Modeling Ordered Choices

It is increasingly common for analysts to seek out the opinions of individuals and organizations using attitudinal scales such as degree of satisfaction or importance attached to an issue. Examples include levels of obesity, seriousness of a health condition, attitudes towards service levels, opinions on products, voting intentions, and the degree of clarity of contracts. Ordered choice models provide a relevant methodology for capturing the sources of influence that explain the choice made among a set of ordered alternatives. The methods have evolved to a level of sophistication that can allow for heterogeneity in the threshold parameters, in the explanatory variables (through random parameters), and in the decomposition of the residual variance. This book brings together contributions in ordered choice modeling from a number of disciplines, synthesizing developments over the last fifty years, and suggests useful extensions to account for the wide range of sources of influence on choice.

William H. Greene is Professor of Economics and Toyota Motor Corp Professor at the Stern School of Business, New York University.

David A. Hensher is Professor of Management and Director of the Institute of Transport and Logistics, University of Sydney.

Modeling Ordered Choices

A Primer

William H. Greene and David A. Hensher

CAMBRIDGE UNIVERSITY PRESS

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

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521142373

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

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

ISBN 978-0-521-19420-4 Hardback
ISBN 978-0-521-14237-3 Paperback

Contents

Tables

Cambridge University Press
978-0-521-19420-4 - Modeling Ordered Choices: A Primer
William H. Greene and David A. Hensher
Frontmatter
More information

Figures

Cambridge University Press
978-0-521-19420-4 - Modeling Ordered Choices: A Primer
William H. Greene and David A. Hensher
Frontmatter
More information

Preface

This book began as a short note to propose the estimator in Section 8.3. In researching the recent developments in ordered choice modeling, we concluded that it would be useful to include some pedagogical material about uses and interpretation of the model at the most basic level. Our review of the literature revealed an impressive breadth and depth of applications of ordered choice modeling, but no single source that provided a comprehensive summary. There are several somewhat narrow surveys of the basic ordered probit/logit model, including Winship and Mare (1984), Becker and Kennedy (1992), Daykin and Moffatt (2002) and Boes and Winkelmann (2006a), and a book-length treatment, by Johnson and Albert (1999) that is focused on Bayesian estimation of the basic model parameters using grouped data. (See, also, Congdon (2005), Ch. 7 and Agresti (2002), Section 7.4.) However, these stop well short of examining the extensive range of variants of the model and the variety of fields of applications, such as bivariate and multivariate models, two-part models, duration models, panel data models, models with anchoring vignettes, semiparametric approaches, and so on. (We have, of necessity, omitted mention of many – perhaps most – of the huge number of applications.) This motivated us to assemble this more complete overview of the topic. As this review proceeded, it struck us that a more thorough survey of the model itself, including its historical development, might also be useful and (we hope) interesting for readers. The following is also a survey of the methodological literature on modeling ordered outcomes and ordered choices.

The development of the ordered choice regression model has emerged in two surprisingly disjointed strands of literature: in its earliest forms in the bioassay literature, and in its modern social science counterpart with the pioneering paper by McKelvey and Zavoina (1975) and its successors, such as Terza (1985). There are a few prominent links between these two literatures, notably Walker and Duncan (1967). However, even up to the contemporary literature, biological scientists and social scientists have largely successfully avoided bumping into each other. For example, the 500+ entry references list of this survey shares only four items with its 100+ entry counterpart in Johnson and Albert (1999).

The earliest applications of modeling ordered outcomes involved aggregate (grouped) data assembled in table format, and with moderate numbers of levels of usually a single stimulus. The fundamental ordered logistic ("cumulative odds") model in its various forms serves well as an appropriate modeling framework for such data. Walker and Duncan (1967) focused on a major limitation of the approach. When data are obtained with large numbers of inputs – the models in Brewer *et al.* (2008), for example, involve over forty covariates – and many levels of those inputs, then cross-tabulations are no longer feasible or adequate. Two requirements become obvious: the use of the individual data and the heavy reliance on what amount to multiple regression-style techniques. McKelvey and Zavoina (1975) added to the model a reliance on a formal underlying "data-generating process," the latent regression. This mechanism makes an occasional appearance in the bioassay treatment, but is never absent from the social science application. The *cumulative odds model* for contingency tables and the fundamental *ordered probit model* for individual data are now standard tools. The recent advances in ordered choice modeling have involved modeling heterogeneity, in cross-sections and in panel data sets. These include a variety of threshold models and models of parameter variation such as latent class and mixed and hierarchical models. The chapters in this book present, in some detail, the full range of varieties of models for ordered choices.

This book is intended to be a survey of a particular class of discrete choice models. We anticipate that it can be used in a graduate level course in applied econometrics or statistics at the level of, say, Greene (2008a) or Wooldridge (2002b) and as a reference in specialized courses such as microeconometrics or discrete choice modeling. We assume that the reader is familiar with basic statistics and econometrics and with modeling techniques somewhat beyond the linear regression model. An introduction to maximum likelihood estimation and the most familiar binary choice models, probit and logit, is assumed, though developed in great detail in Chapter 2. The focus of this book is on areas of application of ordered choice models. The range of applications considered here includes economics, sociology, health economics, finance, political science, statistics in medicine, transportation planning, and many others. We have drawn on all of these in our collection of applications. We leave it to others, e.g., Hayashi (2000), Wooldridge (2002a), or Greene (2008a) to provide background material on, e.g., asymptotic theory for estimators and practical aspects of nonlinear optimization.

All of the computations carried out here were done with *NLOGIT* (see www.nlogit.com). Most of them can also be done with several other packages, such as *Stata* and SAS. Since this book is not a "how to" guide for any particular computer program, we have not provided any instructions on how to obtain the results with *NLOGIT* (or any other program). We assume that the interested reader can follow through on our developments with their favorite software, whatever that might be. Rather, our interest is in the models and techniques.

We would like to thank Joseph Hilbe and Chandra Bhat for their suggestions that have improved this work and Allison Greene for her assistance with the manuscript. Any errors that remain are ours.