Empirical Development Economics

Måns Söderbom and Francis Teal
with Markus Eberhardt, Simon Quinn and Andrew Zeitlin

Routledge
Taylor & Francis Group
LONDON AND NEW YORK
## Contents

*List of figures*  
xviii  
*List of tables*  
xx  
*Notes on authors*  
xxiii  
*Preface*  
xxiv  
*How to use this book*  
xxvii  

### PART I  
**Linking models to data for development**  

1. An introduction to empirical development economics  
1.1 *The objective of the book*  
1.2 *Models and data: the Harris–Todaro model*  
1.3 *Production functions and functional form*  
1.3.1 The Cobb–Douglas production function  
1.3.2 The constant elasticity of substitution (CES) functional form  
1.4 *A model with human capital*  
1.5 *Data and models*  
1.5.1 The macro GDP data  
1.5.2 Interpreting the data  

*References*  
*Exercise*  

### SECTION I  
**Cross-section data and the determinants of incomes**  

2. The linear regression model and the OLS estimator  
2.1 *Introduction: models and causality*  
2.2 *The linear regression model and the OLS estimators*  
2.2.1 The linear regression model as a population model  
2.2.2 The zero conditional mean assumption  
2.2.3 The OLS estimator  
2.3 *The Mincerian earnings function for the South African data*
# Contents

2.4 *Properties of the OLS estimators*  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 *Identifying the causal effect of education*  31  

References  31  

Exercise  32  

3 Using and extending the simple regression model  33  

3.1 *Introduction*  33  

3.2 *Dummy explanatory variables and the return to education*  33  

3.3 *Multiple regression*  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 *Interpreting multiple regressions*  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  

References  45  

Exercise  45  

4 The distribution of the OLS estimators and hypothesis testing  47  

4.1 *Introduction*  47  

4.2 *The distribution of the OLS estimators*  47  
   4.2.1 The normality assumption  47  
   4.2.2 Why normality?  48  

4.3 *Testing hypotheses about a single population parameter*  49  
   4.3.1 The *t*-distribution  49  
   4.3.2 The *t*-test  51  
   4.3.3 Confidence intervals  53  

4.4 *Testing for the overall significance of a regression*  55  

4.5 *Testing for heteroskedasticity*  57  

4.6 *Large sample properties of OLS*  58  
   4.6.1 Consistency  58  
   4.6.2 Asymptotic normality  60  

References  60  

Exercise  61  

5 The determinants of earnings and productivity  62  

5.1 *Introduction*  62  

5.2 *Testing the normality assumption*  62  

5.3 *The earnings function*  65  
   5.3.1 Bringing the tests together  65  
   5.3.2 Robust and clustered standard errors  65
11.3.2 Interpreting the IV estimator 159
11.4 The properties of the IV estimator 160
11.4.1 The IV and OLS estimators compared 160
11.4.2 Inference with the IV estimator 161
11.5 The causes of differences in world incomes 162
Exercise 167
References 168

SECTION IV
An introduction to programme evaluation 169

12 The programme evaluation approach to development policy 171
12.1 Introduction: causal effects and the counterfactual problem 171
12.2 Rubin causal model 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 Selection on observables 177
12.3.1 Ignorability of treatment 177
12.3.2 Overlap 178
12.4 Unconditional unconfoundedness and the experimental approach 179
References 180
Exercise 180

13 Models, experiments and calibration in development policy analysis 182
13.1 Introduction 182
13.2 Empirical estimators under (conditional) unconfoundedness 182
13.2.1 Multivariate regression 183
13.2.2 Panel data methods 184
13.3 A randomised controlled trial (RCT) for conditional cash transfers 185
13.4 Calibrating technology 188
13.5 Education, technology and poverty 190
References 190
Exercise 191

PART 2
Modelling development 193

14 Measurement, models and methods for understanding poverty 195
14.1 Introduction 195
14.2 The causes of poverty 195
14.2.1 Poverty and GDP data 195
14.2.2 Poverty, consumption and incomes 196
Contents

14.2.3 Poverty, inequality and GDP 197

14.3 The Mincerian earnings function, the price of labour and poverty 199

14.4 Modelling impacts 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 An overview: measurement, models and methods 203

References 204

Exercise 205

SECTION V

Modelling choice 207

15 Maximum likelihood estimation 209

15.1 Introduction 209

15.2 The concept of maximum likelihood 209

15.3 The concept of population 211

15.4 Distributional assumptions and the log-likelihood function 211

15.5 Maximising the (log-)likelihood 214

15.6 Maximum likelihood in Stata 215

15.7 Problems and warnings ... 218

15.7.1 Maximum likelihood and endogeneity 218

15.7.2 Maximum likelihood and convergence 219

15.8 Properties of maximum likelihood estimates 220

15.8.1 Consistency 221

15.8.2 Efficiency 221

15.8.3 So what? 221

15.9 Hypothesis testing under maximum likelihood 222

15.10 Overview 224

References 224

Exercise 224

16 Modelling choice: the LPM, probit and logit models 226

16.1 Introduction 226

16.2 Binary choices and interpreting the descriptive statistics 227

16.3 Estimation by OLS: the linear probability model 228

16.4 The probit and logit models as latent variable models 231

16.4.1 The probit model 232

16.4.2 The logit model 234

16.5 Maximum likelihood estimation of probit and logit models 234

16.6 Explaining choice 235

References 237

Exercise 237
17 Using logit and probit models for unemployment and school choice 239
   17.1 Introduction 239
   17.2 Interpreting the probit model and the logit model 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 Goodness of fit 245
   17.4 Indian private and state schools 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 Models of unemployment and school choice 250
References 252
Exercise 252

18 Corner solutions: modelling investing in children and by firms 254
   18.1 Introduction 254
   18.2 OLS estimation of corner response models 255
      18.2.1 Investment in Ghana’s manufacturing sector 255
      18.2.2 Gender discrimination in India 258
   18.3 The Tobit model 260
   18.4 Two-part models 262
      18.4.1 Truncated normal hurdle model 264
      18.4.2 The log-normal hurdle model 265
   18.5 Overview 268
References 268
Exercise 269
Appendix: the Inverse Mills Ratio (IMR) 269

SECTION VI
Structural modelling 271

19 An introduction to structural modelling in development economics 273
   19.1 Introduction: the challenge of using microeconomic theory in empirical research 273
   19.2 Using a structural model to think about risk-sharing 274
   19.3 Building and solving a microeconomic model 276
   19.4 Thinking about unobservables and choosing an estimator 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 Estimating the model 283
23 Heterogeneity, selection and the marginal treatment effect (MTE) 328
23.1 Introduction 328
23.2 Instrumental variables estimates under homogeneous treatment effects 328
23.3 Instrumental variables estimates under heterogeneous treatment effects 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 Selection and the marginal treatment effect 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 The return to education once again 339
23.6 An overview 341
References 342
Exercise 342

SECTION VIII
Dynamic models for micro and macro data 345

24 Estimation of dynamic effects with panel data 347
   24.1 Introduction 347
   24.2 Instrumental variable estimation of dynamic panel-data models 348
   24.3 The Arellano–Bond estimator 349
      24.3.1 No serial correlation in the errors 349
      24.3.2 Serially correlated errors 350
   24.4 The system GMM estimator 351
   24.5 Estimation of dynamic panel-data models using Stata 352
   24.6 The general case 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 Using the estimators 358
References 358
Exercise 359
Appendix: the bias in the fixed effects estimator of a dynamic panel-data model 359

25 Modelling the effects of aid and the determinants of growth 361
   25.1 Introduction 361
   25.2 Dynamic reduced-form models 361
Contents

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 Growth rate effects: a model of endogenous growth 368
25.3.1 Dynamic and growth rate models 368
25.3.2 Is there evidence for endogenous growth? 370

25.4 Aid, policy and growth revisited with annual data 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 A brief overview: aid, policy and growth 372

References 373
Exercise 373

SECTION IX

Dynamics and long panels 375

26 Understanding technology using long panels 377
26.1 Introduction 377
26.2 Parameter heterogeneity in long panels 378
26.3 The mean group estimator 379
26.4 Cross-section dependence due to common factors 383
26.5 Conclusion 386
References 386
Exercise 386

27 Cross-section dependence and nonstationary data 388
27.1 Introduction 388
27.2 Alternative approaches to modelling cross-section dependence 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 Modelling cross-section dependence using cross-section averages 390
27.4 Detecting cross-section dependence 393
27.5 Panel unit root testing 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 (MW) 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 Cointegration testing in panels 396
27.6.1 Residual analysis and error-correction models 396
27.6.2 Tests for panel cointegration 397
27.7 Parameter heterogeneity, nonstationary data and cross-section dependence 397
References 399
Exercise 400

28 Macro production functions for manufacturing and agriculture 402
28.1 Introduction 402
28.2 Estimating a production function for manufacturing 403
   28.2.1 The homogeneous models 403
   28.2.2 The heterogeneous models 405
28.3 Estimating a production function for agriculture 407
   28.3.1 Unit roots 408
   28.3.2 What determines the productivity of agriculture? 409
28.4 Manufacturing and agriculture and the growth of an economy 412
References 412
Exercise 413

SECTION X
An overview 415

29 How can the processes of development best be understood? 417
29.1 Introduction 417
29.2 A range of answers as to the causes of poverty 417
29.3 Macro policy, growth and poverty reduction 419
29.4 Programme evaluation and structural models 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 Skills, technology and the returns on investment 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 A final word 421
References 422

Bibliography 423
Index 431