

Contents

Preface — VII

Acknowledgments — IX

Acronyms — XV

1 Introduction — 1

- 1.1 Some illustrative examples — 1
 - 1.1.1 Predicting the cross-section of stock returns — 1
 - 1.1.2 Explaining flows to mutual funds — 2
 - 1.1.3 The impact of CEO compensation or board structure on firm value — 3
 - 1.1.4 Explaining capital structure choice — 4
 - 1.1.5 Governance and firm value — 5
 - 1.1.6 Dividend policy — 6
- 1.2 The structure of this text — 7

2 Linear static models — 9

- 2.1 The linear regression model with cross-sectional data — 9
 - 2.1.1 The linear model and ordinary least squares — 9
 - 2.1.2 Heteroskedasticity — 15
 - 2.1.3 Hypothesis testing — 17
- 2.2 The linear regression model with time-series data — 19
- 2.3 Introducing panel data — 22
 - 2.3.1 Parameter heterogeneity — 23
 - 2.3.2 Illustration: testing an asset pricing model — 24
- 2.4 The linear regression model with panel data — 26
- 2.5 Clustered standard errors — 29
- 2.6 Random effects estimators — 38
- 2.7 Fixed effects estimators — 42
- 2.8 Clustered standard errors with fixed effects — 50
- 2.9 First-difference estimators — 52
- 2.10 Testing for heteroskedasticity and autocorrelation — 55
- 2.11 Models with fixed time effects — 57
- 2.12 The Fama-MacBeth approach — 58
 - 2.12.1 Cross-sectional regressions — 58
 - 2.12.2 Serial correlation and firm effects — 60
 - 2.12.3 Pooled OLS versus Fama-MacBeth — 62
 - 2.12.4 The errors-in-variables problem — 64
- 2.13 Goodness-of-fit — 65

3 Dealing with heterogeneity and endogeneity: fixed effects, IV and GMM — 68

- 3.1 Fixed effects and instrumental variables — 68
- 3.2 Two-way and interactive fixed effects — 71
- 3.3 How not to control for unobserved heterogeneity — 78
- 3.4 More on instrumental variables — 82
 - 3.4.1 Two-stage least squares — 82
 - 3.4.2 Weak instruments — 89
 - 3.4.3 Standard errors — 91
 - 3.4.4 IV estimators with panel data — 92
- 3.5 Instrumental variables and GMM — 94
 - 3.5.1 Moment conditions — 94
 - 3.5.2 The optimal weighting matrix — 96
 - 3.5.3 Tests for overidentifying restrictions — 99
 - 3.5.4 Panel estimators — 100
- 3.6 Strict exogeneity — 102

4 Outliers, missing values and other data issues — 107

- 4.1 Outliers and influential observations — 107
 - 4.1.1 Identifying and eliminating outliers — 108
 - 4.1.2 Robust estimation methods — 111
- 4.2 Missing values — 112
- 4.3 Incomplete panels — 115
 - 4.3.1 Estimation with randomly missing data — 116
 - 4.3.2 Selection bias and some simple tests — 118
- 4.4 Measurement errors — 120

5 Linear dynamic models — 126

- 5.1 The problem of unobserved heterogeneity — 126
- 5.2 The problem with the fixed effects estimator — 129
- 5.3 Instrumental variables — 130
- 5.4 The Arellano-Bond estimator — 132
 - 5.4.1 Moment conditions — 132
 - 5.4.2 Models with exogenous variables — 135
 - 5.4.3 System GMM — 136
 - 5.4.4 Specification tests — 137
 - 5.4.5 When γ is close to one — 137
 - 5.4.6 Too many instruments — 138
- 5.5 Applications — 140

6 Models with limited dependent variables — 145

- 6.1 Binary choice models — 146

6.1.1	Specification and interpretation — 147
6.1.2	Estimation: maximum likelihood — 154
6.1.3	Pooled estimation approaches — 159
6.1.4	Clustered standard errors — 161
6.1.5	Fama-MacBeth estimation — 163
6.1.6	Binary choice models with random effects — 164
6.1.7	Binary choice models with fixed effects — 168
6.1.8	Goodness-of-fit — 172
6.1.9	Binary choice models and instrumental variables — 173
6.2	Multiple outcomes — 175
6.2.1	Ordered probit and ordered logit — 176
6.2.2	Multinomial models — 179
6.3	Tobit models — 181
6.3.1	Introduction to censoring — 181
6.3.2	Pooled estimation — 185
6.3.3	A tobit model with random effects — 187
6.3.4	Extensions — 189
6.4	Dynamic models — 196
6.4.1	State dependence versus unobserved heterogeneity — 196
6.4.2	A dynamic probit model with unobserved heterogeneity — 197
6.5	Count data — 199
6.6	Duration models — 201
6.6.1	Hazard rates and survival functions — 202
6.6.2	Estimation — 204
7	Estimating average treatment effects — 208
7.1	Potential outcomes — 209
7.1.1	Average treatment effects — 210
7.1.2	Randomised experiments — 211
7.1.3	Biases — 212
7.2	Conditional independence — 215
7.2.1	Weighting — 216
7.2.2	Regression-adjustment — 217
7.2.3	Regression with a treatment indicator — 220
7.2.4	Matching — 221
7.3	Regression discontinuity — 225
7.3.1	Sharp RDD — 226
7.3.2	Challenges to identification — 230
7.3.3	Extensions and variations — 232
7.4	A switching regression model — 234
7.5	Instrumental variables — 236
7.6	Difference-in-differences — 241

7.6.1 The basic difference-in-differences approach — 242

7.6.2 Extensions — 248

Bibliography — 255

Index — 275