Models for Ordered Choices

William Greene
Department of Economics
Stern School of Business
New York University

To appear in

Handbook of Choice Modelling,
Stephane Hess and Andrew Daly
Edward Elgar Publishing
Forthcoming 2013

1. Introduction

Netflix (www.netflix.com) is an internet company that rents movies to subscribers. After a customer rents a movie, the next time they log on to the website, they are invited to rate the movie on a five point scale, where five is the highest, most favorable rating. The ratings of the many thousands of subscribers who rented that movie are averaged to provide a recommendation to prospective viewers. For example, as of April 5, 2009, the average rating of the 2007 movie *National Treasure: Book of Secrets* [2007] given by approximately 12,900 visitors to the site was 3.8. This rating process provides a natural application of the models and methods described in this survey.

The model described here is an *ordered choice model*. Ordered choice models are appropriate for a wide variety of settings in the social and biological sciences. The essential ingredient is the *mapping from an underlying, naturally ordered preference scale to a discrete ordered observed outcome*, such as the rating scheme described above. The model of ordered choice pioneered by Aitcheson and Silvey [1957] and Snell [1964] and articulated in its modern form by Zavoina and McElvey [1969] and McElvey and Zavoina [1971, 1975] and McCullagh [1980] has become a widely used tool in many fields. The number of applications in the current literature is large and increasing rapidly. A search of just the 'ordered probit' model identified applications on:

- academic grades (Butler et al. [1994], Li and Tobias [2006a]),
- bond ratings (Terza [1985]),
- Congressional voting on a Medicare bill (McElvey and Zavoina [1975]),
- credit ratings (Cheung [1996] , Metz and Cantor [2006]),
- driver injury severity in car accidents (Wang and Kockelman [2005], Eluru, Bhat and Hensher [2008]),
- drug reactions (Fu et al.[2004]),
- duration (Han and Hausman [1990], Ridder [1990]),
- education (Machin and Vignoles [2005], Carneiro, Hansen and Heckman [2001, 2003], Cameron and Heckman [1998], Cunha, Heckman and Navarro [2007], Johnson and Albert [1999]),
- eye disease severity (Biswas and Das [2002]),
- financial failure of firms (Jones and Hensher [2004], Hensher and Jones [2007]),
- happiness (Winkelmann [2005], Zigante [2007]),
- health status (Greene [2008a], Riphahn, Wambach and Million [2003]),
- insect resistance to insecticide (Walker and Duncan [1967]),
- job classification in the military (Marcus and Greene [1983]),
- job training (Groot and van den Brink [2002a]),
- labor supply (Heckman and MaCurdy [1981]),
- life satisfaction (Clark et al. [2001], Wim and ven den Brink [2002, 2003b]),
- monetary policy (Eichengreen, Watson and Grossman [1985]),
- nursing labor supply (Brewer et al. [2008]),
- obesity (Greene, Harris, Hollingsworth and Maitra [2008]),
- perceptions of difficulty making left turns while driving (Zhang [2007]),
- pet ownership (Butler and Chatterjee [1997]),
- political efficacy (King et al. [2004]),
- pollution (Wang and Kockelman [2009a]),
- product quality (Prescott and Visscher [1977], Shaked and Sutton [1982]),

- promotion and rank in nursing (Pudney and Shields [2000]),
- self assessed health (Greene, Harris and Hollingsworth [2012]),
- stock price movements (Tsay [2005]),
- tobacco use (Harris and Zhao [2007], Kasteridis, Munkin and Yen [2008]),
- toxicity in pregnant mice (Agresti [2002]),
- trip stops (Bhat [1997]),
- vehicle ownership (Bhat and Pulugurta [1998], Train [1986], Hensher, Smith, Milthorpe and Bernard [1992]),
- work disability (Kapteyn et al. [2007]),

and hundreds more.

This survey will lay out some of the central features of ordered choice models. After developing the basic model, we describe some of the specification issues and model extensions that have appeared in recent studies. There are numerous surveys of ordered choice modeling in the received literature. This one draws heavily on Greene and Hensher [2010]. Some of the ideas developed in Sections 4 and 5 are extended in Greene, Harris, Hollingsworth and Weterings [2012]. Section 2 briefly discusses two foundational elements of the model, random utility models and the model for binary choices. The main development of the ordered choice model is given in Section 3. Sections 4 through 6 detail a number of specification issues, including individual heterogeneity, functional form and panel data modeling.

2. Binary Choice Model

The *random utility* model is one of two essential building blocks that form the foundation for modeling ordered choices. The second fundamental pillar is the *model for binary choices*. The ordered choice model that will be the focus of the rest of this survey is an extension of a model used to analyze the situation of a choice between two alternatives – whether the individual takes an action or does not, or chooses one of two elemental alternatives, and so on.

2.1 Random Utility Formulation of a Model for Binary Choice

An application we will develop is based on a survey question in a large German panel data set, roughly, "on a scale from zero to ten, how satisfied are you with your health?" The full data set consists of from one to seven observations – it is an unbalanced panel – on 7,293 households for a total of 27,326 household year observations. A histogram of the responses appears in Figure 3.2. We might formulate a random utility/ordered choice model for the variable R_i = "*Health Satisfaction*" as

$$
\begin{aligned}
U_i^* &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \\
R_i &= 0 \text{ if } -\infty < U_i^* \leq \mu_0, \\
R_i &= 1 \text{ if } \mu_0 < U_i^* \leq \mu_1, \\
&\cdots \\
R_i &= 10 \text{ if } \mu_9 < U_i^* < +\infty,
\end{aligned}
$$

where $\mathbf{x}_i$ is a set of variables such as gender, income, age, and education that are thought to influence the response to the survey question. (Note that at this point, we are pooling the panel data as if they were a cross section of n = 32,726 independent observations and denoting by i one of those observations.) The average response in the full sample is 6.78 Consider a simple response variable, y_i = "*Healthy*," (i.e., better than average), defined by

$y_i = 1$ if $R_i \geq 7$ and $y_i = 0$ otherwise.

Then, in terms of the original variables, the model for y_i is

$y_i = 0$ if $R_i \in (0, 1, 2, 3, 4, 5, 6)$ and $y_i = 1$ if $R_i \in (7, 8, 9, 10)$.

By adding the terms, we then find, for the two possible outcomes,

$y_i = 0$ if $U_i^* \leq \mu_6$,
$y_i = 1$ if $U_i^* > \mu_6$.

Substituting for U_i^*, we find

$y_i = 1$ if $\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i > \mu_6$
or $y_i = 1$ if $\varepsilon_i > \mu_6 - \boldsymbol{\beta}'\mathbf{x}_i$
and $y_i = 0$ otherwise.

We now assume that the first element of $\boldsymbol{\beta}'\mathbf{x}_i$ is a constant term, α, so that $\boldsymbol{\beta}'\mathbf{x}_i - \mu_6$ equivalent to $\boldsymbol{\gamma}'\mathbf{x}_i$ where the first element of $\boldsymbol{\gamma}$ is a constant that is equal to $\alpha - \mu_6$ and the rest of $\boldsymbol{\gamma}$ is the same as the rest of $\boldsymbol{\beta}$. Then, the binary outcome is determined by

$y_i = 1$ $\boldsymbol{\gamma}'\mathbf{x}_i + \varepsilon_i > 0$
and $y_i = 0$ otherwise.

In general terms, we write the binary choice model in terms of the underlying utility as

$y_i^* = \boldsymbol{\gamma}'\mathbf{x}_i + \varepsilon_i$,
$y_i = 1[y_i^* > 0]$,

where the function 1[condition] equals one if the condition is true and zero if it is false.

2.2 Probability Models for Binary Choices

The observed outcome, y_i, is determined by a *latent regression*,

$y_i^* = \boldsymbol{\gamma}'\mathbf{x}_i + \varepsilon_i$.

The random variable y_i takes two values, one and zero, with probabilities

$$\begin{aligned}\text{Prob}(y_i = 1|\mathbf{x}_i) &= \text{Prob}(y_i^* > 0|\mathbf{x}_i)\\ &= \text{Prob}(\boldsymbol{\gamma}'\mathbf{x}_i + \varepsilon_i > 0)\\ &= \text{Prob}(\varepsilon_i > -\boldsymbol{\gamma}'\mathbf{x}_i).\end{aligned}$$

The model is completed by the specification of a particular probability distribution for ε_i. In terms of building an internally consistent model, we require that the probabilities be between zero and one and that they increase when $\boldsymbol{\gamma}'\mathbf{x}_i$ increases. In principle, any probability distribution defined over the entire real line will suffice. The literature on binary choices is overwhelmingly dominated by two models, the standard normal distribution, which gives rise to the *probit model*, $f(\varepsilon_i) = \exp(-\varepsilon_i^2/2)/(2\pi)^{1/2}$ and the standard logistic distribution, $f(\varepsilon_i) = \exp(\varepsilon_i)/[1 + \exp(\varepsilon_i)]^2$, which produces the *logit model*. The normal distribution can be motivated by an appeal to the central

limit theorem and modeling human behavior as the sum of myriad underlying influences. The logistic distribution has proved to be a useful mathematical form for modeling purposes for several decades. These two are by far the most frequently used in applications. Other distributions, such as the complementary log log and Gompertz distribution that are built into modern software such as *Stata* and *NLOGIT* are sometimes specified as well, though without obvious motivation.

The implication of the model specification is that $y_i|\mathbf{x}_i$ is a Bernoulli random variable with

$$\begin{aligned}\text{Prob}(y_i = 1|\mathbf{x}_i) &= \text{Prob}(y_i^* > 0|\mathbf{x}_i)\\ &= \text{Prob}(\varepsilon_i > -\boldsymbol{\gamma}'\mathbf{x}_i)\\ &= \int_{-\boldsymbol{\gamma}'\mathbf{x}_i}^{\infty} f(\varepsilon_i)d\varepsilon_i\\ &= 1 - F(-\boldsymbol{\gamma}'\mathbf{x}_i),\end{aligned}$$

where $F(.)$ denotes the cumulative density function (CDF) or *distribution function* for ε_i. The standard normal and standard logistic distributions are both *symmetric distributions* that have the property that $F(\boldsymbol{\gamma}'\mathbf{x}_i) = 1 - F(-\boldsymbol{\gamma}'\mathbf{x}_i)$. This produces the convenient result $\text{Prob}(y_i = 1|\mathbf{x}_i) = F(\boldsymbol{\gamma}'\mathbf{x}_i)$. Standard notations for the normal and logistic distribution functions are $\Phi(\boldsymbol{\gamma}'\mathbf{x}_i)$ and $\Lambda(\boldsymbol{\gamma}'\mathbf{x}_i)$, respectively. The resulting probit model for a binary outcome is shown in Figure 2.1. Note that since y_i equals zero and one with probabilities $F(-\boldsymbol{\gamma}'\mathbf{x}_i)$ and $F(\boldsymbol{\gamma}'\mathbf{x}_i)$, $\text{E}[y_i|\boldsymbol{\gamma}'\mathbf{x}_i] = F(\boldsymbol{\gamma}'\mathbf{x}_i)$. Thus, the function in Figure 2.1 is also the regression function of y_i on $\boldsymbol{\gamma}'\mathbf{x}_i$ as well as $\text{E}[y_i|\mathbf{x}_i]$

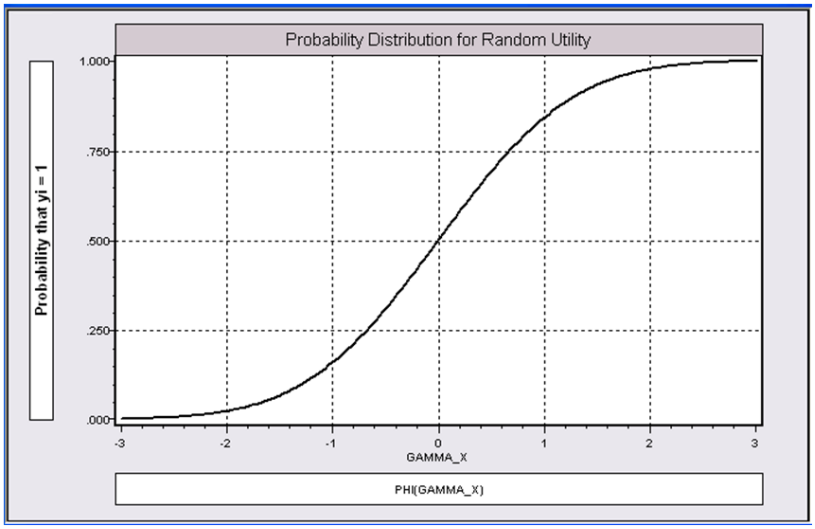

Figure 2.1 Probit Model for Binary Choice

3. A Model for Ordered Choices

The ordered probit model in its contemporary, regression based form was proposed by McElvey and Zavoina [1969, 1971, 1975] for the analysis of ordered, categorical, nonquantitative choices, outcomes and responses. Their application concerned Congressional preferences on a Medicaid bill. Familiar recent examples include bond ratings, discrete opinion surveys such as those on political questions, obesity measures, preferences in consumption, and satisfaction and health status surveys such as those analyzed by Boes and Winkelmann [2006a, 2006b] and other applications mentioned in the introduction. The model is used to describe the data generating process for a random outcome that takes one of a set of discrete, *ordered* outcomes.

3.1 A Latent Regression Model for a Continuous Measure

The model platform is an underlying random utility model or latent regression model,

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ i = 1,...,n,$$

in which the continuous latent utility or 'measure,' y_i^* is observed in discrete form through a *censoring* mechanism;

$$\begin{aligned} y_i &= 0 \text{ if } \mu_{-1} < y_i^* \leq \mu_0, \\ &= 1 \text{ if } \mu_0 < y_i^* \leq \mu_1, \\ &= 2 \text{ if } \mu_1 < y_i^* \leq \mu_2 \\ &= \ldots \\ &= J \text{ if } \mu_{J-1} < y_i^* \leq \mu_J. \end{aligned}$$

Note, for purposes of this introduction, that we have assumed that neither coefficients, $\boldsymbol{\beta}$, nor thresholds, μ_j, differ across individuals. These strong assumptions will be reconsidered and relaxed as the analysis proceeds. The vector $\mathbf{x}_i$ is a set of K covariates that are assumed to be strictly independent of ε_i; $\boldsymbol{\beta}$ is a vector of K parameters that is the object of estimation and inference. The n sample observations are labeled $i = 1,...,n$.

The model contains the unknown marginal utilities, $\boldsymbol{\beta}$, as well as J+2 threshold parameters, μ_j, all to be estimated using a sample of n observations, indexed by $i = 1,...,n$. The data consist of the covariates, $\mathbf{x}_i$ and the observed discrete outcome, $y_i = 0,1,...,J$. The assumption of the properties of the "disturbance," ε_i, completes the model specification. The conventional assumptions are that ε_i is a continuous random disturbance with conventional cumulative distribution function (cdf), $F(\varepsilon_i|\mathbf{x}_i) = F(\varepsilon_i)$ with support equal to the real line, and that the density, $f(\varepsilon_i) = F'(\varepsilon_i)$ is likewise defined over the real line. The assumption of the distribution of ε_i includes independence from, or exogeneity of, $\mathbf{x}_i$.

3.2 Ordered Choice as an Outcome of Utility Maximization

The appearance of the ordered choice model in the transportation literature falls somewhere between a latent regression approach and a more formal discrete choice interpretation. Bhat and Pulugurta [1998] discuss a model for 'ownership propensity,'

$$C_i = k \text{ if and only if } \psi_{k-1} < C_i^* \leq \psi_k,\ k = 0,1,...,K,\ \psi_{-1} = -\infty,\ \psi_K = +\infty,$$

where C_i^* represents the latent auto ownership propensity of household i. The observable counterpart to C_i^* is C_i, typically the number of vehicles owned.[1] Agyemand-Duah and Hall [1997] apply the model to numbers of trips. Bhat [1997] models the number of non-work commute stops with work travel mode choice.] From here, the model can move in several possible directions: A natural platform for the observed number of vehicles owned might seem to be the count data models (e.g., Poisson) detailed in, e.g., Cameron and Trivedi [1998, 2005] or even a choice model defined on a choice set of alternatives, 0,1,2,…[2]

The Poisson model for C_i would not follow from a model of utility maximization, though it would, perhaps, adequately *describe* the data generating process. However, a looser

[1] See, e.g., Hensher, Smith, Milthorpe and Bernard [1992].
[2] Hensher et al. [1992].

interpretation of the vehicle ownership count as a reflection of the underlying preference intensity for ownership suggests an ordered choice model as a plausible alternative platform. Bhat and Pulugurta [1998] provide a utility maximization framework that produces an ordered choice model for the observed count. Their model departs from a random utility framework that assigns separate utility values to different states, e.g., zero car ownership vs. some car ownership, less than or equal to one car owned vs. more than one, and so on (presumably up to the maximum observed in the sample). A suitable set of assumptions about the ranking of utilities produces essentially an unordered choice model for the number of vehicles. A further set of assumptions about the parameterization of the model makes it consistent with the latent regression model above.[1] A wide literature in this area includes applications by Kitamura [1987, 1988], Golub and van Wissen [1988], Kitamura and Bunch [1989], Golob [1990], Bhat and Koppelmann [1993], Bhat [1996], Agyemara-Duan and Hall [1997], Bhat and Pulugurta [1998] and Bhat, Carini and Misra [1999].

One might question the strict ordering of the vehicle count. For example, the vehicles might include different mixtures of cars, SUVs and trucks. Though a somewhat fuzzy ordering might still seem natural, several authors have opted instead, to replace the ordered choice model with an unordered choice framework, the multinomial logit model and variants.[2] Applications include Bhat and Pulugurta [1998], Mannering and Winsten [1985], Train [1986], Bunch and Kitamura [1990], Hensher, et al. [1992], Purvis [1994] and Agostino, Bhat and Pas [1996]. Groot and van den Brink [2003a] encounter the same issue in their analysis of job training sessions. A count model for sessions seems natural, however the length and depth of sessions differs enough to suggest a simple count model will distort the underlying variable of interest, 'training.'

While many applications appear on first consideration to have some 'natural' ordering, this is not necessarily the case when one recognizes that the ordering must have some meaning also in utility or satisfaction space (i.e., a naturally ordered underlying preference scale) if it assumed that the models are essentially driven by the behavioral rule of utility maximization. The number of cars owned is a good example: 0,1,2, >2 is a natural ordering in physical vehicle space, but it is not necessarily so in utility space.

3.3 The Observed Discrete Outcome

A typical social science application might begin from a measured outcome such as:

"Rate your feelings about the proposed legislation as

0	Strongly oppose
1	Mildly oppose
2	Indifferent
3	Mildly support
4	Strongly support.

The latent regression model would describe an underlying continuous, albeit unobservable, preference for the legislation as y_i^*. The surveyed individual, even if they could, does not provide y_i^*, but rather, a censoring of y_i^* into five different ranges, one of which is closest to their own true preferences. By the laws of probability, the probabilities associated with the observed outcomes are

$$\text{Prob}[y_i = j \mid \mathbf{x}_i] = \text{Prob}[\varepsilon_i \leq \mu_j - \boldsymbol{\beta}'\mathbf{x}_i] - \text{Prob}[\varepsilon_i \leq \mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i], j = 0,1,...,J.$$

[1] See Bhat and Pulugurta [1998, page 64].

[2] See, again, Bhat and Pulugurta [1998] who suggest a different utility function for each observed level of vehicle ownership.

It is worth noting, as do many other discrete choice models, the 'model' describes probabilities of outcomes. It does not directly describe the relationship between a y_i and the covariates $\mathbf{x}_i$; there is no obvious regression relationship at work between the observed random variable and the covariates. This calls into question the interpretation of $\boldsymbol{\beta}$, an issue to which we will return at several points below. Though y_i is not described by a regression relationship with $\mathbf{x}_i$ – i.e., y_i is merely a label – one might consider examining the binary variables,

m_{ij} = 1 if $y_i = j$ and 0 if not,

or

M_{ij} = 1 if $y_i \leq j$ and 0 if not,

or

M_{ij}' = 1 if $y_i \geq j$ and 0 if not.

The second and third of these as well as m_{i0} can be described by a simple binary choice (probit or logit) model, though these are usually not of interest. However, in general, there is no obvious regression (conditional mean) relationship between the observed dependent variable(s), y_i, and $\mathbf{x}_i$.

Several normalizations are needed to identify the model parameters. First, in order to preserve the positive signs of all of the probabilities, we require $\mu_j > \mu_{j-1}$. Second, if the support is to be the entire real line, then $\mu_{-1} = -\infty$ and $\mu_J = +\infty$. Since the data contain no unconditional information on scaling of the underlying variable – if y_i^* is scaled by any positive value, then scaling the unknown μ_j and $\boldsymbol{\beta}$ by the same value preserves the observed outcomes – an unconditional, free variance parameter, $\text{Var}[\varepsilon_i] = \sigma_\varepsilon^2$, is not identified (estimable). It is convenient to make the identifying restriction σ_ε = a constant, $\overline{\sigma}$. The usual approach to this normalization is to assume that $\text{Var}[\varepsilon_i|\mathbf{x}_i] = 1$ in the probit case and $\pi^2/3$ in the logit model – in either case to eliminate the free structural scaling parameter. Finally, we will assume that $\mathbf{x}_i$ contains a constant term, which, in turns, requires $\mu_0 = 0$. (If, with the other normalizations, and with a constant term present, this normalization is not imposed, then adding a constant to μ_0 and the same constant to the intercept term in $\boldsymbol{\beta}$ will leave the probability unchanged.)

3.4 Probabilities and the Log Likelihood

With the full set of normalizations in place, the likelihood function for estimation of the model parameters is based on the implied probabilities,

$$\text{Prob}[y_i = j \mid \mathbf{x}_i] = [F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i)] \geq 0, \; j = 0,1,...,J.$$

Figure 3.1 shows the probabilities for an ordered choice model with three outcomes,

$$\begin{aligned}
\text{Prob}[y_i = 0|\mathbf{x}_i] &= F(0 - \boldsymbol{\beta}'\mathbf{x}_i) - F(-\infty - \boldsymbol{\beta}'\mathbf{x}_i) = \text{F}(-\boldsymbol{\beta}'\mathbf{x}_i), \\
\text{Prob}[y_i = 1|\mathbf{x}_i] &= F(\mu_1 - \boldsymbol{\beta}'\mathbf{x}_i) - F(-\boldsymbol{\beta}'\mathbf{x}_i), \\
\text{Prob}[y_i = 2|\mathbf{x}_i] &= F(+\infty - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_1 - \boldsymbol{\beta}'\mathbf{x}_i) = 1 - \text{F}(\mu_1 - \boldsymbol{\beta}'\mathbf{x}_i).
\end{aligned}$$

Estimation of the parameters is a straightforward problem in maximum likelihood estimation. (See, e.g., Pratt [1981] and Greene [2007a, 2008a].) The log likelihood function is

$$\log L = \Sigma_{i=1}^{n}\Sigma_{j=0}^{J} m_{ij} \log[F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i)],$$

where $m_{ij} = 1$ if $y_i = j$ and 0 otherwise. Maximization is done subject to the constraints $\mu_{-1} = -\infty$, $\mu_0 = 0$ and $\mu_J = +\infty$. The remaining constraints, $\mu_{j-1} < \mu_j$ can, in principle, be imposed by a reparameterization in terms of some underlying structural parameters, such as

$$\begin{aligned} \mu_j &= \mu_{j-1} + \exp(\alpha_j) \\ &= \Sigma_{m=1}^{j} \exp(\alpha_m), \end{aligned}$$

however, this is typically unnecessary. See, e.g., Fahrmeier and Tutz [2001]. Expressions for the derivatives of the log likelihood can be found in McElvey and Zavoina [1975], Maddala [1983], Long [1997], *Stata* [2008] and Econometric Software [2012]. The estimator of the asymptotic covariance matrix for the MLE is computed by familiar methods, using the Hessian, outer products of gradients, or in some applications, a 'robust' sandwich estimator.

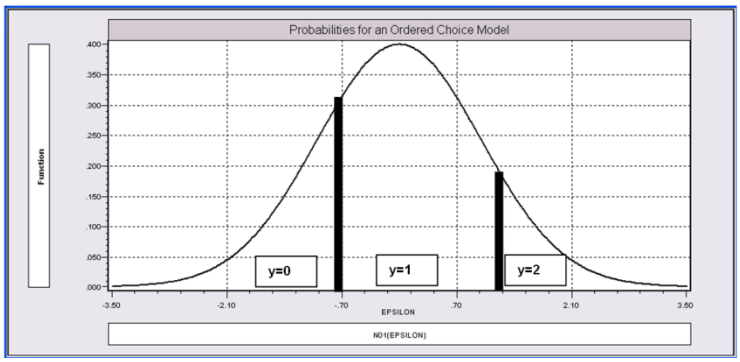

Figure 3.1 Underlying Probabilities for an Ordered Choice Model

The most recent literature (since 2005) includes several applications that use Bayesian methods to analyze ordered choices. Being heavily parametric in nature, they have focused exclusively on the ordered probit model.[1] Some commentary on Bayesian methods and methodology may be found in Koop and Tobias [2006]. Applications to the univariate ordered probit model include Kadam and Lenk [2008], Ando [2006], Zhang et al. [2007] and Tomoyuki and Akira [2006]. In the most basic cases, with diffuse priors, the "Bayesian" methods merely reproduce (with some sampling variability) the maximum likelihood estimator.[2] However, the MCMC methodology is often useful in settings which extend beyond the basic model, for example, applications to a bivariate ordered probit model (Biswas and Das [2002]), a model with autocorrelation (Czado et al. [2005] and Girard and Parent [2001]) and a model that contains a set of endogenous dummy variables in the latent regression (Munkin and Trivedi [2008].)

3.5 Application of the Ordered Choice Model to Self Assessed Health Status

Riphahn, Wambach and Million (RWM, 2003] analyzed individual data on health care utilization (doctor visits and hospital visits) using various models for counts. The data set is an unbalanced panel of 7,293 German households observed from 1 to 7 times for a total of 27,326 observations, extracted from the German Socioeconomic Panel (GSOEP). (See RWM [2003] and Greene [2008a] for discussion of the data set in detail.) Among the variables in this data set is HSAT, a self reported health assessment that is recorded with values 0,1,..,10 (so, $J = 10$).

[1] See Congdon [2005] for brief Bayesian treatment of an ordered logit model.
[2] In this connection, see Train [2003] and Wooldridge and Imbens [2009b] for discussion of the Bernstein – von Mises result.

Figure3.2 shows the distribution of outcomes for the full sample: The figure reports the variable NewHSAT, not the original variable. Forty of the 27,326 observations on HSAT in the original data were coded with noninteger values between 6.5 and 6.95. We have changed these 40 observations to 7s. In order to construct a compact example that is sufficiently general to illustrate the technique, we will aggregate the categories shown as follows: (0-2)=0, (3-5)=1, (6-8)=2, (9)=3, (10)=4. (One might expect collapsing the data in this fashion to sacrifice some information and, in turn, produce a less efficient estimator of the model parameters. See Murad et al. [2003] for some analysis of this issue.) Figure 3.3 shows the result, once again for the full sample, stratified by gender. The families were observed in 1984-1988, 1991 and 1995. For purposes of the application, to maintain as closely as possible the assumptions of the model, at this point, we have selected the most frequently observed year, 1988, for which there are a total of 4,483 observations, 2,313 males and 2,170 females. We will use the following variables in the regression part of the model,

$$\mathbf{x} = (\textit{constant}, \textit{Age}, \textit{Income}, \textit{Education}, \textit{Married}, \textit{Kids}).$$

In the original data set, *Income* is *HHNINC* (household income) and *Kids* is *HHKIDS* (dummy variable for children present in the household). *Married* and *Kids* are binary variables.

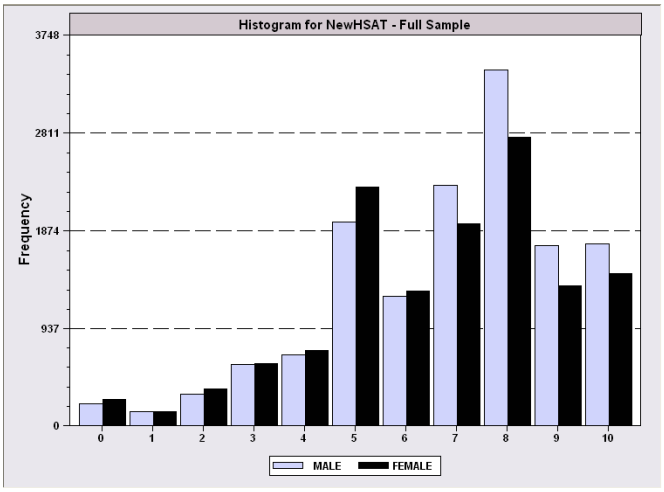

Figure 3.2 Self Reported Health Satisfaction

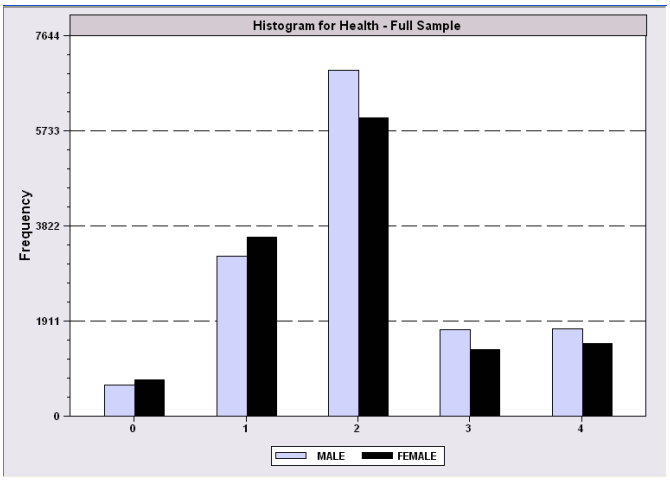

Figure 3.3 Health Satisfaction with Combined Categories

3.5.1 The Estimated Ordered Probit (Logit) Model

Table 3.1 presents estimates of the ordered probit and logit models for the 1988 data set. The estimates for the probit model imply

$$y^* = 1.97882 - .01806Age + .03556Educ + .25869Income - .03100Married + .06065Kids + \varepsilon.$$

$$\begin{aligned} y &= 0 \text{ if } y^* \leq 0 \\ y &= 1 \text{ if } 0 < y^* \leq 1.14835 \\ y &= 2 \text{ if } 1.14835 < y^* \leq 2.54781 \\ y &= 3 \text{ if } 2.54781 < y^* \leq 3.05639 \\ y &= 4 \text{ if } y^* > 3.05639. \end{aligned}$$

Figure 3.4 shows the implied model for a person of average age (43.44 years), education (11.418 years) and income (0.3487) who is married (1) with children (1). The figure shows the implied probability distribution in the population for individuals with these characteristics. As we will examine in the next section, the force of the regression model is that the probabilities change as the characteristics (**x**) change. In terms of the figure, changes in the characteristics induce changes in the placement of the partitions in the distribution and, in turn, in the probabilities of the outcomes.

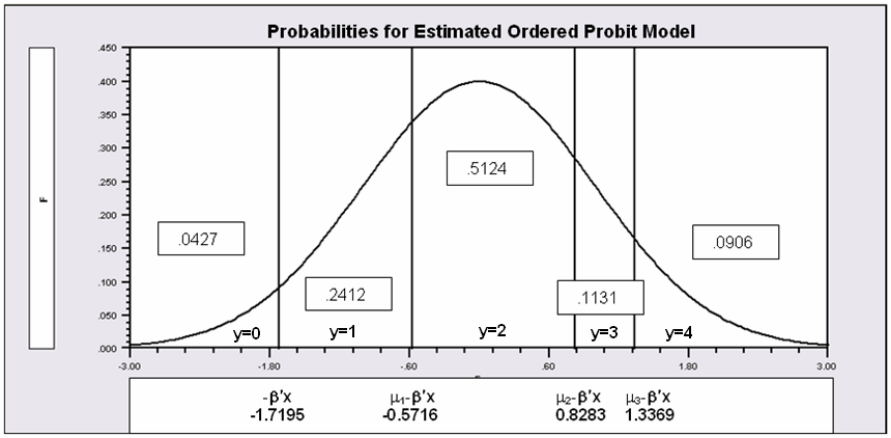

Figure 3.4 Estimated Ordered Probit Model

Table 3.1 Estimated Ordered Choice Models: Probit and Logit

Outcome	Frequency Count	Frequency Percent	Cumulative < = Count	Cumulative < = Percent	Cumulative > = Count	Cumulative > = Percent
HEALTH=00	230	5.1305	230	5.1305	4483	100.0000
HEALTH=01	1113	24.8271	1343	29.9576	4253	94.8695
HEALTH=02	2226	49.6542	3569	79.6119	3140	70.0424
HEALTH=03	500	11.1532	4069	90.7651	914	20.3881
HEALTH=04	413	9.2349	4483	100.0000	414	9.2349

	Ordered Logit				Ordered Probit				
	LogL = -5749.157				LogL = -5752.985				
	LogL0 = -5875.096				LogL0 = -5875.096				
	Chisq = 251.8798				Chisq = 244.2238				
	PseudoRsq = .0214362				PseudoRsq = .0207847				
Variable	Coef.	S.E.	t	P	Coef.	S.E.	t	P	Mean of X
Constant	3.5179	.2038	17.260	.0000	1.9788	.1162	17.034	.0000	1.0000
AGE	-.0321	.0029	-11.178	.0000	-.0181	.0016	-11.166	.0000	43.4401
EDUC	.0645	.0125	5.174	.0000	.0356	.0071	4.986	.0000	11.4181
INCOME	.4263	.1865	2.286	.0223	.2587	.1039	2.490	.0128	.34874
MARRIED	-.0645	.0746	-.865	.3868	-.0310	.0420	-.737	.4608	.75217
KIDS	.1148	.0669	1.717	.0861	.0606	.0382	1.586	.1127	.37943
Mu(1)	2.1213	.0371	57.249	.0000	1.1484	.0212	54.274	.0000	
Mu(2)	4.4346	.0390	113.645	.0000	2.5478	.0216	117.856	.0000	
Mu(3)	5.3771	.0520	103.421	.0000	3.0564	.0267	115.500	.0000	

3.5.2 Interpretation of the Model – Partial Effects and Scaled Coefficients

Interpretation of the coefficients in the ordered probit model is more complicated than in the ordinary regression setting.[1] The outcome variable, *y*, is merely a label for the ordered, non-quantitative outcomes. As such, there is no conditional mean function, $E[y|\mathbf{x}]$ to analyze. In order to interpret the parameters, one typically refers to the probabilities themselves. The partial effects in the ordered choice model are

$$\boldsymbol{\delta}_j(\mathbf{x}_i) = \frac{\partial \text{Prob}(y = j \mid \mathbf{x}_i)}{\partial \mathbf{x}_i} = \left[f(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i) - f(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) \right] \boldsymbol{\beta}.$$

Neither the sign nor the magnitude of the coefficient is informative about the result above, so the direct interpretation of the coefficients is fundamentally ambiguous. A counterpart result for a dummy variable in the model would be obtained by using a difference of probabilities, rather than a derivative.[2] That is, suppose *D* is a dummy variable in the model (such as *Married*) and γ is the coefficient on *D*. We would measure the effect of a change in *D* from 0 to 1 with all other variables held at the values of interest (perhaps their means) using

$$\Delta_j(D) = \left[F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i + \gamma) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i + \gamma) \right] - \left[F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i) \right].$$

The partial effects are shown in Table 3.2. Partial effects are computed using either the derivatives, or first differences for discrete variables;

$$\boldsymbol{\delta}_j(\mathbf{x}_i) = \frac{\partial \text{Prob}(y = j \mid \mathbf{x}_i)}{\partial \mathbf{x}_i} = \left[f(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i) - f(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) \right] \boldsymbol{\beta},$$

[1] See, e.g., Daykin and Moffatt [2002].
[2] See Boes and Winkelmann [2006a] and Greene [2008a, Chapter E22].

Table 3.2 Estimated Partial Effects for Ordered Choice Models

Summary of Marginal Effects for Ordered Probability Model						
Effects computed at means. Effects for binary variables are						
computed as differences of probabilities, other variables at means.						
	Probit			Logit		
Outcome	Effect	dPy<=nn/dX	dPy>=nn/dX	Effect	dPy<=nn/dX	dPy>=nn/dX
	Continuous Variable AGE					
Y = 00	.00173	.00173	.00000	.00145	.00145	.00000
Y = 01	.00450	.00623	-.00173	.00521	.00666	-.00145
Y = 02	-.00124	.00499	-.00623	-.00166	.00500	-.00666
Y = 03	-.00216	.00283	-.00499	-.00250	.00250	-.00500
Y = 04	-.00283	.00000	-.00283	-.00250	.00000	-.00250
	Continuous Variable EDUC					
Y = 00	-.00340	-.00340	.00000	-.00291	-.00291	.00000
Y = 01	-.00885	-.01225	.00340	-.01046	-.01337	.00291
Y = 02	.00244	-.00982	.01225	.00333	-.01004	.01337
Y = 03	.00424	-.00557	.00982	.00502	-.00502	.01004
Y = 04	.00557	.00000	.00557	.00502	.00000	.00502
	Continuous Variable INCOME					
Y = 00	-.02476	-.02476	.00000	-.01922	-.01922	.00000
Y = 01	-.06438	-.08914	.02476	-.06908	-.08830	.01922
Y = 02	.01774	-.07141	.08914	.02197	-.06632	.08830
Y = 03	.03085	-.04055	.07141	.03315	-.03318	.06632
Y = 04	.04055	.00000	.04055	.03318	.00000	.03318
	Binary(0/1) Variable MARRIED					
Y = 00	.00293	.00293	.00000	.00287	.00287	.00000
Y = 01	.00771	.01064	-.00293	.01041	.01327	-.00287
Y = 02	-.00202	.00861	-.01064	-.00313	.01014	-.01327
Y = 03	-.00370	.00491	-.00861	-.00505	.00509	-.01014
Y = 04	-.00491	.00000	-.00491	-.00509	.00000	-.00509
	Binary(0/1) Variable KIDS					
Y = 00	-.00574	-.00574	.00000	-.00511	-.00511	.00000
Y = 01	-.01508	-.02081	.00574	-.01852	-.02363	.00511
Y = 02	.00397	-.01684	.02081	.00562	-.01801	.02363
Y = 03	.00724	-.00960	.01684	.00897	-.00904	.01801
Y = 04	.00960	.00000	.00960	.00904	.00000	.00904

or $$\Delta_j(d,\mathbf{x}_i) = \left[F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i + \gamma) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i + \gamma)\right] - \left[F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i)\right].$$

Since these are functions of the estimated parameters, they are subject to sampling variability and one might desire to obtain appropriate asymptotic covariance matrices and/or confidence intervals. For this purpose, the partial effects are typically computed at the sample means. The delta method is used to obtain the standard errors. Let $\mathbf{V}$ denote the estimated asymptotic covariance matrix for the $(K+J-2)\times 1$ parameter vector $\left(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\mu}}'\right)'$. Then, the estimator of the asymptotic covariance matrix for each vector of partial effects is

$$\mathbf{Q} = \hat{\mathbf{C}}\mathbf{V}\hat{\mathbf{C}}', \text{ where } \hat{\mathbf{C}} = \left[\frac{\partial \hat{\boldsymbol{\delta}}_j(\bar{\mathbf{x}})}{\partial \hat{\boldsymbol{\beta}}'} \quad \frac{\partial \hat{\boldsymbol{\delta}}_j(\bar{\mathbf{x}})}{\partial \hat{\boldsymbol{\mu}}'}\right].$$

The appropriate row of $\hat{\mathbf{C}}$ is replaced with the derivatives of $\Delta_j(d, \bar{\mathbf{x}})$ when the effect is being computed for a discrete variable.

The implication of the preceding result is that the effect of a change in one of the variables in the model depends on all the model parameters, the data, and which probability (cell) is of interest. It can be negative or positive. To illustrate, we consider a change in the education variable on the implied probabilities in Figure 3.5. Since the changes in a probability model are typically 'marginal' (small), we will exaggerate the effect a bit so that it will show up in a figure. Consider, then, the average individual shown in the top panel Figure 3.5, except now, with a Ph.D. (college plus four years of postgraduate work). That is, 20 years of education, instead of the average 11.4 used earlier. The effect of an additional 8.6 years of education is shown in the lower panel of Figure 3.5. All five probabilities have changed. The two at the right end of the distribution have increased while the three at the left have decreased.

The partial effects give the impacts on the specific probabilities per unit change in the stimulus or regressor. For example, for continuous variable *Educ*, we find partial effects for the ordered probit model for the five cells of -.0034, -.00885, .00244, .00424, .00557, respectively, which give the expected change on the probabilities per additional year of education. For the income variable, for the highest cell, the estimated partial effect is .04055. However, some care is needed in interpreting this in terms of a unit change. The income variable has a mean of 0.34874 and a standard deviation of 0.1632. A full unit change in income would put the average individual nearly six standard deviations above the mean. Thus, for the marginal impact of income, one might want to measure a change in standard deviation units. Thus, an assessment of the impact of a change in income on the probability of the highest cell probability might be 0.04055×0.1632 = 0.00662. Precisely how this computation should be done will vary from one application to another.

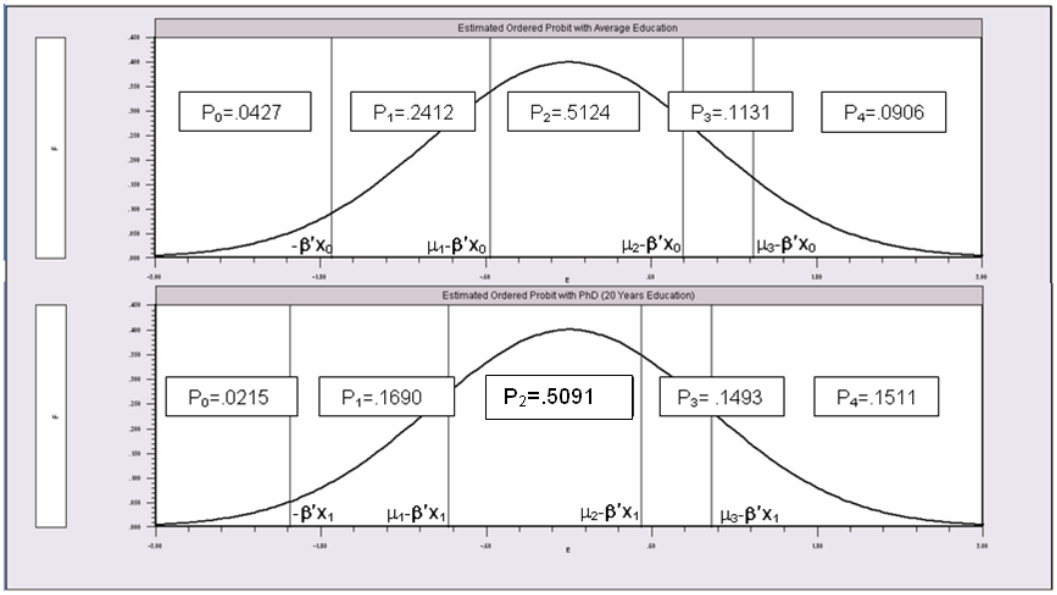

Figure 3.5 Partial Effect in Ordered Probit Model

Note in Table 3.1 there is a large difference in the coefficients obtained for the probit and logit models. The logit coefficients are roughly 1.8 times as large (not uniformly). This difference, which will always be observed, points up one of the risks in attempting to interpret directly the coefficients in the model. This difference reflects an inherent scaling of the underlying variable and in the shape of the distributions. The difference can be traced back (at least in part) to the different underlying variances in the two models. In the probit model, $\sigma_\varepsilon = 1$; in the logit model $\sigma_\varepsilon = \pi/\sqrt{3} = 1.81$. The models are roughly preserving the ratio β/σ_ε in the estimates. The difference is greatly diminished in the partial effects reported in Table 3.2. That

is the virtue of the scaling done to compute the partial effects. The inherent characteristics of the model are essentially the same for the two functional forms.

4 Specification Issues and Generalized Models

It is useful to distinguish between two directions of the contemporary development of the ordered choice model, functional form and heterogeneity. Beginning with Terza [1985], a number of authors have focused on the fact that the model does not account adequately for individual heterogeneity that is likely to be present in micro- level data. This section will consider specification issues. Heterogeneity is examined in Section 5.

4.1 Accommodating Individual Heterogeneity

For a *subjective* well being (SWB) application, the right hand side of the behavioral equation will include variables such as *Income*, *Education*, *Marital Status*, *Children*, *Working Status*, *Health*, and a host of other measurable and unmeasurable, and *measured* and *unmeasured* variables. In individual level behavioral models, such as

$$SWB_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},$$

the relevant question is whether a zero mean, homoscedastic ε_{it}, can be expected to satisfactorily accommodate the likely amount of heterogeneity in the underlying data, and whether it is reasonable to assume that the same thresholds should apply to each individual.

Beginning with Terza [1985], analysts have questioned the adequacy of the ordered choice model from this perspective. As shown below, many of the proposed extensions of the model, such as heteroscedasticity, parameter heterogeneity, etc., parallel developments in other modeling contexts (such as binary choice modeling and modeling counts such as number of doctor visits or hospital visits). The regression based ordered choice model analyzed here does have a unique feature, that the thresholds are part of the behavioral specification. This aspect of the specification has been considered as well.

4.2 Threshold Models – A Generalized Ordered Probit Model

The model analyzed thus far assumes that the thresholds μ_j are the same for every individual in the sample. Terza [1985], Pudney and Shields [2000], Boes and Winkelmann [2006a], Greene, Harris, Hollingsworth and Maitra [2008] and Greene and Hensher [2009], all present cases that suggest individual variation in the set of thresholds is a degree of heterogeneity that is likely to be present in the data, but is not accommodated in the model. Terza's [1985] generalization of the model is equivalent to

$$\mu_{ij} = \mu_j + \boldsymbol{\delta}'\mathbf{z}_i.$$

This is the special case of the 'generalized' model used in his application – his fully general case allows $\boldsymbol{\delta}$ to differ across outcomes. The model is reformulated later to assume that the $\mathbf{z}_i$ in the equation for the thresholds is the same as the $\mathbf{x}_i$ in the regression. For the moment, it is convenient to isolate the constant term from $\mathbf{x}_i$. In Terza's application, in which there were three outcomes,

$$y_i^* = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

and

$$
\begin{aligned}
y_i \;=\;& 0 \text{ if } y_i^* \le 0,\\
& 1 \text{ if } 0 < y_i^* \le \mu + \boldsymbol{\delta}'\mathbf{x}_i,\\
& 2 \text{ if } y_i^* > \mu + \boldsymbol{\delta}'\mathbf{x}_i.
\end{aligned}
$$

There is an ambiguity in the model as specified. In principle, the model for three outcomes has two thresholds, μ_0 and μ_1. With a nonzero overall constant, it is always necessary to normalize the first, $\mu_0 = 0$. Therefore, the model implies the following probabilities:

$$
\begin{aligned}
\text{Prob}(y = 0|\mathbf{x}) &= \Phi(-\alpha - \boldsymbol{\beta}'\mathbf{x}) &&= 1 - \Phi(\alpha_0 + \boldsymbol{\beta}_0'\mathbf{x}),\\
\text{Prob}(y = 1|\mathbf{x}) &= \Phi(\mu + \boldsymbol{\delta}'\mathbf{x}_i - \alpha - \boldsymbol{\beta}'\mathbf{x}) - \Phi(-\alpha - \boldsymbol{\beta}'\mathbf{x}) &&= \Phi(\alpha_0 + \boldsymbol{\beta}_0'\mathbf{x}) - \Phi(\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}),\\
\text{Prob}(y = 2|\mathbf{x}) &= \Phi(\alpha + \boldsymbol{\beta}'\mathbf{x} - \mu - \boldsymbol{\delta}'\mathbf{x}) &&= \Phi(\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}),
\end{aligned}
$$

where $\alpha_0 = \alpha$, $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$, $\alpha_1 = \alpha - \mu$, $\boldsymbol{\beta}_1 = (\boldsymbol{\beta} - \boldsymbol{\delta})$. This is precisely Williams's [2006] "Generalized Ordered Probit Model." That is, at this juncture, Terza's heterogeneous thresholds model and the "generalized ordered probit" model are indistinguishable. For direct applications of Terza's approach, see, e.g., Kerkhofs and Lindeboom [1995], Groot and van den Brink [1999] and Lindeboom and van Doorslayer [2003].

Terza notes (on p. 6) that the model formulation does not impose an ordering on the threshold coefficients. He suggests an inequality constrained maximization of the log likelihood, which is likely to be extremely difficult if there are many variables in $\mathbf{x}$. As a "less rigorous but apparently effective remedy," he proposes to drop from the model variables in the threshold equations that are insignificant in the initial (unconstrained) model.

The analysis of this model continues with Pudney and Shields's [2000] "Generalized Ordered Probit Model," whose motivation, like Terza's was to accommodate *observable* individual heterogeneity in the threshold parameters as well as in the mean of the regression. (Pudney and Shields studied an example in the context of job promotion in which the steps on the promotion ladder for nurses are somewhat individual specific. In their setting, in contrast to Terza's, at least some of the variables in the threshold equations are explicitly different from those in the regression. Their model is parameterized as

$$
\Pr(y_i = g|\mathbf{x}_i,\mathbf{q}_i,t_i) = \Phi[\mathbf{q}_i'\boldsymbol{\beta}_g - \mathbf{x}_i(\boldsymbol{\alpha} + \boldsymbol{\delta}_g)] - \Phi[\mathbf{q}_i'\boldsymbol{\beta}_{g-1} - \mathbf{x}_i(\boldsymbol{\alpha} + \boldsymbol{\delta}_{g-1})].
$$

The resulting equation is now a hybrid with outcome varying parameters in both thresholds and in the regression. The test of threshold constancy is then carried out simply by testing (using an LM test) the null hypothesis that $\boldsymbol{\delta}_g = \mathbf{0}$ for all $\mathbf{g}$. (A normalization, $\boldsymbol{\delta}_0 = \boldsymbol{\delta}_m = \mathbf{0}$, is imposed at the outset.)

Two features of Pudney and Shields' model to be noted are: First, the probabilities in their revised log likelihood [their equation (8)], are not constrained to be positive. Second, the thresholds, $\mathbf{q}_i\boldsymbol{\beta}_g$, are not constrained to be ordered. No restriction on $\boldsymbol{\beta}_g$ will ensure that $\mathbf{q}_i'\boldsymbol{\beta}_g > \mathbf{q}_i'\boldsymbol{\beta}_{g-1}$ for all data vectors $\mathbf{q}_i$.

The equivalence of the Terza and Williams models is only a mathematical means to the end of estimation of the model. The Pudney and Shields model, itself, has constant parameters in the regression model and outcome varying parameters in the thresholds, and clearly stands on the platform of the latent regression. They do note, however, (using a more generic notation) a deeper problem of identification). However it is originally formulated, the model implies that

$$
\text{Prob}[y_i \le j \mid \mathbf{x}_i,\mathbf{z}_i] = F(\mu_j + \boldsymbol{\delta}'\mathbf{z}_i - \boldsymbol{\beta}'\mathbf{x}_i) = F[\mu_j - (\boldsymbol{\delta}^{*\prime}\mathbf{z}_i + \boldsymbol{\beta}'\mathbf{x}_i)],\ \boldsymbol{\delta}^* = -\boldsymbol{\delta}.
$$

In their specification, they had a well defined distinction between the variables, $\mathbf{z}_i$ that should appear only in the thresholds and $\mathbf{x}_i$ that should appear in the regression. More generally, it is less

than obvious whether the variables $\mathbf{z}_i$ are actually in the threshold or in the mean of the regression. Either interpretation is consistent with the estimable model. Pudney and Shields argue that the distinction is of no substantive consequence for their analysis. The consequence is at the theoretical end, not in the implementation. But, this entire development is necessitated by the linear specification of the thresholds. Absent that, most of the preceding construction is of limited relevance.

4.3 Random Parameters Models

Formal modeling of heterogeneity in the parameters as representing a feature of the underlying data, appears in Greene [2002] (version 8.0), Bhat [1999], Bhat and Zhao [2002] and Boes and Winkelmann [2006]. These treatments suggest a full random parameters (RP) approach to the model.

Boes and Winkelmann's [2006] treatment appears as follows:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i,$$

where $\mathbf{u}_i \sim N[\mathbf{0},\boldsymbol{\Omega}]$. Inserting the expression for $\boldsymbol{\beta}_i$ in the latent regression model, we obtain

$$\begin{aligned} y_i^* &= \boldsymbol{\beta}_i'\mathbf{x}_i + \varepsilon_i \\ &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i + \mathbf{x}_i'\mathbf{u}_i. \end{aligned}$$

They propose treating this as a heteroscedastic model – $\text{Var}[\varepsilon_i + \mathbf{x}_i'\mathbf{u}_i] = 1 + \mathbf{x}_i'\boldsymbol{\Omega}\mathbf{x}_i$ – and maximizing the log likelihood directly over $\boldsymbol{\beta}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$. The observation mechanism is the same as earlier. Greene [2002, 2007a, 2008a,b] analyzes the same model, but estimates the parameters by maximum simulated likelihood. First, write the random parameters as

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \mathbf{L}\mathbf{D}\mathbf{w}_i, \tag{8.2}$$

where $\mathbf{w}_i$ has a multivariate standard normal distribution, and $\mathbf{L}\mathbf{D}^2\mathbf{L}' = \boldsymbol{\Omega}$. The Cholesky matrix, $\mathbf{L}$, is lower triangular with ones on the diagonal. The below diagonal elements of $\mathbf{L}$, λ_{mn}, produce the nonzero correlations across parameters. The diagonal matrix, $\mathbf{D}$, provides the scale factors, δ_m, i.e., the standard deviations of the random parameters. The end result is that $\mathbf{L}(\mathbf{D}\mathbf{w}_i)$ is a mixture, $\mathbf{L}\mathbf{w}_i^*$ of random variables, w_{im}^* which have variances δ_m^2. This is a two level 'hierarchical' model (in the more widely used sense). The probability for an observation is

$$\begin{aligned} \text{Prob}(y_i=j|\mathbf{x}_i,\mathbf{w}_i) &= \left[\Phi\left(\mu_j - \boldsymbol{\beta}_i'\mathbf{x}_i\right) - \Phi\left(\mu_{j-1} - \boldsymbol{\beta}_i'\mathbf{x}_i\right)\right] \\ &= \left[\begin{matrix}\Phi\left(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_i)'\mathbf{x}_i\right) - \\ \Phi\left(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_i)'\mathbf{x}_i\right)\end{matrix}\right]. \end{aligned}$$

In order to maximize the log likelihood, we must first integrate out the elements of the unobserved $\mathbf{w}_i$. Thus, the contribution to the unconditional log likelihood for observation i is

$$\log L_i = \log \int_{\mathbf{w}_i} \left[\begin{matrix}\Phi\left(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_i)'\mathbf{x}_i\right) - \\ \Phi\left(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_i)'\mathbf{x}_i\right)\end{matrix}\right] F(\mathbf{w}_i)d\mathbf{w}_i.$$

The log likelihood for the sample is then the sum over the observations. Computing the integrals is an obstacle that must now be overcome. It has been simplified considerably already by decomposing $\boldsymbol{\Omega}$ explicitly in the log likelihood, so that $F(\mathbf{w}_i)$ is the multivariate standard normal density. The *Stata* routine, GLAMM [Rabe-Hesketh, Skrondal and Pickles [2005]) that is used for some discrete choice models does the computation using a form of Hermite quadrature. An alternative, generally substantially faster method of maximizing the log likelihood is maximum simulated likelihood. The integration is replaced with a simulation over R draws from the multivariate standard normal population. The simulated log likelihood is

$$\log L_S = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \begin{bmatrix} \Phi\left(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_{ir})'\mathbf{x}_i\right) - \\ \Phi\left(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_{ir})'\mathbf{x}_i\right) \end{bmatrix}.$$

The simulations are speeded up considerably by using Halton draws.[1] Partial effects and predicted probabilities must be simulated as well. For the partial effects,

$$\frac{\partial \text{Prob}(y_i = j \mid \mathbf{x}_i)}{\partial \mathbf{x}_i} = \int_{\mathbf{w}_i} \begin{bmatrix} \phi\left(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_i)'\mathbf{x}_i\right) - \\ \phi\left(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i - \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_i - (\mathbf{L}\mathbf{D}\mathbf{w}_i)'\mathbf{x}_i\right) \end{bmatrix} (\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i - \mathbf{L}\mathbf{D}\mathbf{w}_i)\, F(\mathbf{w}_i) d\mathbf{w}_i.$$

we use simulation to compute

$$Est.\frac{\partial \text{Prob}(y_i = j \mid \mathbf{x}_i)}{\partial \mathbf{x}_i} = \left\{ \frac{1}{R} \sum_{r=1}^{R} \begin{bmatrix} \phi\left(\hat{\mu}_{j-1} - \hat{\boldsymbol{\beta}}'\mathbf{x}_i - \mathbf{z}_i'\hat{\boldsymbol{\Delta}}'\mathbf{x}_i - (\hat{\mathbf{L}}\hat{\mathbf{D}}\mathbf{w}_{ir})'\mathbf{x}_i\right) - \\ \phi\left(\hat{\mu}_j - \hat{\boldsymbol{\beta}}'\mathbf{x}_i - \mathbf{z}_i'\hat{\boldsymbol{\Delta}}'\mathbf{x}_i - (\hat{\mathbf{L}}\hat{\mathbf{D}}\mathbf{w}_{ir})'\mathbf{x}_i\right) \end{bmatrix} \right\} (\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i - \hat{\mathbf{L}}\hat{\mathbf{D}}\mathbf{w}_{ir}).$$

A similar analysis provides an extension of the latent class model to ordered choice models. The latent class ordered choice model is developed in detail in Greene and Hensher [2010].

5 Ordered Choice Modeling with Panel Data

Development of models for panel data parallels that in other modeling settings. The departure point is the familiar fixed and random effects approaches. We then consider other types of applications including extensions of the random parameters and latent classes formulations, dynamic models and some special treatments that accommodate features peculiar to the ordered choice models.

5.1 Ordered Choice Models with Fixed Effects

[1] See Halton [1970] for the general principle, and Bhat [2001, 2003] and Train [2003] for applications in the estimation of 'mixed logit models') rather than random draws. Further details on this method of estimation are also given in Greene [2007, 2008a].

An ordered choice model with fixed effects formulated in the most familiar fashion would be

$$\text{Prob}[y_{it} = j \mid \mathbf{x}_i] = F(\mu_j - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_{it}) - F(\mu_{j-1} - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_{it}) > 0, j = 0,1,...,J.$$

At the outset, there are two problems that this model shares with other nonlinear fixed effects models. First, regardless of how estimation and analysis are approached, time invariant variables are precluded. Since social science applications typically include demographic variables such as gender and, for some at least, education level, that are time invariant, this is likely to be a significant obstacle. (Several of the variables in the GSOEP analyzed by Boes and Winkelmann [2006b] and others are time invariant.) Second, there is no sufficient statistic available to condition the fixed effects out of the model. That would imply that in order to estimate the model as stated, one must maximize the full log likelihood,

$$\log L = \sum_{i=1}^{N} \log \left\{ \prod_{t=1}^{T_i} \left(\sum_{j=0}^{J} m_{ijt} \left[\Phi\left(\mu_j - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_{it}\right) - \Phi\left(\mu_{j-1} - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_i\right) \right] \right) \right\}.$$

If the sample is small enough, one may simply insert the individual group dummy variables and treat the entire pooled sample as a cross section. See, e.g., Mora [2006] for a cross-country application in banking that includes separate country dummy variables. We are interested, instead, in the longitudinal data case in which this would not be feasible. The data set from which our sample used in the preceding examples is extracted comes from an unbalanced panel of 7,293 households, observed from 1 to 7 times each. The full ordered probit model with fixed effects, including the individual specific constants, can be estimated by unconditional maximum likelihood using the results in Greene [2004a,b and 2008a, Section 16.9.6.c]. The likelihood function is globally concave, so despite its superficial complexity, the estimation is straightforward.[1]

The larger methodological problem with this approach would be at least the potential for the incidental parameters problem that has been widely documented for the binary choice case. [See, e.g., Lancaster [2000).] That is the small *T* bias in the estimated parameters when the full MLE is applied in panel data. For $T = 2$ in the binary logit model, it has been shown analytically [Abrevaya [1997]) that the full MLE converges to $2\boldsymbol{\beta}$. [See, as well, Hsiao [1986, 2003].) No corresponding results have been obtained for larger *T* or for other models. In particular, no theoretical counterpart to the Hsiao [1986, 2003) and Abrevaya [1997] result on the small *T* bias (incidental parameters problem) of the MLE in the presence of fixed effects has been derived for the ordered probit model, even for *T* equal to 2. However, Monte Carlo results have strongly suggested that the small sample bias persists for larger *T* as well, though as might be expected, it diminishes with increasing T. The Monte Carlo results in Greene [2004b] suggest that biases comparable to those in the binary choice models persist in the ordered probit model as well. The values given correspond to estimation of coefficients on a continuous variable (β) and a binary variable (δ) in the equation

Recent proposals for "bias reduction" estimators for binary choice models, including Fernandez-Val and Vella [2007], Fernandez-Val [2008], Carro [2007], Hahn and Newey [2004] and Hahn and Kuersteiner [2003] suggest some directions for further research. However, no counterparts for the ordered choice models have yet been developed. We would note, for this model, the estimation of $\boldsymbol{\beta}$ which is the focus of these estimators, is only a means to the end. As seen earlier, in order to make meaningful statements about the implications of the model for

[1] See Pratt [1981] and Burridge [1981].

behavior, it will be necessary to compute probabilities and derivatives. These, in turn, will require estimation of the constants, or some surrogates. The problem remains to be solved.

5.2 Ordered Choice Models with Random Effects

Save for an ambiguity about the mixture of distributions in an ordered logit model, a random effects version of the ordered choice model is a straightforward extension of the binary choice case developed by Butler and Moffitt [1982]. An interesting application which appears to replicate, but not connect to Butler and Moffitt is Jansen [1990]. Jansen estimates the equivalent of the Butler and Moffitt model with an ordered probit model, using an iterated MLE with quadrature used between iterations.

The structure of the random effects ordered choice model is

$$
\begin{aligned}
y_{it}^* &= \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it}, \\
y_{it} &= j \text{ if } \mu_{j-1} \leq y_{it}^* < \mu_{it}, \\
\varepsilon_{it} &\sim f(.) \text{ with mean zero and constant variance 1 or } \pi^2/3 \text{ (probit or logit)}, \\
u_i &\sim g(.) \text{ with mean zero and constant variance, } \sigma^2, \text{ independent of } \varepsilon_{it} \text{ for all } t.
\end{aligned}
$$

If we maintain the ordered probit form and assume as well that u_i is normally distributed, then, at least superficially, we can see the implications for the estimator of ignoring the heterogeneity. Using the usual approach,

$$
\begin{aligned}
\text{Prob}(y_{it} = j|\mathbf{x}_{it}) &= \text{Prob}(\boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} < \mu_j) - \text{Prob}(\boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} < \mu_{j-1}) \\
&= \Phi\left(\frac{\mu_j}{\sqrt{1+\sigma^2}} - \frac{\boldsymbol{\beta}'\mathbf{x}_{it}}{\sqrt{1+\sigma^2}}\right) - \Phi\left(\frac{\mu_{j-1}}{\sqrt{1+\sigma^2}} - \frac{\boldsymbol{\beta}'\mathbf{x}_{it}}{\sqrt{1+\sigma^2}}\right) \\
&= \Phi(\tau_j - \boldsymbol{\gamma}'\mathbf{x}_{it}) - \Phi(\tau_{j-1} - \boldsymbol{\gamma}'\mathbf{x}_{it}).
\end{aligned}
$$

Unconditionally, then, the result is an ordered probit in the scaled threshold values and scaled coefficients. Evidently, this is what is estimated if the data are pooled and the heterogeneity is ignored.[1] Wooldridge and Imbens [2009c] argue that since the partial effects are $[\phi(\tau_{j-1} - \boldsymbol{\gamma}'\mathbf{x}_{it}) - \phi(\tau_j - \boldsymbol{\gamma}'\mathbf{x}_{it})]\boldsymbol{\gamma}$, the scaled version of the parameter is actually the object of estimation in any event.

5.3 Spatial Autocorrelation

The treatment of spatially correlated discrete data presents several major complications. LeSage [1999, 2004] presents some of the methodological issues. A variety of received applications for binary choice include the geographic pattern of state lotteries (Coughlin, Garrett and Hernandez-Murillo [2004]), Children's Health Insurance Programs (CHIPS) (Franzese and Hays [2007]) and HYV rice adoption (Holloway, Shankar and Rahman [2002]). The extension to ordered choice models has begun to emerge as well, with applications including ozone concentration and land development (Wang and Kockelman [2008, 2009]) and trip generation (Roorda, Páez, Morency, Mercado, and Farber [2009]).

[1] See Wooldridge [2002]. Note that a "robust" covariance matrix estimator does not redeem the estimator. It is still inconsistent.

6 Two Part and Sample Selection Models

Two part models describe situations in which the ordered choice is part of a two stage decision process. In a typical situation, an individual decides whether or not to participate in an activity then, if so, decides how much. The first decision is a binary choice. The intensity outcome can be of several types – what interests us here is an ordered choice. In the example below, an individual decides whether or not to be a smoker. The intensity outcome is how much they smoke. The sample selection model is one in which the participation "decision" relates to whether the data on the outcome variable will be observed, rather than whether the activity is undertaken. This chapter will describe several types of two part and sample selection models

6.1 Inflation Models

Harris and Zhao [2007] analyzed a sample of 28,813 Australian individuals' responses to the question "How often do you now smoke cigarettes, pipes or other tobacco products?" [Data are from the Australian National Drug Strategy Household Survey, NDSHS [2001].) Responses were "zero, low, moderate, high," coded 0,1,2,3. The sample frequencies of the four responses were 0.75, 0.04, 0.14 and 0.07. The spike at zero shows a considerable excess of zeros compared to what might be expected in an ordered choice model. The authors reason that there are numerous explanations for a zero response: "genuine nonsmokers, recent quitters, infrequent smokers who are not currently smoking and potential smokers who might smoke when, say, the price falls." It is also possible that the zero response includes some individuals who prefer to identify themselves as nonsmokers. The question is ambiguously worded, but arguably, the group of interest is the genuine nonsmokers. This suggests a type of latent class arrangement in the population. There are (arguably) two types of zeros, the one of interest, and another type generated by the appearance of the respondent in the latent class of people who respond zero when another response would actually be appropriate. The end result is an inflation of the proportion of zero responses in the data. A "Zero Inflation" model is proposed to accommodate this failure of the base case model. In a recent application, Greene, Harris and Hollingsworth [2012] have extended an ordered probit model of self assessed health (on a zero to four scale) to accommodate "2s and 3s inflation."

6.2 Sample Selection Models

The familiar sample selection model was extended to binary choice models by Wynand and van Praag [1981] and Boyes, Hoffman and Low [1989]. A variety of extensions have also been developed for ordered choice models, both as sample selection (regime) equations and as models for outcomes subject, themselves, to sample selectivity. We consider these two cases and some related extensions.

The models of sample selectivity in this area are built as extensions of Heckman's [1979] canonical model. Estimation of the regression equation by least squares while ignoring the selection issue produces biased and inconsistent estimators of all the model parameters. Estimation of this model by two step methods is documented in a voluminous literature, including Heckman [1979] and Greene [2008a]. The two step method involves estimating $\boldsymbol{\alpha}$ first in the participation equation using an ordinary probit model, then computing an estimate of λ_i, $\hat{\lambda}_i = \phi\left(\hat{\boldsymbol{\beta}}'\mathbf{x}_i\right)/\Phi\left(\hat{\boldsymbol{\beta}}'\mathbf{x}_i\right)$, for each individual in the selected sample. At the second step, an estimate of $(\boldsymbol{\beta},\theta)$ is obtained by linear regression of y_i on $\mathbf{x}_i$ and $\hat{\lambda}_i$. Necessary corrections to the estimated

standard errors are described in Heckman [1979], Greene [1981,2008b], and, in general terms, in Murphy and Topel [2002].

Consider a model of educational attainment or performance in a training or vocational education program (e.g., low, median, high), with selection into the program as an observation mechanism. [Boes [2007] examines a related case, that of a treatment, *D* that acts as an endogenous dummy variable in the ordered outcome model.] In an ordered choice setting, the "second step" model is nonlinear. The received literature contains many applications in which authors have "corrected for selectivity" by following the logic of the Heckman two step estimator, that is, by constructing $\lambda_i = \phi(\boldsymbol{\alpha}'\mathbf{w}_i)/\Phi(\boldsymbol{\alpha}'\mathbf{w}_i)$ from an estimate of the probit selection equation and adding it to the outcome equation.[1] However, this is only appropriate in the linear model with normally distributed disturbances. An explicit expression, which does not involve an inverse Mills ratio, for the case in which the unconditional regression is $E[y|\mathbf{x},\varepsilon] = \exp(\boldsymbol{\beta}'\mathbf{x} + \varepsilon)$ is given in Terza [1998]. A template for nonlinear single index function models subject to selectivity is developed in Terza [1998] and Greene [2006, 2008a, Sec. 24.5.7]. Applications specifically to the Poisson regression appear in several places, including Greene [1995, 2005]. The general case typically involves estimation either using simulation or quadrature to eliminate an integral involving *u* in the conditional density for *y*. Cases in which both variables are discrete, however, are somewhat simpler. A near parallel to the model above is the bivariate probit model with selection developed by Boyes, Hoffman and Low [1989] in which the outcome equation above would be replaced with a second probit model. [Wynand and van Praag [1981] proposed the bivariate probit/selection model, but used the two step approach rather than maximum likelihood.] The log likelihood function for the bivariate probit model is given in Boyes et al. [1989] and Greene [2008a, p. 896]. A straightforward extension of the result provides the log likelihood for the ordered probit case.

Essentially this model is applied in Popuri and Bhat [2003] to a sample of individuals who chose to telecommute ($z = 1$) or not ($z = 0$) then, for those who do telecommute, the number of days that they do. We note two aspects of this application that do depart subtly the sample selection application: (1) the application would more naturally fall into the category of a hurdle model composed of a participation equation and an activity equation given the decision to participate – in the latter, it is known that the activity level is positive.[2] Thus, unlike the familiar choice case, the zero outcome is not possible here. (2) The application would fit more appropriately into the sample selection or hurdle model frameworks for count data such as the Poisson model.[3] Bricka and Bhat [2006] is a similar application applied to a sample of individuals who did ($z=1$) or did not ($z = 0$) underreport the number of trips in a travel based survey. The activity equation is the number of trips underreported for those who did. This study, like its predecessor could be framed in a hurdle model for counts, rather than an ordered choice model.

7 Conclusions

The preceding has developed the standard model for ordered choices as typically analyzed in social science applications (e.g., Johnson and Albert[199]). (There is a parallel, but markedly different stream of literature in biometrics discussed in some detail in Greene and Hensher [2010] and references noted.) Several model extensions, such as outcomes inflation, and specification issues such as modeling heterogeneity are noted as well. These are developed in greater detail in recent surveys such as Boes and Winkelmann [2006a], Greene and Hensher

[1] See, e.g., Greene [1994]. Several other examples are provided in Greene [2008b].
[2] See Cragg [1971] and Mullahy [1986].
[3] See, again, Mullahy [1986], Terza [1994], Greene [1995] and Greene [2007a].

[2010] and Daykin and Moffatt [2002]. Ongoing development, such as nonparametric and Bayesian approaches are noted with some pointers to recent literature is suggested in Greene and Hensher [2010].

References

Abrevaya, J., 1997. "The Equivalence of Two Estimators of the Fixed Effects Logit Model," *Economics Letters*, 55, pp. 41-43.

Agostino, A., C. Bhat and E. Pas, 1996. "A Random Effects Multinomial Probit Model of Car Ownership Choice," *Proceedings of the Third Workshop on Bayesian Statistics in Science and Technology*, Cambridge University Press.

Agresti, A., 2002. *Categorical Data Analysis 2nd Ed.*, New York, John Wiley and Sons.

Aguemang-Duah, K. and F. Hall, 1997. "Spatial Transferability of an Ordered Response Model of Trip Generation," *Transport Research – Series A*, 31, 5, 389-402.

Aitcheson, J. and S. and Silvey, 1957. "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika*, 44, pp. 131-140.

Ando, T., 2006. "Bayesian credit rating analysis based on ordered probit regression model with functional predictor," *Proceeding of The Third IASTED International Conference on Financial Engineering and Applications*, 69-76.

Basu, D. and R. de Jong, 2006. "Dynamic Multinomial Ordered Choice With An Application to the Estimation of Monetary Policy Rules," Department of Economics, Ohio State University, Manuscript.

Bhat, C., 1996. "A Generalized Multiple Durations Proportional Hazard Model with an Application to Activity Behavior During the Work-to-Home Commute," *Transportation Research Part B*, 30, 465-480.

Bhat, C., 1997. "Work Travel Mode Choice and Number of Nonwork Commute Stops," *Transportation Research Part B*, 31(1), 41-54.

Bhat, C., 1999. "An Analysis of Evening Commute Stop-Making Behavior Using Repeated Choice Observations from a Multi-Day Survey," *Transportation Research Part B*, 33(7), 495-510.

Bhat, C., 2001. "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research Part B*, 35(7), 677-693.

Bhat, C., 2003. "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences," *Transportation Research Part B*, 37(9), 837-855.

Bhat, C., J. Carini and R. Misra, 1999. "Modeling the Generation and Organization of Household Activity Stops. *Transportation Research Record*, 1676, 153-161.

Bhat, C. and F. Koppelman, 1993. An Endogenous Switching Simultaneous Equation System of Employment, Income and Car Ownership, *Transportation Research A*, 27, 447-459.

Bhat, C. and V. Pulugurta, 1998. "A Comparison of Two Alternative Behavioral Mechanisms for Car Ownership Decisions," *Transportation Research Part B*, 32(1), 61-75.

Bhat, C. and H. Zhao, 2002. "The Spatial Analysis of Activity Stop Generation," *Transportation Research Part B*, 36(6), 557-575.

Biswas and Das, 2002. **"**A Bayesian Analysis of Bivariate Ordinal Data: Wisconsin Epidemiologic Study of Diabetic Retinopathy Revisited, *Statistics in Medicine*, 21, 4, pp. 549-559.

Boes, S., 2007. "Nonparametric Analysis of Treatment Effects in Ordered Response Models," University of Zurich, Socioeconomic Institute, Working Paper 0709.

Boes, S. and R. Winkelmann, 2004. "Income and Happiness: New Results from Generalized Threshold and Sequential Models," IZA Discussion Paper No. 1175, SOI Working Paper 0407, IZA

Boes, S. and R. Winkelmann, 2006a. "Ordered Response Models," *Allgemeines Statistisches Archiv*, 90, 1, pp. 165-180.

Boes, S. and R. Winkelmann, 2006b. "The Effect of Income on Positive and Negative Subjective Well-Being," University of Zurich, Socioeconomic Institute, Manuscript, IZA Discussion Paper Number 1175.

Boyes, W., D. Hoffman and S. Low, 1989. "An Econometric Analysis of the Bank Credit Scoring Problem," *Journal of Econometrics*, 40, pp. 3-14.

Brant, R., 1990. "Assessing Proportionality in the Proportional Odds Model for Ordered Logistic Regression," *Biometrics*, 46, pp. 1171-1178.

Brewer, C., C. Kovner, W. Greene, Y. Cheng, 2008. **"**Predictors of RNs' Intent to Work and Work Decisions One Year Later in a U.S. National Sample," *The International Journal of Nursing*

Bricka, S., and C. Bhat, 2006. "A Comparative Analysis of GPS-Based and Travel Survey-based Data,". *Transportation Research Record*, 1972, 9-20.

Bunch, D. and R. Kitamura ,1990. Multinomial Probit Estimation Revisited: Testing Estimable Model Specifications, Maximum Likelihood Algorithms and Probit Integral Approximations for Car Ownership, Institute for Transportation Studies Technical Report, University of California, Davis.

Burridge, J., 1981. "A Note On Maximum Likelihood Estimation of Regression Models Using Grouped Data," *Journal of the Royal Statistical Society, Series B*, 43, pp. 41-45.

Butler, J., T. Finegan and J. Siegfried, 1994. "Does More Calculus Improve Student Learning in Intermediate Micro and Macro Economic Theory?" *American Economic Review*, 84, 2, pp. 206-210.

Butler, J. and P. Chatterjee**,** 1997. "Tests of the Specification of Univariate and Bivariate Ordered Probit," *Review of Economics and Statistics*, 79, pp. 343-347.

Butler, J. and R. Moffitt, 1982. "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model," *Econometrica*, 50, pp. 761-764.

Cameron, S. and J. Heckman, 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, pp. 262–333.

Carneiro, P., K. Hansen and J. Heckman, 2001. "Removing the Veil of Ignorance in Assessing the Distributional Impacts of Social Policies," *Swedish Economic Policy Review*, 8, pp. 273-301.

Carneiro, P., K. Hansen and J. Heckman, 2003. "Estimating Distributions of Treatment Effects with an Application to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, pp. 361-422.

Carro, J., 2007. "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," *Journal of Econometrics*, 140, pp. 503-528.

Cheung, S., 1996. "Provincial Credit Rating in Canada: An Ordered Probit Analysis," Bank of Canada, Working Paper 96-6. (http://www.bankofcanada.ca/en/res/wp/1996/wp96-6.pdf)

Clark, A., Y. Georgellis and P. Sanfey, 2001. "Scarring: The Psychological Impact of Past Unemployment," *Economica*, 68, pp. 221-241. et al. 2001

Congden, P., 2005. *Bayesian Models for Categorical Data*, John Wiley and Sons, New York.

Coughlin, C., T. Garrett and R. Hernandez-Murillo, 2004. "Spatial Probit and the Geographic Patterns of State Lotteries," Federal Reserve Bank of St. Louis, Working Paper 2003-042b.

Cragg, J., 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods, Econometrica, 39, 829-844.

Cunha, F., J. Heckman and S. Navarro, 2007. "The Identification & Economic Content of Ordered Choice Models with Stochastic Thresholds," University College Dublin, Gery Institute, Discussion Paper WP/26/2007.

Daykin, A. and P. Moffatt , 2002. "Analyzing Ordered Responses: A Review of the Ordered Probit Model," *Understanding Statistics*, I, 3, pp. 157-166.

Eluru, N., C. Bhat and D. Hensher, 2008. "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Levels in Traffic Crashes," *Accident Analysis and Prevention*, 40, 3, pp. 1033-1054..

Econometric Software, 2007. *NLOGIT: Version 4.0*, Plainview, New York.

Fernández-Val, I. and F. Vella, 2007. "Bias Corrections for Two-Step Fixed Effects Panel Data Estimators," IZA Working Papers Number 2690.

Franzese, R. and J. Hays, 2007. "The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation and Presentation," polmeth.wustl.edu/retrieve.php?id=715, last visited 4.7/09.

Fu, A., M, Gordon, G. Liu B. Dale and R. Christensen, 2004. "Inappropriate Medication Use and Health Outcomes in the Elderly" *Journal of the American Geriatrics Society*, 52, 11, pp. 1934–1939.

Genberg, H. and S. Gerlach, 2004. "Estimating Central Bank Reaction Functions with Ordered Probit: A Note" Graduate Institute of International Studies, Geneva, Manuscript.

Girard, P. and E. Parent, 2001. "Bayesian Analysis of Autocorrelated Ordered Categorical Data for Industrial Quality Monitoring," *Technometrics*, 43, 2, pp. 180-191.

Golob, T. and L. van Wissen, 1998. A Joint Household Travel Distance Generation and Car Ownership Model, Working Paper WP-88-15, Institute of Transportation Studies, University of California, Irvine.

Golob, T., 1990. The Dynamics of Household Travel time Expenditures and Car Ownership Decisions, Transportation Research A, 24, 443-465.

Greene. W., 1981. "Sample Selection Bias As a Specification Error: Comment," *Econometrica*, 49, pp. 795-798.

Greene, W., 1994. "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models," Working Paper 94-10, Department of Economics, Stern School of Business, New York University.

Greene, W., 1995. "Sample Selection in the Poisson Regression Model," Department of Economics, Stern School of Business, New York University, Working paper #95-06, 1995.

Greene, W., 2002. *LIMDEP Version 8.0, Reference Guide*, Plainview, NY, Econometric Software.

Greene, W., 2004a. "Fixed Effects and Bias Due To The Incidental Parameters Problem in the Tobit Model," *Econometric Reviews*, 23, 2, pp. 125-147.

Greene, W., 2004b. "The Behavior of the Fixed Effects Estimator in Nonlinear Models," *The Econometrics Journal* , 7, 1, pp. 98-119.

Greene, W., 2005. "Functional form and Heterogeneity in Models for Count Data," *Foundations and Trends in Econometrics*, 1, 2, pp. 113-218.

Greene, W., 2007. *LIMDEP Version 9.0: Reference Guide*, Plainview, New York, Econometric Software, Inc.

Greene, W., 2008a. *Econometric Analysis, 6th Edition*, Englewood Cliffs, Prentice Hall.

Greene, W., 2008b. "A Stochastic Frontier Model with Correction for Selection," Department of Economics, Stern School of Business, New York University, Working Paper EC-08-09.

Greene, W., M. Harris, B. Hollingsworth, P. Maitra, 2008. "A Bivariate Latent Class Correlated Generalized Ordered Probit Model with an Application to Modeling Observed Obesity Levels," Department of Economics, Stern School of Business, New York University, Working Paper 08-18.

Greene, W. and D. Hensher, 2009. "Ordered Choices and Heterogeneity in Attribute Processing," *Journal of Transport Economics and Policy*, forthcoming..

Greene, W., M. Harris, B. Hollingsworth and T. Weterings, 2012. "Heterogeneity in Ordered Choice Models: A Review with Applications to Self-Assessed Health," forthcoming, *Journal of Economic Surveys.*

Groot, W. and H. van den Brink, 2002. "Sympathy and the Value of Health," *Social Indicators Research*, 61, 1, pp. 97-120.

Groot, W. and H. van den Brink, 2003. "Match Specific Gains to Marriage: A Random Effects Ordered Response Model," *Quality and Quantity*, 37, pp. 317-325.

Hahn, J. and W. Newey, 2004. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 4, pp. 1295-1319.

Halton, J.H., 1970. A Retrospective and Prospective Survey of the Monte Carlo Method. *SIAMReview*, 12, 1-63.

Han, A. and J.A. Hausman, 1990. Flexible Parametric Estimation of Duration and Competing Risk Models, *Journal of Applied Econometrics*, 5, 1-28.

Hahn, J. and G. Kuersteiner, 2003. "Bias Reduction for Dynamic Nonlinear Panel Data Models with Fixed Effects," Department of Economics, UCLA, Manuscript.

Harris, M. and X. Zhao, 2007. "Modeling Tobacco Consumption with a Zero Inflated Ordered Probit Model," *Journal of Econometrics*, 141, pp. 1073-1099.

Heckman, J., 1979. "Sample Selection Bias as a Specification Error," *Econometrica*, 47, pp. 153-161.

Hensher, D. and Jones, S., 2007. Predicting Corporate Failure: Optimizing the Performance of the Mixed Logit Model, *ABACUS*, 43, 3, pp. 241-264.

Hensher, D., N. Smith, N. Milthorpe and P. Barnard, 1992. "Dimensions of Automobile Demand: A longitudinal Study of Household Automobile Ownership and Use," in *Studies in Regional Science and Urban Economics*, Elsevier Science Publishers, Amsterdam.

Holloway, G., Shankar, B. and Rahman, S., 2002. Bayesian Spatial Probit Estimation: A Primer and an Application to HYV Rice Adoption," *Agricultural Economics*, 27, pp. 383-402.

Hsiao, C., 1986. *Analysis of Panel Data*, Cambridge, Cambridge University Press.

Hsiao, C., 2003. *Analysis of Panel Data, 2nd. Ed.*, Cambridge, Cambridge University Press.

Jansen, J., 1990. "On the Statistical Analysis of Ordinal Data when Extravariation is Present," *Applied Statistics*, 39, pp. 75-84.

Johnson, V. and J. Albert, 1999, *Ordinal Data Modeling*, New York, Springer-Verlag.

Jones, S. and D. Hensher, 2004. "Predicting Firm Financial Distress: A Mixed Logit Model," *The Accounting Review (American Accounting Association)*, 79, pp. 1011-1038.

Kadam, A and P. Lenk, 2008. "Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration" . *Journal of Banking and Finance*. (SSRN:ssrn.com/abstract=1084006)

Kapteyn, A., J. Smith and A. van Soest, 2007. "Vignettes and Self-Reports of Work Disability in the United States and the Netherlands," *American Economic Review*, 97, 1, pp. 461-473.

Kasteridis, P., M Munkin, and S. Yen., 2008. "A Binary-Ordered Probit Model of Cigarette Demand." *Applied Economics*, 41.

Kitamura, R., 1987. A Panel Analysis of Household Car Ownership and Mobility, Infrastructure Planning and Management, Proceedings of the Japan Society of Civil Engineers, 383/IV-7, pp. 13-27.

Kitamura, R., 1988. "A Dynamic Model System of Household Car Ownership, Trip Generation and Modal Split, Model Development and Simulation Experiments," In *Proceedings of the 14th Australian Road Research Board Conference, Part 3*, Australian Road Research Board, Vermint South, Victoria, Australia, pp. 96-111.

Kitamura,R. and D. Bunch, 1989. "Heterogeneity and State Dependence in Household Car Ownership: A Panel Analysis Using Ordered-Response Probit Models with Error Components." Research Report, UCD-TRG-RR-89-6, Transportation Research Group, University of California at Davis.

Koop, G. and J. Tobias, 2006. "Semiparametric Bayesian Inference in Smooth Coefficient Models," *Journal of Econometrics*, 134, 1, pp. 283-315.

Lambert, D., 1992. "Zero-inflated Poisson Regression With An Application To Defects In Manufacturing," *Technometrics*, 34, 1, pp. 1-14.

Lancaster, T., 2000. "The Incidental Parameters Problem Since 1948," *Journal of Econometrics*, 95, pp. 391-413.

LeSage, J., 1999. "Spatial Econometrics," at www.rri.wvu.edu/WebBook/LeSage/spatial/spatial.htm.

LeSage, J., 2004. "Lecture 5: Spatial Probit Models," at www4.fe.uc.pt/spatial/doc/lecture5.pdf.

Li, M. and J. Tobias, 2006a. "Calculus Attainment and Grades Received in Intermediate Economic Theory," *Journal of Applied Econometrics*, 21,6, pp. 893-896.

Lindeboom, M. and E. van Doorslayer, 2003. "Cut Point Shift and Index Shift in Self Reported Health," *Ecuity III Project* Working Paper #2.

Long, S. 1997. *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oaks, CA, Sage Publications.

Machin, S. and A. Vignoles, 2005. *What's the Good of Education? The Economics of Education in the UK*, Princeton, N.J., Princeton University Press.

Maddala, J., 1983. *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.

Mannering, F. and Winston, C., 1985. "A Dynamic Analysis of Household Vehicle Ownership and Utilization," *Rand Journal of Economics*, 16, 215-236.

Marcus, A. and W. Greene, 1983. "The Determinants of Rating Assignment and Performance," Working Paper CRC528, Alexandria, VA, Center for Naval Analyses.

McCullagh, P., 1980. "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society, Series B (Methodological)*, 42, pp. 109-142.

McElvey, R. and W. Zavoina, 1971. "An IBM Fortran IV Program to Perform N-Chotomus Multivariate Probit Analysis," *Behavioral Science*, 16, 2, March, pp. 186-187.

McElvey, R. and W. Zavoina, 1975. "A Statistical Model for the Analysis of Ordered Level Dependent Variables," *Journal of Mathematical Sociology*, 4, pp. 103-120.

Metz, A. and R. Cantor, 2006, "Moody's Credit Rating Prediction Model," Moody's, Inc., http://www.moodys.com/cust/content/.../200600000425644.pdf.

Mora, N., 2006., "Sovereign Credit Ratings: Guilty Beyond Reasonable Doubt?" *Journal of Banking and Finance*, 30, pp. 2041-2062.

Mullahy, J., 1986. "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33, pp. 341-365.

Mullahy, J., 1997. "Heterogeneity, Excess Zeros and the Structure of Count Data Models," *Journal of Applied Econometrics*, 12, pp. 337-350.

Munkin, M. and P. Trivedi, 2008. "Bayesian Analysis of the Ordered Probit Model with Endogenous Selection," *Journal of Econometrics*, 143, pp. 334-348.

Murphy, K. and R. Topel, 2002. "Estimation and Inference in Two Stem Econometric Models," *Journal of Business and Economic Statistics*, 20, pp. 88-97 (reprinted from 2, pp. 370-379).

NDSHS, 2001. Computer Files for the Unit Record Data from the National Drug Strategy Household Surveys.

Pratt, J., 1981. "Concavity of the Log Likelihood," *Journal of the American Statistical Association*, 76, pp. 103-116.

Prescott, E. and M. Visscher, 1977. "Sequential Location among Firms with Foresight," *Bell Journal of Economics*, 8, pp. 378–893.

Pudney, S. and M. Shields, 2000. "Gender, Race, Pay and Promotion in the British Nursing Profession: Estimation of a Generalized Ordered Probit Model," *Journal of Applied Econometrics*, 15, pp. 367-399.

Purvis, L., 1994. Using Census Public Use Micro Data Sample to Estimate Demographic and Automobile Ownership Models, Transportation Research Record, 1443, 21-30.

Rabe-Hesketh, S**.**, A. Skrondal, A. and A. Pickles, 2005. "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects," *Journal of Econometrics*, 128, pp. 301-323.

Raudenbusch, S. and A. Bryk, 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, Thousand Oaks, CA.

Ridder, G., 1990. "The Non-parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57 pp. 167–181.

Riphahn, R., A. Wambach and A. Million, 2003. "Incentive Effects on the Demand for Health Care: A Bivariate Panel Count Data Estimation," *Journal of Applied Econometrics*, 18, 4, pp. 387-405.

Roorda, M., Páez, A., Morency, C., Mercado, R. and Farber, S**.**. 2009. "Trip Generation of Vulnerable Populations in Three Canadian Cities: A Spatial Ordered Probit Approach," Manuscript, School of Geography and Earth Sciences, McMaster University.

Shaked, A. and J. Sutton, 1982. "Relaxing Price Competition through Product Differentiation," *Review of Economic Studies*, 49, pp. 3–13.

Snell, E., 1964. "A Scaling Procedure for Ordered Categorical Data," *Biometrics*, 20, pp. 592-607.

Stata, 2008. *Stata, Version 8.0*, College Station TX, Stata Corp.

Terza, J., 1985. "Ordered Probit: A Generalization," *Communications in Statistics – A. Theory and Methods*, 14, pp. 1-11.

Tomoyuki, F. and F. Akira, 2006. "A Quantitative Analysis on Tourists' Consumer Satisfaction Via the Bayesian Ordered Probit Model," Journal of the City Planning Institute of Japan, 41, pp. 2-10. (In Japanese)

Train, K., 1986. Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand, MIT Press

Train, K., 2003. *Discrete Choice Methods with Simulation*, Cambridge, Cambridge University Press.

Tsay, R., 2005. *Analysis of Financial Time Series, 2nd Ed*., New York, John Wiley and Sons.

Walker, S. and D. Duncan, 1967. "Estimation of the Probability of an Event As a Function of Several Independent Variables," *Biometrika*, 54, pp. 167-179.

Wang, X., and K. Kockelman, 2008. "Application of the Dynamic Spatial Ordered Probit Model: Patterns of Land Development Change in Austin, Texas," Manuscript, Department of Civil Engineering, University of Texas, Austin (forthcoming, *Papers in Regional Science*).

Wang, X. and K. Kockelman, 2009. "Application of the Dynamic Spatial Ordered Probit Model: Patterns of Ozone Concentration in Austin, Texas." Manuscript, Department of Civil Engineering, University of Texas, Austin.

Winkelmann, R., 2005. "Subjective Well-being and the Family: Results from an Ordered Probit Model with Multiple Random Effects" *Empirical Economics*, 30, 3, pp 749-761,

Wooldridge, J., 2002b. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MIT Press.

Wooldridge, J. and G. Imbens, 2009a. "Lecture Notes 6, Summer 2007," http://www.nber.org/WNE/lect_6_controlfuncs.pdf.

Wynand, P. and B. van Praag, 1981. "The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection," *Journal of Econometrics*, 17, pp. 229-252.

Zavoina, W. and R. McKelvey, 1969, "A Statistical Model for the Analysis of Legislative voting Behavior," Presented at the meeting of the American Political Science Association.

Zhang, Y, F. Liang and Y. Yuanchang, 2007. "Crash Injury Severity Analysis Using a Bayesian Ordered Probit Model," Transportation Research Board, Annual Meeting, Paper Number 07-2335.

Zhang, J., 2007. "Ordered Probit Modeling of User Perceptions of Protected Left-Turn Signals," *Journal of Transportation Engineering*., 133, 3, pp. 205-214.

Zigante, V., 2007. "Ever Rising Expectations – The Determinants of Subjective Welfare in Croatia," School of Economics and Management, Lund University, Masters Thesis (www.essays.se/about/Ordered+Probit+Model/).