Hierarchical Models for the Analysis of Likert Scales in Regression and Item Response Analysis

Tutz, Gerhard

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Tutz, G. (2021). Hierarchical Models for the Analysis of Likert Scales in Regression and Item Response Analysis. *International Statistical Review*, *89*(1), 18-35. https://doi.org/10.1111/insr.12396

International Statistical Review

International Statistical Review (2021), 89, 1, 18–35 doi:10.1111/insr.12396

Hierarchical Models for the Analysis of Likert Scales in Regression and Item Response Analysis

Gerhard Tutz

Department of Statistics, Ludwig-Maximilians-Universität München, Akademiestraße 1, München, 80799, Germany
E-mail: tutz@stat.uni-muenchen.de

Summary

Appropriate modelling of Likert-type items should account for the scale level and the specific role of the neutral middle category, which is present in most Likert-type items that are in common use. Powerful hierarchical models that account for both aspects are proposed. To avoid biased estimates, the models separate the neutral category when modelling the effects of explanatory variables on the outcome. The main model that is propagated uses binary response models as building blocks in a hierarchical way. It has the advantage that it can be easily extended to include response style effects and non-linear smooth effects of explanatory variables. By simple transformation of the data, available software for binary response variables can be used to fit the model. The proposed hierarchical model can be used to investigate the effects of covariates on single Likert-type items and also for the analysis of a combination of items. For both cases, estimation tools are provided. The usefulness of the approach is illustrated by applying the methodology to a large data set.

Key words: adjacent categories model; cumulative model; hierarchically structured models; ordinal regression; proportional odds model; sequential model.

1 Introduction

Likert scales, which were introduced by Likert (1932), have a long tradition in the social and behavioural sciences to measure attitudes, character and personality traits. Although various versions have been used, the most popular versions are 5-grade and 7-grade Likert scales. The grades $1, \ldots, k$ are assumed to be ordered and reflect agreement/disagreement or approval/disapproval of the respondent with respect to the value statement. In 5-grade Likert scales, the grades are typically interpreted by strongly disagree, disagree, neutral (undecided), agree and strongly agree.

Some authors distinguish between Likert-type items and Likert scales. They refer to Likert-type items if individual items are considered as measurement tools and use the term Likert scale if scores are computed from a number of relating items (see e.g. Kaptein *et al.*, 2010). We will not strictly adhere to this distinction in terminology because a single item also provides a Likert scale. However, we will distinguish between the two cases when modelling responses because differing modelling strategies and estimation tools are needed.

Over the years, it has been extensively debated if Likert scales can be regarded as ordinal or interval scales (see e.g. Carifio & Perla, 2007; Jamieson, 2004). Lantz (2013) explicitly investigated the assumption of equidistance of Likert-type scales and used an experimental design to show that the perceived distance between scale points on a regular 5-point Likert-type scale depends on how the verbal anchors are used. The assumption of the scale level is crucial because it determines the tools that can be used to analyse data, for example, the mean and standard deviation are inappropriate for ordinal data. If Likert scales are designed to measure an underlying latent variable whose values characterise the respondents' attitudes, it is certainly safer to assume not more than ordinal scale level for the items. Various authors have developed instruments that use only ordinal scale level (see Gadermann *et al.*, 2012; Clason & Dormody, 1994; Kaptein *et al.*, 2010; Zumbo *et al.*, 2007). However, most of them confine the analysis to comparing groups of respondents; they do not consider regression models that include categorical and continuous explanatory variables. For example, Gadermann *et al.* (2012) and Zumbo *et al.* (2007) focus on the use of ordinal versions of coefficients alpha (and theta), which seem to work better than Cronbach's alpha.

In item response theory, the scale level has been taken more seriously. Various models for ordinal responses have been proposed; an overview is given, for example, in Van der Linden (2016). Also, the more recently proposed item response trees use ordinal scale level only (De Boeck & Partchev, 2012; Khorramdel & von Davier, 2014; Böckenholt, 2017; Meiser *et al.*, 2019). More recently, mixture models for Likert-type items have been considered by Tijmstra *et al.* (2018). However, item response theory focuses on the investigation of latent traits, whereas the focus here is on the effect of explanatory variables.

In the present paper, we will adopt the weaker assumption of an ordinal scale and assume that item responses are multinomially distributed to account for the discrete nature of the data. The objective is to propose flexible and powerful models that are able to model the effects of explanatory variables on the outcome. We will propose regression models for single Likert-type items as well as models for the combination of items. In contrast to classical item response theory, also in the latter case, we are mainly interested in the explanatory value of covariates. It is argued that the use of classical ordinal models may yield biased estimates if the preference for the neutral category varies across respondents (Section 2). Consequently, in our models, the neutral category is modelled separately. We first consider hierarchical models that combine a binary model and a classical ordinal model (Section 3). Although they can deal adequately with the problem of the neutral category, they do not explicitly make use of the specific structure of Likert scale items. The main contribution of the paper is the symmetric hierarchical model introduced in Section 4. It is composed of binary models in a hierarchical way and accounts for the specific form of Likert-type items, which are always divided into disagreement and agreement categories. The model is very flexible and powerful. It is easy to include response styles, which, when ignored, can yield biased estimates. By using binary models as building blocks, one can also use the potential of additive models to allow for smooth, typically non-linear functions in the predictor. These advantages carry over to the case of several items (Section 6). By defining and building the corresponding binary variables, one can use existing software to fit the symmetric hierarchical latent trait model.

2 Modelling Likert Scale Responses

2.1 Scale Level and Basic Models

Let $Y_i \in \{1, \ldots, k\}$ denote the response on a Likert scale of individual i. Given the structure of Likert scales, one can safely assume that the measurement is at least on an ordinal scale level because each level on the scale refers to a greater or smaller magnitude of the attitude that one wants to measure. What is questionable is that one is measuring on an interval scale, which would mean that distances between successive levels are the same. This is a rather strong assumption that typically lacks empirical foundation. Nevertheless, it is still often made, yielding dubious results.

If one uses the typical toolbox of least squares regression, including ANOVA and *t*-tests, one does not only assume that responses are metrically scaled but often also assumes implicitly that the responses follow a normal distribution. For a response that can take only five values, this is certainly not appropriate and test results may be strongly affected. As Agresti (2010) notes, ordinary least squares regression can be used to identify variables that clearly affect a response, but the approach has distinct limitations. It can yield predicted values outside the range of the responses, and it ignores that the variability of the responses is non-constant for categorical responses. In categorical data, variability is lower at the extreme predictor values 1 and k than at predictor values in the middle. For a discussion of the scale level and further references, see also Jamieson (2004), Lantz (2013), and Göb *et al.* (2007).

Ordinal models

It seems appropriate to use ordinal models to avoid the critical assumption of metrically scaled responses. Let $(Y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$ denote a sample consisting of the categorical responses Y_i and corresponding vectors of explanatory variables $\boldsymbol{x}_i$. Classical regression models for ordinal responses are *cumulative models*. They have the form

$$P(Y_i \geq r|\boldsymbol{x}_i) = F(\beta_{0r} + \boldsymbol{x}_i^T\boldsymbol{\beta}), \quad r = 2, \ldots, k,$$

where $F(.)$ is a distribution function. The most widely used model uses the logistic distribution $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$, yielding the so-called *proportional odds model*,

$$\log \frac{P(Y_i \geq r|\boldsymbol{x})}{P(Y_i < r|\boldsymbol{x})} = \beta_{0r} + \boldsymbol{x}_i^T\boldsymbol{\beta}. \tag{1}$$

Psychometricians are familiar with a quite similar model. In psychometrics, one often models the response on several items without using explanatory variables. Instead, each person has its own parameter. Replacing $\boldsymbol{x}_i^T\boldsymbol{\beta}$ by a general person parameter θ_i and changing the sign of the intercepts by using $\delta_r = -\beta_{0r}$ yields the so-called *graded response model* (Samejima, 1995; 2016),

$$P(Y_i \geq r|\theta_i) = F(\theta_i - \delta_r), \quad r = 2, \ldots, k.$$

The parameter θ_i represents the (latent) attitude of person i while δ_r, $r = 2, \ldots, k$ are item parameters, usually considered as thresholds on the latent continuum.

An alternative class of models are *adjacent categories models*. As regression models, they have the form

$$P(Y_i = r|Y_i \in \{r-1, r\}, \boldsymbol{x}_i) = F(\beta_{0r} + \boldsymbol{x}_i^T\boldsymbol{\beta}), \quad r = 2, \ldots, k,$$

where $F(.)$ again is a distribution function. The models specify the probability of observing category r given the response is in categories $\{r-1, r\}$ by a binary regression model. For the logistic version, one obtains the simple form

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

The corresponding latent trait model in psychometrics is the partial credit model. Substituting θ_i for $\boldsymbol{x}_i^T \boldsymbol{\beta}$ and reparameterising by $\delta_r = -\beta_{0r}$ yields the *partial credit model*

$$\log\left(\frac{P(Y_i = r|\theta_i)}{P(Y_i = r-1|\theta_i)}\right) = \theta_i - \delta_r, \quad r = 2, \ldots, k,.$$

The model propagated by Masters (1982) and Masters & Wright (1984) can also be seen as a polychotomous Rasch model (Andrich, 2010).

2.2 The Problem with the Neutral Category

Likert scales as the 5-grade and 7-grade scales use a midpoint that indicates neutrality ('neutral') or ambivalence ('neither agree nor disagree'). The role of the neutral category is ambivalent. For example, Kulas *et al.* (2008) investigate whether it is used to indicate a moderate standing on a trait/item, or rather is viewed by the respondent as a 'dumping ground' for unsure or non-applicable response. In the latter case, the use of the middle category as part of the integer protocol might yield strongly biased results. For illustration, let us consider the simple case of a binary predictor $x \in \{0, 1\}$. Table 1 shows possible observations with *obs* denoting a flexible number of entries. The number *obs* is an indicator for the use of the middle category in population $x = 1$. Small values indicate that the middle category is avoided; large values indicate that it is preferred. If one uses the proportional odds model or the adjacent categories model, one considers the middle category as one of the categories on the ordinal scales and neglects that it may represent some sort of decision avoidance. Table 2 shows the estimates of β and the corresponding p values if one fits a proportional odds model (left) or an adjacent categories model with logit link (right).

It is seen that for $obs = 10$, estimates suggest that there is an effect of the covariate, but significance is weak. With increasing *obs*, the parameter estimates increase and the p value decreases. For $obs = 100$ and $obs = 200$, which indicate a strong preference for the middle category, one obtains very large values $\hat{\beta}$, which are highly significant. However, the strong effects are due to the strong preference for the middle category in population $x = 1$, which is absent in population $x = 0$.

Table 1. *Simulation data.*

x	1	2	3	4	5
1	10	10	*obs*	15	15
0	15	15	10	10	10

Table 2. *Simulation.*

	Proportional Odds Model		Adjacent Categories Model	
obs	$\hat{\beta}$	p value	$\hat{\beta}$	p value
10	0.616	0.059	0.244	0.059
50	0.640	0.031	0.260	0.052
100	0.745	0.010	0.300	0.037
200	0.980	0.001	0.391	0.017

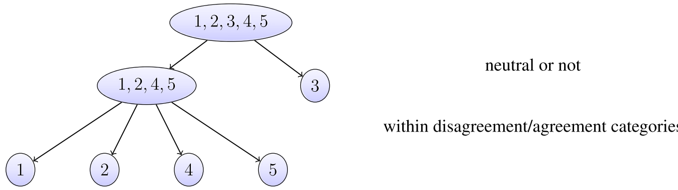

Figure 1. *A tree for five ordered categories. Categories 1,2 represent low response categories; categories 4,5 represent high response categories; and 3 is the neutral middle category. [Colour figure can be viewed at* wileyonlinelibrary.com*]*

3 Hierarchical Models Based on Ordinal Models

As has been demonstrated, if the preference for the neutral category varies across subpopulations, the neutral category should be modelled as a separate response to avoid that the preference for the neutral category is misinterpreted as a statement on the level of agreement. The separation can be obtained by hierarchical modelling. The basic concept is to model in the first step if the respondent chooses the neutral category or not. Subsequently, one models the degree of agreement given that the respondent did not choose the neutral category. The structure of the model is visualised in Figure 1.

To obtain a simple representation of models that treat the middle category as a special response, it is advantageous to rescale the response. Let $m = (k+1)/2$ denote the middle category and $Y_i^{(n)}$ represent the binary variable that codes if the neutral middle category is chosen,

$$Y_i^{(n)} = \begin{cases} 1 & Y_i \neq m \\ 0 & Y_i = m. \end{cases} \tag{3}$$

Thus, $Y_i^{(n)} = 1$ indicates that the respondent has shown some degree of preference, and $Y_i^{(n)} = 0$ indicates that the indecision category was chosen.

If $Y_i^{(n)} = 1$, only categories $1, \dots, m-1, m+1, \dots, k$ can occur. Let the variable that indicates the degree of agreement be given by

$$Y_i^{(a)} = \begin{cases} Y_i & Y_i \leq m-1 \\ Y_i - 1 & Y_i \geq m+1. \end{cases}$$

This amounts to a simple rescaling; the new categories $1, \dots, m-1$ represent the disagreement categories and $m, \dots, k-1$ the agreement categories.

3.1 Combining Classical Ordinal Models and the Preference for the Neutral Category

In the hierarchical structure shown in Figure 1, first, it is distinguished between neutral and disagreement/agreement categories, and then, given that a preference category has been chosen, one models the choice of the remaining categories by an ordinal model. If one chooses the cumulative model in the second step, the total model is given by

$$P(Y_i^{(n)} = 1 | \boldsymbol{x}_i) = F(\beta_o + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(n)}),$$
$$P(Y_i^{(a)} \geq r | Y_i^{(n)} = 1, \boldsymbol{x}_i) = F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}) \quad r = 2, \dots, k-1.$$

In the model, the preference for the middle category is determined by the linear predictor $\boldsymbol{x}_i^T \boldsymbol{\beta}^{(n)}$, whereas the preference for one of the agreement categories is determined by $\boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}$.

The model can also be given by using the original response variable Y_i,

$$P(Y_i \in \{1, \ldots, m-1, m+1, \ldots, k\}|\boldsymbol{x}_i) = F(\beta_0 + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(n)}), \text{ and}$$
$$P(Y_i \geq r|Y_i \in \{1, \ldots, m-1, m+1, \ldots, k\}, \boldsymbol{x}_i) = F(\tilde{\beta}_{0r} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}),$$

for $r \in \{2, \ldots, m-1, m+1, \ldots, k\}$, where the intercepts are rescaled using $\tilde{\beta}_{0r} = \beta_{0r}$, $r = 1, \ldots, m-1$, $\tilde{\beta}_{0,r+1} = \beta_{0r}$, $r = m, \ldots, k-1$. Thus, it is a model for the response Y_i, which consists of two components, one is a binary model and the other an ordinal model with $k-1$ response categories. Because the linear predictors do not share parameters, fitting of the model can be obtained by fitting the binary and the ordinal model separately. More on estimation methods will be given later in a separate section.

The models considered here are hierarchical models that separate the neutral category. General hierarchical regression models for ordinal responses have been considered before: an early reference is Tutz (1989) and alternative versions were given, for example, by Tutz (2012) and Peyhardi *et al.* (2016). Explicit modelling of the neutral category is also found in item response trees (De Boeck & Partchev, 2012; Khorramdel & von Davier, 2014; Böckenholt, 2017; Meiser *et al.*, 2019), which are specific hierarchical models. Structures as given in Figure 1 are also found, for example, in Böckenholt (2017); however, they refer to latent trait models and not models with explanatory variables.

3.2 Political Fears

We apply our modelling strategy to data from the German Longitudinal Election Study, which is a long-term study of the German electoral process (Rattinger *et al.*, 2014). The data we are using originate from the pre-election survey for the German federal election in 2017 and are concerned with political fears. The participants were asked: 'How afraid are you due to the ...' refugee crisis? global climate change? international terrorism? globalisation? use of nuclear energy? The answers were measured on Likert scales from 1 (not afraid at all) to 7 (very afraid). The explanatory variables in the model are *Abitur* (high school leaving certificate, 1: Abitur/A levels; 0: else), *Age* (age of the participant), *EastWest* (1: East Germany/former GDR; 0: West Germany/former FRG), *Gender* (1: female; 0: male) and *Unemployment* (1: currently unemployed; 0: else). The variable *EastWest* refers to the current place of residence where all Berlin residents are assigned to East Germany. We use data consisting of 2036 observations.

We first consider the response referring to fears concerning the use of nuclear energy. Table 3 shows the parameter estimates when fitting a hierarchical model that uses a proportional odds model to model the degrees of agreement (left columns). The variables referring to the choice of the neutral category are denoted by *Neut*. It is seen that age (*AgeNeut*) and the education level (*AbiturNeut*) have an effect on the choice of the neutral category. Respondents with a higher level of education have a stronger tendency to avoid the neutral category than respondents with lower education level. Also, older people avoid the neutral category more often than younger respondents. Given that the neutral category was not chosen, *Age*, *Gender* and *EastWest* have an impact on the degree of agreement. For example, older people tend to show more fear concerning the dangers of nuclear energy than younger people.

For comparison, the right columns of Table 3 give the estimated parameters when fitting a simple proportional odds model, which means that the neutral category is treated simply as one of the ordered categories. It is seen that the parameter estimates differ from the estimates obtained for the hierarchical model, but at least for the significant variables, the differences are not very large. Thus, in this application, the effects on the preferences for the neutral categories seem not to be so strong that the effect sizes in the simple proportional odds model deteriorate.

Table 3. *Modelling fears of the use of nuclear energy (separate fits).*

	Ordinal Model, Separated Fits				Ordinal Model For All Categories			
	Estimate	Standard Error	z value	Pr(>|z|)	Estimate	Standard Error	z value	Pr(>|z|)
Age	0.018142	0.002413	7.517	0.0000	0.016231	0.002167	7.492	0.0000
Gender	0.623739	0.087689	7.113	0.0000	0.581909	0.078974	7.368	0.0000
Unemployment	0.113142	0.278636	0.406	0.6847	–0.029039	0.248390	–0.117	0.9077
EastWest	–0.508401	0.093793	–5.420	0.0000	–0.514423	0.084506	–6.087	0.0000
Abitur	–0.091530	0.089852	–1.019	0.3083	–0.045666	0.081350	–0.561	0.5755
AgeNeut	0.007875	0.003201	2.460	0.0139				
GenderNeut	0.052455	0.117486	0.446	0.6553				
UnemployNeut	0.007706	0.360010	0.021	0.9829				
EastWestNeut	–0.095872	0.125028	–0.767	0.4432				
AbiturNeut	0.308369	0.124211	2.483	0.0130				

Nevertheless, when fitting a proportional odds model, one does not see the effects of variables on the choice of the neutral category, which might be interesting by themselves.

4 The Symmetric Hierarchical Model

Although the combination of an ordinal model and a binary model can be used for Likert scale responses, the model is not explicitly designed for this type of items and does not make efficient use of its structure. For example, it does not use that Likert-type items contain two groups of categories that are similar to each other, namely, disagreement and agreement categories. It is just assumed that categories are ordered. One consequence is that it is less straightforward to include response style effects than in the symmetric model proposed in the following.

The model proposed here is hierarchical and symmetric. Its main feature is that it uses binary models as building blocks, but in a specific way. Binary models as building blocks are not uncommon; all ordinal models contain binary models. The binary models may refer to groups of categories or single categories. For example, the proportional odds model contains binary logit models that compare the categories $\{r, \dots, m\}$ and $\{1, \dots, r-1\}$ see (1), whereas the adjacent categories model compares the adjacent categories r and $r-1$ see (2). An alternative is to compare groups of categories in a hierarchical way. Figure 2 shows how the categories that indicate a degree of agreement can be successively split into groups of categories in a symmetric way. In the first step, the split is into the groups $\{1, 2\}$ and $\{4, 5\}$; in the next step, the former is split into $\{1\}$ and $\{2\}$ and the latter into $\{4\}$ and $\{5\}$. Thus, the first step distinguishes between agreement and disagreement categories while the second step distinguishes between extreme and less extreme preferences. It is crucial that all the splits are *binary* because only then can one use the full potential of binary models, which will be demonstrated later.

The *symmetric hierarchical model* to be considered here has again two components. The first, given by

$$P(Y_i^{(n)} = 1|\boldsymbol{x}_i) = F(\beta_o + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(n)}),$$

is the same as in the hierarchical model considered previously. The difference is in the modelling of the degrees of agreement. For the disagreement categories $1, \dots, m-1$, one uses the binary models

$$P(Y_i \geq r|Y_i \in \{1, \dots, r\}, \boldsymbol{x}_i) = F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}), \quad r = 2, \dots, m-1. \tag{4}$$

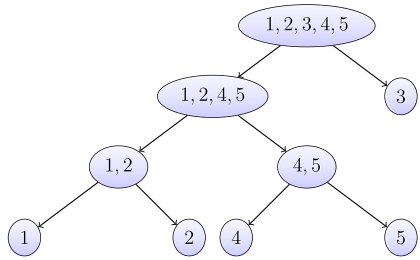

Figure 2. *A tree for five ordered categories. Categories 1,2 represent low response categories; categories 4,5 represent high response categories; 3 is the neutral middle category. [Colour figure can be viewed at* wileyonlinelibrary.com*]*

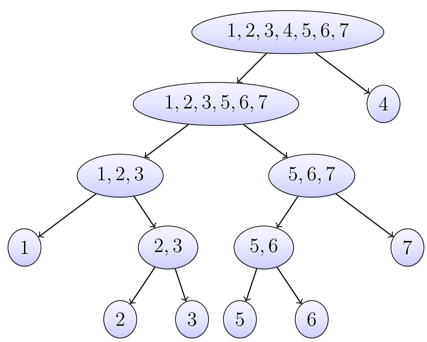

Figure 3. *Tree for seven ordered categories. Categories 1,2,3 represent levels of disagreement and categories 5,6,7 represent levels of agreement. [Colour figure can be viewed at* wileyonlinelibrary.com*]*

For the agreement categories $m+1, \ldots, k-1$, one uses

$$P(Y_i \geq r | Y_i \in \{r-1, \ldots, k\}, \boldsymbol{x}_i) = F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}), \quad r = m+1, \ldots, k. \tag{5}$$

The model is particularly simple for five categories. Then, (4) and (5) are given by

$$P(Y_i = 2 | Y_i \in \{1, 2\}, \boldsymbol{x}_i) = F(\beta_{o2} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}),$$
$$P(Y_i = 4 | Y_i \in \{3, 4\}, \boldsymbol{x}_i) = F(\beta_{o4} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}).$$

The binary models simply determine which of the two categories is preferred within disagreement and agreement categories. However, the model can be used for any odd number of response categories. Then, the binary decisions compare groups of categories within disagreement and agreement categories. For illustration, Figure 3 shows the corresponding tree for $k = 7$, which contains two levels of extremity of responses.

Political fears

Table 4 shows the estimated parameters when fitting the symmetric ordinal model. Again, the preferences for the neutral category is denoted by using *Neut* in addition to the name of the variable. Due to the hierarchical modelling, the estimates of the effects on the choice of the neutral category are the same as in Table 3. The effects of variables on the degrees of agreement differ from the effects in Table 3 because the models have different parameterisations. However,

Table 4. *Symmetric hierarchical model for response NuclearEnergy* .

	Estimate	Standard Error	*z* value	Pr(>|z|)
Age	0.0136	0.0017	7.92	0.0000
Gender	0.3941	0.0624	6.31	0.0000
Unemployment	0.1994	0.1981	1.01	0.3142
EastWest	–0.2871	0.0667	–4.30	0.0000
Abitur	–0.0644	0.0638	–1.01	0.3130
AgeNeut	0.0079	0.0032	2.46	0.0139
GenderNeut	0.0525	0.1175	0.45	0.6553
UnemploymentNeut	0.0077	0.3600	0.02	0.9829
EastWestNeut	–0.0959	0.1250	–0.77	0.4432
AbiturNeut	0.3084	0.1242	2.48	0.0130

although the sizes of effects differ, the same variables are found to have an impact on the response.

4.1 Including Response Styles

One advantage of the symmetric model is that one can easily include response style effects. Response styles have been investigated in the literature for quite some time (see e.g. Messick, 1991; Baumgartner & Steenkamp, 2001; Marin *et al.*, 1992; Meisenberg & Williams, 2008; Van Vaerenbergh & Thomas, 2013 in the social science literature and; Bolt & Newton, 2011, Johnson, 2003; Böckenholt, 2017; Tutz *et al.*, 2018 in item response theory).

In the following, we will consider the extreme response style, which, as used here, means that respondents have a tendency to extreme or middle categories. In the literature, extreme response style sometimes refers only to the tendency to extreme categories. However, symmetric hierarchical models allow to model both tendencies as opposing tendencies simultaneously.

Let z_i denote a vector of explanatory variables that is potentially linked to an extreme response style. The vector z_i can be the same as $\boldsymbol{x}_i$, can be different or contain parts of $\boldsymbol{x}_i$. The structure in Figure 2 suggests a way how to include response style effects. In the last step, the splits distinguish between the extreme categories, $\{1\}$ and $\{2\}$ when $\{1, 2\}$ are split, and $\{4\}$ and $\{5\}$ when $\{4, 5\}$ are split. If persons have a tendency to extreme categories, they will prefer $\{1\}$ in the former split and $\{5\}$ in the latter split. This is obtained by assuming

$$\begin{aligned} P(Y_i = 5|Y_i \in \{4,5\}, \boldsymbol{x}_i) &= F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)} + z_i \boldsymbol{\gamma}^{(a)}), \\ P(Y_i = 2|Y_i \in \{1,2\}, \boldsymbol{x}_i) &= F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)} - z_i \boldsymbol{\gamma}^{(a)}). \end{aligned}$$

Thus, an increase in $z_i \boldsymbol{\gamma}^{(a)}$ increases the tendency to choose category 5 and decreases the tendency to category 2 (and therefore increases the tendency to choose category 1). In summary, it increases the tendency to extreme categories. In general, the model has the form

$$\begin{aligned} P(Y_i \geq r|Y_i \in \{r-1, \ldots, k\}, \boldsymbol{x}_i) &= F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)} + z_i \boldsymbol{\gamma}^{(a)}), \quad r = m+1, \ldots, k, \\ P(Y_i \geq r|Y_i \in \{1, \ldots, r\}, \boldsymbol{x}_i) &= F(\beta_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)} - z_i \boldsymbol{\gamma}^{(a)}), \quad r = m+1, \ldots, k. \end{aligned}$$

It should be noted that the response styles are explicitly linked to explanatory variables, in contrast to the modelling of response styles in item response trees as given by Böckenholt & Meiser (2017). Jeon & De Boeck (2016) include explanatory variables; however, they use a quite different parameterisation.

Table 5. *Hierarchical model for response NuclearEnergy with response style.*

	Estimate	Standard Error	z value	Pr(>|z|)
Age	0.0102	0.0019	5.28	0.0000
Gender	0.5232	0.0715	7.32	0.0000
Unemployment	–0.0764	0.2241	–0.34	0.7330
EastWest	–0.3436	0.0747	-4.60	0.0000
Abitur	–0.0364	0.0722	–0.50	0.6145
AgeStyle	0.0125	0.0025	4.91	0.0000
GenderStyle	–0.2155	0.0967	–2.23	0.0258
UnemploymentStyle	0.5547	0.3054	1.82	0.0693

Political fears

In the following, the model is applied to the fear of nuclear energy. Table 5 shows the estimated parameters of the symmetric model with response style effects. The response style effects are denoted by the ending *Style*. Only variables that show a response style effect are included. It is seen that age and gender definitely contain response style effects; also, for unemployment, it seems to not be negligible. While women have a tendency to prefer more moderate categories, increasing age increases the tendency to extreme categories. It should be noted that estimates of parameters that indicate the placement on the continuum of ordered responses differ from the effects seen in Table 4, where response style effects have been ignored. This is not surprising because ignoring response style effects typically yields biased estimates (see e.g. Tutz & Berger, 2016).

4.2 Non-Linear Effects

Generalised linear models rely on a linear predictor with the consequence that the found effects can be a crude approximation but can also be strongly misleading if one has, for example, u-shaped effects. The last decades have seen strong progress concerning the modelling of non-linear effects. In particular, generalised additive models (GAMs) are effective tools to model smooth effects (see e.g. Hastie & Tibshirani, 1990; Marx & Eilers, 1998; and the extensive treatment in Wood, 2017).

Because the symmetric hierarchical model is constructed from binary models, it can be fitted by using binary response software as the versatile package package *mgcv*, which contains the function *gam*. One only has to transform the ordinal responses into specific binary variables (for details, see Section 5).

Let, as before, $\boldsymbol{x}_i$ denote the vector of variables for which a linear predictor is assumed. Let $v_{i1}, \ldots v_{ip}$ denote p continuous measurements on individual i. The GAM version of the symmetric hierarchical model uses for the modelling of the response in levels of agreement the form

$$P(Y_i \geq r | Y_i \in \{1, \ldots, r\}, \boldsymbol{x}_i) = F(\beta_{or} + s_1(v_{i1}) + \ldots + s_p(v_{ip}) + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}),$$

for $r = 2, \ldots, m-1$, and

$$P(Y_i \geq r | Y_i \in \{r-1, \ldots, k\}, \boldsymbol{x}_i) = F(\beta_{or} + s_1(v_{i1}) + \ldots + s_p(v_{ip}) + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}),$$

for $r = m+1, \ldots, k$, where $s_1(.), \ldots, s_p(.)$ are unknown (smooth) functions. The final form of the functions is determined by the data.

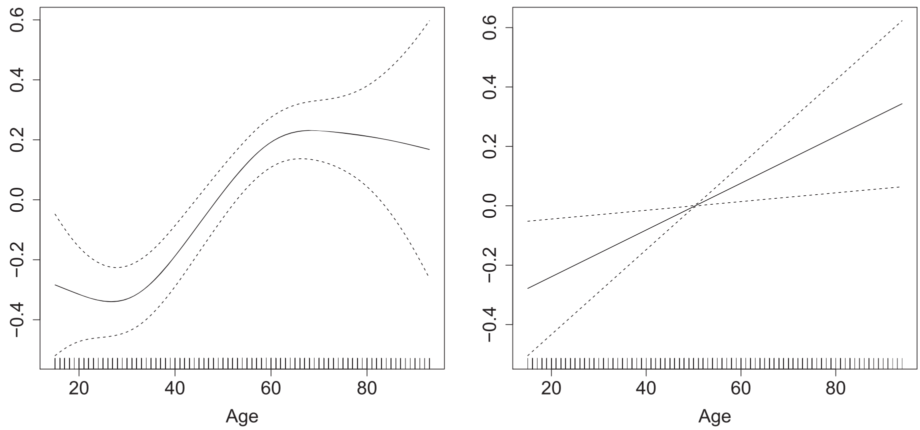

Figure 4. *Response curves for age (response nuclear energy). Left: effect on the choice of the degree of agreement. Right: effect on the choice of the neutral category.*

Table 6. *Estimates of linear effects on response NuclearEnergy when age is modelled non-parametrically.*

	Estimate	Standard Error	*z* value	Pr(>|z|)
Gender	0.3171	0.0448	–7.073	0.000
Unemployment	–0.0204	0.0857	–0.238	0.812
EastWest	0.0033	0.0294	0.113	0.910
Abitur	–0.3552	0.0289	–12.281	0.000

Political fears

In the fear of nuclear energy data, only one non-categorical variable (i.e. age) is available. While the models considered so far assume that it has a linear effect, the corresponding GAM-type model allows to model the effect of age non-parametrically. The choice of the degree of agreement as well as the choice of the middle category can be modelled non-parametrically. Figure 4 shows the resulting curves. The left panel shows the effect of age on the choice of the degree of agreement, and the right panel shows the curve for the choice of the middle category. It is seen that the effect on the choice of the middle category is linear, which is somewhat surprising because given the large number of observation, one could have expected a slightly modified function. For the choice of the degree of agreement, one obtains an increasing but non-linear function. The degree of fear is distinctly increasing only between about 40 and 65 years of age. Below 40 and above 60 years, the effect is flat. Table 6 shows the estimated parameters for the symmetric hierarchical GAM for the response in degrees of agreement. Again, the choice of the middle category is modelled separately by using a binary GAM. It is seen that the fitting of a non-linear function changes the parameter estimates given in Table 5. In particular, the level of education (*abitur*) is now significant, and *EastWest* is not. The estimates for the preference of the middle category do not change; therefore, they are not given.

5 Estimation of Parameters and Inference

The model considered in Section 3.1 uses the binary response variable $Y_i^{(n)}$ and the ordinal response variable $Y_i^{(a)}$. Due to the hierarchical structure, one may fit the two corresponding models separately. This works because the models for $Y_i^{(n)}$ and $Y_i^{(a)}$ do not share any parameters.

The same holds for the symmetric hierarchical model considered in Section 4. The binary model for $Y_i^{(n)}$ and the ordinal model for Y_i given $Y_i \in \{1, \dots, m-1, m+1, \dots, k\}$ can be estimated separately. However, the binary models given $Y_i \in \{1, \dots, m-1, m+1, \dots, k\}$ (seen in the branch below $\{1, 2, 4, 5\}$ in Figure 2) are linked because they contain the same parameter $\boldsymbol{\beta}^{(a)}$. Therefore, one cannot use separate fits of binary models to obtain estimates of the hierarchical model.

As shown in the following, simultaneous fitting can be obtained by recoding of the corresponding responses. Because we are considering estimation of the model for the choice of degrees of agreement only, it is easier to use the response variable $Y_i^{(a)}$, which takes values from $\{1, \dots, k-1\}$. Using $Y_i^{(a)}$, the relevant part of the model has the form

$$P(Y_i^{(a)} \geq r | Y_i^{(a)} \in \{1, \dots, r\}, \boldsymbol{x}_i) = F(\tilde{\beta}_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}), \quad r = 2, \dots, m-1,$$
$$P(Y_i^{(a)*} \geq r | Y_i^{(a)} \in \{r-1, \dots, k-1\}, \boldsymbol{x}_i) = F(\tilde{\beta}_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}), \quad r = m, \dots, k-1,$$

which is just an alternative representation of (4) and (5). For simplicity, we first consider the case of five response categories, in which $k - 1 = 4$. Then, with $\eta_r = \tilde{\beta}_{or} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}$, the probabilities are given by

$$P(Y_i^{(a)} = 1) = P(Y_i^{(a)} = 1 | Y_i^{(a)} \in \{1, 2\}) P(Y_i^{(a)} \in \{1, 2\}) = (1 - F(\eta_2))(1 - F(\eta_3)),$$
$$P(Y_i^{(a)} = 2) = P(Y_i^{(a)} = 2 | Y_i^{(a)} \in \{1, 2\}) P(Y_i^{(a)} \in \{1, 2\}) = F(\eta_2)(1 - F(\eta_3)),$$
$$P(Y_i^{(a)} = 3) = P(Y_i^{(a)} = 2 | Y_i^{(a)} \in \{3, 4\}) P(Y_i^{(a)} \in \{1, 2\}) = (1 - F(\eta_4)) F(\eta_3),$$
$$P(Y_i^{(a)} = 4) = P(Y_i^{(a)} = 2 | Y_i^{(a)} \in \{3, 4\}) P(Y_i^{(a)} \in \{1, 2\}) = F(\eta_4) F(\eta_3).$$

Thus, for binary variables defined by

$$Y_{ir}^{(a)} = 1 \text{ if } \quad Y_i^{(a)} \geq r, \quad Y_{ir}^{(a)} = 0 \text{ otherwise,}$$

one obtains for $Y_i^{(a)} \in \{1, 2\}$ the likelihood

$$L_i = F(\eta_2)^{Y_{i2}^{(a)}} (1 - F(\eta_2))^{1 - Y_{i2}^{(a)}} F(\eta_3)^{Y_{i3}^{(a)}} (1 - F(\eta_3))^{1 - Y_{i3}^{(a)}}$$

because one has $(Y_{i2}^{(a)}, Y_{i3}^{(a)}) = (0, 0)$ if $Y_i^{(a)} = 1$ and $(Y_{i2}^{(a)}, Y_{i3}^{(a)}) = (1, 0)$ if $Y_i^{(a)} = 2$. For $Y_i^{(a)} \in \{3, 4\}$, the likelihood is

$$L_i = F(\eta_3)^{Y_{i3}^{(a)}} (1 - F(\eta_3))^{1 - Y_{i3}^{(a)}} F(\eta_4)^{Y_{i4}^{(a)}} (1 - F(\eta_4))^{1 - Y_{i4}^{(a)}}$$

because one has $(Y_{i3}^{(a)}, Y_{i4}^{(a)}) = (1, 0)$ if $Y_i^{(a)} = 3$ and $(Y_{i2}^{(a)}, Y_{i3}^{(a)}) = (1, 1)$ if $Y_i^{(a)} = 4$. Therefore, the likelihood is the same as the likelihood for independent observation $Y_{i2}^{(a)}, Y_{i3}^{(a)}, Y_{i4}^{(a)}$ for the binary models $P(Y_{ir}^{(a)} = 1) = F(\eta_r)$.

In the general case, one obtains for $Y_{ir}^{(a)} = r \leq m - 1$

$$L_i = \prod_{s = \max\{2, r\}}^{m+1} F(\theta_s + \boldsymbol{x}^T \boldsymbol{\beta})^{Y_{is}^{(a)}} (1 - F(\theta_s + \boldsymbol{x}^T \boldsymbol{\beta}))^{1 - Y_{is}^{(a)}},$$

and for $Y_{ir}^{(a)} = r \geq m$

$$L_i = \prod_{s=m}^{\min\{r+1,k-1\}} F(\theta_s + \boldsymbol{x}^T\boldsymbol{\beta})^{Y_{is}^{(a)}}(1 - F(\theta_s + \boldsymbol{x}^T\boldsymbol{\beta}))^{1-Y_{is}^{(a)}}.$$

In summary, one has to build the binary variables $Y_{ir}^{(a)}$ and then can use the likelihood for binary models. Also, all the inference tools like Wald tests, or likelihood ratio tests, that are available for binary models can be used. Of course, the estimation also works if the linear predictor is extended to include response style effects or smooth components.

6 Item Response Modelling

In the previous sections, regression models for one item have been considered. If responses on more than one item are observed, one may analyse one item at a time by using adequate regression tools. An alternative strategy, which is common in psychometrics and social science studies, is to model responses simultaneously. Corresponding simultaneous hierarchical models are briefly considered in the following.

6.1 Hierarchical Item Response Models

Let now Y_{is} denote the response of respondent i on item s. The variable $Y_{is}^{(n)}$ is defined as before as the binary variable that indicates if the middle category is chosen.

The symmetrical hierarchical model considered here has again two components. The choices of the neutral categories are modelled simultaneously by

$$P(Y_{is}^{(n)} = 1|\boldsymbol{x}_i) = F(\beta_{0s} + b_i^{(n)} + \boldsymbol{x}_i^T\boldsymbol{\beta}^{(n)}), \quad i = 1, \ldots, n, \quad s = 1, \ldots, m,$$

where $b_i^{(n)}$ is a subject-specific parameter. For the disagreement categories $1, \ldots, m-1$, the binary models are

$$P(Y_{is} \geq r|Y_{is} \in \{1, \ldots, r\}, \boldsymbol{x}_i) = F(\beta_{0rs} + b_i^{(a)} + \boldsymbol{x}_i^T\boldsymbol{\beta}^{(a)}), \quad r = 2, \ldots, m-1.$$

For the agreement categories $m+1, \ldots, k$, they are

$$P(Y_{is} \geq r|Y_{is} \in \{r+1, \ldots, k\}, \boldsymbol{x}_i) = F(\beta_{0rs} + b_i^{(a)} + \boldsymbol{x}_i^T\boldsymbol{\beta}^{(a)}, \quad r = m+1, \ldots, k.$$

The subject-specific parameter $(b_i^{(n)}, b_i^{(a)})$ represents the respondent's tendency to choose the neutral category and the level of agreement, respectively. They may be seen as the individual's positioning on the latent attitude scale. If one drops the influence of covariates ($\boldsymbol{x}_i^T\boldsymbol{\beta}^{(n)}$ and $\boldsymbol{x}_i^T\boldsymbol{\beta}^{(a)}$), one obtains a (hierarchical) latent trait model. It should be noted that the model contains a large number of parameters because the item parameters β_{ors} vary across categories, $r = 2, \ldots, k$, and items, $s = 2, \ldots, m$.

The model is a specific hierarchical model that separates the neutral category and is symmetric in the evaluation of the degrees of agreement. A wide variety of hierarchical *latent trait models* has been proposed in the last years under the name IR-Tree models for Item Response-trees (see, among others, De Boeck & Partchev, 2012; Böckenholt, 2012; Khorramdel & von Davier, 2014; Böckenholt, 2017 and Böckenholt & Meiser, 2017). In IR-Trees, typically no explanatory variables are included; they are pure latent trait models. Moreover, each split contains its own parameters. This is different in the model considered here. All splits in the branch that refers to the choice of agreement degrees contain the same parameter vector $\boldsymbol{\beta}^{(a)}$ and the same subject-specific parameter $b_i^{(a)}$. Thus, the ordinality of the response is used more efficiently.

More recently, IR-Tree models that include the same parameters in different steps have been investigated by Meiser *et al.* (2019). Moreover, the whole class of basic IR-Trees has been extended to generalised item response trees by Jeon & De Boeck (2016). Generalised versions do include explanatory variables in the form of linear predictors and can be constrained such that item parameters need not be different across nodes. The model given above can be seen as a specific generalised item response tree model.

One difference to generalised trees is that we use different estimation procedures described in the next section. Moreover, we consider a version in which the linear predictors are replaced by additive predictors, which allows to use unknown smooth functions for metrically scaled explanatory variables. That means that the $\boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}$ in the predictor for the choice of the degree of agreement is replaced by

$$s_1(v_{i1}) + \ldots + s_p(v_{ip}) + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)},$$

where $s_1(.), \ldots, s_p(.)$ are unknown (smooth) functions of the metrically scaled variables. The same modification is used for the choice of the neutral category. The extension corresponds to the additive structure used in Section 4.2 for regression models.

It should be noted that there is another class of models that shows similarities to the hierarchical models considered here, namely, multinomial processing trees as considered by Riefer & Batchelder (1988), Batchelder (1998), Batchelder & Riefer (1999), and Smith & Batchelder (2010). Multinomial process models are trees that reflect a particular type of cognitive architecture. They have typically been used for assessing the cognitive processes in experimental settings. Process models are also represented by trees that are tailored explicitly to particular psychological paradigms (Batchelder & Riefer, 1999). The trees can be designed in a rather general way such that an observed response category can arise from from one or more unobserved processing sequences. In contrast to the trees considered here, terminal nodes are not distinct, and observed response categories can arise from different processes, that is, paths within the tree. Consequently, estimation methods are quite different because they have to account for unobserved branch frequencies.

6.2 Estimation

In latent trait models, full maximum likelihood cannot be recommended because too many parameters are involved. Marginal likelihood estimation typically has much better performance. Therefore, it is assumed that the subject-specific parameters are random effects; more precisely, one assumes $b_i^{(n)} \sim N(0, \sigma^{(n)})$, and $b_i^{(a)} \sim N(0, \sigma^{(a)})$. For the choice between the neutral category and alternatives, one can directly use integration methods to maximise the marginal likelihood

$$L(\boldsymbol{\beta}^{(n)}, \sigma^{(n)}) = \prod_{i=1}^{n} \int \prod_{s=1}^{m} F(\eta_{is})^{Y_{is}^{(n)}} (1 - F(\eta_{is}))^{1-Y_{is}^{(n)}} f(b_i^{(n)}) db_i^{(n)},$$

where $\eta_{is} = \beta_{0s} + b_i^{(n)} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(n)}$ and $f(b_i^{(n)})$ is the density of $N(0, \sigma^{(n)})$.

For the modelling of the degrees of agreement, one has, as in the regression case, to build the corresponding binary variables,

$$Y_{isr}^{(a)} = 1 \text{ if } \quad Y_{is}^{(a)} \geq r, \quad Y_{isr}^{(a)} = 0 \quad \text{otherwise},$$

and maximise

$$L(\boldsymbol{\beta}^{(a)}, \sigma^{(a)}) = \prod_{i=1}^{n} \int \prod_{s=1}^{m} \prod_{r} F(\eta_{is})^{Y_{isr}^{(a)}} (1 - F(\eta_{is}))^{1-Y_{isr}^{(a)}} f(b_i^{(a)}) d b_i^{(n)},$$

where $\eta_{is} = \beta_{0s} + b_i^{(a)} + \boldsymbol{x}_i^T \boldsymbol{\beta}^{(a)}$ and $f(b_i^{(a)})$ is the density of $N(0, \sigma^{(a)})$. The range of the index r depends on the category that has been chosen (see the regression case). After building the binary dummy variables, one can again use existing software for integration methods in binary response models to maximise the marginal likelihood.

For an overview on estimation methods for generalised mixed models and integration techniques, see McCulloch & Searle (2001) and Tutz (2012). In the application, we use Gauss–Hermite integration as provided by the R program package *glmmML*, which allows one to fit generalised linear models with random intercepts and binary responses by maximum likelihood and numerical integration via a Gauss–Hermite quadrature. It can be used after the responses have been transformed to binary responses. More generally, one can allow for correlation between the random effects by using the program *glmer* from the R package *lme4*. However, in the following application, the estimated correlation was –0.07, which suggests that one might assume that random effects can be treated as independent. This is in line with results that are often found when modelling response styles. If correlation between response style and content-related parameters is allowed, it often turns out to be very small (see e.g. Bolt & Newton, 2011; Tutz *et al.*, 2018).

6.3 Fear Data

The simultaneous hierarchical model is applied to the fear data, which include the five items referring to the refugee crisis, the global climate change, the international terrorism, globalisation and the use of nuclear energy. If one fits the hierarchical latent trait model (without covariates), the estimated standard deviation of $b_i^{(n)}$ is 0.579 (standard error 0.058); for $b_i^{(a)}$, it is 1.070 (standard error 0.029). Therefore, heterogeneity with reference to the degrees of agreement is distinctly larger than heterogeneity referring to the preference of the neutral category.

The more interesting analysis is the one in which explanatory variables are included that can explain part of the heterogeneity. There is still substantial heterogeneity in the population. After including all the available explanatory variables, the estimated standard deviation of $b_i^{(n)}$ is 0.567 (standard error 0.058); for $b_i^{(a)}$, one obtains 0.959 (standard error 0.027). The standard deviation referring to the choice of the neutral category is only slightly smaller than the standard deviation for the latent trait model without covariates. Consequently, only one effect (*AgeNeut*) is distinctly significant, as seen from Table 7, which shows the estimated parameters. This is different for the degree of agreement. The reduction in standard deviation is stronger, and more variables (*Age*, *Gender*, *Abitur*) turn out to have an impact on the response.

In addition, we fitted a GAM-type mixed model in which a non-linear effect of age is allowed. Figure 5 shows the resulting plots for the effect on the choice of the degree of agreement (left) and on the choice of the neutral category (right). As in Figure 4, the degree of fear is increasing up to about 65 years of age, but then it is decreasing. For the choice of the middle category, one obtains a linear effect, which is slightly flatter than the effect seen in Figure 4. The effects of the other variables do not change much if age is modelled as a smooth function and are therefore not given.

Table 7. *Parameter estimates for symmetrical hierarchical latent trait model (fear data).*

	Estimate	Standard Error	*z* value	Pr(>|z|)
Age	0.0142	0.0014	9.8111	0.00
Gender	0.5062	0.0530	9.5359	0.00
Unemployment	0.0060	0.1678	0.0357	0.97
EastWest	0.0124	0.0567	0.2186	0.82
Abitur	–0.5194	0.0551	–9.4161	0.00
AgeNeut	0.0050	0.0016	3.0087	0.00
GenderNeut	0.0744	0.0613	1.2135	0.22
UnemployNeut	.3718	0.2110	1.7624	0.07
EastWestNeut	–0.0032	0.0659	–0.0499	0.96
AbiturNeut	0.1160	0.0638	1.8160	0.06

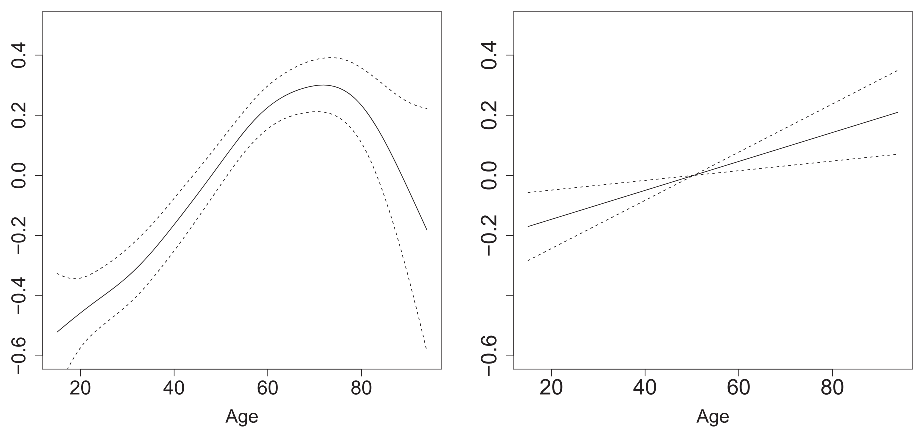

Figure 5. *Response curves for age (response nuclear energy). Left: effect on the choice of the degree of agreement. Right: effect on the choice of the neutral category.*

7 Concluding Remarks

A flexible regression model that is explicitly designed for Likert-type items is proposed. It distinguishes between disagreement and agreement categories and extremity of choices by using binary models as building blocks. The use of binary models offers several advantages; in particular, it is straightforward to include response style parameters and model the effects of covariates in a flexible smooth way. After the original data have been transformed into binary responses, available software can be used to fit the model. Estimation and inference methods for binary models are also provided.

The focus here is on Likert-type items with a neutral category because they are most widely used. However, the proposed hierarchical model can also be used for Likert-type items with an even number of categories, which consist of agreement and disagreement categories only. Then, one simply drops the binary model for the choice of the neutral category and uses only the hierarchical model for $Y_i^{(a)}$.

References

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

Andrich, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika*, **75**(2), 292–308.

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychol. Methods.*, **17**(4), 665–678.

Böckenholt, U. (2017). Measuring response styles in likert items. *Psychol. Methods.*, **22**, 69–83.

Böckenholt, U. & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *Br. J. Math. Stat. Psychol.*, **70**(1), 159–181.

Batchelder, W.H. (1998). Multinomial processing tree models and psychological assessment. *Psychol. Assess.*, **10**(4), 331.

Batchelder, W.H. & Riefer, D.M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.*, **6**(1), 57–86.

Baumgartner, H. & Steenkamp, J.-B.E.M. (2001). Response styles in marketing research: A cross-national investigation. *J. Mark. Res.*, **38**(2), 143–156.

Bolt, D.M. & Newton, J.R. (2011). Multiscale measurement of extreme response style. *Educ. Psychol. Meas.*, **71**(5), 814–833.

Carifio, J. & Perla, R.J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *J. Soc. Sci.*, **3**(3), 106–116.

Clason, D.L. & Dormody, T.J. (1994). Analyzing data measured by individual Likert-type items. *J. Agric. Educ.*, **35**(4), 4.

De Boeck, P. & Partchev, I. (2012). Irtrees: Tree-based item response models of the GLMM family. *J. Stat. Softw.*, **48**(1), 1–28.

Göb, R., McCollin, C. & Ramalhoto, M.F. (2007). Ordinal methodology in the analysis of Likert scales. *Qual. Quant.*, **41**(5), 601–626.

Gadermann, A.M., Guhn, M. & Zumbo, B.D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Pract. Assess. Res. Eval.*, **17**, 3.

Hastie, T. & Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall: London.

Jamieson, S. (2004). Likert scales: How to (ab) use them. *Med. Educ.*, **38**(12), 1217–1218.

Jeon, M. & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behav. Res. Methods*, **48**(3), 1070–1085.

Johnson, T.R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, **68**(4), 563–583.

Kaptein, M.C., Nass, C. & Markopoulos, P. (2010). Powerful and consistent analysis of likert-type ratingscales. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2391–2394, ACM.

Khorramdel, L. & von Davier, M. (2014). Measuring response styles across the big five: A multiscale extension of an approach using multinomial processing trees. *Multivar. Behav. Res.*, **49**(2), 161–177.

Kulas, J.T., Stachowski, A.A. & Haynes, B.A. (2008). Middle response functioning in Likert-responses to personality items. *J. Bus. Psychol.*, **22**(3), 251–259.

Lantz, B. (2013). Equidistance of Likert-type scales and validation of inferential methods using experiments and simulations. *Elec. J. Bus. Res. Methods*, **11**(1), 16–28.

Likert, R. (1932). A technique for the measurement of attitudes. *Arch. psychol.*. **140**, 44–53

Marin, G., Gamba, R.J. & Marin, B.V. (1992). Extreme response style and acquiescence among hispanics the role of acculturation and education. *J. Cross-Cult. Psychol.*, **23**(4), 498–509.

Marx, B.D. & Eilers, P.H.C. (1998). Direct generalized additive modelling with penalized likelihood. *Comput. Stat. Data Anal.*, **28**, 193–209.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, **47**, 149–174.

Masters, G.N. & Wright, B. (1984). The essential process in a family of measurement models. *Psychometrika*, **49**, 529–544.

McCulloch, C. & Searle, S. (2001). *Generalized, Linear, and Mixed Models*. Wiley: New York.

Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personal. Individ. Differ.*, **44**(7), 1539–1550.

Meiser, T., Plieninger, H. & Henninger, M. (2019). Irtree models with ordinal and multidimensional decision nodes for response styles and trait-based rating responses. *Br. J. Math. Stat. Psychol.*, **72**(3), 501–516.

Messick, S. (1991). Psychology and methodology of response styles. In *Improving Inquiry in Social Science: A Volume in Honor of Lee J. Cronbach*, Eds. Snow, R.E. & Wiley, D.E., Lawrence Erlbaum Associates, Publishers: Hillsdale, NJ, pp. 161–200.

Peyhardi, J., Trottier, C. & Guédon, Y. (2016). Partitioned conditional generalized linear models for categorical responses. *Stat. Model.*, **16**(4), 297–321.

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.

Riefer, D.M. & Batchelder, W.H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychol. Rev.*, **95**(3), 318.

Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, **60**(4), 549–572.

Samejima, F. (2016). Graded response model. In *Handbook of Item Response Theory*, Ed. Van der Linden, W., Springer: New York, NY, pp. 95–108.

Smith, J.B. & Batchelder, W.H. (2010). Beta-mpt: Multinomial processing tree models for addressing individual differences. *J. Math. Psychol.*, **54**(1), 167–183.

Tijmstra, J., Bolsinova, M.A. & Jeon, M. (2018). Generalized mixture IRT models with different item-response structures: A case study using likert-scale data. *Behav. Res. Methods*, **55**, 1–20.

Tutz, G. (1989). Compound regression models for categorical ordinal data. *Biom. J.*, **31**, 259–272.

Tutz, G. (2012). *Regression for Categorical Data*. New York: 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., Schauberger, G. & Berger, M. (2018). Response styles in the partial credit model. *Appl. Psychol. Meas.*, **42**, 407–427. http://doi.org/10.1177/0146621617748322

Van Vaerenbergh, Y. & Thomas, T.D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *Int. J. Public Opin. Res.*, **25**(2), 195–217.

Van der Linden, W. (2016). *Handbook of Item Response Theory*. Springer: New York.

Wood, S.N. (2017). *Generalized Additive Models: An Introduction With R*, 2. New York: Chapman and Hall/CRC.

Zumbo, B.D., Gadermann, A.M. & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *J. Mod. App. Stat. Meth.*, **6**(1), 4.

[*Received August 2019, Revised February 2020, Accepted June 2020*]