

Contents

<i>List of figures</i>	xviii
<i>List of tables</i>	xx
<i>Notes on authors</i>	xxiii
<i>Preface</i>	xxv
<i>How to use this book</i>	xxvii

PART 1

Linking models to data for development	1
1 An introduction to empirical development economics	3
1.1 <i>The objective of the book</i>	3
1.2 <i>Models and data: the Harris–Todaro model</i>	4
1.3 <i>Production functions and functional form</i>	6
1.3.1 The Cobb–Douglas production function	6
1.3.2 The constant elasticity of substitution (CES) functional form	10
1.4 <i>A model with human capital</i>	11
1.5 <i>Data and models</i>	13
1.5.1 The macro GDP data	13
1.5.2 Interpreting the data	14
<i>References</i>	14
<i>Exercise</i>	15

SECTION I

Cross-section data and the determinants of incomes	17
2 The linear regression model and the OLS estimator	19
2.1 <i>Introduction: models and causality</i>	19
2.2 <i>The linear regression model and the OLS estimators</i>	20
2.2.1 The linear regression model as a population model	20
2.2.2 The zero conditional mean assumption	21
2.2.3 The OLS estimator	24
2.3 <i>The Mincerian earnings function for the South African data</i>	26

2.4	<i>Properties of the OLS estimators</i>	28	
2.4.1	The assumptions for OLS to be unbiased	28	
2.4.2	The assumptions for OLS to be minimum variance	29	
2.5	<i>Identifying the causal effect of education</i>	31	
	<i>References</i>	31	
	<i>Exercise</i>	32	
3	Using and extending the simple regression model		33
3.1	<i>Introduction</i>	33	
3.2	<i>Dummy explanatory variables and the return to education</i>	33	
3.3	<i>Multiple regression</i>	36	
3.3.1	Earnings and production functions	36	
3.3.2	The OLS estimators for multiple regression	37	
3.3.3	Omitted variables and the bias they may cause	39	
3.4	<i>Interpreting multiple regressions</i>	40	
3.4.1	How much does investing in education increase earnings? Some micro evidence	40	
3.4.2	How much does investing in education increase productivity? Some macro evidence	43	
	<i>References</i>	45	
	<i>Exercise</i>	45	
4	The distribution of the OLS estimators and hypothesis testing		47
4.1	<i>Introduction</i>	47	
4.2	<i>The distribution of the OLS estimators</i>	47	
4.2.1	The normality assumption	47	
4.2.2	Why normality?	48	
4.3	<i>Testing hypotheses about a single population parameter</i>	49	
4.3.1	The <i>t</i> distribution	49	
4.3.2	The <i>t</i> -test	51	
4.3.3	Confidence intervals	53	
4.4	<i>Testing for the overall significance of a regression</i>	55	
4.5	<i>Testing for heteroskedasticity</i>	57	
4.6	<i>Large sample properties of OLS</i>	58	
4.6.1	Consistency	58	
4.6.2	Asymptotic normality	60	
	<i>References</i>	60	
	<i>Exercise</i>	61	
5	The determinants of earnings and productivity		62
5.1	<i>Introduction</i>	62	
5.2	<i>Testing the normality assumption</i>	62	
5.3	<i>The earnings function</i>	65	
5.3.1	Bringing the tests together	65	
5.3.2	Robust and clustered standard errors	65	

5.4	<i>The production function</i>	67
5.4.1	Testing the production function	67
5.4.2	Extending the production function	67
5.5	<i>Interpreting our earnings and production functions</i>	72
5.5.1	Can education be given a causal interpretation?	72
5.5.2	How much does education raise labour productivity?	73
	<i>References</i>	74
	<i>Exercise</i>	74

SECTION II

	Time-series data, growth and development	75
6	Modelling growth with time-series data	77
6.1	<i>Introduction: modelling growth</i>	77
6.2	<i>An introduction to the Solow model</i>	78
6.3	<i>A Solow model for Argentina</i>	80
6.4	<i>OLS estimates under the classical assumptions with time-series data</i>	81
6.4.1	Assumptions for OLS to be unbiased	81
6.4.2	The variance of the OLS estimators	83
6.4.3	Testing for autocorrelation	85
6.5	<i>Static and dynamic time-series models</i>	85
6.6	<i>Assumptions to ensure the OLS estimators are consistent</i>	87
6.7	<i>Spurious regression with nonstationary time-series data</i>	89
6.8	<i>A brief summary</i>	91
	<i>References</i>	92
	<i>Exercise</i>	93
7	The implications of variables having a unit root	95
7.1	<i>Introduction and motivation</i>	95
7.2	<i>Testing for a unit root and the order of integration</i>	96
7.3	<i>Cointegration</i>	100
7.4	<i>How are growth and inflation related in Argentina?</i>	101
7.5	<i>The error-correction model</i>	104
7.6	<i>Causality in time-series models</i>	105
7.7	<i>Cross-section and time-series data</i>	106
	<i>References</i>	107
	<i>Exercise</i>	107
8	Exogenous and endogenous growth	109
8.1	<i>The Solow model and the history of development</i>	109
8.2	<i>Long-term growth and structural change</i>	109
8.3	<i>The Solow model, structural change and endogenous growth</i>	112
8.4	<i>Human capital and the dynamic Solow model</i>	113
8.5	<i>Exogenous and endogenous growth</i>	116

8.6	<i>A Solow interpretation of development patterns</i>	118
	<i>References</i>	118
	<i>Exercise</i>	119
	<i>Appendix: deriving the dynamic Solow model</i>	119

SECTION III

	Panel data	121
9	Panel data: an introduction	123
	9.1 <i>Introduction</i>	123
	9.2 <i>Panel data</i>	123
	9.2.1 <i>The structure of the panel</i>	123
	9.2.2 <i>Panel data and endogeneity</i>	124
	9.3 <i>Panel production functions</i>	127
	9.3.1 <i>A panel macro production function</i>	127
	9.3.2 <i>A panel micro production function</i>	130
	9.4 <i>Interpreting the fixed effect</i>	134
	<i>References</i>	135
	<i>Exercise</i>	135
	<i>Appendix: matrix notation</i>	135
10	Panel estimators: POLS, RE, FE, FD	140
	10.1 <i>Introduction</i>	140
	10.2 <i>Panel estimators</i>	140
	10.2.1 <i>The fixed effects and first difference estimators</i>	140
	10.2.2 <i>The random effects estimator</i>	142
	10.3 <i>Key assumptions for consistency</i>	143
	10.4 <i>Model selection</i>	144
	10.4.1 <i>Testing for correlation between the c_i and the explanatory variables</i>	145
	10.4.2 <i>Testing for the presence of an unobserved effect</i>	146
	10.5 <i>The micro panel production function extended</i>	147
	10.6 <i>What determines the productivity of Ghanaian firms?</i>	148
	<i>References</i>	152
	<i>Exercise</i>	152
11	Instrumental variables and endogeneity	153
	11.1 <i>Introduction</i>	153
	11.2 <i>Sources of bias in the OLS estimates</i>	153
	11.2.1 <i>Bias from omitted variables</i>	153
	11.2.2 <i>Bias from measurement error</i>	154
	11.2.3 <i>Panel data: omitted variables and measurement error</i>	155
	11.3 <i>Instrumental variables</i>	156
	11.3.1 <i>Valid and informative instruments</i>	157

11.3.2	Interpreting the IV estimator	159
11.4	<i>The properties of the IV estimator</i>	160
11.4.1	The IV and OLS estimators compared	160
11.4.2	Inference with the IV estimator	161
11.5	<i>The causes of differences in world incomes</i>	162
	<i>Exercise</i>	167
	<i>References</i>	168

SECTION IV

	An introduction to programme evaluation	169
12	The programme evaluation approach to development policy	171
12.1	<i>Introduction: causal effects and the counterfactual problem</i>	171
12.2	<i>Rubin causal model</i>	172
12.2.1	Potential outcomes	172
12.2.2	Assignment mechanism	173
12.2.3	Defining measures of impact	174
12.2.4	From potential outcomes to regression	174
12.3	<i>Selection on observables</i>	177
12.3.1	Ignorability of treatment	177
12.3.2	Overlap	178
12.4	<i>Unconditional unconfoundedness and the experimental approach</i>	179
	<i>References</i>	180
	<i>Exercise</i>	180
13	Models, experiments and calibration in development policy analysis	182
13.1	<i>Introduction</i>	182
13.2	<i>Empirical estimators under (conditional) unconfoundedness</i>	182
13.2.1	Multivariate regression	183
13.2.2	Panel data methods	184
13.3	<i>A randomised controlled trial (RCT) for conditional cash transfers</i>	185
13.4	<i>Calibrating technology</i>	188
13.5	<i>Education, technology and poverty</i>	190
	<i>References</i>	190
	<i>Exercise</i>	191

PART 2

	Modelling development	193
14	Measurement, models and methods for understanding poverty	195
14.1	<i>Introduction</i>	195
14.2	<i>The causes of poverty</i>	195
14.2.1	Poverty and GDP data	195
14.2.2	Poverty, consumption and incomes	196

14.2.3	Poverty, inequality and GDP	197
14.3	<i>The Mincerian earnings function, the price of labour and poverty</i>	199
14.4	<i>Modelling impacts</i>	201
14.4.1	A generalised Roy model of selection	201
14.4.2	Implications of the Roy model for estimation of treatment effects	202
14.5	<i>An overview: measurement, models and methods</i>	203
	<i>References</i>	204
	<i>Exercise</i>	205

SECTION V

Modelling choice		207
15	Maximum likelihood estimation	209
15.1	<i>Introduction</i>	209
15.2	<i>The concept of maximum likelihood</i>	209
15.3	<i>The concept of population</i>	211
15.4	<i>Distributional assumptions and the log-likelihood function</i>	211
15.5	<i>Maximising the (log-)likelihood</i>	214
15.6	<i>Maximum likelihood in Stata</i>	215
15.7	<i>Problems and warnings ...</i>	218
15.7.1	Maximum likelihood and endogeneity	218
15.7.2	Maximum likelihood and convergence	219
15.8	<i>Properties of maximum likelihood estimates</i>	220
15.8.1	Consistency	221
15.8.2	Efficiency	221
15.8.3	So what?	221
15.9	<i>Hypothesis testing under maximum likelihood</i>	222
15.10	<i>Overview</i>	224
	<i>References</i>	224
	<i>Exercise</i>	224
16	Modelling choice: the LPM, probit and logit models	226
16.1	<i>Introduction</i>	226
16.2	<i>Binary choices and interpreting the descriptive statistics</i>	227
16.3	<i>Estimation by OLS: the linear probability model</i>	228
16.4	<i>The probit and logit models as latent variable models</i>	231
16.4.1	The probit model	232
16.4.2	The logit model	234
16.5	<i>Maximum likelihood estimation of probit and logit models</i>	234
16.6	<i>Explaining choice</i>	235
	<i>References</i>	237
	<i>Exercise</i>	237

17	Using logit and probit models for unemployment and school choice	239
17.1	<i>Introduction</i>	239
17.2	<i>Interpreting the probit model and the logit model</i>	240
17.2.1	A model of unemployment	240
17.2.2	Average partial effects and marginal effects at the mean	240
17.2.3	Age and education as determinants of unemployment in South Africa	245
17.3	<i>Goodness of fit</i>	245
17.4	<i>Indian private and state schools</i>	248
17.4.1	How well do private schools perform?	248
17.4.2	Who attends a private school?	249
17.4.3	Mother's education and wealth as determinants of attending private school in India	250
17.5	<i>Models of unemployment and school choice</i>	250
	<i>References</i>	252
	<i>Exercise</i>	252
18	Corner solutions: modelling investing in children and by firms	254
18.1	<i>Introduction</i>	254
18.2	<i>OLS estimation of corner response models</i>	255
18.2.1	Investment in Ghana's manufacturing sector	255
18.2.2	Gender discrimination in India	258
18.3	<i>The Tobit model</i>	260
18.4	<i>Two-part models</i>	262
18.4.1	Truncated normal hurdle model	264
18.4.2	The log-normal hurdle model	265
18.5	<i>Overview</i>	268
	<i>References</i>	268
	<i>Exercise</i>	269
	<i>Appendix: the Inverse Mills Ratio (IMR)</i>	269

SECTION VI

Structural modelling 271

19	An introduction to structural modelling in development economics	273
19.1	<i>Introduction: the challenge of using microeconomic theory in empirical research</i>	273
19.2	<i>Using a structural model to think about risk-sharing</i>	274
19.3	<i>Building and solving a microeconomic model</i>	276
19.4	<i>Thinking about unobservables and choosing an estimator</i>	281
19.4.1	The model to be estimated	281
19.4.2	Identification in the model	282
19.4.3	Testing the model	282
19.5	<i>Estimating the model</i>	283

19.5.1	The data	283	
19.5.2	Estimation results	283	
19.6	Conclusion	284	
	References	285	
	Exercise	285	
20	Structural methods and the return to education		286
20.1	<i>Introduction: Belzil and Hansen go to Africa</i>	286	
20.2	<i>The question</i>	286	
20.3	<i>A model of investment in education</i>	287	
20.4	<i>Thinking about unobservables and choosing an estimator</i>	292	
20.5	<i>Models and data</i>	296	
20.5.1	'Adolescent econometricians'?	296	
20.5.2	Possible applications for structural modelling in development	297	
20.6	<i>Structural models: hubris or humility?</i>	298	
	References	298	
	Exercise	299	
SECTION VII			
Selection, heterogeneity and programme evaluation			301
21	Sample selection: modelling incomes where occupation is chosen		303
21.1	<i>Introduction</i>	303	
21.2	<i>Sample selection</i>	303	
21.3	<i>A formal exposition</i>	304	
21.3.1	The regression with sample selection	304	
21.3.2	Modelling the correlation of the unobservables	305	
21.4	<i>When is sample selection a problem?</i>	308	
21.5	<i>Selection and earnings in South Africa</i>	309	
21.6	<i>Corner solution and sample selection models</i>	313	
	References	314	
	Exercise	314	
22	Programme evaluation: regression discontinuity and matching		316
22.1	<i>Introduction</i>	316	
22.2	<i>Regression discontinuity design</i>	316	
22.3	<i>Propensity score methods</i>	319	
22.3.1	Regression using the propensity score	319	
22.3.2	Weighting by the propensity score	320	
22.3.3	Matching on the propensity score	321	
22.4	<i>Food aid in Ethiopia: propensity-score matching</i>	322	
22.5	<i>Assessing the consequences of property rights: pipeline identification strategies</i>	323	
22.6	<i>Estimating treatment effects (the plot so far)</i>	326	

23	Heterogeneity, selection and the marginal treatment effect (MTE)	328
23.1	<i>Introduction</i>	328
23.2	<i>Instrumental variables estimates under homogeneous treatment effects</i>	328
23.3	<i>Instrumental variables estimates under heterogeneous treatment effects</i>	330
23.3.1	IV for noncompliance and heterogeneous effects: the LATE Theorem	330
23.3.2	LATE and the compliant subpopulation	332
23.4	<i>Selection and the marginal treatment effect</i>	333
23.4.1	Interpreting the LATE in the context of the Roy model	333
23.4.2	The marginal treatment effect	336
23.4.3	What does IV identify?	337
23.5	<i>The return to education once again</i>	339
23.6	<i>An overview</i>	341
	<i>References</i>	342
	<i>Exercise</i>	342

SECTION VIII

Dynamic models for micro and macro data 345

24	Estimation of dynamic effects with panel data	347
24.1	<i>Introduction</i>	347
24.2	<i>Instrumental variable estimation of dynamic panel-data models</i>	348
24.3	<i>The Arellano–Bond estimator</i>	349
24.3.1	No serial correlation in the errors	349
24.3.2	Serially correlated errors	350
24.4	<i>The system GMM estimator</i>	351
24.5	<i>Estimation of dynamic panel-data models using Stata</i>	352
24.6	<i>The general case</i>	355
24.6.1	The regressors are strictly exogenous	355
24.6.2	The regressors are predetermined	356
24.6.3	The regressors are contemporaneously endogenous	357
24.6.4	Implications of serial correlation in the error term	357
24.7	<i>Using the estimators</i>	358
	<i>References</i>	358
	<i>Exercise</i>	359
	<i>Appendix: the bias in the fixed effects estimator of a dynamic panel-data model</i>	359

25	Modelling the effects of aid and the determinants of growth	361
25.1	<i>Introduction</i>	361
25.2	<i>Dynamic reduced-form models</i>	361

25.2.1	Aid, policy and growth	361
25.2.2	Dynamics and lags	364
25.2.3	Differenced and system GMM estimators	366
25.3	<i>Growth rate effects: a model of endogenous growth</i>	368
25.3.1	Dynamic and growth rate models	368
25.3.2	Is there evidence for endogenous growth?	370
25.4	<i>Aid, policy and growth revisited with annual data</i>	371
25.4.1	Cross section and time-series uses of macro data	371
25.4.2	Growth and levels effects of aid	371
25.5	<i>A brief overview: aid, policy and growth</i>	372
	<i>References</i>	373
	<i>Exercise</i>	373

SECTION IX

Dynamics and long panels 375

26 Understanding technology using long panels 377

26.1	<i>Introduction</i>	377
26.2	<i>Parameter heterogeneity in long panels</i>	378
26.3	<i>The mean group estimator</i>	379
26.4	<i>Cross-section dependence due to common factors</i>	383
26.5	<i>Conclusion</i>	386
	<i>References</i>	386
	<i>Exercise</i>	386

27 Cross-section dependence and nonstationary data 388

27.1	<i>Introduction</i>	388
27.2	<i>Alternative approaches to modelling cross-section dependence</i>	388
27.2.1	Country fixed effects and year dummies	389
27.2.2	Estimating unobserved common factors	389
27.2.3	Constructing weight matrices	390
27.3	<i>Modelling cross-section dependence using cross-section averages</i>	390
27.4	<i>Detecting cross-section dependence</i>	393
27.5	<i>Panel unit root testing</i>	394
27.5.1	First-generation panel unit root test	394
27.5.1.1	The Im, Pesaran and Shin test (IPS)	395
27.5.1.2	The Maddala and Wu test (<i>MW</i>)	395
27.5.2	Second-generation panel unit root test	395
27.5.2.1	The PANIC approach	395
27.5.2.2	The CIPS and CIPSM tests	396
27.6	<i>Cointegration testing in panels</i>	396
27.6.1	Residual analysis and error-correction models	396
27.6.2	Tests for panel cointegration	397

27.7	<i>Parameter heterogeneity, nonstationary data and cross-section dependence</i>	397
	<i>References</i>	399
	<i>Exercise</i>	400
28	Macro production functions for manufacturing and agriculture	402
28.1	<i>Introduction</i>	402
28.2	<i>Estimating a production function for manufacturing</i>	403
28.2.1	The homogeneous models	403
28.2.2	The heterogeneous models	405
28.3	<i>Estimating a production function for agriculture</i>	407
28.3.1	Unit roots	408
28.3.2	What determines the productivity of agriculture?	409
28.4	<i>Manufacturing and agriculture and the growth of an economy</i>	412
	<i>References</i>	412
	<i>Exercise</i>	413
SECTION X		
	An overview	415
29	How can the processes of development best be understood?	417
29.1	<i>Introduction</i>	417
29.2	<i>A range of answers as to the causes of poverty</i>	417
29.3	<i>Macro policy, growth and poverty reduction</i>	419
29.4	<i>Programme evaluation and structural models</i>	419
29.4.1	Programme evaluation and the 'failure' of poverty policies	419
29.4.2	Structural models and understanding the causes of poverty	420
29.5	<i>Skills, technology and the returns on investment</i>	420
29.5.1	The value of skills	420
29.5.2	The role of technology	421
29.5.3	Rates of return on investment	421
29.6	<i>A final word</i>	421
	<i>References</i>	422
	<i>Bibliography</i>	423
	<i>Index</i>	431