

Introduction to **Econometrics**

FIFTH EDITION

Christopher Dougherty

London School of Economics and Political Science

OXFORD
UNIVERSITY PRESS

Contents

INTRODUCTION

Why study econometrics?

Aim of this text

Mathematics and statistics prerequisites for studying econometrics

Additional resources

Econometrics software

REVIEW: RANDOM VARIABLES, SAMPLING, ESTIMATION, AND INFERENCE

R.1 The need for a solid understanding of statistical theory

R.2 Discrete random variables and expectations

Discrete random variables

Expected values of discrete random variables

Expected values of functions of discrete random variables

Expected value rules

Population variance of a discrete random variable

Fixed and random components of a random variable

R.3 Continuous random variables

Probability density

R.4 Population covariance, covariance and variance rules, and correlation

Covariance

Independence of random variables

Covariance rules

Variance rules

Correlation

R.5 Samples, the double structure of a sampled random variable, and estimators

Sampling

Estimators

R.6	Unbiasedness and efficiency	27
	Unbiasedness	27
	Efficiency	28
	Conflicts between unbiasedness and minimum variance	31
R.7	Estimators of variance, covariance, and correlation	33
R.8	The normal distribution	35
R.9	Hypothesis testing	37
	Formulation of a null hypothesis and development of its implications	37
	Compatibility, freakiness, and the significance level	38
R.10	Type II error and the power of a test	43
R.11	t tests	49
	The reject/fail-to-reject terminology	52
R.12	Confidence intervals	53
R.13	One-sided tests	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 t tests	65
	Important special case: $H_0: \mu = 0$	65
	Anomalous results	66
	Justification of the use of a one-sided test	66
R.14	Probability limits and consistency	68
	Probability limits	68
	Consistency	70
	Why is consistency of interest?	71
	Simulations	73
R.15	Convergence in distribution and central limit theorems	76
	Limiting distributions	77
	Key terms	81
Appendix R.1	Unbiased estimators of the population covariance and variance	81
Appendix R.2	Density functions of transformed random variables	83
1	SIMPLE REGRESSION ANALYSIS	85
1.1	The simple linear model	85
1.2	Least squares regression with one explanatory variable	87
1.3	Derivation of the regression coefficients	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
1.4	Interpretation of a regression equation	98
	Changes in the units of measurement	100

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

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

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

8.3 Asymptotic properties of the OLS regression estimators	315
Consistency	316
Asymptotic normality of the OLS regression estimators	317
8.4 The consequences of measurement errors	317
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
8.5 Instrumental variables	327
Asymptotic distribution of the IV estimator	330
Multiple instruments	337
The Durbin–Wu–Hausman specification test	338
Key terms	340
 9 SIMULTANEOUS EQUATIONS ESTIMATION	 343
9.1 Simultaneous equations models: structural and reduced form equations	343
9.2 Simultaneous equations bias	345
A Monte Carlo experiment	348
9.3 Instrumental variables estimation	351
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
Key terms	362
 10 BINARY CHOICE AND LIMITED DEPENDENT VARIABLE MODELS, AND MAXIMUM LIKELIHOOD ESTIMATION	 367
10.1 The linear probability model	367
10.2 Logit analysis	372
Generalization to more than one explanatory variable	374
Goodness of fit and statistical tests	375
10.3 Probit analysis	378
10.4 Censored regressions: tobit analysis	381
10.5 Sample selection bias	386
10.6 An introduction to maximum likelihood estimation	391
Generalization to a sample of n observations	395
Generalization to the case where σ is unknown	395

Application to the simple regression model	398
Goodness of fit and statistical tests	400
Key terms	401
Appendix 10.1 Comparing linear and logarithmic specifications	402
 11 MODELS USING TIME SERIES DATA	 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
<i>t</i> 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
 12 AUTOCORRELATION	 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
13 INTRODUCTION TO NONSTATIONARY TIME SERIES	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
14 INTRODUCTION TO PANEL DATA MODELS	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