Advanced Studies in Theoretical and Applied Econometrics

Volume 46

The titles published in this series are listed at the end of this volume.

László Mátyás • Patrick Sevestre (Eds.)

The Econometrics of Panel Data

Fundamentals and Recent Developments in Theory and Practice

Third Edition

With 13 Figures and 43 Tables

Springer

Editors

Prof. László Mátyás
Central European University
Department of Economics
Nádor u. 9
1051 Budapest
Hungary
matyas@ceu.hu

Prof. Patrick Sevestre
Université Paris 1-Panthéon Sorbonne
Ecole Economique de Paris
(Paris School of Economics)
106-112 Boulevard de l'Hôpital
75013 Paris
France
sevestre@univ-paris1.fr

ISBN: 978-3-540-75889-1 e-ISBN: 978-3-540-75892-1

Advanced Studies in Theoretical and Applied Econometrics ISSN: 1570-5811

Library of Congress Control Number: 2007940369

2nd edition was published by Springer Netherlands in 1995

Cover design: eStudio Calamar S.L., F. Steinen-Broo, Pau/Girona, Spain

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The aim of this third, completely re-written, re-edited and considerably expanded, edition of this book is to provide a general overview of both the basics and recent, more sophisticated, theoretical developments in panel data econometrics. It also aims at covering a number of fields of applications where these methods are used for improving our knowledge and understanding of economic agents' behaviors. Since the pioneering works of Edwin Kuh (1959), Yair Mundlak (1961), Irving Hoch (1962), and Pietro Balestra and Marc Nerlove (1966), the pooling of cross sections and time series data has become an increasingly popular way of quantifying economic relationships. Each series provides information lacking in the other, so a combination of both leads to more accurate, reliable and informative results than would be achievable by one type of series alone. Over the last three decades of the last century, much fundamental work has been done: investigation of the properties of different estimators and test statistics, analysis of dynamic models and the effects of eventual measurement errors, etc.

The more recent years and in particular the ten years elapsed since the second edition of this book have witnessed even more considerable changes. Indeed, our ability to estimate and test nonlinear models have dramatically improved and issues such as the unobserved heterogeneity in nonlinear models, attrition and selectivity bias have received considerable attention. This explains why the number of chapters dealing with such issues has increased in this third edition. Other recent and important developments relate to the issue of unit roots and cointegration in long times series panels as well as that of cross-sectional dependence that occur in particular in spatial models, and else.

The first objective of this book, which takes up Parts I and II, is to give a complete and state of the art presentation of these theoretical developments. Part I is concerned with the basic fixed effects, random effects and random coefficients models, both linear and nonlinear; Part II deals with various extensions: dynamic models with small T panels, dynamic models with large T panels, models with other sources of endogeneity (measurement errors, simultaneity) and also provides an overview of recent developments in several other directions: attrition and selection bias, pseudo-panels, semi- and non-parametric methods, the Bayesian approach to panel data, the poolability of individuals, duration models and point processes, and count data models. The second objective of this volume is to provide insights into the use of panel data in empirical studies. Since the beginnings, interest in panel data has mostly been empirically motivated. Panel data methods have gained an increased importance over time and are now applied in a very large spectrum of economic studies. Part III thus deals with studies in several major fields of applied economics, such as foreign direct investments, production frontiers, linked employer-employee data, labor supply, policy analysis and transitions on the labor market. Some of the chapters in this third edition are revised versions of those already published in the previous ones, while several others are completely new contributions. In this respect, we are particularly happy to welcome aboard our new authors. Their input definitely helped to substantially improve the contents of this volume.

The double emphasis of this book (theoretical and applied), together with the fact that all the chapters have been written by well-known specialists in the field, encourage us to hope that it has now become a standard reference textbook for all those who are concerned with the use of panel data in econometrics, whether they are advanced students, professional economists or researchers. The editors have tried to standardize the notation, language, depth, etc. in order to present a coherent book. However, each chapter is capable of standing on its own as a reference in its own topic.

We must address our thanks to all those who have facilitated the creation of this book: the contributors who produced quality work, then took part in an internal refereeing process to ensure a high overall standard; Kluwer Academic Publishers, who had the foresight to publish in a subject which, at the time of the first edition, had a limited, but expanding, audience; and of course Springer which has become our publisher by now. In particular, Cathelijne van Herwaarden, Herma Drees, Marie Sheldon, Martina Bihn, Ruth Milewski and Isabelle George must be thanked for their help in the realization of this volume. Also, the University of Paris–Val de Marne in France; the Monash Research Fund and the Australian Research Council in Australia, and the Budapest University of Economics and the Hungarian Research Fund (OTKA) in Hungary must be thanked for having provided financial support to the editors for the earlier editions. This third edition has benefited from generous financial support provided by the Central European University and the University of Paris-Val de Marne.

Budapest and Paris
January 2008

László Mátyás
Patrick Sevestre

Contents

Part I Fundamentals

1 Introduction 3
Marc Nerlove, Patrick Sevestre and Pietro Balestra
1.1 Introduction 3
1.2 Data, Data-Generating Processes (DGP), and Inference 4
1.3 History and Dynamics 8
1.4 A Brief Review of Other Methodological Developments 13
1.5 Conclusion 21
References 21

2 Fixed Effects Models and Fixed Coefficients Models 23
Pietro Balestra and Jayalakshmi Krishnakumar
2.1 The Covariance Model: Individual Effects Only 24
2.1.1 Specification 24
2.1.2 Estimation 25
2.1.3 Inference 28
2.2 The Covariance Model: Individual and Time Effects 29
2.2.1 Time Effects Only 29
2.2.2 Time and Individual Effects 30
2.2.3 Inference 32
2.3 Non-spherical Disturbances 33
2.3.1 What Variance–Covariance Stucture? 33
2.3.2 Two General Propositions for Fixed Effects Models 34
2.3.3 Individual Fixed Effects and Serial Correlation 36
2.3.4 Heteroscedasticity in Fixed Effects Models 38
2.4 Extensions 40
2.4.1 Constant Variables in One Dimension 40
2.4.2 Variable Slope Coefficients 41
2.4.3 Unbalanced Panels 44
References 48

3 Error Components Models 49
Badi H. Baltagi, László Mátyás and Patrick Sevestre
3.1 Introduction 49
3.2 The One-Way Error Components Model 50
3.2.1 Definition/Assumptions of the Model 50
3.2.2 The GLS Estimator 52
3.2.3 The Feasible GLS Estimator 55
3.2.4 Some Other Estimators 58
3.2.5 Prediction 63
3.3 More General Structures of the Disturbances 64
3.3.1 The Two-Way Error Components Model 64
3.3.2 Serial Correlation in the Disturbances 70
3.3.3 Two-Way Error Components vs Kmenta's Approach 73
3.3.4 Heteroskedasticity in the Disturbances 74
3.4 Testing 78
3.4.1 Testing for the Absence of Individual Effects 79
3.4.2 Testing for Uncorrelated Effects: Hausman's Test 80
3.4.3 Testing for Serial Correlation 81
3.4.4 Testing for Heteroskedasticity 82
3.5 Estimation Using Unbalanced Panels 84
References 85

4 Endogenous Regressors and Correlated Effects 89
Rachid Boumahdi and Alban Thomas
4.1 Introduction 89
4.2 Estimation of Transformed Linear Panel Data Models 90
4.2.1 Error Structures and Filtering Procedures 91
4.2.2 An IV Representation of the Transformed Linear Model . 93
4.3 Estimation with Time-Invariant Regressors 95
4.3.1 Introduction 95
4.3.2 Instrumental Variable Estimation 96
4.3.3 More Efficient IV Procedures 98
4.4 A Measure of Instrument Relevance 99
4.5 Incorporating Time-Varying Regressors 101
4.5.1 Instrumental Variables Estimation 102
4.6 GMM Estimation of Static Panel Data Models 104
4.6.1 Static Model Estimation 105
4.6.2 GMM Estimation with HT, AM and BMS Instruments . . 107
4.7 Unbalanced Panels 108
References 110

5 The Chamberlain Approach to Panel Data: An Overview and Some Simulations 113
Bruno Crépon and Jacques Mairesse
5.1 Introduction 113
5.2 The Chamberlain Π Matrix Framework 115

5.2.1 The Π Matrix ... 115
5.2.2 Relations Between Π and the Parameters of Interest ... 118
5.2.3 Four Important Cases ... 120
5.2.4 Restrictions on the Covariance Matrix of the Disturbances 124
5.2.5 A Generalization of the Chamberlain Method ... 125
5.2.6 The Vector Representation of the Chamberlain Estimating Equations ... 126
5.2.7 The Estimation of Matrix Π ... 127
5.3 Asymptotic Least Squares ... 130
5.3.1 ALS Estimation ... 130
5.3.2 The Optimal ALS Estimator ... 132
5.3.3 Specification Testing in the ALS Framework ... 135
5.4 The Equivalence of the GMM and the Chamberlain Methods ... 137
5.4.1 A Reminder on the GMM ... 137
5.4.2 Equivalence of the GMM and the Chamberlain Methods . 139
5.4.3 Equivalence in Specific Cases ... 140
5.5 Monte Carlo Simulations ... 144
5.5.1 Design of the Simulation Experiments ... 144
5.5.2 Consistency and Bias ... 147
5.5.3 Efficiency and Robustness ... 152
5.5.4 Standard Errors ... 155
5.5.5 Specification Tests ... 158
5.6 Appendix A: An Extended View of the Chamberlain Method ... 160
5.6.1 Simultaneous Equations Models ... 160
5.6.2 VAR Models ... 160
5.6.3 Endogenous Attrition ... 162
5.7 Appendix B: Vector Representation of the Chamberlain Estimating Equations ... 163
5.7.1 The Vec Operator ... 163
5.7.2 Correlated Effects ... 164
5.7.3 Errors in Variables ... 164
5.7.4 Weak Simultaneity ... 166
5.7.5 Combination of the Different Cases ... 166
5.7.6 Lagged Dependent Variable ... 167
5.7.7 Restrictions on the Covariance Matrix of the Disturbances 167
5.8 Appendix C: Manipulation of Equations and Parameters in the ALS Framework ... 168
5.8.1 Transformation of the Estimating Equations ... 168
5.8.2 Eliminating Parameters of Secondary Interest ... 169
5.8.3 Recovering Parameters of Secondary Interest Once Eliminated ... 170
5.8.4 Elimination of Auxiliary Parameters ... 173
5.9 Appendix D: Equivalence Between Chamberlain's, GMM and Usual Panel Data Estimators ... 174
5.10 Appendix E: Design of Simulation Experiments ... 177

5.10.1 Generating Process of the Variable x . . . 177
5.10.2 Regression Model . . . 178
5.10.3 Calibration of Simulations . . . 179
5.10.4 Three Scenarios . . . 180
5.10.5 The Chamberlain and GMM Estimators . . . 180
5.10.6 Standard Errors and Specification Tests . . . 181
References . . . 181

6 Random Coefficient Models . . . 185
Cheng Hsiao and M. Hashem Pesaran
6.1 Introduction . . . 185
6.2 The Models . . . 186
6.3 Sampling Approach . . . 189
6.4 Mean Group Estimation . . . 192
6.5 Bayesian Approach . . . 193
6.6 Dynamic Random Coefficients Models . . . 197
6.7 Testing for Heterogeneity Under Weak Exogeneity . . . 199
6.8 A Random Coefficient Simultaneous Equation System . . . 203
6.9 Random Coefficient Models with Cross-Section Dependence . . . 206
6.10 Concluding Remarks . . . 208
References . . . 211

7 Parametric Binary Choice Models . . . 215
Michael Lechner, Stéfan Lollivier and Thierry Magnac
7.1 Introduction . . . 215
7.2 Random Effects Models Under Strict Exogeneity . . . 217
7.2.1 Errors are Independent Over Time . . . 218
7.2.2 One Factor Error Terms . . . 219
7.2.3 General Error Structures . . . 221
7.2.4 Simulation Methods . . . 223
7.2.5 How to Choose a Random Effects Estimator for an Application . . . 228
7.2.6 Correlated Effects . . . 229
7.3 Fixed Effects Models Under Strict Exogeneity . . . 230
7.3.1 The Model . . . 231
7.3.2 The Method of Conditional Likelihood . . . 232
7.3.3 Fixed Effects Maximum Score . . . 235
7.3.4 GMM Estimation . . . 236
7.3.5 *Large*-T Approximations . . . 237
7.4 Dynamic Models . . . 238
7.4.1 Dynamic Random Effects Models . . . 238
7.4.2 Dynamic Fixed Effects Models . . . 241
References . . . 242

Part II Advanced Topics

8 Dynamic Models for Short Panels 249
Mark N. Harris, László Mátyás and Patrick Sevestre
8.1 Introduction 249
8.2 The Model 250
8.3 The Inconsistency of Traditional Estimators 252
8.4 IV and GMM Estimators 255
8.4.1 Uncorrelated Individual Effects: The Original Balestra–Nerlove Estimator and its Extensions 256
8.4.2 Correlated Individual Effects 257
8.4.3 Some Monte Carlo Evidence 269
8.5 The Maximum Likelihood Estimator 270
8.6 Testing in Dynamic Models 272
8.6.1 Testing the Validity of Instruments 272
8.6.2 Testing for Unobserved Effects 273
8.6.3 Testing for the Absence of Serial Correlation in ε 274
8.6.4 Significance Testing in Two-Step Variants 275
References 276

9 Unit Roots and Cointegration in Panels 279
Jörg Breitung and M. Hashem Pesaran
9.1 Introduction 279
9.2 First Generation Panel Unit Root Tests 281
9.2.1 The Basic Model 281
9.2.2 Derivation of the Tests 282
9.2.3 Null Distribution of the Tests 284
9.2.4 Asymptotic Power of the Tests 287
9.2.5 Heterogeneous Trends 288
9.2.6 Short-Run Dynamics 291
9.2.7 Other Approaches to Panel Unit Root Testing 293
9.3 Second Generation Panel Unit Root Tests 295
9.3.1 Cross-Section Dependence 295
9.3.2 Tests Based on GLS Regressions 296
9.3.3 Test Statistics Based on OLS Regressions 297
9.3.4 Other Approaches 298
9.4 Cross-Unit Cointegration 299
9.5 Finite Sample Properties of Panel Unit Root Tests 301
9.6 Panel Cointegration: General Considerations 302
9.7 Residual-Based Approaches to Panel Cointegration 306
9.7.1 Spurious Regression 306
9.7.2 Tests of Panel Cointegration 307
9.8 Tests for Multiple Cointegration 308
9.9 Estimation of Cointegrating Relations in Panels 309
9.9.1 Single Equation Estimators 309
9.9.2 System Estimators 312

9.10 Cross-Section Dependence and the Global VAR . . . 313
9.11 Concluding Remarks . . . 316
References . . . 316

10 Measurement Errors and Simultaneity . . . 323
Erik Biørn and Jayalakshmi Krishnakumar
10.1 Introduction . . . 323
10.2 Measurement Errors and Panel Data . . . 323
10.2.1 Model and Orthogonality Conditions . . . 325
10.2.2 Identification and the Structure of the Second Order Moments . . . 327
10.2.3 Moment Conditions . . . 328
10.2.4 Estimators Constructed from Period Means . . . 331
10.2.5 GMM Estimation and Testing in the General Case . . . 332
10.2.6 Estimation by GMM, Combining Differences and Levels 335
10.2.7 Extensions: Modifications . . . 343
10.2.8 Concluding Remarks . . . 343
10.3 Simultaneity and Panel Data . . . 344
10.3.1 SEM with EC . . . 345
10.3.2 Extensions . . . 361
10.4 Conclusion . . . 364
References . . . 365

11 Pseudo-Panels and Repeated Cross-Sections . . . 369
Marno Verbeek
11.1 Introduction . . . 369
11.2 Estimation of a Linear Fixed Effects Model . . . 370
11.3 Estimation of a Linear Dynamic Model . . . 376
11.4 Estimation of a Binary Choice Model . . . 380
11.5 Concluding Remarks . . . 381
References . . . 382

12 Attrition, Selection Bias and Censored Regressions . . . 385
Bo Honoré, Francis Vella and Marno Verbeek
12.1 Introduction . . . 385
12.2 Censoring, Sample Selection and Attrition . . . 386
12.3 Sample Selection and Attrition . . . 389
12.4 Sample Selection Bias and Robustness of Standard Estimators . . . 391
12.5 Tobit and Censored Regression Models . . . 393
12.5.1 Random Effects Tobit . . . 394
12.5.2 Random Effects Tobit with Endogenous Explanatory Variables . . . 396
12.5.3 Dynamic Random Effects Tobit . . . 398
12.5.4 Fixed Effects Tobit Estimation . . . 399
12.5.5 Semi-parametric Estimation . . . 401

12.5.6 Semi-parametric Estimation in the Presence of Lagged Dependent Variables . . . 402
12.6 Models of Sample Selection and Attrition . . . 402
12.6.1 Maximum Likelihood Estimators . . . 403
12.6.2 Two-Step Estimators . . . 404
12.6.3 Alternative Selection Rules . . . 407
12.6.4 Two-Step Estimators with Fixed Effects . . . 408
12.6.5 Semi-parametric Sample Selection Models . . . 409
12.6.6 Semi-parametric Estimation of a Type-3 Tobit Model . . . 410
12.7 Some Empirical Applications . . . 412
12.7.1 Attrition in Experimental Data . . . 412
12.7.2 Real Wages Over the Business Cycle . . . 413
12.7.3 Unions and Wages . . . 415
References . . . 416

13 Simulation Techniques for Panels: Efficient Importance Sampling . . . 419
Roman Liesenfeld and Jean-François Richard
13.1 Introduction . . . 419
13.2 Pseudorandom Number Generation . . . 420
13.2.1 Univariate Distributions . . . 421
13.2.2 Multivariate Distributions . . . 424
13.3 Importance Sampling . . . 426
13.3.1 General Principle . . . 426
13.3.2 Efficient Importance Sampling . . . 428
13.3.3 MC Sampling Variance of (E)IS Estimates . . . 431
13.3.4 GHK Simulator . . . 432
13.3.5 Common Random Numbers . . . 432
13.4 Simulation-Based Inference Procedures . . . 434
13.4.1 Integration in Panel Data Models . . . 434
13.4.2 Simulated Likelihood . . . 435
13.4.3 Simulated Method of Moments . . . 435
13.4.4 Bayesian Posterior Moments . . . 437
13.5 Numerical Properties of Simulated Estimators . . . 437
13.6 EIS Application: Logit Panel with Unobserved Heterogeneity . . . 439
13.6.1 The Model . . . 439
13.6.2 EIS Evaluation of the Likelihood . . . 440
13.6.3 Empirical Application . . . 443
13.7 Conclusion . . . 445
13.8 Appendix: Implementation of EIS for the Logit Panel Model . . . 446
References . . . 448

14 Semi-parametric and Non-parametric Methods in Panel Data Models . . . 451
Chunrong Ai and Qi Li
14.1 Introduction . . . 451
14.2 Linear Panel Data Model . . . 452

14.2.1 Additive Effect . . . 452
14.2.2 Multiplicative Effect . . . 460
14.3 Nonlinear Panel Data Model . . . 462
14.3.1 Censored Regression Model . . . 462
14.3.2 Discrete Choice Model . . . 470
14.3.3 Sample Selection Model . . . 474
14.4 Conclusion . . . 475
References . . . 476

15 Panel Data Modeling and Inference: A Bayesian Primer . . . 479
Siddhartha Chib
15.1 Introduction . . . 479
15.1.1 Hierarchical Prior Modeling . . . 480
15.1.2 Elements of Markov Chain Monte Carlo . . . 483
15.1.3 Some Basic Bayesian Updates . . . 486
15.1.4 Basic Variate Generators . . . 488
15.2 Continuous Responses . . . 489
15.2.1 Gaussian–Gaussian Model . . . 490
15.2.2 Robust Modeling of b_i: Student–Student and Student-Mixture Models . . . 492
15.2.3 Heteroskedasticity . . . 495
15.2.4 Serial Correlation . . . 496
15.3 Binary Responses . . . 497
15.4 Other Outcome Types . . . 501
15.4.1 Censored Outcomes . . . 501
15.4.2 Count Responses . . . 502
15.4.3 Multinomial Responses . . . 503
15.5 Binary Endogenous Regressor . . . 504
15.6 Informative Missingness . . . 507
15.7 Prediction . . . 508
15.8 Residual Analysis . . . 509
15.9 Model Comparisons . . . 509
15.9.1 Gaussian–Gaussian Model . . . 512
15.9.2 Gaussian–Gaussian Tobit model . . . 512
15.9.3 Panel Poisson Model . . . 513
15.10 Conclusion . . . 513
References . . . 514

16 To Pool or Not to Pool? . . . 517
Badi H. Baltagi, Georges Bresson and Alain Pirotte
16.1 Introduction . . . 517
16.2 Tests for Poolability, Pretesting and Stein-Rule Methods . . . 521
16.2.1 Tests for Poolability . . . 521
16.2.2 Pretesting and Stein-Rule Methods . . . 525
16.2.3 Example . . . 526
16.3 Heterogeneous Estimators . . . 527

16.3.1 Averaging Estimators . . . 529
16.3.2 Bayesian Framework . . . 530
16.3.3 An Example . . . 538
16.4 Comments on the Predictive Approach . . . 541
16.4.1 From the Post-sample Predictive Density . . . 541
16.4.2 . . . to the Good Forecast Performance of the Hierarchical Bayes Estimator: An Example . . . 542
16.5 Conclusion . . . 544
References . . . 545

17 Duration Models and Point Processes . . . 547
Jean-Pierre Florens, Denis Fougère and Michel Mouchart
17.1 Marginal Duration Models . . . 548
17.1.1 Distribution, Survivor and Density Functions . . . 548
17.1.2 Truncated Distributions and Hazard Functions . . . 550
17.2 Conditional Models . . . 552
17.2.1 General Considerations . . . 552
17.2.2 The Proportional Hazard or Cox Model . . . 555
17.2.3 The Accelerated Time Model . . . 557
17.2.4 Aggregation and Heterogeneity . . . 558
17.2.5 Endogeneity . . . 560
17.3 Competing Risks and Multivariate Duration Models . . . 561
17.3.1 Multivariate Durations . . . 561
17.3.2 Competing Risks Models: Definitions . . . 563
17.3.3 Identifiability of Competing Risks Models . . . 566
17.3.4 Right-Censoring . . . 568
17.4 Inference in Duration Models . . . 570
17.4.1 Introduction . . . 570
17.4.2 Parametric Models . . . 570
17.4.3 Non-parametric and Semi-parametric Models . . . 576
17.5 Counting Processes and Point Processes . . . 579
17.5.1 Definitions . . . 579
17.5.2 Stochastic Intensity, Compensator and Likelihood of a Counting Process . . . 581
17.6 Poisson, Markov and Semi-Markov Processes . . . 584
17.6.1 Poisson Processes . . . 584
17.6.2 Markov Processes . . . 585
17.6.3 Semi-Markov Processes . . . 592
17.7 Statistical Analysis of Counting Processes . . . 594
17.7.1 The Cox Likelihood . . . 596
17.7.2 The Martingale Estimation of the Integrated Baseline Intensity . . . 597
17.8 Conclusions . . . 600
References . . . 600

18 GMM for Panel Data Count Models 603
Frank Windmeijer
18.1 Introduction 603
18.2 GMM in Cross-Sections 604
18.3 Panel Data Models 606
18.3.1 Strictly Exogenous Regressors 607
18.3.2 Predetermined Regressors 608
18.3.3 Endogenous Regressors 609
18.3.4 Dynamic Models 610
18.4 GMM 612
18.5 Applications and Software 614
18.6 Finite Sample Inference 615
18.6.1 Wald Test and Finite Sample Variance Correction 615
18.6.2 Criterion-Based Tests 617
18.6.3 Continuous Updating Estimator 618
18.6.4 Monte Carlo Results 619
References 623

19 Spatial Panel Econometrics 625
Luc Anselin, Julie Le Gallo and Hubert Jayet
19.1 Introduction 625
19.2 Spatial Effects 626
19.2.1 Spatial Weights and Spatial Lag Operator 628
19.2.2 Spatial Lag Model 630
19.2.3 Spatial Error Model 632
19.3 A Taxonomy of Spatial Panel Model Specifications 636
19.3.1 Temporal Heterogeneity 637
19.3.2 Spatial Heterogeneity 639
19.3.3 Spatio-Temporal Models 644
19.4 Estimation of Spatial Panel Models 648
19.4.1 Maximum Likelihood Estimation 648
19.4.2 Instrumental Variables and GMM 652
19.5 Testing for Spatial Dependence 654
19.5.1 Lagrange Multiplier Tests for Spatial Lag and Spatial Error Dependence in Pooled Models 655
19.5.2 Testing for Spatial Error Correlation in Panel Data Models 655
19.6 Conclusions 656
References 657

Part III Applications

20 Foreign Direct Investment: Lessons from Panel Data 663
Pierre Blanchard, Carl Gaigné and Claude Mathieu
20.1 Introduction 663
20.2 A Simple Model of FDI 664
20.2.1 Assumptions and Preliminary Results 665

20.2.2 Technology and Country Characteristics as Determinants of FDI . . . 666
20.3 Econometric Implementation and Data . . . 668
20.3.1 A General Econometric Model . . . 669
20.3.2 FDI and Data Issues . . . 670
20.4 Empirical Estimations: Selected Applications . . . 672
20.4.1 Testing the Trade-Off Between FDI and Exports . . . 672
20.4.2 Testing the Role of Trade Policy in FDI . . . 677
20.4.3 Testing the Relationship Between FDI and Exchange Rate . . . 683
20.5 Some Recent Econometric Issues . . . 690
20.5.1 FDI, Panel Data and Spatial Econometrics . . . 690
20.5.2 Exchange Rate, Unit Roots and Cointegration . . . 691
References . . . 693

21 Stochastic Frontier Analysis and Efficiency Estimation . . . 697
Christopher Cornwell and Peter Schmidt
21.1 Measurement of Firm Efficiency . . . 698
21.2 Introduction to SFA . . . 700
21.2.1 The Basic SFA Empirical Framework . . . 700
21.2.2 Stochastic vs Deterministic Frontiers . . . 700
21.2.3 Other Frontier Functions . . . 702
21.2.4 SFA with Cross-Section Data . . . 703
21.3 SFA with Panel Data . . . 704
21.3.1 Models with Time-Invariant Inefficiency . . . 704
21.3.2 Models with Time-Varying Inefficiency . . . 714
21.4 Applications . . . 718
21.4.1 Egyptian Tile Manufacturers . . . 718
21.4.2 Indonesian Rice Farmers . . . 720
21.5 Concluding Remarks . . . 723
References . . . 723

22 Econometric Analyses of Linked Employer–Employee Data . . . 727
John M. Abowd, Francis Kramarz and Simon Woodcock
22.1 Introduction . . . 727
22.2 A Prototypical Longitudinal Linked Data Set . . . 729
22.2.1 Missing Data . . . 730
22.2.2 Sampling from Linked Data . . . 732
22.3 Linear Statistical Models with Person and Firm Effects . . . 733
22.3.1 A General Specification . . . 733
22.3.2 The Pure Person and Firm Effects Specification . . . 734
22.4 Definition of Effects of Interest . . . 735
22.4.1 Person Effects and Unobservable Personal Heterogeneity . . . 735
22.4.2 Firm Effects and Unobservable Firm Heterogeneity . . . 736
22.4.3 Firm-Average Person Effect . . . 737

22.4.4 Person-Average Firm Effect 737
22.4.5 Industry Effects 738
22.4.6 Other Firm Characteristic Effects 739
22.4.7 Occupation Effects and Other Person × Firm Interactions 739
22.5 Estimation by Fixed Effects Methods 739
22.5.1 Estimation of the Fixed Effects Model by Direct Least Squares 739
22.5.2 Consistent Methods for β and γ (The Firm-Specific Returns to Seniority) 743
22.6 The Mixed Model 744
22.6.1 REML Estimation of the Mixed Model 746
22.6.2 Estimating the Fixed Effects and Realized Random Effects 747
22.6.3 Mixed Models and Correlated Random Effects Models . . 748
22.7 Models of Heterogeneity Biases in Incomplete Models 750
22.7.1 Omission of the Firm Effects 750
22.7.2 Omission of the Person Effects 751
22.7.3 Inter-industry Wage Differentials 752
22.8 Endogenous Mobility 753
22.8.1 A Generalized Linear Mixed Model 754
22.8.2 A Model of Wages, Endogenous Mobility and Participation with Person and Firm Effects 755
22.8.3 Stochastic Assumptions 756
22.9 Conclusion 758
References 758

23 Life Cycle Labor Supply and Panel Data: A Survey 761
Bertrand Koebel, François Laisney, Winfried Pohlmeier and Matthias Staat
23.1 Introduction 761
23.2 The Basic Model of Life Cycle Labor Supply 762
23.2.1 The Framework 763
23.2.2 First Specifications of the Utility Function 765
23.3 Taking Account of Uncertainty and Risk 768
23.3.1 First Developments 768
23.3.2 Recent Contributions 770
23.3.3 Empirical Results 773
23.4 Voluntary and Involuntary Non-participation 774
23.4.1 Accounting for the Participation Decision 775
23.4.2 Unemployment 778
23.5 Alternative Parameterization and Implications 779
23.6 Relaxing Separability Assumptions 783
23.6.1 Relaxing Within-Period Additive Separability 783
23.6.2 Relaxing Intertemporal Separability in Preferences 784

23.7 Conclusion 790
References 791

24 Dynamic Policy Analysis 795
Jaap H. Abbring and James J. Heckman
24.1 Introduction 795
24.2 Policy Evaluation and Treatment Effects 796
24.2.1 The Evaluation Problem 796
24.2.2 The Treatment Effect Approach 800
24.2.3 Dynamic Policy Evaluation 801
24.3 Dynamic Treatment Effects and Sequential Randomization 803
24.3.1 Dynamic Treatment Effects 803
24.3.2 Policy Evaluation and Dynamic Discrete-Choice Analysis 810
24.3.3 The Information Structure of Policies 813
24.3.4 Selection on Unobservables 815
24.4 The Event-History Approach to Policy Analysis 816
24.4.1 Treatment Effects in Duration Models 817
24.4.2 Treatment Effects in More General Event-History Models 823
24.4.3 A Structural Perspective 828
24.5 Dynamic Discrete Choice and Dynamic Treatment Effects 829
24.5.1 Semi-parametric Duration Models and Counterfactuals . . 831
24.5.2 A Sequential Structural Model with Option Values 844
24.5.3 Identification at Infinity 850
24.5.4 Comparing Reduced-Form and Structural Models 851
24.5.5 A Short Survey of Dynamic Discrete-Choice Models 853
24.6 Conclusion 857
References 857

25 Econometrics of Individual Labor Market Transitions 865
Denis Fougère and Thierry Kamionka
25.1 Introduction 865
25.2 Multi-spell Multi-state Models 867
25.2.1 General framework 867
25.2.2 Non-parametric and Parametric Estimation 872
25.2.3 Unobserved Heterogeneity 878
25.3 Markov Processes Using Discrete-Time Observations 882
25.3.1 The Time-Homogeneous Markovian Model 883
25.3.2 The Mover-Stayer Model 893
25.4 Concluding Remarks 901
References 902

26 Software Review 907
Pierre Blanchard
26.1 Introduction 907
26.2 General-Purpose Econometric Packages 908
26.2.1 EViews (v. 5.1) 908
26.2.2 LIMDEP (v. 8) with NLOGIT (v. 3) 912
26.2.3 RATS (v. 6) 916
26.2.4 SAS (v. 9.1) 920
26.2.5 Stata (v. 9) 923
26.2.6 TSP (v. 5) 927
26.3 High-Level Matrix Programming Languages 930
26.3.1 GAUSS (v. 5) 930
26.3.2 Ox (v. 3.4) 936
26.4 Performance Hints and Numerical Accuracy Evaluation 941
26.4.1 Speed Comparison 941
26.4.2 Numerical Accuracy Evaluations 944
References 949

List of Contributors

Jaap H. Abbring
Department of Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, and Tinbergen Institute,
e-mail: jabbring@econ.vu.nl

John M. Abowd
School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14850, USA,
e-mail: john.abowd@cornell.edu

Chunrong Ai
Department of Economics, University of Florida, Warrington College of Business Administration, 224 MAT, P.O. Box 117140, Gainesville, FL 32611-7140, USA,
e-mail: chunrong.ai@cba.ufl.edu

Luc Anselin
School of Geographical Sciences, Arizona State University, Tempe, AZ 85287, USA,
e-mail: luc.anselin@asu.edu

Pietro Balestra†
Faculty of Economics, University of Lugano, via Giuseppe Buffi 13, CH-6904 Lugano, Switzerland.

Badi H. Baltagi
Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, USA,
e-mail: bbaltagi@maxwell.syr.edu

Erik Biørn
Department of Economics, University of Oslo, P.O. Box 1095, Blindern, 0317 Oslo, Norway,
e-mail: erik.biorn@econ.uio.no

Pierre Blanchard
Erudite, Faculté de Sciences Economiques et de Gestion, Université Paris XII Val de Marne, 61 Av. du Général de Gaulle, 94010 Créteil Cédex, France,
e-mail: blanchard@univ-paris12.fr

Rachid Boumahdi
Toulouse School of Economics, GREMAQ and LIHRE; Université des Sciences Sociales de Toulouse, 21 Allée de Brienne, 31000 Toulouse, France,
e-mail: boumahdi@univtlse1.fr

Jörg Breitung
University of Bonn, Institute of Econometrics, Adenauerallee 24-42, 53113 Bonn, Germany, e-mail: breitung@uni-bonn.de

Georges Bresson
ERMES (UMR 7181, CNRS), Université Paris II and TEPP (FR 3126, CNRS), Institute for Labor Studies and Public Policies, 12, place du Panthéon, 75230 Paris Cedex 05, France,
e-mail: bresson01@aol.com

Siddhartha Chib
Olin Business School, Campus Box 1133, Washington University in St. Louis, 1 Brookings Dr, St. Louis, MO 63130,
e-mail: chib@wustl.edu

Christopher Cornwell
Department of Economics, University of Georgia, Athens, GA 30602, USA,
e-mail: cornwl@terry.uga.edu

Bruno Crépon
CREST-INSEE (Paris), CEPR (London) and IZA (Bonn), CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France,
e-mail: crepon@ensae.fr

Jean-Pierre Florens
Toulouse School of Economics, Institut Universitaire de France, Université des Sciences Sociales de Toulouse, 21 Allée de Brienne, 31000 Toulouse, France,
e-mail: florens@cict.fr

Denis Fougère
CREST-INSEE (Paris), CEPR (London) and IZA (Bonn), CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France,
e-mail: fougere@ensae.fr

Carl Gaigné
INRA, UMR1302, SMART, F-35000 Rennes, France,
e-mail: gaigne@rennes.inra.fr

Mark N. Harris
Department of Econometrics & Business Statistics, Monash University, Victoria 3800, Australia,
e-mail: mark.harris@buseco.monash.edu.au

James J. Heckman
Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, USA; American Bar Foundation and Geary Institute, University College Dublin,
e-mail: jjh@uchicago.edu

Bo Honoré
Department of Economics, Princeton University, Princeton, NJ 08544-1021, USA,
e-mail: honore@Princeton.EDU

Cheng Hsiao
University of Southern California and Nanyang Technological University, Department of Economics, University of Southern California, University Park, Los Angeles, California 90089, USA,
e-mail: chsiao@usc.edu

Hubert Jayet
EQUIPPE, University of Science and Technology of Lille, Faculty of Economics and Social Sciences, 59655 Villeneuve d'Ascq Cedex, France,
e-mail: Hubert.Jayet@univ-lille1.fr

Thierry Kamionka
CNRS and CREST-INSEE (Paris), CREST-INSEE 15, Boulevard Gabriel Péri, 92245 Malakoff Cedex, France,
e-mail: kamionka@ensae.fr

Bertrand Koebel
BETA, Université Louis Pasteur, Strasbourg I and IZA, Bonn, 61 Avenue de la Forêt Noire, F67000 Strasbourg, France,
e-mail: koebel@cournot.u-strasbg.fr

Francis Kramarz
CREST-INSEE (Paris), CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France,
e-mail: kramarz@ensae.fr

Jayalakshmi Krishnakumar
Department of Econometrics, University of Geneva, 40 Bd. du Pont d'Ave, CH-1211 Geneva 4, Switzerland,
e-mail: jaya.krishnakumar@metri.unige.ch

François Laisney
BETA, Université Louis Pasteur, Strasbourg I and ZEW, Mannheim, 61 Avenue de la Forêt Noire, F67000 Strasbourg, France,
e-mail: fla@cournot.u-strasbg.fr

Michael Lechner
Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbühlstr. 14, CH-9000 St. Gallen, Switzerland,
e-mail: Michael.Lechner@unisg.ch

Julie Le Gallo
CRESE, Université de Franche-Comté, 45D Avenue de l'Observatoire, 25030 Besançon Cedex, France,
e-mail: jlegallo@univ-fcomte.fr

Qi Li
Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA,
e-mail: qi@econmail.tamu.edu

Roman Liesenfeld
Department of Economics, Christian-Albrechts-Universität Kiel, Olhshausenstr. 40-60, 24118 Kiel, Germany,
e-mail: liesenfeld@stat-econ.uni-kiel.de

Stéfan Lollivier
INSEE, 18 boulevard Adolphe Pinard, F-75014 Paris, France,
e-mail: stefan.lollivier@insee.fr

Thierry Magnac
Université de Toulouse 1, Toulouse School of Economics, Manufacture des Tabacs, 21, Allée de Brienne, 31000 Toulouse, France,
e-mail: magnac@cict.fr

Jacques Mairesse
CREST-INSEE (Paris), UNU-MERIT (University of Maastricht), and NBER. CREST-INSEE, 15 boulevard Gabriel PERI, 92245 MALAKOFF cedex, France,
e-mail: mairesse@ensae.fr

Claude Mathieu
Erudite, Faculté de Sciences Economiques et de Gestion, Université Paris XII Val de Marne, 61 Av. du Général de Gaulle, 94010 Créteil Cédex, France,
e-mail: mathieu@univ-paris12.fr

László Mátyás
Central European University, Department of Economics, Nádor u. 9, 1051 Budapest, Hungary,
e-mail: matyas@ceu.hu

Michel Mouchart
Emeritus Professor of Statistics and Econometrics, Institut de statistique, 20 Voie du Roman Pays, B-1348 Louvain-La-Neuve (Belgium),
e-mail: Michel.Mouchart@uclouvain.be

Marc Nerlove
Department of Agricultural and Resource Economics, University of Maryland, 2200 Symons Hall, College Park, MD 20742-5535, USA,
e-mail: mnerlove@arec.umd.edu

Hashem Pesaran
Cambridge University and USC, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom,
e-mail: mhp1@econ.cam.ac.uk

Alain Pirotte
ERMES (UMR 7181, CNRS), Université Paris II and TEPP (FR 3126, CNRS), Institute for Labor Studies and Public Policies, 12, place du Panthéon, 75230 Paris Cedex 05, France,
e-mail: apirotte@aol.com

Winfried Pohlmeier
University of Konstanz, Department of Economics, Box D124, D78457 Konstanz, Germany,
e-mail: winfried.pohlmeier@uni-konstanz.de

Jean-François Richard
Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, USA,
e-mail: fantin@pitt.edu

Peter Schmidt
Department of Economics, Michigan State University, East Lansing, MI 48824, USA,
e-mail: schmidtp@msu.edu

Patrick Sevestre
Université Paris 1 – Panthéon Sorbonne, Ecole Economique de Paris (Paris School of Economics), 106-112 Boulevard de l'Hôpital, 75013 Paris, France,
e-mail: sevestre@univ-paris1.fr

Matthias Staat
University of Mannheim, Department of Economics, D68131 Mannheim, Germany,
e-mail: matthias@pool.uni-mannheim.de

Alban Thomas
Toulouse School of Economics, INRA; Université des Sciences Sociales de Toulouse, 21 Allée de Brienne, 31000 Toulouse, France,
e-mail: thomas@toulouse.inra.fr

Francis Vella
Department of Economics, Georgetown University, Washington DC, USA,
e-mail: fgv@georgetown.edu

Marno Verbeek
Department of Financial Management, RSM Erasmus University, Burg. Oudlaan 50, 3062 PA Rotterdam, The Netherlands,
e-mail: mverbeek@rsm.nl

Frank Windmeijer
Department of Economics, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK,
e-mail: f.windmeijer@bristol.ac.uk

Simon Woodcock
Department of Economics, Simon Fraser University, Burnaby, BC, Canada,
e-mail: simon_woodcock@sfu.ca

Part I
Fundamentals

Chapter 1
Introduction

Marc Nerlove, Patrick Sevestre and Pietro Balestra

1.1 Introduction

In his famous and influential monograph, *The Probability Approach in Econometrics*, Haavelmo (1944) laid the foundations for the formulation of *stochastic* econometric models and an approach which has dominated our discipline to this day. He wrote:

> … we shall find that two individuals, or the same individual in two different time periods, may be confronted with exactly the same set of specified influencing factors [and, hence, they have the same y^*, …], and still the two individuals may have different quantities y, neither of which may be equal to y^*. We may try to remove such discrepancies by introducing more "explaining" factors, x. But, usually, we shall soon exhaust the number of factors which could be considered as *common* to all individuals, and which, at the same time, were not merely of negligible influence upon y. The discrepancies $y - y^*$ for each individual may depend upon a great variety of factors, these factors may be different from one individual to another, and they may vary with time for each individual (Haavelmo, 1944, p. 50).

And further that:

> …we find justification for applying them [stochastic approximations] to economic phenomena also in the fact we usually deal only with—and are interested only in—total or average effects of many individual decisions, which are partly guided by common factors, partly by individual specific factors… (Haavelmo, 1944, pp. 51 and 56)

Marc Nerlove
Department of Agricultural and Resource Economics, University of Maryland, 2200 Symons Hall, College Park, MD 20742-5535, USA, e-mail: mnerlove@arec.umd.edu

Patrick Sevestre
Université Paris 1 – Panthéon Sorbonne, Ecole Economique de Paris (Paris School of Economics), 106-112 Boulevard de l'Hôpital, 75013 Paris, France, e-mail: sevestre@univ-paris1.fr

Pietro Balestra
Faculty of Economics, University of Lugano, via Giuseppe Buffi 13, CH-6904 Lugano, Switzerland

This introductory chapter is dedicated to the memory of our late colleague, Pietro Balestra. This chapter largely draws on his and Nerlove's earlier introduction to the second edition of this volume.

L. Mátyás, P. Sevestre (eds.), *The Econometrics of Panel Data*,

Marschak (1950, 1953) further amplified Haavelmo's themes in his introduction to Cowles Commission Monographs 10 and 14, observing that:

> The numerous causes that determine the error incurred ... are not listed separately; instead their joint effect is represented by the probability distribution of the error, a random variable (1950, p. 18) [which] ... is called 'disturbance' or 'shock,' and can be regarded as the joint effect of numerous separately insignificant variables that we are unable or unwilling to specify but presume to be independent of observable exogenous variables. (1953, p. 12).

In this introduction we examine how the basic principle underlying the formulation of econometric models has been carried forward in the development of econometric models and methods for the analysis of panel data. We argue that while fixed effects models may be appropriate in cases in which a population is sampled exhaustively (e.g., data from geographic regions over time) or in which it is desired to predict individual behavior (e.g., the probability that a given individual in a sample will default on a loan), random effects models are more consistent with Haavelmo's view, quoted above, that the "population" we model in econometrics consists not of an infinity of individuals, in general; but of an infinity of decisions. This is not to say, however, that fixed effects models may not be extremely useful as an analytic device.

Moreover, we shall argue, taking a leaf from Knight (1921), that what differentiates the individuals, who make the decisions with which we are concerned, is largely historical, the "three great accumulating funds of inheritance from the past, material goods and appliances, knowledge and skill, and morale." This view has important implications for the relevance and appropriateness of many of the models and methods for the analysis of panel data which have been developed over the past 40 years. We briefly review these developments here and conclude that not only are random effects models most relevant and appropriate but that often our central analytical and modelling concerns are also dynamic. Thus, the most fruitful developments in this enormous literature have been those which deal with the central issues of history and dynamics.

1.2 Data, Data-Generating Processes (DGP), and Inference

In most applications of statistical analysis in the so-called "hard" sciences, the process by which the observed data are generated is transparent, having usually been determined by the investigator by design. In contrast, in many applications in the social sciences, especially in economics, the mechanism by which the data are generated is opaque. In such circumstances, estimation of the parameters of the statistical or econometric model and the testing of specific hypotheses about it are only half the problem of inference. Understanding the process by which the observations at hand are generated is of equal importance. Were the data for example obtained from a sample of firms selected by stratified random sampling from a census of all firms in the United States in 2000? For example, were they obtained from regulatory activity? In the case of time series, the data are almost always "fabricated" in one way or another, by aggregation, interpolation, or extrapolation, or by all three. The nature of the sampling frame or the way in which the data are fabricated must be part of the model specification on which parametric inference or hypothesis testing is based. Nonparametric inference imposes fewer restrictions on the specification of the DGP but incorporation of knowledge about the nature of the DGP is of equal importance. Almost all the methodological papers in Parts 1 and 2 of this volume focus primarily on problems of estimation and inference from a parametrically well-specified model of how the observed data were generated and variously draw their interest from different types of data and their DGPs. In this section we address the issue of why the DGP matters in the context of a specific, although somewhat abstract, example.

Suppose a longitudinal household survey in which the same households are questioned over time about their actions in, say, a number of consecutive months or years and, initially, about various demographic and economic characteristics. These households differ in various ways, some of which we observe and many which we do not. Some of these differences are the result of their past behavior or past circumstances (path dependence), some are differences in tastes or other unobserved characteristics which may be assumed to be permanent (individual heterogeneity), and some are due to peculiarities not permanently associated with time or individual.[1]

What, in the context of these data, can be considered as random, what is the population from which we may consider the data a sample, and what is a parameter, and what a random variable? These issues are central to an understanding of the DGP.

Statistical and, a fortiori, econometric analysis, are usually based on the idea of sampling from a population in order to draw inferences for the underlying population. But what is the population from which economic data may be supposed to be a sample? In his famous 1944 monograph, Haavelmo (1944, p. 56) wrote, "… the class of populations we are dealing with does not consist of an infinity of different individuals, it consists of an infinity of possible decisions which might be taken …". In their recent text, *Econometric Theory and Methods*, Davidson and MacKinnon (2004, pp. 30–31) make the same point: "In econometrics, the use of the term population is simply a metaphor. A better concept is that of a data-generating process, or DGP. By this term, we mean whatever mechanism is at work in the real world of economic activity giving rise to the numbers in our samples, that is, precisely the mechanism that our econometric model is supposed to describe. A DGP is thus the analog of a population in biostatistics. Samples may be drawn from a DGP just as they may be drawn from a population. In both cases, the samples are assumed to be representative of the DGP or population from which they are drawn."

What is a random variable in this context and what is not? Whether or not a particular variable can be considered a random draw from some population or not, in principle can be decided by applying the principle of "exchangeability" introduced by de Finetti (1930). In a nutshell, the idea, very Bayesian in flavor, is to ask whether we can exchange two elements in a sample and still maintain the same subjective distribution. Thus, in a panel study of households, are any two households in the sample exchangeable without affecting the distribution, from which we imagine

[1] In his paper, "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity", Heckman (1991) argues that in general it is not possible to distinguish. The ability to do so rests critically "on maintaining explicit assumptions about the way in which observables and unobservables interact."

household observables and unobservables to be drawn? In a panel of state data, are California and Maryland exchangeable without affecting the subjective distribution of the state effects? It's a dicey question – sometimes.

From the standpoint of a Bayesian there is no real distinction between a parameter and a random variable, but in this context we could say that a parameter is an unobserved variable which affects the distribution of the random variables of the model and is unaffected by the particular values such variables take on. It is what we wish to estimate and about which we wish to make inferences. A related concept is that of an exogenous variable. But note here that such an exogenous variable is still a random variable and not a parameter.

In general, in the formulation of econometric models (i.e., the DGP for the process yielding the particular set of data we want to "explain") the distinction between what can be observed and what is not is fundamental. Linear functions are often used to describe such a DGP. To get more precisely to the issues posed by the formulation of the DGP for a sample of economic data, we need to include several observable variables. Suppose that we draw a random sample of N individuals over T time periods; for example a household survey in which we collect observations on the income, x_{it} and consumption of household i, y_{it}, for many households N, in year t over a brief period T. From the survey we have observations on the pairs (x_{it}, y_{it}). Since the households are chosen at random for the survey, but the years over which they are observed are not, the lists $(x_{i1}, y_{i1}, \ldots, x_{iT}, y_{iT})$, $i = 1, \ldots, N$, are exchangeable, but the order within each list is not.

Imagine we are estimating a consumption function and assume a linear relationship subject to error:

$$y_{it} = a + b\,x_{it} + \varepsilon_{it} \tag{1.1}$$

This would be the case, if for example, the joint distribution of variables could be assumed normal and we were trying to estimate the mean of y_{it} for a particular year t conditional on x_{it}. We might then write ε_{it} as

$$\varepsilon_{it} = \mu_i + \lambda_t + u_{it} \tag{1.2}$$

where ε_{it} is an unobserved random variable which is the sum of three effects, all of which are also unobserved: λ_t is a year effect, arguably nonrandom and therefore a parameter to be estimated for each year, t; μ_i is a household effect, which, in view of the way the observations are drawn, should surely be treated as random, and, finally, u_{it} is a random variable to represent all the rest.

We are far from done yet, however. The question remains as to what we should assume about the observable variables, x_{it}. They are clearly random variables jointly distributed with the variable y_{it}. If not subject to errors of measurement, an assumption difficult to justify in the context of an economic survey, are they also independent of, or at least uncorrelated with, the disturbances ε_{it} in (1.1)? This question clearly affects not only what we can say about the DGP which generates our observations, but also how many and what parameters must be considered. Let us examine the regression with some care. Since λ_t is not a random variable but a parameter, consider it to be a constant for each t and add it to the constant a in the regression equation (1.1):