

# 12 Ordered Response Models \*

Stefan Boes<sup>1</sup> and Rainer Winkelmann<sup>2</sup>

<sup>1</sup> Socioeconomic Institute, University of Zurich  
boes@sts.unizh.ch

<sup>2</sup> Socioeconomic Institute, University of Zurich  
winkelmann@sts.unizh.ch

**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.

## 12.1 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 ordinaly 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

---

\*We are grateful to an anonymous referee for valuable comments.

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 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 nonparametric approaches replace the distributional assumptions of the standard model, or the predictor function, by flexible semi- or nonparametric 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 *et al.* (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.