1 Introduction: random utility and ordered choice models

Netflix (www.netflix.com) is an internet company that rents movies on DVDs to subscribers. The business model works by having subscribers order the DVD online for home delivery and return by regular mail. After a customer returns a DVD, 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* given by approximately 12,900 visitors to the site was 3.8. This rating process provides a natural application of the models and methods that interest us in this book.

For any individual viewer, we might reasonably hypothesize that there is a continuously varying strength of preferences for the movie that would underlie the rating they submit. For convenience and consistency with what follows, we will label that strength of preference "utility," U^*. Given that there are no natural units of measurement, we can describe utility as ranging over the entire real line:

$$-\infty < U^*_{im} < +\infty$$

where i indicates the individual and m indicates the movie. Individuals are invited to "rate" the movie on an integer scale from one to five. Logically, then, the translation from underlying utility to a rating could be viewed as a *censoring* of the underlying utility,

$$\begin{aligned}
R_{im} &= 1 \text{ if } -\infty < U^*_{im} \le \mu_{i1},\\
R_{im} &= 2 \text{ if } \mu_{i1} < U^*_{im} \le \mu_{i2},\\
R_{im} &= 3 \text{ if } \mu_{i2} < U^*_{im} \le \mu_{i3},\\
R_{im} &= 4 \text{ if } \mu_{i3} < U^*_{im} \le \mu_{i4},\\
R_{im} &= 5 \text{ if } \mu_{i4} < U^*_{im} < \infty.
\end{aligned} \tag{1.1}$$

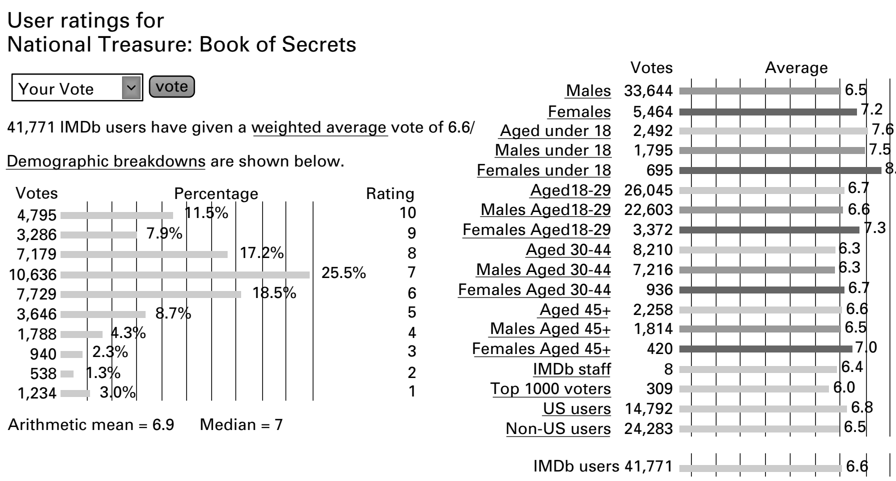
The crucial feature of the description thus far is that the viewer has (and presumably knows) a continuous range of preferences that they could express if they were not forced to provide only an integer from one to five. Therefore, the observed rating represents a censored version of the true underlying preferences. Providing a rating of five could be an outcome ranging from general enjoyment to wild enthusiasm. Note that the *thresholds*, μ_{ij}, are specific to the person, and number (J-1) where J is the number of possible ratings (here, five) with $J-1$ values needed to divide the range of utility into J cells. The thresholds are an important element of the model; they divide the range of utility into cells that are then identified with the observed ratings. One of the admittedly unrealistic assumptions in many applications is that these threshold values are the same for all individuals. Importantly, the difference between two levels of a rating scale (e.g., one compared to two, two compared to three) is not the same on a utility scale; hence we have a strictly nonlinear transformation captured by the thresholds, which are estimable parameters in an ordered choice model.

The model as suggested thus far provides a crude description of the mechanism underlying an observed rating. But it is simple to see how it might be improved. Any individual brings their own set of *characteristics* to the utility function, such as age, income, education, gender, where they live, family situation and so on, which we denote $x_{i1}, x_{i2}, \ldots, x_{iK}$. They also bring their own aggregate of unmeasured and unmeasurable (by the analyst) idiosyncrasies, denoted ε_{im}. How these features enter the utility function is uncertain, but it is conventional to use a linear function, which produces a familiar *random utility function*,

$$U^*_{im} = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \ldots + \beta_{iK}x_{iK} + \varepsilon_{im}. \tag{1.2}$$

Once again, the model accommodates the intrinsic heterogeneity of individuals by allowing the coefficients to vary across them. To see how the heterogeneity across individuals might enter the ordered choice model, consider the user ratings of the same movie noted earlier, posted on December 1, 2008 at a different website, www.IMDb.com, as shown in Figure 1.1. This site uses a ten-point scale. The panel at the left below shows the overall ratings for 41,771 users of the site. The panel at the right shows how the average rating varies across age, gender and whether the rater is a US viewer or not.

An obvious shortcoming of the model is that otherwise similar viewers might naturally feel more enthusiastic about certain genres of movies (action, comedy, crime, etc.) or certain directors, actors or studios. It would be natural



Votes	Percentage	Rating
4,795	11.5%	10
3,286	7.9%	9
7,179	17.2%	8
10,636	25.5%	7
7,729	18.5%	6
3,646	8.7%	5
1,788	4.3%	4
940	2.3%	3
538	1.3%	2
1,234	3.0%	1

Figure 1.1 IMDb.com ratings (www.imdb.com/title/tt0465234/ratings)

for the utility function defined over movies to respond to certain *attributes* $z_1, z_2, \ldots, z_M$. The utility function might then appear, using a vector notation for the characteristics and attributes, as

$$U^*_{im} = \boldsymbol{\beta}'_i \mathbf{x}_i + \boldsymbol{\delta}'_i \mathbf{z}_m + \varepsilon_{im}. \tag{1.3}$$

Note, again, the marginal utilities of the attributes, $\boldsymbol{\delta}_i$, will vary from person to person. We note, finally, two possible refinements to accommodate additional sources of randomness, i.e., individual heterogeneity. Two otherwise *observably* identical individuals (same $\mathbf{x}_i$) seeing the same movie (same $\mathbf{z}_m$) might still react differently because of individual idiosyncrasies that are characteristics of the person that are the same for all movies. Some individuals are drawn to comedies and have low regard for dramas, while others might be uninterested in these two genres and enjoy only action movies. Second, every movie has unique features that are not captured by a simple *hedonic index* of its attributes – a particularly skillful character development, etc. A more complete *random utility function* might appear

$$U^*_{im} = \boldsymbol{\beta}'_i \mathbf{x}_i + \boldsymbol{\delta}'_i \mathbf{z}_m + \varepsilon_{im} + u_i + v_m. \tag{1.4}$$

Finally, note that Netflix maintains a (huge) database of the ratings made by its users, including a complete history for each individual.

To return to the rating mechanism, the model we have constructed is:

$$
\begin{aligned}
R_{im} &= 1 \text{ if } -\infty < \boldsymbol{\beta}_i'\mathbf{x}_i + \boldsymbol{\delta}_i'\mathbf{z}_m + \varepsilon_{im} + u_i + v_m \le \mu_{i1},\\
R_{im} &= 2 \text{ if } \mu_{i1} < \boldsymbol{\beta}_i'\mathbf{x}_i + \boldsymbol{\delta}_i'\mathbf{z}_m + \varepsilon_{im} + u_i + v_m \le \mu_{i2},\\
R_{im} &= 3 \text{ if } \mu_{i2} < \boldsymbol{\beta}_i'\mathbf{x}_i + \boldsymbol{\delta}_i'\mathbf{z}_m + \varepsilon_{im} + u_i + v_m \le \mu_{i3},\\
R_{im} &= 4 \text{ if } \mu_{i3} < \boldsymbol{\beta}_i'\mathbf{x}_i + \boldsymbol{\delta}_i'\mathbf{z}_m + \varepsilon_{im} + u_i + v_m \le \mu_{i4},\\
R_{im} &= 5 \text{ if } \mu_{i4} < \boldsymbol{\beta}_i'\mathbf{x}_i + \boldsymbol{\delta}_i'\mathbf{z}_m + \varepsilon_{im} + u_i + v_m \le \infty.
\end{aligned} \tag{1.5}
$$

Perhaps relying on a central limit theorem to aggregate the innumerable small influences that add up to the individual idiosyncrasies and movie attraction, we assume that the random components, ε_{im}, u_i and v_m are normally distributed with zero means and (for now) constant variances. The assumption of normality will allow us to attach probabilities to the ratings. In particular, arguably the most interesting one is

$$\text{Prob}(R_{im} = 5|\mathbf{x}_i, \mathbf{z}_m, u_i, v_m) = \text{Prob}[\varepsilon_{im} > \mu_{i4} - (\boldsymbol{\beta}_i'\mathbf{x}_i + \boldsymbol{\delta}_i'\mathbf{z}_m + u_i + v_m)]. \tag{1.6}$$

The structure provides the framework for an econometric model of how individuals rate movies (that they rent from Netflix). The resemblance of this model to familiar models of binary choice is more than superficial. For example, one might translate this econometric model directly into a *probit model* by focusing on the variable

$$
\begin{aligned}
E_{im} &= 1 \text{ if } R_{im} = 5\\
E_{im} &= 0 \text{ if } R_{im} < 5.
\end{aligned} \tag{1.7}
$$

Thus, our model is an extension of a binary choice model to a setting of more than two choices. However, the crucial feature of the model is the ordered nature of the observed outcomes and the correspondingly ordered nature of the underlying preference scale.

Beyond the usefulness of understanding the behavior of movie viewers, e.g., whether certain genres are more likely to receive high ratings or whether certain movies appeal to particular demographic groups, such a model has an additional utility to Netflix. Each time a subscriber logs on to the website after returning a movie, a computer program generates recommendations of other movies that it thinks that the viewer would enjoy (i.e., would give a rating of 5). The better the recommendation system is, the more attractive will be the website. Thus, the ability to predict accurately a "5" rating is a model feature that would have business value to Netflix. Netflix is currently (2008 until 2011) running a contest with a $1,000,000 prize to the individual who can devise the best algorithm for matching individual ratings based on ratings of other movies that they have rented. See www.netflixprize.com, Hafner (2006) and Thompson (2008). The Netflix prize and internet rating systems in general, beyond a large popular interest, have attracted a considerable amount of academic attention. See, for example, Ansari *et al.* (2000), Bennett and Lanning (2007) and Umyarov and Tuzhlin (2008).

The model described here is an *ordered choice model*. (The choice of the normal distribution for the random term makes it an *ordered probit 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 Aitchison and Silvey (1957) and Snell (1964) and articulated in its modern form by Zavoina and McKelvey (1975), McKelvey 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 (McKelvey and Zavoina (1975));
- credit ratings (Cheung (1996), Metz and Cantor (2006));
- driver injury severity in car accidents (Eluru *et al.* (2008), Wang and Kockelman (2008));
- drug reactions (Fu *et al.* (2004));
- duration (Han and Hausman (1986, 1990), Ridder (1990));
- education (Carneiro *et al.* (2001, 2003), Machin and Vignoles (2005), Cameron and Heckman (1998), Johnson and Albert (1999), Cunha *et al.* (2007));
- 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 (Riphahn *et al.* (2003), Greene (2008a));
- insect resistance to insecticide (Walker and Duncan (1967));
- job classification in the military (Marcus and Greene (1983));
- job training (Groot and van den Brink (2003c));

Cambridge University Press & Assessment
978-0-521-19420-4 — Modeling Ordered Choices
William H. Greene , David A. Hensher
Excerpt
More Information

- labor supply (Heckman and MaCurdy (1981));
- life satisfaction (Clark *et al.* (2001), Groot and van den Brink (2003c));
- monetary policy (Eichengreen *et al.* (1985));
- nursing labor supply (Brewer *et al.* (2008));
- obesity (Greene *et al.* (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 (2009));
- product quality (Prescott and Visscher (1977), Shaked and Sutton (1982));
- promotion and rank in nursing (Pudney and Shields (2000));
- stock price movements (Tsay (2005));
- tobacco use (Harris and Zhao (2007), Kasteridis *et al.* (2008));
- toxicity in pregnant mice (Agresti (2002));
- trip stops (Bhat (1997));
- vehicle ownership (Train (1986), Bhat and Pulugurta (1998), Hensher *et al.* (1992));
- work disability (Kapteyn *et al.* (2007))

and hundreds more.

This book will survey the development and use of models of ordered choices primarily from the perspective of the social sciences. We will detail the model itself, estimation and inference, interpretation and analysis. We will also survey a wide variety of different kinds of applications, and a wide range of variations and extensions of the basic model that have been proposed in the recent literature.

The practitioner who desires a quick entry-level primer on the model can choose among numerous sources for a satisfactory introduction to the ordered choice model and its uses. Social science-oriented introductions appear in journal articles such as Becker and Kennedy (1992), Winship and Mare (1984), Daykin and Moffatt (2002), and Boes and Winkelmann (2006a), and in textbook and monograph treatments including Maddala (1983), Long (1997), Johnson and Albert (1999), DeMaris (2004), Long and Freese (2006) and Greene (2008a). There are also many surveys and primers for bioassay, including, e.g., Greenland (1994), Ananth and Kleinbaum (1997), Agresti (1999, 2002), and Congdon (2005). This survey is offered as an addition to this list, largely to broaden the discussion of the model and for a number of specific purposes:

- Many interesting extensions of the model already appearing in the literature are not mentioned in any of the surveys listed above.

- Recent analyses of the ordered choice model have uncovered some interesting avenues of generalization.
- The model formulation rests on a number of subtle underlying aspects that are not developed as completely as are the mechanics of using the "technique" (e.g., estimating the parameters). Only a few of the surveys devote substantial space to interpreting the model's components once they are estimated. As made clear here and elsewhere, the coefficients in an ordered choice model provide, in isolation, relatively little directly useful information about the phenomenon under study. Yet, estimation of coefficients and tests of statistical significance are the central (sometimes, only) issue in many of the surveys listed above, and in some of the received applications.
- We will offer our own generalizations of the ordered choice model.
- With the creative development of easy to use contemporary software, many model features and devices are served up because they *can* be computed without much (or any) discussion of *why* they would be computed, or, in some cases, even *how* they are computed. To cite an example, Long and Freese (2006, pp. 195–6) state "several different measures [of fit] can be computed…" [using *Stata*] for the ordered probit model. Their table that follows lists twenty values, seven of which are statistics whose name contains "*R* squared." The values range from 0.047 to 0.432. The discussion to follow provides the reader with a single statement that two Monte Carlo studies have found that one of the measures "closely approximates the R^2 obtained by fitting the linear regression model on the underlying latent variable." We will attempt to draw the focus to a manageable few aspects of the model that appear to have attained some degree of consensus.

The book proceeds as follows. Standard models of binary choice are presented in Chapter 2. The fundamental ordered choice model is developed in some detail in Chapter 3. The historical antecedents to the basic ordered choice model are documented in Chapter 4. In Chapter 5, we return to the latent regression-based form of the model, and develop the different aspects of its use, such as interpreting the model, statistical inference, and fit measures. Some recent generalizations and extensions are presented in Chapters 6–11. Semiparametric models that reach beyond the mainstream of research are discussed in Chapter 12. An application based on a recent study of health care (Riphahn *et al.* (2003)) will be dispersed through the discussion to provide an illustration of the points being presented.

There is a large literature parallel to the social science applications in the areas of biometrics and psychometrics. The distinction is not perfectly neat, but there is a tangible difference in orientation, as will be evident below.

From the beginning with Bliss's (1934a) invention of probit modeling, many of the methodological and statistical developments in the area of ordered choice modeling have taken place in this setting. It will be equally evident that these two areas of application have developed in parallel, but by no means in concert. This book is largely directed toward social science applications. However, the extensions and related features of the models and techniques in biometrics will be integrated into the presentation.

Cambridge University Press & Assessment
978-0-521-19420-4 — Modeling Ordered Choices
William H. Greene , David A. Hensher
Excerpt
More Information

2 Modeling binary choices

The *random utility* model described in Chapter 1 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 book 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. This chapter will develop the standard model for binary choices in considerable detail. Many of the results analyzed in the chapters to follow will then be straightforward extensions.

There are numerous surveys available, including Amemiya (1981), Greene (2008a, Ch. 23) and several book-length treatments such as Cox (1970) and Collett (1991). Our interest here is in the aspects of binary choice modeling that are likely to reappear in the analysis of ordered choices. We have therefore bypassed several topics that do appear in other treatments, notably semiparametric and nonparametric approaches, but whose counterparts have not yet made significant inroads in ordered choice modeling. (Chapter 12 does contain some description of a few early entrants to this nascent literature.) This chapter also contains a long list of topics related to binary choice modeling, such as fit measures, multiple equation models, sample selection, and many others, that are useful as components or building blocks in the analysis of ordered choices. Our intent with this chapter is to extend beyond conventional binary choice modeling, and provide a bridge to the somewhat more involved models for ordered choices. Quite a few of these models, such as the sample selection one, can easily be generalized to the ordered probit model.

The orientation of our treatment is the analysis of individual choice data, as typically appears in social science applications using survey data. An example is the application developed below in which survey data on health satisfaction are transformed into a binary outcome that states whether or not a respondent feels healthier than average. A parallel literature in, e.g., bioassay, such as Cox (1970) and Johnson and Albert (1999) is often focused on *grouped* data in the form of proportions. Two examples would be an experiment to determine the lethality of a new insecticide in which n_i insects are subjected to dosage x_i, and a proportion p_i succumb to the dose, and a state-by-state tally of voting proportions in a (US) presidential election. With only a few exceptions noted in passing, we will not be concerned with grouped data.

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 family year observations. A histogram of the responses appears in Figure 5.1. Consistent with the description in Chapter 1, we might formulate a random utility/ordered choice model for the variable $R_i =$ "*Health Satisfaction*" as

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

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

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

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

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

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

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