

Cambridge University Press  
0521632013 - Regression Analysis of Count Data  
A. Colin Cameron and Pravin K. Trivedi  
Frontmatter  
[More information](#)

---

Econometric Society Monographs No. 30

**Regression analysis  
of count data**

## Regression Analysis of Count Data

Students in both the natural and social sciences often seek regression models to explain the frequency of events, such as visits to a doctor, auto accidents, or new patents awarded. This analysis provides the most comprehensive and up-to-date account of models and methods to interpret such data. The authors have conducted research in the field for nearly 15 years and in this work combine theory and practice to make sophisticated methods of analysis accessible to practitioners working with widely different types of data and software. The treatment will be useful to researchers in areas such as applied statistics, econometrics, marketing, operations research, actuarial studies, demography, biostatistics, and quantitatively oriented sociology and political science. The book may be used as a reference work on count models or by students seeking an authoritative overview. The analysis is complemented by template programs available on the Internet through the authors' homepages.

A. Colin Cameron is Associate Professor of Economics at the University of California, Davis. He has also taught at the Ohio State University and held visiting positions at the Australian National University, Indiana University at Bloomington, and the University of New South Wales. His research on count data and microeconometrics has appeared in many leading econometrics journals.

Pravin K. Trivedi is Professor of Economics at Indiana University at Bloomington and previously taught at the Australian National University and University of Southampton. He has also held visiting positions at the European University Institute and the World Bank. His publications on count data and micro- and macro-econometrics have appeared in most leading econometrics journals.

## Econometric Society Monographs

### *Editors:*

Peter Hammond, Stanford University  
 Alberto Holly, University of Lausanne

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  
 Cheng Hsiao *Analysis of panel data*, 0 521 38933 X  
 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 52136055 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 analysis*, 0 521 25280 6, 0 521 57446 3  
 Christopher Sims, Editor *Advances in econometrics – Sixth World Congress (Volume I)*, 0 521 56610 X

(Series list continues on page after index.)

Cambridge University Press  
0521632013 - Regression Analysis of Count Data  
A. Colin Cameron and Pravin K. Trivedi  
Frontmatter  
[More information](#)

---

# Regression Analysis of Count Data

A. Colin Cameron  
Pravin K. Trivedi



Cambridge University Press  
 0521632013 - Regression Analysis of Count Data  
 A. Colin Cameron and Pravin K. Trivedi  
 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 <http://www.cup.cam.ac.uk>  
 40 West 20th Street, New York, NY 10011-4211, USA <http://www.cup.org>  
 10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© A. Colin Cameron and Pravin K. Trivedi 1998

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 1998

Printed in the United States of America

Typeset in Times Roman 10/12 pt. in L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> [TB]

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

*Library of Congress Cataloging-in-Publication Data*

Cameron, Adrian Colin.

Regression analysis of count data / A. Colin Cameron, Pravin K. Trivedi.

p. cm.

Includes bibliographical references.

ISBN 0-521-63201-3 (hb). – ISBN 0-521-63567-5 (pb)

1. Regression analysis. 2. Econometrics. I. Trivedi, P. K. II. Title.

CA278.2.C36 1998

519.5'36 – dc21

98–15184

CIP

ISBN 0 521 63201 3 hardback

ISBN 0 521 63567 5 paperback

Cambridge University Press  
0521632013 - Regression Analysis of Count Data  
A. Colin Cameron and Pravin K. Trivedi  
Frontmatter  
[More information](#)

---

*To Michelle and Bhavna*

## Contents

	<i>List of Figures</i>	page xi
	<i>List of Tables</i>	xii
	<i>Preface</i>	xv
1	Introduction	1
	1.1 Poisson Distribution	3
	1.2 Poisson Regression	8
	1.3 Examples	10
	1.4 Overview of Major Issues	15
	1.5 Bibliographic Notes	17
2	Model Specification and Estimation	19
	2.1 Introduction	19
	2.2 Example and Definitions	20
	2.3 Likelihood-Based Models	22
	2.4 Generalized Linear Models	27
	2.5 Moment-Based Models	37
	2.6 Testing	44
	2.7 Derivations	50
	2.8 Bibliographic Notes	57
	2.9 Exercises	57
3	Basic Count Regression	59
	3.1 Introduction	59
	3.2 Poisson MLE, PMLE, and GLM	61
	3.3 Negative binomial MLE and QGPMLE	70
	3.4 Overdispersion Tests	77
	3.5 Use of Regression Results	79
	3.6 Ordered and Other Discrete-Choice Models	85
	3.7 Other Models	88
	3.8 Iteratively Reweighted Least Squares	93
	3.9 Bibliographic Notes	94
	3.10 Exercises	95

viii	<b>Contents</b>	
4	Generalized Count Regression	96
4.1	Introduction	96
4.2	Mixture Models for Unobserved Heterogeneity	97
4.3	Models Based on Waiting-Time Distributions	106
4.4	Katz, Double-Poisson, and Generalized Poisson	112
4.5	Truncated Counts	117
4.6	Censored Counts	121
4.7	Hurdle and Zero-Inflated Models	123
4.8	Finite Mixtures and Latent Class Analysis	128
4.9	Estimation by Simulation	134
4.10	Derivations	135
4.11	Bibliographic Notes	136
4.12	Exercises	137
5	Model Evaluation and Testing	139
5.1	Introduction	139
5.2	Residual Analysis	140
5.3	Goodness of Fit	151
5.4	Hypothesis Tests	158
5.5	Inference with Finite Sample Corrections	163
5.6	Conditional Moment Specification Tests	168
5.7	Discriminating among Nonnested Models	182
5.8	Derivations	185
5.9	Bibliographic Notes	187
5.10	Exercises	188
6	Empirical Illustrations	189
6.1	Introduction	189
6.2	Background	190
6.3	Analysis of Demand for Health Services	192
6.4	Analysis of Recreational Trips	206
6.5	LR Test: A Digression	216
6.6	Concluding Remarks	218
6.7	Bibliographic Notes	219
6.8	Exercises	220
7	Time Series Data	221
7.1	Introduction	221
7.2	Models for Time Series Data	222
7.3	Static Regression	226
7.4	Integer-Valued ARMA Models	234
7.5	Autoregressive Models	238
7.6	Serially Correlated Error Models	240
7.7	State-Space Models	242
7.8	Hidden Markov Models	244
7.9	Discrete ARMA Models	245
7.10	Application	246



<b>Contents</b>		ix
	7.11 Derivations: Tests of Serial Correlation	248
	7.12 Bibliographic Notes	250
	7.13 Exercises	250
8	Multivariate Data	251
	8.1 Introduction	251
	8.2 Characterizing Dependence	252
	8.3 Parametric Models	256
	8.4 Moment-Based Estimation	260
	8.5 Orthogonal Polynomial Series Expansions	263
	8.6 Mixed Multivariate Models	269
	8.7 Derivations	272
	8.8 Bibliographic Notes	273
9	Longitudinal Data	275
	9.1 Introduction	275
	9.2 Models for Longitudinal Data	276
	9.3 Fixed Effects Models	280
	9.4 Random Effects Models	287
	9.5 Discussion	290
	9.6 Specification Tests	293
	9.7 Dynamic and Transition Models	294
	9.8 Derivations	299
	9.9 Bibliographic Notes	300
	9.10 Exercises	300
10	Measurement Errors	301
	10.1 Introduction	301
	10.2 Measurement Errors in Exposure	302
	10.3 Measurement Errors in Regressors	307
	10.4 Measurement Errors in Counts	309
	10.5 Underreported Counts	313
	10.6 Derivations	323
	10.7 Bibliographic Notes	324
	10.8 Exercises	325
11	Nonrandom Samples and Simultaneity	326
	11.1 Introduction	326
	11.2 Alternative Sampling Frames	326
	11.3 Simultaneity	331
	11.4 Sample Selection	336
	11.5 Bibliographic Notes	343
12	Flexible Methods for Counts	344
	12.1 Introduction	344
	12.2 Efficient Moment-Based Estimation	345
	12.3 Flexible Distributions Using Series Expansions	350
	12.4 Flexible Models of Conditional Mean	356

**Contents**

12.5	Flexible Models of Conditional Variance	358
12.6	Example and Model Comparison	364
12.7	Derivations	367
12.8	Count Models: Retrospect and Prospect	367
12.9	Bibliographic Notes	369
<i>Appendices:</i>		
A	Notation and Acronyms	371
B	Functions, Distributions, and Moments	374
B.1	Gamma Function	374
B.2	Some Distributions	375
B.3	Moments of Truncated Poisson	376
C	Software	378
	<i>References</i>	379
	<i>Author Index</i>	399
	<i>Subject Index</i>	404

## List of Figures

1.1	Frequency distribution of counts for four types of events.	11
4.1	Negative binomial compared with Poisson.	100
4.2	Two examples of double Poisson.	115
4.3	Two univariate two-component mixtures of Poisson.	130
5.1	Comparison of Pearson, deviance, and Anscombe residuals.	143
5.2	Takeover bids: residual plots.	150
6.1	OFP visits: directional gradients.	200
6.2	OFP visits: component densities from the FMNB2 NB1 model.	203
7.1	Strikes: output (rescaled) and strikes per month.	232
7.2	Strikes: actual and predicted from a static model.	233
7.3	Strikes: actual and predicted from a dynamic model.	248

## List of Tables

3.1	Doctor visits: actual frequency distribution.	68
3.2	Doctor visits: variable definitions and summary statistics.	68
3.3	Doctor visits: Poisson PMLE with different standard error estimates.	69
3.4	Doctor visits: NB2 and NB1 model estimators and standard errors.	76
3.5	Doctor visits: Poisson PMLE mean effects and scaled coefficients.	83
3.6	Doctor visits: alternative estimates and $t$ ratios.	92
4.1	Doctor visits: GEC(k) and gamma MLE and $t$ ratios.	114
5.1	Takeover bids: actual frequency distribution.	147
5.2	Takeover bids: variable definitions and summary statistics.	147
5.3	Takeover bids: Poisson PMLE with NB1 standard errors and $t$ ratios.	148
5.4	Takeover bids: descriptive statistics for various residuals.	149
5.5	Takeover bids: correlations of various residuals.	149
5.6	Takeover bids: Poisson MLE predicted and actual probabilities.	158
6.1	OFP visits: actual frequency distribution.	193
6.2	OFP visits: variable definitions and summary statistics.	194
6.3	OFP visits: likelihood ratio tests.	199
6.4	OFP visits: information criteria (AIC and BIC).	200
6.5	OFP visits: FMNB2 NB1 model, actual, fitted distributions and goodness-of-fit tests.	201
6.6	OFP visits: FMNB2 NB1 model estimates and standard errors.	201
6.7	OFP visits: FMNB2 NB1 model fitted means and variances.	202
6.8	OFP visits: NB2 hurdle model estimates and $t$ ratios.	207
6.9	Recreational trips: actual frequency distribution.	208
6.10	Recreational trips: variable definitions and summary statistics.	208
6.11	Recreational trips: Poisson, NB2, and ZIP model estimates and $t$ ratios.	209

<b>List of Tables</b>		xiii
6.12	Recreational trips: finite mixture estimates and $t$ ratios.	212
6.13	Recreational trips: hurdle model estimates and $t$ ratios.	214
6.14	Recreational boating trips: actual and fitted cumulative frequencies.	215
6.15	Rejection frequencies at nominal 10% significance level.	217
7.1	Strikes: variable definitions and summary statistics.	231
7.2	Strikes: Poisson PMLE with NB1 standard errors and $t$ ratios.	232
7.3	Strikes: residuals autocorrelations and serial correlation tests.	233
7.4	Strikes: Zeger-Qaqish autoregressive model estimates and diagnostics.	247
8.1	Orthogonal polynomials: first and second order.	269
8.2	Health services: pairwise independence tests.	269
9.1	Patents: Poisson PMLE with NB1 standard errors.	286
12.1	Recreational trips: flexible distribution estimators and $t$ ratios.	365
12.2	Recreational trips: cumulative predicted probabilities.	366

## Preface

This book describes regression methods for count data, where the response variable is a nonnegative integer. The methods are relevant for analysis of counts that arise in both social and natural sciences.

Despite their relatively recent origin, count data regression methods build on an impressive body of statistical research on univariate discrete distributions. Many of these methods have now found their way into major statistical packages, which has encouraged their application in a variety of contexts. Such widespread use has itself thrown up numerous interesting research issues and themes, which we explore in this book.

The objective of the book is threefold. First, we wish to provide a synthesis and integrative survey of the literature on count data regressions, covering both the statistical and econometric strands. The former has emphasized the framework of generalized linear models, exponential families of distributions, and generalized estimating equations; the latter has emphasized nonlinear regression and generalized method of moment frameworks. Yet between them there are numerous points of contact that can be fruitfully exploited. Our second objective is to make sophisticated methods of data analysis more accessible to practitioners with different interests and backgrounds. To this end we consider models and methods suitable for cross-section, time series, and longitudinal data. Detailed analyses of several data sets as well as shorter illustrations, implemented from a variety of viewpoints, are scattered throughout the book to put empirical flesh on theoretical or methodological discussion. We draw on examples from, and give references to, works in many applied areas. Our third objective is to highlight the potential for further research by discussion of issues and problems that need more analysis. We do so by embedding count data models in a larger body of econometric and statistical work on discrete variables and, more generally, on nonlinear regression.

The book can be divided into four parts. Chapters 1 and 2 contain introductory material on count data and a comprehensive review of statistical methods for nonlinear regression models. Chapters 3, 4, 5, and 6 present models and applications for cross-section count data. Chapters 7, 8, and 9 present methods for data other than cross-section data, namely time series, multivariate, and

longitudinal or panel data. Chapters 10, 11, and 12 present methods for common complications, including measurement error, sample selection and simultaneity, and semiparametric methods. Thus the coverage of the book is qualitatively similar to that in a complete single book on linear regression models.

The book is directed toward researchers, graduate students, and other practitioners in a wide range of fields. Because of our background in econometrics, the book emphasizes issues arising in econometric applications. Our training and background also influence the organizational structure of the book, but areas outside econometrics are also considered. The essential prerequisite for this book is familiarity with the linear regression model using matrix algebra. The material in the book should be accessible to people with a background in regression and statistical methods up to the level of a standard first-year graduate econometrics text such as Greene's *Econometric Analysis*. Although basic count data methods are included in major statistical packages, more advanced analysis can require programming in languages such as SPLUS, GAUSS, or MATLAB.

Our own entry into the field of count data models dates back to the early 1980s, when we embarked on an empirical study of the demand for health insurance and health care services at the Australian National University. Since then we have been involved in many empirical investigations that have influenced our perceptions of this field. We have included numerous data-analytic discussions in this volume, to reflect our own interests and those of readers interested in real data applications. The data sets, computer programs, and related materials used in this book are available through Internet access to the website <http://www.econ.ucdavis.edu/count.html>. These materials supplement and complement this book and will help new entrants to the field, especially graduate students, to make a relatively easy start.

We have learned much on modeling count data through collaborations with coauthors, notably Partha Deb, Shiferaw Gurmu, Per Johansson, Kajal Mukhopadhyay, and Frank Windmeijer. The burden of writing this book has been eased by help from many colleagues, coauthors, and graduate students. In particular, we thank the following for their generous attention, encouragement, help, and comments on earlier drafts of various chapters: Kurt Brännäs, David Hendry, Primula Kennedy, Tony Lancaster, Scott Long, Xing Ming, Grayham Mizon, Neil Shephard, and Bob Shumway, in addition to the coauthors already mentioned. We especially thank David Hendry and Scott Long for their detailed advice on manuscript preparation using Latex software and Scientific Workplace. The manuscript has also benefited from the comments of a referee and the series editor, Alberto Holly, and from the guidance of Scott Parris of Cambridge University Press.

Work on the book was facilitated by periods spent at various institutions. The first author thanks the Department of Statistics and the Research School of Social Sciences at the Australian National University, the Department of Economics at Indiana University–Bloomington, and the University of California, Davis, for support during extended leaves at these institutions in 1995 and

Cambridge University Press  
0521632013 - Regression Analysis of Count Data  
A. Colin Cameron and Pravin K. Trivedi  
Frontmatter  
[More information](#)

---

**Preface**

xvii

1996. The second author thanks Indiana University and the European University Institute, Florence, for support during his tenure as Jean Monnet Fellow in 1996, which permitted a period away from regular duties. For shorter periods of stay that allowed us to work jointly, we thank the Department of Economics at Indiana University, SELAPO at University of Munich, and the European University Institute.

Finally we would both like to thank our families for their patience and forbearance, especially during the periods of intensive work on the book. This work would not have been possible at all without their constant support.

**A. Colin Cameron**

*Davis, California*

**Pravin K. Trivedi**

*Bloomington, Indiana*