INSTITUT FÜR STATISTIK

Gerhard Tutz, Moritz Berger

Separating Location and Dispersion in Ordinal Regression Models

Separating Location and Dispersion in Ordinal Regression Models

Gerhard Tutz & Moritz Berger

Ludwig-Maximilians-Universität München

Akademiestraße 1, D-80799 München

February 22, 2016

Abstract

In ordinal regression the focus is typically on location effects, potential variation in the distribution of the probability mass over response categories referring to stronger or weaker concentration in the middle is mostly ignored. If dispersion effects are present but ignored goodness-of-fit suffers and, more severely, biased estimates of location effects are to be expected since ordinal regression models are non-linear. A model is proposed that explicitly links varying dispersion to explanatory variables. The embedding into the framework of multivariate generalized linear models allows to use computational tools and asymptotic results that have been developed for this class of models. The model is compared to alternative approaches in applications and simulations. In addition, a visualization tool for the combination of location and dispersion effects is proposed and used in applications.

Keywords: Proportional odds model; scale-dispersion model; ordinal response models; dispersion modelling.

1   Introduction

Since the seminal paper of McCullagh (1980) ordinal regression models have been widely applied in various fields of research, see, for example, Liu and Agresti (2005) and Agresti (2010). An important class of ordinal regression models is the class of cumulative models. The most prominent example is the proportional odds model, which will be considered exemplarily in the following before considering general cumulative models.

Let $Y \in \{1, \dots, k\}$ denote the response and $\boldsymbol{x}$ a vector of explanatory variables. Then the basic form of the proportional odds model is given by

$$\log \frac{P(Y \leq r | \boldsymbol{x})}{P(Y > r | \boldsymbol{x})} = \theta_r + \boldsymbol{x}^T \boldsymbol{\beta}, \quad r = 1, \dots, k - 1, \tag{1}$$

where $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$. An attractive feature of the model is the simple interpretation of parameters, which results from the proportional odds property. This property is seen from considering two sets of explanatory variables $\boldsymbol{x}, \tilde{\boldsymbol{x}}$ and the

corresponding cumulative odds $\gamma(r|\boldsymbol{x}) = P(Y \le r|\boldsymbol{x})/P(Y > r|\boldsymbol{x})$ and $\gamma(r|\tilde{\boldsymbol{x}}) = P(Y \le r|\tilde{\boldsymbol{x}})/P(Y > r|\tilde{\boldsymbol{x}})$. Simple derivation shows that the proportion of the cumulative odds for the two sets of variables is given by

$$\frac{\gamma(r|\boldsymbol{x})}{\gamma(r|\tilde{\boldsymbol{x}})} = \exp((\boldsymbol{x} - \tilde{\boldsymbol{x}})^T \boldsymbol{\beta}),$$

and therefore does not depend on the category r. Consequently, the interpretation of parameters does not depend on the category. More concise, $\exp(\beta_j)$ represents the factor by which all the cumulative odds $P(Y \le r|\boldsymbol{x})/P(Y > r|\boldsymbol{x})$ change if variable x_j increases by one unit.

The simple interpretation gets lost in an extended version of the model in which parameters are category-specific. That means the predictor $\eta_r = \theta_r + \boldsymbol{x}^T \boldsymbol{\beta}$ in model (1) is replaced by $\eta_r = \theta_r + \boldsymbol{x}^T \boldsymbol{\beta}_r$. The corresponding *partial proportional odds model* frequently shows a better fit to the data but interpretation of parameters is more difficult. Moreover, severe restrictions are postulated. While the simple proportional odds model only postulates the ordering of the intercepts $\theta_1 \le \cdots \le \theta_{k-1}$ the extended version postulates $\theta_1 + \boldsymbol{x}^T \boldsymbol{\beta}_1 \le \cdots \le \theta_{k-1} + \boldsymbol{x}^T \boldsymbol{\beta}_{k-1}$ for all values $\boldsymbol{x}$, which can severely restrict the possible values of explanatory variables. Therefore, often simple Fisher scoring does not work and estimation of parameters fails. For special link functions the cumulative model is equivalent to the sequential model, which allows to avoid the ordering of thresholds, see Tutz (1991) and, more recently Peyhardi et al. (2015). The class of partial proportional odds models has been investigated in particular by Brant (1990), Peterson and Harrell (1990) and Bender and Grouven (1998), graphical checks were proposed by Kim (2003) and Liu et al. (2009).

Despite its disadvantages the partial proportional odds model is often used if the fit of the proportional odds model is unsatisfactory. However, the lack-of-fit can also be caused by an insufficient modelling of dispersion effects. In the present paper the focus is on the modelling of varying dispersion in ordinal regression.

For illustration let us consider a simple example that has already been used by McCullagh (1980). Table 1 shows Stuart's (1953) quality of right eye vision data for men and women. From the data it is obvious that women are more concentrated in the middle categories while men have relatively high proportions in the extreme categories. By construction the proportional odds model and other cumulative models without dispersion effects are not able to capture the different variability of subpopulations.

TABLE 1: *Quality of right eye vision in men and women.*

	Highest (1)	Vision Quality 2	3	Lowest (4)
Men	1053	782	893	514
Women	1976	2256	2456	789

Ignoring dispersion effects is less severe in linear models. Varying dispersion, which for linear models is called heteroscedasticity, affects the precision of least squares estimates but they are still unbiased. However, ordinal regression models are non-linear models. For this class of models biased estimates are to be expected if dispersion is not modelled. In general, the modelling of variability is much harder than the modelling

of the mean of the response. For ordinal responses an additional difficulty is that one cannot use the variance of a univariate response because the response is multinomial and therefore multivariate. Although categories are ordered treating it as an univariate response would mean to ignore the scale level.

Here a model is proposed that models dispersion by including special effects in the linear predictor, which yields a model that can be estimated within the generalized linear model framework. In Section 2 the model is introduced and an illustrative application is given. In Section 3 the model is compared to the location-scale model and consequences of ignored dispersion effects are briefly considered. After the consideration of non-symmetric responses in Section 4, in Section 5 alternative strategies to model ordinal response data by including category-specific effects or dispersion effects are discussed.

2 Separating Location and Dispersion

In this section we briefly show how cumulative ordinal models, which include the proportional odds model, and the extended location-scale model can be motivated from an underlying metric response model. Then we consider the model with shifted thresholds, which handles dispersion in a quite different way.

2.1 Cumulative Type Models for Ordinal Responses

Cumulative type models like the proportional odds model can be motivated by latent variables. The basic assumption is that the observed categories represent a coarser (categorical) version of an underlying (continuous) regression model. Let $\tilde{Y}$ be an underlying latent variable that follows a regression model:

$$\tilde{Y} = -\boldsymbol{x}^T\boldsymbol{\beta} + \epsilon,$$

where ϵ is a noise variable with continuous distribution function F. Furthermore, let the link between the observable categories and the latent variable be given by

$$Y = r \quad \Leftrightarrow \quad \theta_{r-1} < \tilde{Y} \le \theta_r,$$

where $-\infty = \theta_0 < \theta_1 < \cdots < \theta_k = \infty$ are thresholds on the latent scale. One obtains immediately

$$P(Y \le r|\boldsymbol{x}) = P(-\boldsymbol{x}^T\boldsymbol{\beta} + \epsilon \le \theta_r) = P(\epsilon \le \theta_r + \boldsymbol{x}^T\boldsymbol{\beta}) = F(\theta_r + \boldsymbol{x}^T\boldsymbol{\beta}).$$

The model is essentially a univariate response model since it is assumed that a univariate response $\tilde{Y}$ is in the background. The response Y is just a coarser version of $\tilde{Y}$ where the thresholds θ_r determine the preference for categories and the covariates produce a shifting on the latent scale. If $F(\cdot)$ is chosen as the logistic distribution function one obtains the proportional odds model (1).

A model that accounts for additional dispersion is obtained by assuming for the latent variable $\tilde{Y} = -\boldsymbol{x}^T\boldsymbol{\beta} + \tau_{\boldsymbol{x}}\epsilon$, where $\tau_{\boldsymbol{x}}$ is the variance of the underlying regression model, which may depend on $\boldsymbol{x}$. The corresponding cumulative model with dispersion, also called *location-scale model*, is given by

$$P(Y \le r|\boldsymbol{x}) = F\left(\frac{\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}}{\tau_{\boldsymbol{x}}}\right), \quad r = 1, \ldots, k-1, \tag{2}$$

see McCullagh (1980). In cases where the concentration in response categories varies across populations, the model is more appropriate than the simple cumulative model. The simple cumulative model is based on the underlying continuous regression model $\tilde{Y} = -\boldsymbol{x}^T\boldsymbol{\beta} + \varepsilon$, where the distribution of ε does not depend on $\boldsymbol{x}$. Thus the model assumes that with varying $\boldsymbol{x}$ the probability mass is merely shifted on the latent scale, therefore $\boldsymbol{x}^T\boldsymbol{\beta}$ is often called the location effect. If the probability mass is more concentrated in one population and spread out in other populations, the simple cumulative model is unable to model the varying dispersion. The inclusion of a variance that varies over populations can capture this effect. Since the model includes a shifting and a dispersion or scaling effect it is often called a location-scale model.

One has to find appropriate ways to link the dispersion parameter to covariates. For example, one can use $\tau_{\boldsymbol{x}} = \exp(\boldsymbol{x}^T\boldsymbol{\gamma})$, which makes $\tau_{\boldsymbol{x}}$ positive. However, the model is highly non-linear and one is no longer within the framework of (multivariate) generalized linear models. Special software is needed to fit the model. For example, Cox (1995) used non-linear regression programs available in SAS. For further investigation of the model see also Nair (1987) and Hamada and Wu (1996).

2.2 Modeling Dispersion by Shifted Thresholds

In the following an alternative way to account for varying dispersion is proposed. Let us consider first the case with an even number of response categories k; then $m = [k/2]$ splits the response categories into equally sized sets $\{1, \ldots, m\}$ and $\{m + 1, \ldots, k\}$. Moreover, it is assumed that the ordered categories refer to a symmetric response, for example by categories of agreement as *strongly disagree, moderately disagree,..., moderately agree, strongly agree*. Let $\boldsymbol{z}$ be an additional vector of explanatory variables, which can be identical to $\boldsymbol{x}$ but does not have to.

Let now the thresholds in the proportional odds model be determined by

$$\theta_r = \beta_{0r} - \boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = 1, \ldots, m - 1$$
$$\theta_m = \beta_{0m}$$
$$\theta_r = \beta_{0r} + \boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = m + 1, \ldots, k - 1.$$

That means the center threshold θ_m remains fixed, but lower and upper thresholds are shifted by $\delta = \boldsymbol{z}^T\boldsymbol{\alpha}$. If δ is positive the intervals defined by thresholds are widened, indicating weaker dispersion, if δ is negative the intervals are shrunk, indicating stronger dispersion. With $\pi(r) = P(Y \le r|\boldsymbol{x}, \boldsymbol{z})$ the model has the form

$$\pi(r) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} - \boldsymbol{z}^T\boldsymbol{\alpha}), \quad r = 1, \ldots, m - 1$$
$$\pi(m) = F(\beta_{0m} + \boldsymbol{x}^T\boldsymbol{\beta})$$
$$\pi(r) = F(\beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta} + \boldsymbol{z}^T\boldsymbol{\alpha}), \quad r = m + 1, \ldots, k - 1,$$

Since it is composed of a location component and a shifting of thresholds it is called the *location-shift model*. It is easily derived that $P(Y = m|\boldsymbol{x}, \boldsymbol{z}) + P(Y = m + 1|\boldsymbol{x}, \boldsymbol{z}) = F(\beta_{0,m+1} + \boldsymbol{x}^T\boldsymbol{\beta} + \delta) - F(\beta_{0,m-1} + \boldsymbol{x}^T\boldsymbol{\beta} - \delta)$. Therefore if $\delta \to \infty$ one obtains $P(Y = m|\boldsymbol{x}, \boldsymbol{z}) + P(Y = m + 1|\boldsymbol{x}, \boldsymbol{z}) \to 1$, which means a tendency toward the middle categories and therefore weak dispersion.

The effect of the additional term $\delta = \boldsymbol{z}^T\boldsymbol{\alpha}$ is illustrated in Figure 1 for a response with $k = 8$ categories and a binary covariate $x \in \{-1, 1\}$ with $\beta = 1$. We set $x = z$ and chose $\theta_1 = -3, \theta_2 = -2, \ldots, \theta_6 = 2, \theta_7 = 3$. Figure 1 shows the distribution

4

FIGURE 1: *Probability distribution of a response with eight categories for several values of α.*

of probabilities without dispersion ($\alpha = 0$) and with dispersion effects $\alpha = 0.4$ and $\alpha = -0.4$. It is seen that for $\alpha = 0.4$ the distribution is more concentrated in the middle if $x = 1$ and stronger dispersed if $x = -1$ when compared to the baseline distribution (first row). For $\alpha = -0.4$ one sees the reverse effect, stronger dispersion if $x = 1$ and more concentration in the middle if $x = -1$.

Effects and Interpretation of Parameters

Let first $\boldsymbol{x}$ and $\boldsymbol{z}$ be distinct. It is easily derived that then the proportional odds assumption still holds for the $\boldsymbol{x}$-variables With $\gamma(r|\boldsymbol{x}, \boldsymbol{z}) = P(Y \leq r|\boldsymbol{x}, \boldsymbol{z})/P(Y > r|\boldsymbol{x}, \boldsymbol{z})$ denoting the cumulative odds for category r one obtains for two sets of explanatory variables $\boldsymbol{x}, \tilde{\boldsymbol{x}}$

$$\log \frac{\gamma(r|\boldsymbol{x}, \boldsymbol{z})}{\gamma(r|\tilde{\boldsymbol{x}}, \boldsymbol{z})} = (\boldsymbol{x} - \tilde{\boldsymbol{x}})^T \boldsymbol{\beta}$$

Therefore the proportion of cumulative odds $\gamma(r|\boldsymbol{x}, \boldsymbol{z})$ and $\gamma(r|\tilde{\boldsymbol{x}}, \boldsymbol{z})$ are the same for all categories r. A consequence is that the parameter β_j from the vector $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_p)$ is given by

$$e^{\beta_j} = \frac{\gamma(r|(x_1, \ldots, x_j + 1, \ldots, x_p), \boldsymbol{z})}{\gamma(r|(x_1, \ldots, x_j, \ldots, x_p), \boldsymbol{z})} \tag{3}$$

That means, if x_j increases by one unit the cumulative odds for each category change by the factor e^{β_j}. For $e^{\beta_j} > 1$ the increase of variable x_j favors low response categories. Thus the main advantage of the proportional odds model, namely simple interpretation of parameters, is kept.

For the $\boldsymbol{z}$-variables the interpretation is different. One obtains for two sets of explanatory variables $\boldsymbol{z}, \tilde{\boldsymbol{z}}$

$$\log \frac{\gamma(r|\boldsymbol{x}, \boldsymbol{z})}{\gamma(\boldsymbol{x}, \tilde{\boldsymbol{z}})} = \begin{cases} -(\boldsymbol{z} - \tilde{\boldsymbol{z}})^T \boldsymbol{\alpha}, & r \in \{1, \ldots, m-1\} \\ (\boldsymbol{z} - \tilde{\boldsymbol{z}})^T \boldsymbol{\alpha}, & r \in \{m+1, \ldots, k-1\}. \end{cases}$$

Thus for α_j from the vector $\boldsymbol{\alpha}^T = (\alpha_1, \ldots, \alpha_q)$ one obtains

$$e^{-\alpha_j} = \frac{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j + 1, \ldots, z_q))}{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j, \ldots, z_q))}, \quad r \in \{1, \ldots, m-1\}$$

$$e^{\alpha_j} = \frac{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j + 1, \ldots, z_q)}{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j, \ldots, z_q))}, \quad r \in \{m+1, \ldots, k-1\}$$

That means, if z_j increases by one unit the cumulative odds for categories $r < m$ change by the factor $e^{-\alpha_j}$ and for categories $r > m$ by the factor e^{α_j}. For $\alpha_j > 0$ the increase of variable z_j decreases the cumulative odds for categories $r < m$ and increases the cumulative odds for categories $r > m$, which means that the response probabilities for extreme categories get smaller. The effect is not a shifting of the probability mass of the response but a stronger concentration in the middle.

If $\boldsymbol{x} = \boldsymbol{z}$ the interpretation of parameters is similar. For simplicity we consider an one dimensional x. It is immediately seen that

$$e^\beta = \frac{\gamma(m|x+1)}{\gamma(m|x)}.$$

Thus e^β represents the odds ratio for categories smaller or equal m if x increases by one unit. It corresponds to the parameter in a binary logit model that distinguishes between categories $\{1, \ldots, m\}$ and $\{m+1, \ldots, k\}$. For the other cumulative odds one obtains

$$\frac{\gamma(r|x+1)}{\gamma(r|x)} = \begin{cases} e^\beta e^{-\alpha}, & r \in \{1, \ldots, m-1\} \\ e^\beta e^\alpha, & r \in \{m+1, \ldots, k-1\}. \end{cases} \tag{4}$$

Thus $e^{-\alpha}$ and e^α modify as factors the basic preference for categories from $\{1, \ldots, m\}$ or $\{m+1, \ldots, k\}$. For symmetric categories as considered here one obtains a more intuitive form by using for large categories the complementary odds defined by $\tilde{\gamma}(r|x) = P(Y \geq r|x)/P(Y < r|x)$, which give the odds for categories larger or equal r. They are linked to the usual cumulative odds by $\tilde{\gamma}(r|x)^{-1} = \gamma(r-1|x)$. One obtains for categories $r \in \{m+1, \ldots, k\}$

$$\frac{\tilde{\gamma}(r|x+1)}{\tilde{\gamma}(r|x)} = e^{-\beta} e^{-\alpha}, \quad r \in \{m+1, \ldots, k-1\}.$$

Thus the scaling factor that modifies the basic preference is again $e^{-\alpha}$. If one considers, for example, only the extreme categories one has

$$\frac{\gamma(1|x+1)}{\gamma(1|x)} = e^\beta e^{-\alpha} \quad \text{and} \quad \frac{\tilde{\gamma}(k|x+1)}{\tilde{\gamma}(k|x)} = e^{-\beta} e^{-\alpha}.$$

Thus the modification of the odds for category 1 as compared to all other categories and the odds for category k as compared to all other categories (the complementary

cumulative odds) are both modified by the factor $e^{-\alpha}$, which means for $\alpha > 0$ that both are shrunk by the factor $e^{-\alpha}$.

The parameter α itself is given by

$$e^{-2\alpha} = \frac{\gamma(s|x+1)/\gamma(s|x)}{\gamma(r|x+1)/\gamma(r|x)} = \frac{\gamma(s|x+1)\tilde{\gamma}(r|x+1)}{\gamma(s|x)\tilde{\gamma}(r|x)}$$

for any $s < m, r > m$. The product $\gamma(s|x)\tilde{\gamma}(r|x)$ is a measure for the concentration of the probabilities in extreme categories. It is large if the probabilities of extreme categories are large. Therefore, $e^{-2\alpha}$ represents the change of the concentration in extreme categories if x increases by one unit.

Eye Vision Example

Let us consider the simple quality of eye vision example from Table 1. The fitted values of the simple proportional odds model and for the location-shift model with dispersion effect are shown in Table 2. It is seen that in both models the location effect ($\hat{\beta} = -0.038$ and $\hat{\beta} = 0.042$) is rather weak and not significant at the 0.05 level. In contrast the dispersion parameter in the model with dispersion $\hat{\alpha} = 0.353$ can definitely not be neglected. The deviance of the proportional odds model is 128.39 on 2 df but reduces to 5.896 on 1 df for the model with location and dispersion effect. The estimated shrinkage factor is $e^{-\hat{\alpha}} = 0.70$, which means that for females the odds for the extreme categories 1 and 4 are shrunk by the factor 0.70 when compared to males.

TABLE 2: *Parameter estimates, standard errors and z-values for the eye vision data.*

Covariate	Proportional Odds Model			Location-Shift Model		
	estimate	std error	z value	estimate	std error	z value
Intercept1	-0.905	0.034	-26.613	-0.721	0.037	-19.397
Intercept2	0.293	0.033	8.911	0.236	0.033	7.104
Intercept3	2.005	0.039	50.398	1.710	0.045	37.563
gender-location	-0.038	0.038	-1.003	0.042	0.038	1.109
gender-dispersion				0.353	0.031	11.348

Model for Odd Number of Response Categories

Let now categories refer to a symmetric response with categories of agreement as *strongly disagree, moderately disagree,..., moderately agree, strongly agree* but with a neutral category in the middle. Then the number of categories k is an odd number. The model with a dispersion component has the same basic structure but now one parameterizes for $m = [k/2] + 1$, which denotes the middle category,

$$\theta_r = \beta_{0r} - \boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = 1, \dots, m-1$$
$$\theta_r = \beta_{0r} + \boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = m, \dots, k-1.$$

The interpretation is similar as in the case with an even number of response categories. For e^{β_j} one obtains the same interpretation, that is, (3) is still the same. Also for the

scaling parameters one obtains the same values , but they hold for different response categories. One obtains

$$e^{-\alpha_j} = \frac{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j + 1, \ldots, z_q))}{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j, \ldots, z_q))}, \quad r \in \{1, \ldots, m - 1\}$$

$$e^{\alpha_j} = \frac{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j + 1, \ldots, z_q))}{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j, \ldots, z_q))}, \quad r \in \{m, \ldots, k - 1\}$$

The same holds for (4), which is still valid but for accordingly modified categories.

2.3 Shifting of Thresholds with Scaling

In the models considered in the previous sections the thresholds have been shifted away from the middle by the value $\delta = \boldsymbol{z}_i^T \boldsymbol{\alpha}$. The effect is a widening of the middle category if k is odd and of the two categories in the middle if k is even. However, the other categories have not been widened. Alternatively one can understand dispersion as a widening of all the categories by using scale values for the widening of the intervals between two thresholds. Let us consider again the case k even and $m = [k/2]$.

Let the thresholds be determined more generally by

$$\theta_r = \beta_{0r} - s_r \boldsymbol{z}^T \boldsymbol{\alpha}, \quad r = 1, \ldots, m - 1$$
$$\theta_m = \beta_{0m}$$
$$\theta_r = \beta_{0r} + s_r \boldsymbol{z}^T \boldsymbol{\alpha}, \quad r = m + 1, \ldots, k - 1,$$

where s_r are scale values that reflect the distance between categories r and m. A simple choice is $s_1 = \ldots = s_{k-1} = 1$, which yields the model used in the previous section.

A particularly attractive choice of scales is obtained by shifting of the thresholds proportional to the distance from the middle threshold. Then one uses $s_r = m - r$ for $r = 1, \ldots, m$ and $s_r = r - m$ for $r = m + 1, \ldots, k - 1$ to obtain the model

$$\begin{aligned} \pi(r) &= F(\beta_{0r} + \boldsymbol{x}^T \boldsymbol{\beta} - (m - r) \boldsymbol{z}^T \boldsymbol{\alpha}), \quad r = 1, \ldots, m \\ \pi(r) &= F(\beta_{0r} + \boldsymbol{x}^T \boldsymbol{\beta} + (r - m) \boldsymbol{z}^T \boldsymbol{\alpha}), \quad r = m + 1, \ldots, k - 1 \end{aligned} \tag{5}$$

The effect is that the intervals between all thresholds are widened by the value $\delta = \boldsymbol{z}_i^T \boldsymbol{\alpha}$. In the case of four response categories the model with scaling is equivalent to the basic model without scaling. However, for more than four categories the models differ. We will refer to the model (5) as the model with scaling.

The interpretation of parameters is similar to the interpretation of parameters in the basic model. If $\boldsymbol{x}$ and $\boldsymbol{z}$ are distinct (3) still holds, which means, if x_j increases by one unit the cumulative odds for each category change by the factor e^{β_j}. For the $\boldsymbol{\alpha}$-parameters one obtains

$$e^{-(m-r)\alpha_j} = \frac{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j + 1, \ldots, z_q))}{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j, \ldots, z_q))}, \quad r \in \{1, \ldots, m\}$$

$$e^{(r-m)\alpha_j} = \frac{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j + 1, \ldots, z_q))}{\gamma(r|\boldsymbol{x}, (z_1, \ldots, z_j, \ldots, z_q))}, \quad r \in \{m + 1, \ldots, k - 1\}$$

For adjacent categories holds

$$e^{\alpha_j} = \frac{\gamma(r|\boldsymbol{x},(z_1,\ldots,z_j+1,\ldots,z_q))/\gamma(r|\boldsymbol{x},(z_1,\ldots,z_j,\ldots,z_q))}{\gamma(r-1|\boldsymbol{x},(z_1,\ldots,z_j+1,\ldots,z_q))/\gamma(r-1|\boldsymbol{x},(z_1,\ldots,z_j,\ldots,z_q))}$$

In the case $x = z$, one obtains now

$$\frac{\gamma(r|x+1)}{\gamma(r|x)} = \begin{cases} e^{\beta}e^{-(m-r)\alpha}, & r \in \{1,\ldots,m\} \\ e^{\beta}e^{(r-m)\alpha}, & r \in \{m+1,\ldots,k-1\}, \end{cases}$$

in particular for middle category m one obtains again

$$e^{\beta} = \frac{\gamma(m|x+1)}{\gamma(m|x)}.$$

For positive α the value $e^{-(m-r)\alpha}$ is smaller than 1, which means it is a shrinkage factor for categories $r < m$. The value $e^{(r-m)\alpha}$ is greater than 1 and therefore increases the odds ratios for large r.

For the case k odd widening of the intervals between thresholds by a fixed value is more difficult. Let again $m = [k/2]+1$ denote the the middle category. The widening of the intervals by the value $\delta = \boldsymbol{z}_i^T\boldsymbol{\alpha}$ is obtained by

$$\theta_r = \beta_{0r} - [(m-r-1)+1/2]\boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = 1,\ldots,m-1$$
$$\theta_r = \beta_{0r} + [(r-m)+1/2]\boldsymbol{z}^T\boldsymbol{\alpha}, \quad r = m,\ldots,k-1$$

Again, for $\boldsymbol{x}$ and $\boldsymbol{z}$ distinct (3) holds and the interpretation of the β parameters are the same.

2.4 Inference and Computation of Estimates

The strength of the proposed modelling of dispersion effects is that the resulting models can be embedded within the framework of multivariate generalized linear models (GLMs). That means they have the form

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

where $\boldsymbol{\pi}^T = (\pi_1,\ldots,\pi_q)$ is the vector the of response probabilities with components $\pi_r = P(Y = r|\boldsymbol{x})$, $\boldsymbol{X}$ is a design matrix constructed from the predictors $\boldsymbol{x}$ and $\boldsymbol{z}$, $\boldsymbol{\beta}$ 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 $\boldsymbol{X}\boldsymbol{\beta}$ are the linear predictors $(\eta_1,\ldots,\eta_q)$, where $q = k-1$. For details of the representation as a multivariate GLMs see Fahrmeir and Tutz (2001), Tutz (2012). Thus the whole machinery of multivariate GLMs, including algorithms, can be used to obtain estimates and standard errors. Also testing of effects, analysis of residuals and goodness-of-fit tests developed for GLMs can be used.

Estimates can be obtained by using the R-package `VGAM` (Yee, 2010; Yee, 2014). The function `vglm()` allows to estimate various multivariate GLMs (Yee and Wild, 1996). By appropriate specification of the design matrix the proposed location-shift model with dispersion effects can be fitted by using `vglm()`. A proportional odds model as

considered here can by specified by the family function `cumulative(reverse=FALSE, parallel=FALSE~1)`, where the second argument ensures that only the thresholds are category-specific. In the location-shift model the z-variables can be seen as a special case of category-specific covariates that differ according to a constant factor (depending on the number of response categories and the type of shifting). For the specification of category-specific covariates argument `xij` can be used when calling `vglm()`. Estimates can easily be obtained after building the design matrix that includes the z-variables in the specific form.

TABLE 3: *Parameter estimates, standard errors and z-values for the government data.*

	Covariate	Proportional Odds Model			Location-Shift Model		
		estimate	std error	z value	estimate	std error	z value
location effects	Gender	-0.157	0.068	-2.303	-0.138	0.069	-1.991
	Income	-0.076	0.021	-3.510	-0.076	0.025	-3.040
	Age	0.076	0.019	4.009	0.079	0.019	4.127
	Age^2	-0.079	0.010	-7.680	-0.079	0.010	-7.703
	Little	-0.874	0.124	-7.071	-0.693	0.130	-5.311
	Medium	-1.129	0.114	-9.889	-0.960	0.121	-7.939
	Strong	-1.267	0.129	-9.843	-1.098	0.135	-8.143
	Very Strong	-0.892	0.148	-6.030	-0.745	0.154	-4.842
dispersion effects	Gender				0.189	0.040	4.699
	Income				0.013	0.013	1.024
	Age				-0.020	0.011	-1.776
	Age^2				0.005	0.006	0.872
	Little				0.265	0.067	3.948
	Medium				0.255	0.060	4.226
	Strong				0.321	0.071	4.488
	Very Strong				0.075	0.074	1.007

2.5 Application: Confidence in the Federal Government

We consider data from the general social survey of social science, in short ALLBUS, a study by the German institute GESIS. The data is available from `http://www.gesis.org/allbus`. Our analysis is based on a subset containing 2935 respondents of the ALLBUS in 2012. The response is the confidence in the federal government measured on a symmetric scale from 1 (no confidence at all/excessive distrust) to 7 (excessive confidence). As explanatory variables we consider the gender (0: male, 1: female), the income in thousands of Euros, the age in decades (centered at 50) with a linear and a quadratic term and the self reported interest in politics from 1 (very strong interest) to 5 (no interest at all). For modelling we chose category "no" as reference. The deviance of the location-shift model (without scaling) is $10,179.51$. For the model with scaled shifting of thresholds one obtains a remarkably smaller value of $10,140.91$. Hence we will present results for the model with scaling. The likelihood ratio test statistic for the null hypotheses $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ is 54.5 on 8 degrees of freedom and therefore dispersion should definitely be taken into account.

FIGURE 2: *Visualization of estimated effects for the government data including point-wise confidence intervals.*

FIGURE 3: *Non-linear location (left) and dispersion (right) effects for covariate age for the government data.*

The estimated coefficients and corresponding standard errors of the simple proportional odds model and the location-shift model with scaling are given in Table 3. It is seen that for both models the location effects of all four covariates should be included in the model. The location-shift model typically yields estimates that are closer to zero. Among the dispersion effects only the variables gender and political interest obtain large z-values and seem to be needed in the model.

To simplify the interpretation of effects, Figure 2 shows the tupel $(e^{\hat{\alpha}}, e^{\hat{\beta}})$ for the linear effects of the model with dispersion. The first value, $e^{\hat{\alpha}}$, represents the multiplicative dispersion effect on the odds, for values larger than one one has larger dispersion, for values smaller than one one has smaller dispersion than in the simple proportional odds model. The second value, $e^{\hat{\beta}}$, represents the multiplicative location effect on the odds obtained by the shifting of the underlying continuous response model. For values larger than one small response categories are favored, for values smaller than one large response categories are favored. The coefficients for gender and income are shown in the left panel, the coefficients for political interest are visualized in the right panel. In Figure 2 we also included pointwise 95% confidence intervals that are represented by stars where the horizontal and vertical lengths corresponds to the confidence intervals of $e^{\hat{\alpha}}$ and $e^{\hat{\beta}}$, respectively. From the left panel it can be seen

11

that females tend to choose higher response categories and therefore show a higher confidence in the government than males. At the same time they show smaller dispersion than males, responses are more concentrated in the middle. The confidence also increases with increasing income. However, the dispersion effect is very close to one and can be neglected. The right panel shows that the confidence is higher among all respondents that had at least some political interest. Furthermore, respondents that did not choose one of the extremes ("no" or "very strong interest") show reduced dispersion. This could be interpreted as a response style, people who tend to choose middle categories have the same tendency in all questions.

For non-linear effects as the effect of age, star plots as in Figure 2 are not useful. Therefore, in Figure 3 the fitted non-linear location (left) and dispersion (right) effects of the variable age, denoted by $f_{loc}(age)$ and $f_{dis}(age)$, are given as a function of age. The dispersion effect is not significant, nevertheless, for illustration we show the corresponding curve. The location curve in the left panel shows that confidence is weakest at about 55 years of age but is definitely stronger for younger and older persons.

3   Comparison of Models and and Consequences of Ignored Dispersion

Varying dispersion can be modelled by the proposed location-shift model but also by the location-scale model (2). In the location-shift model the dispersion is modelled by an explicit shifting of the thresholds which is determined by the parameter $\boldsymbol{\alpha}$. In the location-scale model the dispersion is generated by the variance $\tau_{\boldsymbol{x}} = \exp(\boldsymbol{x}^T\boldsymbol{\gamma})$ of the underlying continuous regression model. The effect, determined by the parameter $\boldsymbol{\gamma}$, is now multiplicative on the thresholds since the predictor has the form $\eta_r = \gamma_{0r}\exp(-\boldsymbol{x}^T\boldsymbol{\gamma}) + \boldsymbol{x}^T\boldsymbol{\beta}\exp(-\boldsymbol{x}^T\boldsymbol{\gamma})$. Moreover, the dispersion also modifies the location term. Although the models are not equivalent in applications we found the differences in terms of goodness-of-fit can be rather small.

For illustration we first consider the eye vision data example. We draw sub samples of size $n = 200$ from the data set and computed the location effect, the dispersion effect and the deviances of the location-scale and the location-shift model. As it is seen from Figure 4 the estimates and deviances of the two models show strong correlation. In particular the deviances of the two models are very close. Therefore, in cases with almost no location effect the models yield similar estimates and goodness-of-fit measures.

Since in the eye vision data example the data generating model is not known, we illustrate the fitting in a small simulation study in which the data generating models are known. We consider two binary covariates with $\boldsymbol{\beta}^T = (0.5, 0.5)$, $k = 5$ response categories and thresholds $\theta_r \in \{-2, \ldots, 2\}$. First data are generated by the location-scale model with varying strength of dispersion in the first variable. Then the location-scale and the location-shift model are fitted. The first row of Figure 5 shows the resulting deviances. In order to match the strength of dispersion we computed the parameter α of the location-shift model that shows approximately the same dispersion as the corresponding parameter γ of the location-scale model. The relation between these two parameters is non-linear, large values of α correspond to small values of γ. Then data were generated by the location-shift and again both models are fitted.

The resulting deviances are shown in the second row of Figure 5. It is seen that the deviances of the two models are quite close with just slightly better fits of the data generating model. If, however, the dispersion is ignored and a simple proportional odds model is fitted, the fit suffers strongly.

FIGURE 4: *Deviances of model fits for sub samples of size $n = 200$ from the eye vision data.*

FIGURE 5: *Deviances of model fits for data generated by the location-scale model (first row) and data generated by the corresponding location-shift model (second row).*

If strong variation is present the omission of corresponding effects might not only yield large deviances but also reduce the accuracy of the estimates of the location effect. The effect is illustrated by using the same data generating model as before but now we focus on the estimation of the first parameter. Figure 6 shows the estimates of the location effect β_1. In the first row the location-scale model was the data generating model, in the second row the location-shift model. It is seen that there is no bias if no dispersion effect is present. However, with increasing dispersion the estimates are biased.

In both models we used $\beta_1 = 0.5$. However, one should be aware that the parameters cannot compared directly since they represent different effects in the two

models. In the location-scale model the predictor has the form $\eta_r = \gamma_{0r}/\exp(\boldsymbol{x}^T\boldsymbol{\gamma}) + \boldsymbol{x}^T\boldsymbol{\beta}/\exp(\boldsymbol{x}^T\boldsymbol{\gamma})$. In particular, the dispersion is also included in the location term. For a simple binary predictor $x \in \{0,1\}$, the location term is $\boldsymbol{x}^T\boldsymbol{\beta}/\exp(\boldsymbol{x}^T\boldsymbol{\gamma}) = x\beta/\exp(x\gamma)$, which for $x = 1$ takes the value $\beta/\exp(\gamma)$. Thus, if one ignores the possible variation and fits a model that does not account for it one estimates the parameter $\beta/\exp(\gamma)$ instead of β. Therefore, if γ is positive one can expect a bias towards zero, if γ is negative, one will overestimate the strength of the location effect. This effect is seen from the first row of Figure 6. The bias can be severe if γ is large, for example, if $\gamma = 1.5$, estimates are very close to zero, which is not surprising since $\beta/\exp(\gamma) = 0.5/4.48 = 0.11$. In the location-shift model the tendency of the bias is different. As is seen from the second row small values of α (stronger dispersion) yield stronger location effects. For positive values of α the estimated effects are weaker ($\alpha = 0.8$), for large values of α ($\alpha > 2$), however, even the sign of the effect changes. The effects are similar if one considers negative values of β_1 (not shown). Overall, it is seen that ignoring dispersion effects may yield strongly biased estimates.

FIGURE 6: *Estimates of the location parameter β_1 for data generated by the location-scale model (first row) and data generated by the corresponding location-shift model (second row).*

4 Non-Symmetric Responses

4.1 Modelling Varying Dispersion in Non-Symmetric Responses

In the previous section we considered symmetric responses, which often occur in survey data if the extent of the agreement to a statement is evaluated. However, also non-symmetric responses may show dispersion that varies over sub populations. The dispersion modelled so far means varying variability centered at a middle category, which is quite natural for a symmetric response. For non-symmetric responses one may pick a category m and model the variability with a centering between m and $m + 1$ as in model (5).

	Covariate	Proportional Odds Model			Loc-Shift Model (m=3)		
		estimate	std error	z value	estimate	std error	z value
	Intercept1	2.541	1.940	1.309	3.980	2.250	1.769
	Intercept2	3.803	1.957	1.943	3.564	2.011	1.773
	Intercept3	4.809	1.971	2.440	3.059	2.598	1.177
	Intercept4	6.823	2.016	3.385	3.729	3.669	1.017
location effects	Treatment	0.938	0.331	2.834	1.309	0.372	3.513
	Age	-0.372	0.129	-2.871	-0.345	0.149	-2.312
	Age^2	0.006	0.002	3.006	0.006	0.002	2.437
dispersion effects	Treatment				0.636	0.254	2.508
	Age				0.032	0.094	0.343
	Age^2				-0.001	0.002	-0.194

For distinct variables $\boldsymbol{x}$ and $\boldsymbol{z}$ the interpretation of the parameters is the same as in model (5) because the derivation of the parameters does not depend on the chosen m. Thus one has several models depending on the chosen category m. The goodness-of-fit of the model measured by the deviance can be used to select a model. It turned out that the estimated location effect $\boldsymbol{\beta}$ depends very weakly on the choice of m whereas the values of the dispersion effects $\boldsymbol{\alpha}$ do depend on m.

However, the case $\boldsymbol{x} = \boldsymbol{z}$ is different. Then it does not matter which category m is chosen, all models (5) with any fixed m are equivalent. The only difference is in the interpretation of parameters. The equivalence is seen by transforming the parameters. Let $\beta_{r0}^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\alpha}^{(m)}$ denote the parameters of the model (5) for fixed category m. It can be shown that for two values m and l

$$\beta_{r0}^{(m)} = \beta_{r0}^{(l)}, r = 1, \ldots, k-1, \quad \boldsymbol{\alpha}^{(m)} = \boldsymbol{\alpha}^{(l)}, \quad \boldsymbol{\beta}^{(m)} = \boldsymbol{\beta} + (m-l)\boldsymbol{\alpha}^{(l)}.$$

That means, the intercepts and the dispersion parameters $\boldsymbol{\alpha}$ do not depend on the choice of m. The only parameters that depend on the choice of m are the $\boldsymbol{\beta}$ parameters, and the transformation uses the $\boldsymbol{\alpha}$ parameters. To obtain the interpretation as dispersion parameters again a middle category is a good choice because one obtains

$$\frac{\gamma(r|x+1)}{\gamma(r|x)} = \begin{cases} e^\beta e^{-(m-r)\alpha}, & r \in \{1, \ldots, m\} \\ e^\beta e^{(r-m)\alpha}, & r \in \{m+1, \ldots, k-1\}, \end{cases}$$

and in particular

$$e^\beta = \frac{\gamma(m|x+1)}{\gamma(m|x)}.$$

That means e^β refers to the increase of x by one unit for the fixed category m and α is determined by

$$e^{-\alpha} = \frac{\gamma(r|x+1)/\gamma(r|x)}{\gamma(r+1|x+1)/\gamma(r+1|x)}.$$

4.2 Application: Knee Injuries

As an application we consider data from a clinical trial ($n = 127$) that investigates the effect of a therapy on the recovery of knee injuries. The response is the pain during

TABLE 5: *Parameter estimates, standard errors and z-values for the retinopathy data.*

	Covariate	Proportional Odds Model			Location-Shift Model		
		estimate	std error	z value	estimate	std error	z value
location effects	SM	-0.254	0.191	-1.328	-0.159	0.198	-0.802
	DIAB	-0.139	0.013	-10.368	-0.148	0.014	-10.524
	GH	-0.459	0.074	-6.175	-0.485	0.076	-6.324
	BP	-0.072	0.013	-5.357	-0.071	0.014	-5.204
dispersion effects	SM				0.491	0.235	2.087
	DIAB				-0.037	0.016	-2.254
	GH				-0.101	0.092	-1.099
	BP				-0.007	0.015	-0.465

Covariate	Partial Proportional Odds Model		
	estimate	std error	z value
SM1	-0.405	0.205	-1.972
SM2	0.086	0.254	0.340
DIAB1	-0.129	0.014	-8.889
DIAB2	-0.166	0.018	-9.264
GH1	-0.435	0.080	-5.426
GH2	-0.535	0.097	-5.470
BP1	-0.068	0.014	-4.627
BP2	-0.075	0.017	-4.432

movement measured on a scale from 1 (no pain) to 5 (severe pain), for more details see Tutz (2012). We model the treatment effect (1: therapy, 0: placebo) and the effect of the covariate age in years. With a quadratic effect of age the predictor of the proportional odds model without dispersion has the form

$$\eta_r = \theta_r + x_{treat}\beta_{treat} + x_{age}\beta_{age} + x_{age}^2\beta_{age^2}, \quad r = 1, \ldots, 4$$

The estimated coefficients and corresponding standard errors for the simple proportional odds model and the location-shift model with scaled shifting of thresholds $m = 3$ are given in Table 4. For the simple proportional odds model the deviance is 362.9 on 501 degrees of freedom and for the models with location and dispersion effects the deviance is 356.3 on 498 degrees of freedom. There are significant location effects for treatment and the linear and the quadratic effect of age. Concerning the dispersion part only the treatment effect with estimate $\hat{\alpha}_{treat} = 0.636$ seems to be relevant. However, the inclusion of dispersion effects yields a stronger location effect of the variable treatment.

5 Partial Proportional Odds Models versus the Modelling of Dispersion

If the proportional odds model does not fit the data well, one option is to introduce category-specific parameters, which corresponds to using the partial proportional odds model, the other option is to include dispersion effects. Both modelling strategies will yield a better fit. In the following we briefly consider these two options.

An interesting case is the modelling of three response categories ($k = 3$). Then the two predictors of the location shift model are $\eta_1 = \beta_{01} + \boldsymbol{x}^T\boldsymbol{\beta} - \boldsymbol{x}^T\boldsymbol{\alpha}$, $\eta_2 =$

TABLE 6: *Parameter estimates, standard errors and z-values for the election data.*

	Covariate	Proportional Odds Model			Location-Shift Model		
		estimate	std error	z value	estimate	std error	z value
location effects	Gender	0.628	0.088	7.137	0.583	0.091	6.391
	Age	-0.012	0.002	-4.385	-0.013	0.002	-4.425
	Age^2	0.001	0.001	5.041	0.001	0.001	4.966
	College	-1.419	0.095	-14.864	-1.466	0.105	-13.954
	Home	-0.410	0.096	-4.234	-0.432	0.097	-4.415
	Length	-1.134	0.149	-7.596	-1.212	0.155	-7.796
dispersion effects	Gender				0.141	0.072	1.966
	Age				-0.001	0.002	-0.445
	Age^2				0.001	0.001	1.788
	College				0.108	0.085	1.279
	Home				0.176	0.077	2.266
	Length				0.217	0.122	1.772

$\beta_{02} + \boldsymbol{x}^T\boldsymbol{\beta} + \boldsymbol{x}^T\boldsymbol{\alpha}$, which is the same as the reparameterized predictors $\eta_r = \beta_{0r} + \boldsymbol{x}^T\boldsymbol{\beta}_r$, where $\boldsymbol{\beta}_1 = \boldsymbol{\beta} - \boldsymbol{\alpha}$, $\boldsymbol{\beta}_1 = \boldsymbol{\beta} + \boldsymbol{\alpha}$. Therefore, the location-shift model is equivalent to the partial proportional odds model. Nevertheless, there are some benefits when using the location-shift parameterization. If the hypothesis $H_0 : \alpha_r = 0$ holds the r-th variable has global and not category-specific effects. The test result is immediately seen from the $z-$ or p-value of the corresponding parameter. Within the partial proportional odds model, one has to test the hypothesis $H_0 : \beta_1 = \beta_2$ to investigate if the r-th variable has global effects, which typically makes refitting of the model under constraints necessary. This is illustrated in a small example.

Application: Retinopathy Data

In a 6-year followup study on diabetes and retinopathy status reported by Bender and Grouven (1998) the interesting question is how the retinopathy status is associated with risk factors. The considered risk factor is smoking (SM = 1: smoker, SM = 0: non-smoker) adjusted for the known risk factors diabetes duration (DIAB) measured in years, glycosylated hemoglobin (GH), which is measured in percent, and diastolic blood pressure (BP) measured in mmHg. The response variable retinopathy status has three categories (1: no retinopathy; 2: nonproliferative retinopathy; 3: advanced retinopathy or blind). The simple proportional odds model yields deviance 904.14, the model with category-specific intercepts yields 892.45, the same as the the location-shift model. The difference, 11.69, on 4 df shows that at least some of the parameters should be category-specific. From the fitted parameters of the location-shift model (Table 5) one sees immediately that smoking and DIAB are susceptible of having category-specific effects but not GH and BP. This is not seen from the estimates of the category-specific model.

The location-shift model also provides a different interpretation of the effects of smoking and DIAB. In the location-shift model DIAB shows a strong shifting effect and also varying dispersion. Smoking shows no significant shifting effect, also in the simple proportional odds model the effect is not significant. In the category-specific model smoking for the first split into categories 1 and $\{2,3\}$ seems to be substantial (z-value -1.972) but not for the other split into categories $\{1,2\}$ and 3. Within the

location-shift model this is explained by a different dispersion over response categories for smokers and non-smokers.

FIGURE 7: *Visualization of estimated effects of the location-shift model for the election data including pointwise confidence intervals.*

FIGURE 8: *Non-linear location (left) and dispersion (right) effects of the location-shift model for the election data for covariate age.*

Application: Information about Politics

In the following we consider an application in which the extension to category-specific effects seems not necessary, however, dispersion effects are present. We use data from the American National Election Study http://www.electionstudies.org/ containing 1790 respondents from the study in 2000, see Jackman (2009). The response is on an ordinal rating scale that represents the general level of information about politics and public affairs from 1 (very low) to 5 (very high). The obtained level was assessed by the interviewer assigned to each respondent. Explanatory variables are gender (0: male, 1: female), age (centered at 47), college degree (College; yes/no), if the respondent or his family owns their home (Home) and the length of the interview (on a log scale).

When fitting a simple proportional odds model one obtains 4891.198 on 7150 df. To evaluate if effects are really global we fitted a model with category-specific effects. The

difference in deviances between the two models was 24.42 on 18 df. Therefore, one can assume that no category-specific effects are needed and the simple proportional odds model seems appropriate. However, one might also investigate if there are dispersion effects. We fitted a location-shift model with only six additional parameters (dispersion effects) to obtain the deviance 4873.526 on 7144 df. When comparing to the simple proportional odds model now one obtains a difference in deviances of 17.636 on 6 df, which indicates that dispersion effects are present. The fitted parameters and corresponding z-values of the location-shift model, given in Table 6, show that the location effects of all variables should be included in the model. Among the dispersion effects the two variables gender and home seem to be relevant. From Figure 7 it is seen that females seem to be less informed about politics and show weaker dispersion. Respondents who own their home also show weaker dispersion but are better informed about politics. Figure 8 shows the non-linear effects of the variable age. The location curve (left panel) shows that the level of information is highest at about 60 and much lower for younger and older respondents.

6 Concluding Remarks

An alternative model for the explicit inclusion of dispersion effects is proposed. In terms of goodness-of-fit the model is frequently quite similar to the location-scale model. Nevertheless the model has some advantages. It can be embedded into the framework of generalized linear models and therefore all the inference techniques and asymptotic results that have been shown to hold for this class of models can be used. The interpretation of parameters differs from that of the parameters of the location-scale model. When interpreting parameters of the location-scale model one typically refers to the underlying latent regression model. While the proportional odds model without dispersion can also be fitted and interpreted without referring to the latent model, with dispersion, however, it seems unavoidable to refer to the latent model. In contrast, parameters of the location-shift model can be interpreted straightforward in terms of log-odds.

We also investigated alternative modelling strategies. One may extend simple models with global effect to more flexible models like the partial proportional odds model or examine if dispersion effects as in the location-scale or location-shift model are present. The former strategy may yield models that are much harder to interpret. Some authors argue that simpler models as the proportional odds model are often to be preferred even if the fit is not too good because the obtained first-order effects are often informative for overall summaries that explain the most important dimension of an effect (Agresti, 2010). The second strategy, investigating if dispersion effects are needed, has the advantage that the first-order effects concerning the location are kept and summary measures concerning the location are still available. In addition, if dispersion effects are present estimates of the location effects will be less biased.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data, 2nd Edition.* New York: Wiley.

Bender, R. and U. Grouven (1998). Using binary logistic regression models for ordinal data with non–proportional odds. *Journal of Clinical Epidemiology 51*, 809–816.

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

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

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

Hamada, M. and C. F. J. Wu (1996). A critical look at accumulation analysis and related methods. *Technometrics 32*, 119–130.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Wiley: Hoboken, New Jersey.

Kim, J.-H. (2003). Assessing practical significance of the proportional odds assumption. *Statistics & probability letters 65*(3), 233–239.

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

Liu, Q. and A. Agresti (2005). The analysis of ordinal categorical data: An overview and a survey of recent developments. *Test 14*, 1–73.

McCullagh, P. (1980). Regression model for ordinal data (with discussion). *Journal of the Royal Statistical Society B 42*, 109–127.

Nair, V. N. (1987). Chi-squared-type tests for ordered alternatives in contingency tables. *Journal of the American Statistical Association 82*, 283–291.

Peterson, B. and F. E. Harrell (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics 39*, 205–217.

Peyhardi, J., C. Trottier, and Y. Guédon (2015). A new specification of generalized linear models for categorical data. *Biometrika*.

Tutz, G. (1991). Sequential models in ordinal regression. *Computational Statistics & Data Analysis 11*, 275–295.

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

Yee, T. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software 32*(10), 1–34.

Yee, T. W. (2014). *VGAM: Vector Generalized Linear and Additive Models*. R package version 0.9-4.

Yee, T. W. and C. J. Wild (1996). Vector generalized additive models. *Journal of the Royal Statistical Society B*, 481–493.