Allgemeines Statistisches Archiv 0, 0–14

Ordered Response Models

By Stefan Boes and Rainer Winkelmann*

Summary: We discuss regression models for ordered responses, such as ratings of bonds, schooling attainment, or measures of subjective well-being. Commonly used models in this context are the ordered logit and ordered probit regression models. They are based on an underlying latent model with single index function and constant thresholds. We argue that these approaches are overly restrictive and preclude a flexible estimation of the effect of regressors on the discrete outcome probabilities. For example, the signs of the marginal probability effects can only change once when moving from the smallest category to the largest one. We then discuss several alternative models that overcome these limitations. An application illustrates the benefit of these alternatives.

Keywords: Marginal effects, generalized threshold, sequential model, random coefficients, latent class analysis, happiness. JEL C25, I25.

1. Introduction

Regression models for ordered responses, i.e. statistical models in which the outcome of an ordered dependent variable is explained by a number of arbitrarily scaled independent variables, have their origin in the biometrics literature. Aitchison and Silvey (1957) proposed the ordered probit model to analyze experiments in which the responses of subjects to various doses of stimulus are divided into ordinally ranked classes. Snell (1964) suggested the use of the logistic instead of the normal distribution as an approximation for mathematical simplification. The first comprehensive treatment of ordered response models in the social sciences appeared with the work of McKelvey and Zavoina (1975) who generalized the model of Aitchison and Silvey to more than one independent variable. Their basic idea was to assume the existence of an underlying continuous latent variable – related to a single index of explanatory variables and an error term – and to obtain the observed categorical outcome by discretizing the real line into a finite number of intervals.

McCullagh (1980) developed independently the so-called *cumulative model* in the statistics literature. He directly modelled the cumulative probabilities of the ordered outcome as a monotonic increasing transformation of a linear predictor onto the unit interval, assuming a logit or probit link function. This specification yields the same probability function as the model of McKelvey and Zavoina, and is therefore observationally equivalent. Both papers spurred a large literature on how to model ordered dependent variables, the former mostly in the social sciences, the latter predominantly in the medical and biostatistics literature.

On the one hand, a number of parametric generalizations have been proposed. These include alternative link functions, prominent examples being

Received: / Revised:

* We are grateful to an anonymous referee for valuable comment.

the log-log or the complementary log-log function (McCullagh, 1980), generalized predictor functions that include, for example, quadratic terms or interactions, or dispersion parameters (Cox, 1995). Olsson (1979) and Ronning and Kukuk (1996) discuss estimation of models in which both dependent and independent variables are ordered in the context of multivariate latent structural models, i.e. an adaptation of log-linear models to ordinal data. On the other hand, semi- and non-parametric approaches replace the distributional assumptions of the standard model, or the predictor function, by flexible semi- or non-parametric functional forms. General surveys of the parametric as well as the semi- and nonparametric literature are given, for example, in Agresti (1999), Barnhart and Sampson (1994), Clogg and Shihadeh (1994), Winship and Mare (1984), Bellemare, Melenberg, and van Soest (2002), and Stewart (2004), the two latter references in particular for the semi- and nonparametric treatments of ordered data.

When thinking about the usefulness of these alternative models, it is inevitable to make up one's mind on the ultimate objective of the analysis. It is our perception that in most applications of ordered response models the parameters of the latent model do not have direct interpretation *per se*. Rather, the interest lies in the shift of the predicted discrete ordered outcome distribution as one or more of the regressors change, i.e. the *marginal probability effects*. Perhaps surprisingly, standard ordered response models are not very well suited to analyze these marginal probability effects, because the answer is to a large extent predetermined by the rigid parametric structure of the model. Therefore, we consider a number of generalizations that allow for flexible analyses of marginal probability effects. In addition to the generalized threshold model (Maddala, 1983; Terza, 1985; Brant, 1990) and the sequential model (Fienberg, 1980; Tutz, 1990, 1991), we show how additional flexibility can be gained by modeling individual heterogeneity either by means of a random coefficients model or as a finite mixture/latent class model.

The remainder of the paper is organized as follows. In the next section we provide a short review of the standard model, before turning to the generalizations in section 3. In section 4 we illustrate the methods with an analysis of the relationship between income and happiness using data from the German Socio-Economic Panel. Our results show that marginal probability effects in the generalized alternatives are substantially different from those in the standard model. For example, the standard model implies that the probability of being *completely satisfied* increases on average by about 0.017 percentage points by a one-percentage increase in income, while it is decreasing or constant in the generalized models. Section 5 concludes.

2. Standard Ordered Response Models

Consider the following examples. In a survey, respondents have been asked about their life-satisfaction, or their change in health status. Answer categories might range from 0 to 10 where 0 means *completely dissatisfied* and

10 means *completely satisfied*, or from 1 to 5, where 1 means *greatly deteriorated* and 5 means *greatly improved*, respectively. The objective is to model these ordered responses as functions of explanatory variables.

Formally, let the ordered categorical outcome y be coded, without loss of generality, in a rank preserving manner, i.e. $y \in \{1, 2, \ldots, J\}$ where J denotes the total number of distinct categories. Furthermore, suppose that a $(k \times 1)$-dimensional vector x of covariates is available. In standard ordered response models, the cumulative probabilities of the discrete outcome are related to a single index of explanatory variables in the following way

$$\Pr[y \leq j|x] = F(\kappa_j - x'\beta) \qquad j = 1, \ldots, J \tag{1}$$

where κ_j and $\beta_{(k\times1)}$ denote unknown model parameters, and F can be any monotonic increasing function mapping the real line onto the unit interval. Although no further restrictions are imposed *a priori* on the transformation F it is standard to replace F by a distribution function, the most commonly used ones being the standard normal (which yields the ordered probit) and the logistic distribution (associated with the ordered logit model), and we assume in what follows that F represents either the standard normal or logistic distribution. In order to ensure well-defined probabilities, we require that $\kappa_j > \kappa_{j-1}$, $\forall j$, and it is understood that $\kappa_J = \infty$ such that $F(\infty) = 1$ as well as $\kappa_0 = -\infty$ such that $F(-\infty) = 0$.

Ordered response models are usually motivated by an underlying continuous but latent process $y^\star$ together with a response mechanism of the form

$$y = j \quad \text{if and only if} \qquad \kappa_{j-1} \;\leq\; y^\star = x'\beta + u \;<\; \kappa_j \qquad j = 1, \ldots, J$$

where $\kappa_0, \ldots, \kappa_J$ are introduced as threshold parameters, discretizing the real line, represented by $y^\star$, into J categories. The latent variable $y^\star$ is related linearly to observable and unobservable factors and the latter have a fully specified distribution function $F(u)$ with zero mean and constant variance.

The cumulative model (1) can be postulated without assuming the existence of a latent part and a threshold mechanism, though. Moreover, since $y^\star$ cannot be observed and is purely artificial, its interpretation is not of interest. The main focus in the analysis of ordered data should be put on the conditional cell probabilities given by

$$\Pr[y = j|x] = F(\kappa_j - x'\beta) - F(\kappa_{j-1} - x'\beta) \tag{2}$$

In order to identify the parameters of the model we have to fix location and scale of the argument in F, the former by assuming that x does not contain a constant term, the latter by normalizing the variance of the distribution function F. Then, equation (2) represents a well-defined probability function which allows for straightforward application of maximum likelihood methods for a random sample of size n of pairs (y, x).

The most natural way to interpret ordered response models (and discrete probability models in general) is to determine how a marginal change in one regressor changes the distribution of the outcome variable, i.e. all the outcome probabilities. These marginal probability effects can be calculated as

$$MPE_{jl}(x) = \frac{\partial \Pr[y=j|x]}{\partial x_l} = \Big[f(\kappa_{j-1} - x'\beta) - f(\kappa_j - x'\beta)\Big]\beta_l \quad (3)$$

where $f(z) = \mathrm{d}F(z)/\mathrm{d}z$ and x_l denotes the l-th (continuous) element in x. With respect to a discrete valued regressor it is more appropriate to calculate the change in the probabilities before and after the discrete change Δx_l,

$$\Delta \Pr[y=j|x] = \Pr[y=j|x+\Delta x_l] - \Pr[y=j|x] \quad (4)$$

In general, the magnitude of these probability changes depends on the specific values of the ith observation's covariates. After taking expectation with respect to x we obtain average marginal probability effects, which can be estimated consistently by replacing the true parameters by their corresponding maximum likelihood estimates and taking the average over all observations.

However, if we take a closer look at (3) and (4) it becomes apparent that marginal probability effects in standard ordered response models have two restrictive properties that limit the usefulness of these models in practice. First, the ratio of marginal probability effects of two distinct continuous covariates on the same outcome, i.e. *relative* marginal probability effects, are constant across individuals and the outcome distribution, because from (3) we have that

$$\frac{MPE_{jl}(x)}{MPE_{jm}(x)} = \frac{\beta_l}{\beta_m}$$

which does not depend on i and j. Second, marginal probability effects change their sign exactly once when moving from the smallest to the largest outcome. More precisely, if we move stepwise from the lowest category $y = 1$ to the highest category $y = J$, the effects are either first negative and then positive ($\beta_l > 0$), or first positive and then negative ($\beta_l < 0$). This "single crossing property" follows directly from the bell-shaped density functions of the standard normal and the logistic distribution. Therefore, if we are interested in the effect of a covariate on the outcome probabilities, i.e. if we turn our attention to the effects on the full distribution of outcomes, the standard models preclude a flexible analysis of marginal probability effects by design.

3. Generalized Ordered Response Models

Three assumptions of the standard model are responsible for its limitations in analyzing marginal probability effects: First, the single index assumption, second, the constant threshold assumption, and third, the distributional assumption which does not allow for additional individual heterogeneity between individual realizations. While relaxing these assumptions we want to retain the possibility of interpreting the model in terms of marginal probability effects. Therefore, we need to search for a richer class of parametric models that does not impose restrictions such as constant relative effects or single crossing. In this section we present four such alternatives.

3.1. Generalized Threshold Model. The first model we consider relaxes the single index assumption and allows for different indices across outcomes. This model was introduced by Maddala (1983) and Terza (1985) who proposed to generalize the threshold parameters by making them dependent on covariates

$$\kappa_j = \tilde{\kappa}_j + x'\gamma_j$$

where γ_j is a $k \times 1$-dimensional vector of response specific parameters. Plugging this into (1) we get the cumulative probabilities in the generalized threshold model

$$\Pr[y \leq j|x] = F(\tilde{\kappa}_j + x'\gamma_j - x'\beta) = F(\tilde{\kappa}_j - x'\beta_j) \qquad j = 1, \ldots, J \tag{5}$$

where it is understood that $\tilde{\kappa}_0 = -\infty$ and $\tilde{\kappa}_J = \infty$, as before. The last equality in (5) follows because γ_j and β cannot be identified separately with the same x entering the index function and the generalized thresholds, and we define $\beta_j \equiv \beta - \gamma_j$. The cumulative probabilities define a probability density function in the same manner as in (2) and parameters can be estimated directly by maximum likelihood. A non-linear specification can be used to ensure that $\tilde{\kappa}_{j-1} - x'\beta_{j-1} < \tilde{\kappa}_j - x'\beta_j$ for all $\tilde{\kappa}$, $\tilde{\beta}$ and x (e.g. Ronning, 1990). We observe that the generalized threshold model nests the standard model under the restrictions $\beta_1 = \ldots = \beta_{J-1}$ and therefore both models can be tested against each other by performing a likelihood ratio (LR) test.

The generalized threshold model provides a framework in which marginal probability effects can be analyzed with much more flexibility than in the standard model, since

$$MPE_{jl}(x) = f(\tilde{\kappa}_{j-1} - x'\beta_{j-1})\beta_{j-1l} - f(\tilde{\kappa}_j - x'\beta_j)\beta_{jl} \tag{6}$$

does not rely anymore on a single crossing property or constant relative effects. Nevertheless, this generalization comes at a cost. The model now contains $(J-2)k$ parameters more than before which reduces the degrees of freedom considerably, in particular when J is large.

3.2. Random Coefficients Model. As a second alternative we discuss the class of random coefficients models. The basic idea is to randomize the parameters of interest by adding an error term that is correlated with the unobserved factors in u. Thus, we translate individual heterogeneity into parameter heterogeneity, writing the vector of slopes as

$$\beta = \tilde{\beta} + \varepsilon$$

where ε is an individual specific $(k \times 1)$-dimensional vector of error terms. Moreover, we assume for the joint error term $\gamma \equiv (\varepsilon' \quad u)'$ that

$$\mathrm{E}[\gamma|x] = 0 \quad \text{and} \quad E[\gamma\gamma'|x] = \Sigma \;\; \text{with} \;\; \Sigma = \begin{pmatrix} \Omega & \psi \\ \psi' & 1 \end{pmatrix}$$

where Ω is the $(k \times k)$-dimensional covariance matrix of ε, ψ is the $(k \times 1)$-dimensional covariance vector between the slope parameters and u, and $\mathrm{Var}[u|x] = 1$, as before. The consequences of this modification are easiest seen from the latent variable representation, where we now have $y^\star = x'\tilde{\beta} + \tilde{u}$ with "new" error term $\tilde{u} \equiv x'\varepsilon + u$, such that

$$\mathrm{E}[\tilde{u}|x] = 0 \quad \text{and} \quad \mathrm{E}[\tilde{u}\tilde{u}'|x] = x'\Omega x + 2x'\psi + 1 \;\equiv \sigma_{\tilde{u}}^2$$

and $\tilde{u}/\sigma_{\tilde{u}}$ is distributed with distribution function F. If ε and u are jointly normal with covariance structure given by Σ, we obtain an ordered probit model with unobserved heterogeneity. However, in principle, we do not need to know the distributions of ε or u, as long as F is a well-defined distribution function. In this case, we can express the cumulative probabilities in the random coefficients model as

$$\Pr[y \leq j|x] = F\left(\frac{\kappa_j - x'\tilde{\beta}}{\sigma_{\tilde{u}}}\right) \;\; \equiv \tilde{F}_j(x) \tag{7}$$

where $\sigma_{\tilde{u}} = \sqrt{x'\Omega x + 2x'\psi + 1}$ can be seen as dispersion parameter. The standard model is a special case of the random coefficients model under the assumption $\Omega = 0$ and $\psi = 0$. Thus, a simple LR test can be used to test for parameter heterogeneity.

The probability density function of y is obtained in the same way as in (2), and one can calculate marginal probability effects in the random coefficients model as

$$\begin{aligned} MPE_{jl}(x) = {} & \left[\tilde{f}_{j-1}(x) - \tilde{f}_j(x)\right] \frac{\tilde{\beta}_l}{\sigma_{\tilde{u}}} \\ & + \left[\tilde{f}_{j-1}(x)\left(\kappa_{j-1} - x'\tilde{\beta}\right) - \tilde{f}_j(x)\left(\kappa_j - x'\tilde{\beta}\right)\right] \frac{x'\Omega_l + \psi_l}{\sigma_{\tilde{u}}^3} \end{aligned} \tag{8}$$

by using product and chain rules. In (8), Ω_l denotes the l-th column in Ω and ψ_l the l-th element in ψ, respectively, and $\tilde{f}(z) = \mathrm{d}\tilde{F}(z)/\mathrm{d}z$. The first term in (8) corresponds to the marginal probability effects in the standard model corrected for the standard deviation of the disturbance $\tilde{u}$. The second term arises because we assume a specific form of heteroscedasticity which makes the error term dependent on x. Consequently, marginal probability effects in the random coefficient model are more flexible than those in the standard model since the sign of the second term is indeterminate.

The random coefficients model can be estimated directly by the method of maximum likelihood with heteroscedasticity corrected index function. However, some caution is required in running the optimization routines. Although the parameters of the model are identified by functional form, the specific structure of the model might cause problems in some datasets. Specifically, certain values of Ω, ψ and x can drive $\sigma^2_{\tilde{u}}$ to be negative or its square root to be almost linear in the parameters, such that the argument in F gets complex or is not identified, respectively. Nevertheless, if the data support the model, we should find reasonable estimates of the elements in Ω and ψ.

3.3. Finite Mixture Model. The third approach is a finite mixture model for ordered data (Everitt, 1988; Everitt and Merette, 1990; Uebersax, 1999) which provides a very flexible way of modeling heterogeneity among groups of individuals. It is supposed that the population is split into C distinct latent classes and each class has its own data-generating process, i.e. we relax the distributional assumption of the standard model and its implied homogeneity. To fix ideas, let $c = 1, \dots, C$ denote the index of classes and write the cumulative probabilities for class c as

$$\Pr[y_c \leq j|x] = F(\kappa_{cj} - x'\beta_c) \ \equiv F_{cj}(x)$$

However, individual class membership is not observable and we assume that each individual belongs to a certain class c with probability π_c. Thus, we can write the cumulative probabilities of the observed outcomes as a mixture of class specific cumulative probabilities

$$\Pr[y \leq j|x] = \sum_{c=1}^{C} \pi_c F_{cj}(x) \tag{9}$$

where the π_c's sum up to unity. The probability density function of the ordered outcome is given by $\Pr[y = j|x] = \sum_c \pi_c \Big(F_{cj}(x) - F_{cj-1}(x)\Big)$ and marginal probability effects can be obtained, as before, by taking the first order derivative with respect to x_l

$$MPE_{jl}(x) = \sum_{c=1}^{C} \pi_c \Big(f_{cj-1}(x) - f_{cj}(x)\Big)\beta_{cl} \tag{10}$$

Again, the sign of marginal probability effects is indeterminate because of the dependence on π_c and β_{cl} which might differ in magnitude and sign among classes. The statistical significance of these differences can be tested by conducting a LR test with restrictions $\pi_1 = \ldots = \pi_C$ and $\beta_1 = \ldots = \beta_C$, that is, a total number of $(C-1)(k+1)$ restrictions. Uebersax (1999) gives conditions for identification of class specific thresholds and slope parameters.

The parameters of the finite mixture model can be estimated directly via maximum likelihood. This requires maximization of a (in general multimodal) log-likelihood function of the form

$$\ln L(\theta, \pi | y, x, z) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \ln \left\{ \sum_{c=1}^{C} \pi_c \Big(F_{cj}(x_i) - F_{cj-1}(x_i) \Big) \right\}$$

where θ and π is shorthand notation for the vectors of class specific parameters θ_c (which include thresholds and slopes) and probabilities π_c, respectively, and y_j is a binary variable indicating whether $y = j$. The multimodality of the log-likelihood function and the large number of parameters for increasing C might cause the optimization routines to be slow in finding the global maximum. Furthermore, although the probability function of the complete mixture might be well-defined, the probabilities in a subset of classes can turn negative. An alternative approach of getting the maximum likelihood estimates that circumvents these problems is to formulate the model as an incomplete data problem and to apply the EM algorithm of Dempster et al. (1977).

To be more specific, let m_c denote a binary variable indicating individual class membership which can be interpreted as independent realizations of a C-component multinomial distribution with component probabilities π_c, the prior probability of belonging to class c. The (complete-data) log-likelihood function for a random sample of size n conditional on observed class membership m can be written as

$$\ln L(\theta, \pi | y, x, m) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \sum_{c=1}^{C} m_{ci} \Big\{ \ln \pi_c + \ln \Big(F_{cj}(x_i) - F_{cj-1}(x_i) \Big) \Big\} \tag{11}$$

Since we cannot observe individual class membership, that is the data are incomplete, we cannot maximize this log-likelihood function directly.

The EM algorithm proceeds iteratively in two steps, based on an E-step in which the expectation of (11) is taken with respect to m given the observed data and the current fit of θ and π, and an M-step in which the log-likelihood function (11) is maximized with respect to θ and π given expected individual class membership. The linearity of the complete-data log-likelihood in m allows for direct calculation of the expected individual class membership given the observed data and the parameters obtained in the q-th iteration step. This expectation corresponds to the probability of the ith entity belonging to class c, henceforth called posterior probability τ_c. From the assumptions above or simply by Bayes' theorem it can be shown that

$$\tau_c \left(y, x; \theta^{(q)}, \pi^{(q)} \right) = \frac{\pi_c^{(q)} \left(F_{cj}^{(q)}(x) - F_{cj-1}^{(q)}(x) \right)}{\sum_{c=1}^{C} \pi_c^{(q)} \left(F_{cj}^{(q)}(x) - F_{cj-1}^{(q)}(x) \right)} \tag{12}$$

where $F_{cj}^{(q)}$ denotes the value of F evaluated at the parameters obtained in the q-th iteration step. These probabilities can be used to anaylze the characteristics of each class, i.e. we can assign each individual to the class for which its probability is the highest and then derive descriptive statistics or marginal probability effects per class.

The M-step replaces m_c in (11) by its expectation, τ_c, and therefore considers the expected log-likelihood to be maximized. Again, the linearity in (11) provides a substantial simplification of the optimization routine. First, updated estimates of $\pi_c^{(q+1)}$ can be obtained directly by taking the sample average $n^{-1}\sum_i \tau_c(.)$ where $0 \leq \tau_c(.) \leq 1$ (see (12)). Secondly, each class can be maximized separately with respect to θ_c to get updated estimates $\theta_c^{(q+1)}$ taking into account the multiplicative factor τ_c. In other words, we can estimate C simple ordered probits or logits while weighting the data appropriately and alter the E- and M-steps repeatedly until the change in the difference between the log-likelihood values is sufficiently small.

3.4. Sequential Model. The last alternative for a flexible ordered response model adopts methods from the literature on discrete time duration data. In this literature, the main quantity of interest is the conditional exit probability (or "hazard rate") $\Pr[y = j|y \geq j, x]$, where y is the duration of the spell and j is the time of exit. The key insight is that such discrete time hazard rate models can be used for any ordered response y. Once the conditional transition probabilities are determined, the unconditional probabilities are obtained from the recursive relationship

$$\Pr[y = j|x] = \Pr[y = j|y \geq j, x] \Pr[y \geq j|x] \qquad j = 1, \dots, J \tag{13}$$

where

$$\begin{aligned} &\Pr[y \geq 1|x] = 1 \\ &\Pr[y \geq j|x] = \prod_{r=1}^{j-1} \Big\{ 1 - \Pr[y = r|y \geq r, x] \Big\} \qquad j = 2, \dots, J \end{aligned} \tag{14}$$

and it is understood that $\Pr[y = J|y \geq J, x] = 1$. Using (13) and (14) the whole probability function of y can be expressed in terms of conditionals, or more precisely, as a sequence of binary choice models where each decision is made for a specific category j conditional on refusing all categories smaller than j. This kind of model can be motivated by a sequential response mechanism where each of the J outcomes can be reached only step-by-step, starting with the lowest category, and therefore the model is refered to as *sequential model*. This model implicitly accounts for the ordering information in y without assuming any cardinality in the threshold mechanism.

To complete the model we specify the conditional transition probabilities as

$$\Pr[y = j|y \geq j, x] = F(\alpha_j + x'\beta_j) = F_j(x) \qquad j = 1, \dots, J \tag{15}$$

where α_j is a category specific constant, β_j is a category specific slope parameter, and it is understood that $\alpha_J = \infty$ such that $F_J(\infty) = 1$. Therefore, in contrast to previously discussed models, we do not parameterize the cumulative probabilities but rather the conditional transition probabilities. The parameters can be estimated by running j consecutive binary choice models where the dependent variable is the binary indicator y_j defined in the previous section, and only observations with $y \geq j$ are included. Therefore, estimation is simplified considerably compared to the generalized threshold and the random coefficients model since no further restrictions on the parameter space are required. The downside is that computation of the marginal probability effects is now more complicated. It can be shown that

$$\begin{aligned} MPE_{1l}(x) &= f_1(x)\beta_{1l} \\ MPE_{jl}(x) &= f_j(x)\beta_{jl}\Pr[y \geq j|x] - F_j(x)\sum_{r=1}^{j-1} MPE_{rl}(x) \qquad j = 2, \ldots, J \end{aligned} \tag{16}$$

Clearly, these effects are very flexible, as they can vary by category and do not rely on a single crossing property or constant relative effects. The sequential model and the standard model are nonnested models and one may use information based measures like the *Akaike Information Criterion* (AIC) as a model selection criterion. Moreover, for the problem of choosing among the generalized alternatives the same strategy is advisable.

4. Empirical Illustration

In order to illustrate the benefit of the generalized ordered response models we analyze the effect of income on happiness using data from the German Socio-Economic Panel (GSOEP; see also Boes and Winkelmann, 2004). The relationship between income and happiness was studied before in a number of papers (see, for example, Easterlin, 1973, 1974; Scitkovsky, 1975; Frey and Stutzer, 2000, 2002; Shields and Wheatley Price, 2004 and the references therein) and has gained renewed interest in the recent literature because of its use for valuation of public goods or intangibles (see, for example, Winkelmann and Winkelmann, 1998; Frey, Luechinger, and Stutzer, 2002; van Praag and Baarsma, 2005).

We used data from the 1997 wave of the GSOEP and selected a sample of 1735 men aged between 25 and 65. The dependent variable *happiness* with originally 11 categories was recoded to avoid cells with low frequency and, after merging the lower categories 0/1/2 and 3/4, we retained a total of $J = 8$ ordered response categories. We included among the regressors logarithmic family income and logarithmic household size as well as a quadratic form in age, and two dummy variables indicating good health status as well as unemployment.

In our regression analysis, we assumed that F is the cumulative density function of the standard normal distribution. The random coefficients model was simplified by restricting Ω and ψ such that $\sigma_{\tilde{u}}^2 = \Omega_{ll}x_l^2 + 2\psi_l x_l + 1$,

TABLE 1. Model Selection

	Ordered Probit	Generalized Threshold	Sequential Probit	Random Coefficients	Finite Mixture
No. of param.	[13]	[49]	[49]	[15]	[26]
$\ln L$	-3040.58	-2999.59	-2999.12	-3035.88	-3024.65
AIC	6107.16	6097.18	6096.24	6101.76	6101.30
No. of obs.	1735				

Notes: The data were drawn from the 1997 wave of the German Socio-Economic Panel, the dependent variable *happiness* with originally eleven categories (0-10) was recoded to avoid cells with low frequency; we subsumed categories 0-2 in j=1, categories 3/4 in j=2, the remaining in ascending order up to j=8.

where x_l is assumed to be logarithmic income, Ω_{ll} denotes the l-th diagonal element in Ω and ψ_l the l-th element in ψ. Thus, we confine our analysis to parameter heterogeneity in the income coefficient, with all other parameters being deterministic. In the finite mixture model, we considered only two latent classes ($C = 2$). The following discussion proceeds in two steps: First, we evaluate the models by means of likelihood ratio tests and selection criteria, and second, we examine the implications for interpretion in terms of marginal probability effects.

The first question we address is whether one of the models presented above uses the information inherent in the data optimally. For this purpose, we perform likelihood ratio tests or AIC comparisons, depending on the situation. For example, the differences between the generalized threshold and the standard ordered probit model are statistically signifcant if we can reject the null hypothesis of no category specific parameters. This can be investigated by running a likelihood ratio test with minus two times the difference between the log-likelihoods of the standard and the generalized model as appropriate test statistic, showing a value of 79.98. The test statistic is asymptotically χ^2-distributed with 36 degrees of freedom. Thus, we can reject the null hypothesis, and thereby the standard orderd probit model. Likewise, we can compare the random coefficients model as well as the finite mixture model with the ordered probit, the latter being rejected in both cases. The sequential model and the standard ordered probit are nonnested models which rules out the application of a LR test. Instead, we may calculate the AIC for each model, showing values of 6107.96 and 6096.24 for the ordered probit and the sequential probit, respectively. A smaller value indicates a better fit while penalizing for the proliferation of parameters, and, although 36 parameters more, we favor the sequential probit to the ordered probit model. Furthermore, among the generalized alternatives the generalized threshold and the sequential model have the smallest AIC values, followed by the finite mixture model and the random coefficients model.

We now turn our attention to average marginal probability effects of income on happiness. The MPE's of the ordered probit model are reported

Table 2. Marginal Probability Effects of Income on Happiness

	Ordered Probit	Generalized Threshold	Sequential Probit	Random Coeff.	Finite Mixture	
					Class 1	Class 2
$j = 1$	-0.0076	-0.0098	-0.0083	-0.0165	-1.3e-07	-0.0245
$j = 2$	-0.0228	-0.0096	-0.0155	-0.0391	-0.0076	-0.0092
$j = 3$	-0.0223	-0.0352	-0.0338	-0.0297	-0.0024	-0.0565
$j = 4$	-0.0160	-0.0444	-0.0410	-0.0140	-0.0026	-0.0285
$j = 5$	-0.0090	0.0039	0.0095	0.0135	-0.0030	0.0198
$j = 6$	0.0328	0.0680	0.0697	0.0589	0.0028	0.0920
$j = 7$	0.0275	0.0403	0.0334	0.0234	0.0073	0.0069
$j = 8$	0.0173	-0.0133	-0.0140	0.0035	0.0056	5.7e-08

Notes: The table reports average marginal probability effects of logarithmic income on happiness responses, $AMPE_{j,\ln(income)}$. For example, in the ordered probit model $AMPE_{6,\ln(income)} = 0.0328$ means that the probability of $j = 6$ increases by about 0.0328 percentage points given an increase in logarithmic income by 0.01 (which corresponds to an increase in income by about 1 percent).

in the first column of table 2. Our results show a positive coefficient of logarithmic income, implying a negative sign of the MPE's for low happiness responses, switching into the positive for $j \geq 6$. The interpretation of, for example, $MPE_6 = 0.0328$ is that a one-percent increase in income raises the probability of $happiness = 6$ by approximately 0.0328 percentage points. Compared to the standard model, the generalized threshold and the sequential model yield substantially different effects (see columns 2 and 3). First, the sign of MPE_5 changes, indicating a positive effect also for the fifth category. Second, the magnitude of some MPE's are clearly underestimated by the standard model. For example, the estimated MPE_6 in the generalized ordered response models is more than twice as large as in the ordered probit. Third, and probably most important, the sign of the marginal probability effect in the utmost right part of the outcome distribution turns out to be negative, violating the single crossing requirement of the simple model. This means that an increase in income actually *reduces* the probability of being very happy, a result consistent with the view that "money does not buy happiness".

The results of the random coefficients model are reported in the fourth column of table 2. The calculated MPE's tend to support the results of the generalized threshold and the sequential model, although there is no negative effect on the highest happiness response. However, the random coefficient specification provides further insights into the relationship between income and happiness. We estimated $\hat{\Omega}_{ll} = 0.60$ and $\hat{\psi}_l = -0.77$, the latter implying that unobservables in the happiness equation are negatively correlated with the random coefficient. This can be interpreted as follows: If unobservables in the happiness equation tend to increase the probability of higher responses, then the effect of income is lower for these individuals.

In the finite mixture model we can make use of the posterior probabilities to obtain marginal probability effects per class (see columns 5 and 6). The results indicate that the effect of income on happiness can be neglected for one class (the relatively happy class with average happiness of 5.71) whereas for the class of relatively unhappy people (average happiness of 4.25) income plays a much more important role.

5. Concluding Remarks

In this paper we argued that the standard ordered probit and ordered logit models, while commonly used in applied work, are characterized by some restrictive and therefore non-desirable properties. We then discussed four generalized models, namely the generalized threshold, the random coefficients, the finite mixture, and the sequential model. All of them are substantially more flexible in analyzing marginal probability effects since they do not rely on constant relative effects or a single crossing property.

An illustrative application with data from the 1997 wave of the GSOEP dealt with the relationship between income and happiness. We asked how a one-percent increase in income is predicted to change the happiness distribution, ceteris paribus. The analysis showed that the estimated marginal probability effects differed markedly between the standard ordered probit model and the probit-specified alternatives. For example, a negative marginal effect for the highest answer category (as predicted by the generalized threshold model) is ruled out by assumption in the standard model.

As is not uncommon with such generalizations, they can be computationally burdensome due to the larger number of parameters, restrictions on the parameter space, or a multimodality of the likelihood function. Nevertheless, the greater flexibility and enhanced interpretation possibilities should render these alternative models indispensable tools in all research situations, where an accurate estimation of the marginal probability effects over the entire range of the outcome distribution is of interest.

References

Aitchison, J., and S.D. Silvey (1957). The Generalization of Probit Analysis to the Case of Multiple Repsonses. *Biometrika* **44** 131–140.

Agresti, A. (1999). Modelling Ordered Categorical Data: Recent Advances and Future Challenges. *Statistics in Medicine* **18** 2191–2207.

Anderson, J.A. (1984). Regression and Ordered Categorical Variables. *Journal of the Royal Statistical Society. Series B (Methodological)* **46** 1–30.

Barnhart, H.X., and A.R. Sampson (1994). Overview of Multinomial Models for Ordered Data. *Communications in Statistics – A. Theory and Methods* **23** 3395–3416.

Bellemare C., B. Melenberg, and A. van Soest (2002). Semi-parametric Models for Satisfaction with Income. *Portuguese Economic Journal* **1** 181–203.

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 No. 0407.

Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordered Logistic Regression. *Biometrics* **46** 1171–1178.

Clogg, C.C., and E.S. Shihadeh (1994). *Statistical Models for Ordered Variables.* Sage Publications, Thousand Oaks.

Cox, C. (1995). Location-Scale Cumulative Odds Models for Ordered Data: A Generalized Non-Linear Model Approach. *Statistics in Medicine* **14** 1191–1203.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39** 1–38.

Easterlin, R. (1973). Does Money Buy Happiness?. *Public Interest* **30** 3–10.

Easterlin, R. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramowitz* P. David, M. Reder, eds. 89–125. Academic Press, New York.

Everitt, B.S. (1988). A Finite Mixture Model for the Clustering of Mixed-Mode Data. *Statistics and Probability Letters* **6** 305–309.

Everitt, B.S., and C. Merette (1990). The Clustering of Mixed-Mode Data: A Comparison of Possible Approaches. *Journal of Applied Statistics* **17** 283–297.

Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data.* MIT PRess, Cambridge, MA.

Frey, B.S., S. Luechinger, and A. Stutzer (2004). Valuing Public Goods: The Life Satisfaction Approach. CESifo Working Paper No. 1158.

Frey, B.S., and A. Stutzer (2000). Happiness, Economy and Institutions. *The Economic Journal* **110** 918–938.

Frey, B.S., and A. Stutzer (2002). *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being.* Princeton University Press, Princeton and Oxford.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge.

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

McKelvey, R., and W. Zavoina (1975). A Statistical Model for the Analysis of Ordered Level Dependent Variables. *Journal of Mathematical Sociology* **4** 103–120.

Olsson, U. (1979). Maximum-Likelihood Estimation of the Polychoric Correlation Coefficient. *Psychometrika* **44** 443–460.

Ronning, G. (1990). The Informational Content of Responses from Business Surveys. In *Microeconometrics. Surveys and Applications* J.P. Florens, M. Ivaldi, J.J. Laffont, F. Laisney, eds. 123–144. Basil Blackwell, Oxford..

Ronning, G., and M. Kukuk (1996). Efficient Estimation of Ordered Probit Models. *Journal of the American Statistical Association* **91** 1120–1129.

Scitovsky, T. (1975). Income and Happiness. *Acta Oeconomica* **15** 45–53.

Shields, M., and S. Wheatley Price (2005). Exploring the Economic and Social Determinants of Psychological Well-Being and Perceived Social Support in England. *Journal of The Royal Statistical Society. Series A* **168** 513–537.

Snell, E.J. (1964). A Scaling Procedure for Ordered Categorical Data. *Biometrics* **20** 592–607.

Stewart, M.B. (2004). A Comparison of Semiparametric Estimators for the Ordered Response Model. *Computational Statistics and Data Analysis* **49** 555–573.

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

Tutz, G. (1990). Sequential Item Response Models with an Ordered Response. *British Journal of Mathematical and Statistical Psychology* **43** 39–55.

Tutz, G. (1991). Sequential Models in Ordered Regression. *Computational Statistics and Data Analysis* **11** 275–295.

Uebersax, J.S. (1999). Probit Latent Class Analysis with Dichotomous or Ordered Category Measures: Conditional Independence/Dependence Models. *Applied Psychological Measurement* **23** 283–297.

van Praag, B.M.S., and B.E. Baarsma (2005). Using Happiness Surveys to Value Intangibles: The Case of Airport Noise. *The Economic Journal* **115** 224–246.

Winkelmann, L., and R. Winkelmann (1998). Why Are the Unemployed So Unhappy? Evidence from Panel Data. *Economica* **65** 1–15.

Winship, C., and R.D. Mare (1984). Regression Models with Ordered Variables. *American Sociological Review* **49** 512-525.

Stefan Boes
Socioeconomic Institute
University of Zurich
Zuerichbergstr. 14, CH-8032 Zurich
Switzerland
boes@sts.unizh.ch

Rainer Winkelmann
Socioeconomic Institute
University of Zurich
Zuerichbergstr. 14, CH-8032 Zurich
Switzerland
winkelmann@sts.unizh.ch