to the one observation, and if we divided by  $n = 1$  we would estimate a variance of zero. By dividing by  $n - 1$  the variance is undefined for a sample of one. Why is  $S^2$  a good estimate of the population variance? The answer is that it is essentially simply another average; hence the law of large numbers applies and it will be a consistent estimate of the true population variance.

Now we are finally in a position to construct a formal hypothesis test. The basic test is known as the student 't' test and is given by:

$$t = \frac{\bar{Y} - \mu}{\sqrt{s^2/n}} \quad (1.8)$$

When the sample is small this will follow a student  $t$ -distribution, which can be looked up in any standard set of statistical tables. In practice, however, once the sample is larger than 30 or 40, the  $t$ -distribution is almost identical to the standard normal distribution, and in econometrics it is common practice simply to use the normal distribution. The value of the normal distribution that implies 0.025 in each tail of the distribution is 1.96. This is the critical value that goes with a two-tailed test at a 5% significance level. So if we want to test the hypothesis that our estimate of the life expectancy of men of 75.1 actually is a random draw from a population with a mean of 70, then the test would be:

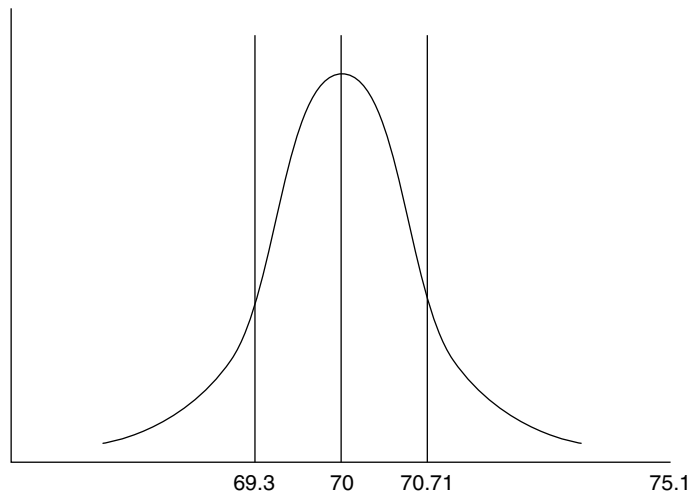
$$t = \frac{75.1 - 70}{\sqrt{s^2/3.87}} = \frac{5.1}{.355} = 14.37$$

This is greater than the 5% significance level of 1.96, and so we would reject the null hypothesis that the true population mean is 70. Equivalently, we could evaluate the proportion of the distribution that is associated with an absolute  $t$ -value greater than 4.1, which would then be the probability value discussed above. Formally, the probability, or  $p$ -value, is given by:

$$p\text{-value} = \Pr_{H_0}(|Y - \mu| > |\bar{Y}^{act} - \mu|) = \Pr_{H_0}(|t| > |t^{act}|)$$

So if the  $t$ -value is exactly 1.96 the  $p$ -value will be 0.05, and when the  $t$ -value is greater than 1.96 the  $p$ -value will be less than 0.05. The two values contain exactly the same information, simply expressed in a different way. The  $p$ -value is useful in other circumstances, however, as it can be calculated for a range of different distributions and can avoid the need to consult statistical tables, as its interpretation is always straightforward.

## 12 Statistical background and basic data handling



**Figure 1.2** A normal distribution for life expectancy around the null

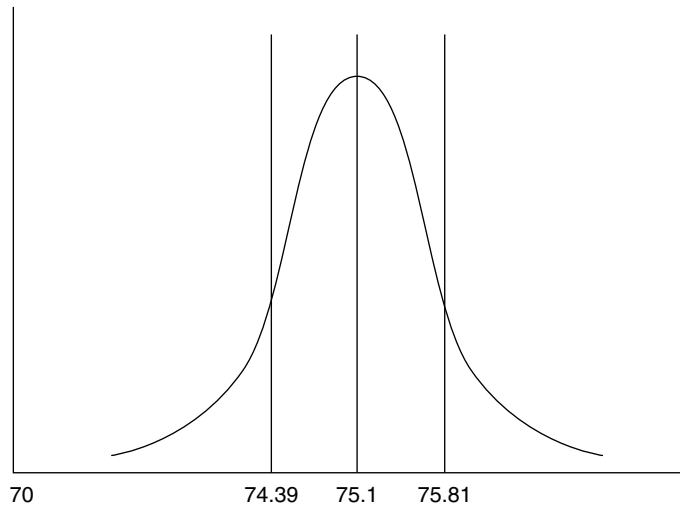
Figure 1.2 illustrates this procedure. It shows an approximately normal distribution centred on the null hypothesis with the two tails of the distribution defined by 69.3 and 70.71. Ninety-five per cent of the area under the distribution lies between these two points. The estimated value of 75.1 lies well outside this central region, and so we can reject the null hypothesis that the true value is 70 and that we observed 75.1 purely by chance. The  $p$ -value is twice the area under the curve which lies beyond 75.1, and clearly this is very small indeed.

One final way to think about the confidence we have in our estimate is to construct a confidence interval around the estimated parameter. We have an estimated mean value of 75.1, but we know there is some uncertainty as to what the true value is. The law of large numbers tells us that this is a consistent estimate of the true value, so with just this one observation our best guess is that the true value is 75.1. The central limit theorem tells us that the distribution around this value is approximately normal, and we know the variance of this distribution. So we can construct an interval around 75.1 that will contain any required amount of the distribution. The convention again is to use a 95% confidence interval, and this may be constructed as follows:

$$CI_{95\%} = \left\{ \bar{Y} + 1.96 \frac{S}{\sqrt{n}}, \bar{Y} - 1.96 \frac{S}{\sqrt{n}} \right\} = \bar{Y} + 0.71, \bar{Y} - 0.71$$

So with 95% confidence we can say that the true mean lies between 74.39 and 75.81. This is shown in Figure 1.3; all that has happened here is that the picture has been moved so that it now centres on the estimated value of 75.1 and 95% of the figure lies inside the confidence interval. Clearly the null value of 70 lies way outside this region, and so we can again conclude that the true value of the mean is highly unlikely to be 70.

The same conclusion arises from calculating the formal  $t$ -test or the  $p$ -value or considering the confidence interval, because they are all simply different ways of expressing the same underlying distribution.



**Figure 1.3** A 95% confidence interval around the estimated mean

## Conclusion

In this chapter we have outlined the basic steps in constructing a theory of estimation and hypothesis testing. We began from the simple idea of random sampling, which gave rise to the proposition that the elements of a sample will have an iid distribution. From this we were able to define a population distribution and to make some inference about this distribution by constructing the mean and then defining the sampling distribution of the mean. By using the law of large numbers and the central limit theorem we were able to define the shape of the sampling distribution, and finally, given this, we were able to outline the basic testing procedure used in classical econometrics.

While at first sight this may appear to relate specifically to a simple estimation procedure, the mean, the same steps may be applied to almost any estimation procedure, as we will see in later chapters of this book. So when we estimate a parameter in a model from a data set we are essentially following the same steps. Any estimation procedure is essentially just taking a sample of data and averaging it together in some way. We have a sampling distribution for the parameter and we can investigate the unbiasedness and consistency of the estimation procedure. We can go on to apply the central limit theorem, which will establish that this sampling distribution will tend to a normal distribution as the sample size grows. Finally, we can use this result to construct hypothesis tests about the parameters that have been estimated and to calculate  $p$ -values and confidence intervals.

# Index

## A

- Adaptive expectations model, 244, 247–249
- Adjusted  $R^2$ , defined, 72–73
- Akaike information criterion, defined, 73
  - ADF tests and, 372, 375, 378, 379
  - ARIMA modelling and, 300
- Almon lag procedure, 241, 243–244
- AR(1), *see* First-order autoregressive process (AR(1)),  $p$ th-order autoregressive process ARCH effects and ARCH models
- ARCH models, general approach, 311–321
  - computer example, 314–318
  - mathematical approach, 318–321
- ARCH tests for heteroskedasticity, 148–150
  - computer example, 149–150
  - steps for, 149
- ARIMA models, defined
  - computer example, 301–305
  - estimating and forecasting with, 300
- Augmented Dickey–Fuller test (ADF)
  - cointegration equation and, 392–393
  - computer example, 377–380
  - general, 372
  - steps for, 374–375
- Autocorrelation, defined, 160
  - ARCH tests and, 148
  - causes of, 160
  - computer examples, 165, 170, 172, 175, 178, 180
  - consequences of ignoring, 162–164
  - detecting, graphical method, 165
  - detecting, tests for, 167–176
  - first-order, 161
  - higher-order, 161
  - lagged dependent variable and, 174–175
  - residual plot and, 165
  - resolving, 177–181
- Autoregressive (AR) models, 289–294
- Auxiliary regressions, 108–109
  - heteroskedasticity and, 128–139
  - LM test approach and, 128–139

## B

- Bar graph, 18
- Base year, defined, 21–22
- Best linear unbiased estimators
  - in multiple regression, 70–72
  - in simple regression, 38–40
- Box–Cox transformation, 193–205
- Breusch–Godfrey test, 171–173
  - computer example, 172
- Breusch–Pagan test, 127–130
  - computer example, 129–130

## C

- CAPM, 87–97
  - advanced EViews programming, 96–97
  - EViews commands, 93–96
  - EViews programming, 90–93
  - steps in the application, 89–90
- Causality, defined, 349
  - computer example, 351–354
  - Granger causality test, 349–351
  - Sims causality test, 351
  - testing for, 347–348
- Central limit theorem, 8–10
- Chi-square test, defined, 80
- Chow test, defined, 231–232
- Cobb–Douglas production function, 23, 51, 76–77
  - double log model and, 192–193
- Cochrane–Orcutt iterative procedure, 178–179
  - vs Hildreth–Lu search procedure, 179
- Coefficients
  - correlation, 109
  - drivers, 446–447
  - dummy, 220–224
  - first-order autocorrelation, 161
  - testing linear restrictions of, 75–77
  - testing the significance of the OLS, 46–47
- Cointegration, defined, 384
  - computer example, 408–415
  - and the ECM, general approach, 387



- Cointegration, defined – *continued*
  - and the ECM, mathematical approach, 388
  - Engle and Granger approach, 396–406
  - Johansen approach, 384–391
  - mathematical approach, 388–392
  - in panel data, 491–496
  - testing for, 392–395, 403–407
- Conditional variance, 98–301
- Confidence interval, 12–14, 45–46
- Consistency of OLS estimators
  - in multiple regression, 70
  - in simple regression, 42–43
- Constant returns to scale, 75–76
- Constant term
  - dummy variables and, 215
- Consumer price index, 21
- Consumption income relationship, 52
- Correlation
  - first-order serial, *see* Autocorrelation
- Correlogram
  - for ARIMA models, 290, 294, 296
  - in EViews, 301–305
- Cross-sectional data
  - defined, 15
  - in EViews, 48
  - in Stata, 512
- D**
- Data, 14–16
  - base period and, 21–22
  - basic handling, 17–21
  - cross-sectional, 15
  - entering in EViews, 506–507
  - entering in Stata, 511–515
  - panel, 16–17
  - time series, 15–16
- Day-of-the-week effect, 233–238
- Diagnostic checking in ARIMA modelling, 298, 300
- Dickey–Fuller tests, 370–372
  - computer example, 377–380
  - performing in EViews, 374
  - performing in Stata, 376
- Differencing, 23
  - generalized differencing approach, 176
  - in spurious regressions, 366
- Distributed lag models, 240–244
  - Almon transformation and, 243–244
  - Koyck transformation and, 241–243
- Double-log model, 192–193
- Dummy variables, defined, 214
  - Chow test, 231–232
  - day-of-the-week effect, 233–238
  - dependent variables, 252–272
  - multiple categories, 226–228
  - seasonal, 225–226
  - slope shift using, 220–221
  - structural change testing, 229–232
  - trap, 223–224
  - trap and exact multicollinearity, 106
- Durbin *h*-test, 173–175
- Durbin–Watson test for serial correlation, 167–170
- Dynamic models
  - in panel data, 476
  - in time series data, 240–249
- E**
- Efficiency of the OLS coefficients, 40–41
- E-GARCH model, 329–330
  - computer example, 330–331
- Equation
  - reduced form, 253–254
  - simultaneous, 252–261
- Error(s)
  - correction model, *see* Error-correction model, measurement
  - normality of, 198
  - specification, 185
- Error-correction model, defined, 387
  - cointegration and, 386
  - computer example, 408–415
- Estimation
  - with ARIMA models, 300
  - OLS method, 32–35
  - of simultaneous equation models, 256–257
  - time-varying coefficient method, 444–446
  - using dummy variables, 219–226
- Estimator(s)
  - best linear unbiased, 38–40, 69–72
  - consistency, 42–43, 70
  - efficiency, 40
  - unbiasedness, 39–40, 70
- EViews software, basics, 504–508
- Exact identification, 254
- Exact multicollinearity, 104–106
- Excel and OLS estimation, 53–55
- F**
- F-form of the likelihood ratio test, 77
- Finite prediction error, 73
- First-order autocorrelation coefficient, 161
- First-order autoregressive process (AR(1))
  - defined, 289
  - miscellaneous derivations with, 289–290
- Fitted straight line, 31–33
- Fixed effects model, 462–463
- F-test for overall significance, 78–79
- Function(s)
  - autocorrelation, 293–294
  - Cobb–Douglas production, 23, 51, 76–77
  - impulse response, 434
- Functional forms, 189–193
  - double-log, 192–193
  - including interaction terms, 191

- linear-log, 189–190
  - logarithmic, 184
  - log-linear, 192
  - polynomial, 190–191
  - reciprocal, 190
- G**
- GARCH models, defined, 415
    - advanced EViews programming, 334–338
    - application, 338–342
    - computer examples, 314–338
    - in Stata, 332–334
  - Generalized least squares, 150–151
  - General to specific approach, 207–208
  - Glesjer test, 130–132
    - computer example in EViews, 131
    - computer example in Stata, 131–132
  - Goldfeld–Quandt test, 135–137
  - Goodness of fit, defined, 43–44
    - in limited dependent variable model, 271–272
    - measurement of, 43–44
  - Granger causality test, 349–351
    - application, 351–354
    - computer example, 356–357
    - steps for, 349–351
- H**
- Hendry/LSE approach, 207–208
  - Heteroskedasticity
    - computer examples, 139–151
    - consequences of, 122–124
    - defined, 119
    - illustration of, 119–120
    - resolving, 151–155
    - testing for, 126–137
  - Heteroskedasticity consistent estimation method, 152
  - Hildreth–Lu search procedure, 178–179
    - vs Cochrane–Orcutt research procedure, 178
  - Histogram, 18–20
    - and normality tests, 198–199
  - Homoskedasticity, 36, 120, 310
  - Hypothesis testing
    - and the central limit theorem, 8–10
    - confidence intervals and, 45–46
    - p*-value approach, 47
    - rule of thumb, 46–47
    - steps to, 46
    - testing individual coefficients, 75
    - testing linear restrictions, 75
    - testing the significance of OLS coefficients, 46
- I**
- Identification problem, defined, 253–254, 299–300
    - cointegrated systems and, 422–427
    - computer example, 258–261, 426–427
    - conditions for, 254–255
    - example of, 255–256
    - order condition and, 421
    - rank condition and, 422
  - Im, Pesaran and Shin panel unit-root test, 488
  - Imperfect multicollinearity, 106–107
  - Impulse response functions, 434
  - Indirect least squares, 256
  - Instrumental variables, 444, 474
  - Integrated of order *d*, 298, 366
  - Integrated of order one, 365–366
  - Integration
    - Dickey–Fuller tests of, 355–357
    - Phillips–Perron tests of, 357–361
    - testing for the order of, 355
  - Intercept term
    - dummy variables and, 214–216, 218–219
  - Invertibility in MA models, 295–296
- J**
- January effect application, 228–229
  - Joint significance, 78–79
- K**
- Kao panel cointegration test, 492–493
  - Keynesian model, 53–57
  - Koyck lag model, 241–243
- L**
- Lagged dependent variables
    - adaptive expectations model and, 247–249
    - partial adjustment model and, 244–245
    - serial correlation and, 173
  - Lagrange multiplier (LM) test, 81–82
    - in EViews, 82
    - for heteroskedasticity, 127–135
    - for serial correlation, 170–171
    - for testing linear restrictions, 75–76
  - Larsson *et al.* panel cointegration test, 495–496
  - Least squares method, defined, 32
    - derivation of solutions for the multiple model, 64–66
    - derivation of solutions for the simple model, 32–35
    - derivation of solutions with matrix algebra, 66–67
  - Levin and Lin panel unit-root test, 487–488
  - Likelihood ratio test, 77
    - F*-form of, 77–78
  - Limited dependent variable regression, 265–282
  - Linear-log model, 189–190
  - Linear probability model, 266–267
    - problems with, 267–268
  - Ljung–Box test statistic, 300

- Logit model, 269  
 computer example, 278–281  
 general approach, 269–270  
 goodness of fit and, 271–272  
 interpretation of estimates, 270–271  
 mathematical approach, 272–273  
 in Stata, 278–282
- Log-linear model, 192
- Log-log model, 192–193
- Long-run behaviour, error-correction model and, 386
- M**
- Maddala and Wu panel unit-root test, 489
- Marginal effect  
 of functional forms, 189–193  
 interpretation of, 189, 193
- Marginal propensity to consume, 52, 191, 217, 249
- McCoskey and Kao panel cointegration test, 493
- Misspecification, 31, 37, 77, 185–208  
 RESET test, 199  
 tests for, 197–201
- Model(s)  
 adaptive expectations, 247–249  
 ARCH, 310–342  
 ARIMA, 287–307  
 autoregressive, 244–249, 289–294  
 distributed lag, 240–244  
 double-log, 192–193  
 dynamic, 240  
 E-GARCH, 329–330  
 error correction, 386–390  
 fixed effects, 462–463  
 GARCH, 321–323  
 GARCH-M, 324  
 Hendry/LSE approach, 207–208  
 Keynesian, 53  
 Koyck, 241–243  
 with lagged dependent variables, 174, 244–247  
 linear-log, 189–190  
 linear panel data, 434  
 logit, 269–273  
 log-linear, 192  
 log-log, 192–193  
 multinomial logit, 276  
 multinomial probit, 277  
 partial adjustment, 244–245  
 polynomial, 190–191  
 probit, 274–275  
 random effect, 463–464  
 reciprocal, 190  
 solving, 431–440  
 TGARCH, 328–329  
 Tobit, 278
- Modelling  
 average economic regression, 206–207  
 general to specific, 206–207  
 Hendry/LSE approach, 207–208  
 simple to general, 198  
 traditional view, 206–207
- Moving average models, 294–296
- Multicollinearity  
 computer examples, 110–115  
 consequences of, 105–109  
 defined, 104  
 detecting, 109–110  
 exact, 106–107  
 imperfect, 106–107  
 perfect, 104–106
- Multinomial logit model, 276
- Multinomial probit model, 277
- Multiple regression computer examples, 82–84  
 defined, 64  
 in EViews, 74  
 goodness of fit and, 72  
 hypothesis testing and, 75–82  
 in Stata, 74
- N**
- Non-nested models, tests for, 201–203
- Nonstationarity, defined, 363  
 and spurious regression, 366  
 and unit roots, 370–374
- O**
- OLS, *see* Ordinary least squares
- Omitted variables, 84–85  
 bias, 186  
 LM tests for, 84  
 and the plug-in solution, 187–188
- Order condition, 254, 421
- Ordinary least squares, defined, 32  
 GLS procedure and, 151  
 heteroskedasticity and consequences, 122–126  
 serial correlation and consequences, 162–165
- Overidentification, defined, 254
- Overparametrized model, 298
- P**
- Panel data, defined, 16–17  
 advantages of, 460  
 cointegration and, 491–496  
 common constant, 461  
 different methods of estimation, 461  
 estimating in EViews, 469  
 estimating in Stata, 474  
 fixed effects, 462–463  
 Hausman test, 464–465  
 heterogeneous, 452  
 inserting in EViews, 465–468  
 inserting in Stata, 471–473  
 linear model, 461  
 random effects, 463–464

unit root tests, 486–491

Park test, 144

Partial adjustment model, 244–245  
computer example, 245–247

Partial autocorrelation function, 294, 300, 307

Pedroni panel cointegration test, 494–495

Perfect multicollinearity, 106–107

Pooled mean group estimator, 480–482

Pooling assumption, 460, 478

Probit model, 274–275  
computer example, 278–282  
general approach, 274–275  
mathematical approach, 275–276

Production function, 51

$p$ th-order autoregressive process (AR( $p$ )), 291–292

$p$ th-order serial correlation, 161–162

$p$ -value approach, 47

**Q**

Qualitative information, defined, 214  
dummy variables and, 214–218  
with multiple categories, 222–224, 226  
slope term and, 216–217

**R**

$R^2$ , 44  
problems with, 44–45

$R^2$  adjusted for degrees of freedom, 72

Random effects model, 463–464

Rank condition, 254–255, 422

Reciprocal functional form, 190

Redundant variables, 79–84, 186–187

Regression  
Dickey–Fuller, 370–372  
multiple, 64–96  
simple, 29–58  
spurious, 363  
sum of squares, 43–44

Regressions specification error test (RESET), 199

Residual, defined, 21  
test of normality, 197

Robust inference, 125, 165

**S**

Scatter plots, 18, 31  
detecting autocorrelation, 165–167  
detecting heteroskedasticity, 126–128  
in EViews, 18  
simple regression, 31  
spurious regressions, 367–368  
in Stata, 19

Seasonal dummies, 225–226  
application, 228–229

Serial correlation, *see* Autocorrelation

Significantly different from zero, 5, 47

Simple linear regression model, 30

computer examples, 53–57  
interpretation of coefficients, 30

Simple to general modelling, 206–207

Simultaneous equation model, 252  
consequences of ignoring simultaneity, 253  
estimation of, 256–257  
identification problem, 253  
structure of reduced forms, 253–254

Specification error, defined, 185, 199

Spurious correlation, 363–365

Spurious regression, 363–365

Stata software, basics, 509–517

Stationarity, defined, 289

Stationary time series, 289

Structural breaks, 17, 18, 299

Structural change, 230

**T**

Test(ing)  
approach, 392–395  
For ARCH effects, 313  
for autocorrelation, 165–177  
for causality, 349–351  
for cointegration, Engle–Granger approach, 392–395  
for cointegration, Johansen approach, 395–403  
of goodness of fit, 73, 271  
for heteroskedasticity, 126–135  
hypothesis, 8–10, 45–47  
individual coefficients, 46, 75  
for the joint significance of the  $X_s$ , 78–79  
linear restrictions, 75–77  
for misspecification, 197–201  
for structural change, 232–233

$t$ -test, 47, 80–81

Time series data, 15–16

Time series models, *see* ARIMA models

Time-varying coefficient models, 443–456  
choosing coefficient drivers, 447–451  
coefficient drivers, 446–447  
estimation, 444–446

Total sum of squares, 43–44

$t$ -test, 47, 80–81

TVC models, *see* Time-varying coefficient models

**U**

Unbiasedness of OLS coefficients  
multiple regression, 70  
simple regression, 39–40

Unit roots, defined, 363–370  
Dickey–Fuller test and, 370–372  
in EViews, 374  
panel data and, 460–463  
Phillips–Perron test, 372–374  
in Stata, 376

536 *Index***V**

VAR models, *see* Vector autoregressive models

Variable(s)

- dummy, 214
- instrumental, 444, 448, 477
- lagged dependent, 174–176
- omitted, 185–186
- qualitative, 214
- redundant, 186–187

Variation

- explained, 44
- total, 44
- unexplained, 44

Vector autoregressive (VAR) models, 347–349

in EViews, 354–357

pros and cons, 348–349

in Stata, 357

**W**

Wald test, 77

computer example, 82–84

performing in EViews, 80

Weighted least squares, 150–152

White's heteroskedasticity consisted

estimation, 152

White's test, 137–139, 146

computer example, 146

in EViews, 138

in Stata, 139