

Analysis of Panel Data, Fourth Edition

Now in its fourth edition, this comprehensive introduction to fundamental panel data methodologies provides insights on what is most essential in panel literature. A capstone to the 40-year career of a pioneer of panel data analysis, this new edition's primary contribution will be the coverage of advancements in panel data analysis, a statistical method widely used to analyze two- or higher-dimensional panel data. The topics discussed in early editions have been reorganized and streamlined to comprehensively introduce panel econometric methodologies useful for identifying causal relationships among variables, supported by interdisciplinary examples and case studies. This book, to be featured in Cambridge's Econometric Society Monographs series, has been the leader in the field since the first edition. It is essential reading for researchers, practitioners, and graduate students interested in the analysis of microeconomic behavior.

Cheng Hsiao is Professor of Economics at the University of Southern California.





Econometric Society Monographs Series

Editors:

Andrea Prat, Columbia University Stéphane Bonhomme, University of Chicago

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.

Books in the Series

- C. Hsiao, Analysis of Panel Data: 4th edn, 2022
- D. M. Kreps, The Black-Scholes-Merton Model as an Idealization of Discrete-Time Economies, 2019
- W. Thomson, How to Divide When There Isn't Enough: From Aristotle, the Talmud, and Maimonides to the Axiomatics of Resource Allocation, 2019
- O. Compte & A. Postlewaite, Ignorance and Uncertainty, 2018
- I. Molchanov & F. Molinari, Random Sets in Econometrics, 2018
- B. Honoré, A. Pakes, M. Piazzesi, & L. Samuelson (eds.), Advances in Economics and Econometrics: Eleventh World Congress, Vols. I & II, 2017
- S. Maurer, On the Shoulders of Giants: Colleagues Remember Suzanne Scotchmer's Contributions to Economics, 2017
- C. P. Chambers & F. Echenique, Revealed Preference Theory, 2016
- J.-F. Mertens, S. Sorins, & S. Samir, Repeated Games, 2015
- C. Hsiao, Analysis of Panel Data: 3rd edn, 2014
- C. Cameron & P. Trivedi, Regression Analysis of Count Data, 2nd edn, 2013
- A. Harvey, Dynamic Models for Volatility and Heavy Tails, with Applications to Financial and Economic Time Series, 2013
- D. Acemoglu, M. Areilano, & E. Dekel (eds.), Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Vols. I, II, & III, 2013
- M. Fleurbaey & F. Maniquet, A Theory of Fairness and Social Justice, 2011
- R. Vohra, Mechanism Design: A Linear Programming Approach, 2011
- K. Samphantharak & R. Townsend, Households as Corporate Firms: An Analysis of Household Finance Using Integrated Household Surveys and Corporate Financial Accounting, 2009
- I. Gilboa, Theory of Decision under Uncertainty, 2009
- F. Vega-Redondo, Complex Networks, 2007
- R. Blundell, W. Newey, & T. Persson, (eds.), Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Vols. I, II, & III, 2006
- J. Roemer, Democracy, Education, and Equality, 2006
- C. Blackorby, W. Bossert, & D. Donaldson, Population Issues in Social Choice Theory, Welfare Economics and Ethics, 2005
- R. Koenker, Quantile Regression, 2005
- C. Hsiao, Analysis of Panel Data, 2nd ed., 2003
- M. Dewatripont, L. P. Hausen, & S. J. Turnovsky (eds.), Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vols. I, II, & III, 2003
- E. Ghysels, N. Swanson, & M. Watson (eds.), Essays in Econometrics: Collected Papers of Clive W. J. Granger, Vols. 1 &II, 2001
- S. Strøm (ed.), Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium, 1999
- A. C. Cameron & P. K. Trivedi, Regression Analysis of Count-Data, 1998
- D. Jacobs, E. Kalai, & M. Kamien (eds.), Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1998
- D. M. Kreps & K. F. Wallis (eds.), Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Vols. I, II, & III, 1997

Continued on page following the index





Analysis of Panel Data

Fourth Edition

Cheng Hsiao University of Southern California





CAMBRIDGEUNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi $-110025, India\,$

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781316512104

DOI: 10.1017/9781009057745

© Cheng Hsiao 1986, 2003, 2014, 2022

This publication 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 edition published 1986 Second edition 2003 Third edition 2014 Fourth edition 2022

A catalogue record for this publication is available from the British Library.

ISBN 978-1-316-51210-4 Hardback ISBN 978-1-009-06075-2 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.



To my parents,
Tseng Hsiao and Y. N. Shu Hsiao
and my family,
Amy Meiyun, Irene C., Allen C., Michael C. and Wendy C. Hsiao





Contents

List of Figures pa		page xiii	
Lis	t of To	ables	XV
Preface to the Fourth Edition		xviii	
1	Introduction		1
	1.1	Introduction	1
	1.2	Advantages of Panel Data	4
	1.3	Challenges to Panel Data Analysis	8
	1.4	Outline of the Monograph	12
2	Line	ar Static Models with Additive Effects	15
	2.1	Introduction	15
	2.2	Fixed-Effects Models: Dummy-Variable Approach	18
	2.3	Random-Effects Models: Variance-Components Models	22
	2.4	Fixed Effects or Random Effects	29
	2.5	Tests for Misspecification	37
	2.6	Models with Time- and/or Individual-Invariant Explanatory Variables and	
		both Individual- and Time-Specific Effects	38
	2.7	Analysis of Covariance Tests for the Presence of Individual- or	
		Time- Specific Effects	43
	2.8	Heteroscedasticity and Autocorrelation	50
	2.9	Models with Arbitrary Error Structure: Chamberlain π -Approach	54
	App	endix 2A: Consistency and Asymptotic Normality of the	
	M	inimum-Distance Estimator	59
	Appe	endix 2B: Characteristic Vectors and the Inverse of the	
	Va	riance-Covariance Matrix of a Three-Component Model	61
3	Dyna	amic Models with Additive Specific Effects	63
	3.1	Introduction	63
	3.2	The Least Squares and the Least Squares Dummy Variable (Covariance)	
		Estimator	64
	3.3	Method of Moments Estimator	67
	3.4	The Quasi-Likelihood Approach for Random-Effects Models	75
	3.5	The Likelihood Approach-Fixed-Effects Models	85



x Contents

	3.6	Relations between the Likelihood Based Estimator and the GMM	89
	3.7	Estimation of Dynamic Models with Arbitrary Serial Correlations in the	0.5
	2.0	Residuals	95
	3.8	Models with Both Individual- and Time-Specific Additive Effects	96
		endix 3A: Derivation of the Asymptotic Covariance Matrix of the	101
		easible MDE	101
	App	endix 3B: Large N and T Asymptotics	102
4	Sim	ultaneous-Equations Models	107
	4.1	Introduction	107
	4.2	Joint Generalized Least Squares Estimation Technique	110
	4.3	Estimation of Structural Equations	114
	4.4	Triangular System	121
		endix 4A: The Determinant and Inverse of the Triangular System	
	С	ovariance Matrix	131
5	Dyn	amic System	133
	5.1	Introduction	133
	5.2	Panel Vector Autoregressive Models	134
	5.3	Cointegrated Panel Models and Vector Error Correction	142
	5.4	Unit Root and Cointegration Tests	148
	5.5	Dynamic Simultaneous Equation Models	157
6	Disc	rete Data	165
	6.1	Introduction	165
	6.2	Qualitative Response Models for Cross-Sectional Data	165
	6.3	Panel Parametric Approach to Static Models with Heterogeneity	170
	6.4	Semiparametric Approach to Static Models	180
	6.5	Dynamic Models	183
	6.6	Alternative Approaches for Identifying Dynamic Dependence	203
7	Limited Dependent and Sample Selection Models		
	7.1	Cross-Sectional Data Approach	214
	7.2	A Sample Selection Example: Nonrandomly Missing Data in a Panel	223
	7.3	Tobit Models with Random Individual Effects	229
	7.4	Fixed-Effects Estimator	230
	7.5	An Example: Housing Expenditure	241
	7.6	Dynamic Tobit Models	244
8	Some Nonlinear Models		251
	8.1	Duration Model	251
	8.2	Count Data Model	262
	8.3	Nonparametric Models	268
9	Miso	cellaneous Topics	273
	9.1	Quantile Regression Analysis	273



		Contents	X
	9.2	Simulation Methods	276
	9.3	Data with Multilevel Structures	281
	9.4	Errors of Measurement	283
	9.5	Estimating Distributed Lags in Short Panels	288
	9.6	Rotating or Randomly Missing Data	297
	9.7	Pseudo Panels (or Repeated Cross-Sectional Data)	302
	9.8	Discretizing Unobserved Heterogeneity	305
10	Interactive Effects Models		307
	10.1	Introduction	307
	10.2	Fixed-Effects Linear Static Models	309
	10.3	Fixed-Effects Linear Dynamic Models	315
	10.4	Models with Mixed Fixed and Random Effects	323
	10.5	Quantile Models	328
	10.6	Factor Dimension Determination	330
11	Spatial Models and Tests for Cross-Sectional Dependence		339
	11.1	Introduction – Weak or Strong Cross-Correlations	339
	11.2	The Basic Formulation on Spatial Weight Matrix and Spatial	
		(Dependence) Coefficients	341
	11.3	Spatial Error Models	343
	11.4	Spatial Regressive Models	346
	11.5	Some Extensions	348
	11.6	Mixed Spatial and Factor Process	351
	11.7	Cross-Sectional Dependence Tests	354
12	Program Evaluation Using Panel Data		361
	12.1	Introduction	361
	12.2	Cross-Sectional Data Approach	362
	12.3	Panel Data Approach	368
	12.4	Some Examples	374
	12.5	Multiple Treated Units	381
	12.6	Simulating or Ranking Program Outcomes with Different Policy Options	384
13	Variable Coefficients Models		389
	13.1	Introduction	389
	13.2	Coefficients that Vary over Cross-Sectional Units	391
	13.3	Coefficients that Vary over Time and Cross-Sectional Units	401
	13.4	A Mixed Fixed- and Random-Coefficients Model	407
	13.5	Dynamic Random-Coefficients Models	414
	13.6	An Example: Liquidity Constraints and Firm Investment Expenditure	418
	13.7	Coefficients that Vary over Time	422
	13.8	Coefficients that Are Functions of Other Exogenous Variables	428
	13.9	Correlated Random-Coefficients Models	431
	Anne	endix 13A: Combination of Two Normal Distributions	438



xii Contents

14	4 Big Data Analytics		440
	14.1	Machine Learning Algorithms	440
	14.2	Inference with High-Dimensional Data	442
	14.3	Inference for Low-Dimensional Parameters in the Presence of	
		High-Dimensional Data	446
	14.4	Prediction	453
References		469	
Author Index		498	
Subject Index		502	



Figures

1.1	Homogeneous stope, neterogeneous intercept, a	page 9
1.2	Homogeneous slope, heterogeneous intercept, b	9
1.3	Homogeneous slope, heterogeneous intercept, c	9
1.4	Heterogeneous slope and intercept, a	10
1.5	Heterogeneous slope and intercept, b	10
1.6	Earning and Education	12
6.1	Male dynamic response path for the health shock that leads to activity	
	limiting $(ALC = 1)$ vs. not $(ALC = 0)$.	204
6.2	Male dynamic response path for the health shock that leads to activity	
	limiting $(ALC = 1)$ vs. not $(ALC = 0)$.	205
7.1	Density of y^* censored or truncated at 0.	217
7.2	Distribution of y and y^* under symmetric trimming.	218
7.3	Probability density of u_{it} .	232
7.4	Conditional densities of y_{i1}^* and y_{i2}^* given $(x_{i1}, x_{i2}, \alpha_i)$, assuming $\Delta x_i' \beta < 0$. 232
7.5	Conditional densities of y_{i1}^* and y_{i2}^* given $(x_{i1}, x_{i2}, \alpha_i)$, assuming $\Delta x_i' \beta > 0$.	233
7.6	The distribution of (y_{i1}^*, y_{i2}^*) assuming $\Delta x_i' \beta < 0$. $A_1 = \{(y_{i1}^*, y_{i2}^*) : y_{i1}^* > 0\}$	
	$-\Delta x_i' \boldsymbol{\beta}, \ \ y_{i2}^* > y_{i1}^* + \Delta x_i' \boldsymbol{\beta}\}, A_2 = \{(y_{i1}^*, y_{i2}^*) : y_{i1}^* \le -\Delta x_i' \boldsymbol{\beta}, \ \ y_{i2}^* > 0\},$	
	$B_1 = \{(y_{i1}^*, y_{i2}^*) : y_{i1}^* > -\Delta \mathbf{x}_i' \boldsymbol{\beta}, \ 0 < y_{i2}^* < y_{i1}^* + \Delta \mathbf{x}_i' \boldsymbol{\beta}\}, B_2 = \{(y_{i1}^*, y_{i2}^*) : \mathbf{x}_i^* \in \mathcal{S}_{i1}^* \}$	
	$y_{i1}^* > -\Delta x_i' \beta, \ y_{i2}^* \le 0$.	233
7.7	The distribution of (y_{i1}^*, y_{i2}^*) assuming $\Delta x_i' \beta > 0$.	
	$A_{1} = \{(y_{i1}^{*}, y_{i2}^{*}) : y_{i1}^{*} > 0, y_{i2}^{*} > y_{i1}^{*} + \Delta x_{i}' \beta\}, A_{2} = \{(y_{i1}^{*}, y_{i2}^{*}) : y_{i1}^{*} \leq 0, y_{i2}^{*} > \Delta x_{i}' \beta\}, B_{1} = \{(y_{i1}^{*}, y_{i2}^{*}) : y_{i1}^{*} > 0, \Delta x_{i}' \beta < y_{i2}^{*} < 0\}$	
	0, $y_{i2}^* > \Delta x_i' \beta$, $B_1 = \{(y_{i1}^*, y_{i2}^*) : y_{i1}^* > 0, \Delta x_i' \beta < y_{i2}^* < 0\}$	
	$y_{i1}^* + \Delta x_i' \beta$, $B_2 = \{(y_{i1}^*, y_{i2}^*) : y_{i1}^* > 0, y_{i2}^* \le \Delta x_i' \beta$.	234
7.8	$\gamma > 0, \gamma y_{i0}^* - \boldsymbol{\beta}' \Delta x_{i2} > 0.$	245
7.9	$\gamma > 0, \gamma y_{i0}^{*} - \boldsymbol{\beta}' \Delta x_{i2} < 0.$	246
8.1	Parameter estimates for the forward default intensity function	261
8.2	Lehman Brothers' term structure of forward and cumulative	
	default probabilities	262
12.1	Actual and AICC predicted real GDP growth rate from 1993Q1 to 2003Q4	375
12.2	AICC – Actual and counterfactual real GDP growth rate from	
	2004Q1 to 2008Q1	376
12.3	Actual and AIC predicted real GDP growth rate from 1993Q1 to 2003Q4	377
12.4	AIC – Actual and counterfactual real GDP growth rate from	
	2004Q1 to 2008Q1	378
12.5	Difference between the actual and counterfactual cigarette consumption in	
	the treatment period	380

xiii



xiv List of Figures

12.6	The impact of China's high speed rail projects on the per capita growth rate	
	of some Chinese cities	382
12.7	Counterfactual number of infected and active cases for Germany and UK.	
	Counterfactual number of infected and active cases for Germany and UK	387
14.1	Comparison of aggregation of the stock price changes of top 15	
	US companies	459
14.2	Relative maximum likelihood for the parameters δ_1	466



Tables

1.1	The span and sample sizes of the National Longitudinal Surveys	page 2
2.1	Least-squares estimates of elasticity of Minnesota farm production	
	function based on alternative assumptions	17
2.2	Characteristics of firms grouped on the basis of the firm constant	17
2.3	Wage equations (dependent variable: log wage)	30
2.4	Covariance tests for homogeneity	47
3.1	Simulation results for the biases of the MLEs for dynamic	
	random-effects models	82
3.2	Various estimates of the parameters of Balestra and Nerlove's	
	demand-for-gas model (3.4.24) from the pooled sample, 1957–1962.	84
3.3	Monte Carlo design	92
3.4	Bias of estimators ($T = 5$ and $N = 50$)	92
3.5	Root mean square error $(T = 5 \text{ and } N = 50)$	93
3.6	GMM with many IVs	93
3.7	Simple IV	94
3.8	Fully fledged MLE	94
3.9	Asymptotic properties of the likelihood based estimator and methods of	
	moment estimator for dynamic models	95
6.1	Average values of β for the fixed-effects probit model	176
6.2	Monte Carlo results for first-order Markov process	188
6.3	Estimates of employment models for women aged 45–59 in 1968	196
6.4	Comparisons of employment models using run data: Women aged 45–59	
	in 1968	199
6.5	Estimates of brand choices using various approaches (standard errors in	
	parentheses)	201
6.6	Maximum-likelihood estimators of dynamic model with random effects	202
6.7	Maximum-likelihood estimators of dynamic model with fixed effects	
	(bandwidth parameter = 8)	203
7.1	Parameter estimates of the earnings-function structural model with and	
	without a correction for attrition	228
7.2	Estimation results for the budget share equations without correction for	
	selection (standard errors in parentheses)	242
7.3	Estimation results for the budget share equations using panel data	
	models taking selection into account (standard errors in parentheses)	243
7.4	Estimates of AR(1) coefficients of log real annual earnings (in thousands)	248



xvi List of Tables

with 1 with	259 320 321 322 335
	321 322
	322
	335
	335
l with	
	336
with	
	337
3 with	
	338
	359
	374
on	371
Oli	375
	376
based	370
bascu	377
2	379
	381
xpenditures	
	412
	417
. 1 1	
viduai	410
	419
	420
!coefficients	420
_	420
- and	
ixed	
	421
l	
i	
ıdom	
	422
	451
	451
al assets	452
	on based n xpenditures vidual es vicoefficients - and ixed d sion, d ndom



	List of Tables	xvii
14.4	Estimation of β_x for the effects of unemployment insurance bonus on unemployment duration	452
14.5	unemployment duration Root-mean-square prediction error of log kilowatt-hours	432
14.5	(one-period-ahead forecast)	462
14.6	Root-mean-square prediction error of log kilowatts	
	(one-period-ahead forecast)	462
14.7	Least squares estimation of aggregate money demand function	463
14.8	Random-coefficient estimates of Japan prefecture money demand equation	463
14.9	Error sum of squares (ESS) and predicted error sum of squares (PES) for	
	disaggregate and aggregate data	464



Preface to the Fourth Edition

This is a substantially revised and expanded version of *Analysis of Panel Data* that considers new advancements in panel data analysis and streamlines and reorganizes topics discussed in earlier editions. The aim of this monograph remains the same: to comprehensively introduce panel econometric methodologies that are useful for identifying causal relationships among variables of interest to empirical investigators. Some empirical examples are also provided to give readers a perspective of the development of the field in the last half-century.

We start with modeling unobserved heterogeneity in additive forms for linear static and dynamic models, static and dynamic systems in structural and reduced forms, qualitative choice, limited dependent models, and some nonlinear models such as duration, count data, and nonparametric models. We then move to more complicated approaches to modeling unobserved heterogeneity, such as factor approach, spatial approach, and varying parameter models. We conclude with a chapter on the challenges of big data analytics to panel data analysis. However, no single monograph can cover the enormous contributions to panel data analysis. I apologize for the omissions.

I would like to thank Fatima Perez for patiently typing numerous versions of the manuscript, Dongwook Kim for carefully checking for errors, and China National Science Foundation grants 71631004 and 72033008 for partial research support. Comments by Xiaoyi Han on recent advances in spatial models, Cynthia Yang on stochastic SIR models, and Yuwei Xie on the big data approach to predictions are also much appreciated. I would also like to thank Robert Dreesen for support of this project, Laura Simmons for coordinating the production process, Matthew Van Atta for excellent copy editing, and the typesetting team led by Matthew Rohit Jayakumarxavier for their outstanding work; the journals *Computational Statistics, Econometrica, Econometrics Review*, the *Journal of the American Statistical Association*, the *Journal of Applied Econometrics*, the *Journal of Business and Economic Statistics*, and the *Journal of Econometrics*; the University of Chicago Press; and J.J. Heckman, D. McFadden, C. Manski, M.H. Pesaran, and C. Yang for their permissions to reproduce some of the materials published here.

I have received immeasurable help and support throughout my academic career from my family, teachers, colleagues, and students. I am grateful to my parents, Teng Hsiao and Y.N. Shu Hsiao, for providing a loving and safe environment during the tumultuous years of growing up; Leong Tan and Margaret Tan for giving a warm family environment to me as a lonely foreign student; Frank and Cynthia Kozin for advancing my understanding of mathematics in my youth; Helen Tu for typing up hundreds of job application letters; and the rest of my family members for their understanding and emotional support. I would like to thank Takeshi Amemiya and T.W. Anderson for being my lifelong mentors in

xviii



Preface to the Fourth Edition

xix

econometrics, and Dennis J. Aigner for his continuous support. I have also benefitted enormously from interactions with colleagues, collaborators, and friends, Jane Ayers, Pietro Balestra, Michael Binder, Badi Baltagi, Richard Blundell, Georges Bressons, Zongwu Cai, Luke M.W. Chan, Songnian Chen, H.S. Ching, In Choi, Hsinhsin Chou, Gregory Chow, Kannika Damronplasit, Hiroshi Fujiki, Ronald Gallant, Li Gan, Jiti Gao, John Geweke, John C. Ham, Toshihiko Hayashi, Yongmiao Hong, Arie Kapteyn, Xiao Ke, Kenneth C.W. Keng, Jae-Young Kim, Roger Klein, Roger Koenker, Chungming Kuan, Naoto Kunitomo, Ekaterini Kyriazidou, Lawrence J. Lau, Lungfei Lee, Lingfeng Li, Qi Li, Tong Li, Chongming Lin, Echu Liu, Esfandier Maasoumi, Jacques Mairesse, Michael McAleer, Roger H. Moon, Kimio Morimuni, Dean Mountain, Fred Nold, Jeffrey Nugent, Mario Tello Pacheco, Euston Quah, Scott Parris, Isabelle Perrigne, M. Hashem Pesaran, Dale J. Poirer, Geert Ridder, Peter M. Robinson, Patrick Sevester, Yan Shen, Baohong Sun, Kamil Tahmiscioglu, Derek Terrell, Kai Y. Tsui, Quang Vuong, Shui Ki Wan, Cindy SH Wang, Liqun Wang, Tongsan Wang, Yuenkang Wang, Thomas J. Wansbeek, Thomas Wilson, Zhije Xiao, Fengbei Yau, Eden Yu, Shihti Yu, Arnold Zellner, Xueyen Zhao, and Qiankun Zhou. To all of them, I am eternally grateful.

