

A class of Multidimensional Latent Class IRT models for ordinal
polytomous item responses

Silvia Bacci*, Francesco Bartolucci†, Michela Gnaldi‡

Department of Economics, Finance, and Statistics, University of Perugia

July 18, 2013

*Corresponding author: Silvia Bacci, Department of Economics, Finance and Statistics, University of Perugia,
Via A. Pascoli, 20, 06123 Perugia, Italy; *email*: silvia.bacci@stat.unipg.it.

†*email*: bart@stat.unipg.it

‡*email*: gnaldi@stat.unipg.it

Abstract

We propose a class of multidimensional Item Response Theory models for polytomously-scored items with ordinal response categories. This class extends an existing class of multidimensional models for dichotomously-scored items in which the latent abilities are represented by a random vector assumed to have a discrete distribution, with support points corresponding to different latent classes in the population. In the proposed approach, we allow for different parameterizations for the conditional distribution of the response variables given the latent traits, which depend on the type of link function and the constraints imposed on the item parameters. Moreover, we suggest a strategy for model selection that is based on a series of steps consisting of selecting specific features, such as the dimension of the model (number of latent traits), the number of latent classes, and the specific parametrization. In order to illustrate the proposed approach, we analyze a dataset from a study on anxiety and depression on a sample of oncological patients.

KEYWORDS: Graded Response Model; Hospital Anxiety and Depression Scale; Partial Credit Model; Rating Scale Model; unidimensionality.

1 Introduction

Item Response Theory (IRT) models are commonly used to analyze data from the administration of questionnaires made of items with dichotomously- or polytomously-scored responses. In this paper, in particular, we focus on polytomously-scored items with ordered response categories. Items of this type are used in several contexts, such as education, marketing, and psychology. For a review see Hambleton and Swaminathan (1985), Van der Linden and Hambleton (1997), and Nering and Ostini (2010).

It is well known that the most common IRT models are based on the unidimensionality assumption; this means that a single latent trait is measured by the items of the adopted questionnaire. Moreover, many formulations of such models rely on the assumption that this latent trait has a normal distribution. Several extensions have been proposed in the literature to overcome these restrictive assumptions and to make traditional IRT models more flexible and realistic. Firstly, multidimensional IRT models have been proposed in the literature to account for more than one latent trait at the same time; each single trait corresponds to a separate latent dimension. We refer the reader to Reckase (2009) for a thorough overview on this topic and Duncan and Sten-

beck (1987), Agresti (1993), Kelderman and Rijkes (1994), Kelderman (1996), and Adams et al. (1997) for specific proposals. A second relevant proposal in the IRT literature is based on the assumption that the random variable representing the latent trait has a discrete distribution. Each support point of this distribution corresponds to a class of individuals that is homogeneous in terms of ability; these groups are named *latent classes* (Lazarsfeld and Henry, 1968; Goodman, 1974). Therefore, this is a convenient assumption when the model is used to group individuals in separate clusters. Moreover, the resulting model is semi-parametric (Lindsay et al., 1991), in the sense that no parametric assumptions are formulated on the distribution of the latent trait in the population. We also have to consider that adopting a discrete distribution make easy the computation of the maximum estimates via the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) because the marginal distribution of the response variables may be obtained without the need to solve an integral as with a continuous latent distribution. Comparison between these two formulations (continuous and discrete) are provided by Masters (1985), Langheine and Rost (1988), and Heinen (1996). For some examples of discretized variants of IRT models we refer the reader to Lindsay et al. (1991), Formann (1992), Hoijsink and Molenaar (1997), Vermunt (2001), Smit et al. (2003), Rost (1991), and von Davier and Rost (1995).

The two extensions described above have been combined by Bartolucci (2007), who proposed a class of multidimensional latent class (LC) IRT models for binary items, based on: (*i*) between-item multidimensionality of the latent traits (Adams et al., 1997; Zhang, 2004) and (*ii*) discreteness of the latent traits. Moreover, either a Rasch (Rasch, 1960) or a two-parameter logistic (2PL) parameterization (Birnbaum, 1968) may be adopted for the probability of a correct response to each item. A similar proposal is due to von Davier (2008), who formulated the so-called diagnostic model based on fixed rather than free abilities. See also Haberman et al. (2008), who proposed an interesting comparison of multidimensional IRT models based on continuous and discrete latent traits.

The main aim of the present article is to extend the class of models of Bartolucci (2007) to the case of items with ordinal polytomous responses. The proposed extension is formulated so that different parameterizations may be adopted for the conditional distribution of the response variables, given the latent traits. In the present class of models, different types of link function may be adopted. Using the terminology of Molenaar (1983), see also Agresti (2002) and Van der Ark

(2001), the following models result: (i) graded response models, based on global (or cumulative) logits, (ii) partial credit models, which make use of local (or adjacent category) logits, and (iii) sequential models, based on continuation ratio logits. For each of these link functions, we explicitly consider the possible presence of constraints on the item discrimination parameters and/or on the threshold difficulties. We show how all the possible parameterizations result in an extension of traditional IRT models, by introducing assumptions of multidimensionality and discreteness of the latent trait distribution.

Special attention is also devoted to model selection that, for a certain application, aims at choosing the suitable number of latent classes, the type of link function, the multidimensional structure of items, and the parameterization for the item discriminating and difficulty parameters.

The remainder of this paper is organized as follows. In Section 2 we describe some basic parameterizations for IRT models for items with ordinal responses. In Section 3 we describe the proposed class of multidimensional LC IRT models and briefly outline their maximum likelihood estimation. In Section 4 we illustrate the suggested model selection procedure and, in Section 5, the proposed class of models is illustrated through the analysis of a real dataset concerning the measurement of anxiety and depression in oncological patients. Some final remarks are reported in Section 6.

2 Models for polytomous item responses

Let X_j denote the response variable for the j -th item of the questionnaire, with $j = 1, \dots, r$. This variable has l_j categories, indexed from 0 to $l_j - 1$. Moreover, in the unidimensional case, let

$$\lambda_{jx}(\theta) = p(X_j = x | \Theta = \theta), \quad x = 0, \dots, l_j - 1,$$

denote the probability that a subject with latent trait (or ability) level θ responds with category x to this item. Also let $\boldsymbol{\lambda}_j(\theta)$ denote the probability vector $(\lambda_{j0}(\theta), \dots, \lambda_{j,l_j-1}(\theta))'$, the elements of which sum up to 1.

The IRT models for polytomous responses that are of interest in the present article are expressed

through the general formulation

$$g_x[\boldsymbol{\lambda}_j(\theta)] = \gamma_j(\theta - \beta_{jx}), \quad j = 1, \dots, r, \quad x = 1, \dots, l_j - 1, \quad (1)$$

where $g_x(\cdot)$ is a link function specific of category x and γ_j and β_{jx} are item parameters which are usually identified as *discrimination indices* and *difficulty levels*, respectively, and on which suitable constraints may be assumed. The first type of constraint is that

$$\gamma_j = 1, \quad j = 1, \dots, r, \quad (2)$$

so that all items have the same discriminating power. Another interesting constraint consists in expressing β_{jx} as sum of two components, that is,

$$\beta_{jx} = \beta_j + \tau_x, \quad j = 1, \dots, r, \quad x = 1, \dots, l_j - 1, \quad (3)$$

where β_j may be interpreted as the general difficulty level of item j and τ_x is the specific difficulty level referred to response category x , which is common to all items. This constraint makes sense only when the all items have the same number of response categories, and this number is larger than 2 (i.e., $l_1 = \dots = l_r = l$, $l > 2$). The underlying hypothesis is that the distance between two consecutive response categories, in terms of difficulty, is the same for all items, even if the items may have different general difficulty levels.

On the basis of the specification of the link function in (1) and on the basis of the constraints adopted on the item parameters, different unidimensional IRT models for polytomous responses result. In particular, the formulation of each of these models depends on:

1. *Type of link function:* We consider the link function based on: (i) global (or cumulative) logits, (ii) local (or adjacent categories) logits, and (iii) continuation ratio logits. In the first case we have that

$$g_x[\boldsymbol{\lambda}_j(\theta)] = \log \frac{\lambda_{jx}(\theta) + \dots + \lambda_{j,l_j-1}(\theta)}{\lambda_{j0}(\theta) + \dots + \lambda_{j,x-1}(\theta)} = \log \frac{p(X_j \geq x | \Theta = \theta)}{p(X_j < x | \Theta = \theta)}, \quad x = 1, \dots, l_j - 1,$$

and the corresponding models are also known as graded response models. With local logits

we define the partial credit models, which are based on

$$g_x[\boldsymbol{\lambda}_j(\theta)] = \log \frac{\lambda_{jx}(\theta)}{\lambda_{j,x-1}(\theta)} = \log \frac{p(X_j = x|\Theta = \theta)}{p(X_j = x-1|\Theta = \theta)}, \quad x = 1, \dots, l_j - 1.$$

Finally, sequential models result with continuation ratio logits, that is,

$$g_x[\boldsymbol{\lambda}_j(\theta)] = \log \frac{\lambda_{jx}(\theta) + \dots + \lambda_{j,l_j-1}(\theta)}{\lambda_{j,x-1}(\theta)} = \log \frac{p(X_j \geq x|\Theta = \theta)}{p(X_j = x-1|\Theta = \theta)}, \quad x = 1, \dots, l_j - 1.$$

2. *Constraints on the discrimination parameters:* We consider: (i) a general situation in which each item may discriminate differently from the others and (ii) a special case in which all the items discriminate in the same way, as defined in (2).
3. *Formulation of item difficulty parameters:* We consider: (i) a general situation in which the parameters β_{jx} are unconstrained and (ii) a special case in which these parameters are constrained so that the distance between difficulty levels from category to category is the same for each item (rating scale parameterization), as defined in (3).

By combining the above constraints, we obtain four different item parametrizations, which are based on free or constrained discrimination parameters and on rating scale or free parameterization for difficulties. Therefore, also according to the type of link function, twelve different types of unidimensional IRT models for ordinal responses result. These models are listed in Table 1.

TABLE 1 HERE

Abbreviations used for the models specified in Table 1 refer to the way these models are known in the literature. Thus, it is possible to identify: the Graded Response Model (GRM; Samejima, 1969), the Partial Credit model (PCM; Masters, 1982), the Rating Scale model (RSM; Andrich, 1978), the Generalized Partial Credit Model (GPCM; Muraki, 1992), and the Sequential Model (SM) obtained as special case of the acceleration model of Samejima (1995). Moreover, RS-GRM denotes the rating scale version of the GRM (Muraki, 1990); RS-GPCM and RS-SM denote the rating scale versions of GPCM (Muraki, 1997) and SM, respectively; 1P-GRM (Van der Ark,

2001), 1P-RS-GRM (Van der Ark, 2001), and SRM (Sequential Rasch Model; Tutz, 1990) refer to versions with constant discrimination index corresponding to the GRM, RS-GRM, and SM models, respectively. Finally, by SRSM we indicate the Sequential Rating Scale Model of Tutz (1990).

3 The proposed class of models

In the following, we describe the multidimensional extension of the unidimensional IRT models for ordinal responses outlined in the previous section, which is based on latent traits with a discrete distribution. We recall that the proposed class of models represents a generalization of the class of multidimensional models proposed by Bartolucci (2007) for dichotomously-scored items.

Let s be the number of dimensions, that is, different latent traits measured by the items, let $\Theta = (\Theta_1, \dots, \Theta_s)'$ be a vector of latent variables corresponding to these latent traits, and let $\theta = (\theta_1, \dots, \theta_s)'$ denote one of its possible realizations. The random vector Θ is assumed to have a discrete distribution with k support points, denoted by ξ_1, \dots, ξ_k , and mass probabilities π_1, \dots, π_k , with $\pi_c = p(\Theta = \xi_c)$. Moreover, let δ_{jd} be a dummy variable equal to 1 if item j measures latent trait of type d and to 0 otherwise, with $j = 1, \dots, r$ and $d = 1, \dots, s$.

It is important to note that assuming that the random vector Θ has a discrete distribution is equivalent to assume that the population from which the sample comes is divided into a finite number of latent classes, so that subjects in the same class have the same ability level. For the subjects in latent class c , $c = 1, \dots, k$, the level of each ability corresponds to a specific element of the vector ξ_c defined above.

In this multidimensional setting, we redefine the conditional response probabilities as

$$\lambda_{jx}(\theta) = p(X_j = x | \Theta = \theta), \quad x = 0, \dots, l_j - 1,$$

and we let $\lambda_j(\theta) = (\lambda_{j0}(\theta), \dots, \lambda_{j,l_j-1}(\theta))'$. Then, assumption (1) is generalized as follows

$$g_x[\lambda_j(\theta)] = \gamma_j \left(\sum_{d=1}^s \delta_{jd} \theta_d - \beta_{jx} \right), \quad j = 1, \dots, r, \quad x = 1, \dots, l_j - 1, \quad (4)$$

where the item parameters γ_j and β_{jx} may be subjected to the same constraints detailed in Section 2. More precisely, on the basis of the constraints assumed on these parameters, we obtain different

specifications of equation (1) that are reported in Table 2, where we distinguish the unidimensional case ($s = 1$) from the multidimensional case ($s > 1$).

TABLE 2 HERE

Each of the item parameterizations reported in Table 2 may be independently combined with global, local, and continuation ratio logit link functions to obtain different types of multidimensional LC IRT models for ordinal responses, representing as many generalizations of models as in Table 1. For instance, we may define a multidimensional LC version of GRM as

$$\log \frac{p(X_j \geq x | \Theta = \theta)}{p(X_j < x | \Theta = \theta)} = \gamma_j \left(\sum_{d=1}^s \delta_{jd} \theta_d - \beta_{jx} \right), \quad x = 1, \dots, l_j - 1, \quad (5)$$

and a multidimensional LC version of RSM as

$$\log \frac{p(X_j = x | \Theta = \theta)}{p(X_j = x - 1 | \Theta = \theta)} = \sum_{d=1}^s \delta_{jd} \theta_d - (\beta_j + \tau_x), \quad x = 1, \dots, l - 1. \quad (6)$$

Note that when $l_j = 2$, $j = 1, \dots, r$, so that item responses are binary, equations (5) and (6) specialize, respectively, in the multidimensional LC 2PL model and in the multidimensional LC Rasch model, both of which are described by Bartolucci (2007).

Regarding the *manifest distribution* of $\mathbf{X} = (X_1, \dots, X_r)'$, that is the marginal distribution of this random vector, the discreteness assumption for the random vector Θ implies that

$$p(\mathbf{x}) = p(\mathbf{X} = \mathbf{x}) = \sum_{c=1}^k p(\mathbf{X} = \mathbf{x} | \Theta = \xi_c) \pi_c,$$

where, due to the classical assumption of *local independence*, we have

$$\begin{aligned} p(\mathbf{x}|c) = p(\mathbf{X} = \mathbf{x} | \Theta = \xi_c) &= \prod_{j=1}^r p(X_j = x_j | \Theta = \xi_c) = \\ &= \prod_{d=1}^s \prod_{j \in \mathcal{J}_d} p(X_j = x_j | \Theta_d = \xi_{cd}). \end{aligned}$$

In the above expression, \mathcal{J}_d denotes the subset of $\mathcal{J} = \{1, \dots, r\}$ containing the indices of the

items measuring the d -th latent trait, with $d = 1, \dots, s$ and ξ_{cd} denoting the d -th elements of $\boldsymbol{\xi}_c$.

In order to ensure the identifiability of the proposed models, suitable constraints on the parameters are required. For this aim, in equation (4) we constrain one discriminant index to be equal to 1 and one difficulty parameter is equal to 0 for each dimension. More precisely, let j_d be a specific element of \mathcal{J}_d , say the first. When the discrimination indices are not constrained to be constant, we assume that $\gamma_{j_d} = 1$ ($d = 1, \dots, s$). Moreover, with free item difficulties we assume that $\beta_{j_d 1} = 0$ ($d = 1, \dots, s$), whereas with a rating scale parameterization we assume $\beta_{j_d} = 0$ ($d = 1, \dots, s$) and $\tau_1 = 0$.

In summary, a multidimensional LC IRT model with ordinal responses is based on a number of free parameters obtained by summing the number of free probabilities π_c , the number of ability parameters ξ_{cd} , the number of free item difficulty parameters β_{j_x} , and that of free item discrimination parameters γ_j . As shown in Table 3, it is clear that the number of free parameters does not depend on the type of logit, but only on the adopted item parametrization.

TABLE 3 HERE

The proposed models may be efficiently estimated by maximizing the model log-likelihood through an EM algorithm (Dempster et al., 1977) that may be implemented in a similar way as described in Bartolucci (2007), to which we refer the reader for details. As usual, this algorithm is based on alternating two steps, named E-step and M-step, until convergence. The E-step consists in computing the *posterior probability* of every latent class and sample unit. This is the conditional probability that the individual belongs to a certain latent class given the response configuration he/she provided. The M-step consists in maximizing the expected value of the *complete data log-likelihood*. This is the log-likelihood that could be computed knowing the latent class from which every sample unit comes. Its expected value is computed on the basis of the posterior probabilities obtained at the E-step. An implementation in R is available in the `MultiLCIRT` package by Bartolucci et al. (2013).

4 Model selection

In order to select a particular model in the proposed class of multidimensional LC IRT models for ordinal responses, we need to specify the following features: (i) number of latent classes (k), (ii) link function ($g_x(\cdot)$), (iii) constraints on the item parameters γ_j and β_{jx} , and (iv) number of latent dimensions (s) and the corresponding allocation of items within each dimension (δ_{jd} , $j = 1, \dots, r$, $d = 1, \dots, s$). Therefore, we need to make a number of choices for each mentioned aspect, relying on a suitable criterion. In particular, we adopt the Bayesian Information Criterion (BIC; Schwarz, 1978) that, differently from the criterion based on the likelihood ratio (LR) statistic, can be also used to compare non-nested models. Note that, in the multidimensional approach of Bartolucci (2007) for dichotomously-scored items, model selection is mainly based on the Akaike Information Criterion (AIC; Akaike, 1973). However, we prefer BIC in the present approach so that more parsimonious models tend to be selected, even if with moderate sample sizes the two criteria typically lead to the selection of the same model.

We propose to base model selection on the following sequence of steps:

1. *Selection of the number of latent classes.* To select the number of latent classes (k), it is useful to proceed by comparing models that differ only in this characteristic. We suggest to select k with reference to the standard LC model (Goodman, 1974). In this way, it is not necessary to choose any specific link function or item parameterization. Moreover, restrictive assumptions on the dimension of the model (number of latent traits) are not necessary, since the LC model may be also seen as an IRT model in which a specific dimension is associated to each item. In fact, the distribution of the response variables given the latent class is completely free. In practice, in order to select the number of latent classes we fit the LC model with increasing values k ; we stop to increase k when we observe an increase of the BIC index with respect to the previous value of k . Then, the value just before the first increase of the BIC index is taken as the optimal number of latent classes because it corresponds to the minimum of this index among all the fitted models. To avoid the problem of multimodality of the likelihood function, which is typical with LC models, we suggest to repeat the estimation process for each k by randomly choosing the starting values of the model parameters.
2. *Selection of the logit link function.* In selecting the type of logits (global, local, or continuation

ratio) we still use BIC, while retaining the value of k selected at the previous step. Moreover, we adopt the same multidimensional latent structure as above, that is, with one dimension for each item, so that the latent vector Θ has dimension r . Since it may happen that no relevant differences in the goodness of fit of the competing models are observed (i.e., the BIC index assumes very similar values for different link functions), the choice of the type of logit should also take into account the different interpretations behind the types of logit; see also Maydeu-Olivares et al. (1994) and Samejima (1996).

3. *Selection of the number of dimensions.* This aspect is of main interest when estimating multidimensional IRT models and is connected to testing for a certain number of dimensions. In particular, several authors (Martin-Löf, 1973; Christensen et al., 2002) have dealt with testing unidimensionality in connection with Rasch type models using LR test statistics. On the basis of this principle, Bartolucci (2007) proposed a model-based hierarchical clustering procedure that can also be applied for the extended models proposed in the present article, and then with ordinal items, and that allows us to detect groups of items that measure the same latent trait. Coherently with the general principle outlined at the beginning of this section, we select the number of dimensions on the basis of BIC. Therefore, we fit a series of models corresponding to different dimensional structures, retaining the number of classes and the type of logit selected above. The dimensional structure corresponding to the smallest BIC index is then selected.
4. *Selection of the item discriminating and difficulty parameterization.* This step consists in choosing possible constraints on the discriminating and difficulty parameters. Four different types of nested models may be defined by combining free or constrained γ_j parameters with free or constrained β_{jx} parameters. Once the other elements of the model have been defined through the previous steps, we perform the comparison among the four specification by BIC.

5 Application to measurement of anxiety and depression

The data used to illustrate the proposed class of polytomous LC IRT models consist of a sample of 201 oncological Italian patients, who were asked to fill in the “Hospital Anxiety and Depression

Scale” (HADS; Zigmond and Snaith, 1983), concerning the measurement of these pathologies. The questionnaire is composed of 14 polytomous items equally divided between the two dimensions. All items of the HADS questionnaire have four response categories: from 0 corresponding to the lowest level of anxiety or depression to 3 corresponding to the highest level of anxiety or depression. Table 4 shows the distribution of the item responses among the four categories, distinguishing between the two dimensions.

TABLE 4 HERE

Altogether, responses are mainly concentrated in categories 0 and 1 both for anxiety and depression, whereas category 3, corresponding to the highest level of psycho-pathological disturbances, is selected less than 10% of the times for each item. By summing item responses, it is possible to obtain, for each patient, a score indicating a raw measure of anxiety and depression: the closer the raw score is to the minimum value of 0, the lower the level of anxiety or depression. The mean raw score observed for the entire sample is very similar through the two dimensions, being 7.1 for anxiety and 7.2 for depression (standard deviation is equal to 4.2 in both cases). Correlation between scores on anxiety and scores on depression is very high; it is equal to 0.98.

In the present context, the assumption of unidimensionality might be not realistic. Thus, the adoption of one model in the proposed class of models, rather than a strictly unidimensional IRT model, appears more suitable and significantly more convenient, as it allows us to detect homogeneous classes of individuals who have similar latent characteristics, so that patients in the same class will receive the same clinical treatment.

To proceed with model selection, the four steps suggested in Section 4 are sequentially performed. We recall that the first step consists in detecting the optimal number of latent classes. For this aim, a comparison among LC models with different number of latent classes is performed; in particular, we consider $k = 1, \dots, 4$. The results of this preliminary fitting are reported in Table 5, where, to avoid the problem of the likelihood multimodality, results are obtained by both deterministic and random starting values.

TABLE 5 HERE

On the basis of the adopted selection criterion, we choose $k = 3$ as optimal number of latent classes, which corresponds to the smallest BIC value with both deterministic and random initializations of the EM estimation algorithm.

Regarding the second step and the choice of the suitable logit link function, a comparison between a graded response type model and a partial credit type model is carried out with $k = 3$ latent classes, free item discriminating and difficulties parameters, and a completely general multi-dimensional structure for the data (i.e., one dimension for each item). Note that the continuation ratio logit link function is not suitable in this context, because the item response process does not consist of a sequence of successive steps. Table 6 shows that a global logit link is preferable to a local logit link. It can be also observed that a graded response type model has a better fit than the standard LC model, as the BIC value observed for the former is smaller than that detected for the latter (see Table 5).

TABLE 6 HERE

Once we have chosen the global logit as the best link function, we carry on with the choice of the dimensionality structure. BIC is again used to compare models which differ in terms of this structure, all other elements being equal (i.e., free item discriminating and difficulty parameters), that is (i) a graded response model with r -dimensional structure, (ii) a graded response model with bidimensional structure (i.e., anxiety and depression), and (iii) a graded response model with unidimensional structure (i.e., all the items belong to the same dimension). For the sake of completeness, log-likelihood and BIC values are also provided for each model. According to these results, the hypothesis of unidimensionality is included in the model (see Table 7). Note that the same choice would be adopted by the LR test.

TABLE 7 HERE

As previously outlined, the number of free parameters per item depends on the adopted constraints on the discriminating indices (γ_j) and on the difficulty parameters (β_{jx}). In our application, this implies a comparison between four models, according to the classification adopted in Table 1. The results of this comparison are reported in Table 8 that shows that the preferred model among the four considered is the unidimensional 1P-GRM model, that is, the graded response type model with free β_{jx} parameters and constant γ_j parameters. Such a result is achieved on the basis of either the BIC criterion or the LR test.

TABLE 8 HERE

As the sequence of the previously described steps may be considered partly arguable, we repeated the analysis by changing the ordering of the model selection steps. It is worth noting that the unidimensional 1P-GRM is selected also through this procedure.

The main results obtained under the selected unidimensional 1P-GRM model are shown in Table 9. On the basis of the estimates of support points $\hat{\xi}_c$ and probabilities $\hat{\pi}_c$, we identify a class of patients (class 1) characterized by the least severe psycho-pathological disturbances and another class (class 3) which represent patients with the worst conditions. The last class represents the 16.7% of all patients.

TABLE 9 HERE

6 Concluding remarks

In this article, we extend the class of multidimensional latent class (LC) Item Response Theory (IRT) models (Bartolucci, 2007) for dichotomously-scored items to the case of ordinal polytomously-scored items. At the same time, the proposed class of models allows for (i) ordinal polytomous responses of different nature, (ii) multidimensionality, and (iii) discreteness of latent traits. Within the proposed approach, different link functions may be adopted. These are based on global, local, or

continuation ratio logits. Moreover, different constraints on the item discriminating and difficulty parameters may be included in the model.

It is worth noting that assuming that the latent traits have a discrete distribution is equivalent to assume that the population from which the observed sample comes is divided into a certain number of latent classes. The proposed formulation allows for estimating the abilities and probabilities of these classes. Moreover, a semi-parametric approach results which simplifies the maximization of log-likelihood function in the estimation process.

Since the specific model adopted in a certain application depends on several features, we pay particular attention to the model selection procedure which is based on a sequence of steps starting from the choice of the number of latent classes and ending with the choice of the constraints on the item parameters. In this way, we provide a unifying framework in which several aspects connected with IRT analyses are considered at the same time.

Acknowledgments

Francesco Bartolucci and Silvia Bacci acknowledge the financial support from the grant RBFR12SHVV of the Italian Government (FIRB project “Mixture and latent variable models for causal inference and analysis of socio-economic data”).

References

- Adams, R., Wilson, M., Wang, W. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement* 21:1–24.
- Agresti, A. (1993). Computing conditional maximum likelihood estimates for generalized Rasch models using simple loglinear models with diagonals parameters. *Scandinavian Journal of Statistics* 20:63–71.
- Agresti, A. (2002). *Categorical data analysis*. New York: Wiley, 2nd ed.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N., Csaki, F., eds., *Second International symposium of information theory*. Budapest: Akademiai Kiado, pp. 267–281.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* 43:561–573.

- Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika* 72:141–157.
- Bartolucci, F., Bacci, S., Gnaldi, M. (2013). MultiLCIRT: An R package for multidimensional latent class item response models. *Computational Statistics and Data Analysis, in press* .
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee’s ability. In: Lord, F. M., Novick, M. R., eds., *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA, pp. 395–479.
- Christensen, K., Bjorner, J., Kreiner, S., Petersen, J. (2002). Testing unidimensionality in polytomous Rasch models. *Psychometrika* 67:563–574.
- Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 39:1–38.
- Duncan, O., Stenbeck, M. (1987). Are Likert scales unidimensional? *Social Science Research* 16:245–259.
- Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association* 87:476–486.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61:215–231.
- Haberman, S. J., von Davier, M., Lee, Y. (2008). Comparison of multidimensional item response models: multivariate normal ability distributions versus multivariate polytomous ability distributions. Tech. Rep., ETS Research Rep. No. RR-08-45, Princeton, NJ: ETS.
- Hambleton, R. K., Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff.
- Heinen, T. (1996). *Latent class and discrete latent traits models: similarities and differences*. Thousand Oaks, CA: Sage.
- Hojtink, H., Molenaar, I. (1997). A multidimensional item response model: constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika* 62:171–190.
- Kelderman, H. (1996). Multidimensional rasch models for partial-credit scoring. *Applied Psychological Measurement* 20:155–168.

- Kelderman, H., Rijkes, J. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika* 59(2):149–176.
- Langheine, R., Rost, J. (1988). *Latent trait and latent class models*. New York: Plenum.
- Lazarsfeld, P. F., Henry, N. W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.
- Lindsay, B., Clogg, C., Greco, J. (1991). Semiparametric estimation in the rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* 86:96–107.
- Martin-Löf, P. (1973). *Statistiska modeller*. Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistisk vid Stockholms Universitet.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika* 47:149–174.
- Masters, G. (1985). A comparison of latent trait and latent class analyses of Likert-type data. *Psychometrika* 50(1):69–82.
- Maydeu-Olivares, A., Drasgow, F., Mead, A. (1994). Distinguishing among parametric Item Response models for polychotomous ordered data. *Applied Psychological Measurement* 18(3):245–256.
- Molenaar, I. (1983). *Item steps (Heymans Bulletin 83-630-OX)*. Groningen, The Netherlands: University of Groningen.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement* 14:59–71.
- Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement* 16:159–176.
- Muraki, E. (1997). A generalized partial credit model. In: Van der Linden, W., Hambleton, R. K., eds., *Handbook of modern item response theory*. Springer, pp. 153–164.
- Nering, M. L., Ostini, R. (2010). *Handbook of polytomous item response theory models*. New York: Taylor and Francis.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. (2009). *Multidimensional Item Response Theory*. Springer.

- Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *The British Journal of Mathematical and Statistical Psychology* 44:75–92.
- Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph* 17.
- Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika* 60:549–472.
- Samejima, F. (1996). Evaluation of mathematical models for ordered polychotomous responses. *Behaviormetrika* 23:17–35.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Smit, A., Kelderman, H., van der Flier, H. (2003). Latent trait latent class analysis of an eysenck personality questionnaire. *Methods of Psychological Research Online* 8(3):23–50.
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology* 43:39–55.
- Van der Ark, L. A. (2001). Relationships and properties of polytomous Item Response Theory models. *Applied Psychological Measurement* 25:273–282.
- Van der Linden, W., Hambleton, R. K. (1997). *Handbook of modern item response theory*. Springer.
- Vermunt, J. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement* 25:283–294.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology* 61(2):287–307.
- von Davier, M., Rost, J. (1995). Polytomous mixed Rasch models. In: Fischer, G., Molenaar, I., eds., *Rasch models. Foundations, recent developments, and applications*. Springer-Verlag: New York, pp. 371–379.
- Zhang, J. (2004). Comparison of unidimensional and multidimensional approaches to irt parameter estimation. Tech. Rep., ETS Research Rep. No. RR-04-44, Princeton, NJ: ETS.
- Zigmond, A., Snaith, R. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* 67(6):361–370.

discrimination indices	difficulty levels	resulting parameterization	resulting model (depending on the type of logit)		
			global	local	continuation
free	free	$\gamma_j(\theta - \beta_{jx})$	GRM	GPCM	SM
free	constrained	$\gamma_j[\theta - (\beta_j + \tau_x)]$	RS-GRM	RS-RSM	RS-SM
constrained	free	$\theta - \beta_{jx}$	1P-GRM	PCM	SRM
constrained	constrained	$\theta - (\beta_j + \tau_x)$	1P-RS-GRM	RSM	SRSM

Table 1: *List of unidimensional IRT models for ordinal polytomous responses which result from the different choices of the link function, constraints on the discrimination indices, and constraints on the difficulty levels.*

discrimination indices	difficulty levels	Number of latent traits	
		$s = 1$	$s > 1$
free	free	$\gamma_j(\theta - \beta_{jx})$	$\gamma_j(\sum_d \delta_{jd}\theta_d - \beta_{jx})$
free	constrained	$\gamma_j[\theta - (\beta_j + \tau_x)]$	$\gamma_j[\sum_d \delta_{jd}\theta_d - (\beta_j + \tau_x)]$
constrained	free	$\theta - \beta_{jx}$	$\sum_d \delta_{jd}\theta_d - \beta_{jx}$
constrained	constrained	$\theta - (\beta_j + \tau_x)$	$\sum_d \delta_{jd}\theta_d - (\beta_j + \tau_x)$

Table 2: *Item parameterizations for unidimensional ($s = 1$) and multidimensional ($s > 1$) IRT model for ordinal polytomous responses.*

discrimination indices	difficulty levels	Number of free parameters
		(#par)
free	free	$(k - 1) + sk + [\sum_{j=1}^r (l_j - 1) - s] + (r - s)$
free	constrained	$(k - 1) + sk + [(r - s) + (l - 2)] + (r - s)$
constrained	free	$(k - 1) + sk + [\sum_{j=1}^r (l_j - 1) - s]$
constrained	constrained	$(k - 1) + sk + [(r - s) + (l - 2)]$

Table 3: *Number of free parameters for different constraints on the item discrimination and difficulty parameters.*

Item	Response category			
	0	1	2	3
2	35.3	52.7	8.0	4.0
6	39.8	46.3	10.0	4.0
7	46.3	22.4	21.9	9.5
8	19.4	49.3	24.9	6.5
10	7.0	40.8	44.3	8.0
11	30.8	49.8	11.4	8.0
12	34.3	46.3	14.9	4.5
Anxiety	30.4	43.9	19.3	6.3
1	43.8	32.8	16.4	7.0
3	56.7	29.9	9.0	4.5
4	31.8	54.7	11.9	1.5
5	46.3	38.8	13.4	1.5
9	9.0	27.9	55.2	8.0
13	42.3	42.3	11.4	4.0
14	30.8	37.3	28.9	3.0
Depression	37.2	37.7	20.9	4.2

Table 4: *Distribution of HADS item responses (row percentage frequencies).*

k	Deterministic start			Random start		
	$\hat{\ell}$	#par	BIC	$\hat{\ell}(\max)$	#par	BIC(min)
1	-3153.151	42	6529.040	-3153.151	42	6529.040
2	-2814.635	85	6080.051	-2814.635	85	6080.051
3	-2677.822	128	6034.468	-2674.484	128	6027.791
4	-2645.435	171	6197.736	-2608.570	171	6104.805

Table 5: *Standard LC models: log-likelihood ($\hat{\ell}$), number of parameters, and BIC values for $k = 1, \dots, 4$ latent classes (the smallest BIC value, selected with deterministic and random starts, is in boldface).*

	Global logit	Local logit
$\hat{\ell}$	-2726.348	-2741.321
#par	72	72
BIC	5834.534	5864.479

Table 6: *Graded response and partial credit type models with $k = 3$: log-likelihood ($\hat{\ell}$), number of parameters, and BIC values (the smallest BIC value is in boldface).*

Model	$\hat{\ell}$	#par	BIC	Deviance	p -value
r -dimensional	-2726.348	72	5834.534	–	–
bidimensional	-2731.249	60	5780.696	9.802	0.633
unidimensional	-2731.894	59	5776.682	1.290	0.256

Table 7: r -dimensional, bidimensional, and unidimensional graded response models with $k = 3$: log-likelihood, number of parameters, BIC value, and LR test results (deviance and p -value); (the smallest BIC value is in boldface).

Model	$\hat{\ell}$	#par	BIC	Deviance	p -value
GRM	-2731.894	59	5776.682	–	–
RS-GRM	-2795.570	33	5766.149	127.353 (vs GRM)	0.000
1P-GRM	-2741.285	46	5726.521	18.782 (vs GRM)	0.130
1P-RS-GRM	-2844.518	20	5795.102	206.467 (vs 1P-GRM)	0.000

Table 8: Item parameters selection: log-likelihood, number of parameters, BIC values, and LR test results (deviance and p -value) between nested graded response models with $k = 3$ and $s = 1$ (the smallest BIC value is in boldface).

Dimension	Latent class c		
	1	2	3
Psycho-pathological disturbances	-0.776	1.183	3.418
Probability	0.342	0.491	0.167

Table 9: Estimated support points $\hat{\xi}_c$ and probabilities $\hat{\pi}_c$ of latent classes for the unidimensional 1P-GRM.