LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK

Gerhard Tutz and Clemens Draxler

A Common Framework for Classical and Tree-Based Item Response Models Including Extended Hierarchically Structured Models

Technical Report Number 227, 2019
Department of Statistics
University of Munich

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

A Common Framework for Classical and Tree-Based Item Response Models Including Extended Hierarchically Structured Models

Gerhard Tutz
Ludwig-Maximilians-Universität München
Akademiestraße 1, 80799 München
Clemens Draxler
UMIT – The Health and Life Sciences University
Eduard-Wallnöfer-Zentrum 1, 6060 Hall in Tirol

October 30, 2019

Abstract

A common framework is provided that comprises classical ordinal item response models as the cumulative, sequential and adjacent categories models as well as the more recently propagated item response tree models. The obtained taxonomy is based on the role that binary models play as building blocks of the various models. The study of the binary models contained in ordinal latent trait models clarifies the interpretation of item parameters in classical models. The taxonomy for ordinal models also contains a new general class of hierarchically structured models, which can be seen as a generalization of item response tree models. For this class of models estimation methods are developed, which make use of commonly available program packages.

Keywords: Ordered responses, latent trait models, item response theory, graded response model, partial credit model, sequential model, Rasch model, item response trees

1 Introduction

Various latent trait models for ordered response data have been proposed in the literature, for an overview see, for example, Van der Linden (2016). One can in particular distinguish between three basic types of models, cumulative models, sequential models and adjacent categories models. One of the objectives of the present paper is to show how these models are easily built from binary latent trait models. The way how the binary models are used to construct models helps to understand the structure of the models and to clarify the meaning of the parameters. It also provides a framework that allows to embed more recently developed ordinal item response models as the tree-based models, yielding a general taxonomy of ordinal item response models.

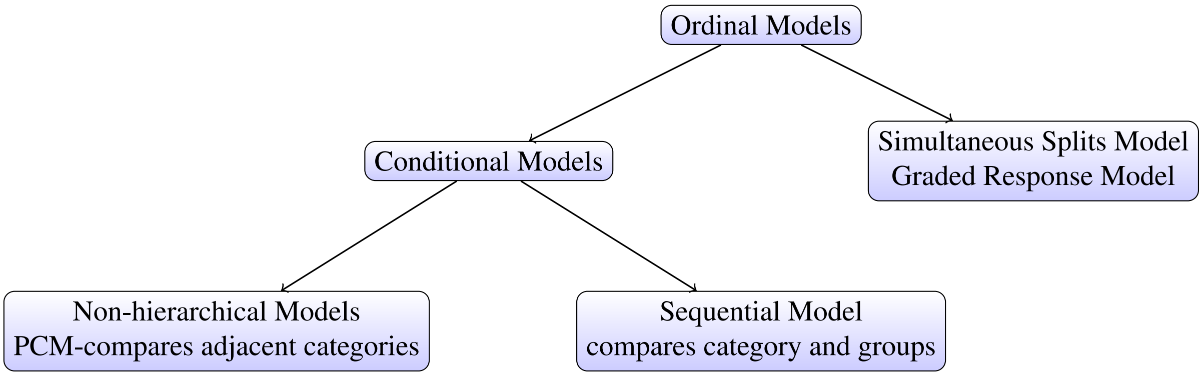

FIGURE 1: *Structure of classical ordinal latent trait models.*

Binary models for person p and item i have the common form

$$P(Y_{pi} = 1) = F(\theta_p - \delta_i), \quad (1)$$

where $F(.)$ is a cumulative distribution function, θ_p is the person parameter, and δ_i is the item parameter, typically a difficulty or threshold. An important member of this class of models is the Rasch model, which is obtained if $F(.)$ is the logistic distribution function $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$. It is straightforward to include an item discrimination parameter by using $\alpha_i(\theta_p - \delta_i)$ instead of $\theta_p - \delta_i$. For simplicity we will mostly use the simple form without an item discrimination parameter.

Given one has a response in *ordered* categories $\{0, 1, \ldots, k\}$ there are several ways to construct an ordinal model from binary models of the form (1). The binary models can be used to compare specific categories or groups of categories from $\{0, 1, \ldots, k\}$. One can, in particular,

- compare groups of categories that result from splitting the categories into the subsets $\{0, 1, \ldots, r - 1\}$ and $\{r, \ldots, k\}$,
- compare (conditionally) between two categories, for example, adjacent categories,
- compare (conditionally) between a category and a set of adjacent categories, for example, $\{r - 1\}$ and $\{r, \ldots, k\}$.

The different ways to compare categories correspond to cumulative models, adjacent categories and sequential models in that order. The way how binary models are used yields a taxonomy of classical ordered latent trait models. As a preview Figure 1 shows that the consideration of the binary models contained in ordinal models yields a distinction between conditional and non-conditional models. The cumulative or graded response model is the only non-conditional model from this class of models. The other two use some sort of conditioning on responses when considering the binary building blocks within the models.

In the second part of the paper we consider tree-based model, which have been propagated more recently, for example, by De Boeck and Partchev (2012), Böckenholt (2017). Tree based models use the binary choices conditionally in a hierarchical way. We extend existing approaches to find parsimonious models that efficiently use the information in the ordering of categories.

The characterization of ordinal models by the binary models that are contained as building blocks yields a general taxonomy of ordinal item response models that includes tree-based models. The obtained hierarchical structure differs from the taxonomy proposed by Thissen and Steinberg (1986). Their classification into "divide-by-total" models and "difference" models focuses on the form of the response probabilities. They do not distinguish between conditional and un-conditional models, a distinction that arises quite naturally if one focuses on the binary models as the elements that are behind complex item response models. An additional advantage of the proposed structure is that it clarifies the interpretation of parameters in complex models.

2 Classical Ordered Response Models

In the following let $Y_{pi} \in \{0, 1, \ldots, k\}$, $p = 1, \ldots, P$, $i = 1, \ldots, I$, denote the ordinal response of person p on item i. An important partition of the response categories is the partition into the subsets $\{0, \ldots, r-1\}$ and $\{r, \ldots, k\}$, which can be represented by the binary variable

$$Y_{pi}^{(r)} = \begin{cases} 1 & Y_{pi} \geq r \\ 0 & Y_{pi} < r. \end{cases} \tag{2}$$

The variables $Y_{pi}^{(1)}, \ldots, Y_{pi}^{(k)}$ are called *split variables* because they split the response categories into two subsets. They play a major role in the construction of the traditional ordered latent trait models.

2.1 Simultaneous Modelling of Splits: The Graded Response Model

Let us assume that the response categories represent levels of performance in an achievement test. Then one can consider two groups of categories, $\{0, 1, \ldots, r-1\}$ for low performance and $\{r, \ldots, k\}$ for high performance, where low and high are relative terms that refer to "below category r" and "above or in category r". One might assume that the split into low and high performance is determined by a binary model with person ability θ_p and a threshold that depends on the category at which the categories have been split. Therefore, one assumes

$$P(Y_{pi}^{(r)} = 1) = F(\theta_p - \delta_{ir},) \quad, k = 1, \ldots, k. \tag{3}$$

Thus, for each dichotomization into categories $\{0, 1, \ldots, r-1\}$ and $\{r, \ldots, k\}$ a binary model is assumed to hold. Importantly, the models are assumed to hold *simultaneously* with the same person ability θ_p but different item difficulties δ_{ir}. Simple rewriting yields the *cumulative model*

$$P(Y_{pi} \geq r) = F(\theta_p - \delta_{ir}), \quad k = 1, \ldots, k, \tag{4}$$

which is equivalent to a version of Samejima's *graded response model* (Samejima, 1995, 2016). Thus, the graded response model can be seen as a model for which the dichotomizations into the categories $Y_{pi} < r$ and $Y_{pi} \geq r$ are *simultaneously* modeled. One consequence is that item difficulties are ordered. Since $P(Y_{pi} = r) = P(Y_{pi} \geq r) - P(Y_{pi} \geq r+1) = F(\theta_p - \delta_{ir}) - F(\theta_p - \delta_{i,r+1}) \geq 0$, one obtains that $\delta_{ir} \leq \delta_{i,r+1}$has to hold for all categories.

The strong link between the binary responses and the ordinal response yields a specific view of the graded response model that differs from traditional ones. In an achievement test the sequence of binary responses $(Y_{pi}^{(1)}, \ldots, Y_{pi}^{(k)})$ can be seen as referring to tasks with increasing difficulties. More concrete, because item difficulties are ordered, one has $P(Y_{pi}^{(r)} = 1) \geq P(Y_{pi}^{(r+1)} = 1)$, which means the "task" represented by $Y_{pi}^{(r)}$ is simpler than the "task" $Y_{pi}^{(r+1)}$. Moreover, if the task $Y_{pi}^{(r)}$ was completed ($Y_{pi}^{(r)} = 1$ or, equivalently, $Y_{pi} \geq r$), the preceding tasks $Y_{pi}^{(s)}, s < r$ ($Y_{pi}^{(s)} = 1$ or, equivalently, $Y_{pi} \geq s$) were also completed. Therefore, the outcome of the sequence of binary variables has the specific form

$$(Y_{pi}^{(1)} \ldots, Y_{pi}^{(k)}) = (1, \ldots, 1, 0, \ldots, 0),$$

which means a sequence of ones is followed by a sequence of zeros. Binary variables that follow this pattern have been called *Guttman variables* and the resulting response space is usually referred to as Guttman space (Andrich, 2013). The increasing sequence of difficulties $\delta_{i1} \leq \cdots \leq \delta_{ik}$ may also be seen as thresholds that have to be exceeded to obtain a higher level of performance. $Y_{pi}^{(r)} = 1$ means that threshold δ_{ir} has been exceeded. A nice feature is that the ordinal response is simply given as the sum of the binary variables

$$Y_{pi} = Y_{pi}^{(1)} + \cdots + Y_{pi}^{(k)},$$

which means the ordinal response is the number of thresholds that have been exceeded. It is also the number of tasks that have been successfully completed.

The interpretation of item parameters as thresholds is also supported by an alternative derivation of the cumulative model. Let $\tilde{Y}_{pi} = \theta_p + \varepsilon_{pi}$, where ε_{pi} is a noise variable with symmetric continuous distribution function $F(.)$, denote a latent variable that is invoked if person p tries to solve item i. Y_{pi} is essentially the ability of the person plus a noise variable and can be seen as the random ability of the person. The category boundaries approach assumes that category r is observed if the latent variable is between thresholds δ_{ir} and $\delta_{i,r+1}$. More formally, one has $Y_{pi} = r \quad \Leftrightarrow \quad \delta_{ir} \leq \tilde{Y}_{pi} < \delta_{i,r+1}$. It is immediately seen that one obtains the cumulative model and thresholds have to be ordered.

It is noteworthy that the derivation of the cumulative model by Samejima (1995, 2016) is different from the derivations given here. Samejima considers steps in the problem solving process. It is assumed that a graded item score r is assigned to an examinee who successfully completes up to step r but fails to complete step $r+1$. The conceptualization is very similar to that of the sequential model to be considered later. Contrary to this motivation Andrich (2015) states that there is no concept of steps in the cumulative model. It is indeed hard to see why the dichotomization specified in the model representations (9) and (4) should be linked to steps. Certainly the variables $Y_{pi}^{(r)}$ should not be seen as steps. $Y_{pi}^{(r)} = 1$ simply denotes that a person has at least performance level r. Since performance levels are ordered, that means, its performance cannot be below level r, or, in split variables, $Y_{pi}^{(1)} = \cdots = Y_{pi}^{(r-1)} = 1$, which is the Guttman property of the binary responses. One observes $Y_{pi} = r$, if, in addition $Y_{pi}^{(r+1)} = 0$, which means that the performance is below level r. However, no steps or transitions are needed to explain the level of performance. As Andrich (2015) argues, if a performance like acting is to be classified according to some protocol, the judge places the person's performance in one of the categories on the trait, not how the person transitioned in getting to the category. Moreover, even in simple binary models for problem solving one observes if the problem was solved or not, but not the transition. Thus, when considering ordinal models and the binary models contained in them there is no reason to construct a transition. It might be misleading and is not compatible with the underlying process, which is determined by *simultaneous* dichotomizations or the placing on the continuum of the latent scales, which is divided by the thresholds $\delta_{i1} \leq \cdots \leq \delta_{i,k}$.

Thissen and Steinberg (1986) called the graded response models "difference" models because the probabilities are given as differences, $P(Y_{pi} = r) = F(\theta_p - \delta_{ir}) - F(\theta_p - \delta_{i,r+1})$. Although they also start with binary models they do not further investigate that the models have to hold simultaneously.

2.2 Conditional Comparison of Categories: the Partial Credit and Other Adjacent Categories Models

Rather than compare groups of categories by utilizing a binary model one can also compare two categories from the set of categories $\{0, 1, \dots, k\}$. A choice that suggests itself are adjacent categories. Let the binary models that compare two adjacent categories be given by

$$P(Y_{pi} = r | Y_{pi} \in \{r-1, r\}) = F(\theta_p - \delta_{ir}), \quad r = 1, \dots, k. \tag{5}$$

Again all the models contain the same person parameter but model-specific item parameters. It is straightforward to derive that for the logistic distribution function one obtains the *partial credit model*

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

which was propagated by Masters (1982) and Masters and Wright (1984). It is also equivalent to the *polytomous Rasch model*, which is just a different parameterization, see, for example, Andrich (2010). Thissen and Steinberg (1986) called the graded response model a "divide-by-total" model because of the denominator in the probabilities.

If one uses binary models as building blocks, the question arises why one should confine oneself to adjacent categories, although they seem a natural choice. An alternative would be to use binary models for a set of pairs of categories $(s_1, r_1), \dots, (s_k, r_k)$, $s_i < r_i$. It can be shown that postulating binary models for pairs of categories also yields the partial credit model. Therefore, the partial credit model can be seen as a general model for pairs of categories.

An alternative form of the model, which emphasizes the implicit comparison of categories is

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)}\right) = \theta_p - \delta_{ir}, \quad r = 1, \dots, k. \tag{6}$$

That means, the PCM directly compares two adjacent categories, and θ_p determines the strength of the preference for the higher category.

It should be emphasized that the binary models used as building blocks are *conditional* models, it is assumed that a binary model holds *given the response is in two categories from the set of available categories*. This is seen from the representation (5) but hidden in the representation (6). However, it has consequences for the interpretation of parameters. The item parameters represent thresholds *given* the response is in categories $\{r-1, r\}$ and the trait parameters are the abilities to score r rather than $r-1$ *given* the response is in categories $\{r-1, r\}$. Therefore, the parameters refer to a local conditional decision or preference although changing the item parameter changes the probabilities of all possible outcome values since the PCM assumes that the binary models hold simultaneously. Nevertheless the binary models are conditional models and parameters should be interpreted with reference to the conditional structure. One consequence of the conditional parameterization is that thresholds do not have to be ordered though there has been some discussion on the ordering of thresholds, see, for example, Adams et al. (2012), Andrich (2013), Andrich (2015).

The class of adjacent categories model also contains simplified versions that use a sparser parameterization. By assuming that the item parameters can be decomposed into two terms in the form $\delta_{il} = \delta_i + \tau_l$, one obtains the Rasch rating scale model (RSM), see Andrich (1978), Andrich (2016).

2.3 Conditional Comparison of a Single Category and a Group of Categories: Sequential Models

In achievement tests frequently items are used that are solved in consecutive observed steps. For example, a mathematical problem may have the form: $(\sqrt{49}-9)^3 =?$. One can distinguish four levels: no problem solved (level 0), $\sqrt{49} = 7$ solved (level 1), $7-9 = -2$ solved (level 2), $(-2)^3 = -8$ solved (level 3). Obviously the sub problems have to be solved in a consecutive way. A sub problem can only be solved if the all the previous sub problems have been solved. A model that explicitly models the solving of sub problems has the form

$$P(Y_{pi} \geq r | Y_{pi} \geq r-1) = F(\theta_p - \delta_{ir}), \quad r = 1, \ldots, k. \tag{7}$$

The model is known as *sequential model* (Tutz, 1990) or *step model* (Verhelst et al., 1997). It is a process model for consecutive steps. One models the transition to higher categories given the previous step was successful. The first step is the only non-conditional step. If it fails, the response is in category 0 (first sub problem not solved), if it is successful, the response is larger than 0 (first sub problem solved). In the latter case the person tries to take the second step. If it is not successful, the response is in category 1 (second sub problem not solved), if it is successful, the response is larger than 1 (second sub problem solved), etc. In the r-th step it is distinguished between $Y_{pi} = r-1$ and $Y_{pi} \geq r$ *given* at least level $r-1$ is reached ($Y_{pi} \geq r-1$). In the model the parameter θ_p represents the person's ability to successfully perform each of the steps while δ_{ir} is the difficulty in step r. Of course, later steps can be easier than early steps, thus item difficulties are not necessarily ordered. In the example step 2 ($7-9$) is certainly easier to master than step 1 ($\sqrt{49} = 7$). However, sub problem 2 can be only solved after step 1 was successful. Therefore, the item parameters have

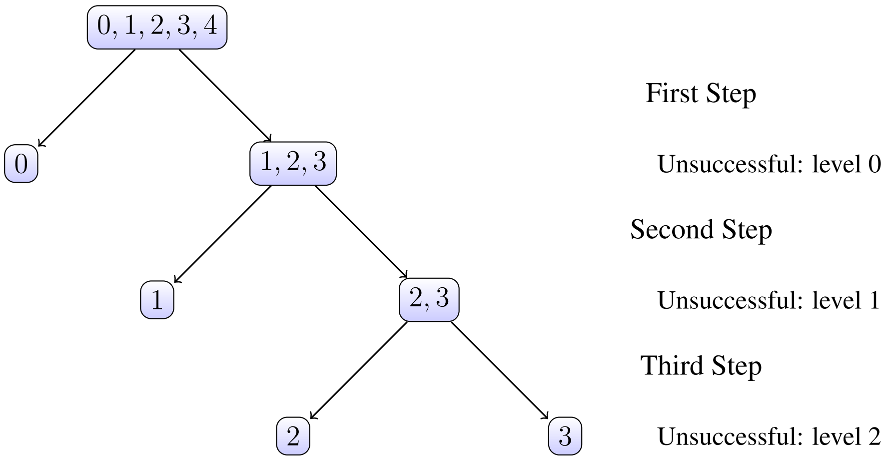

FIGURE 2: *The sequential model as a hierarchicaly structured model.*

local meaning, they refer to the difficulty in a step given that all previous steps were successful. In contrast, the same ability parameter is present in each of the steps, which makes the model uni-dimensional in terms of person parameters.

The logistic version of the model, also called logistic sequential model, can be given in the alternative form of a *continuation ratio model*,

$$\log\left(\frac{P(Y_{pi} \geq r)}{P(Y_{pi} = r-1)}\right) = \theta_p - \delta_{ir}, \quad r = 1, \dots, k. \tag{8}$$

The logits on the left hand side compare the categories the probability of a response in the categories $\{r, \dots, k\}$ to the probability of a response in category $\{r-1\}$. In this sense the binary models contained in the sequential model compare groups of categories to single categories. This comparison is also seen from the tree representation of the model given in Figure 2, which shows a sequential model with four categories. It shows the sequence of (conditional) binary splits. In the r-th step a decision between category $\{r-1\}$ and categories $\{r, \dots, k\}$ is obtained. The split is conditional, given categories $\{r-1, \dots, k\}$, that means, under the condition that the previous step was successful.

A disadvantage of the model representation (8) is that it does not directly show the underlying process. The implicit conditioning on responses $Y_{pi} \geq r$, which is essential for the interpretation of the model parameters, gets lost. It is however seen in the model representation with *split variables* given by

$$P(Y_{pi}^{(r)} = 1 | Y_{pi}^{(r-1)} = 0, \dots, Y_{pi}^{(1)} = 0) = F(\theta_p - \delta_{ir}) \quad , r = 1, \dots, k, \tag{9}$$

which again shows that the split variables $(Y_{pi}^{(1)} \dots, Y_{pi}^{(k)})$ form a Guttman space.

2.4 Overview on Classical Ordinal Models

It has been shown that all the models contain binary models that split categories into two subsets. In the partial credit model and the sequential model the splits are condi-

TABLE 1: *Overview of traditional ordinal models.*

	Category Representation Logistic Version $\log(.) = \theta_p - \delta_{ir}$	Conditional Representation General Version $P(.) = F(\theta_p - \delta_{ir})$	Conditional Representation With split variables $P(.) = F(\theta_p - \delta_{ir})$
Cumulative	$\log\left(\frac{P(Y_{pi}\geq r)}{P(Y_{pi}<r)}\right)$	$P(Y_{pi} \geq r)$	$P(Y_{pi}^{(r)} = 1)$
Partial Credit	$\log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}\right)$	$P(Y_{pi} = r \mid Y_{pi} \in \{r-1, r\})$	$P(Y_{pi}^{(r)} = 1 \mid Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0)$
Sequential	$\log\left(\frac{P(Y_{pi}\geq r)}{P(Y_{pi}=r-1)}\right)$	$P(Y_{pi} \geq r \mid Y_{pi} \geq r-1)$	$P(Y_{pi}^{(r)} = 1 \mid Y_{pi}^{(r-1)}) = 0$

tional whereas in the cumulative model the splits are simultaneous but not conditional. Figure 1 visualizes the hierarchy of models.

In Table 1 the models are given in various representations. The left column shows the logistic versions of the models. It shows which categories or groups of categories are compared. In particular it is seen which type of logits are determined by the difference between person parameter and item parameter, $\theta_p - \delta_{ir}$. For example, in the partial credit model one has the adjacent categories logits $\log(P(Y_{pi} = r)/P(Y_{pi} = r-1))$, in the sequential model one has the continuation ratios $\log(P(Y_{pi} \geq r)/P(Y_{pi} = r-1))$. In the middle column the general conditional representations of the models are given. In these representations the distribution function $F(.)$ can be any strictly monotonic distribution function. It shows which *conditional* binary response models are contained in the ordinal model. In the case of the graded response model the condition is empty since it is a non-conditional model. The right column shows the representation of the general models with split variables. It also shows clearly the conditioning implicitly contained in the models. It should be noted that in all the models the split variables $(Y_{pi}^{(1)} \ldots, Y_{pi}^{(k)})$ form a Guttman space and the ordinal response is given by $Y_{pi} = Y_{pi}^{(1)} + \cdots + Y_{pi}^{(k)}$.

3 Hierarchically Structured Modeling: Tree-Based Models

The classical models considered in the previous section represent different types of modelling concerning the conditioning. While the graded response model is a model that does not rely on conditioning, the partial credit model conditions on a response in adjacent categories. The sequential model is conditional but, in contrast to the partial credit model, it can be represented as a tree (see Figure 2). This makes it a special model, it is *hierarchical*, that means, it can be represented by a sequence of conditional splits. Neither the graded response models nor the partial credit model are hierarchical. More recently with IR-Trees a whole class of hierarchical models has been introduced. IR-Trees will be considered briefly in the following and extended afterwards. For simplicity in the following the response categories are $\{1, \ldots, k\}$, which is the common notation in IR-Trees.

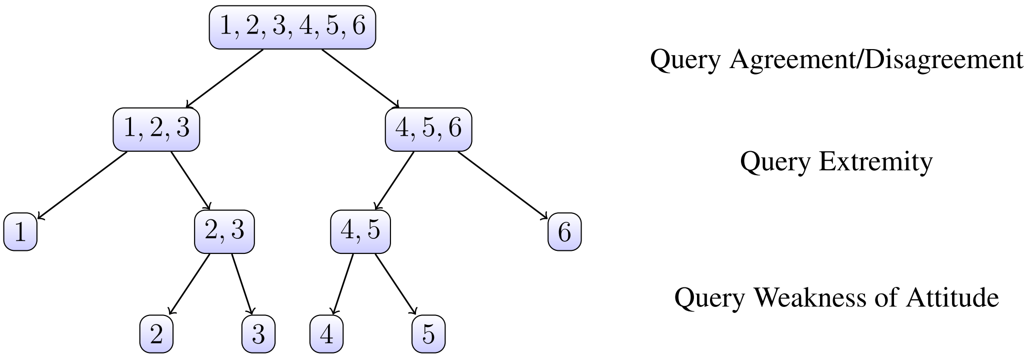

FIGURE 3: *A tree for six ordered categories, categories 1,2,3 represent levels of disagreement, categories 4,5,6 represent levels of agreement (compare Figure 3 in Böckenholt (2017)).*

3.1 IR-Tree Models

Tree-based models assume a nested structure with the building blocks given as binary models. They were considered by De Boeck and Partchev (2012), Böckenholt (2012), Khorramdel and von Davier (2014), Böckenholt (2017) and Böckenholt and Meiser (2017). In the following we use the presentation of IR-tree models given by Böckenholt (2017). IR-tree models are sequential process model, a response is constructed based on a series of mental questions. For illustration we consider an ordinal response with six categories that represent an ordinal response from “strongly disagree” to “strongly agree”. Figure 3 shows the corresponding tree, which is equivalent to Figure 3 in Böckenholt (2017). The first query determines a respondent’s agreement or disagreement. The second query determines the extremity of the (dis)agreement and the third query assesses whether the agreement is weak or not. For each query in the tree, which corresponds to a conditional binary decision one uses a binary model. For query q the model is given by

$$P(Y_{pi}^{(q)} = 1) = F(\theta_p^{(q)} - \delta_i^{(q)}). \tag{10}$$

Although the resulting model is rather flexible and easy to estimate a disadvantage is that for each query one has a new person parameter. That makes the model multidimensional in terms of person parameters and person parameters are interpreted with reference to the specific query, that is, the conditional decision. The model seems not to efficiently use the information in the ordered categories. In particular, the basic propensity to agree or disagree is only modelled in the first query. The person parameters in the next queries refer to response styles, whether a person prefers extreme or middle categories. However, the basic propensity to agree or disagree is not present in later queries though it might also determine the choice between category groups $\{1\}$ and $\{2, 3\}$. In contrast, a model like the sequential model uses the same person parameter in all binary decisions. Also in the partial credit model, which is not hierarchical, the same person parameter is present in the conditional binary decisions. In the next section we consider trees that include the same parameters in different levels yielding more parsimonious models.

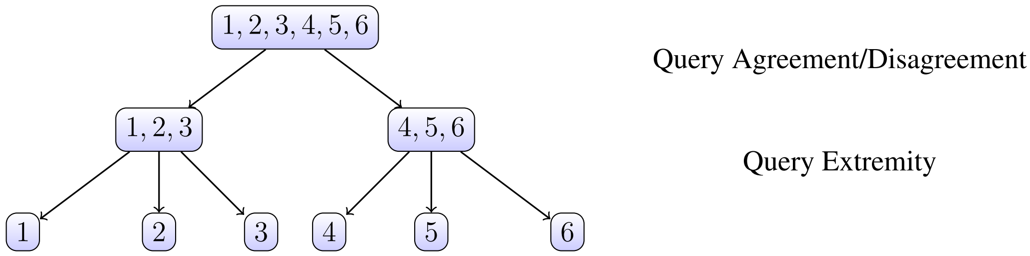

FIGURE 4: *A tree for six ordered categories with three levels .*

3.2 Hierarchical Partitioning

Let us consider the tree for six ordered categories given in Figure 4. The tree has only two levels in addition to the 0-level with all categories. One can model the propensity to agree or disagree by

$$P(Y_{pi} \geq 4) = F(\theta_p - \delta_i^{(1)}), \tag{11}$$

where $\delta_i^{(1)}$ is the level 1 item parameter. The conditional propensity to choose from one of the categories in level 2 can be specified, for example, by conditional graded response models

$$P(Y_{pi} \geq r | Y_{pi} \leq 3) = F(\theta_p - \delta_{ir}^{(2)}), r = 2, 3 \tag{12}$$

$$P(Y_{pi} \geq r | Y_{pi} \geq 4) = F(\theta_p - \delta_{ir}^{(2)}), r = 5, 6. \tag{13}$$

The model has as many parameters as a simple graded response model, however, order restrictions are weaker. One just has $\delta_{i2}^{(2)} \leq \delta_{i3}^{(2)}$ and $\delta_{i5}^{(2)} \leq \delta_{i6}^{(2)}$ whereas in the simple cumulative model five thresholds have to be ordered.

Uni-dimensionality

The two level model considered here has the advantage that it is uni-dimensional with regard to the person parameter. The same person parameter is present on each level to exploit the information in the ordered categories efficiently. That means if one compares, for example, category $\{3\}$ and categories $\{1, 2\}$ by using the odds ratio

$$\log \frac{P(Y_{pi} = 3)}{P(Y_{pi} \in \{1, 2\})} = \theta_p - \delta_{i3}^{(2)},$$

one has a linear function that increases with increasing person parameter θ_p. Therefore, also comparisons between probabilities of categories from the disagreement categories depend on the parameter θ_p. Actually, comparisons between any categories s and r are functions of θ_p, and, of course, item parameters. In contrast, in IR-tree models, which use a different item parameter in each query/split comparisons between groups depend on quite different item parameters.

Including Response Styles

A further strength of the model is that it is easily extended to include response style parameters by using in level 2 the parameterization

$$P(Y_{pi} \geq r|Y_{pi} \leq 3) = F(\theta_p + \gamma_p - \delta_{ir}^{(2)}), r = 2, 3$$
$$P(Y_{pi} \geq r|Y_{pi} \geq 4) = F(\theta_p - \gamma_p - \delta_{ir}^{(2)}), r = 5, 6.$$

The parameter γ_p is a response style parameter that contains the tendency to middle categories. Figure 5 illustrates the effect of the response style parameter. In the first row the number of response categories is four, parameters are $\delta^{(1)} = 0.5, \delta_2^{(2)} = -1.5$, $\delta_2^{(2)} = -1.5, \delta_3^{(2)} = 1.5$. In the second row the number of response categories is six. In the middle column there is no response style parameter, $\gamma_p = 0$, in the left column it is positive, $\gamma_p = 1$, in the right column negative, $\gamma_p = -1$. The effect of the response style is obvious. A person with positive value γ_p has a tendency to middle categories whereas a person with negative value γ_p has a tendency to extreme responses. In contrast to IR-tree models one keeps the attitude parameter θ_p and adds a response style parameter instead of dropping the attitude parameter and using a new response style parameter in the higher levels.

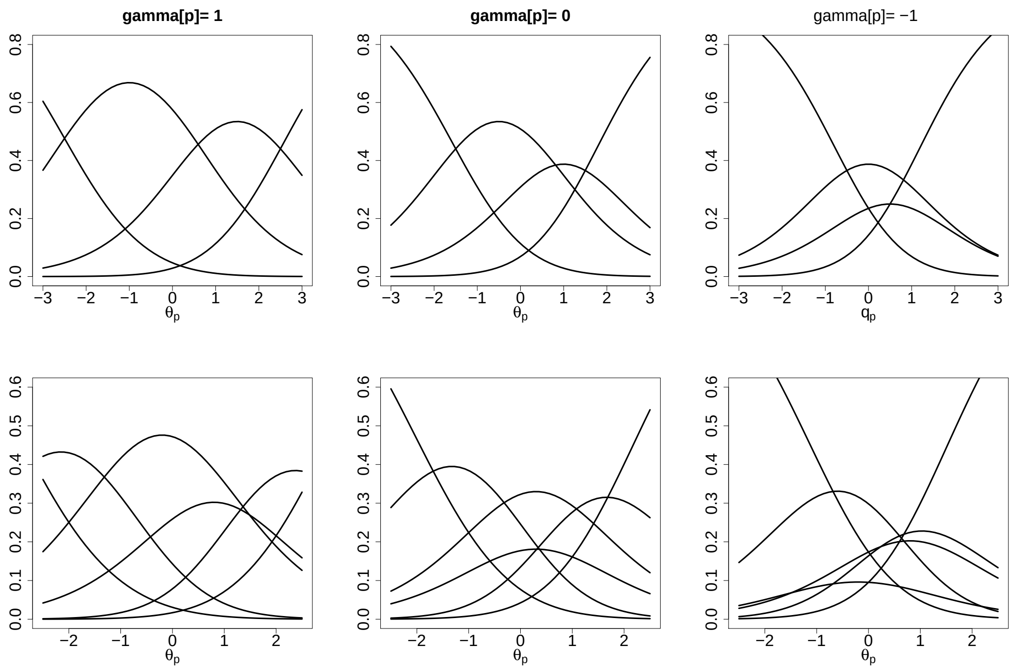

FIGURE 5: *Probabilities $P(Y_{pi} = r)$ against θ_p for positive (left), negative γ_p (right) and $\gamma_p = 0$ (middle); in the upper panels the number of categories is four, in the lower panels it is six.*

Symmetry of Response Categories

Hierarchically structured model are in particular useful in Likert items that represent a symmetric form of agreement and disagreement in the form "strongly disagree", "weakly disagree", ... "weakly agree", "strongly agree". For such symmetrical response categories it is sensible to account for the symmetry by using a more parsimonious parameterization. A person with parameter $\theta_p = \delta^{(1)}$ is "neutral", she has the same probability of choosing agreement categories or disagreement categories. For this neutral person and six response categories as in Figure 4 one also expects that $P(Y_{pi} = 1) = P(Y_{pi} = 6)$, $P(Y_{pi} = 2) = P(Y_{pi} = 5)$, $P(Y_{pi} = 3) = P(Y_{pi} = 4)$, which yields $\delta_2^{(2)} - \delta^{(1)} = -(\delta_6^{(2)} - \delta^{(1)})$, and $\delta_3^{(2)} - \delta^{(1)} = -(\delta_5^{(2)} - \delta^{(1)})$. Thus, the free item parameters are $\delta^{(1)}, \delta_2^{(2)}, \delta_3^{(2)}$.

General Ordinal Model with Two Levels

If the number of agreement and disagreement categories is even it is natural to split in the first step into categories $\{1, \ldots, k/2\}$ and $\{k/2 + 1, \ldots, k\}$ by utilizing a binary model. If the number of response categories is odd, there is a neutral category in the middle. Then it seems sensible to use a three categories model in the first step.

In general a hierarchical model is obtained by first modeling the response groups of homogenous response categories and then modeling the response within groups. Generally, let the categories $1, \ldots, k$ be subdivided into basic sets $S_1, \ldots, S_t$, where $S_i = \{m_{i-1} + 1, \ldots, m_i\}, m_0 = 0, m_t = k$. In the *first step* the response in one of the sets is determined by an ordinal model with item parameters $\delta_r^{(1)}$. In the *second step* the conditional response given S_i is determined by an ordinal model with parameters that are linked to S_i. From these two steps one obtains, if one uses the cumulative model,

$$\begin{aligned} P(Y_{pi} \in T_j) &= F(\theta_p - \delta_{ij}^{(1)}), \\ P(Y_{pi} \le r | Y \in S_j) &= F(\theta_p - \delta_{ir}^{(2)}), \end{aligned} \tag{14}$$

where $T_j = S_1 \cup \cdots \cup S_j, \delta_{i1}^{(1)} < \ldots < \delta_{i,j-1}^{(1)} < \delta_j^{(1)} = \infty, \delta_{i,m_{j-1}+1}^{(2)} < \ldots < \delta_{i,m_{j-1}}^{(2)}, j = 1, \ldots \ldots, t$.

For even categories one uses $S_1 = \{1, \ldots, k/2\}$ and $S_2 = \{k/2 + 1, \ldots, k\}$, for odd categories one uses $S_1 = \{1, \ldots, (k-1)/2\}$ and $S_2 = \{(k+1)/2\}$, $S_3 = \{(k+1)/2, \ldots, k\}$. The inclusion of response style parameters is in the same way as in the example with six categories. The presentation given here uses the cumulative model in both steps, however one may also use the adjacent categories model in both steps. Hierarchically structured models of this type were considered by Thissen-Roe and Thissen (2013) and in a regression context by Tutz (1989).

The inclusion of response style parameters in the hierarchically structured model provides an alternative to response style modeling as that for the partial credit model given by Tutz et al. (2018).

3.3 Alternative Representations of Hierarchical Partitioning Models and Parameter Estimation

Parameter estimates of a hierarchical tree model can be obtained by using ordinary program packages for multidimensional models assuming a vector-valued person parameter along with a particular arrangement of the data. For simplicity let us consider a categorical ordinal response with four categories, $\{1, 2, 3, 4\}$. The two-dimensional model that can be used has the form

$$P(Y_{ip} \geq r | Y_{ip} \in T) = F(\boldsymbol{a}^T \boldsymbol{\theta}_p - \delta_{ir}), r = 2, 3, 4,$$

with $\boldsymbol{\theta}_p^T = (\theta_{p1}, \theta_{p2})$ and the vector-valued constant $\boldsymbol{a}^T = (a_1, a_2)$ representing the loadings of the two dimensions (or factors) on the items. The condition $Y_{ip} \in T$ is kept general and will be specified from case to case.

The level 1 binary model can be specified by using $\boldsymbol{a}^T = (a_1, a_2) = (1, 0)$ and no condition, that is,

$$P(Y_{pi} \geq 3) = F(\theta_{p1} - \delta_{i3}),$$

where $\theta_{p1} = \theta_p$ and $\delta_{i3} = \delta_i^{(1)}$. The conditional level 2 model that distinguishes between categories 1 and 2 given $T \in \{1, 2\}$ is obtained by choosing $\boldsymbol{a}^T = (a_1, a_2) = (1, 1)$ yielding

$$P(Y_{pi} \geq 2 \mid Y_{pi} \leq 2) = F(\theta_{p1} + \theta_{p2} - \delta_{i2}),$$

where $\theta_{p1} = \theta_p$, $\theta_{p2} = \gamma_p$, $\delta_{i2} = \delta_{i2}^{(2)}$, therefore a response style parameter is included. The conditional level 2 model that distinguishes between categories 3 and 4 given $T \in \{3, 4\}$ is obtained by choosing $\boldsymbol{a}^T = (a_1, a_2) = (1, -1)$ yielding

$$P(Y_{pi} \geq 4 \mid Y_{pi} \geq 3) = F(\theta_{p1} - \theta_{p2} - \delta_{i4}),$$

where $\theta_{p1} = \theta_p$, $\theta_{p2} = \gamma_p$, $\delta_{i4} = \delta_{i4}^{(2)}$.

The hierarchical structure of the model allows to write the likelihood contributions for person p and item i as products of conditional probabilities. For the ordinal response $Y_{pi} = 1$ one obtains the likelihood contribution

$$\begin{aligned} L_{pi}(\theta_{p1}, \theta_{p2}, \delta_{i3}, \delta_{i2}) &= (1 - P(Y_{pi} \geq 3))(1 - P(Y_{pi} \geq 2 \mid Y_{pi} \leq 2)) \\ &= (1 - F(\theta_{p1} - \delta_{i3}))(1 - F(\theta_{p1} + \theta_{p2} - \delta_{i2})), \end{aligned}$$

for $Y_{pi} = 2$ one obtains

$$\begin{aligned} L_{pi}(\theta_{p1}, \theta_{p2}, \delta_{i3}, \delta_{i2}) &= (1 - P(Y_{pi} \geq 3))P(Y_{pi} \geq 2 \mid Y_{pi} \leq 2) \\ &= (1 - F(\theta_{p1} - \delta_{i3}))F(\theta_{p1} + \theta_{p2} - \delta_{i2}), \end{aligned}$$

for $Y_{pi} = 3$ one obtains

$$\begin{aligned} L_{pi}(\theta_{p1}, \theta_{p2}, \delta_{i3}, \delta_{i4}) &= P(Y_{pi} \geq 3)(1 - P(Y_{pi} \geq 4 \mid Y_{pi} \geq 3)) \\ &= F(\theta_{p1} - \delta_{i3})(1 - F(\theta_{p1} + \theta_{p2} - \delta_{i4})), \end{aligned}$$

and for $Y_{pi} = 4$ one obtains

$$\begin{aligned} L_{pi}(\theta_{p1}, \theta_{p2}, \delta_{i3}, \delta_{i4}) &= P(Y_{pi} \geq 3)P(Y_{pi} \geq 4 \mid Y_{pi} \geq 3) \\ &= F(\theta_{p1} - \delta_{i3})F(\theta_{p1} + \theta_{p2} - \delta_{i4}). \end{aligned}$$

Note again that $\theta_{p1} = \theta_p, \theta_{p2} = \gamma_p$ and $\delta_{i3} = \delta_i^{(1)}, \delta_{i2} = \delta_{i2}^{(2)}, \delta_{i4} = \delta_{i4}^{(2)}$. As is seen from the likelihood contributions level 1 and level 2 parameters do not occur simultaneously. In addition, the likelihood is not a function of both level 2 item parameters. For $Y_{pi} \leq 2$ the likelihood is not a function of $\delta_{i4} = \delta_{i4}^{(2)}$, and for $Y_{pi} \geq 3$ it is not a function of $\delta_{i2} = \delta_{i2}^{(2)}$.

These properties of the likelihood allow to represent it by defining binary data with missing values. Let each ordinal response Y_{pi} be represented by a row vector of Bernoulli variables $(Z_{pi1}, Z_{pi2}, Z_{pi3})$. The first entry $Z_{pi1} \in \{0, 1\}$ refers to the level 1 binary decision (or split) $\{1, 2\}$ or $\{3, 4\}$, the second entry $Z_{pi2} \in \{0, 1\}$ to the level 2 conditional binary decision $\{1\}$ or $\{2\}$ given the level 1 decision is $\{1, 2\}$, and the third entry $Z_{pi3} \in \{0, 1\}$ corresponds to the level 2 conditional binary decision $\{3\}$ or $\{4\}$ given the level 1 decision is $\{3, 4\}$. If $Y_{pi} = 1$ then set $(Z_{pi1}, Z_{pi2}, Z_{pi3}) = (0, 0, \text{NA})$, if $Y_{pi} = 2$ let the values of the binary variables be given by $(0, 1, \text{NA})$, if $Y_{pi} = 3$ by $(1, \text{NA}, 0)$, and if $Y_{pi} = 4$ by $(1, \text{NA}, 1)$, with NA denoting that the respective Bernoulli variable or conditional binary decision is not available.

Then, for $Y_{pi} \leq 2$ the likelihood can be written as

$$\begin{aligned}
&L_{pi}(\theta_{p1}, \theta_{p2}, \delta_{i3}, \delta_{i2}) \\
&= P(Y_{pi} \geq 3)^{z_{pi1}} (1 - P(Y_{pi} \geq 3))^{1-z_{pi1}} \\
&* P(Y_{pi} \geq 2 \mid Y_{pi} \leq 2)^{z_{pi2}} (1 - P(Y_{pi} \geq 2 \mid Y_{pi} \leq 2))^{1-z_{pi2}},
\end{aligned}$$

where the factor referring to the conditional binary decision $\{3\}$ or $\{4\}$ given $\{3, 4\}$, and depending on $\delta_{i4} = \delta_{i4}^{(2)}$, is omitted since the binary decision Z_{pi3} is not available for $Y_{pi} \leq 2$. For $Y_{pi} \geq 3$ one obtains

$$\begin{aligned}
&L_{pi}(\theta_{p1}, \theta_{p2}, \delta_{i3}, \delta_{i4}) \\
&= P(Y_{pi} \geq 3)^{z_{pi1}} (1 - P(Y_{pi} \geq 3))^{1-z_{pi1}} \\
&* P(Y_{pi} \geq 4 \mid Y_{pi} \geq 3)^{z_{pi3}} (1 - P(Y_{pi} \geq 4 \mid Y_{pi} \geq 3))^{1-z_{pi3}},
\end{aligned}$$

where the factor referring to the conditional binary decision $\{1\}$ or $\{2\}$ given $\{1, 2\}$, and depending on $\delta_{i2} = \delta_{i2}^{(2)}$, is omitted since the binary decision Z_{pi2} is not available for $Y_{pi} \geq 3$.

After transforming the original ordinal observations into binary observations (including missing values) the data matrix consists of P rows and $3I$ columns, i.e. three columns for each item i. Together with the corresponding specification of the loadings a_1, a_2 one can simply use ordinary existing program packages for multidimensional models, like Conquest (Adams, Wu, and Wilson, 2015) or the R (R Core Team, 2018) packages TAM (Robitzsch, Kiefer, and Wu, 2018) and MCMCpack (Martin, Quinn, and Park, 2011), to obtain parameter estimates. The former two provide marginal ML estimation and the latter a Bayesian approach. The program packages assume for each of the 3 columns per item a single item parameter. These are equivalent with the level 1 item parameter $\delta_i^{(1)}$ and the two level 2 item parameters $\delta_{i2}^{(2)}$ and $\delta_{i4}^{(2)}$ of the present tree model. For example, for $I = 2$ items the so-called Q-matrix with $3I$ rows and 2 columns (for the 2 dimensions) containing the choices of the loadings is given by

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}$$

.

3.4 Illustrative Example

To illustrate the estimation approach for hierarchical tree models we use a real-world data example from the R package ltm (Rizopoulos, 2006). The data originate from the 1992 Euro-Barometer Survey measuring the attitude towards the advantages of science and technology. The data contain categorical ordinal responses in four categories to seven questions obtained from 392 persons. The four categories are labeled with "strongly disagree", "disagree to some extent", "agree to some extent", and "strongly agree" and thus provide a symmetrical response format. Four of the seven questions are formulated positively, i.e. expressing an advantage of science and technology. Three items are negatively formulated, i.e. expressing a disadvantage. For these the order of categories has been reversed.

The R package TAM has been used to estimate the parameters of a tree model with binary splits (or decisions) exactly as decsribed in Section 3.3 and assuming the binary Rasch model for these binary splits. Two variants of the tree model have been considered, one with response style parameter γ_p and one without by setting $\gamma_p = 0$ for all persons. Marginal ML estimation has been used by assuming that (θ_i, γ_i) are drawn independently from a bivariate Gaussian distribution with expected value zero. A bivariate Gauss-Hermite procedure with 21 quadrature points for each dimension has been utilized for the numerical approximation of the integral involved in the marginal likelihood. The variances and the covariance of the assumed bivariate normal distribution are estimated jointly with the level 1 and level 2 item parameters of the model. Table 2 shows the parameter estimates and provides the obtained information criteria.

TABLE 2: *Marginal ML estimates with standard errors in parenthesis of survey measuring the attitude towards science and technolgy*

	with resp. sty. par.			without resp. sty. par.		
Item name	Level 1	Level 2 low	Level 2 high	Level 1	Level 2 low	Level 2 high
Comfort	-2.46 (0.18)	-2.94 (0.56)	1.58 (0.14)	-2.43 (0.18)	-2.31 (0.5)	1.21 (0.13)
Environment	0.92 (0.12)	-0.37 (0.14)	1.81 (0.25)	0.91 (0.11)	-0.25 (0.13)	1.51 (0.22)
Work	-0.77 (0.11)	-1.64 (0.24)	2.1 (0.18)	-0.75 (0.11)	-1.44 (0.21)	1.56 (0.16)
Future	-1.4 (0.13)	-2.51 (0.33)	1.29 (0.15)	-1.38 (0.13)	-2.13 (0.3)	0.95 (0.13)
Technology	1.06 (0.12)	-0.49 (0.14)	2.53 (0.3)	1.04 (0.12)	-0.35 (0.12)	2.06 (0.27)
Industry	1.94 (0.15)	-0.19 (0.13)	2.31 (0.4)	1.92 (0.15)	-0.12 (0.11)	1.98 (0.36)
Benefit	-0.89 (0.11)	-2.28 (0.28)	1.51 (0.16)	-0.88 (0.11)	-1.95 (0.25)	1.1 (0.14)

Note. Model with response style parameter: AIC = 5883, BIC = 5979, variances and covariance of person parameters $\mathrm{var}(\theta_p) = .51$, $\mathrm{var}(\gamma_p)$ = 1.92, $\mathrm{cov}(\theta_p, \gamma_p)$ = -.03. Model without response style parameter: AIC = 6106, BIC = 6194, variance of person parameters $\mathrm{var}(\theta_p)$ = .43.

Comparison of AIC and BIC values are in favor of the model that includes response style parameters. For the AIC one has 5883 for the model with response style parameters and 6106 for the model without. The same ordering is obtained for the BIC. The relevance of the response style parameters is also seen from the variances of the person parameters. The variance of the content-related parameter θ_p is substantially smaller than the variance of the response style parameter γ_p. The respondents seem to differ stronger in their response styles (tendency to middle vs. extreme categories) than with respect to their attitudes (towards science and technology). As can be seen from Table 2, the level 1 parameters hardly change if response style parameters are included. In contrast, level 2 parameters do change. Ignoring the response style parameter seems to yield biased level 2 item parameter estimates with a bias towards 0. In Figure 6 the item parameter estimates of the two models are plotted to visualize the differences. In the present illustrative example the correlation between response style parameters and content-related parameters is very small, the correlation coefficient is close to zero ($\rho = 0.03$).

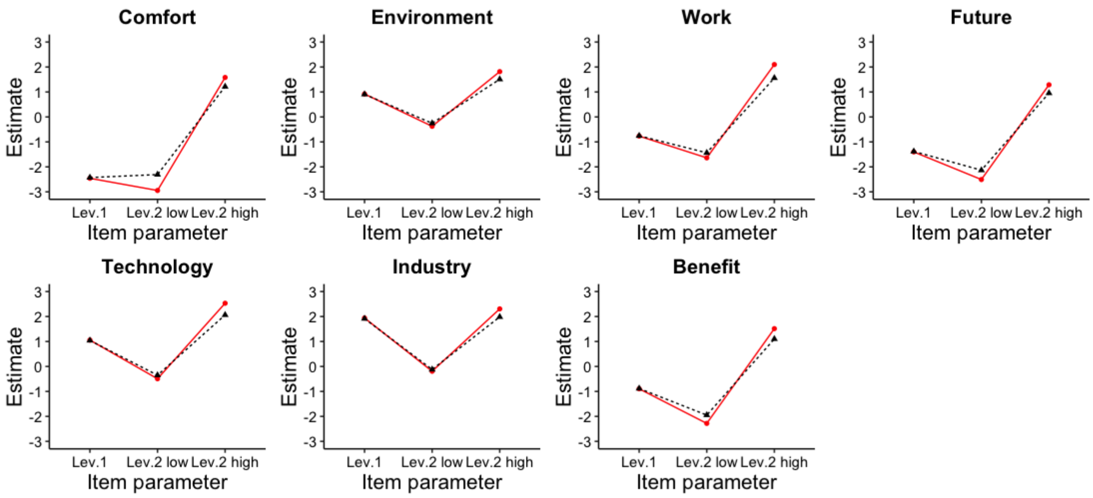

FIGURE 6: *Marginal ML estimates of item parameters. Solid (red) line visualizes estimates for model with resp. sty. par., dashed (black) line visualizes estimates for model without resp. sty. par.*

4 A Taxonomy of Ordinal Item Response Models

The taxonomy of ordinal item response models obtained by specifying the building blocks in ordinal models is visualized in Figure 7. At the outset one can distinguish between conditional models and simultaneous splits models. The former use binary models in a conditional way, by assuming that the choice between categories has already been narrowed down to a reduced set of categories. In contrast, the latter assume no conditioning but assume that the splits between categories are simultaneously determined by the same person parameter.

There are two groups of conditional models. In the first group pairs of categories are compared by utilizing a binary response model to obtain, for example, the partial credit model and its simplified version, the rating scale model. The second group is formed by hierarchical models. The crucial difference between non-hierarchical and hierarchical model is that in the former the conditions under which binary models are assumed to hold are overlapping. For example, in the partial credit model one binary sub model conditions on the the categories $\{0, 1\}$ another sub model conditions on $\{1, 2\}$. Both conditions contain the category 1. This overlapping prevents a representation as a hierarchical model.

Hierarchically structured models can be divided into three types of models, which are not exclusive. The sequential model assumes a sequential process and therefore

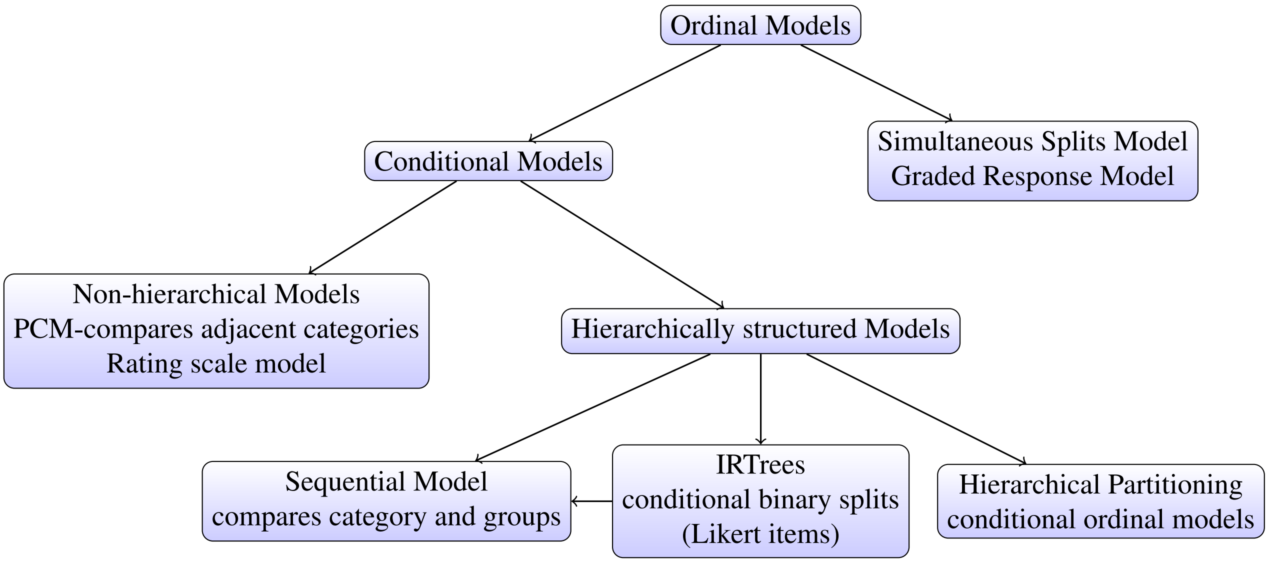

FIGURE 7: *Model overview.*

conditions on the level that has been reached. IRTrees are general hierarchical models that allow almost arbitrary binary splits. They seem useful in particular for Likert items in which the categories are divided into agreement and disagreement categories. Of course, if one does not restrict the construction to Likert type items the sequential model can be considered a special case of general IRTree models, which is visualized by the arrow that connects the two types of models. Nevertheless it seems appropriate to see it as a specific hierarchically structured model, therefore the arrow between hierarchically structured models and sequential models. Hierarchical partitioning models are general models that also allow ordinal models in the split levels. In specific cases, for example, if one has only four categories in a disagreement - agreement item, of course all the splits are binary and one obtains an IRTree model.

The structure proposed here has the purpose of characterizing modeling approaches and describing relationships. As described above the models are not always exclusive, in particular for hierarchically structured models there is some overlap between the three types of hierarchically structured models. However, most types of identified models are distinct. A graded response model cannot be represented as a partial credit model, and the partial credit model is not a hierarchically structured model.

5 Concluding Remarks

It has been shown that an easily comprehensible taxonomy of ordinal item response models is obtained by investigating the role of binary models within the structure of ordinal models. The obtained structure contains IR-Tree models and the wider class of hierarchical partitioning models. The latter has been illustrated exemplarily by investigating response styles in ordinal response data. Many more applications are feasible since the class contains a variety of models that seem worth studying and comparing in future research.

One of the advantages of having a distinct taxonomy of models is that the meaning of parameters becomes clear. In particular, parameters in conditional models should be interpreted with regard to the conditioning.

References

Adams, R. J., M. L. Wu, and M. Wilson (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement 72*(4), 547–573.

Adams, R. J., M. L. Wu, and M. Wilson (2015). Acer conquest: Generalised item response modelling software, version 4 [computer software]. *Camberwell, Victoria, Australia: Australian Council for Educational Research.*

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika 43*(4), 561–573.

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

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any 'threshold disorder controversy'. *Educational and Psychological Measurement 73*(1), 78–124.

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

Andrich, D. (2016). Rasch rating-scale model. In W. Van der Linden (Ed.), *Handbook of Modern Item Response Theory*, pp. 75–94. Springer.

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

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

Böckenholt, U. and T. Meiser (2017). Response style analysis with threshold and multi-process irt models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology 70*(1), 159–181.

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.

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.

Martin, A. D., K. M. Quinn, and J. H. Park (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software 42*(9), 22.

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

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

R Core Team (2018). R: A language and environment for statistical computing, version 3.5. 0. *R Foundation for Statistical Computing: Vienna, Austria*.

Rizopoulos, D. (2006). ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software 17*(5), 1–25.

Robitzsch, A., T. Kiefer, and M. Wu (2018). *TAM: Test analysis modules*. R package version 3.0-21.

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 W. Van der Linden (Ed.), *Handbook of Item Response Theory*, pp. 95–108.

Thissen, D. and L. Steinberg (1986). A taxonomy of item response models. *Psychometrika 51*(4), 567–577.

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). Compound regression models for categorical ordinal data. *Biometrical Journal 31*, 259–272.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Statistical and Mathematical Psychology 43*, 39–55.

Tutz, G., G. Schauberger, and M. Berger (2018). Response styles in the partial credit model. *Applied Psychological Measurement 42*, 407–427.

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

Verhelst, N. D., C. Glas, and H. De Vries (1997). A steps model to analyze partial credit. In W. Van der Linden (Ed.), *Handbook of Modern Item Response Theory*, pp. 123–138. Springer.