LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Gerhard Tutz & Gunther Schauberger & Moritz Berger

Response Styles in the Partial Credit Model

Technical Report Number 196, 2016
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Response Styles in the Partial Credit Model

Gerhard Tutz, Gunther Schauberger & Moritz Berger

Ludwig-Maximilians-Universität München
Akademiestraße 1, 80799 München

August 23, 2016

Abstract

In the modelling of ordinal responses in psychological measurement and survey-based research, response styles that represent specific answering patterns of respondents are typically ignored. One consequence is that estimates of item parameters can be poor and considerably biased. The focus here is on the modelling of a tendency to extreme or middle categories. An extension of the Partial Credit Model is proposed that explicitly accounts for this specific response style. In contrast to existing approaches, which are based on finite mixtures, explicit person-specific response style parameters are introduced. The resulting model can be estimated within the framework of generalized mixed linear models. It is shown that estimates can be seriously biased if the response style is ignored. In applications it is demonstrated that a tendency to extreme or middle categories is not uncommon. A software tool is developed that makes the model easy to apply.

Keywords: Partial credit model; Likert-type scales; Rating scales; Response styles; Ordinal data; Generalized linear models

1 Introduction

Response styles are an important problem in psychological measurement and survey data. Various response styles have been identified, for an overview see, for example, Messick (1991); Baumgartner and Steenkamp (2001). In Likert-type scales, which represent the level of agreement in the form *strongly disagree, moderately disagree,..., moderately agree, strongly agree* a particularly interesting response style is the extreme response style and its counterpart, the tendency to favor middle categories. The problem with response styles is that they can affect the validity of scale scores because estimates of the substantive trait may be biased if the response style is ignored. Models that explicitly account for response styles are able to reduce the bias. They account for additional heterogeneity in the population, which in some applications is itself of interest, in particular if it is linked to explanatory variables.

Various methods for investigating response styles have been proposed in survey research with a focus on the dependence of the response styles on covariates, for an overview see Van Vaerenbergh and Thomas (2013). Here, the focus is on the modelling of response styles in item response models. In item response data respondents rate their level of agreement on a series of items and response styles can be considered a consistent pattern of responses that is independent of the content of a response (Johnson, 2003). In latent trait theory one can distinguish several approaches to account for response styles by incorporating it into a psychometric model. One approach uses the nominal response model proposed by Bock (1972). Bolt and Johnson (2009) and Bolt and Newton (2011) use the multi-trait model to investigate the presence of a response style dimension. Johnson (2003) considered a cumulative type model for extreme response styles. An alternative strategy for measuring response style is the use of mixture item response theory. For example, Eid and Rauber (2000) considered a mixture of partial credit models. It is assumed that the whole population can be divided in disjunctive latent classes. After classes have been identified it is investigated if item characteristics differ between classes, potentially revealing differing response styles. Finite mixture models for item response data were also considered by Gollwitzer et al. (2005) and Maij-de Meij et al. (2008). Related latent class approaches were used by Moors (2004), Kankaraš and Moors (2009), Moors (2010) and Van Rosmalen et al. (2010). As Bolt and Johnson (2009) pointed out, in these models response style is viewed as a discrete qualitative difference in which each respondent is a member of one class, which might be a disadvantage if response style is viewed as a continuous trait.

A quite different more recent strategy uses tree methodology to investigate response styles. Trees, in general, assume a nested structure where first a decision about the direction of the response and then about the strength is obtained. Models of this type have been proposed by Suh and Bolt (2010), De Boeck and Partchev (2012), Thissen-Roe and Thissen (2013), Jeon and De Boeck (2015), Böckenholt (2012), Khorramdel and von Davier (2014) and Plieninger and Meiser (2014).

The approach proposed here differs from all these strategies. In the proposed model, for each person an additional parameter is included that indicates if the person shows a specific tendency to extreme or middle categories. In contrast to mixture models, the response style is a continuous trait that can take any value. The simultaneous estimation of ability parameters and response style parameters shows if there is some association between the substantive trait and the response style. We explicitly consider an extension of the partial credit model but the method can also be used to model response styles in alternative ordinal latent trait models. The basic concept, explicit modelling of a tendency to middle or extreme categories, has been used before by Tutz and Berger (2016). However, Tutz and Berger (2016) considered just one item and modelled the effect of covariates. Therefore, no explicit response style parameter, which varies in the population, is present and estimation methods are quite different.

In Section 2 first the Partial Credit Model is briefly considered and then the extended model which contains explicit response style parameters is introduced. In Section 3 estimation of parameters is discussed by considering alternative methods. An illustrative example is given in Section 4. In Section 5 it is investigated how the estimates suffer from the ignorance of response styles. In Section 6 further applications that illustrate the method are given. Section 7 concludes the paper, it includes in particular a short discussion of the advantages of the model over common mixture models.

2 The Extended Partial Credit Model

In this section, before introducing the model that accounts for the response style, the basic Partial Credit Model is briefly considered.

2.1 The Partial Credit Model

Let $Y_{pi} \in \{0, 1, \dots, k\}$, $p = 1, \dots, P$, $i = 1, \dots, I$ denote the ordinal response of person p on item i. The partial credit model (PCM) assumes for the probabilities

$$P(Y_{pi} = r) = \frac{\exp(\sum_{l=1}^{r} \theta_p - \delta_{il})}{\sum_{s=0}^{k} \exp(\sum_{l=1}^{s} \theta_p - \delta_{il})}, \quad r = 1, \dots, k,$$

where θ_p is the person parameter and $(\delta_{i1}, \dots, \delta_{ik})$ are the item parameters of item i. For notational convenience the definition of the model implicitly uses $\sum_{k=1}^{0} \theta_p - \delta_{ik} = 0$. With this convention an alternative form is given by

$$P(Y_{pi} = r) = \frac{\exp(r\theta_p - \sum_{k=1}^{r} \delta_{ik})}{\sum_{s=0}^{k} \exp(\sum_{k=1}^{s} \theta_p - \delta_{ik})}.$$

The PCM was proposed by Masters (1982), see also Masters and Wright (1984).

The defining property of the partial credit model is seen if one considers adjacent categories. The resulting presentation

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)}\right) = \theta_p - \delta_{ir}, \quad r = 1, \dots, k$$

shows that the model is locally (given response categories $r-1$, r) a binary Rasch model with person parameter θ_p and item difficulty δ_{ir}. It is immediately seen that for $\theta_p = \delta_{ir}$ the probabilities of adjacent categories are equal, that is, $P(Y_{pi} = r) = P(Y_{pi} = r - 1)$. That means item response curves never cross, see also middle column of Figure 1.

The partial credit model inherits from the Rasch model a specific property, namely that the comparison of items does not depend on the person parameters and the comparison of persons does not depend on the item parameters. This special property is often referred to as the specific objectivity of the Rasch model (Rasch, 1966; Rasch, 1977; Irtel, 1995). More precisely, one obtains for the comparison of items i and j

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)}\right) - \log\left(\frac{P(Y_{pj} = r)}{P(Y_{pj} = r - 1)}\right)$$
$$= \log\left(\frac{P(Y_{pi} = r)/P(Y_{pi} = r - 1)}{P(Y_{pj} = r)/P(Y_{pj} = r - 1)}\right) = -(\delta_{ir} - \delta_{jr}),$$

which does not depend on the person parameter θ_p. Thus, items i and j can be compared in terms of odds ratios with the odds defined for adjacent categories without the necessity to refer to specific persons. In this sense, it is in accordance with the general principle of specific objectivity that the results of any comparison of two "objects" (items) is independent of the choice of the "agent" (person).

In the same way, the comparison between person p and person $\tilde{p}$ is obtained by

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)}\right) - \log\left(\frac{P(X_{\tilde{p}i} = r)}{P(X_{\tilde{p}i} = r - 1)}\right)$$
$$= \log\left(\frac{P(Y_{pi} = r)/P(Y_{pi} = r - 1)}{P(X_{\tilde{p}i} = r)/P(X_{\tilde{p}i} = r - 1)}\right) = \theta_p - \theta_{\tilde{p}},$$

which does not depend on the the item parameters. Thus, the persons p and $\tilde{p}$ can be compared without reference to the item that is used in the measurement.

We mention specific objectivity because it is a strong property that separates the measurement from the tool that is used. It will be shown that the extended versions of the model considered in the following are also measurement tools that share the "specific objectivity" property.

2.2 Explicit Modelling of Response Styles

Let the categories $0, \dots, k$ represent graded agree-disagree attitudes with a natural symmetry like *strongly disagree, moderately disagree,..., moderately agree, strongly agree.* There are two possibilities, the number of categories can be odd with a neutral middle category, or the number of categories can be even so that persons have to commit themselves to a positive or negative tendency.

Odd Number of Response Categories

Let us start with an odd number of categories, that means k is even, and let $m = k/2$ denote the middle category. In the partial credit model the predictor, when choosing between categories $r-1$ and r, has the form $\eta_{pir} = \theta_p - \delta_{ir}$. The parameter δ_{ir} determines the choice between categories $r-1$ and r. Response styles that account for a tendency to extreme categories or a tendency to middle categories are modeled by modifying the parameter δ_{ir}. For the disagreement categories $1, \dots, m$ an additional person parameter γ_p is included in the predictor, for the agreement categories the person parameter $-\gamma_p$ is included. The resulting partial credit model with response style (PCMRS) has the form

$$\log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}\right) = \theta_p + \gamma_p - \delta_{ir} = \theta_p - (\delta_{ir} - \gamma_p), \quad r = 1, \dots, m$$
$$\log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}\right) = \theta_p - \gamma_p - \delta_{ir} = \theta_p - (\delta_{ir} + \gamma_p), \quad r = m+1, \dots, k.$$

The parameter γ_p can be seen as a shifting of thresholds. If γ_p is positive one has for categories $1, \dots, m$ a shifting of the thresholds δ_{ir} to the left yielding the new thresholds $\delta_{ir} - \gamma_p$, for the agreement categories $m+1, \dots, k$ one has a shifting to the right yielding the new thresholds $\delta_{ir} + \gamma_p$. The effect is that categories in the middle have higher probabilities of being chosen. The extreme case $\gamma_p \to \infty$ yields $P(Y_{pi} = m) \to 1$. If γ_p is negative one has the reverse effect; the person has a tendency to the extreme categories. For $\gamma_p \to -\infty$ the whole probability mass is in the extreme categories 0 and k. Therefore,

positive γ_p means the person has a tendency to middle categories and

negative γ_p indicates that the person has a tendency to extreme categories.

The predictor, when choosing between categories $r-1$ and r, has the closed form

$$\eta_{pir} = \theta_p + \text{sgn}(m - r + 0.5)\gamma_p - \delta_{ir}, \quad r = 1, \dots, k,$$

where $\text{sgn}(\cdot)$ denotes the sign function with $\text{sgn}(x) = 1$ if $x > 0$, $\text{sgn}(x) = -1$ if $x < 0$ and $\text{sgn}(x) = 0$ otherwise. In applications we found a scaled version typically yields better fits. Then the predictor is replaced by

$$\eta_{pir} = \theta_p + (m - r + 0.5)\gamma_p - \delta_{ir}, \quad r = 1, \dots, k,$$

or

$$\eta_{pir} = \theta_p - \tilde{\delta}_{ir}, \quad r = 1, \dots, k,$$

for the modified thresholds $\tilde{\delta}_{ir} = \delta_{ir} - (m - r + 0.5)\gamma_p$. The modification of the thresholds is done in a symmetric way with a centering in the middle of the response scale. For example, if $k = 4$, one obtains $\tilde{\delta}_{i1} = \delta_{i1} - 1.5\gamma_p, \tilde{\delta}_{i2} = \delta_{i2} - 0.5\gamma_p, \tilde{\delta}_{i3} = \delta_{i3} + 0.5\gamma_p, \tilde{\delta}_{i4} = \delta_{i4} + 1.5\gamma_p$, and the difference between adjacent thresholds changes by the same value. In general, the scaled version implies that the difference between the modified thresholds of adjacent categories changes by a constant, that is,

$$\tilde{\delta}_{ir} - \tilde{\delta}_{i,r-1} = \delta_{ir} - \delta_{i,r-1} + \gamma_p.$$

That means that for positive γ_p the difference becomes larger and smaller for negative γ_p. Therefore, for positive γ_p the probability mass is shifted to the middle categories, for negative γ_p it is shifted to the extreme categories. It should be noted that for the un-scaled version with modified thresholds $\tilde{\delta}_{ir} = \delta_{ir} - \text{sgn}(m - r + 0.5)\gamma_p$ the shifting is un-balanced. For example, if $k = 4$, one obtains $\tilde{\delta}_{i1} = \delta_{i1} - \gamma_p, \tilde{\delta}_{i2} = \delta_{i2} - \gamma_p, \tilde{\delta}_{i3} = \delta_{i3} + \gamma_p, \tilde{\delta}_{i4} = \delta_{i4} + \gamma_p$, and the difference between thresholds linked to categories 2 and 3 is $\tilde{\delta}_{i3} - \tilde{\delta}_{i2} = \delta_{i3} - \delta_{i2} + 2\gamma_p$ whereas the difference between all other adjacent thresholds is unchanged by the introduction of γ_p. Consequently we will use the scaled version in the following.

For illustration we show in Figure 1 the probabilities of response categories (upper row) and the cumulative probability functions $P(Y_{pi} > r)$ (lower row) as functions of the person abilities for varying response style parameter γ_p. The upper panels show the case $k = 2$ (three response categories), the lower show $k = 3$ (four response categories). It is immediately seen that for $\gamma_p = -1.5$ the extreme categories have much higher probabilities than for $\gamma_p = 0$. The inverse is seen for $\gamma_p = 1.5$. It is noteworthy that the probabilities of adjacent categories are now equal, that is, $P(Y_{pi} = r) = P(Y_{pi} = r - 1)$, if $\theta_p = \tilde{\delta}_{ir}$.

The extended PCMRS model still shows specific objectivity. For the comparison of items i and j one obtains again

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)}\right) - \log\left(\frac{P(Y_{pj} = r)}{P(Y_{pj} = r - 1)}\right) = -(\delta_{ir} - \delta_{jr}),$$

which does not depend on the person parameter θ_p or the response style parameter γ_p. Therefore, the comparison of item difficulties does not depend on the persons. For the comparison between person p and person $\tilde{p}$ one obtains

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)}\right) - \log\left(\frac{P(Y_{\tilde{p}i} = r)}{P(Y_{\tilde{p}i} = r - 1)}\right) = \theta_p - \theta_{\tilde{p}} + (m - r + 0.5)(\gamma_p - \gamma_{\tilde{p}}),$$

which does not depend on the the item parameters. Therefore, the comparison of persons is just a function of the person parameters, which now includes the ability parameters $\theta_p, \theta_{\tilde{p}}$ and the response style parameters $\gamma_p, \gamma_{\tilde{p}}$.

It should be noted that a constraint is needed to obtain identifiable parameters. One can use, for example, $\theta_P = 0$, or the symmetric side constraint $\sum_{p=1}^{P} \theta_p = 0$. If not mentioned otherwise we will use the symmetric side constraint.

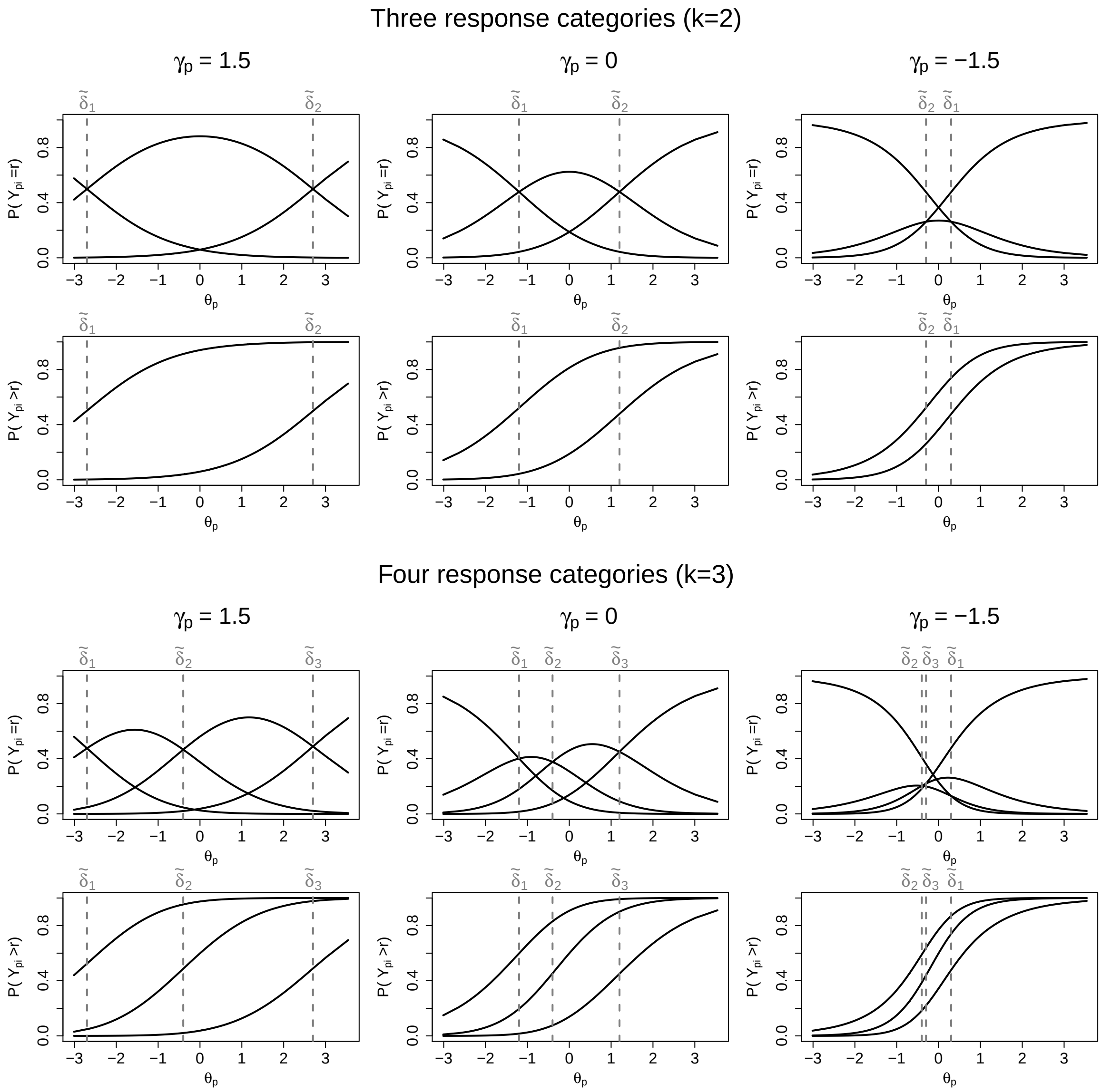

FIGURE 1: *Probabilities $P(Y_{pi} = r)$ and item response curves $P(Y_{pi} > r)$ against θ_p for positive, negative γ_p and $\gamma_p = 0$; in the upper panels the number of categories is three, in the lower panels it is four.*

Even Number of Response Categories

If k is odd, that means the number of categories is even one has a split into agreement categories at category $m = [k/2] + 1$. The corresponding model with an additional person parameter that allows to model the tendency to middle or extreme categories is given by

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)}\right) = \theta_p + \gamma_p - \delta_{ir} = \theta_p - (\delta_{ir} - \gamma_p), \quad r = 1, \ldots, m-1$$
$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)}\right) = \theta_p - \delta_{ir}, \quad r = m$$
$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)}\right) = \theta_p - \gamma_p - \delta_{ir} = \theta_p - (\delta_{ir} + \gamma_p), \quad r = m+1, \ldots, k.$$

In closed form the predictor, when choosing between categories $r-1$ and r, is given by

$$\eta_{pir} = \theta_p + \text{sgn}(m-r)\gamma_p - \delta_{ir}, \quad r = 1, \ldots, k$$

where $\text{sgn}(m-r)$ is again the sign function. In the model, thresholds for low categories are shifted to the left, others to the right. For the extreme case $\gamma_p \to \infty$ one obtains $P(Y_{pi} = m-1) + P(Y_{pi} = m) \to 1$, therefore a tendency to middle categories. For $\gamma_p \to -\infty$ one obtains $P(Y_{pi} = m-1) + P(Y_{pi} = m) \to 0$, and, therefore, a tendency to extreme categories. In the scaled version, which is considered here, one uses $(m-r)\gamma_p$ instead of $\text{sgn}(m-r)\gamma_p$.

3 Estimation

Joint Maximum Likelihood Estimation

As the PCM, also the PCMRS can be embedded into the framework of (multivariate) generalized linear models (GLMs). Then estimates can be obtained by using program packages that fit multivariate GLMs. This joint likelihood approach yields estimates for all the parameters.

The embedding into the framework of GLMs is obtained in the following way. Let the parameters be collected in the vectors $\boldsymbol{\theta}^T = (\theta_1, \ldots, \theta_{P-1})$, $\boldsymbol{\delta}_i^T = (\delta_{i1}, \ldots, \delta_{ik})$, $\boldsymbol{\gamma}^T = (\gamma_1, \ldots, \gamma_{P-1})$. With $\mathbf{1}_m^{(q)}$ denoting a unit vector of length q with a 1 in component m one has

$$\log \frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)} = (\mathbf{1}_p^{(P-1)})^T \boldsymbol{\theta} - (\mathbf{1}_r^{(k)})^T \boldsymbol{\delta}_i + (m - r + 0.5)(\mathbf{1}_p^{(P)})^T \boldsymbol{\gamma}.$$

Therefore, by appropriate specifications of the model components joint ML estimates can in principle be obtained by using the `R` package `VGAM` (Yee, 2010; Yee, 2014).

Marginal Likelihood Estimation

A disadvantage of the joint likelihood estimation is that many parameters have to be estimated, which makes estimates unstable. Moreover, estimates typically are asymptotically biased since the increase of the number of persons also increases the number of parameters to be estimated. An alternative method, which works much better, is marginal likelihood estimation. In order to reduce the number of parameters one assumes that the response style parameters are drawn from a normal distribution $N(0, \sigma^2)$. The corresponding marginal likelihood with $\boldsymbol{\delta}^T = (\boldsymbol{\delta}_1^T, \ldots, \boldsymbol{\delta}_I^T)$ is

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2) = \prod_{p=1}^{P} \int P(\{Y_{p1}, \ldots, Y_{pI}\}) f(\gamma_p) d\gamma_p,$$

where $f(\gamma_p)$ is the density $N(0, \sigma^2)$ of the random effects. The corresponding log-likelihood simplifies to

$$l(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2) = \sum_{p=1}^{P} \log \left(\int \prod_{i=1}^{I} \prod_{r=1}^{k} \{ \frac{\exp(\sum_{l=1}^{r} \theta_p + \text{sgn}(r-m)\gamma_p - \delta_{il})}{\sum_{s=0}^{k} \exp(\sum_{l=1}^{s} \theta_p + \text{sgn}(r-m)\gamma_p - \delta_{il})} \}^{y_{pir}} f(\gamma_p) d\gamma_p \right),$$

where $y_{pir} = 1$ if $Y_{pi} = r$ and $y_{pir} = 0$ otherwise.

Maximization of the marginal log-likelihood can be obtained by integration techniques.

Typically on first wants to obtain good estimates of the item parameters and estimate person parameters later for the validated test tool. Therefore, one also assumes a distribution for the person effects, which yields the marginal likelihood

$$L(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \prod_{p=1}^{P} \int P(\{Y_{p1}, \ldots, Y_{pI}\}) f(\gamma_p, \theta_p) d\gamma_p \, d\theta_p,$$

where $f(\gamma_p, \theta_p)$ now denotes the two-dimensional density of the person parameters, $N(\mathbf{0}, \boldsymbol{\Sigma})$. The diagonals of the matrix $\boldsymbol{\Sigma}$ contain the variance of the response style parameters σ_γ^2 and the variance of the person effects, σ_θ^2, the off diagonals are the covariances between response style and location effects, $\text{cov}_{\gamma\theta}$.

The corresponding log-likelihood is

$$l(\boldsymbol{\delta}, \boldsymbol{\Sigma}) =$$
$$\sum_{p=1}^{P} \log \left(\int \prod_{i=1}^{I} \prod_{r=1}^{k} \{ \frac{\exp(\sum_{l=1}^{r} \theta_p + \text{sgn}(r-m)\gamma_p - \delta_{il})}{\sum_{s=0}^{k} \exp(\sum_{l=1}^{s} \theta_p + \text{sgn}(r-m)\gamma_p - \delta_{il})} \}^{y_{pir}} f(\gamma_p, \theta_p) d\gamma_p \, d\theta_p \right).$$

The embedding into the framework of generalized mixed models allows to use methods that have been developed for this class of models. One strategy is to use joint maximization of a penalized log-likelihood with respect to parameters and random effects appended by estimation of the variance of random effects, see Breslow and Clayton (1993), Wolfinger (1993) and McCulloch and Searle (2001). However, joint maximization algorithms tend to underestimate the variances and, therefore, the true values of the random effects. An alternative strategy, which is used here, is numerical integration by Gauss-Hermite integration methods. Early versions for univariate random effects date back to Hinde (1982) and Anderson and Aitkin (1985). For an overview on estimation methods for generalized mixed model see McCulloch and Searle (2001) and Tutz (2012).

4 An Illustrative Example

As an example, we consider data from the ALLBUS, the general survey of social science carried out by the German institute GESIS. They are available from `http://www.gesis.org/allbus`. We use data containing the answers of 2535 respondents from the questionnaire in 2012. One part of the survey, comprising 8 questions, asks for the degree of confidence in public institutions and organizations. Examples are the federal constitutional court, the justice or the political parties. The answers are all measured on a scale from 1 (no confidence at all) to 7 (excessive confidence).

We fitted a simple PCM and the extended PCMRS using scaled shifting of thresholds. In both cases, marginal estimation is applied assuming normally distributed person parameters for the PCM and two-dimensional normally distributed person and response style parameters for the PCMRS. The estimated variance of the person parameters when fitting the PCM is $\hat{\sigma}^2 = 0.682$. When fitting the PCMRS one obtains the estimated covariance matrix between person and response style parameters:

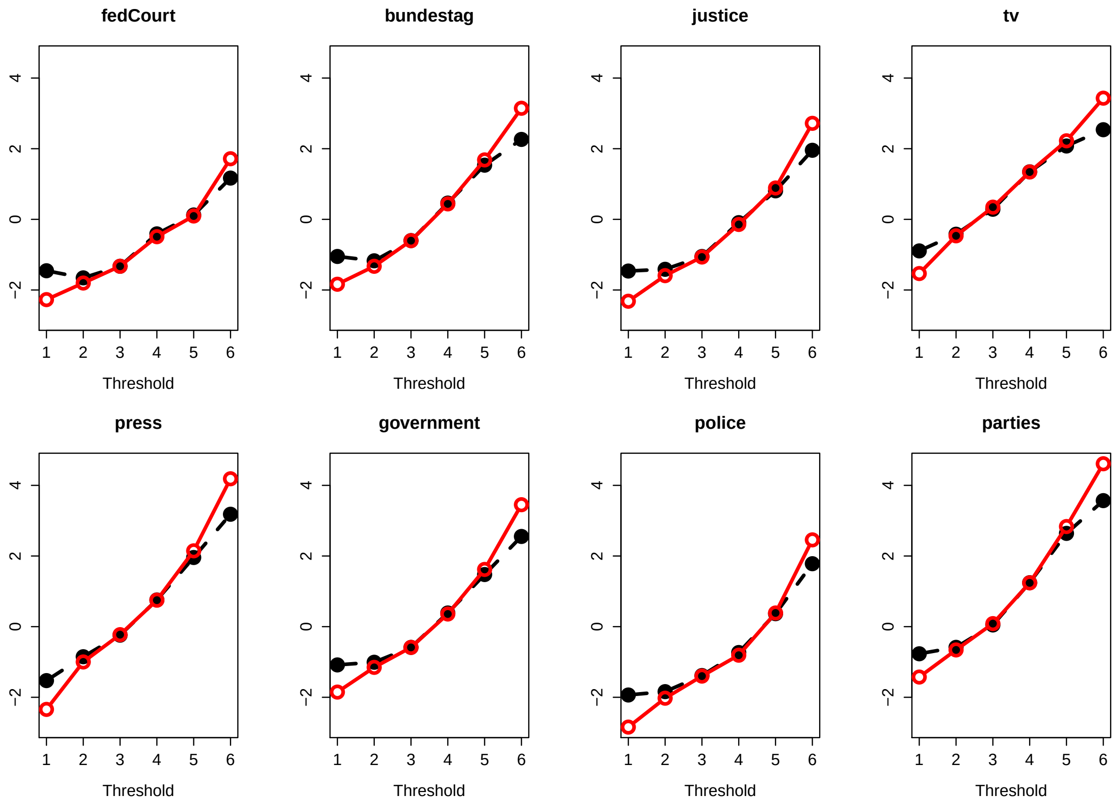

FIGURE 2: *Estimates of item parameters (ALLBUS). Solid (red) line represents estimates for PCMRS, dashed (black) line represents estimates for PCM.*

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} \hat{\sigma}_\theta^2 & \hat{cov}_{\gamma\theta} \\ \hat{cov}_{\gamma\theta} & \hat{\sigma}_\gamma^2 \end{pmatrix} = \begin{pmatrix} 0.772 & 0.023 \\ 0.023 & 0.245 \end{pmatrix}.$$

In this application the correlation between the ability parameters and the response style parameters is rather small ($\hat{\rho}_{\gamma\theta} = 0.052$). However, the standard deviation of the response style parameters ($\hat{\sigma}_\gamma = 0.495$) indicates that response styles should not be neglected when analysing the data. The estimates of the item parameters are shown in Figure 2, separately for each item. With seven response categories, there are six thresholds each. The red, solid lines correspond to the estimates for the extended PCMRS and the black, dashed lines correspond to the estimates for the simple PCM. It is striking that the estimates for the first and the last threshold strongly differ for all items, while the middle thresholds are fairly equal. If the presence of response style parameters is ignored in particular the parameters of extreme categories seem to be attenuated. As will be shown later, ignoring the response style has also consequences for the estimation of person parameters.

Although, many estimates of the item parameters coincide, there are still big differences between both models due to the presence of the response styles. For illustration, the item response curves for the item "justice" are given in Figure 3. The curves are plotted against person parameters, which are chosen between the 10% and the 90% quantile of the person parameters according to its estimated one-dimensional normal distribution $\hat{\theta} \sim N(0, 0.772)$. Again, the solid lines show the estimates for the extended PCMRS and the black lines show the estimates for the simple PCM. The first row corresponds to the curves for category 1, the second row to the curves for category

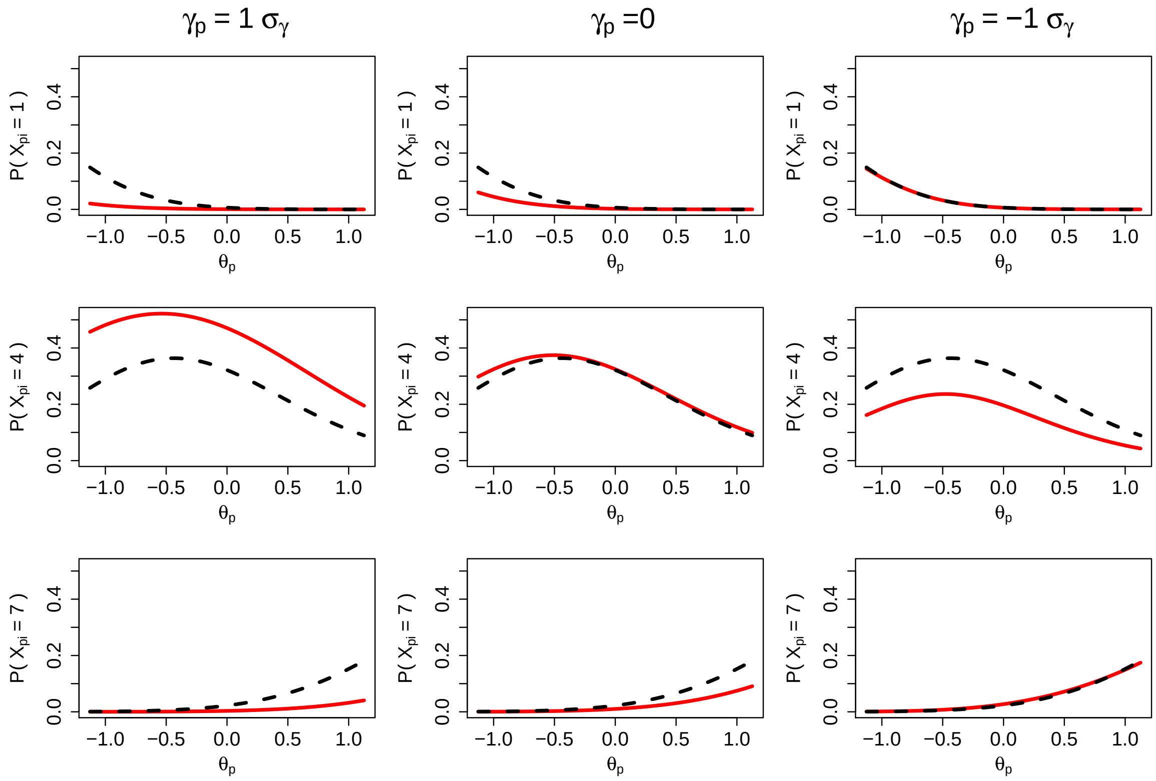

FIGURE 3: *Response curves for item "justice" (ALLBUS) along person parameter θ (between 10% and 90% quantile) for category 1 (upper panel), category 4 (middle panel) and category 7 (lower panel). Columns represent different response styles: tendency to the middle (left), no response style (middle) tendency to the extremes (right). Solid (red) lines represent estimates for PCMRS, dashed (black) lines represent estimates for PCM.*

4, and the third row shows the curves for category 7. The middle panel represents persons without response style ($\gamma_p = 0$), the left panel represents persons with a tendency to the middle ($\gamma_p = \sigma_\gamma$) and the right panel represents persons with a tendency to the extremes ($\gamma_p = -\sigma_\gamma$). With values $\gamma_p = -\hat{\sigma}_\gamma$ and $\gamma_p = \hat{\sigma}_\gamma$, the left and the right figures represent the extremes of a continuum containing 68% of the population. Obviously the curves obtained by the PCM are the same in each case. There are only minor differences between the two models for persons without response styles. However, for persons with $\gamma_p = \sigma_\gamma$ the probability for the extreme categories decreases while the probability for category 4 strongly increases when fitting the extended PCMRS. For persons with $\gamma_p = -\sigma_\gamma$ the effect is the opposite.

5 Ignoring the Response Style

The illustrative example showed that there are notable differences between the fits of the simple PCM and the extended PCMRS if response styles are present. In particular the item parameters of the extreme categories were attenuated in the PCM. In the following simulations it is demonstrated that this is the result of strongly biased estimates of the $\boldsymbol{\delta}$-parameters, which are the parameters of interest in most studies.

In the simulation study we consider exemplarily several settings with $P = 500$

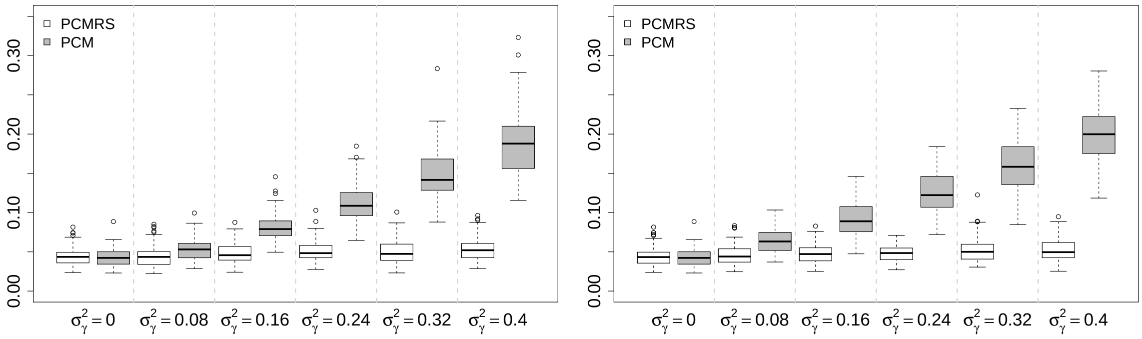

FIGURE 4: *MSEs of all item parameters for the simulations with ordered item parameters, the correlation is $\rho_{\gamma\theta} = 0$ (left) and $\rho_{\gamma\theta} = 0.3$ (right).*

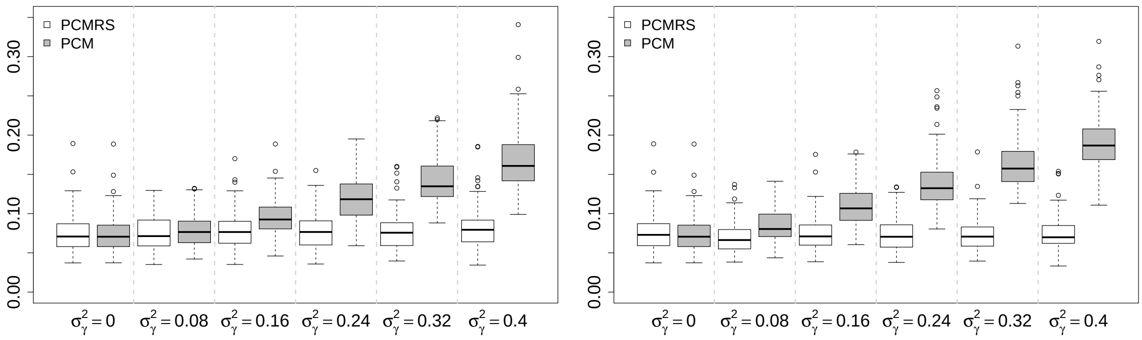

FIGURE 5: *MSEs of all item parameters for the simulations with non-ordered item parameters, the correlation is $\rho_{\gamma\theta} = 0$ (left) and $\rho_{\gamma\theta} = 0.3$ (right).*

persons and $I = 10$ items composed of 7 categories ($k = 6$). The data generating model is the PCMRS with k-dimensional normally distributed item parameters, $\boldsymbol{\delta}_i \sim N_k(\mathbf{0}, \boldsymbol{I})$. The person parameters are drawn from a two-dimensional normal distribution $(\theta_p, \gamma_p) \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$, where $\sigma^2_\theta = 1$. In each setting the variance of the response style parameters varies: $\sigma^2_\gamma \in \{0, \ldots, 0.4\}$. For $\sigma^2_\gamma = 0$ the data generating model corresponds to the simple PCM. We consider simulations with and without correlation between the ability parameters and the response style parameters ($\rho_{\gamma\theta} = 0$ or $\rho_{\gamma\theta} = 0.3$). In many applications the item parameters are ordered, that is, $\delta_{i1} \leq \cdots \leq \delta_{ik}$, but that is not necessarily the case. Therefore, we distinguish between ordered (ascending) and non-ordered item parameters. To obtain ordered item parameters, they are first drawn from the multivariate normal distribution and subsequently ordered by size. In each setting 100 data sets were generated.

Figure 4 shows the boxplots of the mean squared errors (MSEs) summarized over all item parameters, computed by $\frac{1}{60}(\sum_{i=1}^{10} \sum_{r=1}^{6} (\hat{\delta}_{ir} - \delta_{ir})^2)$, for the two settings with ordered item parameters. For each value of σ^2_γ the results of the PCMRS are given on the left (not coloured) and the results of the PCM are given on the right (gray coloured). In both settings (with and without correlation) it is seen that the MSEs are very similar for very small values of σ^2_γ, but for large values of σ^2_γ the MSEs are much larger if the response style is ignored. The picture is very similar in the case of non-ordered item parameters (Figure 5) but the increase of the MSEs with growing σ^2_γ is slightly weaker.

The poor estimation accuracy of the PCM is mainly caused by the bias, which is

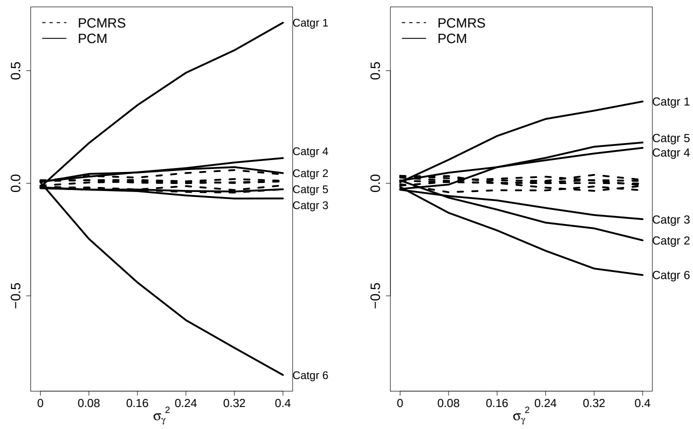

FIGURE 6: *Bias of the item parameters of the first item for the simulation with ordered item parameters (left) and non-ordered item parameters (right) and* $\rho_{\gamma\theta} = 0$.

illustrated for one item in Figure 6. The figure shows the bias of the item parameters $\delta_{1r}, r = 1, \ldots, 6$ for the setting with ordered item parameters (left panel) and for the setting with non-ordered item parameters (right panel) without correlation ($\rho_{\gamma\theta} = 0$). One obtains strongly biased estimates even for moderate values of σ_γ^2 when fitting the PCM (solid lines). Conspicuously, in the ordered case mainly the parameters of the extreme categories 1 and 6 are affected. The positive bias for parameter $\delta_{11} = -1.25$ and the negative bias for parameter $\delta_{16} = 1.71$ both indicate that the parameter sizes are strongly underestimated. These effects were already seen in the illustrative example.

In the non-ordered case one observes biased results for all categories and the bias is less systematic. Again a positive bias for the parameter $\delta_{11} = -0.84$ corresponds to an attenuation of the effect, but, for example, the positive bias for parameter $\delta_{15} = 1.71$ indicates an overestimation of the effect. These findings are very similar for the settings with correlation $\rho_{\gamma\theta} = 0.3$ (not shown).

The estimates of the components of the covariance matrix $\boldsymbol{\Sigma}$ of the person parameters (θ_p, γ_p) are given in Figure 7. It shows the results for the settings with ordered item parameters, without correlation (left panel) and with correlation (right panel). The true values are marked by red crosses. Both models yield estimates of the variance of the ability parameters σ_θ^2 (given in the first row), whereas only the extended PCMRS yields estimates of the variance of the response style parameters σ_γ^2 and the covariance $\text{cov}_{\gamma\theta}$. It is seen that both components are estimated with sufficient accuracy in both settings also for high values of σ_γ^2. In the setting with correlation the covariance (given in the third row) increases with increasing σ_γ^2. The variance of the ability parameters is slightly underestimated by both models, for the simple PCM the bias is a bit stronger. In summary, the structure parameters contained in the covariance matrix are estimated rather well.

Effect on Estimated Person Parameters

If the response style is ignored, estimates of the item parameters can be strongly biased. Consequently also the estimated person parameters will be affected. Posterior

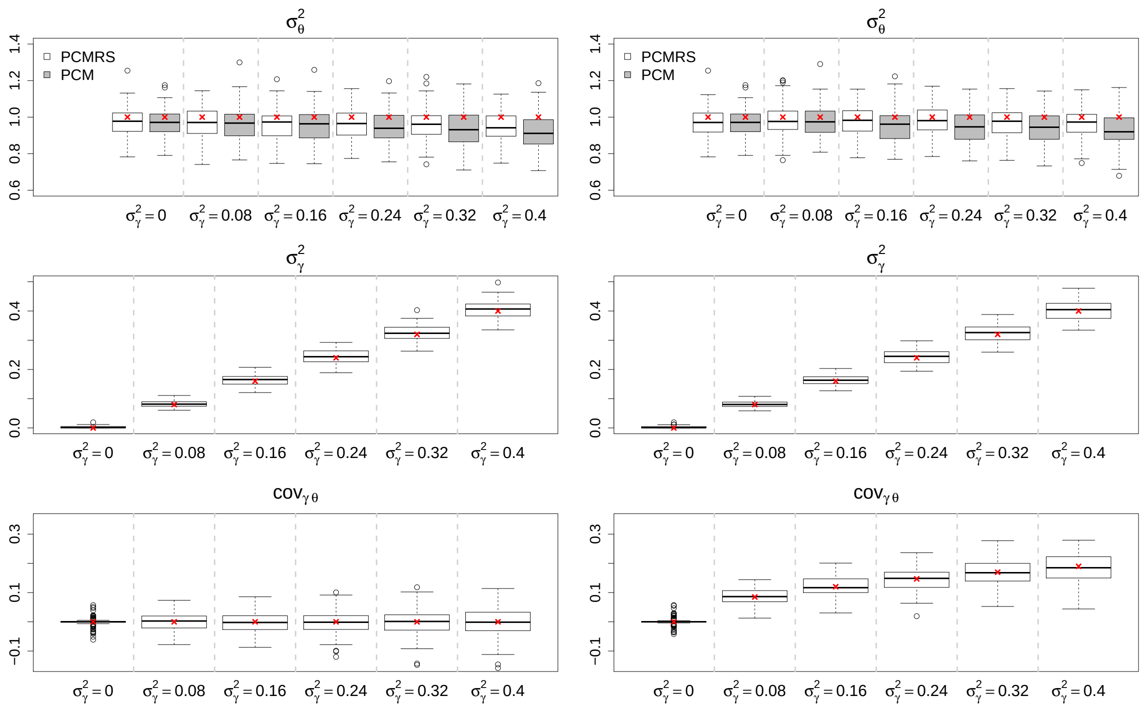

FIGURE 7: *Estimates of the variance components for the simulations with ordered item parameters and $\rho_{\gamma\theta} = 0$ (left) and $\rho_{\gamma\theta} = 0.3$ (right). The true values are marked by red crosses.*

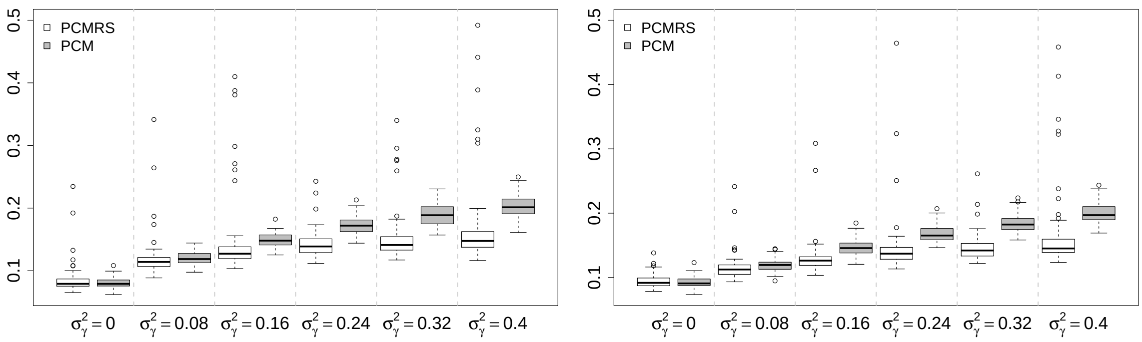

FIGURE 8: *MSEs of person parameters for the simulations with ordered item parameters, the correlation is $\rho_{\gamma\theta} = 0$ (left) and $\rho_{\gamma\theta} = 0.3$ (right).*

mean estimates of person parameters $\boldsymbol{\alpha}_p = (\theta_p^m, \gamma_p^m)^T$ can be obtained as

$$\boldsymbol{\alpha}_p^m = \mathrm{E}(\boldsymbol{\alpha}_p | \boldsymbol{Y}_{p.}) = \int \boldsymbol{\alpha}_p f_P(\boldsymbol{\alpha}_p | \boldsymbol{Y}_{p.}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Sigma}}) d\boldsymbol{\alpha}_p$$

where $\boldsymbol{Y}_{p.}^T = (Y_{p1}, \dots, Y_{pI})$, $\hat{\boldsymbol{\delta}}^T = (\hat{\delta}_{11}, \dots, \hat{\delta}_{Ik})$, and

$$f_P(\boldsymbol{\alpha}_p | \boldsymbol{Y}_{p.}, \boldsymbol{\delta}, \boldsymbol{\Sigma}) = \frac{P(\boldsymbol{Y}_{p.} | \boldsymbol{\alpha}_p, \boldsymbol{\delta}) f(\boldsymbol{\alpha}_p | \boldsymbol{\Sigma})}{\int P(\boldsymbol{Y}_{p.} | \boldsymbol{\alpha}_p,) f(\boldsymbol{\alpha}_p | \boldsymbol{\delta}, \boldsymbol{\Sigma}) d\boldsymbol{\alpha}_p},$$

$$P(\boldsymbol{Y}_{p.} | \boldsymbol{\alpha}_p, \boldsymbol{\delta}) = \prod_{i=1}^{I} P(Y_{pi} | \boldsymbol{\alpha}_p, \boldsymbol{\delta}).$$

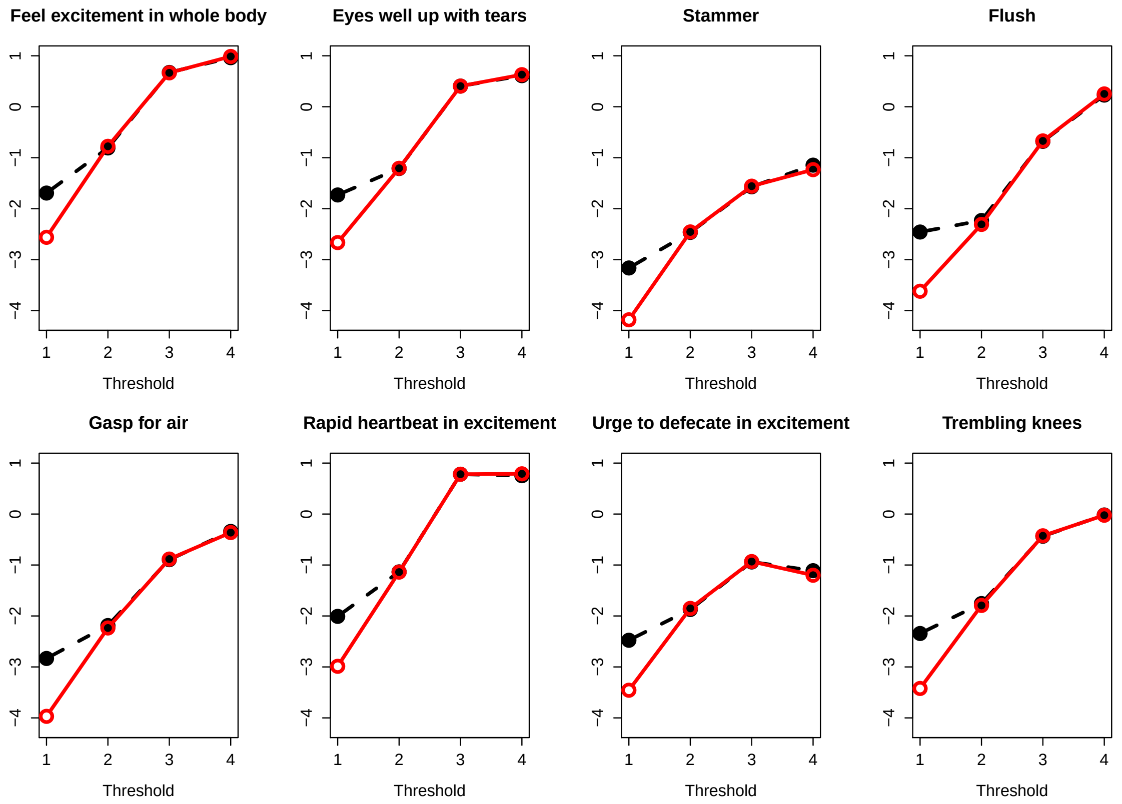

FIGURE 9: *Estimates of item parameters for items on emotional reactivity (FCC), separately for each item. Solid (red) line represents estimates for PCMRS, dashed (black) line represents estimates for PCM.*

For the computation numerical integration techniques are needed, for details, see, for example, Fahrmeir and Tutz (1997).

For illustration, Figure 8 shows the mean squared errors (MSEs) summarized over all person parameters for the two simulation settings with ordered item parameters. Again, for each value of σ_γ^2 the results for the PCMRS are given on the left (not coloured) and the results for the PCM are given on the right (gray coloured). The effect on the person parameters is similar to the effect on the item parameters, the MSEs are very similar for very small values of σ_γ^2, but for large values of σ_γ^2 the MSEs are much larger if the response style is ignored. The picture is very similar in the case of non-ordered item parameters (not given).

6 Applications

Although the model was introduced for symmetric response categories a response style as a tendency to middle or extreme categories is often also found for non-symmetric responses. In the following we will consider response categories that represent the frequency of complaints ranging from "never" to "almost every day".

In our applications we use data from the standardiziation sample of the Freiburg Complaint Checklist (FCC) (Fahrenberg, 2010; ZPID, 2013). The FCC is a questionnaire that is used to assess physical complaints of adults. The revised version of the FCC contains 71 items that can measure complaints on 9 different scales, as for example the scales *general condition*, *tenseness* or *emotional reactivity*. The primary

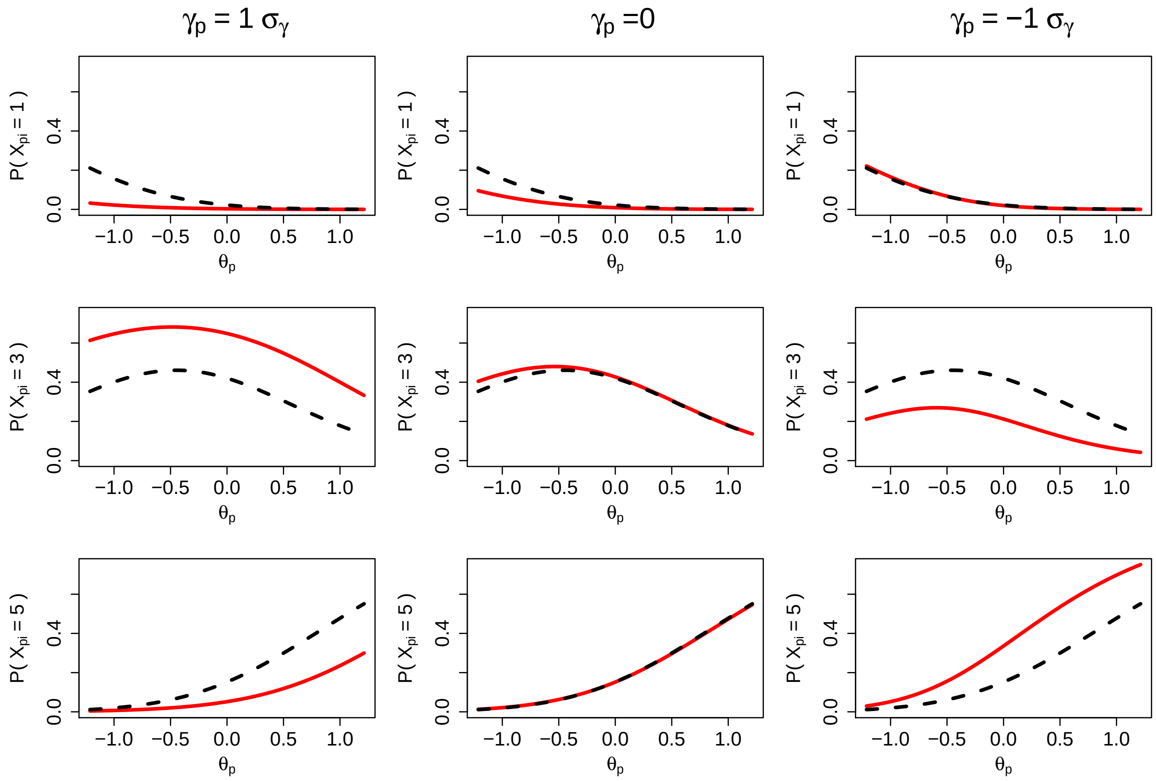

FIGURE 10: *Response curves for item "Eyes well up with tears" (FCC) along person parameter θ (between 10% and 90% quantile) for category 1 (upper panel), category 3 (middle panel) and category 5 (lower panel). Columns represent different response styles: tendency to the middle (left), no response style (middle) tendency to the extremes (right). Solid (red) lines represent estimates for PCMRS, dashed (black) lines represent estimates for PCM.*

data of the standardization sample contains data on 2070 participants (2032 complete cases). Each of the 71 items is measured on a 5-point response scale that refers to the frequency of the complaint: "never", "about 2 times a year", "about 2 times a month", "approximately 3 times a week" or "almost every day".

6.1 Emotional Reactivity

First, the PCMRS and, for comparison, also the simple PCM is applied to the items referring to the scale *emotional reactivity*. When fitting a simple PCM one obtains $\hat{\sigma}^2_\theta = 0.785$, for the extended PCMRS one obtains the covariance matrix:

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} \hat{\sigma}^2_\theta & c\hat{o}v_{\gamma\theta} \\ c\hat{o}v_{\gamma\theta} & \hat{\sigma}^2_\gamma \end{pmatrix} = \begin{pmatrix} 0.896 & 0.260 \\ 0.260 & 0.559 \end{pmatrix}.$$

The estimated standard deviation of the response style parameters ($\hat{\sigma}_\gamma = 0.747$) is quite high and, therefore, response styles are definitely present. Figure 9 shows the estimates for the item parameters, separately for each item. The solid (red) lines represent estimates for the PCMRS and the dashed (black) lines represent estimates for the PCM.

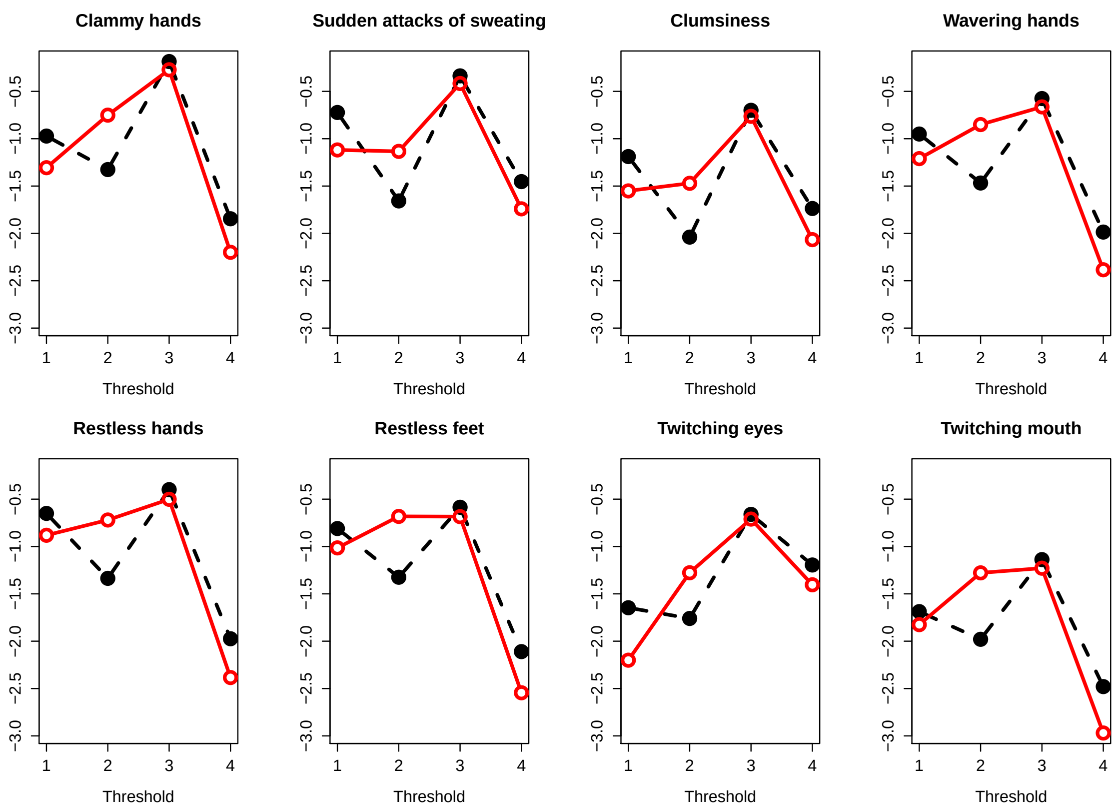

FIGURE 11: *Estimates of item parameters for items on tenseness (FCC), separately for each item. Solid (red) line represents estimates for PCMRS, dashed (black) line represents estimates for PCM.*

It can be seen that the estimates for the first category differ strongly while the estimates for the other categories mostly coincide. Except for the item "Urge to defecate in excitement", all items show ascending category-specific parameters.

To further illustrate the differences between both models due to the presence of the response style parameters in the PCMRS, Figure 10 shows the item response curves for the second item "Eyes well up with tears" for categories 1, 3 and 5 (rows). In analogy to Figure 3 the curves are plotted along different values of the person parameters, which are chosen between the 10% and the 90% quantile of the person parameters according to its estimated uni-dimensional normal distribution $\hat{\theta} \sim N(0, 0.896)$. The middle panel represents persons without response style ($\gamma_p = 0$), the left panel represents persons with a tendency to the middle ($\gamma_p = \sigma_\gamma$) and the right panel represents persons with a tendency to the extremes ($\gamma_p = -\sigma_\gamma$).

For $\gamma_p = 0$, small differences only show up for lower values of the person parameter θ_p. However huge differences appear for $\gamma_p = -\hat{\sigma}_\gamma$ and $\gamma_p = \hat{\sigma}_\gamma$. While for $\gamma_p = -\hat{\sigma}_\gamma$ the highest category becomes quite dominant, for $\gamma_p = \hat{\sigma}_\gamma$ the middle category is the most probable category along the whole range of θ-values.

6.2 Tenseness

In a second analysis, the items corresponding to the scale tenseness are considered. The estimated variance of the person parameters in the PCM is $\hat{\sigma}^2 = 0.546$ while the estimated covariance matrix between person and response style parameters for the

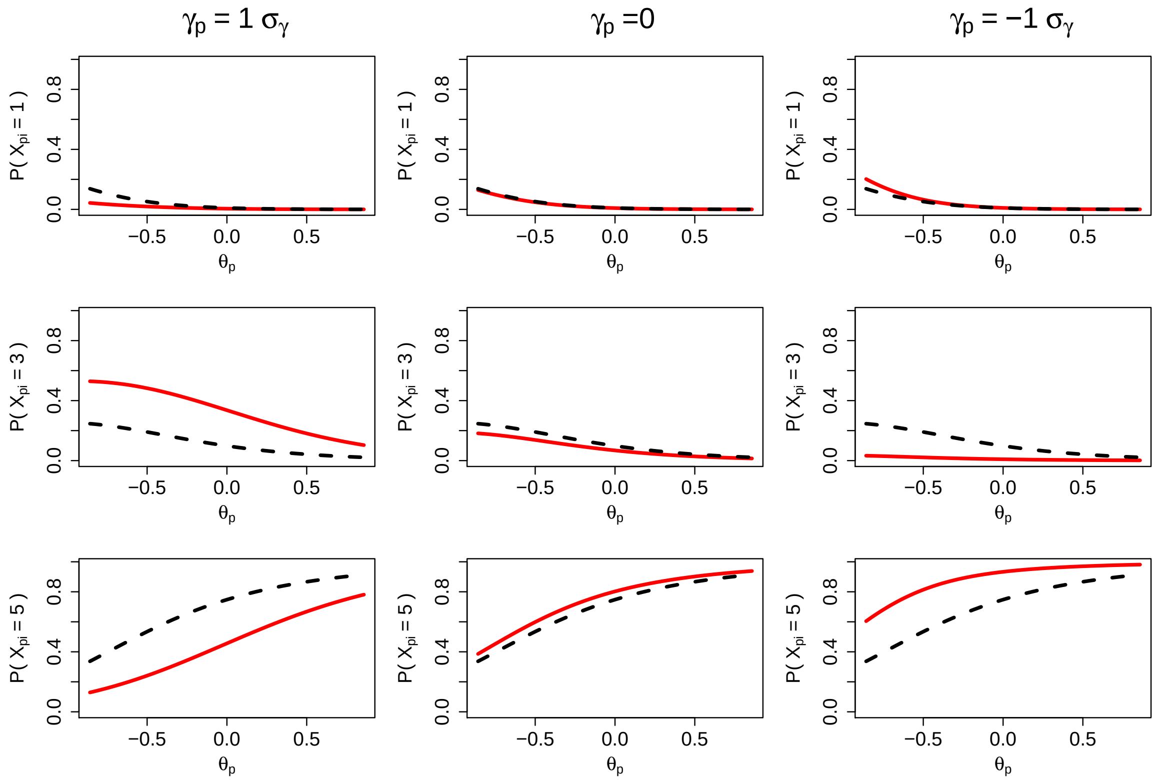

FIGURE 12: *Response curves for item "Clammy hands" (FCC) along person parameter θ (between 10% and 90% quantile) for category 1 (upper panel), category 4 (middle panel) and category 7 (lower panel). Columns represent different response styles: tendency to the middle (left), no response style (middle) tendency to the extremes (right). Solid (red) lines represent estimates for PCMRS, dashed (black) lines represent estimates for PCM.*

PCMRS is

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} \hat{\sigma}_\theta^2 & c\hat{o}v_{\gamma\theta} \\ c\hat{o}v_{\gamma\theta} & \hat{\sigma}_\gamma^2 \end{pmatrix} = \begin{pmatrix} 0.449 & 0.263 \\ 0.263 & 1.172 \end{pmatrix}.$$

In contrast to the previous example, the variance (i.e. magnitude) of the response style effects is much higher now. Figure 11 shows the estimates for the item parameters. Again, the solid (red) lines represent estimates for the PCMRS and the dashed (black) lines represent estimates for the PCM.

In contrast to Figure 9, here the parameters are not ascending along the categories. In fact, the parameters for the PCM all have the same tendency with category 2 lower than 1, category 3 higher than 2 and category 4 lower than category 3. The estimates for the PCMRS are more stable across the categories.

Looking into a single item now gives also somewhat different results than for the previous example. Figure 12 shows the response curves, exemplary for the item "Clammy hands", for three different response style parameters (columns) along the person parameter θ_p. Overall, the last category has very high probabilities, whereas the first category has very low probabilities. In particular, for a person with a rather high tendency to the extremes ($\gamma_p = -\hat{\sigma}_\gamma$), the probability for the last (fifth) category is above 0.6 throughout the whole range of person parameters.

7 Concluding Remarks

We explicitly considered a partial credit model that accounts for response styles. However, the basic concept can also be used to model response styles in alternative ordinal latent trait models like the graded response model (Samejima, 1997). The only complication in graded response models is that it is based on a threshold concept with thresholds that have to be ordered. The ordering has to hold also after incorporating response style parameters.

The proposed PCMRS model has several advantages, in particular in comparison to mixture models. Mixture models assume that respondents come from different latent classes. Different item response models are fitted within different classes, some may represent the substantive trait, some may represent response style behaviour. One of the problems with mixture models is always the number of classes, which is unknown. Typically one gets quite different models if one fits, for example, two or three classes, since all the parameters change when considering one more class. If one has chosen a number of classes it is still difficult to interpret the difference between classes and explain what feature is represented by a class, it might be a response style or some other dimension that is involved when responding to items. Since the classes are not pre-specified, for example, by explicitly modelling response style behaviour, there is much uncertainty involved and the interpretation of the model within classes often tends to be vague. In contrast, in the PCMRS model the explicit modelling of the response style allows to decide if it is present, and if, how strong it is.

The other problem with mixture models is that it is viewed as a discrete trait, respondents are classified into multiple classes representing different response behaviour (Bolt and Johnson, 2009). If one considers response style as a continuous trait, as is quite common in parts of the literature, alternative models have to be used. The models proposed by Bolt and Johnson (2009) and Bolt and Newton (2011) are constructed as continuous response style models as is the PCMRS. Instead of using a mixture they use a multi-trait model. However, similar problems as in mixture models arise. One has to decide how many latent dimensions are needed and how the results have to be interpreted. Moreover, the basic model is a model for nominal categories.

One final remark concerns the use of partial credit models in achievement tests. The partial credit model is well suited for attitude test. Nevertheless, it is also used in achievement tests. The original motivation was based on the consideration of steps in the process of solving an item (Masters, 1982). However, the step metaphor is not an appropriate description of the model, see Verhelst and Verstralen (2008) and, more recently, Andrich (2015), it has not been used in later representations of the model (Masters and Wright, 1997). If one wants to model steps an appropriate model is the sequential model or step model outlined in Tutz (1989) and Verhelst et al. (1997).

In attitude tests with categories that are not symmetric the response style considered here reflects a tendency of persons to middle or extreme categories, which can be considered as a personality trait. If the ordinal response is used in achievement tests the interpretation of the γ_p-parameters is slightly different. If it is present it may represent a way of working on items. While θ_p represents the ability to solve items, γ_p reflects the way items are solved. Two persons may have the same ability but differ in the way they have obtained it. If γ_p is large (tendency to the middle) the person has obtained its overall score by working on many items with average success, if γ_p is small (tendency to extreme categories) the person may have obtained it by solving some of the items with maximal score and others not at all or very low score. Thus, for achievement tests it is more a working style than a response style.

The proposed method has been implemented in R (R Core Team, 2015), it is available on request from the authors and will be available from CRAN soon. For the joint normal distribution of the person parameters and the response style parameters two-dimensional Gauss-Hermite integration is used. For faster performance it is (in a parallel manner) implemented in C++ and integrated into R by using the package Rcpp (Eddelbuettel et al., 2011; Eddelbuettel, 2013). Optimization of the marginal likelihood is done numerically by using the algorithm L-BFGS-B, see Byrd et al. (1995).

References

Anderson, D. A. and M. Aitkin (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society Series B 47*(2), 203–210.

Andrich, D. (2015). The problem with the step metaphor for polytomous models for ordinal assessments. *Educational Measurement: Issues and Practice 34*(2), 8–14.

Baumgartner, H. and J.-B. E. Steenkamp (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research 38*(2), 143–156.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika 37*(1), 29–51.

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

Bolt, D. M. and T. R. Johnson (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*.

Bolt, D. M. and J. R. Newton (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement 71*(5), 814–833.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association 88*(421), 9–25.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing 16*(5), 1190–1208.

De Boeck, P. and I. Partchev (2012). Irtrees: Tree-based item response models of the glmm family. *Journal of Statistical Software 48*(1), 1–28.

Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. Springer.

Eddelbuettel, D., R. François, J. Allaire, J. Chambers, D. Bates, and K. Ushey (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software 40*(8), 1–18.

Eid, M. and M. Rauber (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment 16*(1), 20.

Fahrenberg, J. (2010). Freiburg Complaint Checklist [Freiburger Beschwerdenliste (FBL)]. *Göttingen, Hogrefe*.

Fahrmeir, L. and G. Tutz (1997). *Multivariate Statistical Modelling based on Generalized Linear Models.* New York: Springer-Verlag.

Gollwitzer, M., M. Eid, and R. Jürgensen (2005). Response styles in the assessment of anger expression. *Psychological assessment 17*(1), 56.

Hinde, J. (1982). Compound poisson regression models. In R. Gilchrist (Ed.), *GLIM 1982 International Conference on Generalized Linear Models*, pp. 109–121. New York: Springer-Verlag.

Irtel, H. (1995). An extension of the concept of specific objectivity. *Psychometrika 60*(1), 115–118.

Jeon, M. and P. De Boeck (2015). A generalized item response tree model for psychological assessments. *Behavior Research Methods, published online*. doi: 10.3758/s13428-015-0631-y.

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.

Kankaraš, M. and G. Moors (2009). Measurement equivalence in solidarity attitudes in europe insights from a multiple-group latent-class factor approach. *International Sociology 24*(4), 557–579.

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

Maij-de Meij, A. M., H. Kelderman, and H. van der Flier (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement 32*(8), 611–631.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika 47*(2), 149–174.

Masters, G. N. and B. Wright (1984). The essential process in a family of measurement models. *Psychometrika 49*(4), 529–544.

Masters, G. N. and B. D. Wright (1997). The partial credit model. In *Handbook of modern item response theory*, pp. 101–121. Springer.

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

Messick, S. (1991). Psychology and methodology of response styles. *Improving inquiry in social science: A volume in honor of Lee J. Cronbach*, 161–200.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. a multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review 20*(4), 303–320.

Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research 22*(1), 93–119.

Plieninger, H. and T. Meiser (2014). Validity of multiprocess irt models for separating content and response styles. *Educational and Psychological Measurement 74*(5), 875–899.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British journal of mathematical and statistical psychology 19*(1), 49–57.

Rasch, G. (1977). On specific objectivity. an attempt at formalizing the request for generality and validity of scientific statements in symposium on scientific objectivity. *Danish Year-Book of Philosophy Kobenhavn 14*, 58–94.

Samejima, F. (1997). Graded response model. *Handbook of modern item response theory*, 85–100.

Suh, Y. and D. M. Bolt (2010). Nested logit models for multiple-choice item response data. *Psychometrika 75*(3), 454–473.

Thissen-Roe, A. and D. Thissen (2013). A two-decision model for responses to likert-type items. *Journal of Educational and Behavioral Statistics 38*(5), 522–547.

Tutz, G. (1989). Sequential item response models with an ordered response. *British Journal of Statistical and Mathematical Psychology 43*(1), 39–55.

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

Tutz, G. and M. Berger (2016). Response styles in rating scales - simultaneous modelling of content-related effects and the tendency to middle or extreme categories. *Journal of Educational and Behavioral Statistics 41*(3), 239–268.

Van Rosmalen, J., H. Van Herk, and P. Groenen (2010). Identifying response styles: A latent-class bilinear multinomial logit model. *Journal of Marketing Research 47*(1), 157–172.

Van Vaerenbergh, Y. and T. D. Thomas (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research 25*(2), 195–217.

Verhelst, N. D., C. Glas, and H. De Vries (1997). A steps model to analyze partial credit. In *Handbook of modern item response theory*, pp. 123–138. Springer.

Verhelst, N. D. and H. Verstralen (2008). Some considerations on the partial credit model. *Psicológica: Revista de metodología y psicología experimental 29*(2), 229–254.

Wolfinger, R. W. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika 80*(4), 791–795.

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.

ZPID (2013). PsychData of the Leibniz Institute for Psychology Information ZPID. *Trier: Center for Research Data in Psychology*.