Regression Analysis of Count Data

Second Edition

A. Colin Cameron
Department of Economics
University of California
Davis, CA 95616, U.S.A.
Telephone: 530-752–8396
Fax: 530-752-9382
E-mail: accameron@ucdavis.edu

Pravin K. Trivedi
Department of Economics
Indiana University
Bloomington, IN 47405, U.S.A.
Telephone: 812-855-3567
Fax: 812-855-3736
E-mail: trivedi@indiana.edu

April 2012

Contents

Preface

Since *Regression Analysis of Count Data* was published in 1998 significant new research has contributed to the range and scope of count data models. This growth is reflected in many new journal articles, fuller coverage in textbooks, and wide interest in and availability of software for handling count data models. These developments (to which we have also contributed) have motivated us to revise and expand the first edition. Like the first edition, the current version reflects an orientation towards practical data analysis.

The revisions in this edition have affected all chapters. First, we have corrected the typographical and other errors in the first edition, improved the graphics throughout, and where appropriate we have provided a cleaner and simpler exposition. Second we have revised and relocated material that seemed better placed in a different location, mostly within the same chapter though occasionally in a different chapter. For example material in Chapter 4 (generalized count models), chapter 8 (multivariate counts), and Chapter 13 (measurement errors) has been pruned and rearranged so the more mainstream topics appear earlier while the more marginal topics have disappeared altogether. For similar reasons bootstrap inference has moved from Chapter 5 to Chapter 2. Our goal here has been to improve quality of synthesis and accessibility of material to the reader. Third, the final few chapters have been reordered. Chapter 10 (endogeneity and selection) has moved up from Chapter 11. It replaces the measurement error chapter which now appears as chapter 13. Chapter 11 now covers flexible parametric models (previously Chapter 12). And the current Chapter 12, which covers Bayesian methods, is a new addition. Fourth, we have removed material that was of marginal interest and replaced it with material of potentially greater interest, especially to practitioners. For example, as barriers to implementation of more computer-intensive methods have come down, we have liberally sprinkled illustrations of simulation-based methods throughout the book. Fifth, bibliographic notes at the end of every chapter have been refreshed to include newer references and topics. Sixth, we have developed an almost complete set of computer code for the examples in this book.

The first edition has been expanded by about 25 per cent. This expansion reflects the addition of a new chapter 12 on Bayesian methods as well as significant additions to most other chapters. Chapter 2 has new sections on robust inference and empirical likelihood, and material on the bootstrap and generalized estimating equations now appears in this chapter. In Chapter 3 and throughout the book, the term pseudo-ML has been changed to quasi-ML and robust standard errors are computed using the robust sandwich form. Chapter 4 improves the coverage and discussion of how many alternative count models relate to each other. Censored, truncated, hurdle, zero-inflated and, especially, finite mixture models are now covered in greater depth, with a more uniform notation, and hierarchical count models and models with cross-sectional and spatial dependence have been newly added. Chapter 5 moves up presentation of methods for discrimination among nonnested models. Chapter 6 adds a new empirical example of fertility data that poses a fresh challenge to count data modelers. The time series coverage in Chapter 7 has been expanded to include more recently developed models, and there is some rearrangement so that the most often used models appear first. The coverage of multivariate count models in Chapter 8 uses a broader

and more modern range of dependence concepts, and provides a lengthy treatment of parametric copula-based models. The survey of count data panel models in Chapter 9 gives greater emphasis to moment-based approaches and has a more comprehensive coverage of dynamic panels, the role of initial conditions, conditionally correlated random effects, flexible functional forms and specification tests. Chapter 10 provides an improved exposition of models with endogeneity and selection, including consideration of latent factor and two-part models as well as simulation-based inference and control function estimators. A major new topic in Chapter 11 is quantile regression models for count data, and the coverage of semiparametric and nonparametric methods has been considerably expanded and updated. As previously mentioned, the new Chapter 12 covers Bayesian analysis of count model, providing an entry to the world of Markov chain Monte Carlo analysis of count models. Finally, Chapter 13 provides a comprehensive survey of measurement error models for count data. As a result of the expanded coverage of old topics and appearance of new ones, the bibliography is now significantly larger and includes more than a hundred additional new references.

To emphasize its empirical orientation the book has added many new examples based on real data. These examples are scattered throughout the book, especially in Chapters 6-12. In addition we have a number of examples based on simulated data. Researchers, instructors and students interested in replicating our results can obtain all the data and computer programs used to produce the results given in this book via Internet from our respective personal web sites.

This revised and expanded second edition draws extensively from our jointly authored research undertaken with Partha Deb, Jie Qun Guo, Judex Hyppolite, Tong Li, Doug Miller, Murat Munkin, and David Zimmer. Jeff Racine provided valuable advice for Chapter 11. We thank them all.

A. Colin Cameron

Davis, CA

Pravin K. Trivedi

Bloomington, IN

April 2012