

Analysis of Panel Data

Panel data models have become increasingly popular among applied researchers due to their heightened capacity for capturing the complexity of human behavior as compared to cross-sectional or time-series data models. As a consequence, more and richer panel data sets also have become increasingly available. This second edition is a substantial revision of the highly successful first edition of 1986. Recent advances in panel data research are presented in a rigorous and accessible manner and are carefully integrated with the older material. The thorough discussion of theory and the judicious use of empirical examples will make this book useful to graduate students and advanced researchers in economics, business, sociology, political science, etc. Other specific revisions include the introduction of Bayes method and the notion of strict exogeneity with estimators presented in a generalized method of moments framework to link the identification of various models, intuitive explanations of semiparametric methods of estimating discrete choice models and methods of pairwise trimming for the estimation of panel sample selection models, etc.

Cheng Hsiao is Professor of Economics at the University of Southern California. His first edition of *Analysis of Panel Data* (Econometric Society Monograph 11, Cambridge University Press, 1986) has been the standard introduction for students to the subject in the economics literature. Professor Hsiao is also a coauthor of *Econometric Models, Techniques, and Applications*, Second Edition (Prentice Hall, 1996, with M. Intriligator and R. Bodkin), coeditor of *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, 1999, with K. Lahiri, L.F. Lee, and M.H. Pesaran), and coeditor of *Nonlinear Statistical Inference* (Cambridge University Press, 2001, with K. Morimune and J.L. Powell). He is a Fellow of the Econometric Society and coeditor and Fellow of the *Journal of Econometrics*.

Econometric Society Monographs No. 34

Editors:

Andrew Chesher, University College London
 Matthew O. Jackson, California Institute of Technology

The Econometric Society is an international society for the advancement of economic theory in relation to statistics and mathematics. The Econometric Society Monograph Series is designed to promote the publication of original research contributions of high quality in mathematical economics and theoretical and applied econometrics.

Other titles in the series:

- G. S. Maddala *Limited-dependent and qualitative variables in econometrics*, 0 521 33825 5
 Gerard Debreu *Mathematical economics: Twenty papers of Gerard Debreu*, 0 521 33561 2
 Jean-Michel Grandmont *Money and value: A reconsideration of classical and neoclassical monetary economics*, 0 521 31364 3
 Franklin M. Fisher *Disequilibrium foundations of equilibrium economics*, 0 521 37856 7
 Andreu Mas-Colell *The theory of general economic equilibrium: A differentiable approach*, 0 521 26514 2, 0 521 38870 8
 Truman F. Bewley, Editor *Advances in econometrics – Fifth World Congress (Volume I)*, 0 521 46726 8
 Truman F. Bewley, Editor *Advances in econometrics – Fifth World Congress (Volume II)*, 0 521 46725 X
 Hervé Moulin *Axioms of cooperative decision making*, 0 521 36055 2, 0 521 42458 5
 L. G. Godfrey *Misspecification tests in econometrics: The Lagrange multiplier principle and other approaches*, 0 521 42459 3
 Tony Lancaster *The econometric analysis of transition data*, 0 521 43789 X
 Alvin E. Roth and Marilda A. Oliviera Sotomayor, Editors *Two-sided matching: A study in game-theoretic modeling and analysis*, 0 521 43788 1
 Wolfgang Härdle, *Applied nonparametric regression*, 0 521 42950 1
 Jean-Jacques Laffont, Editor *Advances in economic theory – Sixth World Congress (Volume I)*, 0 521 48459 6
 Jean-Jacques Laffont, Editor *Advances in economic theory – Sixth World Congress (Volume II)*, 0 521 48460 X
 Halbert White *Estimation, inference and specification*, 0 521 25280 6, 0 521 57446 3
 Christopher Sims, Editor *Advances in econometrics – Sixth World Congress (Volume I)*, 0 521 56610 X
 Christopher Sims, Editor *Advances in econometrics – Sixth World Congress (Volume II)*, 0 521 56609 6
 Roger Guesnerie *A contribution to the pure theory of taxation*, 0 521 23689 4, 0 521 62956 X
 David M. Kreps and Kenneth F. Wallis, Editors *Advances in economics and econometrics – Seventh World Congress (Volume I)*, 0 521 58011 0, 0 521 58983 5
 David M. Kreps and Kenneth F. Wallis, Editors *Advances in economics and econometrics – Seventh World Congress (Volume II)*, 0 521 58012 9, 0 521 58982 7
 David M. Kreps and Kenneth F. Wallis, Editors *Advances in economics and econometrics – Seventh World Congress (Volume III)*, 0 521 58013 7, 0 521 58981 9
 Donald P. Jacobs, Ehud Kalai, and Morton I. Kamien, Editors *Frontiers of research in economic theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, 0 521 63222 6, 0 521 63538 1
 A. Colin Cameron and Pravin K. Trivedi, *Regression analysis of count data*, 0 521 63201 3, 0 521 63567 5
 Steinar Strøm, Editor *Econometrics and economic theory in the 20th century: The Ragnar Frisch Centennial Symposium*, 0 521 63323 0, 0 521 63365 6
 Eric Ghysels, Norman R. Swanson, and Mark Watson, Editors *Essays in econometrics: Collected papers of Clive W. J. Granger (Volume I)*, 0 521 77496 9
 Eric Ghysels, Norman R. Swanson, and Mark Watson, Editors *Essays in econometrics: Collected papers of Clive W. J. Granger (Volume II)*, 0 521 79649 0

Cambridge University Press
0521818559 - Analysis of Panel Data, Second Edition
Cheng Hsiao
Frontmatter
[More information](#)

Analysis of Panel Data

Second Edition

CHENG HSIAO

University of Southern California



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
 0521818559 - Analysis of Panel Data, Second Edition
 Cheng Hsiao
 Frontmatter
[More information](#)

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
 The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
 The Edinburgh Building, Cambridge CB2 2RU, UK
 40 West 20th Street, New York, NY 10011-4211, USA
 477 Williamstown Road, Port Melbourne, VIC 3207, Australia
 Ruiz de Alarcón 13, 28014 Madrid, Spain
 Dock House, The Waterfront, Cape Town 8001, South Africa
<http://www.cambridge.org>

© Cheng Hsiao 2003

This book is in copyright. Subject to statutory exception
 and to the provisions of relevant collective licensing agreements,
 no reproduction of any part may take place without
 the written permission of Cambridge University Press.

First published 2003

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times Roman 10/12 pt. *System* L^AT_EX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Hsiao, Cheng, 1943–

Analysis of panel data / Cheng Hsiao. – 2nd ed.

p. cm. – (Econometric Society monographs ; no. 34)

Includes bibliographical references and index.

ISBN 0-521-81855-9 – ISBN 0-521-52271-4 (pb.)

1. Econometrics. 2. Panel analysis. 3. Analysis of variance. I. Title. II. Series.

HB139 .H75 2002

330'.01'5195 – dc21

2002023348

ISBN 0 521 81855 9 hardback

ISBN 0 521 52271 4 paperback

Cambridge University Press
0521818559 - Analysis of Panel Data, Second Edition
Cheng Hsiao
Frontmatter
[More information](#)

*To my wife, Amy Mei-Yun
and my children,
Irene Chiayun
Allen Chenwen
Michael Chenyee
Wendy Chiawen*

Contents

<i>Preface to the Second Edition</i>	<i>page</i> xiii
<i>Preface to the First Edition</i>	xv
Chapter 1. Introduction	1
1.1 Advantages of Panel Data	1
1.2 Issues Involved in Utilizing Panel Data	8
1.2.1 Heterogeneity Bias	8
1.2.2 Selectivity Bias	9
1.3 Outline of the Monograph	11
Chapter 2. Analysis of Covariance	14
2.1 Introduction	14
2.2 Analysis of Covariance	15
2.3 An Example	21
Chapter 3. Simple Regression with Variable Intercepts	27
3.1 Introduction	27
3.2 Fixed-Effects Models: Least-Squares Dummy-Variable Approach	30
3.3 Random-Effects Models: Estimation of Variance-Components Models	34
3.3.1 Covariance Estimation	35
3.3.2 Generalized-Least-Squares Estimation	35
3.3.3 Maximum Likelihood Estimation	39
3.4 Fixed Effects or Random Effects	41
3.4.1 An Example	41
3.4.2 Conditional Inference or Unconditional (Marginal) Inference	43
3.4.2.a Mundlak's Formulation	44
3.4.2.b Conditional and Unconditional Inferences in the Presence or Absence of Correlation between Individual Effects and Attributes	46

viii	Contents	
	3.5 Tests for Misspecification	49
	3.6 Models with Specific Variables and Both Individual- and Time-Specific Effects	51
	3.6.1 Estimation of Models with Individual-Specific Variables	51
	3.6.2 Estimation of Models with Both Individual and Time Effects	53
	3.7 Heteroscedasticity	55
	3.8 Models with Serially Correlated Errors	57
	3.9 Models with Arbitrary Error Structure – Chamberlain π Approach	60
	Appendix 3A: Consistency and Asymptotic Normality of the Minimum-Distance Estimator	65
	Appendix 3B: Characteristic Vectors and the Inverse of the Variance–Covariance Matrix of a Three-Component Model	67
	Chapter 4. Dynamic Models with Variable Intercepts	69
	4.1 Introduction	69
	4.2 The Covariance Estimator	71
	4.3 Random-Effects Models	73
	4.3.1 Bias in the OLS Estimator	73
	4.3.2 Model Formulation	75
	4.3.3 Estimation of Random-Effects Models	78
	4.3.3.a Maximum Likelihood Estimator	78
	4.3.3.b Generalized-Least-Squares Estimator	84
	4.3.3.c Instrumental-Variable Estimator	85
	4.3.3.d Generalized Method of Moments Estimator	86
	4.3.4 Testing Some Maintained Hypotheses on Initial Conditions	90
	4.3.5 Simulation Evidence	91
	4.4 An Example	92
	4.5 Fixed-Effects Models	95
	4.5.1 Transformed Likelihood Approach	96
	4.5.2 Minimum-Distance Estimator	98
	4.5.3 Relations between the Likelihood-Based Estimator and the Generalized Method of Moments Estimator (GMM)	99
	4.5.4 Random- versus Fixed-Effects Specification	101
	4.6 Estimation of Dynamic Models with Arbitrary Correlations in the Residuals	103
	4.7 Fixed-Effects Vector Autoregressive Models	105
	4.7.1 Model Formulation	105
	4.7.2 Generalized Method of Moments (GMM) Estimation	107

Contents	ix
4.7.3 (Transformed) Maximum Likelihood Estimator	109
4.7.4 Minimum-Distance Estimator	109
Appendix 4A: Derivation of the Asymptotic Covariance Matrix of the Feasible MDE	111
Chapter 5. Simultaneous-Equations Models	113
5.1 Introduction	113
5.2 Joint Generalized-Least-Squares Estimation Technique	116
5.3 Estimation of Structural Equations	119
5.3.1 Estimation of a Single Equation in the Structural Model	119
5.3.2 Estimation of the Complete Structural System	124
5.4 Triangular System	127
5.4.1 Identification	127
5.4.2 Estimation	129
5.4.2.a Instrumental-Variable Method	130
5.4.2.b Maximum-Likelihood Method	133
5.4.3 An Example	136
Appendix 5A	138
Chapter 6. Variable-Coefficient Models	141
6.1 Introduction	141
6.2 Coefficients That Vary over Cross-Sectional Units	143
6.2.1 Fixed-Coefficient Model	144
6.2.2 Random-Coefficient Model	144
6.2.2.a The Model	144
6.2.2.b Estimation	145
6.2.2.c Predicting Individual Coefficients	147
6.2.2.d Testing for Coefficient Variation	147
6.2.2.e Fixed or Random Coefficients	149
6.2.2.f An Example	150
6.3 Coefficients That Vary over Time and Cross-Sectional Units	151
6.3.1 The Model	151
6.3.2 Fixed-Coefficient Model	153
6.3.3 Random-Coefficient Model	153
6.4 Coefficients That Evolve over Time	156
6.4.1 The Model	156
6.4.2 Predicting β_t by the Kalman Filter	158
6.4.3 Maximum Likelihood Estimation	161
6.4.4 Tests for Parameter Constancy	162
6.5 Coefficients That Are Functions of Other Exogenous Variables	163
6.6 A Mixed Fixed- and Random-Coefficients Model	165
6.6.1 Model Formulation	165
6.6.2 A Bayes Solution	168
6.6.3 An Example	170

x	Contents	
	6.6.4 Random or Fixed Parameters	172
	6.6.4.a An Example	172
	6.6.4.b Model Selection	173
	6.7 Dynamic Random-Coefficient Models	175
	6.8 An Example – Liquidity Constraints and Firm Investment Expenditure	180
	Appendix 6A: Combination of Two Normal Distributions	185
	Chapter 7. Discrete Data	188
	7.1 Introduction	188
	7.2 Some Discrete-Response Models	188
	7.3 Parametric Approach to Static Models with Heterogeneity	193
	7.3.1 Fixed-Effects Models	194
	7.3.1.a Maximum Likelihood Estimator	194
	7.3.1.b Conditions for the Existence of a Consistent Estimator	195
	7.3.1.c Some Monte Carlo Evidence	198
	7.3.2 Random-Effects Models	199
	7.4 Semiparametric Approach to Static Models	202
	7.4.1 Maximum Score Estimator	203
	7.4.2 A Root- N Consistent Semiparametric Estimator	205
	7.5 Dynamic Models	206
	7.5.1 The General Model	206
	7.5.2 Initial Conditions	208
	7.5.3 A Conditional Approach	211
	7.5.4 State Dependence versus Heterogeneity	216
	7.5.5 Two Examples	218
	7.5.5.a Female Employment	218
	7.5.5.b Household Brand Choices	221
	Chapter 8. Truncated and Censored Data	225
	8.1 Introduction	225
	8.2 An Example – Nonrandomly Missing Data	234
	8.2.1 Introduction	234
	8.2.2 A Probability Model of Attrition and Selection Bias	235
	8.2.3 Attrition in the Gary Income-Maintenance Experiment	238
	8.3 Tobit Models with Random Individual Effects	240
	8.4 Fixed-Effects Estimator	243
	8.4.1 Pairwise Trimmed Least-Squares and Least-Absolute-Deviation Estimators for Truncated and Censored Regressions	243
	8.4.1.a Truncated Regression	243
	8.4.1.b Censored Regressions	249
	8.4.2 A Semiparametric Two-Step Estimator for the Endogenously Determined Sample Selection Model	253

Contents	xi
8.5 An Example: Housing Expenditure	255
8.6 Dynamic Tobit Models	259
8.6.1 Dynamic Censored Models	259
8.6.2 Dynamic Sample Selection Models	265
Chapter 9. Incomplete Panel Data	268
9.1 Estimating Distributed Lags in Short Panels	268
9.1.1 Introduction	268
9.1.2 Common Assumptions	270
9.1.3 Identification Using Prior Structure of the Process of the Exogenous Variable	271
9.1.4 Identification Using Prior Structure of the Lag Coefficients	275
9.1.5 Estimation and Testing	277
9.2 Rotating or Randomly Missing Data	279
9.3 Pseudopanel (or Repeated Cross-Sectional Data)	283
9.4 Pooling of a Single Cross-Sectional and a Single Time-Series Data Set	285
9.4.1 Introduction	285
9.4.2 The Likelihood Approach to Pooling Cross-Sectional and Time-Series Data	287
9.4.3 An Example	288
Chapter 10. Miscellaneous Topics	291
10.1 Simulation Methods	291
10.2 Panels with Large N and T	295
10.3 Unit-Root Tests	298
10.4 Data with Multilevel Structures	302
10.5 Errors of Measurement	304
10.6 Modeling Cross-Sectional Dependence	309
Chapter 11. A Summary View	311
11.1 Introduction	311
11.2 Benefits and Limitations of Panel Data	311
11.2.1 Increasing Degrees of Freedom and Lessening the Problem of Multicollinearity	311
11.2.2 Identification and Discrimination between Competing Hypotheses	312
11.2.3 Reducing Estimation Bias	313
11.2.3.a Omitted-Variable Bias	313
11.2.3.b Bias Induced by the Dynamic Structure of a Model	315
11.2.3.c Simultaneity Bias	316
11.2.3.d Bias Induced by Measurement Errors	316

xii **Contents**

11.2.4 Providing Micro Foundations for Aggregate Data Analysis	316
11.3 Efficiency of the Estimates	317
<i>Notes</i>	319
<i>References</i>	331
<i>Author Index</i>	353
<i>Subject Index</i>	359

Preface to the Second Edition

Since the publication of the first edition of this monograph in 1986, there has been a phenomenal growth of articles dealing with panel data. According to the *Social Science Citation Index*, there were 29 articles related to panel data in 1989. But in 1997 there were 518; in 1998, 553; and in 1999, 650. The increasing attention is partly due to the greater availability of panel data sets, which can better answer questions of substantial interest than a single set of cross-section or time series data can, and partly due to the rapid growth in computational power of the individual researcher. It is furthermore motivated by the internal methodological logic of the subject (e.g., Trognon (2000)).

The current version is a substantial revision of the first edition. The major additions are essentially on nonlinear panel data models of discrete choice (Chapter 7) and sample selection (Chapter 8); a new Chapter 10 on miscellaneous topics such as simulation techniques, large N and T theory, unit root and cointegration tests, multiple level structure, and cross-sectional dependence; and new sections on estimation of dynamic models (4.5–4.7), Bayesian treatment of models with fixed and random coefficients (6.6–6.8), and repeated cross-sectional data (or pseudopanel), etc. In addition, many of the discussions in old chapters have been updated. For instance, the notion of strict exogeneity is introduced, and estimators are also presented in a generalized method of moments framework to help link the assumptions that are required for the identification of various models. The discussion of fixed and random effects is updated in regard to restrictions on the assumption about unobserved specific effects, etc.

The goal of this revision remains the same as that of the first edition. It aims to bring up to date a comprehensive analytical framework for the analysis of a greater variety of data. The emphasis is on formulating appropriate statistical inference for issues shaped by important policy concerns. The revised edition of this monograph is intended neither as an encyclopedia nor as a history of panel data econometrics. I apologize for the omissions of many important contributions. A recount of the history of panel data econometrics can be found in Nerlove (2000). Some additional issues and references can also be found in a survey by Arellano and Honoré (2001) and in four recent

edited volumes – Matyás and Sevester (1996); Hsiao, Lahiri, Lee, and Pesaran (1999); Hsiao, Morimune, and Powell (2001); and Krishnakumar and Ronchetti (2000). Software is reviewed by Blanchard (1996).

I would like to thank the editor, Scott Parris, for his encouragement and assistance in preparing the revision, and Andrew Chesher and two anonymous readers for helpful comments on an early draft. I am also very grateful to E. Kyriazidou for her careful and detailed comments on Chapters 7 and 8, S. Chen and J. Powell for their helpful comments and suggestions on Chapter 8, and H.R. Moon for the section on large panels, Sena Schlessinger for her expert typing of the manuscript except for Chapter 7, Yan Shen for carefully proofreading the manuscript and for expertly typing Chapter 7, and Siyan Wang for drawing the figures for Chapter 8. Of course, all remaining errors are mine. The kind permissions to reproduce parts of articles by James Heckman, C. Manski, Daniel McFadden, Ariel Pakes, *Econometrica*, *Journal of the American Statistical Association*, *Journal of Econometrics*, *Regional Science and Urban Economics*, *Review of Economic Studies*, The University of Chicago Press, and Elsevier Science are also gratefully acknowledged.

Preface to the First Edition

Recently, empirical research in economics has been enriched by the availability of a wealth of new sources of data: cross sections of individuals observed over time. These allow us to construct and test more realistic behavioral models that could not be identified using only a cross section or a single time series data set. Nevertheless, the availability of new data sources raises new issues. New methods are constantly being introduced, and points of view are changing. An author preparing an introductory monograph has to select the topics to be included. My selection involves controlling for unobserved individual and/or time characteristics to avoid specification bias and to improve the efficiency of the estimates. The more basic and more commonly used methods are treated here, although to some extent the coverage is a matter of taste. Some examples of applications of the methods are also given, and the uses, computational approaches, and interpretations are discussed.

I am very much indebted to C. Manski and to a reader for Cambridge University Press, as well as to G. Chamberlain and J. Ham, for helpful comments and suggestions. I am also grateful to Mario Tello Pacheco, who read through the manuscript and made numerous suggestions concerning matters of exposition and corrections of errors of every magnitude. My appreciation also goes to V. Bencivenga, A.C. Cameron, T. Crawley, A. Deaton, E. Kuh, B. Ma, D. McFadden, D. Mountain, G. Solon, G. Taylor, and K.Y. Tsui, for helpful comments, and Sophia Knapik and Jennifer Johnson, who patiently typed and retyped innumerable drafts and revisions. Of course, in material like this it is easy to generate errors, and the reader should put the blame on the author for any remaining errors.

Various parts of this monograph were written while I was associated with Bell Laboratories, Murray Hill, Princeton University, Stanford University, the University of Southern California, and the University of Toronto. I am grateful to these institutions for providing me with secretarial and research facilities and, most of all, stimulating colleagues. Financial support from the National Science Foundation, U.S.A., and from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.