Sparser Ordinal Regression Models Based on Parametric and Additive Location-Shift Approaches

Tutz, Gerhard; Berger, Moritz

International Statistical Review

International Statistical Review (2022), **90**, 2, 306–327 doi: 10.1111/insr.12484

Sparser Ordinal Regression Models Based on Parametric and Additive Location-Shift Approaches

Gerhard Tutz[1] and Moritz Berger[2]

[1]*Institut für Statistik, Ludwig-Maximilians-Universität München, Akademiestraße 1, Munich, D-80779, Germany*
[2]*Institut für Medizinische Biometrie, Informatik und Epidemiologie, Medizinische Fakultät, Universität Bonn, Venusberg-Campus 1, Bonn, D-53127, Germany*
E-mail: moritz.berger@imbie.uni-bonn.de

Summary

The potential of location-shift models to find adequate models between the proportional odds model and the non-proportional odds model is investigated. It is demonstrated that these models are very useful in ordinal modelling. While proportional odds models are often too simple, non-proportional odds models are typically unnecessary complicated and seem widely dispensable. In addition, the class of location-shift models is extended to allow for smooth effects. The additive location-shift model contains two functions for each explanatory variable, one for the location and one for dispersion. It is much sparser than hard-to-handle additive models with category-specific covariate functions but more flexible than common vector generalised additive models. An R package is provided that is able to fit parametric and additive location-shift models.

Key words: Adjacent categories model; cumulative model; dispersion; location-shift model; ordinal regression; proportional odds model.

1 Introduction

The proportional odds model, which was propagated by McCullagh (1980), is probably the most widely used ordinal regression model. The assumption that effects of covariates are not category-specific makes it a simply structured model that allows to interpret parameters in terms of cumulative odds. However, in many applications, the model shows poor goodness-of-fit and does not adequately represent the underlying probability structure. As alternatives, non-proportional and partial proportional odds models were proposed. They allow for category-specific effects of explanatory variables and typically show much better fit than proportional odds models see, for example, Brant (1990), Peterson & Harrell (1990), Cox (1995), Bender & Grouven (1998), Kim (2003), Williams (2006), Liu *et al*. (2009) and Williams (2016).

A major disadvantage of non-proportional odds models is that many parameters are involved, which makes interpretation of parameters much harder than in the simple proportional odds model. Moreover, the space of explanatory variables can be almost empty, and estimation of parameters tends to fail in cases with a larger number of response categories. In the present paper, models between the proportional and non-proportional odds models are propagated. They are sufficiently complex to provide an adequate fit but contain much less parameters than the non-proportional odds model.

Non-proportional and proportional odds models are logistic versions of cumulative ordinal models with category-specific or global, that is, not category-specific effects of variables, respectively. We consider the more general class of cumulative models, which may use any response function that is determined by a strictly increasing distribution function. In addition, we consider the alternative class of adjacent categories models with general link functions. For all of these models, it is essential to find an adequate representation of data that does not involve too many parameters.

The class of models that is investigated contains a location term in the tradition of the proportional odds model (and other models with global parameters), but instead of using a multitude of category-specific parameters, the location term is complemented by a linear term that represents variability of the response, which may be seen as dispersion or, in questionnaires, the tendency of respondents to prefer extreme or middle categories. Parametric models of this type were considered by Tutz and Berger (2016; 2017a).

The paper has two main objectives. It is demonstrated that location-shift versions of cumulative and adjacent categories models are often adequate when modelling ordinal responses. They can be seen as a natural extension of proportional odds models that avoid the complexity of non-proportional odds models. Indeed, non-proportional odds type models turn out to be a frequently dispensable class of models. They are unnecessarily complicated and hardly needed in ordinal modelling. In contrast to most statistical papers, which propagate more complex modelling, in the first part of the paper, we plead for a simpler class of models instead of a complexer one. In the second part of the paper, location-shift models are extended to allow for smooth effects of covariates. Extensions of additive models that include general category-specific effects are rather hard to obtain. The proposed additive location-shift model offers a way to go beyond the simple global effects model without adding too many functions.

In Section 2, parametric ordinal models and their location-shift versions are considered, including details on inference. Section 3 contains illustrative simulations. In Section 4, it is demonstrated that in applications, non-proportional odds models are often not needed. Traditional additive ordinal models are briefly considered in Section 5, and the additive location-shift model is introduced as an alternative. The section is complemented by further simulations and applications. Available software is described in Section 6.

2 Ordinal Regression Models

2.1 Proportional and Non-Proportional Odds Models

The most widely used ordinal regression model is the proportional odds model, which is a member of the class of cumulative models. Cumulative models can be derived from an underlying latent variable. Let Y^* be an underlying latent variable for which the regression model $Y^* = -\boldsymbol{x}^T\boldsymbol{\beta} + \epsilon$ holds, where ϵ is a noise variable with continuous distribution function $F(.)$, $\boldsymbol{x}$ is a vector of explanatory variables, and $\boldsymbol{\beta}$ a vector of coefficients. If one assumes that the link between the observable categorical response Y and the latent trait is specified by $Y = r \Leftrightarrow \theta_{r-1} < Y^* \leq \theta_r$, where $-\infty = \theta_0 < \theta_1 < \ldots < \theta_k = \infty$, one obtains the *cumulative model*

$$P(Y \leq r|\boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}), \quad r = 1, \ldots, k-1, \tag{1}$$

where the category-specific intercepts β_{0r} are identical to the thresholds on the latent scale, that is, $\beta_{0r} = \theta_r$. If one uses the logistic distribution $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$, one obtains the *proportional odds model*

$$\text{logit}P(Y \leq r|\boldsymbol{x}) = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}.$$

The strength of the model is that interpretation of parameters is very simple. Let $\gamma_r(\boldsymbol{x}) = P(Y \leq r|\boldsymbol{x})/P(Y > r|\boldsymbol{x})$ denote the cumulative odds, then e^{β_j} can be directly interpreted as the odds ratio that compares the cumulative odds with value $x_j + 1$ in the j-th variable to the odds with value x_j in the j-th variable, when all other variables are kept fixed,

$$e^{\beta_j} = \frac{\gamma_r(x_1, \dots, x_j + 1, \dots, x_p)}{\gamma_r(x_1, \dots, x_j, \dots, x_p)}. \tag{2}$$

It is important that the interpretation does not depend on the category, e^{β_j} is the same for all odds γ_r, $r = 1, \dots, k - 1$. The independence of parameters on categories holds for the whole class of cumulative models (1) because they share the stochastic ordering property, which means that for two sets of explanatory variables $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ the term

$$F^{-1}(P(Y \leq r|\boldsymbol{x})) - F^{-1}(P(Y \leq r|\tilde{\boldsymbol{x}})) = (\boldsymbol{x} - \tilde{\boldsymbol{x}})^T\boldsymbol{\gamma},$$

does not depend on the category r.

Early versions of the cumulative logistic model were given by Snell (1964), Walker & Duncan (1967) and Williams & Grizzle (1972). More general cumulative models were considered, among others, by Genter & Farewell (1985), Armstrong & Sloan (1989), Ananth & Kleinbaum (1997) and Steadman & Weissfeld (1998). Campbell & Donner (1989) and Rudolfer *et al*. (1995) investigated their use in prediction, and more recently, robust estimators have been proposed by Iannario *et al*. (2017).

The problem with cumulative models of the form (1) is that they often do not fit the data well, which calls for more complicated models. A class of models that has been considered in the literature is the *cumulative model with category-specific effects*

$$P(Y \leq r|\boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}_r), \;\; r = 1, \dots, k - 1, \tag{3}$$

which uses the parameter vectors $\boldsymbol{\beta}_r^T = (\beta_{1r}, \dots, \beta_{pr})$ and allows that parameters vary across categories. Logistic models of this type are also called *non-proportional odds models* to distinguish them from the simpler versions. If $\boldsymbol{\beta}_1 = \dots = \boldsymbol{\beta}_{k-1}$, the model simplifies to the simple cumulative model (1).

Of course, not for all variables, the parameters have to vary over categories, which might yield two types of variables, variables with a *global* effect, for which $\beta_{j1} = \dots = \beta_{j,k-1} = \beta_j$, and variables with *category-specific effects*, that is, $\beta_{js} \neq \beta_{jr}$ for at least two categories s, r. Peterson & Harrell (1990) distinguished between these two effect types when considering the so-called *partial proportional odds model*. To distinguish between the effects of variables, it is helpful to consider two distinct vectors of explanatory variables $\boldsymbol{w}$ and $\boldsymbol{z}$. Then, the model has the form

$$P(Y \leq r|\boldsymbol{x}, \boldsymbol{z}) = F(\beta_{0r} + \boldsymbol{w}^T\boldsymbol{\beta}^w + \boldsymbol{z}^T\boldsymbol{\beta}_r^z), \;\; r = 1, \dots, k - 1. \tag{4}$$

While the effects of $\boldsymbol{w}$ are global the effects of $\boldsymbol{z}$ are category-specific. If no z-variables are included, one obtains the proportional odds model, if no w-variables are present, one obtains the general model with category-specific effects (3) (with renamed variables). The interpretation of the effects of w-variables is the same as in the proportional odds model, whereas interpretation of the effects of z-variables is harder and has to refer to specific response categories.

Partial proportional odds models have been investigated, for example, by Brant (1990), Peterson & Harrell (1990), Cox (1995), Bender & Grouven (1998), Kim (2003) and Liu *et al*. (2009). In particular, tests are provided that can distinguish between variables with global and category-specific effects. In sociology, they are also referred to as generalised ordered logit models (Williams, 2006; 2016).

The general model with category-specific effects is attractive because it usually provides a better fit to the data. In addition, if effects vary strongly across categories, one might miss some effects that show only if one allows for category-specific effects, see, for example, the retinopathy study given in Tutz (2012), Example 9.1, p. 251. However, the whole class of models has some serious disadvantages. One is that one has many parameters, which are much harder to interpret. More seriously, the possible values of explanatory variables can be strongly restricted because it is postulated that $\beta_{01} + \boldsymbol{x}^T\boldsymbol{\beta}_1 \leq \ldots \leq \beta_{0,k-1} + \boldsymbol{x}^T\boldsymbol{\beta}_{k-1}$ for all values $\boldsymbol{x}$. Even if estimates exist, in future observations with more extreme values in the explanatory variables, the estimated probabilities can be negative. For problems with the model, see also Walker (2016) who even concludes that it is impossible to generalise the cumulative class of ordered regression models in ways consistent with the spirit of generalised cumulative regression models.

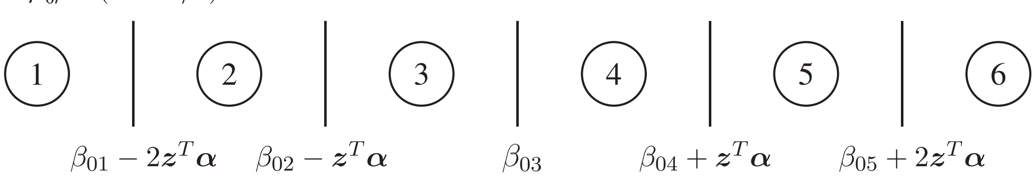
2.2 Cumulative Location-Shift Models

An alternative extension of the non-proportional odds model was proposed by Tutz & Berger (2017a). They assume that variables may change the thresholds of the underlying latent trait. Then, thresholds β_{0r} in the proportional odds model are replaced by $\beta_{0r} + (k/2 - r)\boldsymbol{z}^T\boldsymbol{\alpha}$, where $\boldsymbol{z}$ denotes a vector of covariates, possibly containing components of $\boldsymbol{x}$. The replacement yields the so-called location-shift model, which is given in closed form by

$$P(Y \leq r|\boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} + (r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha}), \ \ r = 1, \ldots, k-1. \tag{5}$$

It contains the familiar location term $\boldsymbol{x}^T\boldsymbol{\beta}$, which models the location on the latent continuum and therefore the tendency to low or high response categories. In addition, it contains the scaled shifting term $(r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha}$, which modifies the thresholds and has a quite different interpretation.

The term $\boldsymbol{z}^T\boldsymbol{\alpha}$ determines the shifting of thresholds, whereas the scaling factor $(r - k/2)$ is an additional weight chosen such that the difference between thresholds are widened or shrunk by the same amount. For illustration, let us consider the case $k = 6$, for which the modified thresholds $\beta_{0r} + (r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha}$ have the form



In general, thresholds are widened if $\boldsymbol{z}^T\boldsymbol{\alpha}$ is positive, and shrunk if it is negative. The consequence is that one observes more concentration in the middle or extreme categories, respectively. Because more concentration in the middle means less variability, the term $\boldsymbol{z}^T\boldsymbol{\alpha}$ can also be seen as representing dispersion. For positive values, the distribution is more concentrated, meaning smaller dispersion, whereas for negative values, one has larger dispersion. The effect is also seen from considering the differences between adjacent predictors,

$$\eta_r - \eta_{r-1} = \beta_{0r} - \beta_{0,r-1} + \boldsymbol{z}^T\boldsymbol{\alpha}, \ \ r = 2, \dots, k-1, \tag{6}$$

where $\eta_r = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} + (r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha}$ is the r-th predictor. For positive values of $\boldsymbol{z}^T\boldsymbol{\alpha}$, the difference between adjacent predictors becomes greater, whereas for negative values, it becomes smaller. Thus, $\boldsymbol{\alpha}$ represents the tendency to middle or extreme categories linked to covariates $\boldsymbol{z}$, which is separated from the location effect $\boldsymbol{x}^T\boldsymbol{\beta}$. It can be derived that the interpretation of the β parameters is the same as in the proportional odds model if $\boldsymbol{x}$ and $\boldsymbol{z}$ are distinct (Tutz & Berger, 2017a).

2.3 Nested Structure of Cumulative Models

A different view of the cumulative model is obtained by seeing it as a non-proportional odds model with specific constraints on the parameters. Let us consider the general case $\boldsymbol{x} = \boldsymbol{z}$. Then, one has

$$P(Y \le r|\boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T(\boldsymbol{\beta} + (r - k/2)\boldsymbol{\alpha})) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}_r),$$

where $\boldsymbol{\beta}_r = \boldsymbol{\beta} + (r - k/2)\boldsymbol{\alpha}$. The model is equivalent to a category-specific model with constraints

$$(\boldsymbol{\beta}_r - \boldsymbol{\beta})/(r - k/2) = \boldsymbol{c}, \ \ r = 1, \dots, k-1,$$

where $\boldsymbol{\beta} = \sum_{r=1}^{k-1}\boldsymbol{\beta}_r$ and $\boldsymbol{c}$ is a vector of constants. If the category-specific model with constraints is assumed to hold, the vector $\boldsymbol{c}$ turns out to be $\boldsymbol{\alpha}$.

That means, in particular, that the location-shift model is a submodel of the model with category-specific effects. Because the proportional odds model is a submodel of the location-shift model, one has the nested structure

proportional odds model ⊂ location-shift model ⊂ non-proportional odds model

or, more generally,

model with global effects ⊂ location-shift model ⊂ model with category-specific effects.

Because the location-shift model is a (multivariate) generalised linear model, one can investigate if the models can be simplified by testing the sequence of nested models, see, for example, Tutz (2012).

To clarify the relation between the partial proportional odds model and the location-shift model, let $\boldsymbol{x}$ and $\boldsymbol{\beta}$ be partitioned into two subvectors, $\boldsymbol{x}^T = (\boldsymbol{w}^T, \boldsymbol{z}^T)$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}^{wT}, \boldsymbol{\beta}^{zT})$, such that $\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{w}^T\boldsymbol{\beta}^w + \boldsymbol{z}^T\boldsymbol{\beta}^z$. Then, one obtains for the linear predictor

$$\eta_r = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} + (r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha} = \beta_{0r} + \boldsymbol{w}^T\boldsymbol{\beta}^w + \boldsymbol{z}^T\boldsymbol{\beta}^z + (r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha} = \beta_{0r} + \boldsymbol{w}^T\boldsymbol{\beta}^w + \boldsymbol{z}^T\boldsymbol{\beta}^z_r,$$

where $\boldsymbol{\beta}^z_r = \boldsymbol{\beta}^z + (r - k/2)\boldsymbol{\alpha}$. Thus, the predictor is very similar to the predictor of the partial proportional odds model (4), which has the general form $\eta_r = \beta_{0r} + \boldsymbol{w}^T\boldsymbol{\beta}^w + \boldsymbol{z}^T\boldsymbol{\beta}^z_r$. However, there is one crucial difference. The category-specific parameters in the partial proportional odds model are not constrained whereas the parameters in the location-shift model are constrained to have the form $\boldsymbol{\beta}^z_r = \boldsymbol{\beta}^z + (r - k/2)\boldsymbol{\alpha}$. This constraint makes the location-shift model a more restricted model with the advantage that the α-parameters can be interpreted as effects that represent a tendency to middle or extreme categories. Thus, the location-shift model is a submodel of the partial proportional odds model that allows for easier interpretation of effects.

One of the disadvantages of non-constrained model versions with category-specific effects is that the simple interpretation of parameters gets lost. One has a multitude of parameters for which one might easily lose track. For example, if one has just 4 variables and 10 categories (see the example in Section 4.1), the model contains 45 parameters, for each variable one has 9 parameters. In contrast, the proportional odds model contains only 13 parameters, and the impact of one variable is described by just one parameter. The models propagated here are models that are between the most general model and the model with global effects, in them the impact of a single variable is described by just two parameters (instead of $k-1$ parameters as in the general model and one in the model with global effects).

2.4 Adjacent Categories Models

An alternative class of models for ordinal responses are *adjacent categories models*. In its simple version, they assume

$$P(Y > r | Y \in \{r, r+1\}, \boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}), \quad r = 1, \ldots, k - 1. \tag{7}$$

where $F(.)$ again is a strictly increasing distribution function but no ordering of intercepts has to be postulated. The logistic version has the form

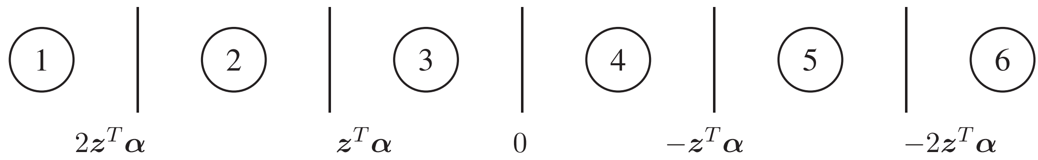
$$\log\left(\frac{P(Y = r+1|\boldsymbol{x})}{P(Y = r|\boldsymbol{x})}\right) = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}, \quad r = 2, \ldots, k - 1. \tag{8}$$

The interpretation of parameters is as simple as for basic cumulative models; e^{β_j} is the odds ratio that compares the odds with value $x_j + 1$ in the j-th variable to the odds with value x_j in the j-th variable; however, the odds are not cumulative odds but adjacent categories odds, $\gamma_r(\boldsymbol{x}) = P(Y = r+1|\boldsymbol{x})/P(Y = r|\boldsymbol{x})$.

In the same way as in the cumulative models, the linear predictor can be replaced by a predictor with category-specific parameters, that is, predictors $\eta_r = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}_r$ to obtain a better fit. The corresponding model contains many parameters, which are harder to interpret. A sparser model is the *location-shift version of the adjacent categories model*

$$\log\left(\frac{P(Y = r+1|\boldsymbol{x})}{P(Y = r|\boldsymbol{x})}\right) = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} + (k/2 - r)\boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = 1, \ldots, k - 1. \tag{9}$$

For $k = 6$, one obtains for the term $(k/2 - r)\boldsymbol{z}^T\boldsymbol{\alpha}$, which distinguishes between category r and $r+1$



Thus, one obtains the same effects as in the cumulative location-shift model, if $z^T\alpha$ is large the person has a tendency to choose middle categories, if $z^T\alpha$ is small there is a tendency to extreme categories.

For the adjacent categories model, the same hierarchy holds as for the cumulative models. The model with global effects is a submodel of the adjacent categories location-shift model, which is a submodel of the general model with category-specific effects.

2.5 Inference

All the models considered in the previous sections can be embedded within the framework of multivariate generalised linear models (GLMs). That means they have the form

$$g(\boldsymbol{\pi}) = \boldsymbol{X\delta} \text{ or } \boldsymbol{\pi} = h(\boldsymbol{X\delta}),$$

where $\boldsymbol{\pi}^T = (\pi_1, \ldots, \pi_q),\ q = k - 1$, is the vector of the response probabilities with components $\pi_r = P(Y = r|\boldsymbol{x}, \boldsymbol{z})$, $\boldsymbol{X}$ is a design matrix constructed from the predictors $\boldsymbol{x}$ and $\boldsymbol{z}$, $\boldsymbol{\delta}$ is the total parameter vector, $g = (g_1, \ldots, g_q): \mathbb{R}^q \to \mathbb{R}^q$ is a vector-valued *link function* and $h(\cdot) = g(\cdot)^{-1}$ is the response function. The components of the vector $\boldsymbol{X\beta}$ are the linear predictors $(\eta_1, \ldots, \eta_q)$. Details of the representation of classical cumulative and adjacent categories models as multivariate GLMs are found in Fahrmeir & Tutz (2001) and Tutz (2012), and for the shifted versions, see Tutz & Berger (2016) and Tutz & Berger (2017b). The representation as multivarate GLMs allows to use all the tools that have been developed for that class of models, including algorithms to obtain estimates and standard errors. Also testing of effects, analysis of residuals and goodness-of-fit tests developed for GLMs can be used. Programme packages that can be used are described in Section 6.

3 Simulation Study

To illustrate the properties of the models, we show the results of a small simulation study when the data-generating model is known, starting with the cumulative model.

We consider an ordinal response with $k = 5$ response categories and a normally distributed covariate $x \sim N(0, 0.5)$. We generated data from (a) a proportional odds model with $\beta = 1$, (b) from a non-proportional odds model with parameter vector $(\beta_1, \beta_2, \beta_3, \beta_4)^\top = (-1, 0, 1, 2)$, (c) a non-proportional odds model with parameter vector $(\beta_1, \beta_2, \beta_3, \beta_4)^\top = (-1, 0, 0.8, 1)$ and (d) a non-proportional odds model with parameter vector $(\beta_1, \beta_2, \beta_3, \beta_4)^\top = (-0.5, 0.5, -0.5, 0.5)$. Importantly, the model in scenario (b) corresponds to a cumulative location-shift model with parameters $\beta = 0.5$ and $\alpha = 1$. In all four scenarios, we considered $n = 500$ observations (100 replications) and set the category-specific intercepts to $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})^\top = (-3, -1, 1, 3)$.

To evaluate the performance of the models, we computed the fitted linear predictors $\hat{\eta}_{ir},\ i = 1, \ldots, n,\ r = 1, \ldots, 4$, respectively, and investigated if the models can be simplified using likelihood ratio tests. Figure 1 shows the squared differences $(\hat{\eta}_{ir} - \eta_{ir})^2$ averaged over observations and categories. If the proportional odds model is the data-generating model

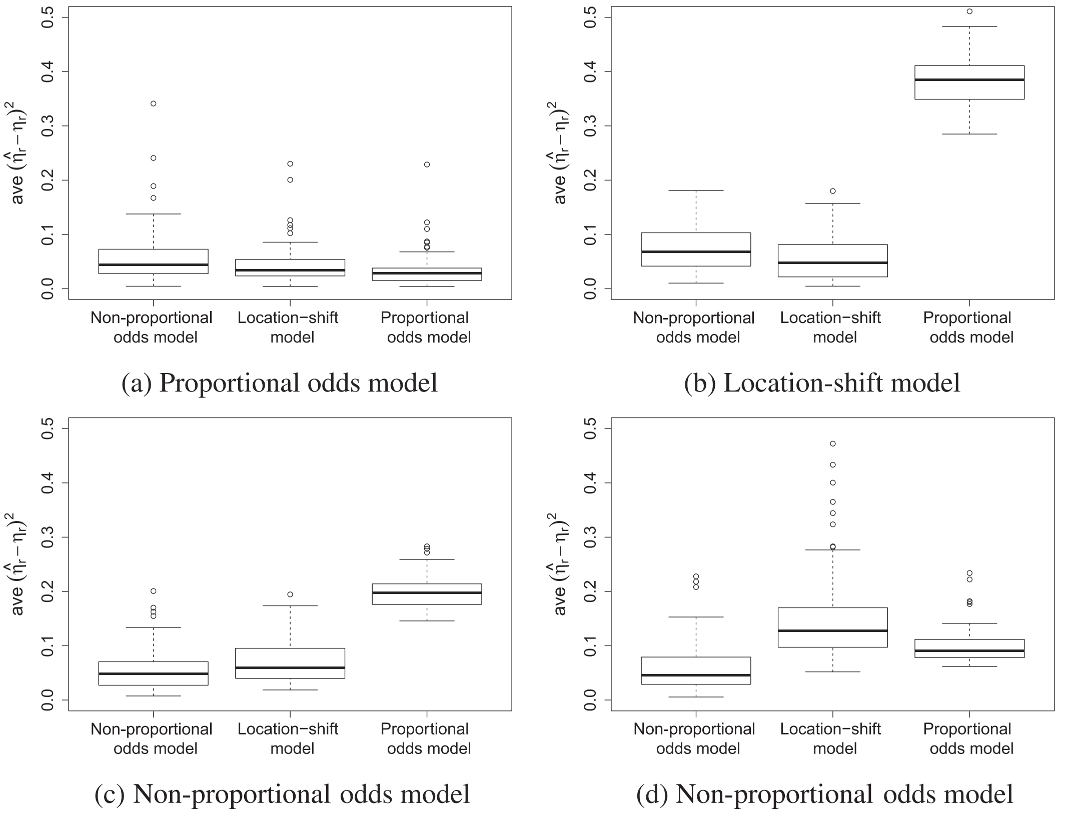


Figure 1. *Results of the simulation study (cumulative models). The boxplots show the squared differences* $(\hat{\eta}_{ir} - \eta_{ir})^2$ *averaged over all 500 observations and all four categories, when fitting the proportional odds model (right), the location-shift model (middle) and the non-proportional odds model (left). The true data-generating model is indicated in the subheadings. Note that due to the ordering constraint of the non-proportional odds model, we restricted the range of values of x to [− 2, 2]*

(Scenario a), it shows the best performance (selected in 91 replications), but the location-shift model is close. If the location-shift model is the data-generating model (Scenario b), the proportional odds model performs very poorly. If the data stem from a non-proportional odds model with increasing parameter (Scenario c), the proportional odds model performs poorly while the performance of the location-shift model is comparable with the fitting of the non-proportional odds model. Only if the data come from a non-proportional odds model with strongly varying coefficients (Scenario d), it distinctly outperforms the other models (selected in 96 replications). It is seen that the proportional odds model is a good choice only if it really holds. In all other cases, the performance deteriorates, in some cases very strongly. Table 1 (upper part) shows which models are chosen if one tests the hierarchy of models with significance level 0.05. If the proportional odds model holds (Scenario a), the non-proportional odds model is retained only in 5% of the cases, which corresponds to the significance level. If the location-shift model holds, it is chosen in 87% of the cases. If the non-proportional odds model holds but with weakly varying coefficients (Scenario c), the location-shift model is often chosen as an acceptable approximation to the non-proportional odds model. Larger sample sizes would be necessary to distinguish between the two models. However, if a non-proportional odds model with distinctly varying coefficients holds, it is chosen in almost all of the cases.

In the second part, we generated data from logistic adjacent categories models with the following changes: as no ordering of predictors is postulated, we considered multivariable models

Table 1. *Results of the simulation study*

Scenario	*Non-proportional odds model*	*Location-shift model*	*Proportional odds model*
Cumulative models			
a	5	4	91
b	8	87	5
c	14	86	0
d	96	0	4
Adjacent categories models			
a	10	3	87
b	5	95	0
c	23	77	0
d	98	1	1

Number of simulation runs in which the models have been selected when comparing the sequence of nested models using likelihood ratio tests.

with five normally distributed covariates $x_1, \dots, x_5 \sim N(0, 0.5)$. For each covariate, we chose the same β-parameters as before, with the exception of Scenario (b) where we used the reversed parameter vectors $(\beta_{j1}, \beta_{j2}, \beta_{j3}, \beta_{j4})^\top = (2, 1, 0, -1)$, $j = 1, \dots, 5$. The category-specific intercepts were set to $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})^\top = (2, 2/3, -2/3, -2)$. The results are shown in Table 1 (lower part) and Figure 2, which confirm the previous findings obtained for the cumulative class of models.

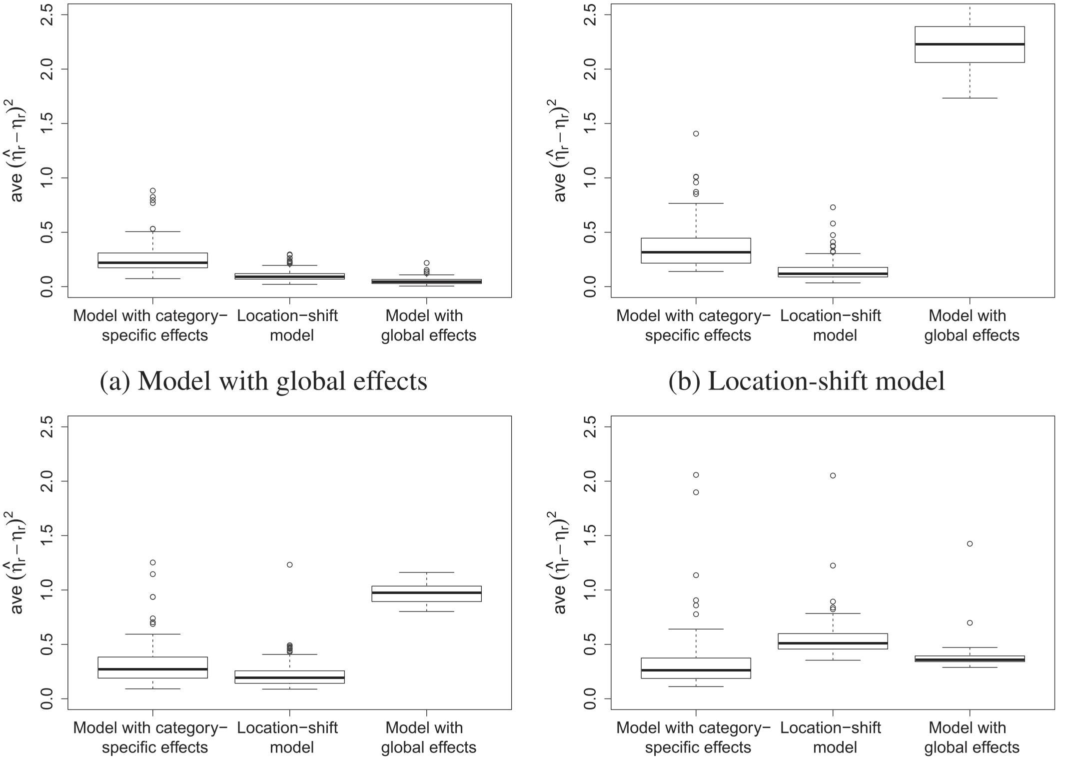
4 Applications

To demonstrate that the location-shift model is frequently a good choice that shows satisfactory goodness-of-fit while being comparably sparse in parameters, we consider several applications.

4.1 Safety in Naples

The package CUB (Iannario *et al*., 2015) contains the data set relgoods, which provides results of a survey aimed at measuring the subjective extent of feeling safe in the streets. The data were collected in the metropolitan area of Naples, Italy. Every participant was asked to assess on a 10-point ordinal scale his/her personal score for feeling safe with large categories referring to feeling safe. There are $n = 2225$ observations and four variables, *Age*, *Gender* (0: male, 1: female), *Residence* (1: City of Naples, 2: District of Naples, 3: Others Campania, 4: Others Italia) and the educational degree (*EduDegree*; 1: compulsory school, 2: high school diploma, 3: Graduated-Bachelor degree, 4: Graduated-Master degree, 5: Post graduated).

Table 2 (upper part) shows the deviances of the fitted models and the differences for the cumulative models. The full model with category-specific effects has 90 parameters, which reduces to 27 parameters in the location-shift model. The difference in deviances suggests that the full model can be simplified to the location-shift model; however, it certainly does not simplify to the model with global effects (difference of deviances 49.32 on 9 df). That means the location-shift model contains enough structure to explain the effect of covariates on the response, but the simpler structure without the term $(r - k/2)\boldsymbol{z}^T\boldsymbol{\alpha}$ is too simple, that is, relevant effects are missing. It is noteworthy that in this application, as in the other applications used here, the sample size is rather large ($n = 2225$). Typically, if sample sizes are large, one finds more significant effects. Therefore, it is remarkable that the complex model with category-specific effects can be simplified in spite of the large sample size.



(a) Model with global effects (b) Location-shift model

(c) Model with category-specific effects (d) Model with category-specific effects

Figure 2. *Results of the simulation study (adjacent categories models). The boxplots show the squared differences* $(\hat{\eta}_{ir} - \eta_{ir})^2$ *averaged over all 500 observations and all four categories, when fitting the model with global effects (right), the location-shift model (middle) and the model with category-specific effects (left). The true data-generating model is indicated in the subheadings*

Table 2. *Fits of models with logistic link for safety data*

	Deviance	*df*	*Difference in deviances*	*df*	p *value*
Cumulative models					
Non-proportional odds model	9 825.78	19 935			
Location-shift model	9 899.67	19 998	73.89	63	0.1640
Proportional odds model	9 948.99	20 007	49.32	9	0.0000
Adjacent categories models					
Model with category-specific effects	9 828.07	19 935			
Location-shift model	9 902.43	19 998	74.36	63	0.1549
Model with global effects	9 959.00	20 007	56.57	9	0.0000

Similar results are found when using the adjacent categories approach. In the lower part of Table 2, the fits for logistic adjacent categories models are given. It is seen that there is no need to use the general model with category-specific effects because the difference in deviances between the general model and the location-shift version is not significant. However, the location-shift model can not be reduced to the model with global effects. The fits are well comparable with the fits obtained for the cumulative models given in Table 2. For the adjacent categories as well as for the cumulative modelling approach, the location-shift versions turn out to be the best compromise between goodness-of-fit and sparsity. The fits of the cumulative versions are slightly better than that of the adjacent categories location-shift version.

In the following, we briefly compare the estimates of the alternative modelling approaches. The estimates of the proportional odds model and the cumulative logistic location-shift model are given in Table 3. It is seen that the dispersion effects are not negligible. All variables show rather small p values in the dispersion component, which explains the strong difference in deviances between the location-shift model and the simple model with global effects, which does not account for varying dispersion. It is seen that the simple proportional odds model yields stronger location effects than the location-shift model, which is a hint that estimates might be biased if dispersion effects are ignored. The same pattern is found for the adjacent categories models (not given).

Instead of showing all the parameters, we use the plotting tool provided by our R package (see Section 6). In Figure 3, the location effects of age, gender and residence are plotted against the dispersion effects (left: cumulative model, right: adjacent categories model). The abscissa represents the multiplicative dispersion effect on the odds $e^{\hat{\alpha}}$, and the ordinate axis represents the multiplicative location effect $e^{\hat{\beta}}$ for the variables. In addition to the point estimates, pointwise 95% confidence intervals are included. The horizontal and vertical lengths of the stars correspond to the confidence intervals of $e^{\hat{\alpha}}$ and $e^{\hat{\beta}}$, respectively. Thus, the stars also show the significance of effects. If the stars cross the line $y = 1$, location effects have to be considered as significant, and if they cross the line $x = 1$, dispersion effects have to be considered as significant.

To make the models comparable, we did not use the classical representation of the cumulative model. We used the reverse categories representation $P(Y \geq r|\boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} + (k/2 - r)\boldsymbol{z}^T\boldsymbol{\alpha})$. Then, the location effects have the same interpretation as in the adjacent categories model, large values of $\boldsymbol{x}^T\boldsymbol{\beta}$ indicate a preference for high response categories while small values indicate a preference for low categories. The resulting star plots for both models, the cumulative and the adjacent categories model, are very similar. It is seen that people living outside

Table 3. *Estimates of proportional odds model and cumulative logistic location-shift model for safety data*

	Proportional odds model				*Location-shift model*			
	Coef	*SE*	z *value*	p *value*	*Coef*	*SE*	z *value*	p *value*
Location effects								
Age	−0.045	0.026	−1.713	0.086	−0.041	0.026	−1.578	0.114
Gender	−0.343	0.075	−4.563	0.000	−0.327	0.075	−4.343	0.000
Residence2	0.518	0.090	5.705	0.000	0.572	0.092	6.199	0.000
Residence3	0.899	0.117	7.644	0.000	0.938	0.119	7.859	0.000
Residence4	1.397	0.141	9.885	0.000	1.339	0.148	9.039	0.000
EduDegree2	−0.307	0.111	−2.748	0.006	0.274	0.112	−2.449	0.014
EduDegree3	−0.319	0.150	−2.118	0.034	0.294	0.151	−1.947	0.051
EduDegree4	−0.162	0.159	−1.021	0.307	0.112	0.160	−0.704	0.481
EduDegree5	−0.292	0.221	−1.319	0.187	−0.261	0.221	−1.177	0.239
Dispersion effects								
Age					−0.018	0.007	−2.442	0.014
Gender					0.045	0.022	2.039	0.041
Residence2					−0.085	0.028	−2.984	0.002
Residence3					−0.090	0.036	−2.458	0.013
Residence4					−0.155	0.042	−3.703	0.000
EduDegree2					0.099	0.030	3.297	0.000
EduDegree3					0.142	0.044	3.176	0.001
EduDegree4					0.163	0.047	3.437	0.000
EduDegree5					0.107	0.064	1.668	0.095

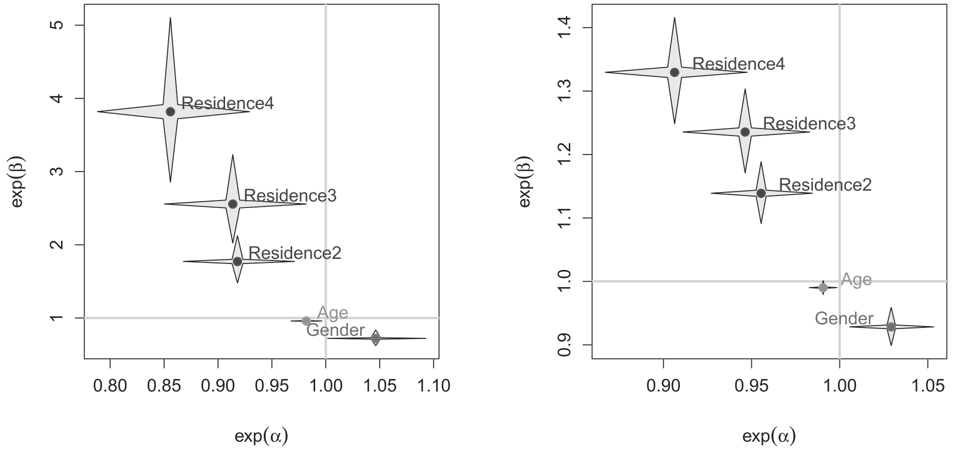


Figure 3. *Plots of* $(e^{\hat{\alpha}}, e^{\hat{\beta}})$ *for safety data, y-axis represents location, x-axis represents dispersion, left: cumulative location-shift model, right: adjacent categories location-shift model*

of the city of Naples feel much safer (y-axis, reference category: city of Naples), effects are ordered and the larger the distance to the city, the safer they feel. Females and older people feel less safe (below the line $y = 1$). As far as concentration of responses is concerned, people living outside the city of Naples and older people have stronger dispersion (below $x = 1$) while females show less dispersion than men. It should be noted that age is measured in decades, otherwise the age effect would be too close to zero in the star plot. Although the estimated parameter values are quite different for the cumulative and the adjacent categories models, the conclusions one draws are very similar, as are the goodness-of-fits. Thus, one might use either of the two models to investigate the impact of variables. However, it is certainly warranted to account for dispersion effects.

4.2 Nuclear Energy

The German Longitudinal Election Study (GLES) is a long-term study of the German electoral process (Rattinger *et al*., 2014). The data consist of 2 036 observations and originate from the pre-election survey for the German federal election in 2017 and are concerned with political fears. In particular, the participants were asked: ‘How afraid are you due to the use of nuclear energy?’ The answers were measured on Likert scales from 1 (*not afraid at all*) to 7 (*very afraid*). The explanatory variables used in the model are *Aged* (age of the participant), *Gender* (1: female; 0: male) and *EastWest* (1: East Germany/former GDR; 0: West Germany/former FRG). Table 4 shows the fits of cumulative and adjacent categories models, respectively.

Table 4. *Fits of models with logistic link for response fear of nuclear energy*

	Deviance	*df*	*Difference in deviances*	*df*	p *value*
Cumulative models					
Model with category-specific effects	7 499.61	12 192			
Location-shift model	7 506.36	12 204	6.75	12	0.873
Model with global effects	7 544.60	12 206	38.24	2	0.000
Adjacent categories models					
Model with category-specific effects	7 500.77	12 192			
Location-shift model	7 508.72	12 204	7.95	12	0.997
Model with global effects	7 545.41	12 206	36.69	2	0.000

Comparison between the full models with category-specific effects and location-shift versions yields p values greater than 0.8. It is obvious that the location-shift versions of the models represent satisfying approximations while models with global parameters should not be used to describe the underlying response structure.

The strength of effects is seen from the stars in Figure 4, which shows parameter estimates for the cumulative location-shift model on the left and the adjacent categories model on the right-hand side. Again, we used the inverse order of categories in the cumulative model and age measured in decades. It is seen that all parameters have significant location and dispersion effects with the exception of EastWest, for which the dispersion effect is not distinctly significant. It is seen that females and older people are more afraid of the consequences of the use of nuclear energy while residents of the Eastern part are less afraid. Females show stronger dispersion than men, and older people less dispersion than younger respondents.

4.3 Climate Change

Let us again consider the GLES data but now the response to the item 'How afraid are you due to the climate change?'. Table 5 shows the fits of cumulative and adjacent categories models, respectively. Also for this question, the full models can be simplified to location-shift versions of the models although the reduction is not so obvious as in the question that refers to the use of nuclear energy.

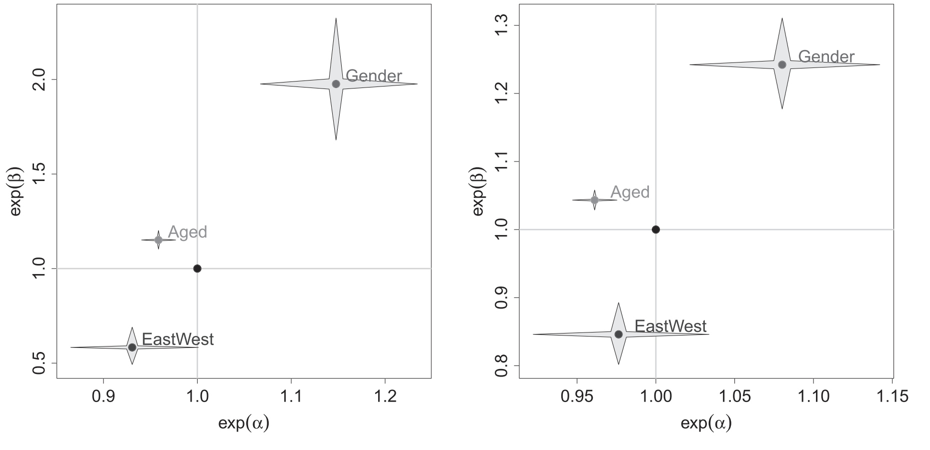


Figure 4. *Plots of* $(e^{\hat{\alpha}}, e^{\hat{\beta}})$ *for response fear of nuclear energy, y-axis represents location, x-axis represents dispersion, left: cumulative location-shift model, right: adjacent categories location-shift model*

Table 5. *Fits of models with logistic link for response fear of climate change*

	Deviance	*df*	*Difference in deviances*	*df*	p *value*
Cumulative models					
Model with category-specific effects	7 152.12	12 192			
Location-shift model	7 170.56	12 204	18.34	12	0.1057
Model with global effects	7 178.06	12 206	7.50	2	0.0235
Adjacent categories models					
Model with category-specific effects	7 153.42	12 192			
Location-shift model	7 171.93	12 204	18.51	12	0.1010
Model with global effects	7 177.17	12 206	5.24	2	0.0723

4.4 Demand for Medical Care

Deb & Trivedi (1997) analysed the demand for medical care for individuals, aged 66 and over, based on a data set from the US National Medical Expenditure survey in 1987/88. The data ('NMES1988') are available from the R package AER (Kleiber & Zeileis, 2008). The response is the number of physician/non-physician office and hospital outpatient visits, which is categorised with categories given by 1: zero, 2: 1–3, 3: 4–6, 4: 7–10, 5: 11–20, 6: above 20. The available covariates include *Age*, the self-perceived health status (*Health*; 0: poor, 1: average, 2: excellent) and the number of chronic conditions (*Numchron*). Because the effects vary across gender, we consider only male, married patients ($n = 1388$). The data set is interesting because it is one of the applications in which cumulative models show fitting problems. The model with category-specific effects can not be fitted at all, the cumulative location-shift model yields unstable estimates and no standard errors are available. In contrast, for the adjacent categories model, maximum likelihood estimates and standard errors are obtained by regular software. The big advantage of the adjacent categories model over the cumulative model that shows here is that parameter values are not restricted in the adjacent categories model. Table 6 shows the fits for the adjacent categories models. It is again seen that one might use the location-shift model but the simple model with global parameters is not appropriate.

5 Generalised Additive Models

In the following traditional additive models for ordinal responses are considered briefly. Then, the additive location-shift model is introduced.

5.1 Generalised Additive Models for Ordinal Responses

Parametric models as the non-proportional odds model are rather restrictive. They assume a simple linear predictor, which might be very misleading if, for example, U-shaped effects are present. A very flexible class of models that avoids these restrictions are generalised additive models, which are well developed for continuous and univariate responses, see, for example, Hastie & Tibshirani (1986), Buja *et al*. (1989) and Friedman & Silverman (1989).

In generalised additive models, the linear predictor $\boldsymbol{x}^T\boldsymbol{\beta}$ is replaced by the additive term

$$f_{(1)}(x_1)+\ldots+f_{(p)}(x_p),$$

where the $f_{(j)}(.)$ are unspecified functions. The unknown functions may be expanded in basis functions (Eilers & Marx, 1996), smoothing splines (Gu, 2002) or thin-plate splines (Wood, 2004); all of them have been used to model binary or continuous responses.

Ordinal models with additive predictors were considered by Yee & Wild (1996) and Yee (2010,2015) within the framework of vector generalised additive models. For ordinal models, one has to replace the whole predictor $\eta_r = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}$ by

Table 6. *Fits of adjacent categories models with logistic link for demand for medical care data*

	Deviance	*df*	*Difference in deviances*	*df*	p *value*
Non-proportional odds model	4 258.41	6 910			
Location-shift model	4 282.05	6 925	23.64	15	0.0714
Proportional odds model	4 303.64	6 930	19.69	5	0.0014

$$\eta_r = \beta_{0r} + f_{(1)}(x_1) + \ldots + f_{(p)}(x_p),$$

which contains a category-specific intercept but fixed smooth variable effects. The essential trait is that it is assumed that the functions $f_{(j)}(x_p)$ do not vary across categories, they are *global* effects. Thus, if one uses the cumulative approach, the models can be considered as additive versions of proportional odds models with accordingly simple interpretation of effects. If the j-th variable increases by one unit from x_j to $x_j + 1$ and all other variables remain fixed, one obtains

$$F^{-1}(P(Y \leq r|x_1, \ldots, x_j + 1, \ldots, x_p)) - F^{-1}(P(Y \leq r|x_1, \ldots, x_j, \ldots, x_p)) = \\ = f_{(j)}(x_j + 1) - f_{(j)}(x_j),$$

which contains only the function $f_{(j)}(.)$. In the logistic version, the inverse distribution function is equivalent to the cumulative log-odds, and one obtains

$$\log\left(\frac{\gamma_r(x_1, \ldots, x_j + 1, \ldots, x_p)}{\gamma_r(x_1, \ldots, x_j, \ldots, x_p)}\right) = f_{(j)}(x_j + 1) - f_{(j)}(x_j),$$

where the $\gamma_r(\boldsymbol{x}) = P(Y \leq r|\boldsymbol{x})/P(Y > r|\boldsymbol{x})$ are the cumulative odds. After transformation, one has

$$\frac{\gamma_r(x_1, \ldots, x_j + 1, \ldots, x_p)}{\gamma_r(x_1, \ldots, x_j, \ldots, x_p)} = e^{f_{(j)}(x_j + 1) - f_{(j)}(x_j)},$$

which can be interpreted as the change in odds if the j-th variable increases by one unit from x_j to $x_j + 1$. If the function is linear, that is, $f_{(j)}(x_j) = x_j\beta_j$, one obtains on the right-hand side e^{β_j}, which is equivalent to Equation (2). Then, the effect strength does not depend on the baseline value x_j. This is different in the general additive case, in which the change depends on the 'starting' value x_j, which is increased by one unit. Nevertheless, the effect strength is not affected by the values of the other covariates. Similar properties hold if one uses the adjacent categories model with additive predictor structure.

5.2 Additive Location-Shift Models

The additive ordinal models with global effects considered in the previous section share some problems with the parametric model with global effects. Although it is more flexible by allowing for smooth effects, it is rather restrictive by assuming that the effects of covariates do not depend on the category. Consequently, it might show bad goodness-of-fit. One can extend the model in the same way as linear models by allowing that the smooth functions are category-specific. Then, one postulates

$$\eta_r = \beta_{0r} + f_{(1),r}(x_1) + \ldots + f_{(p),r}(x_p),$$

where the functions $f_{(j),r}(.)$ depend on r. However, for each covariate, one has to fit $k - 1$ functions, which can lead to a confusing number of functions if one has, for example, 10 response categories, which is not unusual in questionnaires. Moreover, the functions are severely restricted because $\eta_r \leq \eta_{r+1}$ has to hold for all r, which is hard to control in estimation. If it is not accounted for, resulting estimates might yield negative probabilities.

One can try to restrict the variation of the functions by assuming that they are varying not too strongly across categories, see Tutz (2003), but this approach calls for complicated regularisation methods, and still one has $k-1$ functions to interpret for each explanatory variable.

The model proposed here is an additive version of the location-shift model, which avoids the large number of functions but typically fits much better than the simple additive model. The *additive location-shift model* uses the predictor

$$\eta_r = \beta_{0r} + f_{(1)}(x_1) + \ldots + f_{(p)}(x_p) + (r - k/2)\{f_{(1)}^{(S)}(z_1) + \ldots + f_{(m)}^{(S)}(z_m)\},$$

where $f_{(1)}^{(S)}(.), \ldots, f_{(m)}^{(S)}(.)$ are unspecified dispersion functions. The predictor contains two types of smooth functions, the ones in the location term $f_{(1)}(x_1) + \ldots + f_{(p)}(x_p)$ and the ones in the dispersion term $f_{(1)}^{(S)}(z_1) + \ldots + f_{(m)}^{(S)}(z_m)$.

In particular, when $\boldsymbol{x}$ and $\boldsymbol{z}$ are distinct, the functions have a simple interpretation. If the j-th x-variable increases by one unit from x_j to $x_j + 1$ and all other variables remain fixed, one obtains for the cumulative model the same property as in the simple additive model,

$$\frac{\gamma_r(x_1, \ldots, x_j + 1, \ldots, x_p)}{\gamma_r(x_1, \ldots, x_j, \ldots, x_p)} = e^{f_{(j)}(x_j + 1) - f_{(j)}(x_j)},$$

which means that the functions can be interpreted as change in (cumulative) odds ratios. For the differences between adjacent predictors, one obtains

$$\eta_r - \eta_{r-1} = \beta_{0r} - \beta_{0,r-1} + \{f_{(1)}^{(S)}(z_1) + \ldots + f_{(m)}^{(S)}(z_m)\}.$$

That means that large values of $f_{(j)}^{(S)}(.)$ widen the distance between adjacent predictors, while small values shrink the distance between adjacent predictors. Therefore, large values indicate a tendency to middle categories or smaller dispersion while small values indicate a tendency to extreme categories or strong dispersion.

As the parametric model, the additive location-shift model accounts for dispersion without being too complex. In the general case $\boldsymbol{x} = \boldsymbol{z}$, the additive location-shift model contains just two smooth functions per variable that characterise the effect of explanatory variables on the response, one for the location and one for the dispersion. That means one has to fit only $2p$ smooth functions instead of $(k-1)p$, which would be needed in the general model with category-specific covariate functions.

For the fitting of the unknown functions $f_{(j)}(.), f_{(j)}^{(S)}(.)$, we use an expansion in basis functions. Thus, functions are approximated by

$$f_{(j)}(x) = \sum_{s=1}^{M} \beta_{js}\Phi_s(x) \text{ and } f_{(j)}^{(S)}(z) = \sum_{s=1}^{M} \alpha_{js}\Phi_s(z),$$

where $\Phi_1(.), \ldots, \Phi_M(.)$ are basis functions. A widely used set of basis functions are B-splines (Eilers & Marx, 1996), which are also implemented in our R package **ordDisp** to be described in detail in Section 6.

5.3 Simulation Study

We generated data from the additive adjacent categories location-shift model. Again, we considered $k = 5$ response categories and one normally distributed covariate $x \sim N(0, 0.5)$. As in the adjacent categories model in Section 3, we set the category-specific intercepts to $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})^\top = (2, 2/3, -2/3, -2)$. The smooth functions of the location and the dispersion term had the form

$$f_{(x)}(x) = 0.5 \arctan(x) \text{ and } f_{(x)}(z)^S = \arctan(x).$$

The models with global effects and the category-specific models were fitted by applying the function `vgam()` of the R package **VGAM**, and the location-shift models were fitted using **ordDisp** with four cubic B-splines. The squared differences $(\hat{\eta}_{ir} - \eta_{ir})^2$ from fitting the three nested models are depicted in Figure 5. It is seen that the location-shift model on average performs best, whereas the category-specific model performs worse and shows high variability. When testing the hierarchy of models, the location-shift model is chosen in 79% of the cases, while the category-specific model is retained in 16% of the cases. The simple model with global effects is found to be sufficient in only 5% of the cases. The results are therefore comparable with those in Section 3, Scenarios (b) and (c).

5.4 Safety Data

It has been shown in Section 4.1 that the parametric location-shift model provided a good compromise between sparsity and goodness-of-fit for the response feeling safe in Naples. The only continuous variable was age, which had p values 0.086 (cumulative model) and 0.114 (adjacent categories model). The p values are greater than 0.05 but not so far away that one can be sure that there is no effect of age. In the following, age is included as a smooth function approximated by four cubic B-splines in the location term and as a linear function in the dispersion term (a linear function turned out as a good approximation in the dispersion term). Figure 6 shows the resulting curves (left: location effect $f_{(age)}(age)$, right: dispersion effect $f^{(S)}_{(age)}(age)$, upper panel: cumulative model, lower panel: adjacent categories model). It is seen that in both

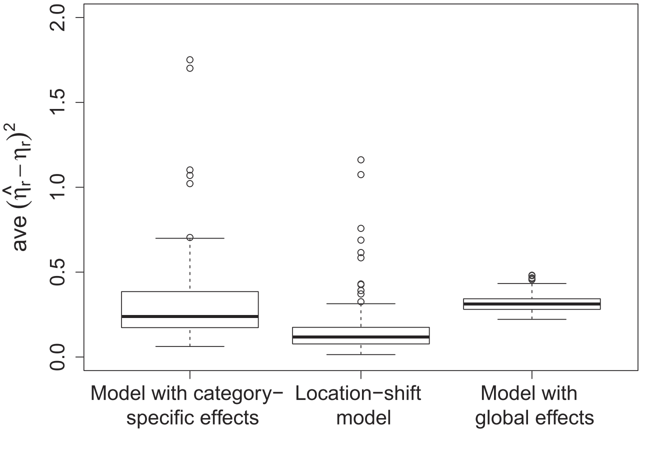


Figure 5. *Results of the simulation study (additive adjacent categories models). The boxplots show the squared differences $(\hat{\eta}_{ir} - \eta_{ir})^2$ averaged over all 500 observations and all four categories, when fitting the model with global effects (right), the location-shift model (middle) and the model with category-specific effects (left)*

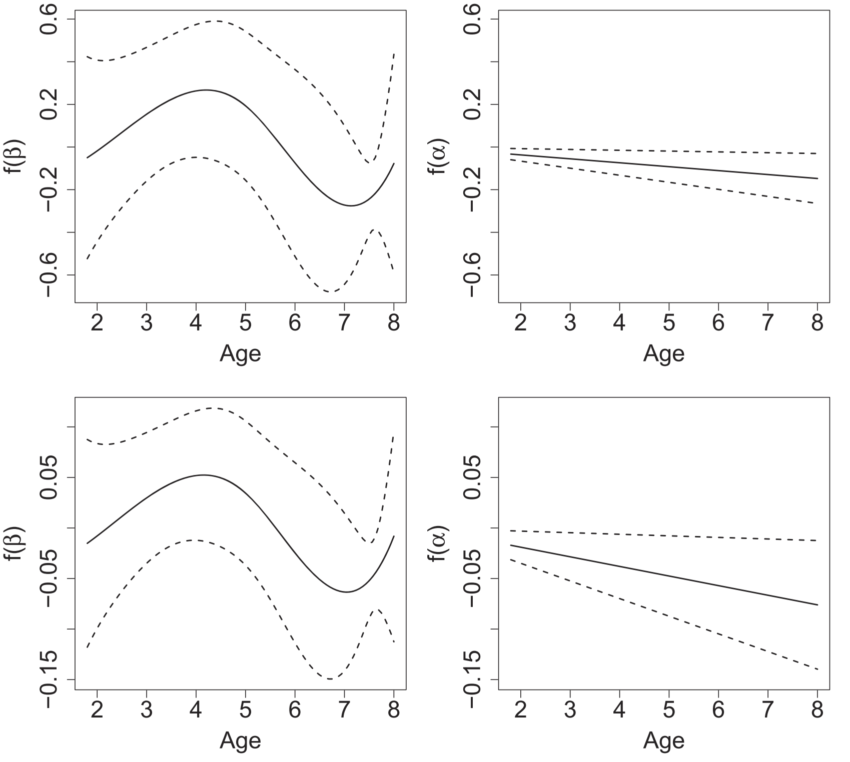


Figure 6. *Safety data (left: location effect $f_{(age)}$(age), right: dispersion effect $f^{(S)}_{(age)}(age)$, upper panel: cumulative model, lower panel: adjacent categories model). Note that age is measured in decades*

models, a linear effect of age seems not appropriate. In particular, young and older persons seem to feel less safe than persons in their forties. Testing if the smooth effect of age is needed yields a *p* value of 0.046 (cumulative model), which indicates that age should not be neglected.

5.5 Nuclear Energy

As a second example, the effect of age on the fear of the use of nuclear energy is considered. Figure 7 shows the estimated location and dispersion effects for the additive cumulative and adjacent categories model. The estimates of the location effect indicate that the fear of the use of nuclear energy is strongest for people in their sixties and weakest for people around thirty. The estimates of the dispersion term indicate that older people tend to have less dispersion than younger respondents. Likelihood ratio tests show that location as well as dispersion effects are not to be neglected. The likelihood ratio test for the location effects is 64.62 on 4 df, and for the dispersion effect, 21.24 on 1 df if the cumulative model is fitted. Similar values result for the adjacent categories model.

6 Programme Packages

Classical cumulative models that do not contain a shifting component can be fitted by using the function `vglm()` of the R package **VGAM** (Yee, 2010). Parametric and additive location-shift models can be fitted by using the R add-on package **ordDisp** (Berger, 2020), which is described in the following.

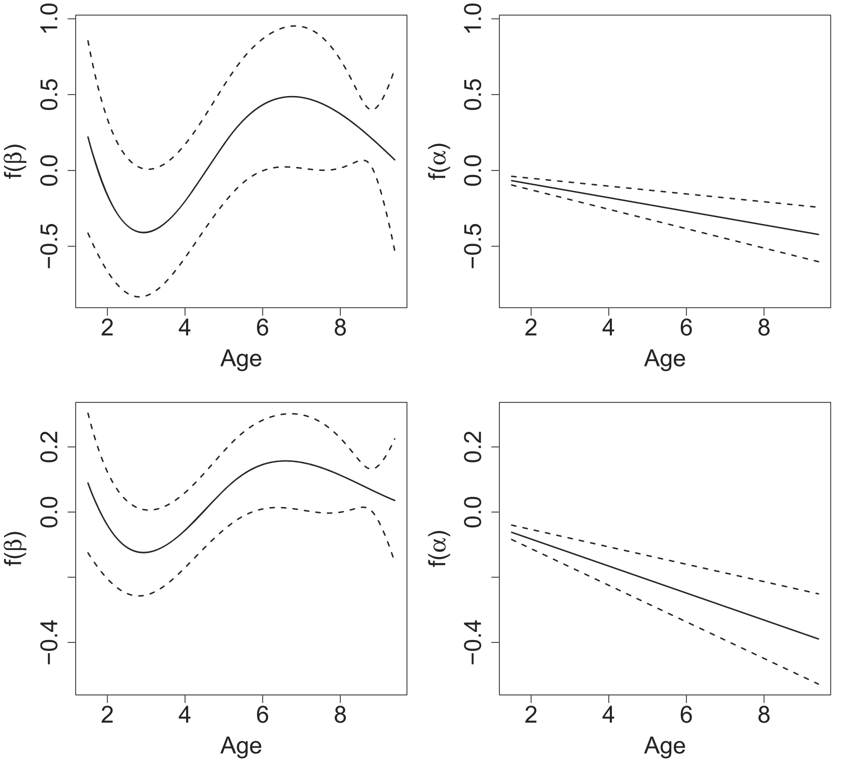


Figure 7. *Nuclear energy data (left: location effect $f_{(age)}(age)$, right: dispersion effect $f^{(S)}_{(age)}(age)$, upper panel: cumulative model, lower panel: adjacent categories model). Note that age is measured in decades*

The call of the main fitting function (in parts) is

ordDisp(formula, data, family = c(``cumulative'', ``acat''), n_bs = 6,
reverse = FALSE, …).

The formula needs to have the form `y` ~ x1 + … + xp | z1 + … + zq, where on the right-hand side of the formula, the *x*-variables of the location term and the *z*-variables of the dispersion term are separated by the `|`-operator. The function allows to fit smooth effects $f(.)$ and $f^{(S)}(.)$ by using `s(x)` and `s(z)` in the respective part of the formula. The functions are then fitted using `n_bs` B-spline basis functions. In the case of nominal covariates, `ordDisp()` generates 0-1-coded dummy variables. If `reverse=TRUE`, the function uses the reverse categories representation $P(Y \geq r|\boldsymbol{x})/P(Y < r|\boldsymbol{x})$ for the cumulative model and $P(Y = r|\boldsymbol{x})/P(Y = r+1|\boldsymbol{x})$ for the adjacent categories model. To keep the interpretation of the dispersion effects, the scaling factor is reversed to $(k/2 - r)$ in the cumulative case and $(r - k/2)$ in the adjacent categories case.

Function `ordDisp()` internally calls function `vglm()` of the R package **VGAM** (Yee, 2010). Thus, the fitted object inherits all the values of a `vglm`-object, and importantly, all the methods implemented for objects of class `vglm`, like `print`, `summary`, `predict` and `plot` can be applied. Additionally, star plots depicting the location effects against the dispersion effects including pointwise 95% confidence intervals (cf. Figure 3) can be generated using the function `plotordDisp(object, names, …)`, where the variables to be plotted are passed to the function by the `names`-argument. Note that the use of `plotordDisp()` is only meaningful for variables with both a location effect and a dispersion effect.

7 Concluding Remarks

The main messages of the paper can be summarises as follows:

- Proportional odds models, or, more general, models with category-specific parameters are widely dispensable. In many applications, a simpler version is appropriate. It has in particular been demonstrated that parametric location-shift models typically are sufficiently complex to approximate the underlying probability structure in ordinal regression.
- Location-shift models, which are propagated here, have the advantage that they show not only location effects but also dispersion effects or tendencies to respond, which are typically present in applications.
- If linear effects are questionable, the smooth location-shift model provides an alternative to simple global effect models. The models allow to account for smooth dispersion effects.

The wide applicability of location-shift models has the additional advantage that variable selection becomes easier. In general category-specific models, there is a hierarchy that describes how single variables can be influential. The effects of a variable can be category-specific, global or zero. Variable selection means one has to determine which variables have which of the three possible influence structures, category-specific, global or zero. In principle, this can be obtained by using tests, although several tests are needed for each variable. More advanced and attractive methods of variable selection that are widely used nowadays are based on penalty methods like the lasso or the elastic net (Tibshirani, 1996; Zou, 2006; Zou & Hastie, 2005). They can not be used directly to the hierarchical selection problem in ordinal models. A hierarchical method that works for ordinal models has been proposed by Pössnecker & Tutz (2016), but rather difficult penalty terms are needed. In location-shift models, variable selection is much easier because effects are separated in a location effect and a dispersion effect. Then, variable selection by tests means one tests if coefficients are zero. Future research on advanced selection methods based on regularisation methods can use the simple structure of the linear predictor with separate coefficients for effects.

Implicitly, we also compared two approaches to modelling data, the cumulative approach and the adjacent categories approach. Typically, the models yield similar goodness-of-fit. A distinct advantage of the adjacent categories model is that no restrictions on the parameter space are postulated, which makes it more adequate when allowing for a more complex predictor structure.

There is a third class of ordinal models, namely, sequential models, which have not been considered here. Parametric sequential models have the form $P(Y \geq r | Y \geq r-1, \boldsymbol{x}) = F(\beta_{0r} + \boldsymbol{x}^T \boldsymbol{\beta}_r)$. They reflect the successive transition to higher categories in a stepwise fashion since $Y \geq r$ given $Y \geq r-1$ can be interpreted as the transition to categories higher than category $r-1$ given at least category $r-1$ has been reached. Sequential models are strongly linked to discrete survival and have been considered, for example, by Armstrong & Sloan (1989), Tutz (1991) and Ananth & Kleinbaum (1997). Location-shift models for this type of model seem less useful because of the structure of the model. In sequential models, the category-specific parameters $\boldsymbol{\beta}_r$ have a distinct meaning, and they represent the impact of covariates on the transition to higher categories given lower categories have already been reached. Including a shift term, which represents a tendency to middle or extreme categories seems less useful.

All three types of models, cumulative, adjacent categories and sequential models have also been used to model marginal responses in contingency tables. For example, Bartolucci *et al*. (2007) extended the results of Colombi & Forcina (2001) and Bergsma & Rudas (2002) and considered general interaction modelling including higher order interactions. If all variables are categorical, these modelling approaches provide alternative parameterisations, which possibly could be simplified by using shifting approaches.

ACKNOWLEDGEMENTS

Open access funding enabled and organized by Projekt DEAL.

References

Ananth, C.V. & Kleinbaum, D.G. (1997). Regression models for ordinal responses: A review of methods and applications. *Int. J. Epidemiol.*, **26**, 1323–1333.

Armstrong, B. & Sloan, M. (1989). Ordinal regression models for epidemiologic data. *Am. J. Epidemiol.*, **129**, 191–204.

Bartolucci, F., Colombi, R. & Forcina, A. (2007). An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Stat. Sin.*, **17**, 691–711.

Bender, R. & Grouven, U. (1998). Using binary logistic regression models for ordinal data with non–proportional odds. *J. Clin. Epidemiol.*, **51**, 809–816.

Berger, M. (2020). orddisp: Separating location and dispersion in ordinal regression models. R package version 2.1.1.

Bergsma, W.P. & Rudas, T. (2002). Marginal models for categorical data. *Ann. Stat.*, **30**(1), 140–159.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, **46**, 1171–1178.

Buja, A., Hastie, T. & Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Stat.*, **17**, 453–510.

Campbell, M.K. & Donner, A.P. (1989). Classification efficiency of multinomial logistic-regression relative to ordinal logistic-regression. *J. Am. Stat. Assoc.*, **84**(406), 587–591.

Colombi, R. & Forcina, A. (2001). Marginal regression models for the analysis of positive association of ordinal response variables. *Biometrika*, **88**(4), 1007–1019.

Cox, C. (1995). Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Stat. Med.*, **14**, 1191–1203.

Deb, P. & Trivedi, P.K. (1997). Demand for medical care by the elderly: A finite mixture approach. *J. Appl. Econ.*, **12**(3), 313–336.

Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and Penalties. *Stat. Sci.*, **11**, 89–121.

Fahrmeir, L. & Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer: New York.

Friedman, J.H. & Silverman, B. (1989). Flexible parsimonious smoothing and additive modelling (with discussion). *Technometrics*, **31**, 3–39.

Genter, F.C. & Farewell, V.T. (1985). Goodness-of-link testing in ordinal regression models. *Can. J. Stat.*, **13**, 37–44.

Gu, C. (2002). *Smoothing Splines ANOVA Models*. Springer–Verlag: New York.

Hastie, T. & Tibshirani, R. (1986). Generalized additive models (c/r: p. 310–318). *Statist. Sci.*, **1**, 297–310.

Iannario, M., Monti, A.C., Piccolo, D., Ronchetti, E. et al. (2017). Robust inference for ordinal response models. *Electron. J. Stat.*, **11**(2), 3407–3445.

Iannario, M., Piccolo, D. & Simone, R. (2015). CUB: A class of mixture models for ordinal data. r package version 1.1.3. http://cran.r-project.org/package=cub

Kim, J.-H. (2003). Assessing practical significance of the proportional odds assumption. *Stat. Probab. Lett.*, **65**(3), 233–239.

Kleiber, C. & Zeileis, A. (2008). *Applied Econometrics with r*. Springer–Verlag: New York.

Liu, I., Mukherjee, B., Suesse, T., Sparrow, D. & Park, S.K. (2009). Graphical diagnostics to check model misspecification for the proportional odds regression model. *Stat. Med.*, **28**(3), 412–429.

McCullagh, P. (1980). Regression model for ordinal data (with discussion). *J. R. Stat. Soc.*, **B 42**, 109–127.

Peterson, B. & Harrell, F.E. (1990). Partial proportional odds models for ordinal response variables. *Appl. Stat.*, **39**, 205–217.

Pössnecker, W. & Tutz, G. (2016). A general framework for the selection of effect type in ordinal regression, Technical Report 186, Department of Statistics LMU.

Rattinger, H., Roßteutscher, S., Schmitt-Beck, R., Weßels, B. & Wolf, C. (2014). Pre-election cross section (GLES 2013). GESIS Data Archive, Cologne ZA5700 Data file Version 2.0.0.

Rudolfer, S.M., Watson, P.C. & Lesaffre, E. (1995). Are ordinal models useful for classification? a revised analysis. *J. Stat. Comput. Simul.*, **52**(2), 105–132.

Snell, E.J. (1964). A scaling procedure for ordered categorical data. *Biometrics*, **20**, 592–607.

Steadman, S. & Weissfeld, L. (1998). A study of the effect of dichotomizing ordinal data upon modelling. *Commun. Stat. Simul. Comput.*, **27(4)**, 871–887.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.*, **B 58**, 267–288.

Tutz, G. (1991). Sequential models in ordinal regression. *Comput. Stat. Data Anal.*, **11**, 275–295.

Tutz, G. (2003). Generalized semiparametrically structured ordinal models. *Biometrics*, **59**, 263–273.

Tutz, G. (2012). *Regression for Categorical Data*. Cambridge University Press.

Tutz, G. & Berger, M. (2016). Response styles in rating scales—Simultaneous modelling of content-related effects and the tendency to middle or extreme categories. *J. Educ. Behav. Stat.*, **41**, 239–268.

Tutz, G. & Berger, M. (2017a). Separating location and dispersion in ordinal regression models. *Econ. Stat.*, **2**, 131–148.

Tutz, G. & Berger, M. (2017b). Separating location and dispersion in ordinal regression models. *Econ. Stat.*, **2**, 131–148.

Walker, R.W. (2016). On generalizing cumulative ordered regression models. *J. Mod. Appl. Stat. Methods*, **15**(2), 28.

Walker, S.H. & Duncan, D.B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, **54**, 167–178.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata J.*, **6**(1), 58.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *J. Math. Sociol.*, **40**(1), 7–20.

Williams, O.D. & Grizzle, J.E. (1972). Analysis of contingency tables having ordered response categories. *J. Am. Stat. Assoc.*, **67**, 55–63.

Wood, S.N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.*, **99**, 673–686.

Yee, T.W. (2010). The VGAM package for categorical data analysis. *J. Stat. Softw.*, **32**(10), 1–34.

Yee, T.W. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in r*. Springer.

Yee, T.W. & Wild, C.J. (1996). Vector generalized additive models. *J. R. Stat. Soc.*, **B**, 481–493.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.*, **101**(476), 1418–1429.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc.*, **B 67**, 301–320.

[*Received July 2020; accepted December 2021*]