Brief contents

What's new in this edition

- Several new models for count data
  - » Three-parameter negative binomial
  - » Poisson inverse Gaussian
  - » Zero-inflated generalized Poisson
- New chapter on data synthesis and simulation
- Marginal effects and discrete change for GLMs

Generalized Linear Models and Extensions, Third Edition
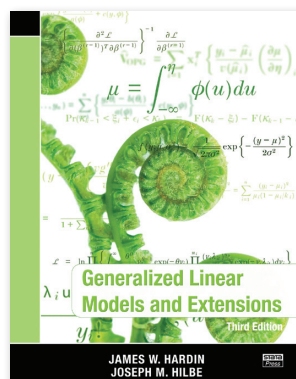
By James W. Hardin and Joseph M. Hilbe

Publisher: Stata Press
Copyright: 2012
ISBN-13: 978-1-59718-105-1
Pages: 455; paperback
Price: **$58**

For more details or to order, visit us online at
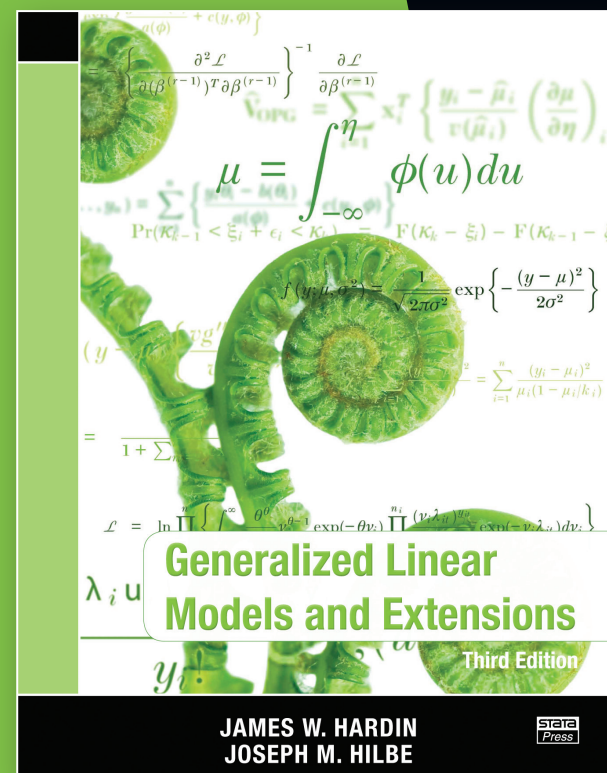**stata-press.com/books/generalized-linear-models-and-extensions.**

How to contact us

Stata Press presents

Generalized Linear Models and Extensions

Third Edition

JAMES W. HARDIN
JOSEPH M. HILBE

About the authors

James W. Hardin is an associate professor in and the Biostatistics Division Director of the Department of Epidemiology and Biostatistics at the University of South Carolina. He is also an affiliate faculty member of the Institute for Families in Society at the University of South Carolina. Dr. Hardin is an active researcher in laboratory outcomes, disparities, sexual identity/health, and self-reported health data. He is coauthor with Phillip Good of four editions of *Common Errors in Statistics (And How to Avoid Them)*. Hardin is on the editorial board of the *Stata Journal* and continues to develop user-written commands for the Stata community. With Hilbe, he teaches courses through Statistics.com on generalized linear models and generalized estimating equations.

Joseph M. Hilbe is an emeritus professor at the University of Hawaii, an adjunct professor of statistics at Arizona State University, and a Solar System Ambassador at NASA's Jet Propulsion Laboratory. An elected Fellow of the American Statistical Association and elected member (Fellow) of the International Statistical Institute, Professor Hilbe is chair of the International Astrostatistics Network, is currently on the editorial boards of six statistics journals, and is author of several popular texts in statistical modeling, including two editions of *Negative Binomial Regression* (Cambridge University Press), *Logistic Regression Models* (Chapman & Hall/CRC), and with Professor Hardin, *Generalized Estimating Equations*.

The book's audience

- Applied researchers, especially those who analyze count data
- Researchers familiar with generalized linear models but who are new to Stata
- Stata users looking for a theoretical reference for generalized linear models

Comment from the Stata technical group

Generalized linear models (GLMs) extend linear regression to models with a non-Gaussian, or even discrete, response. GLM theory is predicated on the exponential family of distributions—a class so rich that it includes the commonly used logit, probit, and Poisson models. Although one can fit these models in Stata by using specialized commands (for example, `logit` for logit models), fitting them as GLMs with Stata's `glm` command offers some advantages. For example, model diagnostics may be calculated and interpreted similarly regardless of the assumed distribution.

This text thoroughly covers GLMs, both theoretically and computationally, with an emphasis on Stata. The theory consists of showing how the various GLMs are special cases of the exponential family, showing general properties of this family of distributions, and showing the derivation of maximum likelihood (ML) estimators and standard errors. Hardin and Hilbe show how iteratively reweighted least squares, another method of parameter estimation, is a consequence of ML estimation using Fisher scoring. The authors also discuss different methods of estimating standard errors, including robust methods, robust methods with clustering, Newey–West, outer product of the gradient, bootstrap, and jackknife. The thorough coverage of model diagnostics includes measures of influence such as Cook's distance, several forms of residuals, the Akaike and Bayesian information criteria, and various R^2-type measures of explained variability.

After presenting general theory, Hardin and Hilbe then break down each distribution. Each distribution has its own chapter that explains the computational details of applying the general theory to that particular distribution. Pseudocode plays a valuable role here, because it lets the authors describe computational algorithms relatively simply. Devoting an entire chapter to each distribution (or family, in GLM terms) also allows

for including real-data examples showing how Stata fits such models, as well as presenting certain diagnostics and analytical strategies that are unique to that family. The chapters on binary data and on count (Poisson) data are excellent in this regard. Hardin and Hilbe give ample attention to the problems of overdispersion and zero inflation in count-data models.

The final part of the text concerns extensions of GLMs, which come in three forms. First, the authors cover multinomial responses, both ordered and unordered. Although multinomial responses are not strictly a part of GLM, the theory is similar in that one can think of a multinomial response as an extension of a binary response. The examples presented in these chapters often use the authors' own Stata programs, augmenting official Stata's capabilities. Second, GLMs may be extended to clustered data through generalized estimating equations (GEEs), and one chapter covers GEE theory and examples. Finally, GLMs may be extended by programming one's own family and link functions for use with Stata's official `glm` command, and the authors detail this process.

In addition to other enhancements—for example, a new section on marginal effects—the third edition contains several new extended GLMs, giving Stata users new ways to capture the complexity of count data. New count models include a three-parameter negative binomial known as NB-P, Poisson inverse Gaussian (PIG), zero-inflated generalized Poisson (ZIGP), a rewritten generalized Poisson (GP), two- and three-component finite mixture models, and a generalized censored Poisson and negative binomial. This edition has a new chapter on simulation and data synthesis, but also shows how to construct a wide variety of synthetic and Monte Carlo models throughout the book.