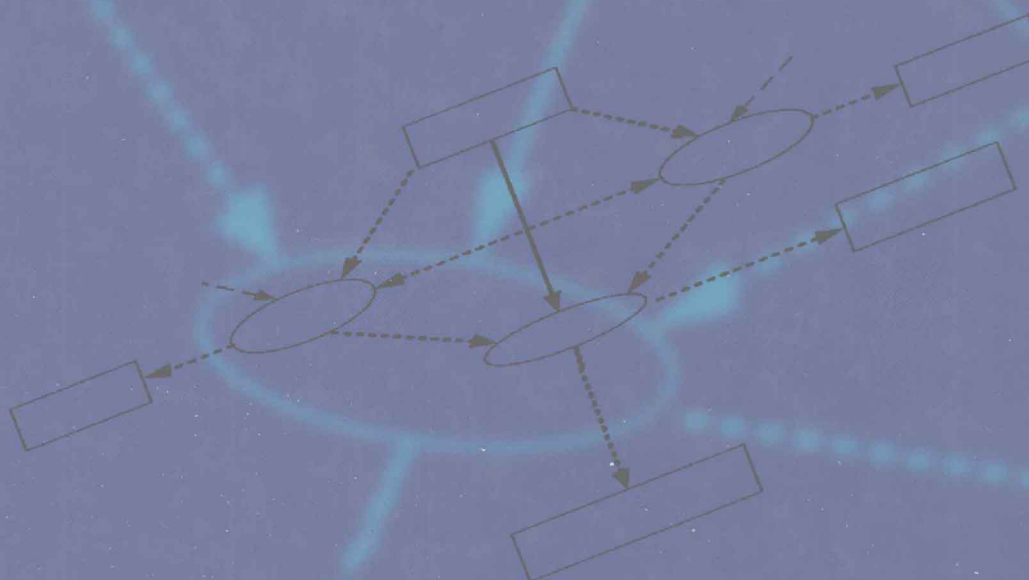


A. COLIN CAMERON • PRAVIN K. TRIVEDI



MICROECONOMETRICS

METHODS AND APPLICATIONS

CAMBRIDGE

Microeconometrics

Methods and Applications

A. Colin Cameron
*University of California,
Davis*

Pravin K. Trivedi
Indiana University



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521848053

© A. Colin Cameron and Pravin K. Trivedi 2005

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 2005

Printed in the United States of America

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

Library of Congress Cataloging in Publication Data

Cameron, Adrian Colin.

Microeconometrics : methods and applications / A. Colin Cameron, Pravin K. Trivedi.
p. cm.

Includes bibliographical references and index.

ISBN 0-521-84805-9 (hardcover)

1. Microeconomics – Econometric models. I. Trivedi, P. K. II. Title.

HB172.C343 2005

338.5'01'5195 – dc22

2004022273

ISBN-13 978-0-521-84805-3 hardback

ISBN-10 0-521-84805-9 hardback

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

Microeconometrics

This book provides a comprehensive treatment of microeconometrics, the analysis of individual-level data on the economic behavior of individuals or firms using regression methods applied to cross-section and panel data. The book is oriented to the practitioner. A good understanding of the linear regression model with matrix algebra is assumed. The text can be used for Ph.D. courses in microeconometrics, in applied econometrics, or in data-oriented microeconomics sub-disciplines; and as a reference work for graduate students and applied researchers who wish to fill in gaps in their tool kit. Distinguishing features include emphasis on nonlinear models and robust inference, as well as chapter-length treatments of GMM estimation, nonparametric regression, simulation-based estimation, bootstrap methods, Bayesian methods, stratified and clustered samples, treatment evaluation, measurement error, and missing data. The book makes frequent use of empirical illustrations, many based on seven large and rich data sets.

A. Colin Cameron is Professor of Economics at the University of California, Davis. He currently serves as Director of that university's Center on Quantitative Social Science Research. He has also taught at The Ohio State University and has held short-term visiting positions at Indiana University at Bloomington and at a number of Australian and European universities. His research in microeconometrics has appeared in leading econometrics and economics journals. He is coauthor with Pravin Trivedi of *Regression Analysis of Count Data*.

Pravin K. Trivedi is John H. Rudy Professor of Economics at Indiana University at Bloomington. He has also taught at The Australian National University and University of Southampton and has held short-term visiting positions at a number of European universities. His research in microeconometrics has appeared in most leading econometrics and health economics journals. He coauthored *Regression Analysis of Count Data* with A. Colin Cameron and is on the editorial boards of the *Econometrics Journal* and the *Journal of Applied Econometrics*.

To
my mother and the memory of my father
the memory of my parents

Preface

This book provides a detailed treatment of microeconomic analysis, the analysis of individual-level data on the economic behavior of individuals or firms. This type of analysis usually entails applying regression methods to cross-section and panel data.

The book aims at providing the practitioner with a comprehensive coverage of statistical methods and their application in modern applied microeconometrics research. These methods include nonlinear modeling, inference under minimal distributional assumptions, identifying and measuring causation rather than mere association, and correcting departures from simple random sampling. Many of these features are of relevance to individual-level data analysis throughout the social sciences.

The ambitious agenda has determined the characteristics of this book. First, although oriented to the practitioner, the book is relatively advanced in places. A cookbook approach is inadequate because when two or more complications occur simultaneously – a common situation – the practitioner must know enough to be able to adapt available methods. Second, the book provides considerable coverage of practical data problems (see especially the last three chapters). Third, the book includes substantial empirical examples in many chapters to illustrate some of the methods covered. Finally, the book is unusually long. Despite this length we have been space-constrained. We had intended to include even more empirical examples, and abbreviated presentations will at times fail to recognize the accomplishments of researchers who have made substantive contributions.

The book assumes a good understanding of the linear regression model with matrix algebra. It is written at the mathematical level of the first-year economics Ph.D. sequence, comparable to Greene (2003). We have two types of readers in mind. First, the book can be used as a course text for a microeconometrics course, typically taught in the second year of the Ph.D., or for data-oriented microeconomics field courses such as labor economics, public economics, and industrial organization. Second, the book can be used as a reference work for graduate students and applied researchers who despite training in microeconometrics will inevitably have gaps that they wish to fill.

For instructors using this book as an econometrics course text it is best to introduce the basic nonlinear cross-section and linear panel data models as early as possible,

PREFACE

initially skipping many of the methods chapters. The key methods chapter (Chapter 5) covers maximum-likelihood and nonlinear least-squares estimation. Knowledge of maximum likelihood and nonlinear least-squares estimators provides adequate background for the most commonly used nonlinear cross-section models (Chapters 14–17 and 20), basic linear panel data models (Chapter 21), and treatment evaluation methods (Chapter 25). Generalized method of moments estimation (Chapter 6) is needed especially for advanced linear panel data methods (Chapter 22).

For readers using this book as a reference work, the chapters have been written to be as self-contained as possible. The notable exception is that some command of general estimation results in Chapter 5, and occasionally Chapter 6, will be necessary. Most chapters on models are structured to begin with a discussion and example that is accessible to a wide audience.

The Web site www.econ.ucdavis.edu/faculty/cameron provides all the data and computer programs used in this book and related materials useful for instructional purposes.

This project has been long and arduous, and at times seemingly without an end. Its completion has been greatly aided by our colleagues, friends, and graduate students. We thank especially the following for reading and commenting on specific chapters: Bijan Borah, Kurt Brännäs, Pian Chen, Tim Cogley, Partha Deb, Massimiliano De Santis, David Drukker, Jeff Gill, Tue Gorgens, Shiferaw Gurmu, Lu Ji, Oscar Jorda, Roger Koenker, Chenghui Li, Tong Li, Doug Miller, Murat Munkin, Jim Priege, Ahmed Rahman, Sunil Sapra, Haruki Seitani, Yacheng Sun, Xiaoyong Zheng, and David Zimmer. Pian Chen gave detailed comments on most of the book. We thank Rajeev Dehejia, Bronwyn Hall, Cathy Kling, Jeffrey Kling, Will Manning, Brian McCall, and Jim Ziliak for making their data available for empirical illustrations. We thank our respective departments for facilitating our collaboration and for the production and distribution of the draft manuscript at various stages. We benefited from the comments of two anonymous reviewers. Guidance, advice, and encouragement from our Cambridge editor, Scott Parris, have been invaluable.

Our interest in econometrics owes much to the training and environments we encountered as students and in the initial stages of our academic careers. The first author thanks The Australian National University; Stanford University, especially Takeshi Amemiya and Tom MaCurdy; and The Ohio State University. The second author thanks the London School of Economics and The Australian National University.

Our interest in writing a book oriented to the practitioner owes much to our exposure to the research of graduate students and colleagues at our respective institutions, UC-Davis and IU-Bloomington.

Finally, we thank our families for their patience and understanding without which completion of this project would not have been possible.

A. Colin Cameron
Davis, California

Pravin K. Trivedi
Bloomington, Indiana

Contents

List of Figures	<i>page</i> xv
List of Tables	xvii
Preface	xxi

I Preliminaries

1 Overview	3
1.1 Introduction	3
1.2 Distinctive Aspects of Microeconometrics	5
1.3 Book Outline	10
1.4 How to Use This Book	14
1.5 Software	15
1.6 Notation and Conventions	16
2 Causal and Noncausal Models	18
2.1 Introduction	18
2.2 Structural Models	20
2.3 Exogeneity	22
2.4 Linear Simultaneous Equations Model	23
2.5 Identification Concepts	29
2.6 Single-Equation Models	31
2.7 Potential Outcome Model	31
2.8 Causal Modeling and Estimation Strategies	35
2.9 Bibliographic Notes	38
3 Microeconomic Data Structures	39
3.1 Introduction	39
3.2 Observational Data	40
3.3 Data from Social Experiments	48
3.4 Data from Natural Experiments	54

CONTENTS

3.5	Practical Considerations	58
3.6	Bibliographic Notes	61

II Core Methods

4	Linear Models	65
4.1	Introduction	65
4.2	Regressions and Loss Functions	66
4.3	Example: Returns to Schooling	69
4.4	Ordinary Least Squares	70
4.5	Weighted Least Squares	81
4.6	Median and Quantile Regression	85
4.7	Model Misspecification	90
4.8	Instrumental Variables	95
4.9	Instrumental Variables in Practice	103
4.10	Practical Considerations	112
4.11	Bibliographic Notes	112
5	Maximum Likelihood and Nonlinear Least-Squares Estimation	116
5.1	Introduction	116
5.2	Overview of Nonlinear Estimators	117
5.3	Extremum Estimators	124
5.4	Estimating Equations	133
5.5	Statistical Inference	135
5.6	Maximum Likelihood	139
5.7	Quasi-Maximum Likelihood	146
5.8	Nonlinear Least Squares	150
5.9	Example: ML and NLS Estimation	159
5.10	Practical Considerations	163
5.11	Bibliographic Notes	163
6	Generalized Method of Moments and Systems Estimation	166
6.1	Introduction	166
6.2	Examples	167
6.3	Generalized Method of Moments	172
6.4	Linear Instrumental Variables	183
6.5	Nonlinear Instrumental Variables	192
6.6	Sequential Two-Step m-Estimation	200
6.7	Minimum Distance Estimation	202
6.8	Empirical Likelihood	203
6.9	Linear Systems of Equations	206
6.10	Nonlinear Sets of Equations	214
6.11	Practical Considerations	219
6.12	Bibliographic Notes	220

CONTENTS

7 Hypothesis Tests	223
7.1 Introduction	223
7.2 Wald Test	224
7.3 Likelihood-Based Tests	233
7.4 Example: Likelihood-Based Hypothesis Tests	241
7.5 Tests in Non-ML Settings	243
7.6 Power and Size of Tests	246
7.7 Monte Carlo Studies	250
7.8 Bootstrap Example	254
7.9 Practical Considerations	256
7.10 Bibliographic Notes	257
8 Specification Tests and Model Selection	259
8.1 Introduction	259
8.2 m-Tests	260
8.3 Hausman Test	271
8.4 Tests for Some Common Misspecifications	274
8.5 Discriminating between Nonnested Models	278
8.6 Consequences of Testing	285
8.7 Model Diagnostics	287
8.8 Practical Considerations	291
8.9 Bibliographic Notes	292
9 Semiparametric Methods	294
9.1 Introduction	294
9.2 Nonparametric Example: Hourly Wage	295
9.3 Kernel Density Estimation	298
9.4 Nonparametric Local Regression	307
9.5 Kernel Regression	311
9.6 Alternative Nonparametric Regression Estimators	319
9.7 Semiparametric Regression	322
9.8 Derivations of Mean and Variance of Kernel Estimators	330
9.9 Practical Considerations	333
9.10 Bibliographic Notes	333
10 Numerical Optimization	336
10.1 Introduction	336
10.2 General Considerations	336
10.3 Specific Methods	341
10.4 Practical Considerations	348
10.5 Bibliographic Notes	352

III Simulation-Based Methods

11 Bootstrap Methods	357
11.1 Introduction	357
11.2 Bootstrap Summary	358
11.3 Bootstrap Example	366
11.4 Bootstrap Theory	368
11.5 Bootstrap Extensions	373
11.6 Bootstrap Applications	376
11.7 Practical Considerations	382
11.8 Bibliographic Notes	382
12 Simulation-Based Methods	384
12.1 Introduction	384
12.2 Examples	385
12.3 Basics of Computing Integrals	387
12.4 Maximum Simulated Likelihood Estimation	393
12.5 Moment-Based Simulation Estimation	398
12.6 Indirect Inference	404
12.7 Simulators	406
12.8 Methods of Drawing Random Variates	410
12.9 Bibliographic Notes	416
13 Bayesian Methods	419
13.1 Introduction	419
13.2 Bayesian Approach	420
13.3 Bayesian Analysis of Linear Regression	435
13.4 Monte Carlo Integration	443
13.5 Markov Chain Monte Carlo Simulation	445
13.6 MCMC Example: Gibbs Sampler for SUR	452
13.7 Data Augmentation	454
13.8 Bayesian Model Selection	456
13.9 Practical Considerations	458
13.10 Bibliographic Notes	458

IV Models for Cross-Section Data

14 Binary Outcome Models	463
14.1 Introduction	463
14.2 Binary Outcome Example: Fishing Mode Choice	464
14.3 Logit and Probit Models	465
14.4 Latent Variable Models	475
14.5 Choice-Based Samples	478
14.6 Grouped and Aggregate Data	480
14.7 Semiparametric Estimation	482

CONTENTS

14.8	Derivation of Logit from Type I Extreme Value	486
14.9	Practical Considerations	487
14.10	Bibliographic Notes	487
15	Multinomial Models	490
15.1	Introduction	490
15.2	Example: Choice of Fishing Mode	491
15.3	General Results	495
15.4	Multinomial Logit	500
15.5	Additive Random Utility Models	504
15.6	Nested Logit	507
15.7	Random Parameters Logit	512
15.8	Multinomial Probit	516
15.9	Ordered, Sequential, and Ranked Outcomes	519
15.10	Multivariate Discrete Outcomes	521
15.11	Semiparametric Estimation	523
15.12	Derivations for MNL, CL, and NL Models	524
15.13	Practical Considerations	527
15.14	Bibliographic Notes	528
16	Tobit and Selection Models	529
16.1	Introduction	529
16.2	Censored and Truncated Models	530
16.3	Tobit Model	536
16.4	Two-Part Model	544
16.5	Sample Selection Models	546
16.6	Selection Example: Health Expenditures	553
16.7	Roy Model	555
16.8	Structural Models	558
16.9	Semiparametric Estimation	562
16.10	Derivations for the Tobit Model	566
16.11	Practical Considerations	568
16.12	Bibliographic Notes	569
17	Transition Data: Survival Analysis	573
17.1	Introduction	573
17.2	Example: Duration of Strikes	574
17.3	Basic Concepts	576
17.4	Censoring	579
17.5	Nonparametric Models	580
17.6	Parametric Regression Models	584
17.7	Some Important Duration Models	591
17.8	Cox PH Model	592
17.9	Time-Varying Regressors	597
17.10	Discrete-Time Proportional Hazards	600
17.11	Duration Example: Unemployment Duration	603

CONTENTS

17.12	Practical Considerations	608
17.13	Bibliographic Notes	608
18	Mixture Models and Unobserved Heterogeneity	611
18.1	Introduction	611
18.2	Unobserved Heterogeneity and Dispersion	612
18.3	Identification in Mixture Models	618
18.4	Specification of the Heterogeneity Distribution	620
18.5	Discrete Heterogeneity and Latent Class Analysis	621
18.6	Stock and Flow Sampling	625
18.7	Specification Testing	628
18.8	Unobserved Heterogeneity Example: Unemployment Duration	632
18.9	Practical Considerations	637
18.10	Bibliographic Notes	637
19	Models of Multiple Hazards	640
19.1	Introduction	640
19.2	Competing Risks	642
19.3	Joint Duration Distributions	648
19.4	Multiple Spells	655
19.5	Competing Risks Example: Unemployment Duration	658
19.6	Practical Considerations	662
19.7	Bibliographic Notes	663
20	Models of Count Data	665
20.1	Introduction	665
20.2	Basic Count Data Regression	666
20.3	Count Example: Contacts with Medical Doctor	671
20.4	Parametric Count Regression Models	674
20.5	Partially Parametric Models	682
20.6	Multivariate Counts and Endogenous Regressors	685
20.7	Count Example: Further Analysis	690
20.8	Practical Considerations	690
20.9	Bibliographic Notes	691
V	Models for Panel Data	
21	Linear Panel Models: Basics	697
21.1	Introduction	697
21.2	Overview of Models and Estimators	698
21.3	Linear Panel Example: Hours and Wages	708
21.4	Fixed Effects versus Random Effects Models	715
21.5	Pooled Models	720
21.6	Fixed Effects Model	726
21.7	Random Effects Model	734

CONTENTS

21.8	Modeling Issues	737
21.9	Practical Considerations	740
21.10	Bibliographic Notes	740
22	Linear Panel Models: Extensions	743
22.1	Introduction	743
22.2	GMM Estimation of Linear Panel Models	744
22.3	Panel GMM Example: Hours and Wages	754
22.4	Random and Fixed Effects Panel GMM	756
22.5	Dynamic Models	763
22.6	Difference-in-Differences Estimator	768
22.7	Repeated Cross Sections and Pseudo Panels	770
22.8	Mixed Linear Models	774
22.9	Practical Considerations	776
22.10	Bibliographic Notes	777
23	Nonlinear Panel Models	779
23.1	Introduction	779
23.2	General Results	779
23.3	Nonlinear Panel Example: Patents and R&D	762
23.4	Binary Outcome Data	795
23.5	Tobit and Selection Models	800
23.6	Transition Data	801
23.7	Count Data	802
23.8	Semiparametric Estimation	808
23.9	Practical Considerations	808
23.10	Bibliographic Notes	809
VI	Further Topics	
24	Stratified and Clustered Samples	813
24.1	Introduction	813
24.2	Survey Sampling	814
24.3	Weighting	817
24.4	Endogenous Stratification	822
24.5	Clustering	829
24.6	Hierarchical Linear Models	845
24.7	Clustering Example: Vietnam Health Care Use	848
24.8	Complex Surveys	853
24.9	Practical Considerations	857
24.10	Bibliographic Notes	857
25	Treatment Evaluation	860
25.1	Introduction	860
25.2	Setup and Assumptions	862

CONTENTS

25.3	Treatment Effects and Selection Bias	865
25.4	Matching and Propensity Score Estimators	871
25.5	Differences-in-Differences Estimators	878
25.6	Regression Discontinuity Design	879
25.7	Instrumental Variable Methods	883
25.8	Example: The Effect of Training on Earnings	889
25.9	Bibliographic Notes	896
26	Measurement Error Models	899
26.1	Introduction	899
26.2	Measurement Error in Linear Regression	900
26.3	Identification Strategies	905
26.4	Measurement Errors in Nonlinear Models	911
26.5	Attenuation Bias Simulation Examples	919
26.6	Bibliographic Notes	920
27	Missing Data and Imputation	923
27.1	Introduction	923
27.2	Missing Data Assumptions	925
27.3	Handling Missing Data without Models	928
27.4	Observed-Data Likelihood	929
27.5	Regression-Based Imputation	930
27.6	Data Augmentation and MCMC	932
27.7	Multiple Imputation	934
27.8	Missing Data MCMC Imputation Example	935
27.9	Practical Considerations	939
27.10	Bibliographic Notes	940
A	Asymptotic Theory	943
A.1	Introduction	943
A.2	Convergence in Probability	944
A.3	Laws of Large Numbers	947
A.4	Convergence in Distribution	948
A.5	Central Limit Theorems	949
A.6	Multivariate Normal Limit Distributions	951
A.7	Stochastic Order of Magnitude	954
A.8	Other Results	955
A.9	Bibliographic Notes	956
B	Making Pseudo-Random Draws	957
	References	961
	Index	999

List of Figures

3.1	Social experiment with random assignment	<i>page</i> 50
4.1	Quantile regression estimates of slope coefficient	89
4.2	Quantile regression estimated lines	90
7.1	Power of Wald chi-square test	249
7.2	Density of Wald test on slope coefficient	253
9.1	Histogram for log wage	296
9.2	Kernel density estimates for log wage	296
9.3	Nonparametric regression of log wage on education	297
9.4	Kernel density estimates using different kernels	300
9.5	k -nearest neighbors regression	309
9.6	Nonparametric regression using Lowess	310
9.7	Nonparametric estimate of derivative of y with respect to x	317
11.1	Bootstrap estimate of the density of t -test statistic	368
12.1	Halton sequence draws compared to pseudo-random draws	411
12.2	Inverse transformation method for unit exponential draws	413
12.3	Accept–reject method for random draws	414
13.1	Bayesian analysis for mean parameter of normal density	424
14.1	Charter boat fishing: probit and logit predictions	466
15.1	Generalized random utility model	516
16.1	Tobit regression example	531
16.2	Inverse Mills ratio as censoring point c increases	540
17.1	Strike duration: Kaplan–Meier survival function	575
17.2	Weibull distribution: density, survivor, hazard, and cumulative hazard functions	585
17.3	Unemployment duration: Kaplan–Meier survival function	604
17.4	Unemployment duration: survival functions by unemployment insurance	605
17.5	Unemployment duration: Nelson–Aalen cumulated hazard function	606
17.6	Unemployment duration: cumulative hazard function by unemployment insurance	606

LIST OF FIGURES

18.1	Length-biased sampling under stock sampling: example	627
18.2	Unemployment duration: exponential model generalized residuals	633
18.3	Unemployment duration: exponential-gamma model generalized residuals	633
18.4	Unemployment duration: Weibull model generalized residuals	635
18.5	Unemployment duration: Weibull-IG model generalized residuals	636
19.1	Unemployment duration: Cox CR baseline survival functions	661
19.2	Unemployment duration: Cox CR baseline cumulative hazards	662
21.1	Hours and wages: pooled (overall) regression	712
21.2	Hours and wages: between regression	713
21.3	Hours and wages: within (fixed effects) regression	713
21.4	Hours and wages: first differences regression	714
23.1	Patents and R&D: pooled (overall) regression	793
25.1	Regression-discontinuity design: example	880
25.2	RD design: treatment assignment in sharp and fuzzy designs	883
25.3	Training impact: earnings against propensity score by treatment	892
27.1	Missing data: examples of missing regressors	924