

# Contents

<b>INTRODUCTION</b>	<b>1</b>
Why study econometrics?	1
Aim of this text	2
Mathematics and statistics prerequisites for studying econometrics	2
Additional resources	3
Econometrics software	4
<b>REVIEW: RANDOM VARIABLES, SAMPLING, ESTIMATION, AND INFERENCE</b>	<b>5</b>
<b>R.1 The need for a solid understanding of statistical theory</b>	<b>5</b>
<b>R.2 Discrete random variables and expectations</b>	<b>7</b>
Discrete random variables	7
Expected values of discrete random variables	8
Expected values of functions of discrete random variables	9
Expected value rules	10
Population variance of a discrete random variable	11
Fixed and random components of a random variable	12
<b>R.3 Continuous random variables</b>	<b>14</b>
Probability density	14
<b>R.4 Population covariance, covariance and variance rules, and correlation</b>	<b>19</b>
Covariance	19
Independence of random variables	19
Covariance rules	20
Variance rules	21
Correlation	22
<b>R.5 Samples, the double structure of a sampled random variable, and estimators</b>	<b>23</b>
Sampling	23
Estimators	24



<b>R.6</b>	<b>Unbiasedness and efficiency</b>	27
	Unbiasedness	27
	Efficiency	28
	Conflicts between unbiasedness and minimum variance	31
<b>R.7</b>	<b>Estimators of variance, covariance, and correlation</b>	33
<b>R.8</b>	<b>The normal distribution</b>	35
<b>R.9</b>	<b>Hypothesis testing</b>	37
	Formulation of a null hypothesis and development of its implications	37
	Compatibility, freakiness, and the significance level	38
<b>R.10</b>	<b>Type II error and the power of a test</b>	43
<b>R.11</b>	<b><i>t</i> tests</b>	49
	The reject/fail-to-reject terminology	52
<b>R.12</b>	<b>Confidence intervals</b>	53
<b>R.13</b>	<b>One-sided tests</b>	58
	$H_0: \mu = \mu_0, H_1: \mu = \mu_1$	58
	Generalizing from $H_0: \mu = \mu_0, H_1: \mu = \mu_1$ to $H_0: \mu = \mu_0, H_1: \mu > \mu_0$	64
	$H_0: \mu = \mu_0, H_1: \mu < \mu_0$	64
	One-sided <i>t</i> tests	65
	Important special case: $H_0: \mu = 0$	65
	Anomalous results	66
	Justification of the use of a one-sided test	66
<b>R.14</b>	<b>Probability limits and consistency</b>	68
	Probability limits	68
	Consistency	70
	Why is consistency of interest?	71
	Simulations	73
<b>R.15</b>	<b>Convergence in distribution and central limit theorems</b>	76
	Limiting distributions	77
	<b>Key terms</b>	81
<b>Appendix R.1</b>	<b>Unbiased estimators of the population covariance and variance</b>	81
<b>Appendix R.2</b>	<b>Density functions of transformed random variables</b>	83
<b>1</b>	<b>SIMPLE REGRESSION ANALYSIS</b>	85
<b>1.1</b>	<b>The simple linear model</b>	85
<b>1.2</b>	<b>Least squares regression with one explanatory variable</b>	87
<b>1.3</b>	<b>Derivation of the regression coefficients</b>	89
	Least squares regression with one explanatory variable: the general case	92
	Two decompositions of the dependent variable	95
	Regression model without an intercept	96
<b>1.4</b>	<b>Interpretation of a regression equation</b>	98
	Changes in the units of measurement	100



<b>1.5 Two important results relating to OLS regressions</b>	105
The mean value of the residuals is zero	106
The sample correlation between the observations on $X$ and the residuals is zero	106
<b>1.6 Goodness of fit: <math>R^2</math></b>	107
Example of how $R^2$ is calculated	109
Alternative interpretation of $R^2$	110
<b>Key terms</b>	111
<b>2 PROPERTIES OF THE REGRESSION COEFFICIENTS AND HYPOTHESIS TESTING</b>	113
<b>2.1 Types of data and regression model</b>	113
<b>2.2 Assumptions for regression models with nonstochastic regressors</b>	114
<b>2.3 The random components and unbiasedness of the OLS regression coefficients</b>	118
The random components of the OLS regression coefficients	118
The unbiasedness of the OLS regression coefficients	122
Normal distribution of the regression coefficients	124
<b>2.4 A Monte Carlo experiment</b>	126
<b>2.5 Precision of the regression coefficients</b>	130
Variances of the regression coefficients	130
Standard errors of the regression coefficients	133
The Gauss–Markov theorem	137
<b>2.6 Testing hypotheses relating to the regression coefficients</b>	139
0.1 percent tests	144
<p><math>p</math> values</p>	144
One-sided tests	145
Confidence intervals	147
<b>2.7 The <math>F</math> test of goodness of fit</b>	150
Relationship between the $F$ test of goodness of fit and the $t$ test on the slope coefficient in simple regression analysis	152
<b>Key terms</b>	153
<b>Appendix 2.1 The Gauss–Markov theorem</b>	154
<b>3 MULTIPLE REGRESSION ANALYSIS</b>	156
<b>3.1 Illustration: a model with two explanatory variables</b>	156
<b>3.2 Derivation of the multiple regression coefficients</b>	158
The general model	160
Interpretation of the multiple regression coefficients	161
<b>3.3 Properties of the multiple regression coefficients</b>	164
Unbiasedness	165
Efficiency	166



Precision of the multiple regression coefficients	166
<i>t</i> tests and confidence intervals	169
<b>3.4 Multicollinearity</b>	171
Multicollinearity in models with more than two explanatory variables	174
Examples of multicollinearity	174
What can you do about multicollinearity?	175
<b>3.5 Goodness of fit: <math>R^2</math></b>	180
<i>F</i> tests	182
Further analysis of variance	184
Relationship between <i>F</i> statistic and <i>t</i> statistic	186
<b>3.6 Prediction</b>	189
Properties of least squares predictors	191
<b>Key terms</b>	195
<b>4 NONLINEAR MODELS AND TRANSFORMATIONS OF VARIABLES</b>	197
<b>4.1 Linearity and nonlinearity</b>	197
<b>4.2 Logarithmic transformations</b>	201
Logarithmic models	201
Semilogarithmic models	205
The disturbance term	208
Comparing linear and logarithmic specifications	209
<b>4.3 Models with quadratic and interactive variables</b>	214
Quadratic variables	215
Higher-order polynomials	217
Interactive explanatory variables	218
Ramsey's RESET test of functional misspecification	222
<b>4.4 Nonlinear regression</b>	225
<b>Key terms</b>	228
<b>5 DUMMY VARIABLES</b>	230
<b>5.1 Illustration of the use of a dummy variable</b>	230
Standard errors and hypothesis testing	234
<b>5.2 Extension to more than two categories and to multiple sets of dummy variables</b>	237
Joint explanatory power of a group of dummy variables	240
Change of reference category	240
The dummy variable trap	242
Multiple sets of dummy variables	244
<b>5.3 Slope dummy variables</b>	250
Joint explanatory power of the intercept and slope dummy variables	252



<b>5.4 The Chow test</b>	255
Relationship between the Chow test and the $F$ test of the explanatory power of a set of dummy variables	258
<b>Key terms</b>	259
<b>6 SPECIFICATION OF REGRESSION VARIABLES</b>	261
<b>6.1 Model specification</b>	261
<b>6.2 The effect of omitting a variable that ought to be included</b>	262
The problem of bias	262
Invalidation of the statistical tests	265
$R^2$ in the presence of omitted variable bias	267
<b>6.3 The effect of including a variable that ought not to be included</b>	272
<b>6.4 Proxy variables</b>	276
Unintentional proxies	278
<b>6.5 Testing a linear restriction</b>	280
$F$ test of a linear restriction	281
The reparameterization of a regression model	282
$t$ test of a linear restriction	284
Multiple restrictions	285
Zero restrictions	285
<b>Key terms</b>	286
<b>7 HETEROSKEDASTICITY</b>	290
<b>7.1 Heteroskedasticity and its implications</b>	290
Possible causes of heteroskedasticity	293
<b>7.2 Detection of heteroskedasticity</b>	295
The Goldfeld–Quandt test	296
The White test	297
<b>7.3 Remedies for heteroskedasticity</b>	299
Weighted least squares	299
Mathematical misspecification	303
Robust standard errors	305
How serious are the consequences of heteroskedasticity?	306
<b>Key terms</b>	308
<b>8 STOCHASTIC REGRESSORS AND MEASUREMENT ERRORS</b>	311
<b>8.1 Assumptions for models with stochastic regressors</b>	311
<b>8.2 Finite sample properties of the OLS regression estimators</b>	313
Unbiasedness of the OLS regression estimators	313
Precision and efficiency	314



<b>8.3 Asymptotic properties of the OLS regression estimators</b>	<b>315</b>
Consistency	316
Asymptotic normality of the OLS regression estimators	317
<b>8.4 The consequences of measurement errors</b>	<b>317</b>
Measurement errors in the explanatory variable(s)	318
Measurement errors in the dependent variable	320
Imperfect proxy variables	322
Example: Friedman's permanent income hypothesis	322
<b>8.5 Instrumental variables</b>	<b>327</b>
Asymptotic distribution of the IV estimator	330
Multiple instruments	337
The Durbin–Wu–Hausman specification test	338
<b>Key terms</b>	<b>340</b>
 <b>9 SIMULTANEOUS EQUATIONS ESTIMATION</b>	 <b>343</b>
<b>9.1 Simultaneous equations models: structural and reduced form equations</b>	<b>343</b>
<b>9.2 Simultaneous equations bias</b>	<b>345</b>
A Monte Carlo experiment	348
<b>9.3 Instrumental variables estimation</b>	<b>351</b>
Underidentification	354
Exact identification	355
Overidentification	356
Two-stage least squares	357
The order condition for identification	358
Unobserved heterogeneity	360
Durbin–Wu–Hausman test	361
<b>Key terms</b>	<b>362</b>
 <b>10 BINARY CHOICE AND LIMITED DEPENDENT VARIABLE MODELS, AND MAXIMUM LIKELIHOOD ESTIMATION</b>	 <b>367</b>
<b>10.1 The linear probability model</b>	<b>367</b>
<b>10.2 Logit analysis</b>	<b>372</b>
Generalization to more than one explanatory variable	374
Goodness of fit and statistical tests	375
<b>10.3 Probit analysis</b>	<b>378</b>
<b>10.4 Censored regressions: tobit analysis</b>	<b>381</b>
<b>10.5 Sample selection bias</b>	<b>386</b>
<b>10.6 An introduction to maximum likelihood estimation</b>	<b>391</b>
Generalization to a sample of $n$ observations	395
Generalization to the case where $\sigma$ is unknown	395



Application to the simple regression model	398
Goodness of fit and statistical tests	400
Key terms	401
<b>Appendix 10.1 Comparing linear and logarithmic specifications</b>	402
<b>11 MODELS USING TIME SERIES DATA</b>	405
11.1 Assumptions for regressions with time series data	405
11.2 Static models	408
11.3 Models with lagged explanatory variables	413
Estimating long-run effects	415
11.4 Models with a lagged dependent variable	416
The partial adjustment model	419
The error correction model	421
The adaptive expectations model	421
More general autoregressive models	424
11.5 Assumption C.7 and the properties of estimators in autoregressive models	427
Consistency	429
Limiting distributions	431
$t$ tests in an autoregressive model	432
11.6 Simultaneous equations models	435
11.7 Alternative dynamic representations of time series processes	438
Time series analysis	439
Vector autoregressions	441
Key terms	443
<b>12 AUTOCORRELATION</b>	445
12.1 Definition and consequences of autocorrelation	445
Consequences of autocorrelation	447
Autocorrelation with a lagged dependent variable	449
12.2 Detection of autocorrelation	449
The Breusch–Godfrey test	450
The Durbin–Watson test	451
12.3 Fitting a model subject to AR(1) autocorrelation	455
Issues	456
Inference	457
The common factor test	460
12.4 Apparent autocorrelation	467
12.5 Model specification: specific-to-general versus general-to-specific	472
Comparison of alternative models	473
The general-to-specific approach to model specification	475



Key terms	476
Appendix 12.1 Demonstration that the Durbin–Watson $d$ statistic approximates $2 - 2\rho$ in large samples	477
<b>13 INTRODUCTION TO NONSTATIONARY TIME SERIES</b>	478
13.1 Stationarity and nonstationarity	478
Stationary time series	478
Nonstationary time series	484
Deterministic trend	487
Difference-stationarity and trend-stationarity	488
13.2 Spurious regressions	490
Spurious regressions with variables possessing deterministic trends	491
Spurious regressions with variables that are random walks	491
13.3 Graphical techniques for detecting nonstationarity	501
13.4 Tests of nonstationarity: the augmented Dickey–Fuller $t$ test	506
Untrended process	507
Trended process	510
13.5 Tests of nonstationarity: other tests	513
The Dickey–Fuller test using the scaled estimator of the slope coefficient	513
The Dickey–Fuller $F$ test	516
Power of the tests	516
Further tests	518
Tests of deterministic trends	518
Further complications	518
13.6 Cointegration	519
13.7 Fitting models with nonstationary time series	524
Detrending	524
Differencing	525
Error correction models	526
Key terms	528
<b>14 INTRODUCTION TO PANEL DATA MODELS</b>	529
14.1 Reasons for interest in panel data sets	529
14.2 Fixed effects regressions	531
Within-groups fixed effects	533
First differences fixed effects	534
Least squares dummy variable fixed effects	535
14.3 Random effects regressions	537
Assessing the appropriateness of fixed effects and random effects estimation	539
Random effects or OLS?	541
A note on the random effects and fixed effects terminology	541



14.4 Differences in differences	544
Key terms	546
<i>APPENDIX A: Statistical tables</i>	547
<i>APPENDIX B: Data sets</i>	565
<i>Bibliography</i>	577
<i>Author index</i>	581
<i>Subject index</i>	582

## Why study econometrics?

Econometrics is the term used to describe the application of statistical methods to the quantification and critical assessment of hypothetical relationships using data. The term 'econometrics' suggests that the methods relate only to economic analysis. In fact, applications will be found far more broadly, in virtually all the social sciences and elsewhere. It is true that econometrics has been responsible for much of the development of econometrics, but other disciplines have also made substantial contributions. Indeed, regression analysis, the core technique, appears initially to have been developed in applications to astronomy by Legendre and Gauss in the first few years of the nineteenth century.

It is with the aid of econometrics that we discriminate between competing theories and put numerical clothing onto the successful ones. For economists, econometric analysis may be motivated by a simple desire to improve our understanding of how the economy works, at either the microeconomic or the 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 be those that have the greatest impact.

It is now generally recognized that nearly all professional economists, not just those actually working with data, should have a basic understanding of econometrics. There used to be a view that microeconomics and macroeconomics comprised the core training of an economist, and that econometrics was an optional extra to be pursued by those with a flair for numbers and an inclination to get their hands dirty with data. In particular, much of early macroeconomic theory was in reality no more than conjecture propounded by (over-)confident theorists who thought that the job of quantifying their theories could safely be left to others with lesser vision and a greater willingness to apply themselves to empirical detail.

That view is long gone. Microeconomic and macroeconomic theories are generally considered to be of little interest if they are not supported by robust empirical analysis. As a consequence of the recognition of its importance, an understanding