Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN METODOLOGIA STATISTICA
PER LA RICERCA SCIENTIFICA

Ciclo XXVI

Settore Concorsuale di afferenza: 13/D1
Settore Scientifico disciplinare: SECS-S/01

MULTIDIMENSIONAL ITEM RESPONSE THEORY MODELS WITH GENERAL AND SPECIFIC LATENT TRAITS FOR ORDINAL DATA

Presentata da: Irene Martelli

Esame finale anno 2014

Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN METODOLOGIA STATISTICA
PER LA RICERCA SCIENTIFICA

Ciclo XXVI

Settore Concorsuale di afferenza: 13/D1
Settore Scientifico disciplinare: SECS-S/01

MULTIDIMENSIONAL ITEM RESPONSE THEORY MODELS WITH GENERAL AND SPECIFIC LATENT TRAITS FOR ORDINAL DATA

Presentata da: Irene Martelli

Coordinatore Dottorato:
Chiar.mo Prof. Angela Montanari

Relatore:
Chiar.mo Prof. Stefania Mignani

Esame finale anno 2014

To my family and Lorenzo,
for their love and support.

Abstract

The aim of the thesis is to propose a Bayesian estimation through Markov chain Monte Carlo of multidimensional item response theory models for graded responses with complex structures and correlated traits. In particular, this work focuses on the multiunidimensional and the additive underlying latent structures, considering that the first one is widely used and represents a classical approach in multidimensional item response analysis, while the second one is able to reflect the complexity of real interactions between items and respondents.

A simulation study is conducted to evaluate the parameter recovery for the proposed models under different conditions (sample size, test and subtest length, number of response categories, and correlation structure). The results show that the parameter recovery is particularly sensitive to the sample size, due to the model complexity and the high number of parameters to be estimated. For a sufficiently large sample size the parameters of the multiunidimensional and additive graded response models are well reproduced. The results are also affected by the trade-off between the number of items constituting the test and the number of item categories.

An application of the proposed models on response data collected to investigate Romagna and San Marino residents' perceptions and attitudes towards the tourism industry is also presented.

Acknowledgements

First and foremost I want to thank my supervisor Stefania Mignani for her constant attention, care and belief. Then, I would like to express my gratitude to Mariagiulia Matteucci for her precious and fundamental suggestions and supervision during the preparation of the thesis. My work would not have been successful without her. Appreciation is extended to Cristina Bernini who has provided the data. My special thanks to my friends and colleagues Lucia and Violeta for their support during the whole period of the PhD.

Preface

Item response theory (IRT) falls within the wide context of the measurement of theoretical latent constructs, which are not observable by definition and can only be determined indirectly, through the use of other manifest variables.

IRT is extensively used in educational and psychological fields, where usually a test consisting of a set of items is submitted to a sample of examinees to infer the individuals' unobservable characteristics (abilities). To this aim, IRT (Hambleton and Swaminathan, 1985; van der Linden and Hambleton, 1997) represents the main methodological approach that allows to estimate both the item psychometric properties and the subjects' scores. Moreover, IRT shows a great potential in applications within behavioral sciences.

In the past, unidimensionality, i.e. the presence of a unique construct underlying the response process, was one of the most common assumption. Nevertheless, real data often suggest a multidimensional structure and, with the aim to infer such distinct latent traits, tests should include different subtests.

For this reason, models that allow the presence of more than one latent trait have been recently developed. The so called multidimensional IRT (MIRT) models (see e.g., Reckase, 2009) are able to describe the complexity of the data, taking into account correlated abilities and also a possible hierarchical structure of latent traits. This is the reason why MIRT models perform better in fitting the subtests if compared to separate unidimensional models.

Several approaches are possible within the multidimensional perspective: explorative models, where all latent traits are allowed to affect all the item responses, or confirmatory models, where all the relations between observed and latent variables need to be specified in advance. By using a confirmatory approach, it is also possible to assume the simultaneous presence of general and specific latent traits underlying the response process (Sheng and Wikle, 2008). A further distinction can be made between non compensatory and compensatory models, where a lack in one trait naturally compensates for the other (Reckase, 2009).

In several applications, data are characterized by hierarchical structures and the introduction of different levels for latent dimensions permits to specify more general models. Specifically, a proper hierarchy can be assumed to underlie the response process, where the highest level is associated with the overall trait, while dimensions representing more specific traits are located on lower level of the hierarchy.

High-order and additive models are two approaches that allow to include a general trait in addition to multiple specific traits. Particularly, in additive models, we can analyze the strength of the relationships between the specific latent traits and the associated test items directly as well as the strength of the relationships between the general latent trait and all the test items. This feature is particularly appealing for complex applications.

A final distinction can be made according to the nature of the observable variables. Usually, in an educational testing framework we deal with binary items (i.e. correct/incorrect) while in psychological and behavioral researches items are typically ordinal, representing judgments or agreements. Different models for ordinal data have been developed according to the number of item parameters (e.g. partial credit models, graded response models) in a unidimensional context. On the contrary, within a multidimensional context, models for binary data are usually applied and, often, the available ordinal data are dichotomized, with a consequent loss of information. Models for ordinal data remain uncommon and were developed only for uncorrelated latent traits.

For these reasons, in this work we propose an extension of the unidimensional graded response model (Samejima, 1969) for ordinal data to multidimensional structures with correlated traits, namely the multiunidimensional and the additive structures. A further innovative and important aspect of our proposal deals with the estimation procedure, in fact, we propose a Markov chain Monte Carlo (MCMC) procedure for parameter estimation which we implement using the open-source software OpenBUGS.

Structure of the thesis

In the first chapter some fundamental notions about IRT are introduced. A first section illustrates the basic concepts and definitions characterizing the IRT approach, with a brief description of unidimensional models for binary data. A second section focuses on unidimensional models for ordinal data and, in particular, on the Samejima's model for graded responses. A final section explains the reasons that have driven several developments of IRT towards its multidimensional generalization.

The second chapter introduces the MIRT approach. In the first section the main features of these models are described, while in the second section a brief review on MIRT models for both binary and ordinal response is reported, together with a brief description of their most common estimation methods.

In the third chapter the main principles characterizing the Bayesian estimation in MIRT context are introduced. The first section describes the general Bayesian framework, while the second section presents the available Bayesian estimation methods based on MCMC techniques. The third section briefly introduces the functioning of OpenBUGS, which permits to easily run the most common MCMC algorithm, i.e. the Gibbs sampler.

In the fourth chapter two MIRT models for ordinal data with a complex structure are introduced in terms of specification, interpretation and estimation. The focus is on two MIRT models for graded responses and correlated latent traits: the multiunidimensional model, where items in each subtest characterize a single ability, and the additive model, where each item measures a general and a specific ability directly.

The fifth chapter describes a simulation study that has been conducted in order to evaluate the parameter recovery of the estimation method for the proposed models. The simulation study design is illustrated in the first section, while the second and the third sections report the results of the simulations performed for the multiunidimensional and the additive models for ordinal data, respectively.

In the sixth chapter an application of the proposed models to real data is presented. The application focuses on the investigation of residents' perceptions and attitudes towards the tourism industry.

In the seventh chapter conclusions and further research on applicative and methodological aspects are discussed.

Contents

List of Tables

List of Figures

Chapter 1

An introduction to item response theory (IRT)

In this chapter we introduce the fundamental notions concerning item response theory (IRT). A brief description of IRT models for binary and ordinal data is carried out. Particular attention is given to the unidimensional Samejima's model for graded responses, which represents the starting point towards a generalization into a multidimensional context.

1.1 Basic concepts and definitions

IRT falls within the wide context of the measurement of theoretical latent constructs. A latent construct is not observable by definition and it can only be determined indirectly, through the use of other manifest variables. Examples of latent constructs are the mathematics achievements of students, the satisfaction of a costumer about a product or service, the psychological status and all the situations that may refer to the concept of perception, e.g. depression and happiness. Another relevant field of application of IRT methods is represented by the behavioral sciences, where the manifest variables, that are often ordinal, express a judge or an agreement to the phenomenon of interest.

If we consider the educational and psychological fields, where IRT is extensively used, we can say that IRT has the final aim to measure abilities and attitudes of individuals through the responses on a number of test items. In other words, by using IRT models, we wish to determine the position of the individual along some latent dimensions, representing the unobservable characteristics of the individuals.

In IRT literature the latent traits are commonly called abilities, for the intensive use of IRT methods in the educational field, where the constructs are represented by the students' latent abilities. The analysis of the relation between latent continuous variables and observed categorical variables is known in the statistical literature as latent trait analysis, that is the reason why in this thesis the words "abilities", "latent abilities" and "latent traits" are all referred to the same concept.

The use of IRT as a measurement theory is fairly recent: in the pioneer work of Lord and Novick (1968) a first formalization of the theory is expressed, on the basis of ideas and principles that raised in the thirties and forties. Improvements of IRT were due to the necessity to overtake the lacks of the classical test theory (CTT), for instance the sensitivity to sample conditions and the fact that in CTT individual abilities and test characteristics can be interpreted only in the same context (Hambleton et al., 1991). Moreover, IRT focuses on item rather than on individual score, while in the CTT the evaluation of test properties and item characteristics are not included. On the other side, IRT permits to evaluate individual ability and to describe the performances of the items on the test simultaneously. For these reasons, IRT seemed to be an alternative and promising method to substitute CTT in theoretical and application fields, showing a wide and effective framework.

1.1.1 The concept of model in IRT

In IRT a model is defined by a mathematical function used to describe the conditional probability of a response given the latent ability, for an item with categorical responses (Thissen and Steinberg, 1986). The mathematical function expresses how an examinee with a high position on a latent trait is likely to provide a different response to an examinee with a low position on the trait (Ostini and Nering, 2006). The parametric model describes the relationship between the "observable", i.e. the examinee's performance in the test, and the "unobservable", the latent ability.

In general, different models can be specified depending on:

- The structure of the data: binary or polytomous (nominal or ordinal) responses;
- The number of latent dimensions: unidimensional or multidimensional models;
- The distribution functions used to link responses and ability(ies);
- The number of item parameters introduced in the model.

Concerning the first point, IRT permits to specify different models depending on the kind of items we are dealing with, i.e. items with two response categories or items with more than two response categories (that, in turn, can be odered or not). The second point is a crucial choice in the model specification procedure: when only one ability affects the responses we are assuming unidimensionality, while when we need two or more latent traits to describe the correlation among the responses we are assuming multidimensionality. Moreover, the model depends on the probability distribution used to describe the relationship between the response and the examinee's ability(ies) and the number of parameters describing the item characteristics introduced. The most common probability models used are the normal distribution function (normal ogive models) and the logistic distribution function (logit models). Finally, a distinction can be made with reference to the number of item parameters, one, two or three, introduced in the model.

1.1.2 IRT unidimensional models for binary data

In order to illustrate the basic concepts and assumptions of IRT and to introduce the notation, we start from the simplest models: the unidimensional models for dichotomous responses (i.e. correct and incorrect). In this context there are three fundamental assumptions.

The first assumption states that only one latent ability affects the item responses (unidimensionality assumption).

The second assumption states that a change in the probability of a correct response, due to a change in the examinee latent ability, is completely described by the item characteristic curve (ICC). Thus, the ICC describes how the probability of a response to an item changes relative to a change in the latent trait. As illustrated before, different distribution functions used to link responses and ability, i.e. different mathematical forms of the ICC, lead to different IRT models. In any case the probability of a correct response is expressed as a function of person and item parameters.

The third assumption is the so called local independence assumption: responses to a pair of items are statistically independent given the underlying latent ability. Local independence holds when the assumption of unidimensionality is true. Let consider a random vector of p item responses for the i-th subject ($i = 1, \dots, n$), denoted by $\mathbf{Y}_i$, and the corresponding observed responses, $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})$. θ_i is the ability of the examinee i. The assumption of local independence can be stated as:

$$P(\mathbf{y}_i|\theta_i) = P(y_{i1}|\theta_i)P(y_{i2}|\theta_i)\dots P(y_{ip}|\theta_i) = \prod_{j=1}^{p} P(y_{ij}|\theta_i) \; .$$

When local independence holds, there is one latent variable underlying the responses and, conditionally to this latent variable, responses are assumed to be independent.

The unidimensional IRT model for binary data expresses the probability π_{ij} of a correct response by the subject i to the item j as a function of the predictor η_{ij}, which depends on θ_i and on $\boldsymbol{\xi}_j$, the vector of parameters characterizing item j, for $j = 1, \dots, p$:

$$\eta_{ij} = f(\theta_i, \boldsymbol{\xi}_j) \; . \tag{1.1}$$

The so called probit or normal ogive model is obtained when a normal distribution is used (1.2), whereas when we use the logistic distribution we get the logit model (1.3)[1]:

[1]Normal ogive models and logistic models have different ICCs for equivalent set of item parameters values. It can be proved (Haley, 1952; Birnbaum, 1968) that the two formulations are equivalent in terms of predicted probability through the introduction of a scaling constant 1.702 into the logistic model, in order to balance for differences in ICCs. When this constant is introduced in the model, the predicted probabilities differ by less than 0.01 for each level of ability (Haley, 1952): $|\,\Phi(\eta_{ij}) - \exp(1.702\,\eta_{ij})/[1 + \exp(1.702\,\eta_{ij})]\,| < 0.01$.

$$\pi_{ij} = \Phi(\eta_{ij}) \Rightarrow \Phi^{-1}(\pi_{ij}) = \eta_j \tag{1.2}$$

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \Rightarrow \text{logit}(\pi_{ij}) = \eta_{ij} \ , \tag{1.3}$$

where Φ is the standard normal cumulative distribution function. Different unidimensional models can then be obtained by introducing a different number of item parameters $\boldsymbol{\xi}_j$ describing the item characteristics. The simplest case has only one item parameter $\boldsymbol{\xi}_j = \{\beta_j\}$, and β_j is called *difficulty parameter*. An example of one-parameter logistic model is the Rasch model (Rasch, 1960) and if we consider a logarithmic transformations of the scale of person and item parameters (Fischer, 1995), the predictor becomes $\eta_{ij} = \theta_i - \beta_j$.

If $\boldsymbol{\xi}_j = \{\alpha_j, \beta_j\}$ a *discrimination parameter* α_j is added to the model and we are in the case of two-parameter models. The predictor (1.1) becomes $\eta_{ij} = \alpha_j\theta_i - \beta_j$: model (1.2) becomes the two-parameter normal ogive model (Lord, 1952) while model (1.3) becomes the two-parameter logistic model (Birnbaum, 1968).

A further extension can finally be done by introducing a *guessing parameter* γ_j for each item, leading to three-parameter models where $\boldsymbol{\xi}_j = \{\alpha_j, \beta_j \, \gamma_j\}$ (Lord, 1980). See Reckase (2009) for an exhaustive description of such models.

With respect to the ICC, the parameters α_j, β_j and γ_j represent the slope, the location and the lower asymptote, respectively.

1.2 IRT unidimensional models for ordinal data

Models briefly presented above are all referred to dichotomous responses, nevertheless items with multiple response options exist and their use is quite common in behavioral sciences. IRT models for polytomous items operate in a different way from binary models. In the latter case the knowledge of the characteristics of a response determines also the characteristics of the other complementary response, while for polytomous items this feature does not hold anymore and each category function must be modeled separately (Samejima, 1996). In Figure 1.1 the ICC for a binary item is reported, while Figure 1.2 shows different response functions for an item with five categories.

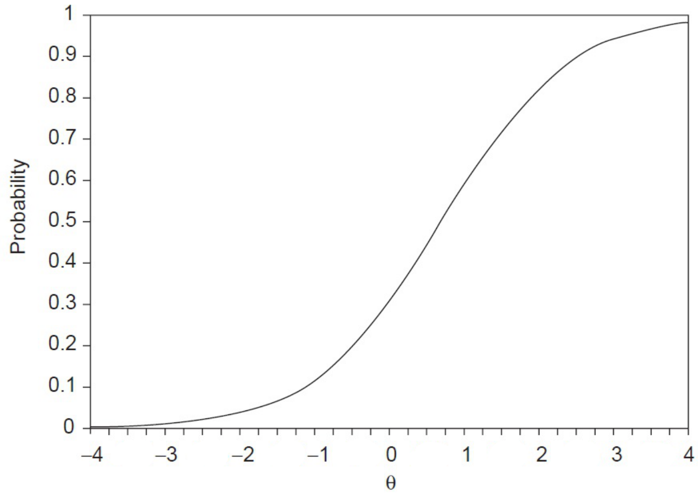


Figure 1.1. Item characteristic curve for a binary item

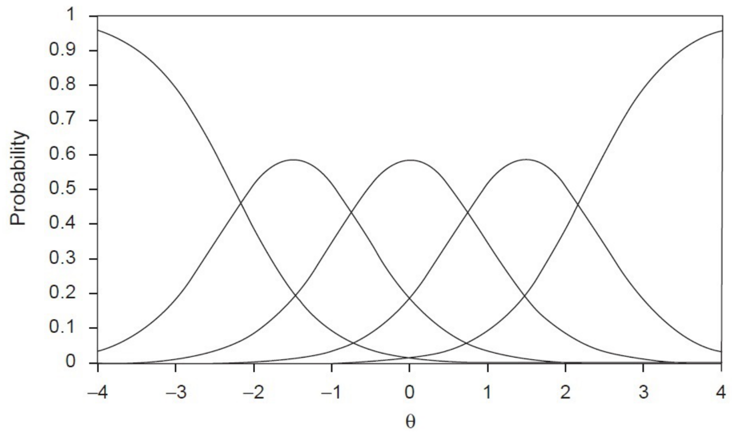


Figure 1.2. Item response functions for an item with five categories

From Figure 1.2 we can see how, for ordered items, the category response functions are not all monotonic: only the curves related to the first and the last categories are, respectively, monotonically decreasing and increasing. The presence of non-monotonic functions raises some complications: these functions cannot be described only in terms of discrimination and difficulty parameter, as in the binary case. The choice of the proper mathematical form and the estimation of parameters for such unimodal functions is a relevant issue. For ordered polytomous items this problem has been solved by treating polytomous items basically as 'concatenated dichotomous' items (Samejima, 1969, 1996): dichotomizations of item response data are combined in order to get suitable response functions for each item category.

As we will illustrate more in detail later, several models for ordinal data exist as result of extensions of the models for binary data. The simplest model for ordinal items is the partial credit model (Masters, 1982), which is an extension of the Rash model for binary items, i.e. with one item parameter. Despite its wide use, it focuses on the scoring of the individuals and its restrictive assumptions make it inadequate for modeling purposes, especially in complex contests. In this work we focus on the Samejima's graded response model, which is the generalization of the two-parameter IRT model for binary data. This choice has been lead by the consideration that models that include also the guessing parameter, even if they are appropriate educational field, do not suit well in the context of behavioral science, where individuals typically express opinions.

1.2.1 Samejima's unidimensional graded response model

The graded response model for ordinal data was developed by Samejima in 1969. Examples of graded responses are Lykert-type scales ("strongly-disagree", "disagree", "neutral", "agree", and "strongly agree") and responses ordered on the basis of a range of scores.

Let consider a set of p ordinal items, $Y_1, \ldots, Y_j, \ldots, Y_p$, where each item has K_j categories, indexed by k. In the parametrization of the model we consider that the lowest score on item j is 1, while the highest score is K_j and each item is characterized by $K_j - 1$ thresholds or boundaries $\kappa_{j1}, \ldots, \kappa_{j,K_j-1}$. The probability of achieving k or higher categories is assumed to increase monotonically with a growth in the latent ability (Samejima, 1996; Reckase, 2009), therefore the thresholds must satisfy the so called order constraint: $\kappa_{j1} < \cdots < \kappa_{j,K_j-1}$.

Concerning the dichotomization procedure mentioned above, Samejima's (1969) graded model is based on the probability that an item response will be observed in *category k or higher*: the probability π_{ijk} that the i-th subject will select the k-th category on item j is equal to the probability of answering above the lower boundary for the category (κ_{k-1}) minus the probability of answering above the category's upper boundary (κ_k). Figure 1.3[2] describes the dichotomization method used in Samejima's models, a dashed line is used to represent an hypothetical response in category $k = 4$: the probability to have a response in such category can be computed as $P^*_{i4} - P^*_{i5}$, where in general with $P^*_{ik} = P(Y_{ij} \geq k|\theta_i)$ we denote the probability of accomplishing step k at a given level of θ.

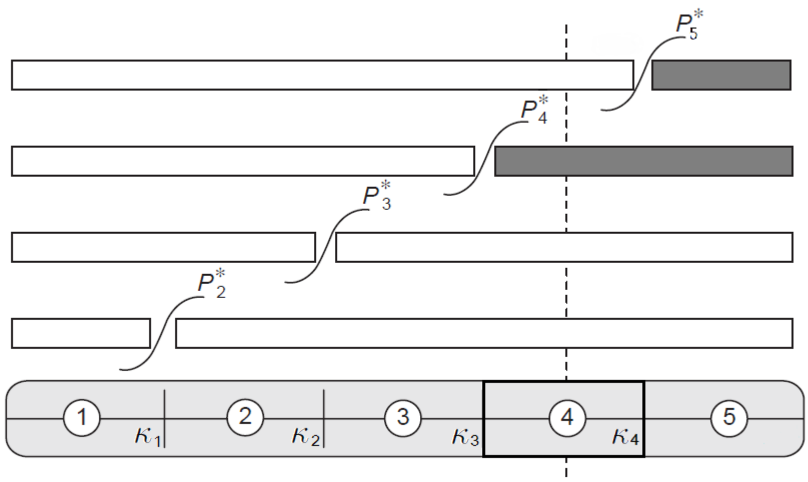


Figure 1.3. Dichotomization of polytomous item responses, the dashed line indicates the observed category response.

The probability that the i-th examinee's response will fall in the k-th category on item j can thus be written as:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i) = P^*_{ik} - P^*_{i,k+1} \ , \tag{1.4}$$

where P^*_{i1} and P^*_{i,K_j+1} are assumed to be respectively 1 and 0, in order to ensure that the probability of each category can be determined from (1.4). The two-parameter normal ogive and logistic formulations of the model can be obtained from (1.4). The normal ogive form of the Samejima's model for graded responses

[2]Figure adapted from Ostini and Nering (2006).

is given by:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i, \kappa_{jk}, \kappa_{j,k+1}) = \frac{1}{\sqrt{2\pi}} \int\limits_{\alpha_j\theta_i - \kappa_{j,k+1}}^{\alpha_j\theta_i - \kappa_{jk}} e^{-t^2/2}\, dt \ . \tag{1.5}$$

From expression (1.5), we can observe that the discrimination parameter α_j, i.e. the slope of the response functions, is constant between all different category responses of a given item. This constraint ensure to avoid negative probabilities (Steinberg and Thissen, 1995). The boundary parameters κ_{jk} vary within an item, according to the order constraint $\kappa_{j,k-1} < \kappa_{jk} < \kappa_{j,k+1}$, and at each level of $\theta = \kappa_{jk}$, the examinee has a probability of 0.5 of endorsing the category (Reeve, 2002). P^*_{ik} is the trace line reflecting the probability that an examinee's response will fall in that scoring category or a higher, at any specific level of latent ability θ. The graded model response function $P(Y_{ij} = k|\theta_i)$ reflects the rate of examinees responding to the k-th category through the different levels of θ, that is a non-monotonic curve, with the exception of the curves associated to the extreme categories, as previously pointed out in Figure 1.2 (Thissen et al., 2001).

1.2.2 Other unidimensional IRT models for graded responses

Several models for items with two or more ordered responses have been developed. An assortment of these models, together with their features, has been introduced by van der Linden and Hambleton (1997) and van der Ark (2001). In addition to Samejima's graded response model (1969), other widely applied IRT models for ordinal data are the partial credit model (Masters, 1982) and its extension, the generalized partial credit model (Muraki, 1992). The partial credit model is an extension to the case of ordinal items of the Rash model for binary items, i.e. with one item parameter. On the other side, the Samejima's graded response model is the generalization of the two-parameter IRT model for binary data.

In partial credit model and in its generalization, the category responses on the item represent the "levels of performance" (Reckase, 2009). As well as in the graded response model, we have thresholds between adjacent scores: an examinee's performance is on the left or the right side of a threshold with a specific probability. Here the dichotomization procedure involves only two category boundaries for a given item, see Ostini and Nering (2006) for a detailed discussion about differences between Samejima and Rasch dichotomization approaches.

Mathematical expressions for the partial credit model and the generalized partial credit model are presented in (1.6) and (1.7), where $\boldsymbol{D} = 1.702$ is the scaling constant:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i) = \frac{exp\left\{\sum_{u=1}^{k}(\theta_i - \kappa_{ju})\right\}}{\sum_{v=1}^{K_j} exp\left\{\sum_{u=1}^{k}(\theta_i - \kappa_{ju})\right\}} \tag{1.6}$$

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i) = \frac{exp\left\{\sum_{u=1}^{k}\boldsymbol{D}\alpha_j(\theta_i - \beta_j + \kappa_{ju})\right\}}{\sum_{v=1}^{K_j} exp\left\{\sum_{u=1}^{k}\boldsymbol{D}\alpha_j(\theta_i - \beta_j + \kappa_{ju})\right\}}\ . \tag{1.7}$$

In the generalized partial credit model the assumption of constant discrimination parameter of test items is relaxed, in fact α_j parameters may vary across items. Reckase (2009) provides an exhaustive illustration of such models. Other IRT models for polytomous items have been proposed by Bock (1972), Andrich (1978, 1982), Thissen and Steinberg (1984), and Rost (1988). All these models refer to an unidimensional underlying ability structure.

1.3 Towards multidimensional models

Unidimensional models are suitable when tests are made to measure only one latent ability (Sheng and Wikle, 2009). There are some advantages in the use of such unidimensional models: i) they have quite simple mathematical forms; ii) they perform well in fitting the data in several empirical applications; and iii) they are rather robust to violations of assumptions (Reckase, 2009).

Nevertheless, real interactions between examinees and test items are not simple as described in unidimensional models. A person is likely to use more than a single ability in the response process, on one hand, and the problems posed in a test can require several abilities in order to get the right solution, on the other side.

Multidimensional IRT (MIRT) models were developed to have a more accurate description of interactions between persons and test items. In particular, in MIRT models a vector of latent abilities is introduced, instead of assuming a single person parameter.

In other words, MIRT models deal with quite common circumstances where an examinee requires multiple abilities in order to respond to an item. In this case, more than one latent construct is measured by that item. One of the most famous example in the educational field is a "mathematical test item presented as story that requires both mathematical and verbal abilities to arrive at a correct score" (Fox, 2010), where both mathematical and reading comprehension skills are involved in the answering process.

Chapter 2

Multidimensional IRT (MIRT) models: a review

As previously pointed out, the latent space that has to be measured may be more complex than the one underlying unidimensional IRT models. The so called MIRT models are used when separate latent abilities are encompassed in the observed responses for an item.

In this chapter we introduce the MIRT approach. In particular, we show how different models can be specified depending on the latent ability structure hypothesized to underlie the response process. A literature review on MIRT models for both binary and ordinal data is reported. A final section describes the most common estimation methods in IRT and MIRT frameworks.

2.1 Main features of MIRT models

The assessment of dimensionality is a key topic in IRT and in the latent variable framework. A review of methods for an empirical detection of the structure of tests with binary items was made by Tate (2003). In his work, a particular attention is given to the assessment of the test statistical structure as subtended from the relations between examinees and items. This aspect should be an important part of the development, evaluation, and maintenance of large-scale test.

Several IRT models are based on a common postulate: the assumption of unidimensionality. However, the local independence assumption holds only if the latent space is entirely specified. For this reason, many efforts for the characterization of the concept of dimensionality and for its detection have been made. We can say that an accurate and unequivocal definition of dimensionality does not exist yet. This is due to the fact that the phenomenon is latent by nature, hence a direct comparison with observed results is not possible.

Hambleton and Swaminathan (1985) justified the unidimensionality assumption with the presence of a dominant trait able to explain the examinees' responses. In this sense, we can imagine that a single trait always exists but crucial points are if the dominant trait is sufficiently strong and in which way it dominates the others. Conversely, Traub (1983) argued that unidimensionality is probably more the exception than the rule, with respect to the skills necessary to answer to the items on most cognitive tests.

Some weak features of the unidimensionality assumption have been reviewed by Adams et al. (1997), with the aim to propose a MIRT model. The use of unidimensional models might be improper for tests intentionally built from subcomponents that are assumed to measure different abilities. IRT models seem to be robust to these violations of unidimensionality, especially with highly correlated latent constructs. In fact, if we assume the existence of a single latent ability, it can be seen as the dominant factor reflecting the different composition of the items. On the other hand, when a test is made by mutually exclusive subtests of items or when the underlying dimensions are not highly correlated, the use of a unidimensional model can bias the parameter estimation, adaptive item selection and trait estimation. The problem is highlighted especially in adaptive testing, when the examinees are administrated different combinations of items and the traits underlying the performance may reflect the different composition of the items (Matteucci, 2007).

Finally, as shortly described at the end of Chapter 1, the assessment of knowledge, competencies and achievement is going more and more towards a multidimensional evaluation. The reason of the widely use of MIRT models in recent studied is that the actual interactions between examinees and test items are complex and necessitate to be framed in a multidimensional background. A clarifying example reported in Matteucci (2007) concerns the assessment of proficiency in the University context, where the student's evaluation is typically multidimensional at each level: within a single course and during all the University career, students are evaluated on the basis of multiple competencies.

2.1.1 Compensatory and noncompensatory approaches

MIRT models can be classified in two main groups: compensatory and noncompensatory models, depending on the way the vector of latent abilities, $\boldsymbol{\theta}$, is combined with item parameters to obtain the probability of responses to the item.

In *compensatory* models we use a linear combination of the values of $\boldsymbol{\theta}$ in the specification of the response probabilities, by using a logistic or a normal ogive form. This approach implies that different combinations of elements in $\boldsymbol{\theta}$ can yield the same sum, and the direct consequence is a compensation effect: if a θ-value is low, but another one is appropriately high, the sum can be the same.

In *noncompensatory* models, different latent abilities used to solve an item are separated and each part is used as an unidimensional model. Then the global probability is obtained as the product of the probabilities of each unidimensional part. Nonlinearity raises in relation to the use of the product of such probabilities, and the compensation property does not hold (Reckase, 2009).

2.1.2 Confirmatory and exploratory approaches

Another classification of MIRT models can be done with reference to the available information at the model specification step. Mainly, the investigation of multidimensionality can be conducted by using two different approaches: the exploratory and the confirmatory approaches.

In the exploratory approach no prior knowledge is included in the model, in terms of relationship between items and latent traits.

When the number of latent abilities is specified in advance, the method is not merely explorative and we are in a confirmatory context. In line with the confirmatory approach, not only the number of latent variables is pre-specified but also their relationships with the items. In fact, the researcher can use prior knowledge to define which items load on which factors.

2.1.3 Underlying latent structures

In this paragraph a brief review of different multidimensional latent structures is reported. For simplicity, figures are referred to the simplest case of a test consisting of two subtests. Circles represent latent traits and squares represent observed item responses. Subtests are indicated with dashed lines.

Consecutive unidimensional model In Figure 2.1 is illustrated the so called *consecutive unidimensional* approach, where simple unidimensional IRT models are fitted to each subtest in a sequential way. Fitting this model, we obtain person measures for every specific ability, but a direct estimation for the relation between them is not feasible (Huang et al., 2013).

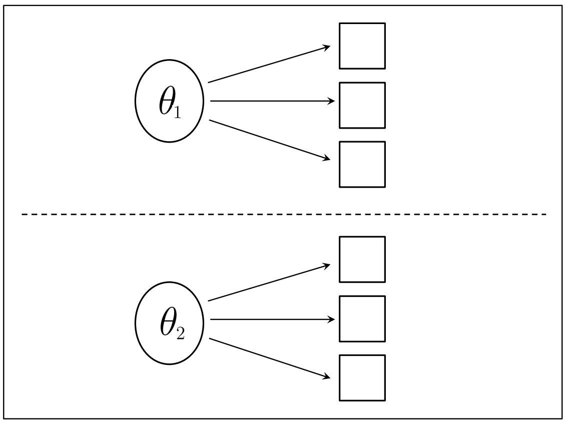


Figure 2.1. Consecutive unidimensional latent structure.

Multiunidimensional model Figure 2.2 reports the underlying structure for the between-item MIRT model (Wang et al., 2004), also called *multiunidimensional* approach (Sheng and Wikle, 2007), where abilities are allowed to correlate and the intensity of such associations can be obtained directly.

Bi-factor model The well known *bi-factor* model, first introduced by Holzinger and Swineford (1937), where a general (or common) ability, θ_0, and a specific ability are assumed to affect the response to each item, is illustrated in Figure 2.3. This is a case where there is within-item multidimensionality, i.e. single

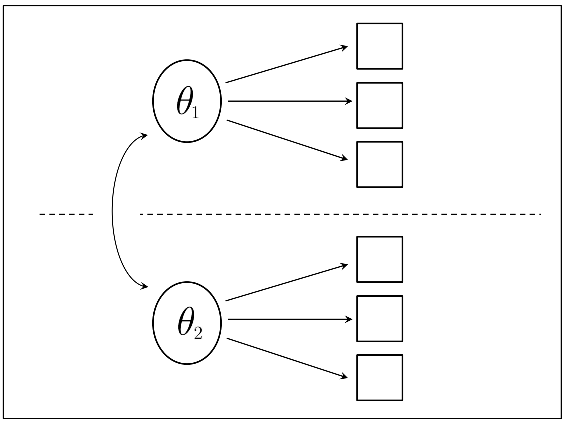


Figure 2.2. Multiunidimensional latent structure.

items measure more than one latent trait. This approach ignores the association between latent abilities.

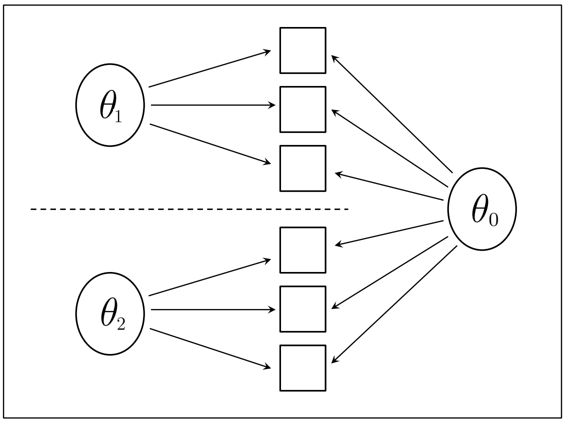


Figure 2.3. Bi-factor latent structure.

Hierarchical models Figure 2.4 shows the latent structure assumed for MIRT *hierarchical* models, where the hierarchical structure in general and specific latent constructs is modeled explicitly: items in the same subtest measure a specific ability and, in turn, each specific ability is influenced by a general ability. Different hierarchical models can be specified depending on the relation between specific and overall abilities: if each specific ability is a linear function of the overall ability we are in the case illustrated in **(a)**, while if each specific ability linearly combines to form the overall ability we are in the case showed by **(b)** (Schmid and Leiman, 1957; Sheng and Wikle, 2008).

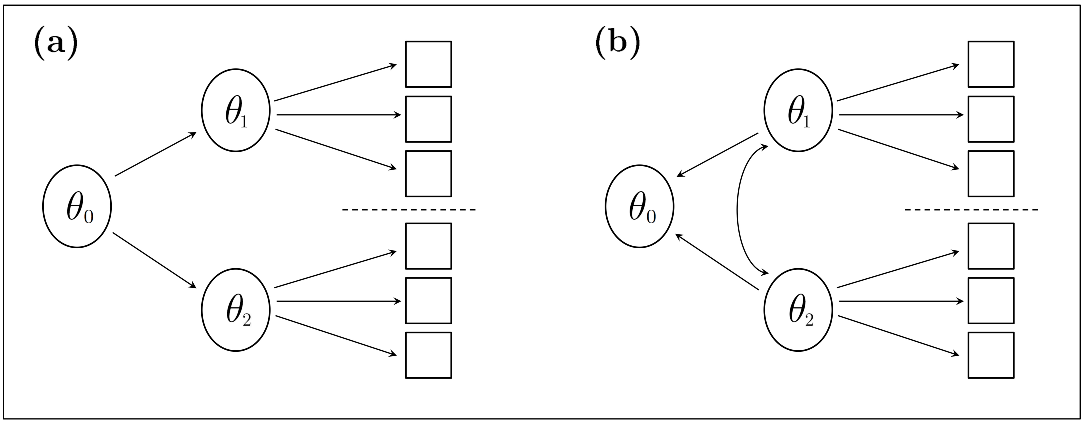


Figure 2.4. Hierarchical latent structures.

Additive model In the *additive* model presented in Figure 2.5 the latent structure is such that the response to a test item is affected both by the general and the specific latent traits, so that the latent abilities form an additive structure (Sheng and Wikle, 2009). This model has a latent structure similar to the bi-factor model, but here all the latent constructs are allowed to correlate.

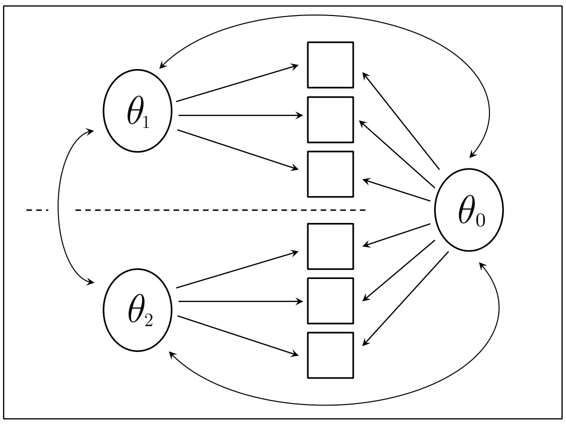


Figure 2.5. Additive latent structure.

2.2 MIRT models for binary data

MIRT is a methodology that has been developed with the principal aim of dealing with the situation of complexity in psychological measurement when several latent abilities influence the individual's performance on a given item (Reckase, 1997). By introducing a person trait and item discrimination parameters for each ability measured by a test item, MIRT models permit separate inferences with reference to each distinct latent dimension of an examinee (Ackerman, 1993).

Two parameter normal ogive model for binary data

Let consider a test consists of p multiple choice items, each measuring m latent abilities, $\theta_{1i}, \ldots, \theta_{mi}$. Let $\boldsymbol{Y} = [Y_{ij}]_{n \times p}$ represents the data matrix, i.e. a matrix containing n examinees' responses to p binary items, so that, for $i = 1, \ldots, n$ and $j = 1, \ldots, p$, Y_{ij} is defined as:

$$Y_{ij} = \begin{cases} 1, & \text{if examinee } i \text{ answers item } j \text{ correctly} \\ 0, & \text{if examinee } i \text{ answers item } j \text{ incorrectly.} \end{cases}$$

Reckase (1985) derived a multidimensional extension of the compensatory unidimensional two-parameter model, that in its normal ogive formulation becomes:

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \boldsymbol{\alpha}_j, \beta_j) = \Phi\left(\sum_{\nu=1}^{m} \alpha_{\nu j}\theta_{\nu i} - \beta_j\right) = \\ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sum_{\nu=1}^{m} \alpha_{\nu j}\theta_{\nu i} - \beta_j} e^{-t^2/2}\, dt\ . \quad (2.1)$$

Each individual is characterized by a vector $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})$ of latent abilities, where m is the number of latent dimensions measured by a generic item, in contrast to the unidimensional case, where they are classified by only one latent ability θ_i.

Item discrimination parameters are also represented by a vector, reflecting multiple dimensions: $\boldsymbol{\alpha}_j = (\alpha_{1j}, \ldots, \alpha_{mj})$, where j represents the item number and m shows the dimension to which the discrimination value is related. If the discrimination parameter related to dimension ν, $\alpha_{\nu j}$, is high, it means that such dimension has a great influence in determining an examinee's success on item j. Finally, β_j is a scalar parameter determining the location in the latent space where the item provides maximum information.

Multiunidimensional model for binary data

As illustrated in the work of Sheng and Wikle (2007), the elements in the vector of discrimination parameters $\boldsymbol{\alpha}_j = (\alpha_{1j}, \ldots, \alpha_{mj})$ can be considered as factor loadings in factor analysis. If a rotation is performed so that each item loads on one factor only, the vector of discrimination parameters can be simplified to $\boldsymbol{\alpha}_j = (0, \ldots, 0, \alpha_{\nu j}, 0, \ldots, 0)$, and we can get the expression for the multiunidimensional model for binary data, where each latent trait is related to a single set of items, from (2.1). The underlying latent structure of such model is illustrated in Figure 2.2. Let consider a test consisting of p items. The test is structured into m subtests, each one composed by p_ν items that measure one latent trait. The probability that the individual i will obtain a correct response to item j belonging to the ν-th subtest is given by:

$$P(Y_{\nu ij} = 1|\theta_{\nu i}, \alpha_{\nu j}, \beta_j) = \Phi\left(\alpha_{\nu j}\theta_{\nu i} - \beta_j\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\alpha_{\nu j}\theta_{\nu i}-\beta_j} e^{-t^2/2}\, dt \ ,$$

where $\alpha_{\nu j}$ is a scalar parameter reflecting the item discrimination, $\theta_{\nu i}$ is a scalar parameter reflecting the individual's ν-th ability, and β_j is a scalar parameter representing the location in the latent space where the item provides maximum information.

Additive model for binary data

The additive MIRT model for dichotomous data proposed by Sheng and Wikle (2009) assumes an underlying latent structure such that both specific abilities and an overall ability affect directly the individual response to a test item, resulting in an additive structure (see Figure 2.5).

If we consider again a test containing p items structured into m subtests (each one composed by p_ν items), according to the additive MIRT model for binary data, the probability that the individual i will obtain a correct response to item j belonging to the ν-th subtest is given by:

$$P(Y_{\nu ij} = 1|\theta_{0i}, \theta_{\nu i}, \alpha_{0\nu j}, \alpha_{\nu j}, \beta_j) = \\ = \Phi(\alpha_{0\nu j}\theta_{0i} + \alpha_{\nu j}\theta_{\nu i} - \beta_j) = \int_{-\infty}^{\alpha_{0\nu j}\theta_{0i}+\alpha_{\nu j}\theta_{\nu i}-\beta_j} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \ , \quad (2.2)$$

where $\theta_{\nu i}$ is a scalar parameter representing the examinee's ability in the ν-th dimension, θ_{0i} is the i-th individual parameter related to the overall ability, $\alpha_{0\nu j}$ is the j-th item discrimination parameter with reference to the overall ability θ_{0i}, $\alpha_{\nu j}$ is the item discrimination parameter with reference to the specific ability $\theta_{\nu i}$, and β_j is a scalar parameter representing the location in the latent space where the item provides maximum information.

The expression in (2.2) implies that the probability that an individual endorses an item is directly influenced by two latent traits: a general ability and a specific one (Sheng and Wikle, 2009).

A more detailed description of the models for binary data presented above goes beyond the purpose of this study. Our decision to focus the analysis on the additive structure has been driven by the fact that this latent structure, according to which both the specific and general latent traits directly underlie all the test items, represents a plausible and fairly detailed approximation of the real interactions between individuals and item responses. On the other hand, the multiunidimensional model is simpler than the additive, but it is regularly used in MIRT applications.

The exposition of these two models has been done in order to furnish a more complete background on the latent structures that we will discuss in detail for the case of ordered responses.

2.3 MIRT models for ordinal data

A multidimensional formalization of IRT models for graded responses has been developed as an extension of the unidimensional version by several authors. In this section we present some works that focus on multidimensional models for ordered items. These works have not necessary developed in an IRT context, but also in the framework of confirmatory factor analysis. Basically, the interest in adopting such models raised to face the widespread use of Likert items (Likert, 1932), and in general other ordered scales, on questionnaires in sociological and psychological measurement. The extensive availability of such data has led, in the last two decades, to the need of new progressions towards a multidimensional version of IRT model for graded responses.

We begin by introducing some notation. As in the case for dichotomous items, let consider a test made by p multiple choice items, each measuring m latent traits, $\theta_1, \ldots, \theta_m$. Now the data are collected in a matrix, $\boldsymbol{Y} = [Y_{ij}]_{n \times p}$, containing n examinees' responses to p ordered items, thus, for $i = 1, \ldots, n$ and $j = 1, \ldots, p$, Y_{ij} is defined as:

$$Y_{ij} = \begin{cases} 1, & \text{if the answer of examinee } i \text{ to item } j \text{ falls in category } 1 \\ 2, & \text{if the answer of examinee } i \text{ to item } j \text{ falls in category } 2 \\ \vdots & \quad \vdots \\ K_j, & \text{if the answer of examinee } i \text{ to item } j \text{ falls in category } K_j \end{cases}$$

where 1 and K_j are the lowest and the highest score for item j, respectively.

Muraki and Carlson (1995) developed a MIRT model for polytomously scored items on the basis of Samejima's graded response model in the full information factor analysis context. In their work, they show how the factor analytic model for categorical variables is based on the assumption that the response process, say Z_{ij}, is an underlying not observable variable and, for each subject i, realized into the vector of observed ordered item responses $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ip})$. They also model the response process variable Z_{ij} as a linear combination of the m

latent traits, $\theta_{1i}, \theta_{2i}, \ldots, \theta_{mi}$, and the factor loadings $\alpha_{j1}, \alpha_{j2} \ldots, \alpha_{jm}$. Thus:

$$Z_{ij} = \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + \cdots + \alpha_{jm}\theta_{mi} + \varepsilon_{ij} = \boldsymbol{\alpha}'_j\boldsymbol{\theta}_i + \varepsilon_{ij} \ ,$$

where ε_{ij} is an unobserved random variable that is assumed to be distributed as $N\left(0, \sigma_j^2\right)$. Muraki and Carlson (1995) introduced the threshold parameter γ_{jk} associated with the k-th category of item j, and modeled the unobservable response process according to the psychological mechanism, that is $Y_{ij} = k$ if $\gamma_{j,k-1} \leq Z_{ij} < \gamma_{jk}$, for $k = 1, \ldots, K_j$, $\gamma_{j0} = -\infty$, and $\gamma_{jK_j} = +\infty$. The probability to get the response category k of item j by examinee i, given the examinee's m-dimensional latent trait and assuming a normal ogive model, is formalized as:

$$P(Y_{ij} = k|\boldsymbol{\theta}_i) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_j} \int\limits_{\gamma_{j,k-1}}^{\gamma_{jk}} \exp\left\{-\frac{1}{2}\left(\frac{Z_{ij} - \boldsymbol{\alpha}'_j\boldsymbol{\theta}_i}{\sigma_j}\right)^2\right\} dZ \ . \tag{2.3}$$

Model (2.3) can be rewritten in a more familiar way with item response models, by applying some transformation of the variables (see Muraki and Carlson (1995) for the detailed procedure). The authors focus on uncorrelated latent dimensions (bi-factor latent structure) and furnish a detailed procedure of the Expectation Maximization (EM) algorithm in a marginal maximum likelihood estimation context (the matter of estimation methods will be covered in the next section). The proposed algorithm has been implemented in the POLYFACT computer program (Muraki, 1993), which calculates the factor loadings via the principal factor method adopted to the product-moment correlation matrix. The program treats the observed responses as continuous variables (Muraki and Carlson, 1995).

In the study by Ferrando (1999) a comparison between three different item response models for graded responses has been made, focusing on a continuous response model based on linear factor analysis, a censored response model, where the graded responses are considered to be censored continuous variables, and a multidimensional graded response model in the formulation given by Muraki and Carlson (1995). They observed that, even though there have been several applications of the unidimensional graded response model to attitude and personality data, applications of the multidimensional version of the model are not common.

Ferrando (1999) concludes showing that the solutions were similar for the three models considered, but that the estimation method could affect the results.

A more recent work by Edwards (2010a) falls within the context of confirmatory item factor analysis models. He developed a relatively user friendly package, MultiNorm (Edwards, 2010b), where the user can fit multidimensional graded (or dichotomous) response models characterized by a multiunidimensional or a bi-factor underlying latent structure. The estimation technique used in this work belong to the Markov chain Monte Carlo (MCMC) techinques. Again, for a further discussion on estimation methods see the next section.

Other applications of MIRT models for graded responses, with empirical examples regarding mainly the field of educational assessment and the psychological reactance, can be found in Yao and Schwarz (2006), Fu et al. (2010), Brown et al. (2011) and van der Ark et al. (2011). It is worth to remark that the latent structures assumed in these studies were prevalently the multiunidimensional structure (Figure 2.2) and the bi-factor structure (Figure 2.3).

Considering this scarcity of existing research about MIRT models for ordinal outcomes, especially for complex cases, in this work we take into consideration the one represented by an additive underlying latent structure (Figure 2.5), after having introduced the multiunidimensional case (Figure 2.2).

2.4 Estimation methods

In IRT models, as well as in MIRT models, the characteristics of interest are the person's abilities and the item parameters: different values of these parameters lead to different response probability. Nevertheless, these two important characteristics are both unknown and the available data are represented only by a collection of responses given by a sample of examinees.

Concerning the estimation procedure, we need to consider two relevant features: the first one is that the response model is not linear and the second one is that is not possible to observe the latent trait θ. It implies that the estimation is similar to perform a nonlinear regression with unknown predictor values.

Starting from the available data, the focal objective is in the determination of the θ values for every individual and the item parameters from the item responses. We can perform a simultaneous estimation of ability and item parameters in a context of the maximum likelihood (ML) or in a Bayesian framework.

The estimation procedure is in general affected by the way the probabilities of the responses are theorized. There are two main interpretations of probability: one of them is the *stochastic subject* interpretation, where the observed examinees are considered as fixed and probabilities reflect the unpredictability of specific events. Here the latent variables are constructed as unknown fixed parameters. The other interpretation of probability is the *random sampling*, where the examinees are considered as a representative random sample from a population, so that it raises the needing to specify a specific distribution of the latent trait and the latent variables are constructed as random.

In the framework of ML estimation, three main methods can be identified:

- The joint maximum likelihood (JML);
- The conditional maximum likelihood (CML);
- The marginal maximum likelihood (MML).

In the JML and CML methods we are in the context of the stochastic subjects interpretation of probability, i.e. fixed latent variables, whereas in the MML method we are in the random sampling interpretation framework and the latent variables are treated as random.

The applicability of JML and CML is pretty limited. The JML method works by simultaneously estimating item and person parameter through an iterative procedure. This method is quite simple but the complexity of the algorithm increases with the number of observations. The standard limit theorems do not apply and the resulting parameter estimators are not consistent (Andersen, 1970). The CML was a method suggested by Andersen (1970) and based on the availability of a sufficient statistic for the ability in order to simplify the maximum likelihood conditioning on it. There is a relevant problem which limits the applicability of such method: most models, including the quite simple unidimensional two parameter model, do not have simple sufficient statistics (Johnson, 2007).

The MML estimation method is the most widely applied and, by considering the joint probability of a certain response pattern given the latent trait and integrating out of the individual likelihoods, it defines the marginal probability of observing the item response pattern. To obtain the parameters estimates, the EM algorithm is used (Ayala, 2009).

A single estimated latent trait value can be associated to each individual through maximum a posteriori or expected a posteriori techniques. In general, all the ML estimation methods consider fixed item parameters. Conversely, in the Bayesian context, both the latent abilities and the item parameters are regarded as random variables.

As we will see later more in detail in the next chapter, the adoption of a fully Bayesian approach implies several advantages. It allows a joint estimation of item parameters and individual abilities and it permits to include uncertainties about item parameters and abilities, and in general prior beliefs, in the prior distributions. MCMC estimation of IRT and MIRT models can be then viewed as an alternative to MML estimation, where the approximation of multiple integrals involved in the likelihood function, especially for increasingly complex models, may represent a serious problem.

Chapter 3

Bayesian estimation of MIRT models

This chapter introduces the main ideas and functioning characterizing the Bayesian approach for estimation purposes, with a particular focus on the simulation-based methods for parameter estimation. Available Bayesian estimation methods based on MCMC techniques for MIRT models are also presented.

3.1 Elements of Bayesian statistics in MIRT context

According to the Bayesian approach, all the model parameters, i.e. person and item parameters in our case, are random variables, each one with its prior distribution reflecting the prior information available and the uncertainty about their real values before the observation of the data.

All the MIRT models so far illustrated (for both binary and ordinal items) are specified with the final aim to express the data-generating process as a function of the unknown person and item parameters. These are likelihood models and present the density of the data conditional on the model parameters. In order to formulate a Bayesian model, we need to specify:

- A prior distribution for each unknown model parameter;
- A likelihood model reflecting the data-generating process.

Once the data are observed, the prior information is updated with the information contained in the observed data and a posterior distribution is made, which permits to perform direct inference about parameters.

3.1.1 Prior distribution choice

A key point in Bayesian framework is the possibility to specify prior distributions for the unknown model parameters with the aim to exploit background information and beliefs available before the collection of the sample. All these context information are expressed as probability distributions and, as a result, are reflected in a prior distribution. On the other hand, the conditional probability distribution is specified to reflect the observed data.

One of the main objection to the Bayesian framework regards the specification of these prior distributions, that can be considered extensively subjective and arbitrary (Gelman, 2008). It has to be noticed that the choice of the prior distributions, made at the moment of model specification, is subjective by definition.

Therefore, only prior distributions expressing prior ideas can be considered correct in this setting and, even if the choice is subjective, it cannot be considered arbitrary since it reflects the researcher's thought (Fox, 2010). In addition, it is possible to specify the so called vague priors, that are objective non informative prior distributions indicating ignorance around the unknown parameter values.

The branch of objective Bayesian statistics rely on the specification of objective prior distributions. Even though it does not need any subjective contribution, we have to consider that a specific point of strength of Bayesian methodology is the possibility of including beliefs and prior information in model specification, and objective Bayesian methods do not allow to do that.

The inclusion of prior beliefs can increase the reliability of the statistical inference. In IRT and MIRT frameworks, item responses represent the observed data and we can include other sources of information in the model through the a priori model. These are circumstances where data-based information is slight, and where prior information can significantly improve the statistical inference (Fox, 2010).

3.1.2 Bayes' Theorem

Let consider a set of N observations, denoted by $\boldsymbol{y} = (y_1, \dots, y_N)$, that are the numerical realization of the random vector $\boldsymbol{Y} = (Y_1, \dots, Y_N)$, which follows some probability distribution. Let denote with $p(\boldsymbol{y})$ the probability density (mass) function of the continuous (discrete) variable $\boldsymbol{Y}$.

Now let assume that, starting from the observed responses, we are interested in measuring the unknown person ($\boldsymbol{\theta}$) and item ($\boldsymbol{\xi}$) parameters, denoted by $\boldsymbol{\lambda} = (\boldsymbol{\theta}, \boldsymbol{\xi})$. We denote with $p(\boldsymbol{\lambda})$ the prior distribution reflecting the beliefs on unknown parameters. The term $p(\boldsymbol{y}|\boldsymbol{\lambda})$ reflects the information about $\boldsymbol{\lambda}$ from the vector of observed values $\boldsymbol{y}$. In general, we can be interested in the sampling distribution and the likelihood function if we consider it as a function of the data or as a function of the parameters, respectively. Usually, the distribution of the parameters given the data is of main interest. According to the Bayes' Theorem, the conditional distribution of $\boldsymbol{\lambda}$ given the response data is

$$p(\boldsymbol{\lambda}|\boldsymbol{y}) \;=\; \frac{p(\boldsymbol{y}|\boldsymbol{\lambda})\, p(\boldsymbol{\lambda})}{p(\boldsymbol{y})} \;\propto\; p(\boldsymbol{y}|\boldsymbol{\lambda})\, p(\boldsymbol{\lambda}) \;, \tag{3.1}$$

where $\propto$ denotes proportionality. The term $p(\boldsymbol{\lambda}|\boldsymbol{y})$ is the posterior density of the parameter $\boldsymbol{\lambda}$ given both prior and sample information and, for continuous quantities, $p(\boldsymbol{y}) = \int_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} p(\boldsymbol{y}|\boldsymbol{\lambda}) p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$ (where $\boldsymbol{\Lambda}$ denotes the set of all the possible values of $\boldsymbol{\lambda}$).

Since we are interested in person and item parameters, we replace expression (3.1) with

$$p(\boldsymbol{\theta}, \boldsymbol{\xi}|\boldsymbol{y}) \;=\; \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})\, p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})}{p(\boldsymbol{y})} \;\propto\; p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})\, p(\boldsymbol{\theta})\, p(\boldsymbol{\xi}) \;, \tag{3.2}$$

where $p(\boldsymbol{\theta})$ is the prior for person parameters $\boldsymbol{\theta}$, $p(\boldsymbol{\xi})$ is the prior for item parameters $\boldsymbol{\xi}$ and these prior densities are assumed to be independent from each other, thus $p(\boldsymbol{\theta}, \boldsymbol{\xi}) = p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})$.

The denominator of expression (3.2) is called the data marginal density, marginal likelihood, or integrated likelihood. Its evaluation can be a time costly process, so that, when the knowledge of the shape of the posterior $p(\boldsymbol{\theta}, \boldsymbol{\xi}|\boldsymbol{y})$ is enough for the study purposes, we can focus on the unnormalized density function: $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\theta})p(\boldsymbol{\xi})$ (Fox, 2010).

The statement of the well-known Bayes' Theorem (Bayes and Price, 1763) is represented by the expression reported in (3.2). In particular, the expression $p(\boldsymbol{\theta}, \boldsymbol{\xi}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})\; p(\boldsymbol{\theta})\; p(\boldsymbol{\xi})$ is a factorization representing the product of the likelihood $\mathcal{L}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\xi})$ and the prior density, as typically $\mathcal{L}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\xi}) = p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})$. All the sample information regarding person and item parameters is contained in this likelihood function.

A relevant distribution for the inference process is the so called joint posterior density $p(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\xi})$. This density can be factorized as follow:

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\xi}) &= p(\boldsymbol{\theta}, \boldsymbol{\xi}|\boldsymbol{y})\; p(\boldsymbol{y}) && (3.3)\\ &= p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})\; p(\boldsymbol{\theta})\; p(\boldsymbol{\xi})\;. && (3.4) \end{aligned}$$

From the expressions above we can observe that the joint posterior distribution can be factorized in two different ways: *(i)* as the marginal density of the data and the posterior of the unknown parameters (3.3), and *(ii)* as the prior distributions of the parameters and the likelihood of $(\boldsymbol{\theta}, \boldsymbol{\xi})$ given $\boldsymbol{y}$ (3.4).

3.1.3 Marginal posterior distributions for model parameters

In order to make inference, the joint posterior distribution reported in (3.2) is used. Since this high-dimensional distribution has a complex form, and consequently it usually shows an analytically intractable expression, we need to focus on one of the unknown parameters, and consider the other as a nuisance parameter.

More precisely, if we are interested in the distribution of $\boldsymbol{\theta}$, we assume $\boldsymbol{\xi}$ as a nuisance parameter and, integrating out all the possible values of $\boldsymbol{\xi}$, from (3.2) we obtain the marginal posterior density for person parameters:

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y}) &= \int_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} p(\boldsymbol{\theta},\boldsymbol{\xi}|\boldsymbol{y})\, d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\, p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})}{p(\boldsymbol{y})}\, d\boldsymbol{\xi} \\ &\propto \int_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\, p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})\, d\boldsymbol{\xi}\,. \end{aligned} \tag{3.5}$$

When we are interested in the distribution of $\boldsymbol{\xi}$, we consider $\boldsymbol{\theta}$ as a nuisance parameter and thus we integrate out all the values of $\boldsymbol{\theta}$, getting the marginal posterior density for item parameters:

$$\begin{aligned} p(\boldsymbol{\xi}|\boldsymbol{y}) &= \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} p(\boldsymbol{\theta},\boldsymbol{\xi}|\boldsymbol{y})\, d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\, p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})}{p(\boldsymbol{y})}\, d\boldsymbol{\theta} \\ &\propto \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\, p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})\, d\boldsymbol{\theta}\,. \end{aligned} \tag{3.6}$$

In general, the information contained in the joint and/or marginal posterior distributions are summarized by the posterior mean (median) and standard deviation. Concerning the joint posterior distribution of person and item parameters, as previously pointed out, several difficulties arise as a result of its high-dimensionality and analytical intractability. Nonetheless, with reference to the marginal posterior densities of person (3.5) and item (3.6) parameters, the same difficulties remain, as the mathematical expressions are not always known.

These computational problems can be solved by the use of simulation based techniques. In particular, the MCMC method is a very useful technique that we will be briefly describe in the next section.

3.2 Markov chain Monte Carlo methods

The Bayesian approach based on MCMC techniques has increased its popularity in the estimation of unidimensional and multidimensional item response models.

A twofold motivation can drive the use of such method. First of all, it can represent an effective substitute to the classical EM algorithm implemented in the MML estimation. In fact, it works with simulation and introduces an informative prior distribution in the estimation process and, unlike the MML method, the Bayesian approach considers both the person parameters and item parameters as random variables. Secondly, it can also be seen as a compensatory instrument to the EM algorithm. The posterior distribution generated through the MCMC techniques can be used to evaluate the suitability of the normal approximations in the MML, so that we can compare the two approaches with reference to the accuracy of parameter recovery.

As we will see in this section, MCMC is a very useful and relatively straightforward method to make inference when we have to face with a very complex model, where it is actually difficult to sample or directly simulate from the posterior distribution. This represents a common situation in a MIRT context.

In particular, the Gibbs sampler is a widely used MCMC algorithm consisting in a quite precise scheme to create suitable samples from the posterior density. Moreover, this method is not very constraining and fairly simple to implement, if compared with other methods. For the motivations mentioned above, MCMC strategies have been implemented in IRT background by several researchers and many studies have been made in order to investigate the properties of these methods. Of particular interest is also the evaluation of model parameter recovery in comparison with the classical methods.

If we perform a comparison between the MCMC technique and the classical MML estimation, we can summarize the main advantages of the MCMC approach in:

- the flexibility regarding the modeling of all the connections between latent and observed variables;
- the appropriateness for more complex models;
- the non-sensitivity to the choice of starting values (unlike the EM algorithm).

Two relevant works that perform Bayesian estimation using the Gibbs sampler in an IRT context are the works of Albert (1992) and Béguin and Glas (2001). In the first one, the Gibbs sampler for the unidimensional two parameter model for binary data is implemented, and the MCMC algorithm is compared with the

EM algorithm through an application in the educational assessment context. In the second work, an extension to the multidimensional case has been done with respect to the work of Albert. Other item response applications of MCMC can be found in Fox and Glas (2001); Patz and Junker (1999a,b).

As previously highlighted, from a Bayesian point of view, the leading purpose of the researcher is to analyze the properties of the posterior distribution $p(\boldsymbol{\lambda}|\boldsymbol{y})$ which, as we can see from (3.1), is proportional to the product between the likelihood function and the prior distribution (recall that $\boldsymbol{\lambda}$ is the vector representing the unknown parameters of interest, and $\boldsymbol{y}$ represents the observed data).

For exposition purpose and without loss of generality, in this section we will consider the simplest case where the vector of unknown parameters is unidimensional, namely $\boldsymbol{\lambda} = \lambda$. When the posterior distribution does not have a familiar functional form and/or it is not possible to perform a direct simulation because of the complexity of the model, simulation methods based on Markov chains seem to be an easy way to get samples from the posterior density $p(\lambda|\boldsymbol{y})$.

The MCMC is a class of techniques developed with the final aim of reproducing a target distribution by simulating one or more sequences of correlated random variables. In our context the target distribution is represented by the posterior density $p(\lambda|\boldsymbol{y})$. A random walk in the space of the parameter λ is simulated by the MCMC algorithm, where at each iteration t, for $t = 1, \ldots, T$, the value of $\lambda^{[t]}$ is drawn from a probability function which depends on the value of λ at the previous step, $\lambda^{[t-1]}$. The underlying idea is that the regions of the state space are touched by the random walk in a proportional way with respect to their posterior probabilities and, for a sufficiently large number of iterations, it might approximate the target distribution. MCMC methods differ from the Monte Carlo methods because the simulated values are correlated, rather than being statistically independent. The generated Markov chain converges to an unique and stationary distribution that corresponds to the target distribution (Gelman et al., 2003). Therefore, with reference to the reproduction of the marginal posterior densities of IRT model parameters with a complex structure, this method is able to furnish reliable results, and overtakes the problem of analytically intractable distributions.

One of the key point concerning all the MCMC techniques is the creation of a chain sufficiently long to approximate the target distribution. Considering that we are in the context of iterative based methods, the time of convergence also represents a relevant topic. Usually, a so called "burn-in" period, containing a fixed number of first iterations, is defined and excluded from the analysis.

The chain length is affected by the complexity of the posterior distribution, the initial values and the speed of convergence. Gelman et al. (2003) recommend to use half of the sample as burn-in period. On the contrary, other authors prefer to directly choose the number of iterations as burn-in period, for example in one of the analyses illustrated in Béguin and Glas (2001), the burn-in period is of 1000 iterations against a run length of 30000 iterations.

What we suggest from a practical point of view is to control the behavior of the sampled parameters through a plot in the sequence of iterations, and then decide subsequently.

Moreover, another significant (but still not very clear, as illustrated in Gilks et al. (1996)) topic concerns the number of distinct chains needed to implement the MCMC algorithm. Mainly, there are three different approaches. According to the first one, only one long chain is created, considering that the longer the chain is, the higher the possibility to find new modes is.

The second approach is based on the creation of several quite long chains. The main advantage of this approach is that multiple chains allow the comparison between the results, that can permit to detect some significant differences and symptom of non-stationarity.

The use of the third approach, consisting of the utilization of many short chains, is driven by the aim of creating independent samples. Actually, this approach is not advisable because chains can take a long time to reach the convergence and independent samples are not required.

Several MCMC algorithm exist, depending on the features of the problem and the specific attributes of the Markov chains. Each MCMC algorithm defines a transition distribution $p(\lambda^{[t]}|\lambda^{[t-1]})$, representing the probability of a parameter, say λ, to move from a state to the following, starting from a proper initial values $\lambda^{[0]}$.

Examples of detailed essays about MCMC are Gelman et al. (2003), Gamerman (1997) and Gilks et al. (1996).

3.2.1 Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm (Hastings, 1970) is one of the most popular MCMC mehtods and it can be directly implemented in a Bayesian framework. Our aim is the generation of a sample of size T from the target distribution represented, in our context, by the posterior distribution $p(\lambda|\boldsymbol{y})$. We can summarize the M-H algorithm functioning in the following way (Ntzoufras, 2011):

1. Set initial values $\lambda^{[0]}$;

2. Then reiterate the following steps for $t = 1, \ldots, T$:

   (i) Set $\lambda = \lambda^{[t-1]}$

   (ii) Generate a new candidate parameter value λ' from a proposal (jumping) distribution $q(\lambda'|\lambda)$

   (iii) Calculate the ratio $\alpha = min\left(1, \frac{p(\lambda'|\boldsymbol{y})q(\lambda|\lambda')}{p(\lambda|\boldsymbol{y})q(\lambda'|\lambda)}\right)$

   (iv) Update $\lambda^{[t]} = \lambda'$ with probability α; otherwise set $\lambda^{[t]} = \lambda$.

Let focus on the case where λ is a vector of parameters that can assume only continuous values.

According to step (i), suitable starting values have been provided. Let suppose to be in the state $\lambda^{[s-1]}$ of the chain.

In the step (ii) of the algorithm, a new candidate λ' is sampled by using a "proposal distribution" $q(\lambda'|\lambda^{[s-1]})$. The proposal distribution is also called "jumping distribution", in order to emphasize the concept of movement from the current value to the next one of the chain. It is also possible to define the probability of "jumping" in the opposite direction, i.e. from λ' to $\lambda^{[s-1]}$, that is $q(\lambda^{[s-1]}|\lambda')$ Even if in the original M-H algorithm (Metropolis et al., 1953) only symmetric proposals were considered, this property is not compulsory in the more recent versions of the algorithm (Ntzoufras, 2011).

Furthermore, the proposal $q(\cdot)$ should be defined in a proper way. In fact, the resulted chain needs to satisfy some specific characteristics, namely: irreducibility, aperiodicity and not transitoriness. A chain is irreducible if it is possible to move from one state to any other state in a finite number of steps with positive probability, aperiodic if all the states are acyclic, and not transient if all the states are recurrent (i.e. the probability to return to a state from the same state is equal to one). Moreover, the ratio $r = \frac{q(\lambda' | \lambda^{[s-1]})}{q(\lambda^{[s-1]} | \lambda')}$ must be strictly positive, for every value of λ such that both the numerator and the denominator are nonzero.

In the step (iii) the "acceptance probability" α is computed. The higher the α is, the more probable the acceptance of the candidate value λ' will be. The quantity r consists of two components: the ratio of the posterior probabilities, which drives the algorithm towards the λ-value with higher posterior density, and the ratio of the "proposal densities", which also has an influence in determining the direction to one or the other λ-value.

Step (iv) of the M-H algorithm is about the acceptance or the rejection of the candidate value λ'. To make this choice, we draw a random number u from the uniform distribution in the $[0, 1]$ interval. Then we set:

$$\lambda^{[s]} = \begin{cases} \lambda' , & \text{if} \quad u < \alpha \\ \lambda^{[s-1]} , & \text{if} \quad u \geq \alpha . \end{cases} \tag{3.7}$$

Thus, the candidate value λ' is accepted with probability α and rejected in case of $u \geq \alpha$. In both cases (acceptance or rejection) the iterations progress and the algorithm proceeds to generate the next value.

The M-H algorithm can be also applied in case of discrete-values parameters where the $q(\cdot)$ proposal distribution becomes the probability mass function used to generate candidate points.

3.2.2 Gibbs sampler

The Gibbs sampler was first introduced by Geman and Geman (1984) and then formalized by Gelfand and Smith (1990). It can be obtained as a special case of the M-H algorithm by using as a proposal distribution the so called "full conditional posterior distribution":

$$p(\lambda_j \mid \lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_d, \, \boldsymbol{y}) \;=\; p(\lambda_j \mid \boldsymbol{\lambda}_{*j}, \, \boldsymbol{y}). \tag{3.8}$$

Such proposal distribution implies a probability of acceptance α equal to one, due to the fact that the ratio r is 1 (see Gelman et al., 2003). With an acceptance probability equal to one, at each iteration the algorithm performs the jump provided by step (iv) in the M-H algorithm. The Gibbs sampler is based on iterative sampling of the conditional distributions resulting from the decomposition of the full posterior density.

A first advantage of the Gibbs sampler is that, for every iteration, the values are randomly generated from unidimensional distributions for which a wide variety of computational tools exists (Gilks et al., 1996). Another important advantage is that it does not require the specification of a proposal distribution. This is a key point, because an inaccurate choice of the proposal $q(\cdot)$ in the M-H algorithm may lead to a very slow algorithm.

Thus, if it is difficult to sample from a complex and/or high-parameterized posterior distribution and it is possible to decompose the vector of parameters, we can proceed to generate the parameter values from the single conditional distribution in a sequential way.

Let suppose that we are interested in producing a sample of size T from the target distribution, represented here by the posterior distribution $p(\boldsymbol{\lambda}|\boldsymbol{y})$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$. The functioning of the Gibbs sampler algorithm can be described with the following steps (Ntzoufras, 2011):

1. Set initial values $\lambda^{[0]}$;

2. Then reiterate the following steps for $t = 1, \ldots, T$:

   (i) Set $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{[t-1]}$

   (ii) For $j = 1, \ldots, p$, update λ_j from $\lambda_j \sim p(\lambda_j \mid \boldsymbol{\lambda}_{*j}, \boldsymbol{y})$

   (iii) Set $\boldsymbol{\lambda}^{[t]} = \boldsymbol{\lambda}$ and save it as the generated set of values at $t+1$ iteration of the algorithm.

Hence, given a particular state of the chain $\boldsymbol{\lambda}^{[t]}$, we generate the new parameter values by:

$$
\begin{array}{lll}
\lambda_1^{[t]} & \text{from} & p(\lambda_1 \mid \lambda_2^{[t-1]},\ \lambda_3^{[t-1]},\ \lambda_4^{[t-1]}, \ldots, \lambda_p^{[t-1]},\ \boldsymbol{y}) \\
\lambda_2^{[t]} & \text{from} & p(\lambda_2 \mid \lambda_1^{[t]},\ \lambda_3^{[t-1]},\ \lambda_4^{[t-1]}, \ldots, \lambda_p^{[t-1]},\ \boldsymbol{y}) \\
\lambda_3^{[t]} & \text{from} & p(\lambda_3 \mid \lambda_1^{[t]},\ \lambda_2^{[t]},\ \lambda_4^{[t-1]}, \ldots, \lambda_p^{[t-1]},\ \boldsymbol{y}) \\
\vdots & & \vdots \\
\lambda_p^{[t]} & \text{from} & p(\lambda_p \mid \lambda_1^{[t]},\ \lambda_2^{[t]},\ \lambda_4^{[t]}, \ldots, \lambda_{p-1}^{[t]},\ \boldsymbol{y}) \ .
\end{array}
$$

Generating values from the single conditional distributions is relatively easy, since those are univariate distributions. Moreover, under appropriate conditions of regularity, the $\boldsymbol{\lambda}^{[t]}$-distribution will converge to the target distribution. Usually, this convergence process is fast and the complete sequence $\{\boldsymbol{\lambda}^{[t]}\}$ can be considered as the simulated sample of the distribution of interest (Matteucci, 2007).

For a more detailed exposition of the Gibbs sampler, see Gamerman (1997) and Gelman et al. (2003), or Gelfand and Smith (1990) for early presentations of this widely used MCMC algorithm.

3.3 Bayesian computation using OpenBUGS

In the following, the simulation study and the application on real data will be performed using OpenBUGS (`http://www.openbugs.net`), an open-source version of the famous software package BUGS (Bayesian inference Using Gibbs Sampling) that permits an user-friendly implementation of the Gibbs sampler.

The software package BUGS was developed in the context of the BUGS project. The BUGS project started in 1989 in the MRC Biostatistic Unit in Cambridge and the last version of the resulting software developed by Spiegelhalter et al. (1996) became very popular in the 1990s. WinBUGS, an available windows-based version of BUGS, has finished to be further upgraded in 2012 hence OpenBUGS, which basically contains all the features of its ancestor WinBUGS, represents nowadays the future of the BUGS project.

A detailed description of the software goes beyond the scope of this work, nevertheless, useful tools to understand the theoretical ideas that are the foundations of BUGS and its functioning are the book of Ntzoufras (2011) and Lunn et al. (2013).

As we can find in Lunn et al. (2009), there are several reasons behind the success of the BUGS software. These appealing features can be strictly summarized in:

- **Flexibility**. Flexibility is quite probably the principal reason for BUGS's popularity. BUGS runs the Gibbs sampling method to any directed acyclic graph specified in its language, moreover it allows the user to add new distributions and functions.

- **Easy implementation**. The model implementation using BUGS is fairly simple because the package itself run the MCMC algorithm. It is not necessary for the user to write down all the full conditional distributions. Moreover, measures, plots and statistics to check the convergence and the fit of the model are automatically computed.

These aspects notwithstanding, the user must always be careful because BUGS does not perform any control about the model identification, thus several mistakes can be made without any alert from the program. As the manual clearly remark:

Gibbs sampling can be dangerous!

Chapter 4

MIRT graded response models with complex structures

In this chapter we specify two MIRT models for graded responses with a complex structure. After having established a dichotomization method, we focus on models with a multiunidimensional structure, where items in each subtest characterize a single ability, and on models with an additive structure, where each item measures a general and a specific ability directly. In the MIRT model presented, all the latent traits are allowed to correlate. The main scientific contribution of this work is the multidimensional additive model for graded responses with correlated traits, estimated with MCMC tecniques. Due to the adoption of Bayesian estimation methods, particular attention is paid to the model building phases.

4.1 MIRT graded response models (GRMs)

A multidimensional generalization of IRT graded response model (GRM) can be obtained from its unidimensional counterpart. Let consider: *(i)* n individuals; *(ii)* a set of p ordinal items where the response Y_{ij} of the i-th subject to the j-th item can take values in the set $\{1, \dots, K_j\}$. Each item thus has $K_j - 1$ thresholds $\kappa_{j1}, \dots, \kappa_{j,K_j-1}$ that have to satisfy the order constraint $\kappa_{j1} < \cdots < \kappa_{j,K_j-1}$;

and *(iii)* the existence of multiple, say m, latent abilities $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})'$ underlying the responses to the items.

For simplicity, in this paragraph we do not consider the number of latent dimensions, even if we have always to take in mind that $\boldsymbol{\theta}_i$ is a vector, so we are dealing with the presence of concomitant latent dimensions. The key point of the choice of the underlying latent structure will be examined more closely later.

Assumptions are quite similar to the unidimensional version of the model: it is assumed that an individual can reach a specific category level of an ordinal test item only if he/she is also able to reach all the lower categories on the same item. In other words, the item necessitates an amount of steps and the accomplishment of a step requires the achievement of the previous one. This type of model is then appropriate for rating scales where a rating category includes all previous categories (Reckase, 2009).

The notation introduced above implies that the lowest score on item j is 1 and the highest score is K_j. The probability that the i-th examinee will select the k-th category or higher on item j is assumed to increase monotonically with an increase in any component of the $\boldsymbol{\theta}_i$ vector, i.e. an increase in any of the latent abilities underlying the test.

We have used a dichotomization procedure by adapting Samejima's approach (see section 1.2.1): in order to make the implementation of the models more clear and easy, our models are specified on the basis of the probability that an item response will fall in *category k or lower*, denoted by P (while in section 1.2.1 we have used the probability that an item will fall in *category k or higher*, denoted by P^*). The probability π_{ijk} that the i-th subject will select the k-th category on item j is equal to the probability of answering below the upper boundary for the category (κ_k) minus the probability of answering below the category's lower boundary (κ_{k-1}). Figure 4.1 illustrates the dichotomization method used. The dashed line, that represents the hypothetical response, falls in category $k = 4$: the probability to observe a response in that category can be easily calculated as $P_{i4} - P_{i3}$.

Generalizing the example presented in Figure 4.1, the probability that the i-th examinee's response will fall in the k-th category on item j can be constructed from the cumulative probabilities $P_{ijk} = P(Y_{ij} \leq k|\boldsymbol{\theta}_i)$, for $k = 2, \ldots, K_j$. We

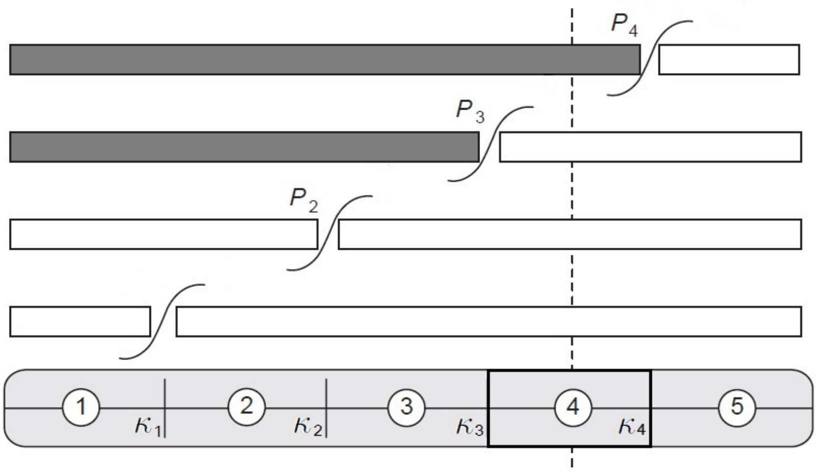


Figure 4.1. Dichotomization used for the MIRT graded response model specification. The dashed line indicates the observed category response.

obtain that:

$$\pi_{ijk} = P_{ijk} - P_{i,j,k-1} = P(Y_{ij} \leq k|\boldsymbol{\theta}_i) - P(Y_{ij} \leq k-1|\boldsymbol{\theta}_i) \ , \tag{4.1}$$

and, with the aim to guarantee that the probability of each category can be determined from (4.1), it is assumed that $\pi_{ij1} = P_{ij1} = P(Y_{ij} \leq 1|\boldsymbol{\theta}_i)$ and $\pi_{ijK_j} = 1 - P_{i,j,K_j-1} = 1 - P(Y_{ij} \leq K_j - 1|\boldsymbol{\theta}_i)$.

A normal ogive or a logistic formulation of the model can be obtained from expressions (1.2) and (1.3), but a previous step is needed to get an expression for the predictor η_{ij}. In the multidimensional case the predictor becomes a function of the $\boldsymbol{\theta}_i$ vector of person parameters and the $\boldsymbol{\xi}_j$ vector of item parameters, $\eta_{ij} = f(\boldsymbol{\theta}_i, \boldsymbol{\xi}_j)$. In particular, to have an explicit formulation for the predictor we need to make some assumptions reflecting the underlying latent structure hypothesized. Among the different underlying latent structures that can be assumed (see Paragraph 2.1.3), in this thesis we focus on:

- models with a **multiunidimensional** structure, where items in each subtest characterize a single ability;

- models with an **additive** structure, where each item measures a general and a specific ability directly.

As previously mentioned, the choice of these two latent structures has been driven by the fact that the first one is widely used and represents a classical approach in MIRT analysis, while the second one is able to reflect the complexity of real interactions between items and individuals.

4.1.1 Specification of the multiunidimensional GRM

As previously mentioned, according to the multiunidimensional structure, each individual i is assumed to be characterized by a vector of latent traits $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})$ where each latent dimension is measured by a specific set of test items. Thus, considering a test consisting of p items, the test is structured into m subtests indexed by ν, each one composed by p_ν items that measure one latent trait. The cumulative probability that the individual i will select the k-th category or lower on item j belonging to the ν-th subtest is given by:

$$P_{\nu ijk} = P(Y_{\nu ij} \leq k | \theta_{\nu i}, \alpha_{\nu j}, \kappa_{jk}) =$$
$$= \Phi(\kappa_{jk} - \alpha_{\nu j}\theta_{\nu i}) = \int\limits_{-\infty}^{\kappa_{jk}-\alpha_{\nu j}\theta_{\nu i}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \ , \quad (4.2)$$

where $\alpha_{\nu j}$ and κ_{jk} are item parameters representing the item discrimination and the threshold between categories k and $k+1$, respectively. The parameter $\theta_{\nu i}$ represents the i-th examinee ability in the ν-th ability dimension. We can observe that the predictor $\eta_{\nu ij} = f(\boldsymbol{\theta}_i, \boldsymbol{\xi}_j)$ assumes the form: $\eta_{\nu ij} = \kappa_{jk} - \alpha_{\nu j}\theta_{\nu i}$.

The multiunidimensional model for graded response can be specified in a normal ogive formulation *(i)* by considering the cumulative probabilities obtained from (4.2) and *(ii)* by applying the dichotomization procedure represented in Figure 4.1, according to which the probability $\pi_{\nu ijk}$ that the i-th examinee will select the k-th category on item j in subtest ν is:

$$\pi_{\nu ijk} = \begin{cases} P_{\nu ij1} & \text{for } k = 1 \\ P_{\nu ijk} - P_{\nu,i,j,k-1} & \text{for } k = 2, \ldots, K_j - 1 \\ 1 - P_{\nu,i,j,K_j-1} & \text{for } k = K_j \ . \end{cases} \quad (4.3)$$

It has to be noticed how in (4.2) only one specific ability affects the response to a specific item. This structure reminds the unidimensional version of the GRM: we can imagine to fit a sequence of unidimensional models, each one for a specific subtest. Nevertheless, a relevant difference consists in the fact that that distinct latent traits are now allowed to correlate.

4.1.2 Specification of the additive GRM

A relevant aim of this work is to propose a new additive model for ordinal data, estimated by Bayesian MCMC techniques, where the general and specific latent traits are allowed to correlate. In this section, we provide the simple, but very effective, specification for the additive GRM.

Let consider again a test consisting of p items and structured into m subtests, each one composed by p_ν items ($\nu = 1, \ldots, m$). The responses to items belonging to a specific subtest are assumed to be influenced by a specific ability and a general ability, according to the underlying latent structure illustrated in Figure 2.5. The cumulative probability that the individual i will select the k-th category or lower on item j belonging to the ν-th subtest is given by:

$$\begin{aligned} P_{\nu ijk} &= P(Y_{\nu ij} \leq k | \theta_{0i}, \theta_{\nu i}, \alpha_{0\nu j}, \alpha_{\nu j}, \kappa_{jk}) = \\ &= \Phi(\kappa_{jk} - \alpha_{0\nu j}\theta_{0i} - \alpha_{\nu j}\theta_{\nu i}) = \int\limits_{-\infty}^{\kappa_{jk} - \alpha_{0\nu j}\theta_{0i} - \alpha_{\nu j}\theta_{\nu i}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \ . \end{aligned} \quad (4.4)$$

Here, θ_{0i} represents the i-th overall ability and $\theta_{\nu i}$ represent the specific abilities (with $\nu = 1, \ldots, m$). For each item j of the subtest ν: $\alpha_{0\nu j}$ reflects the item discrimination with reference to the overall ability, $\alpha_{\nu j}$ reflects the item discrimination with reference to the specific ability and κ_{jk} is an item parameter that reflect the threshold between categories k and $k+1$. The predictor $\eta_{\nu ij}$ now depends on both specific and general latent traits: $\eta_{\nu ij} = \kappa_{jk} - \alpha_{0\nu j}\theta_{0i} - \alpha_{\nu j}\theta_{\nu i}$.

The probability $\pi_{\nu ijk}$ that the i-th examinee will select the k-th category on item j in subtest ν is obtained recursively from (4.3), as in the multiunidimensional GRM.

It has to be noticed that both general and specific abilities are involved in determining the response probability by following a compensatory approach. Finally, all the latent traits underlying the item responses are allowed to correlate.

4.2 Person and item parameters: interpretation

The aim of this section is to briefly illustrate the meaning of the parameters introduced in the MIRT models for graded responses described in paragraphs 4.1.1 and 4.1.2. Contents of this section are particularly helpful for practical applications.

4.2.1 Ability parameters

The presence of more than one latent trait affecting the response process to a test is on the basis of the use of multidimensional item response theory models. The $\boldsymbol{\theta}_i$-vector of the latent space parameters for person i contains all the information about the measurement of these latent abilities. Higher levels of abilities lead to higher values in the elements of $\boldsymbol{\theta}_i$.

Of course, the composition, and consequently the dimension, of the $\boldsymbol{\theta}_i$-vector depends on the underlying structure we are assuming. As often mentioned before, when we are dealing with a multiunidimensional structure the vector of person parameters has the form $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})$. While in an additive context, a parameter reflecting the general ability is added, and we get: $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i}, \ldots, \theta_{mi})$, where m still denotes the number of specific abilities. One lack in one specific dimension is compensated by the general dimension and viceversa.

4.2.2 Multidimensional item discrimination

Moving towards the significance of the discrimination item parameters, when considered individually, $\alpha_{\nu j}$ reflects the capability of a generic item j to discriminate between individuals with different levels of ability θ_ν, both for multiunidimensional and additive models. Analogously, $\alpha_{0\nu j}$ reproduces the aptitude of the item j to differentiate individuals with different levels of general ability θ_0.

Muraki and Carlson (1995) and Yao and Schwarz (2006) define the multidimensional item discrimination (MDISC) as the maximum discrimination of a test item in a particular direction of the latent space.

Hence, considering the multiunidimensional and additive latent structures assumed for the MIRT models presented in this work, we can define two MDISC measures. The first one (MDISC) is defined with reference to each one of the latent dimensions $\nu = 1, \ldots, m$. For $j = 1, \ldots, p$ it is expressed as:

$$\text{MDISC}_j = \left(\sum_{\nu=1}^{m} \alpha_{\nu j}^2 \right)^{1/2} . \tag{4.5}$$

The second one (MDISC*) include a further dimension, represented by the general ability. For $j = 1, \ldots, p$, MDISC* is expressed by:

$$\text{MDISC}_j^* = \left(\sum_{\nu=1}^{m} \alpha_{\nu j}^2 + \alpha_{0\nu j}^2 \right)^{1/2} . \tag{4.6}$$

With reference to a given item, the higher a value of MDISC (MDISC*) is, the grater is the discrimination power of that item, independently from the assumed underlying latent structure.

4.3 Multiunidimensional GRM implementation

In order to implement the multiunidimensional model for graded responses by using OpenBUGS, the first step that we have to face is the so called *model building phase* (Ntzoufras, 2011). We can summarize the functioning of this phase through several sub-steps, namely:

1. identify the main variable of interest and the corresponding (observed) data;

2. build a structure for the parameters of the distribution;

3. specify the prior distributions;

4. find a distribution that adequately describes the observed data and formulate the likelihood of the model.

Considering that our observed variables of interest (point 1) are the responses, given from a group of examinees, to a test consisting of graded response items, in this section we will define all the elements listed above, according to the model characterized by a multiunidimensional latent ability structure, i.e. according to the probability function defined in (4.2).

4.3.1 Model specification

The probability model is specified according to the multiunidimensional structure (point 2). Recalling the expression in (4.2), a generic cumulative probability $P_{\nu ijk}$ is a function of the item discrimination parameter ($\alpha_{\nu j}$), the threshold parameter between categories k and $k+1$ (κ_{jk}), and the specific ability measured by the j-th item ($\theta_{\nu i}$). Thus, for $\nu = 1, \ldots, m$, $j = 1, \ldots, p$ and $k = 1, \ldots, K_j - 1$, it holds that $P_{\nu ijk} = \Phi(\kappa_{jk} - \alpha_{\nu j}\theta_{\nu i})$.

As previously described, we set $P_{\nu ijK_j} = 1$, and we obtain by difference the probability that the response of individual i to item j will fall in category k: $\pi_{\nu ij1} = P_{\nu ij1}$ and $\pi_{\nu ijk} = P_{\nu ijk} - P_{\nu,i,j,k-1}$, for $\nu = 1, \ldots, m$, $j = 1, \ldots, p$ and $k = 2, \ldots, K_j$.

The model parameters, treated in a Bayesian context as proper random variables, for which we need to specify prior distributions are the person parameters $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})$ and the item parameters $\alpha_{\nu j}$ and $\kappa_{j1}, \ \ldots, \ \kappa_{j,K_j-1}$.

4.3.2 Prior distributions

Getting on to point 3 of the model building phase, in the multiunidimensional GRM we assume that the latent traits $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ are independent and multivariate normally distributed:

$$\boldsymbol{\theta}_i \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ ,$$

where $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})$ is the vector of latent traits for examinee i, $\boldsymbol{\mu}$ is the m-dimensional mean vector and $\boldsymbol{\Sigma}$ is the $m \times m$ constrained variance-covariance matrix with diagonal elements being 1 and off-diagonal elements being the ability correlations.

Thus, for $i = 1, \ldots, n$, the prior distribution for $\boldsymbol{\theta}_i$ is defined as:

$$p(\boldsymbol{\theta}_i) = \frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} \left(\boldsymbol{\theta}_i - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\theta}_i - \boldsymbol{\mu}\right) \right) , \qquad (4.7)$$

where m is the number of specific latent traits (subtests).

Moreover, normal distributions are assumed for item discrimination parameters, that is $\alpha_{\nu j} \sim N(\mu_\alpha, \sigma^2_\alpha)$, for $\nu = 1, \ldots, m$ and $j = 1, \ldots, p$. In addition, considering that the parameter which reflects the power of the item to discriminate between examinees is significantly positive, the truncated version of the normal distribution is taken into account:

$$\alpha_{\nu j} \sim N(\mu_\alpha, \sigma^2_\alpha)\ I(\alpha_{\nu j} > 0) ,$$

where I indicates the indicator function.

The priors for the threshold parameters must account for the order constraint $\kappa_{j1} < \cdots < \kappa_{j,K_j-1}$, hence we proceed first introducing unconstrained auxiliary parameters $\kappa^*_{j1}, \ldots, \kappa^*_{j,K_j-1}$ such that $\kappa^*_{jk} \sim N(\mu_\kappa, \sigma^2_\kappa)$ for $j = 1, \ldots, p$ and $k = 1, \ldots, K_j - 1$ (Curtis, 2010). Then, prior distributions on the thresholds for the j-th item can be obtained considering the order statistics for the auxiliary variables:

$$\begin{aligned} \kappa_{j1} &= \kappa^*_{j,[1]} \\ \kappa_{j2} &= \kappa^*_{j,[2]} \\ &\vdots \\ \kappa_{j,K_j-1} &= \kappa^*_{j,[K_j-1]} \ , \end{aligned}$$

where with $\kappa^*_{j,[s]}$ is denoted the s-th order statistic of $\kappa^*_{j1}, \ldots, \kappa^*_{j,K_j-1}$. As reported in Curtis (2010), this approach is also recommended by Plummer (2010).

Identification issues

Particular attention should be paid to the restrictions that have to be imposed on hyperparameters in order to ensure the model identification. In general, Bayesian item response models can be identified (Fox, 2010) by imposing restrictions on the hyperparameters or via a (standard) scale transformation in estimation procedure.

According to the first approach, for identification purposes we set $\boldsymbol{\mu} = \mathbf{0}$, $\mu_\alpha = 0$, $\mu_\kappa = 0$, $\sigma^2_\alpha = 1$ and $\sigma^2_\kappa = 1$. Moreover, a multivariate normal prior distribution with a fixed correlation structure is assumed for abilities: $\boldsymbol{\theta}_i \sim N_m(\mathbf{0}, \boldsymbol{\Sigma})$, for $i = 1, \ldots, n$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix defined before.

Even if this choice can be viewed as very restrictive, it reflects the common beliefs and usual assumption we find in literature. In fact, a point of strength of the Bayesian approach is the possibility to formulate particular prior distributions depending on the information available a priori.

4.3.3 Likelihood function for responses

A categorical or generalized Bernoulli distribution of parameters $\pi_{\nu ij1}, \ldots, \pi_{\nu ijK_j}$ is assumed for responses (point 4 of the model building phase), thus for $\nu = 1, \ldots, m$, $j = 1, \ldots, p$ and $i = 1, \ldots, n$, it holds that:

$$Y_{ij}|\bullet \ \sim \mathrm{Cat}(\pi_{\nu ij1}, \ \ldots, \ \pi_{\nu ijK_j}) \ , \tag{4.8}$$

therefore:

$$P(Y_{ij} = k|\bullet) = \pi_{\nu ij1}^{[k=1]} \ \cdot \ \pi_{\nu ij2}^{[k=2]} \ \cdot \ \ldots \ \cdot \ \pi_{\nu ijK_j}^{[k=K_j]} . \tag{4.9}$$

Once the likelihood function for observed data is defined, the model is specified and we can perform the Bayesian estimation of the parameters of interest through an easy implementation in OpenBUGS, which run the Gibbs sampler algorithm. In particular the main advantage is due to fact that the joint posterior distribution has an untractable form, while the full conditional distributions are well defined. In fact:

$$P(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\Sigma}|\boldsymbol{Y}) \ \propto \ \boldsymbol{\mathcal{L}}(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\Sigma}) \ P(\boldsymbol{\theta}|\boldsymbol{\Sigma}) \ P(\boldsymbol{\alpha}) \ P(\boldsymbol{\kappa}) \ . \tag{4.10}$$

Expression (4.10) represents the joint posterior distribution of interest, where $\boldsymbol{\mathcal{L}}$ is the likelihood function and $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}$ are assumed to be independent.

Details about the code used to implement the model in OpenBUGS are reported in Appendix A.

4.4 Additive GRM implementation

As mentioned before for the multiunidimensional model, the implementation in OpenBUGS of the additive GRM needs the specification of the model according to the probability function defined in (4.4), the definition of the prior distributions, and the formulation of the likelihood function for the observed responses.

4.4.1 Model specification

The existence of a general ability in addition to the specific abilities implies the introduction of the further component θ_{0i} in the vector of person parameters for individual i: $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i}, \ldots, \theta_{mi})$, therefore the dimension of this vector is now $m+1$.

According to the additive structure presented in Figure 2.5, where each item measures an overall and a specific ability directly, and translated in expression (4.4), a generic cumulative probability $P_{\nu ijk}$ is a function of the item discrimination parameter related to the general ability ($\alpha_{0\nu j}$), the item discrimination parameter related to the specific ability ($\alpha_{\nu j}$), the threshold parameter between categories k and $k+1$ (κ_{jk}), the general ability of the individual ($\theta_{0\nu i}$), and the specific ability ($\theta_{\nu i}$) measured by the j-th item. We remind that each item belonging to a given subtest measures the general ability and only one specific ability. Hence, for $\nu = 1, \ldots, m$, $j = 1, \ldots, p$ and $k = 1, \ldots, K_j - 1$, it holds that $P_{\nu ijk} = \Phi(\kappa_{jk} - \alpha_{0\nu j}\theta_{0i} - \alpha_{\nu j}\theta_{\nu i})$.

Again, we set $P_{\nu ijK_j} = 1$, and the probability that the response of individual i to item j will fall in category k can be obtained by difference, as in the multiunidimensional case: $\pi_{\nu ij1} = P_{\nu ij1}$ and $\pi_{\nu ijk} = P_{\nu ijk} - P_{\nu,i,j,k-1}$, for $\nu = 1, \ldots, m$, $j = 1, \ldots, p$ and $k = 2, \ldots, K_j$.

4.4.2 Prior distributions

Also in the additive GRM we assume that the latent traits $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ are independent and multivariate normally distribuited:

$$\boldsymbol{\theta}_i \sim N_{m+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ ,$$

where $\boldsymbol{\mu}$ is the $(m+1)$-dimensional mean vector and $\boldsymbol{\Sigma}$ is the $(m+1) \times (m+1)$ constrained variance-covariance matrix with diagonal elements being 1 and off-diagonal elements being the ability correlations.

As explained above, θ_{0i} represents the unobservable general ability for examinee i which affects all the responses given from this examinee to the test items, while the specific abilities for the individual i, $\theta_{1i}, \ldots, \theta_{mi}$, affect every item contained in the corresponding subtest ν, for $\nu = 1, \ldots, m$. The prior distribution for $\boldsymbol{\theta}_i$ is then defined by expression (4.7), for $i = 1, \ldots, n$, where m is the number of both subtests and specific latent traits. For identification purposes, we set $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}$ fixed variance-covariance matrix.

Normal distributions are assumed for item discrimination parameters $\alpha_{0\nu j}$ and $\alpha_{\nu j}$, for $\nu = 1, \ldots, m$ and $j = 1, \ldots, p$:

$$\alpha_{0\nu j} \sim N(\mu_{\alpha_0}, \sigma^2_{\alpha_0}) \qquad \alpha_{\nu j} \sim N(\mu_{\alpha}, \sigma^2_{\alpha}) \ ,$$

and after having limited these parameters to be positive and having considered the restraints due to the identification issues (i.e. $\mu_{\alpha_0} = \mu_\alpha = 0$ and $\sigma^2_{\alpha_0} = \sigma^2_\alpha = 1$), we obtain truncated normal prior distributions for $\alpha_{0\nu j}$ and $\alpha_{\nu j}$:

$$\alpha_{0\nu j} \sim N(0,1) \ I(\alpha_{0\nu j} > 0) \qquad \alpha_{\nu j} \sim N(0,1) \ I(\alpha_{\nu j} > 0) \ .$$

Finally, concerning the threshold parameters, again we obtain an ordered series κ_{jk}, $\ldots$, κ_{j,K_j-1} starting from the unconstrained variables κ^*_{jk} (with identification constraints on hyperparmeters $\mu_\kappa = 0$ and $\sigma^2_\kappa = 1$) distributed as $\kappa^*_{jk} \sim N(0,1)$, and applying the transformation:

$$\{\kappa_{jk}, \ldots \kappa_{j,K_j-1}\} = \text{ranked}\{\kappa^*_{j1}, \ldots, \kappa^*_{j,K_j-1}\} \ .$$

4.4.3 Likelihood function for responses

Likewise the multiunidimensional GRM, a categorical distribution of parameters $\pi_{\nu ij1}, \ldots, \pi_{\nu ijK_j}$ is assumed for responses, therefore, for $\nu = 1, \ldots, m$, $j = 1, \ldots, p$ and $i = 1, \ldots, n$, expressions (4.8) and (4.9) hold also for the additive GRM.

See Appendix A for details about the code used to implement the additive GRM in OpenBUGS.

Summarizing, Table 4.1 reports the main characteristics of the multiunidimensional and the additive model considered in this work.

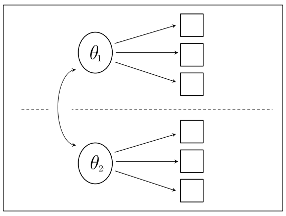
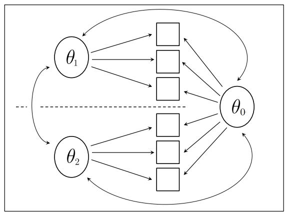
	Multiunidimensional Model for Graded Responses	Additive Model for Graded Responses
	Underlying latent structure	
		
	Model specification	
	$P_{\nu ijk} = \Phi(\kappa_{jk} - \alpha_{\nu j}\theta_{\nu i})$ $P_{\nu ijK_j} = 1$ $\pi_{\nu ij1} = P_{\nu ij1}$ $\pi_{\nu ijk} = P_{\nu ijk} - P_{\nu,i,j,k-1}$	$P_{\nu ijk} = \Phi(\kappa_{jk} - \alpha_{0\nu j}\theta_{0i} - \alpha_{\nu j}\theta_{\nu i})$ $P_{\nu ijK_j} = 1$ $\pi_{\nu ij1} = P_{\nu ij1}$ $\pi_{\nu ijk} = P_{\nu ijk} - P_{\nu,i,j,k-1}$
	Prior distributions on person parameters	
$\boldsymbol{\theta}_i$-vector	$\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{mi})$	$\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i}, \ldots, \theta_{mi})$
Prior on $\boldsymbol{\theta}_i$	$\boldsymbol{\theta}_i \sim N_m(\mathbf{0}, \boldsymbol{\Sigma})$	$\boldsymbol{\theta}_i \sim N_{m+1}(\mathbf{0}, \boldsymbol{\Sigma})$
	Prior distributions on item parameters	
Item discrimination (for a specific ability)	$\alpha_{\nu j} \sim N(0,1)\ I(\alpha_{\nu j} > 0)$	$\alpha_{\nu j} \sim N(0,1)\ I(\alpha_{\nu j} > 0)$
Item discrimination (for the general ability)	none	$\alpha_{0\nu j} \sim N(0,1)\ I(\alpha_{0\nu j} > 0)$
Threshold parameters $\{\kappa_{jk}, \ldots \kappa_{j,K_j-1}\}$	$\kappa^*_{jk} \sim N(0,1)$ $= \text{ranked}\{\kappa^*_{j1}, \ldots, \kappa^*_{j,K_j-1}\}$	$\kappa^*_{jk} \sim N(0,1)$ $= \text{ranked}\{\kappa^*_{j1}, \ldots, \kappa^*_{j,K_j-1}\}$
	Response likelihood	
	$Y_{ij}\vert\bullet \sim \text{Cat}(\pi_{\nu ij1}, \ldots, \pi_{\nu ijK_j})$ $P(Y_{ij} = k\vert\bullet) = \pi_{\nu ij1}^{[k=1]} \cdot \pi_{\nu ij2}^{[k=2]} \cdot \ldots \cdot \pi_{\nu ijK_j}^{[k=K_j]}$	

Table 4.1. Main features of the proposed multiunidimensional and additive models for graded responses.

Chapter 5

Simulation Study

In this chapter we present the simulation study performed to assess the item parameter recovery for both multiunidimensional and additive GRMs. The simulation study is conducted on a bidimensional case by varying the number of response categories, the sample size, the test and subtest lengths and the ability correlation structure. Two distinct simulation analyses have been designed in order to evaluate the parameter recovery of he multiunidimensional and the additive GRMs, respectively. A first series of simulations was carried out with the same simulation conditions for both models (Block 1). Then further conditions were analyzed in order to better understand the behavior of the additive model (Block 2). Several works on MIRT models focus on the accuracy of parameter estimation, and, through the manipulation of simulation conditions, it is possible to assess parameter recovery (Sheng, 2008; Sheng, 2010; Edwards, 2010a).

The first section of the chapter describes the simulation study design, while in the second and third sections are illustrated the simulation conditions and results for the multiunidimensional and additive models, respectively.

5.1 Simulation study design

The aim is the evaluation of the item parameter recovery of the multiunidimensional and the additive GRMs under several conditions. We consider the bidimensional case, $m = 2$, which, in particular, implies the presence of two specific abilities θ_1 and θ_2 for the multiunidimensional model, and the presence of two specific abilities θ_1 and θ_2 and an overall ability θ_0 for the additive model (recalling the graphical notation introduced before, the latent structures are summarized in Figure 5.1).

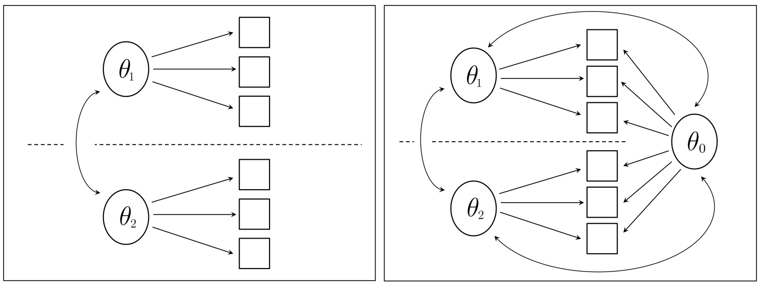


Figure 5.1. Bidimensional case for multiunidimensional and additive structures.

The model parameters and the ability correlations are estimated through OpenBUGS version 3.2.2. The fundamental scheme for each simulation is the following (for more details about the procedure and the codes used for the implementation, see Appendix B):

- Simulate the vectors of 'real' parameters, taking into account the conditions we are testing. We perform this step using an R GUI procedure.

- Perform $Q = 10$ replications of the computation procedure for each simulation. In each replication we sample the data matrix using the parameters obtained at the previous step, and we run OpenBUGS through the R GUI package BRugs (Thomas et al., 2006), which basically permits to recall OpenBUGS automatically from R.

- Proceed to the evaluation of parameter recovery and the computation of the reproduced correlations between the latent traits by using the Q estimates gained at the previous step.

5.1.1 Parameter recovery

In order to evaluate the recovery of the generated item parameters (which in our simulation context correspond to the "real" population values), we compute the absolute bias and the root mean square error (RMSE) for each estimated parameter, taking account the Q replications for each simulation. If we denote with $\hat{\omega}$ a generic parameter estimate, i.e. the mean of the posterior distribution gained in each replication, and with ω^* the real generated value, biases and RMSE are computed as follow:

$$\text{Bias}(\omega) = \frac{1}{Q}\sum_{q=1}^{Q}(\hat{\omega}_q - \omega^*) \tag{5.1}$$

$$\text{RMSE}(\omega) = \frac{1}{Q}\sqrt{\sum_{q=1}^{Q}(\hat{\omega}_q - \omega^*)^2}\,, \tag{5.2}$$

where lower levels of bias and RMSE indicate better precision in parameter recovery.

5.1.2 Estimated ability correlations

Considering that the two models have been specified allowing the latent traits to correlate, and that the correlation structure is reflected in the variance-covariance matrix of the latent abilities $\mathbf{\Sigma}$, we are not interested only in item parameters recovery, but also in the way the models are able to reproduce such ability correlations.

For this reason, for each simulation, we report also the estimated ability Pearson correlations: $\hat{r}_{12}$ for the multiunidimensional model, and $\hat{r}_{01}, \hat{r}_{02}$ and $\hat{r}_{12}$ for the additive model (remind that with 0 we refer to the overall ability).

5.1.3 Convergence detection

In Lunn et al. (2013) is clearly described how important is the detection of the chain convergence. An easy, but effective, strategy is the detection of convergence *informally* by eye. Anyway, the model could include many parameters and, consequently, it can be quite hard to check all of them by eye. Figure 5.2[1] shows two examples of chains that have reached the convergence. The initial part of the chain, i.e. the non-stationary part, is called *burn-in* and the iterations belonging to it must be discharged to be sure that the successive realisations can be considered as a sample from the stationary distribution. The burn-in period is easily recognizable in the first chain reported in Figure 5.2.

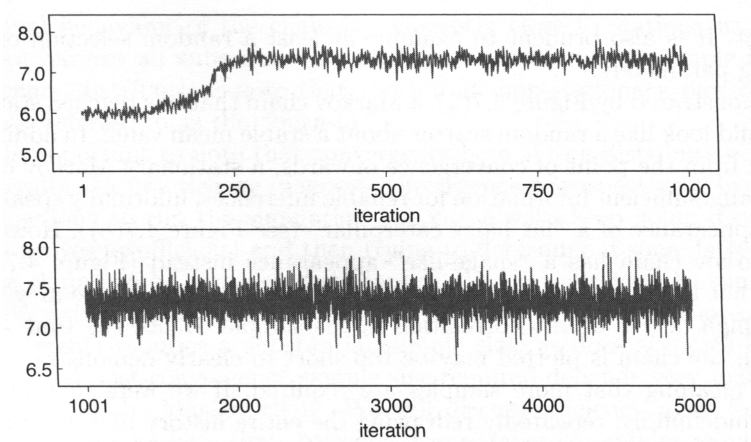


Figure 5.2. Examples of stationary chains.

The so called R statistic of Gelman and Rubin, proposed by Gelman and Rubin (1992) and further developed by Brooks and Gelman (1998), represents an useful instrument adopted to check the convergence of the Markov chains, and hence the reliability of the estimates. This convergence diagnostic can be constructed only when more than one chain are run simultaneously. This aspect lead to our decision of running two distinct chains for each simulation (see the next section).

Basically, the convergence is reached when the chains follow an indistinguishable, not recognizable, trajectory from the initial values. The method is based on the between and within sample variabilities (Ntzoufras, 2011) and the diagnostic statistic is given by:

$$\hat{R} = \frac{\hat{V}}{W} = \frac{T'-1}{T'} + \frac{B/T'}{W}\,\frac{M+1}{M}\,,$$

[1]Source: Lunn et al., 2013.

where T' represents the number of iterations in each chain, M is the number of chains, B/T' is the between-sample variance, that is the variance of the posterior mean values taking into account all the chains, W is the within-sample variance, that is the mean of variances within each chain, and the pooled posterior variance is given by (Ntzoufras, 2011):

$$\hat{V} = \frac{T'-1}{T'} W + \frac{B}{T'} \frac{M+1}{M} .$$

Once the chains are stationary and the convergence is reached, $\hat{R} \rightarrow 1$. A corrected version of the R statistic also exists, see Brooks and Gelman (1998).

5.1.4 Bayesian fit

Additionally to the calculation and examination of $\hat{R}$, other well known indicators for the fit evaluation are the Bayesian deviance and the deviance information criterion (DIC) (Lunn et al., 2013). Their use is appropriate to obtain some measures of fit and complexity of the model considered.
The Bayesian deviance is defined as:

$$D(\theta) = -2 \log p(y|\theta) ,$$

where θ denotes the model parameters and with $p(y|\theta)$ is denoted the full sampling distribution. OpenBUGS considers it as a node (created automatically), so that it has its own posterior distribution and can be considered like the other model parameters. Combining the mean posterior deviance, $\overline{D(\theta)}$, and the number of model parameters, p_D, we can compute the DIC through the expression

$$DIC = \overline{D(\theta)} + p_D .$$

It can be proved that the DIC is an approximation of the Akaike's information criterion, $AIC = \overline{D(\theta)} + 2p_D$. Also in this case, OpenBUGS permits to easily compute the DIC for each model implemented.

5.1.5 General simulation conditions

For all the simulations conducted in this work, $Q = 10$ replications have been performed. For each one, we have considered a chain length of 30,000 iterations, with a burn-in phase of 15,000 iterations. Moreover, two chains have been generated, in order to be able to set in OpenBUGS the computation of the R and the DIC statistics.

These choices may be penalizing with reference to the computational time needed to run the Gibbs sampling algorithm for each simulation (a single replication needs about 13 hours to be completed), nevertheless, after an examination of the R diagnostic illustrated above, they ensure the reaching of the convergence.

For each distinct case, we perform different simulations according to a sample size of $n = 500$, and a larger sample size of $n = 1000$.

5.2 Multiunidimensional GRM: simulations and results

In this section we report the conditions and the results about the simulations made to assess the parameter recovery of MCMC estimation for the multiunidimensional model for graded responses. All the simulations conducted are characterized by the general conditions reported in section 5.1.5 and other specific conditions, with the aim to evaluate the sensitivity of the model.

5.2.1 Simulation conditions

We consider n individuals and a set of p ordinal items, divided into 2 subtests, each one consisting of p_1 and p_2 items. The response Y_{ij} of the i-th individual to the j-th item can take values in the set $1, \ldots, K_j$, hence each item is characterized by $K_j - 1$ thresholds satisfying the order constraint $\kappa_{j1} < \ldots \kappa_{j,K_j-1}$. Moreover, we assume that all the test items have the same number of categories, i.e. $K_1 = ... = K_p = K$. Additionally, we assume the existence of $m = 2$ latent abilities, θ_1 and θ_2, underlying the responses to the items, which follow a multiunidimensional latent structure (see Figure 5.1, left part). Thus, the test consists of two subtests

Simulation	p	p_1	p_2	K_j	n	$\boldsymbol{\Sigma}$
$\sharp$ 1	15	5	10	3	500	Σ_A
$\sharp$ 2	15	5	10	3	500	Σ_B
$\sharp$ 3	15	5	10	4	500	Σ_A
$\sharp$ 4	15	5	10	4	500	Σ_B
$\sharp$ 5	15	5	10	3	1000	Σ_A
$\sharp$ 6	15	5	10	3	1000	Σ_B
$\sharp$ 7	15	5	10	4	1000	Σ_A
$\sharp$ 8	15	5	10	4	1000	Σ_B

Table 5.1. Simulation conditions for the multiunidimensional model for graded responses.

and the items in each subtest characterize a single specific ability. Moreover, the specific abilities are allowed to correlate and the model follows a compensatory approach.

We perform a block of simulations (Block 1) referred to the case where a test length of $p = 15$ is divided into a first subtest made of $p_1 = 5$ items and a second subtest made of $p_2 = 10$ items. A further distinction has been made about the number of item categories, varying from $K = 3$ to $K = 4$. Furthermore, each case was analyzed by using two different correlation matrices among the abilities: $\boldsymbol{\Sigma}_A$ and $\boldsymbol{\Sigma}_B$. $\boldsymbol{\Sigma}_A$ is a 2×2 identity matrix, where the correlation among the specific abilities is set to zero ($r_{12} = 0$). The second correlation matrix $\boldsymbol{\Sigma}_B$ introduces a moderate correlation between the latent abilities ($r_{12} = 0.4$).

By combining all the conditions, we obtain 8 different scenarios, listed in Table 5.1, to investigate the parameter recovery for the multiunidimensional GRM.

5.2.2 Results

In this section we report the results we obtained for each of the 8 simulations conducted for the multiunidimensional model for graded responses. In the following, for each item parameter type within a subtest, median absolute bias and median root mean square error are reported for each scenario, as well as the

Simulations Block 1 - Subtest 1 (5 items)

			α_1		κ_1		κ_2		κ_3	
n	(p_1, p_2) K		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
500	(5,10) 3	Σ_A	**0.11**	**0.09**	0.08	0.05	0.08	0.04		
		Σ_B	**0.11**	**0.07**	0.10	0.02	0.07	0.03		
	(5,10) 4	Σ_A	**0.11**	**0.09**	**0.12**	**0.06**	0.07	0.03	0.09	0.04
		Σ_B	0.09	0.05	0.10	**0.10**	0.08	0.02	0.09	0.05
1000	(5,10) 3	Σ_A	0.06	0.02	0.05	0.03	0.08	0.05		
		Σ_B	0.06	0.04	0.05	0.04	0.05	0.02		
	(5,10) 4	Σ_A	0.05	0.02	0.06	0.04	0.05	0.01	0.04	0.01
		Σ_B	0.05	0.01	0.04	0.01	0.03	0.01	0.05	0.02

Table 5.2. Multiunidimensional model: block 1 simulation results for subtest 1 (median RMSEs and median absolute biases).

Simulations Block 1 - Subtest 2 (10 items)

			α_1		κ_1		κ_2		κ_3	
n	(p_1, p_2) K		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
500	(5,10) 3	Σ_A	0.09	0.02	0.07	0.02	0.07	0.02		
		Σ_B	0.08	0.01	0.07	0.03	0.08	0.02		
	(5,10) 4	Σ_A	0.07	0.03	0.08	**0.06**	0.09	0.05	**0.12**	**0.07**
		Σ_B	0.09	0.02	0.08	**0.08**	0.08	0.01	**0.11**	0.02
1000	(5,10) 3	Σ_A	0.07	0.02	0.06	0.02	0.05	0.03		
		Σ_B	0.08	0.03	0.06	0.03	0.06	0.03		
	(5,10) 4	Σ_A	0.07	0.02	0.05	0.02	0.05	0.02	0.06	0.01
		Σ_B	0.06	0.02	0.05	0.01	0.05	0.00	0.08	0.03

Table 5.3. Multiunidimensional model: block 1 simulation results for subtest 2 (median RMSEs and median absolute biases).

ability correlation estimates.

In Tables 5.2 and 5.3 we present RMSE and absolute bias for the item parameters (discrimination and thresholds parameters) characterizing, respectively, the items belonging to the first subtest and the items belonging to second subtest. Values of the RMSE greater than 0.10 and values of the absolute bias greater than 0.05 are highlighted in bold, identifying cases where the parameter recovery could be improved.

With reference to the first subtest, Table 5.2 shows how the worst performances are related to the smaller sample sizes ($n = 500$). In fact, when we increased the sample size to $n = 1000$, the RMSEs and biases noticeably decreased, other things being equal. The presence of an underlying correlation between the two latent traits does not seem to affect the item parameters recovery.

Results are similar for items belonging to the second subtest (Table 5.3). Higher biases are noticed when sample sizes are smaller, even though the overall parameter reproduction is better if compared to the first subtest. This aspect should be due to the greater number of items included in the second subtest ($p_2 = 10$ versus $p_1 = 5$). For $n = 1000$ item parameters are recovered very precisely.

Multiunidimensional model:
Real and estimated ability correlations

			r_{12}	$\hat{r}_{12}$
500	(5,10) 3	Σ_A	0.00	0.02
		Σ_B	0.40	0.46
	(5,10) 4	Σ_A	0.00	0.03
		Σ_B	0.40	0.45
1000	(5,10) 3	Σ_A	0.00	-0.01
		Σ_B	0.40	0.49
	(5,10) 4	Σ_A	0.00	-0.01
		Σ_B	0.40	0.48

Table 5.4. Multiunidimensional model: real (r) and estimated ($\hat{r}$) ability correlations.

Table 5.4 illustrates the estimated ability correlations for each scenario. It can be noticed that the differences between the generated real values r_{12} and the estimated values $\hat{r}_{12}$ are quite low, indicating good performances of the model.

Here, unlike what we observe for item parameters, the underlying ability correlation structure seems to influence the correlation reproduction. In fact we observe a worst reproduction in correspondence to the model characterized by the more complex latent correlation structure $\mathbf{\Sigma}_B$. The sample size seems to influence the latent traits correlation reproduction: the reproduction accuracy increases with the increase of sample size for the simple case ($\mathbf{\Sigma}_A$), while it decreases with the increase of sample size for the complex case ($\mathbf{\Sigma}_B$).

As a conclusive remark, what emerges from the simulation study conducted to assess the multiunidimensional model for graded responses with correlated latent traits is that item parameters and ability correlations are well reproduced.

5.3 Additive GRM: simulations and results

In this section we report the conditions and the results related to the simulation study conducted to evaluate the multidimensional additive GRM with correlated abilities, estimated within a Bayesian context. In addition to the first block of simulations designed also for the multiunidimensional model, a further block of simulations has been performed, in order to better understand the behavior of our proposed model.

5.3.1 Simulation conditions

The general simulation conditions for the additive model for graded responses are the same as the multiunidimensional model. We still assume the existence of of $m = 2$ specific latent abilities, θ_1 and θ_2 , but now we consider also an overall latent ability θ_0. Accordingly, we are focusing on the bidimensional case, for which the latent structure is represented in Figure 5.1, right part.

We indicate with p the total number of ordinal test items, with p_1 and p_2 the number of items belonging to the first and the second subtest, respectively. K_j

indicates the greater category for the j-th item and we consider that all the items have the same number of categories $K_j = K,\ \forall j$.

We start from a first block of simulations (Block 1) referred to the case where a test length of $p = 15$ is divided into a first subtest made of $p_1 = 5$ items and a second subtest made of $p_2 = 10$ items. A further distinction has been made about the number of item categories, varying in the first block from $K = 3$ to $K = 4$. Furthermore, each case was analyzed by using two different correlation matrices among the abilities: $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$. $\mathbf{\Sigma}_A$ is a 3×3 identity matrix, where all the correlations among the abilities are set to zero ($r_{01} = r_{02} = r_{12} = 0$). In this case, the additive model with orthogonal traits has the same latent structure of the well known bi-factor model and the three latent traits (the general and specific abilities) are separate and well distinguished from each other. The second correlation matrix $\mathbf{\Sigma}_B$ introduce moderate correlations between all the latent abilities ($r_{01} = 0.4,\ r_{02} = 0.3,\ r_{12} = 0.2$). The choice to consider not particularly high levels of correlation has been driven by the consideration that high correlations among the latent abilities may lead to the existence of a dominant latent trait, redirecting to a unidimensional model.

In order to investigate further conditions, we designed a second block of simulations (Block 2), where we increase both the length of the test and the number of item categories. We consider a case characterized by a test length of $p = 50$ (divided into $p_1 = 20$ and $p_2 = 30$ items for subtest 1 and 2, respectively) and $K = 4$ categories for each test item; and a last case where the test length is $p = 30$ ($p_1 = 10$ and $p_2 = 20$) and items have $K = 5$ categories. Again, with respect to the correlation matrix, the two cases of $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$ are distinguished as above.

By combining all the simulation conditions, we obtain 16 different scenarios, illustrated in Table 5.5, to investigate the parameter recovery for the proposed model.

Simulation	p	p_1	p_2	K_j	n	$\boldsymbol{\Sigma}$
$\sharp$ 1	15	5	10	3	500	Σ_A
$\sharp$ 2	15	5	10	3	500	Σ_B
$\sharp$ 3	15	5	10	4	500	Σ_A
$\sharp$ 4	15	5	10	4	500	Σ_B
$\sharp$ 5	15	5	10	3	1000	Σ_A
$\sharp$ 6	15	5	10	3	1000	Σ_B
$\sharp$ 7	15	5	10	4	1000	Σ_A
$\sharp$ 8	15	5	10	4	1000	Σ_B
$\sharp$ 9	50	20	30	4	500	Σ_A
$\sharp$ 10	50	20	30	4	500	Σ_B
$\sharp$ 11	30	10	20	5	500	Σ_A
$\sharp$ 12	30	10	20	5	500	Σ_B
$\sharp$ 13	50	20	30	4	1000	Σ_A
$\sharp$ 14	50	20	30	4	1000	Σ_B
$\sharp$ 15	30	10	20	5	1000	Σ_A
$\sharp$ 16	30	10	20	5	1000	Σ_B

Table 5.5. Simulation conditions for the additive model for graded responses.

5.3.2 Results

Tables 5.6 and 5.7 show the item parameter recovery for the first block of simulations where $p = 15$ ($p_1 = 5$ and $p_2 = 10$), respectively for subset 1 and subtest 2. It emerges that all parameters are quite well recovered when the number of categories for each item is $K = 3$ and a sample size of $n = 500$ is enough to get accurate estimates. Results are slightly better for the Σ_A correlation matrix, rather than Σ_B.

On the other hand, when the number of item categories is $K = 4$ we obtain less accurate estimates, for both Σ_A and Σ_B ability correlation structures. Estimates get better after increasing the sample size, but median RMSEs and biases remain rather high, especially for α_0 and α_v discrimination parameters. Considering that this result is more evident for the first subtest where $p_1 = 5$, rather than the second one where $p_2 = 10$, this may be due to the small number of item compared to the increased number of categories.

Results about the second block of simulations are reported in Tables 5.8 and 5.9. Focusing on the case where $p = 50$ ($p_1 = 20$ and $p_2 = 30$) and $K = 4$, we observe that in both subtests the item parameters are not well recovered, particularly the discrimination parameters.

Nevertheless, these shortcomings are overtaken by increasing the sample size. In fact, when $n = 1000$ all the parameters are recovered rather precisely. Different correlation structures seem not to affect parameter recovery, with an exception of the discrimination parameters for the second subtest, where we register higher median RMSEs in association to the more complex correlation structure.
Analogously, the cases where $p = 30$ ($p_1 = 10$ and $p_2 = 20$) and $K = 5$ benefit from the enlarged sample size. For $n = 1000$, item parameters are recovered with care, with slightly better accuracy with respect to Σ_A correlation matrix.

Simulations Block 1 - Subtest 1 (5 items)

				α_0		α_1		κ_1		κ_2		κ_3		κ_4	
n	(p_1, p_2)	K		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
500	(5,10)	3	Σ_A	0.08	0.05	0.08	0.03	0.08	0.01	0.07	0.04				
			Σ_B	0.09	0.02	**0.15**	**0.13**	0.09	0.03	0.09	0.02				
	(5,10)	4	Σ_A	**0.13**	**0.07**	**0.12**	**0.09**	**0.15**	**0.12**	**0.15**	**0.10**	**0.13**	0.04		
			Σ_B	**0.17**	0.05	**0.23**	**0.10**	**0.16**	**0.16**	0.09	0.02	**0.15**	0.03		
1000	(5,10)	3	Σ_A	0.07	0.03	0.09	0.03	0.08	**0.06**	0.07	0.03				
			Σ_B	0.09	0.03	**0.14**	**0.07**	0.08	0.03	0.06	0.03				
	(5,10)	4	Σ_A	0.09	0.02	**0.16**	**0.12**	**0.15**	0.05	0.06	0.03	**0.16**	**0.10**		
			Σ_B	**0.14**	**0.08**	**0.16**	**0.12**	0.08	0.05	0.08	0.04	**0.15**	**0.10**		

Table 5.6. Additive model: block 1 simulation results for subtest 1 (median RMSEs and median absolute biases).

Simulations Block 1 - Subtest 2 (10 items)

			α_0		α_2		κ_1		κ_2		κ_3		κ_4	
n	(p_1, p_2) K		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
500	(5,10) 3	Σ_A	0.09	0.05	0.10	0.02	0.08	0.02	0.08	0.03				
		Σ_B	**0.11**	0.04	0.00	0.05	0.09	0.04	0.10	0.00				
	(5,10) 4	Σ_A	**0.12**	0.02	**0.14**	0.05	0.10	0.03	0.10	0.04	**0.13**	**0.06**		
		Σ_B	**0.14**	0.04	**0.14**	0.04	0.10	**0.10**	0.09	0.02	**0.11**	0.03		
1000	(5,10) 3	Σ_A	0.09	0.04	0.09	0.05	0.06	0.02	0.06	0.02				
		Σ_B	**0.15**	0.05	**0.18**	0.03	0.05	0.02	0.05	0.01				
	(5,10) 4	Σ_A	**0.16**	**0.12**	**0.16**	**0.09**	0.07	0.03	0.07	0.03	0.08	0.03		
		Σ_B	**0.23**	**0.13**	**0.19**	**0.11**	0.07	0.04	0.06	0.03	0.09	0.04		

Table 5.7. Additive model: block 1 simulation results for subtest 2 (median RMSEs and median absolute biases).

Simulations Block 2 - Subtest 1 (20 and 10 items)

n	(p_1, p_2)	K		α_0 RMSE	α_0 Bias	α_1 RMSE	α_1 Bias	κ_1 RMSE	κ_1 Bias	κ_2 RMSE	κ_2 Bias	κ_3 RMSE	κ_3 Bias	κ_4 RMSE	κ_4 Bias
500	(20,30)	4	Σ_A	**0.14**	**0.07**	**0.14**	**0.08**	0.10	**0.06**	0.10	0.05	0.10	0.04		
			Σ_B	**0.15**	**0.07**	**0.18**	**0.10**	0.09	0.03	0.08	0.03	0.09	0.03		
	(10,20)	5	Σ_A	**0.20**	0.05	**0.21**	**0.06**	0.09	0.01	0.09	0.03	0.08	0.04	0.09	0.04
			Σ_B	**0.17**	**0.07**	**0.22**	**0.07**	0.10	0.02	0.09	0.02	0.08	0.02	0.08	0.02
1000	(20,30)	4	Σ_A	0.07	0.05	0.08	0.04	0.06	0.01	0.06	0.01	0.05	0.01		
			Σ_B	0.06	0.02	0.08	0.04	0.06	0.04	0.06	0.04	0.06	0.03		
	(10,20)	5	Σ_A	0.07	0.04	0.07	0.04	0.08	0.02	0.06	0.02	0.05	0.02	0.05	0.01
			Σ_B	**0.19**	0.04	**0.27**	0.05	0.07	0.03	0.06	0.03	0.05	0.03	0.07	0.02

Table 5.8. Additive model: block 2 simulation results for subtest 1 (median RMSEs and median absolute biases).

Simulations Block 2 - Subtest 2 (30 and 20 items)

				α_0		α_2		κ_1		κ_2		κ_3		κ_4	
n	(p_1, p_2)	K		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
500	(20,30)	4	Σ_A	**0.16**	**0.07**	**0.20**	**0.08**	0.10	0.03	0.09	0.03	0.10	0.03		
			Σ_B	0.07	0.02	0.10	0.05	0.10	**0.06**	0.10	**0.06**	0.10	0.05		
	(10,20)	5	Σ_A	**0.23**	**0.08**	**0.18**	0.05	0.09	0.03	0.07	0.03	0.07	0.02	0.09	0.03
			Σ_B	**0.20**	**0.09**	**0.19**	**0.08**	0.08	0.03	0.07	0.03	0.08	0.02	0.08	0.02
1000	(20,30)	4	Σ_A	0.06	0.05	0.06	0.02	0.06	0.02	0.06	0.02	0.06	0.01		
			Σ_B	**0.14**	0.06	**0.17**	0.03	0.07	0.02	0.05	0.02	0.07	0.02		
	(10,20)	5	Σ_A	0.07	0.05	0.07	0.03	0.05	0.01	0.05	0.01	0.05	0.01	0.06	0.02
			Σ_B	0.06	0.01	0.08	0.03	0.06	0.02	0.05	0.02	0.05	0.01	0.06	0.02

Table 5.9. Additive model: block 2 simulation results for subtest 2 (median RMSEs and median absolute biases).

Additive model: Real and estimated ability correlations

			r_{01}	$\hat{r}_{01}$	r_{02}	$\hat{r}_{02}$	r_{12}	$\hat{r}_{12}$
500	(5,10) 3	Σ_A	0.00	0.07	0.00	0.16	0.00	-0.07
		Σ_B	0.40	0.62	0.30	0.49	0.20	0.24
	(5,10) 4	Σ_A	0.00	0.09	0.00	0.29	0.00	-0.07
		Σ_B	0.40	0.60	0.30	0.56	0.20	0.27
1000	(5,10) 3	Σ_A	0.00	0.11	0.00	0.16	0.00	-0.05
		Σ_B	0.40	0.60	0.30	0.54	0.20	0.27
	(5,10) 4	Σ_A	0.00	0.13	0.00	0.36	0.00	-0.05
		Σ_B	0.40	0.58	0.30	0.65	0.20	0.30
500	(20,30) 4	Σ_A	0.00	0.00	0.00	0.29	0.00	-0.05
		Σ_B	0.40	0.50	0.30	0.36	0.20	0.21
	(10,20) 5	Σ_A	0.00	0.02	0.00	0.15	0.00	-0.03
		Σ_B	0.40	0.51	0.30	0.48	0.20	0.24
1000	(20,30) 4	Σ_A	0.00	0.03	0.00	0.07	0.00	-0.02
		Σ_B	0.40	0.45	0.30	0.37	0.20	0.20
	(10,20) 5	Σ_A	0.00	0.06	0.00	0.11	0.00	-0.05
		Σ_B	0.40	0.52	0.30	0.41	0.20	0.24

Table 5.10. Additive model: real (r) and estimated ($\hat{r}$) ability correlations.

Table 5.10 illustrates the estimated ability correlations for each scenario. Their correspondent true values are also reported and we can observe how the correlations are reproduced. In particular, the results are coherent with the ones observed in relation to the item parameters: the best performance is associated to the cases of the highest sample size, a reasonable number of items (totally 50) and a number of categories equal to 4, even in case of slightly high correlations.

To conclude, the main results showed that the algorithm is particularly sensitive to the sample size due to the model complexity and the high number of parameters to be estimated. In fact, when the sample size is sufficiently large ($n = 1000$), all the parameters are well reproduced. The results are also affected by the trade-off between the test length and the number of categories: the worst results are associated to a high number of categories and a low test length. Analogous evidences apply for the correlation estimates.

Chapter 6

Application to real data: residents' attitudes towards tourism

In this chapter we illustrate an implementation of the proposed models on data collected with the aim to investigate Romagna and San Marino residents' perceptions and attitudes towards the tourism industry. After having introduced the interpretation of model parameters in this new context, we illustrate the research design. Results about the multiunidimensional and additive GRMs estimations are reported in the final two sections.

6.1 Interpretation of model parameters

In the present application, the opinions of a sample of respondents on a set of aspects referred to the tourism industry represent our observed variables. Therefore, latent traits can be defined as 'perceptions'. The investigation involves two distinct aspects of the phenomenon, namely perceived benefits and costs of tourism. Therefore, it is possible to identify two specific perceptions and the overall attitude of respondents as latent variables.

Within this framework, discrimination parameters represent the capability of the items to differentiate between respondents with different levels of agreement, whereas the threshold parameters can be interpreted as 'criticity levels' of the corresponding item. For a given item, high values for the criticity parameters correspond to lower probabilities to observe responses in positive categories.

6.2 Research design

Data analyzed are the result of a research conducted by the University of Bologna with the aim to study the subjective well-being (Bernini et al., 2013). Data were collected in the end of 2010 from residents in the Romagna area and in the State of San Marino (Italy). The Romagna area consists of the provinces of Forlì-Cesena, Rimini, and Ravenna, and is located in the southeast of the Emilia-Romagna region. The independent republic of San Marino borders the Rimini Province.

The tourism industry has a relevant weight in this area: it contains the 7% of Italian accommodation structures and the 5% of Italian entertainment activities. Moreover, it is one of the main Italian tourism destinations, hosting in 2010 almost 27.5 million overnight stays (7.3% of the total national overnights) and 5.3 million arrivals (5.3% of the total national arrivals).

The sampling design was carried out taking into account a stratification of the provinces and the demographic characteristics of the tourists (age and gender). The final sample is representative of the population at the provincial level, with a margin of error of $\pm 5\%$ at a 95% level of confidence. A total of 794 questionnaires were obtained through a telephone survey.

The questionnaire was created with the aim to collect residents' evaluations about costs and benefits of the tourism industry, a personal benefit from tourism, the quality of life in the area, the degree of involvement in the tourism industry, the residents' satisfaction with either their leisure or non-leisure domains, their quality of life, and the degree of support for future development of the tourism industry. Furthermore, personal information (age, gender, nationality, residence and occupation) were also collected (see Appendix C for the submitted questionnaire). Some characteristics of the sample are summarised in Table 6.1.

In particular, among all the aspects investigated through the survey, the object of our analysis is the perception of benefits and costs associated to the tourism industry.

The perceived benefits of tourism were assessed by five items: the support in

	Number	%
Provinces		
Forlì-Cesena	246	31.0
Ravenna	245	30.9
Rimini	243	30.6
San Marino	60	7.6
Gender		
Female	413	52.0
Male	381	48.0
Age		
< 25	65	8.2
25 - 35	115	14.5
35 - 45	171	21.5
45 - 55	127	16.0
55 - 65	95	12.0
≥ 65	221	27.8
Education		
Primary	105	13.2
Lower secondary	196	24.7
Upper secondary	192	24.2
University	301	37.9

Table 6.1. Profile of respondents.

local economic development [B1], quality of life [B2], public services improvement [B3], employment prospects [B4], and opportunities for cultural activities [B5]. Respondents were asked to indicate whether those items would improve for their community as a result of increasing tourism activity on a 7-point anchor scale, from "strongly disagree" to "strongly agree".

On the other hand, the perceived costs of tourism were assessed by other five items: the cost of living [C1], crime [C2], environment damage [C3], traffic congestion [C4], and pollution [C5]. In this case residents were asked to express if those aspects would worsen for their community as a result of increasing tourism activity on the 7-point scale mentioned above. Scales of the items with respect to costs were inverted in order to eliminate reverse scoring and make the low and high scores be associated with high and low perceptions of costs, respectively.

In Table 6.2, the response frequencies are reported for each item. Items B1-B5 refer to the benefits, while items C1-C5 refer to the costs that were perceived by residents about the tourism industry.

		Responses						
		Low benefits ⟵				⟶ High benefits		
Item	Item description	1	2	3	4	5	6	7
B1	Econ. support	12	51	58	157	149	235	132
B2	Quality of life	24	49	78	184	227	155	77
B3	Public services	16	45	97	186	190	171	89
B4	Job opportunities	16	36	69	157	187	198	131
B5	Cultural act.	30	54	76	186	188	157	103
		Responses						
		High costs ⟵				⟶ Low costs		
Item	Item description	1	2	3	4	5	6	7
C1	Cost of life	64	151	182	139	119	100	39
C2	Crime rate	145	169	157	155	69	71	28
C3	Env. damage	117	151	166	187	96	59	18
C4	Traffic	193	152	158	130	89	45	27
C5	Pollution	158	173	164	136	63	81	19

Table 6.2. Response frequencies for items about tourism benefits (B1-B5) and items about tourism costs (C1-C5).

6.3 Results for the multiunidimensional GRM

The parameters of the bidimensional version of the multiunidimensional GRM have been estimated on the basis of the residents' responses to the 5 items on benefits (B1-B5) and the 5 items on costs (C1-C5).

By following a confirmatory approach, we assume that the item responses on benefits are related to the first latent variable θ_1, while the item responses on costs are related to the second latent variable θ_2. The two traits are allowed to correlate.

Concerning the definition of the latent traits, θ_1 can be expressed as the "perception of tourism benefits", while θ_2 can be defined as the "perception of the tourism costs". These interpretations strictly derives from the meaning of the items included in the questionnaire. A positive perception of the effect of the tourism industry is reflected by high resident scores on θ_1 and θ_2. In particular, the higher the positive perception of the effect of tourism on the local environment is, the higher the score is on θ_1. Conversely, the higher the score is on θ_2, the lower the perception of a negative impact of tourism on the environment is.

The model parameters were estimated by using the proposed OpenBUGS procedure for the multiunidimensional GRM, with two chains and 30,000 total iterations (15,000 as burn-in) for each one. Table 6.3 illustrates the item parameter estimates for the test items.

The strength of the relationship among the observed responses and the related latent trait is expressed by the discrimination parameters α. From Table 6.3 we can see that these parameters are all largely positive, suggesting that there is a coherent choice for the chosen latent structure.

Particularly, the capability of an item to differentiate individuals with different perceptions of the impact of tourism increases as the discrimination parameters increases. This relationship means that public services, job opportunities and cultural activities (items B3, B4 and B5, respectively) are the most informative on the perception of the tourism advantage, whereas traffic and pollution (items C4 and C5) can better discriminate between residents who have different perceptions of the environmental impact of tourism. Among all the items, the cost of life (C1) presents the lower discrimination capability.

Item	Item description	$\widehat{\alpha}_\nu$	SD($\widehat{\alpha}_\nu$)	MCSE($\widehat{\alpha}_\nu$)	$\widehat{\kappa}_1$	SD($\widehat{\kappa}_1$)	MCSE($\widehat{\kappa}_1$)	$\widehat{\kappa}_2$	SD($\widehat{\kappa}_2$)	MCSE($\widehat{\kappa}_2$)	$\widehat{\kappa}_3$	SD($\widehat{\kappa}_3$)	MCSE($\widehat{\kappa}_3$)
B1	Econ. support	1.103	0.074	0.001	-2.904	0.155	0.001	-1.920	0.095	0.001	-1.400	0.080	0.001
B2	Quality of life	1.204	0.078	0.001	-2.632	0.134	0.001	-1.856	0.096	0.001	-1.225	0.078	0.001
B3	Public services	1.485	0.096	0.001	-3.240	0.184	0.002	-2.178	0.116	0.001	-1.301	0.088	0.002
B4	Job. opp.	1.423	0.094	0.002	-3.231	0.183	0.002	-2.315	0.123	0.001	-1.549	0.095	0.002
B5	Cultural act.	1.339	0.087	0.001	-2.629	0.134	0.001	-1.844	0.099	0.001	-1.221	0.082	0.001
C1	Cost of life	0.286	0.105	0.004	-1.468	0.115	0.001	-0.633	0.065	0.001	0.003	0.046	0.001
C2	Crime rate	1.563	0.109	0.002	-1.603	0.106	0.001	-0.432	0.080	0.001	0.450	0.080	0.001
C3	Env. damage	1.440	0.100	0.003	-1.744	0.105	0.001	-0.658	0.078	0.001	0.222	0.074	0.001
C4	Traffic	1.638	0.117	0.003	-1.268	0.097	0.001	-0.322	0.082	0.001	0.618	0.085	0.001
C5	Pollution	1.793	0.131	0.003	-1.580	0.114	0.001	-0.413	0.087	0.001	0.566	0.088	0.001

Item	Item description	$\widehat{\kappa}_4$	SD($\widehat{\kappa}_4$)	MCSE($\widehat{\kappa}_4$)	$\widehat{\kappa}_5$	SD($\widehat{\kappa}_5$)	MCSE($\widehat{\kappa}_5$)	$\widehat{\kappa}_6$	SD($\widehat{\kappa}_6$)	MCSE($\widehat{\kappa}_6$)
B1	Econ. support	-0.496	0.065	0.001	0.171	0.063	0.001	1.348	0.077	0.001
B2	Quality of life	-0.274	0.066	0.001	0.807	0.072	0.001	1.906	0.099	0.001
B3	Public services	-0.284	0.075	0.002	0.723	0.080	0.002	2.010	0.113	0.001
B4	Job. opp.	-0.560	0.076	0.002	0.366	0.074	0.001	1.543	0.094	0.002
B5	Cultural act.	-0.235	0.070	0.001	0.679	0.074	0.001	1.749	0.098	0.001
C1	Cost of life	0.471	0.052	0.001	0.962	0.071	0.001	1.670	0.107	0.001
C2	Crime rate	1.345	0.096	0.001	1.867	0.109	0.001	2.775	0.146	0.001
C3	Env. damage	1.235	0.088	0.001	1.982	0.108	0.001	2.893	0.151	0.001
C4	Traffic	1.449	0.102	0.001	2.234	0.126	0.001	2.946	0.158	0.001
C5	Pollution	1.512	0.111	0.001	2.082	0.130	0.001	3.367	0.191	0.001

NOTE: $\nu = 1$ for the items on benefits and $\nu = 2$ for the items on costs, SD = standard deviation, MCSE = Monte Carlo standard error.

Table 6.3. Item parameter estimates for the multiunidimensional GRM.

The thresholds' parameters κ for each are able to reflect the criticity level of the specific aspect considered. In fact, high values for the criticity parameters correspond to lower probabilities to observe responses in higher categories, which means that the items characterized by higher criticity parameters are answered in lower categories more frequently.

For this model, it is not possible to unambiguously order the items by the response probability on the basis of the criticity parameters. But, by fixing $\hat{\theta}_1$ and $\hat{\theta}_2$ at the mean value 0, we can use these parameters to compare the probabilities of category responses for each item, and to compare probabilities to observe a response in a particular category or higher (lower) for each item. The first comparison can be carried out by calculating differences in parameters associated to adjacent thresholds, while the second comparison, which is mainly meaningful in a context of interpretation, can be carried out directly through the thresholds' parameters.

Figure 6.1[1] graphically illustrates the estimated probabilities to observe each category for each test item for a resident with an average perception of tourism benefits and costs, i.e. $\hat{\theta}_1 = 0$ and $\hat{\theta}_2 = 0$.

As an example, an individual with an average perception of tourism benefits will have a higher probability of responding higher categories to item B1 than to item B3, in fact thresholds' parameters associated to higher categories, κ_3, κ_4, κ_5 and κ_6, are regularly lower for item B1 ($\hat{\kappa}_{B1,3} = -1.40$, $\hat{\kappa}_{B1,4} = -0.50$, $\hat{\kappa}_{B1,5} = 0.17$, $\hat{\kappa}_{B1,6} = 1.35$) than to item B3 ($\hat{\kappa}_{B3,3} = -1.30$, $\hat{\kappa}_{B3,4} = -0.28$, $\hat{\kappa}_{B3,5} = 0.72$, $\hat{\kappa}_{B3,6} = 2.01$). This means that, between the advantages of economic support and public services, the first aspect is considered mainly relevant by an individual with an average perception of benefits. From Table 6.3 emerges that thresholds' parameters for item B1 related to the highest categories κ_5 and κ_6 are the lowest in the group of items on benefits. This means that the main and immediate advantages of tourism are identified by the residents in the economic support.

Analogously, a resident with an average perception of the environmental impact of tourism will have a higher probability of answering higher categories to item C1 than to the other items ($\hat{\kappa}_{C1,3} = 0.003$, $\hat{\kappa}_{C1,4} = 0.471$, $\hat{\kappa}_{C1,5} = 0.962$,

[1]NOTE: in order to represent the probabilities associated to categories 1 and 7 for each item, a lower bound of -4 and an upper bound of 4 have been fixed.

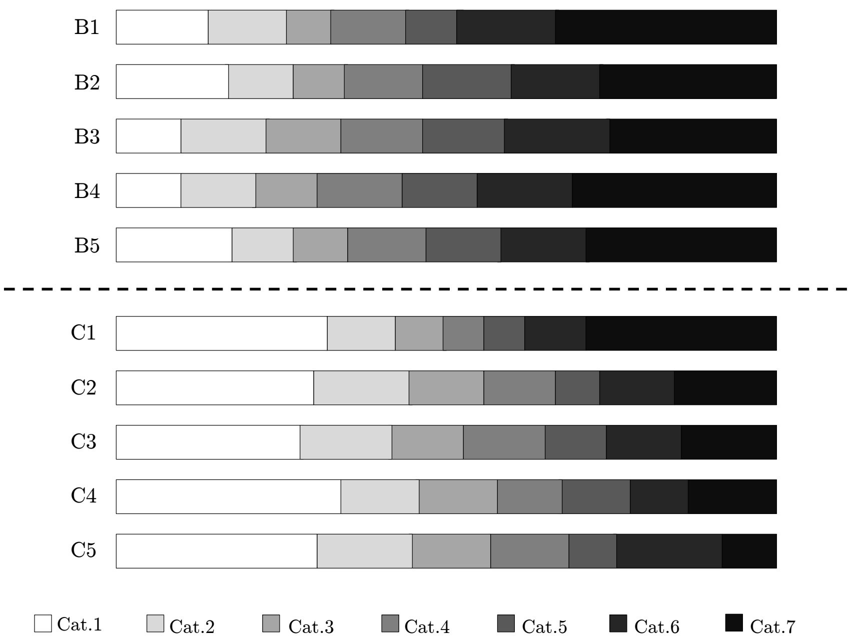


Figure 6.1. Representation of the thresholds' parameter estimates for the multiunidimensional model.

$\hat{\kappa}_{C1,6} = 1.670$). Hence, the cost of life can be regarded as a marginal negative aspect of tourism in comparison with the other issues.

The estimated correlation between the two latent traits is $\hat{r}_{12} = -0.37$. The correlation is negative and relatively high, indicating that the perception of a high economic advantage of tourism is associated with a strongly negative environmental impact. As a conclusive remark, we can observe that the individuals show a different evaluation of the benefits vs. costs, revealing a critical view of the tourism industry, and the multiunidimensional GRM is able to capture this peculiarity.

6.4 Results for the additive GRM

In order to extend the structure of the multiunidimensional model with the inclusion of a general trait that directly affects all the item responses, we estimated the parameters of the additive GRM. A general latent trait θ_0 is added to the specific traits θ_1 and θ_2. The two specific traits have the same interpretation as in the multiunidimensional model, namely perceptions of benefits and costs, while the general trait can be defined as the overall attitude towards tourism.

The foundation is that the general trait is estimated on the basis of the perception of either benefits and costs but conditionally on the specific effects of the two traits, and allowing other residual factors (age, gender, place of residence, occupation,...) to influence the measure of the overall attitude. Concerning the score interpretation, higher scores in the attitude are related to residents who perceive higher advantages and a lower negative impact of tourism.

The model parameters were estimated by using the proposed OpenBUGS procedure for the additive GRM, with two chains and 30,000 iterations (15,000 burn-in) for each one. The item parameter estimates for the additive model are illustrated in Table 6.4. The additive model requires, for each item, the estimation of the general discrimination parameter ($\alpha_{0\nu}$), the specific discrimination parameter (α_ν) and the criticity parameters (κ). Again, items cannot be unambiguously ordered on the basis of the response probabilities.

Figure 6.2 graphically illustrates the estimated probabilities to observe each category for each test item for a resident with an average perception of tourism benefits and costs, i.e. a resident characterized by $\hat{\theta}_0 = 0$ and $\hat{\theta}_1 = 0$ for items on benefits and a resident characterized by $\hat{\theta}_0 = 0$ and $\hat{\theta}_2 = 0$ for items on costs.

Concerning the group of items on benefits, the economic support (B1) and job opportunities (B4) are associated with higher probabilities of responses in higher categories, because the corresponding estimates for the thresholds' parameters are generally lower than for the remaining items ($\hat{\kappa}_{B1,3} = -1.47$, $\hat{\kappa}_{B1,4} = -0.51$, $\hat{\kappa}_{B1,5} = 0.20$, $\hat{\kappa}_{B1,6} = 1.46$ and $\hat{\kappa}_{B4,3} = -1.60$, $\hat{\kappa}_{B4,4} = -0.56$, $\hat{\kappa}_{B4,5} = 0.41$, $\hat{\kappa}_{B4,6} = 1.64$)). This arrangement means that residents who have an average general perception of advantages and an average specific perception of advantages consider the economic development and the job opportunities as the main advantages of tourism.

Moreover, among the items on costs, again the cost of life (C1) is characterised by generally lower thresholds' parameters in comparison to the estimated criticity levels of other items, especially with reference to higher categories ($\hat{\kappa}_{C1,3} = 0.00$, $\hat{\kappa}_{C1,4} = 0.47$, $\hat{\kappa}_{C1,5} = 0.96$ and $\hat{\kappa}_{C1,6} = 1.68$). So that, the cost of life seems to be the least important impact of the tourism industry for a typical respondent.

Item	Item description	$\widehat{\alpha}_\nu$	SD($\widehat{\alpha}_\nu$)	MCSE($\widehat{\alpha}_\nu$)	$\widehat{\kappa}_1$	SD($\widehat{\kappa}_1$)	MCSE($\widehat{\kappa}_1$)	$\widehat{\kappa}_2$	SD($\widehat{\kappa}_2$)	MCSE($\widehat{\kappa}_2$)	$\widehat{\kappa}_3$	SD($\widehat{\kappa}_3$)	MCSE($\widehat{\kappa}_3$)
B1	Econ. support	1.047	0.074	0.001	- 3.049	0.169	0.002	- 2.014	0.105	0.001	- 1.469	0.087	0.001
B2	Quality of life	0.946	0.063	0.001	- 2.539	0.125	0.001	- 1.789	0.090	0.001	- 1.176	0.073	0.001
B3	Public services	1.247	0.082	0.001	- 3.278	0.188	0.002	- 2.200	0.119	0.001	- 1.306	0.088	0.001
B4	Job. opp.	1.290	0.082	0.001	- 3.342	0.187	0.002	- 2.390	0.125	0.001	- 1.595	0.094	0.001
B5	Cultural act.	1.194	0.077	0.001	- 2.713	0.141	0.001	- 1.901	0.103	0.001	- 1.256	0.084	0.001
C1	Cost of life	0.284	0.042	0.000	- 1.491	0.069	0.000	- 0.644	0.049	0.000	- 0.002	0.046	0.000
C2	Crime rate	1.534	0.109	0.002	- 1.824	0.123	0.002	- 0.534	0.085	0.002	0.461	0.083	0.001
C3	Env. damage	1.343	0.090	0.001	- 1.901	0.114	0.002	- 0.745	0.080	0.001	0.205	0.073	0.001
C4	Traffic	1.487	0.126	0.004	- 1.509	0.134	0.005	- 0.397	0.092	0.003	0.700	0.098	0.002
C5	Pollution	1.425	0.103	0.002	- 1.646	0.114	0.003	- 0.452	0.083	0.002	0.548	0.085	0.002

Item	Item description	$\widehat{\alpha}_{0\nu}$	SD($\widehat{\alpha}_{0\nu}$)	MCSE($\widehat{\alpha}_{0\nu}$)	$\widehat{\kappa}_4$	SD($\widehat{\kappa}_4$)	MCSE($\widehat{\kappa}_4$)	$\widehat{\kappa}_5$	SD($\widehat{\kappa}_5$)	MCSE($\widehat{\kappa}_5$)	$\widehat{\kappa}_6$	SD($\widehat{\kappa}_6$)	MCSE($\widehat{\kappa}_6$)
B1	Econ. support	0.013	0.012	0.000	- 0.507	0.066	0.001	0.204	0.064	0.001	1.458	0.086	0.001
B2	Quality of life	0.250	0.073	0.001	- 0.250	0.061	0.001	0.802	0.067	0.001	1.871	0.094	0.001
B3	Public services	0.446	0.095	0.002	- 0.264	0.071	0.001	0.760	0.078	0.001	2.066	0.116	0.002
B4	Job. opp.	0.144	0.083	0.002	- 0.560	0.073	0.001	0.405	0.071	0.001	1.644	0.097	0.002
B5	Cultural act.	0.343	0.094	0.002	- 0.228	0.069	0.001	0.732	0.074	0.001	1.856	0.103	0.001
C1	Cost of life	0.017	0.016	0.000	0.470	0.048	0.000	0.964	0.054	0.000	1.676	0.075	0.000
C2	Crime rate	0.074	0.060	0.001	1.470	0.106	0.002	2.068	0.128	0.002	3.083	0.179	0.003
C3	Env. damage	0.051	0.049	0.001	1.299	0.091	0.001	2.093	0.116	0.002	3.089	0.167	0.002
C4	Traffic	0.906	0.144	0.005	1.660	0.137	0.004	2.566	0.186	0.006	3.381	0.240	0.007
C5	Pollution	0.668	0.113	0.003	1.520	0.110	0.002	2.102	0.133	0.003	3.413	0.200	0.004

NOTE: $\nu = 1$ for the items on benefits and $\nu = 2$ for the items on costs, SD = standard deviation, MCSE = Monte Carlo standard error.

Table 6.4. Item parameter estimates for the additive GRM.

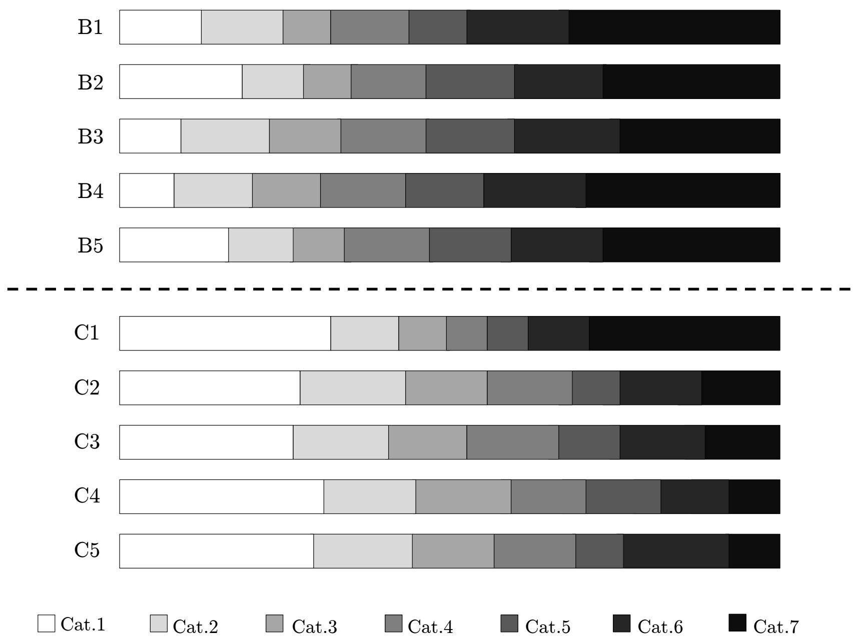


Figure 6.2. Representation of the thresholds' parameter estimates for the additive model.

Focusing on the discrimination parameters, concerning the estimated specific discrimination parameters, results are similar to the multiunidimensional case: the most informative items on the specific perception of tourism benefits are public services, job opportunities and cultural activities (items B3, B4 and B5, respectively), while crime rate, traffic and pollution (items C2, C4 and C5, respectively) are the items that better discriminate respondents with different levels of specific perception of tourism costs.

Higher values of estimated general discrimination parameters are associated to public services (item B3) and cultural activities (item B5) regarding the benefits, and to traffic (item C4) and pollution (item C5) among the items on costs of tourism industry. Consequently, these aspects principally influence the general residents' attitude towards tourism.

Usually, the additive model fits the data better than the multiunidimensional model because the presence of an overall latent trait is generally supported by data. In fact, also for our data a lower DIC is associated with the additive model (DIC= 8945) in comparison to the multiunidimensional model (DIC=10950).

Analogously to the previous model, we estimated the correlations between the latent variables (θ_1, θ_2, and θ_0) of the additive model. The results are $\hat{r}_{01} = 0.03$, $\hat{r}_{02} = 0.18$ and $\hat{r}_{12} = -0.62$. The correlation between the benefit and cost latent traits is negative as in the multiunidimensional model.

The correlation between the benefit latent trait and the attitude is very low, and slightly higher is the estimated correlation between the cost latent trait and the general attitude.

6.5 Heterogeneity in resident perceptions

The multiunidimensional and additive models presented in this work are specified without considering the presence of covariates. Of course, once the measurement process is carried out, latent constructs may result in different scores according to the characteristics of the examinees. In order to face this issue, we perform an analysis of the scores for the general and specific latent traits obtained from the additive model.

Therefore, to investigate the importance of the individuals' heterogeneity in the evaluation of tourism attitudes, the score distributions[2] of the general and specific latent traits are calculated and compared on the basis of some socio-demographic characteristics (Table 6.5).

Residents show, on average, a positive attitude towards tourism (0.62) and a higher perception of benefits (0.57) compared with the costs (0.48).

From Table 6.5 we can observe how the youngest people have both a significant personal attitude tward tourism and a critical perception of the tourism industry: a high score in the perception of benefits is associated to a low score on the perception of costs. This means that the youngest are conscious of the advantages related to tourism, but at the same time, they strongly evaluate the negative effects of the industry on the community. On the contrary, respondents with a low level of education and elderly people show a high attitude towards tourism and a small gap between the benefit and cost scores.

The area of residence also affects the evaluation of the tourism industry. In fact, residents in the tourism municipalities and provinces (Rimini and San

[2] As the scores have a different range, they have been normalized to the range of 0 to 1.

	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$
Age			
< 25	0.57	0.61	0.45
25 - 35	0.61	0.56	0.53
35 - 45	0.61	0.55	0.46
45 - 55	0.63	0.57	0.50
55 - 65	0.64	0.58	0.50
≥ 65	0.62	0.57	0.48
Gender			
Female	0.62	0.58	0.47
Male	0.61	0.55	0.50
Education			
Primary	0.62	0.57	0.48
Lower secondary	0.65	0.57	0.52
Upper secondary	0.60	0.56	0.45
University	0.60	0.57	0.48
Provinces			
Forlì-Cesena	0.65	0.54	0.52
Ravenna	0.65	0.53	0.53
Rimini	0.58	0.62	0.41
San Marino	0.50	0.62	0.40
Typological locality			
Main town	0.60	0.57	0.49
Tourism municipality	0.63	0.59	0.47
Other urban city	0.64	0.55	0.49
Total	0.62	0.57	0.48

Table 6.5. Normalized mean perception and attitude scores by age, gender, education, province and typological area.

Marino), where the seaside tourism is relevant, present a high gap between the benefit and cost scores.

This first research, that has been repeated in 2013, furnishes interesting suggestions for the development of incentive tourism policies, which are also related to the well-being.

Chapter 7

Conclusions

This work falls within the context of item response theory (IRT). In particular, it focuses on models for ordinal data. The importance of developing models for ordinal data is relevant not only from a theoretical perspective. Actually, several fields of application are characterized by ordinal manifest variables and the use of proper models for ordinal data allows to avoid the loss of information due to the dichotomization process. IRT is widely used in psychological and educational fields, but it also shows a great potential in applications within behavioral sciences, where data are often ordinal.

In the past, a common assumption was the presence of a single latent construct underlying the response process. However, real data typically suggest a multidimensional structure. So that, multidimensional IRT (MIRT) models have been recently developed, taking into account the complexity of real data and allowing for the presence of more than one latent trait.

In this work we focus on MIRT models for ordinal data with complex latent structures. Indeed, numerous MIRT models can be specified according to several conditions, and one of them is the hypothesized underlying latent structure. The models proposed in this work are extensions of the unidimensional graded response model (GRM) (Samejima, 1969) and are characterized by multidimensional latent structures with correlated traits. In particular, we consider the multiunidimensional structure, where the item responses are affected by specific traits, and the additive structure, where the item responses are simultaneously affected by a general and specific traits.

Then, we considered two model: the multiunidimensional and the additive GRMs with correlated traits. This choice has been driven by the fact that the first one is widely used and represents a classical approach in MIRT analysis, while the second one is able to reflect the complexity of real interactions between items and respondents.

Due to the complexity of the models proposed, another important aspect of this work concerns the estimation procedure. Within a Bayesian approach, we propose a Markov chain Monte Carlo (MCMC) procedure for parameter estimation, which permits to overtake the problem of analytically intractable expressions. Models are implemented using the open-source software OpenBUGS. This software, allowing for a flexible and rather easy implementation, represents a good solution for estimation issues.

In order to assess the item parameter recovery for both multiunidimensional and additive GRMs we perform a simulation study. The simulation study is conducted on a bidimensional case by varying the simulation conditions, that are: the number of response categories, the sample size, the test and subtest lengths and the latent trait correlation structure. Concisely, the main simulation results showed that the parameter recovery is particularly sensitive to the sample size, due to the model complexity and the high number of parameters to be estimated. For a sufficiently large sample size the parameters of the multiunidimensional and additive GRMs are well reproduced. The results are also affected by the trade-off between the number of items constituting the test and the number of item categories: the worst results are associated to a high number of categories and a low test length. Analogous evidences apply for the latent trait correlation estimates.

In order to verify the actual applicability of the proposed models in real situations, we estimated them on empirical data. Data were collected with the aim to investigate Romagna and San Marino residents' perceptions and attitudes towards the tourism industry. A relevant advantage of the proposed models concerns the possibility to use the data collected without any preliminary transformation, hence without any loss of information.

Some limitations of the research regarding the application study exist, in particular the choice of the prior distributions, the sample size, the number of item categories, the test and subtests lengths, are important issues that have to be always considered and checked.

Lastly, concerning the future works to be done on the MIRT models for ordinal data and correlated traits, first of all it could be interesting to perform further simulations with an increased number of latent dimensions. Secondly, this work focuses on two specific underlying latent structures, hence an extension to different (i.e. hierarchical or high-orders) structures represent a stimulating issue. A final extension could consider the introduction of covariates in the model specification, independently from the underlying structure considered.

Bibliography

T.A. Ackerman. Insuring the validity of the reported score scale by reporting multiple scores. *Paper presented at the North American Meeting of the Prychometric Society*, 1993.

A.J. Adams, M. Wilson, and W-C Wang. The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1): 1–23, 1997.

J.H. Albert. Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, 17(3):251–269, 1992.

E.B. Andersen. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32: 283–301, 1970.

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

D. Andrich. An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, 47(1):105–113, 1982.

R.J. De Ayala. *The theory and practice of item response theory.* Guilford Press, 2009.

Mr. Bayes and MR. Price. An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions (1683-1775)*, pages 370–418, 1763.

A.A. Béguin and C.A.W. Glas. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66(4):541–562, 2001.

C. Bernini, A. Guizzardi, and G. Angelini. DEA-like model and common weights approach for the construction of a subjective community well-being indicator. *Social indicators research*, 114(2):405–424, 2013.

A. Birnbaum. Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and M.R. Novick, editors, *Statistical theories of mental test scores*. Addison-Wesley, Reading MA, 1968.

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

S.P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4): 434–455, 1998.

A.R. Brown, S.J. Finney, and M.K. France. Using the bifactor model to assess the dimensionality of the Hong Psychological Reactance Scale. *Educational and Psychological Measurement*, 71(1):170–185, 2011.

F. F. Chen, S. G. West, and K. H. Sousa. A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2):189–225, 2006.

S.M. Curtis. BUGS code for item response theory. *Journal of Statistical Software*, 36(1):1–34, 2010.

M.C. Edwards. A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika*, 75(3):474–497, 2010a.

M.C. Edwards. MultiNorm User's Guide: multidimensional normal ogive item response theory models. *mimeo*, 2010b.

P.J. Ferrando. Likert scaling using continuous, censored, and graded response models: effects on criterion-related validity. *Applied Psychological Measurement*, 23(2):161–175, 1999.

G.H. Fischer. Derivations of the rasch model. In G.H. Fischer and I.W. Molenaar, editors, *Rasch models: Foundations, recent developments and applications*. Springer-Verlag, New York, 1995.

J-P Fox. *Bayesian item response modeling*. Springer, 2010.

J-P Fox and C.A.W. Glas. Bayesian estimation of a multilevel of an IRT model using Gibbs sampling. *Psychometrika*, 66(2):271–288, 2001.

Z-H Fu, J. Tao, and N-Z Shi. Bayesian estimation of the multidimensional graded response model with nonignorable missing data. *Journal of Statistical Computation and Simulation*, 80(11):1237–1252, 2010.

D. Gamerman. *Markov Chain Monte Carlo*. Chapman and Hall, London, 1997.

A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of American Statistical Association*, 85:398–409, 1990.

A. Gelman. Objections to Bayesian statistics. *Bayesian Analysis*, 3(3):445–449, 2008.

A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. CRC press, London, 2003.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactons on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

R.D. Gibbons, A.J. Rush, and J.C. Immekus. On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. *Journal of psychiatric research*, 43(4):401–410, 2009.

W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall, London, 1996.

H. Gu and C. Ryan. Place attachment, identity and community impacts of tourism-the case of a Beijing hutong. *Tourism management*, 29(4):637–647, 2008.

D.C. Haley. Estimation of the dosage mortality relationship when the dose is subject to error. Technical Report 15 (Office of Naval Research Contract No. 25140, NR-342-022), Stanford University: Applied Mathematics and Statistics Laboratory, 1952.

R.K. Hambleton and H. Swaminathan. *Item Response Theory: Principles and applications.* Kluwer Nijhoff Publishing, Boston, 1985.

R.K. Hambleton, H. Swaminathan, and H.J. Rogers. *Fundamentals of Item Response Theory.* Sage Publications, Newbury Park, CA, 1991.

W.K. Hastings. Monte Carlo simulation methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

K.J. Holzinger and F. Swineford. The bi-factor method. *Psychometrika*, 2(1): 41–54, 1937.

H-Y Huang, W-C Wang, P-H Chen, and C-M Su. Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement*, 37 (8):619–637, 2013.

M.S. Johnson. Marginal maximum likelihood estimation of item response models in R. *Journal of Statistical Software*, 20(10):1–24, 2007.

R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 22(140), 1932.

F.M. Lord. A theory of test scores. *Psychometric Monograph No. 7*, 1952.

F.M. Lord. *Applications of item response theory to practical testing problems.* Lawrence Erlbaum, Hillsdale, NJ, 1980.

F.M. Lord and M.R. Novick. *Statistical theories of mental test scores.* Addison-Wesley, Reading, MA, 1968.

D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067, 2009.

D. Lunn, C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter. *The BUGS Book*. CRC Press, Taylor & Francis Group, 2013.

D.J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337, 2000.

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

M. Matteucci. *Item Response Theory models for the competence evaluation: towards a multidimensional approach in the University guidance*. PhD thesis, University of Bologna, 2007.

N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E.Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1953.

E. Muraki. A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, 16(2):159–176, 1992.

E. Muraki. POLYFACT [computer software]. *Princeton, NJ: Educational Testing Service*, 1993.

E. Muraki and J.E. Carlson. Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, 19(1):73–90, 1995.

I. Ntzoufras. *Bayesian modeling using WinBUGS*, volume 698. Wiley. com, 2011.

R. Ostini and M.L. Nering. *Polytomous Item Response Theory Models (Quantitative Applications in the Social Sciences, Vol. 144)*. Sage Publications, Thousand Oaks, CA, 2006.

R.J. Patz and B.W. Junker. Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, 24(4):342–366, 1999a.

R.J. Patz and B.W. Junker. A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24(2):146–178, 1999b.

M. Plummer. JAGS Version 2.2.0 user manual, 2010.

G. Rasch. *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen, 1960.

M.D. Reckase. The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9(4):401–412, 1985.

M.D. Reckase. The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21(1):25–36, 1997.

M.D. Reckase. *Multidimensional Item Response Theory*. Springer, London, 2009.

B.B. Reeve. An introduction to modern measurement theory. *National Cancer Institute*, 2002.

J. Rost. Measuring attitudes with a threshold model drawing on a traditional scaling concept. *Applied Psychological Measurement*, 12(4):397–409, 1988.

F. Samejima. Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph Supplement No. 17*, 1969.

F. Samejima. Evaluation of mathematical models for ordered polychotomous responses. *Behaviormetrika*, 23(1):17–35, 1996.

J. Schmid and J.M. Leiman. The development of hierarchical factor solutions. *Psychometrika*, 22(1):53–61, 1957.

Y. Sheng. A MATLAB package for Markov chain Monte Carlo with a multi-unidimensional IRT model. *Journal of Statistical Software*, 28(10):1–20, 2008.

Y. Sheng. Bayesian estimation of MIRT models with general and specific latent traits in MATLAB. *Journal of Statistical Software*, 34(3):1–26, 2010.

Y. Sheng and C.K. Wikle. Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 67 (6):899–919, 2007.

Y. Sheng and C.K. Wikle. Bayesian multidimensional IRT models with a hierarchical structure. *Educational and Psychological Measurement*, 68(3):413–430, 2008.

Y. Sheng and C.K. Wikle. Bayesian IRT models incorporating general and specific abilities. *Behaviormetrika*, 36(1):27–48, 2009.

D. Spiegelhalter, A. Thomas, N. Best, and W. Gilks. BUGS 0.5: Bayesian inference using Gibbs sampling - Manual (version ii). *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*, 1996.

L. Steinberg and D. Thissen. *Item response theory in personality research.* Lawrence Erlbaum Associates Hillsdale, NJ, 1995.

R. Tate. A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement*, 27(3):159–203, 2003.

D. Thissen and L. Steinberg. A response model for multiple choice items. *Psychometrika*, 49(4):501–519, 1984.

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

D. Thissen, L. Nelson, K. Rosa, and L.D. McLeod. Item response theory for items scored in more than two categories. In D. Thissen and H. Wainer, editors, *Test scoring*, pages 141–186. Psychology Press, 2001.

A. Thomas, B. O'Hara, U. Ligges, and S. Sturtz. Making BUGS Open. *R News*, 6(1):12–17, 2006. URL `http://cran.r-project.org/doc/Rnews/`.

R.E. Traub. A priori considerations in choosing an item response model. In R.K. Hambleton, editor, *Applications of item response theory*, pages 57–70. Educational Research Institute of British Columbia, Vancouver, BC, 1983.

L.A. van der Ark. Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement*, 25(3):273–282, 2001.

L.A. van der Ark, D.W. van der Palm, and K. Sijtsma. A latent class approach to estimating test-score reliability. *Applied Psychological Measurement*, 35(5): 380–392, 2011.

W.J. van der Linden and R.K. Hambleton. *Handbook of Modern Item Response Theory*. Springer-Verlag, New York, 1997.

W-C Wang, P-H Chen, and Y-Y Cheng. Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, 9(1):116–136, 2004.

L. Yao and R.D. Schwarz. A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied psychological measurement*, 30(6):469–492, 2006.

Appendix A

OpenBUGS code for implemented models

A.1 OpenBUGS code: multiunidimensional and additive models for graded responses

In this section we report the codes used to implement the multiunidimensional model and the additive model for graded responses.

Initial values for the following quantities have to be set and loaded from the user before to run the models: m.theta, Sigma.theta, m.alpha, s.alpha, m.alpha0, s.alpha0, m.kappa and s.kappa (of course, m.alpha0 and s.alpha0 are referred only to the additive model).

Multiunidimensional Model for Graded Responses - OpenBUGS code

```
model{
	pr.alpha <- pow(s.alpha, -2)			# alpha-precision
	pr.kappa <- pow(s.kappa, -2)			# kappa-precision
	Pr.theta[1:m, 1:m] <- inverse(Sigma.theta[,])	# theta-precision

	for (j in 1:p){

		alpha[j] ~ dnorm(m.alpha, pr.alpha)I(0,)		# alpha-prior: iid truncated normal

		for (k in 1:(K[j]-1)){
			kappa.star[j,k] ~ dnorm(m.kappa, pr.kappa)	# k.star-prior: iid normal
			kappa[j,k] <- ranked(kappa.star[j, 1:(K[j]-1)],k)	# obtain k from k.star
		}
	}

	for (i in 1:n) {

		theta[i, 1:m] ~ dmnorm(m.theta[],Pr.theta[,])	# theta-prior: multivariate normal

			for (j in 1:p){			# cumulative probabilities computation
				for (k in 1:(K[j]-1)){	# probit model according to
							# multiunidimensional structure

					P[i,j,k] <- phi(kappa[j,k] - alpha[j] * theta[i, vv[j]])
				}
				P[i,j,K[j]] <- 1
			}

			for(j in 1:p){			# probability computation from
							# cumulative probabilities
				prob[i,j,1] <- P[i,j,1]
				for (k in 2:K[j]){
					prob[i,j,k] <- P[i,j,k] - P[i,j,k-1]
				}
				Y[i,j] ~ dcat(prob[i,j, 1:K[j]])	# likelihood: categorical (generalized Bernoulli)
							# distribution
			}
	}
}
```

Additive Model for Graded Responses - OpenBUGS code

```
model{
	pr.alpha <- pow(s.alpha, -2)				# alpha-precision
	pr.alpha0 <- pow(s.alpha0, -2)				# alpha0-precision
	pr.kappa <- pow(s.kappa, -2)				# kappa-precision
	Pr.theta[1:m+1, 1:m+1] <- inverse(Sigma.theta[,])	# theta-precision

	for (j in 1:p){

		alpha[j] ~ dnorm(m.alpha, pr.alpha)I(0,)		# alpha-prior: iid truncated normal
		alpha0[j] ~ dnorm(m.alpha0, pr.alpha0)I(0,)	# alpha0-prior: iid truncated normal

		for (k in 1:(K[j]-1)){
			kappa.star[j,k] ~ dnorm(m.kappa, pr.kappa)	# k.star-prior: iid normal
			kappa[j,k] <- ranked(kappa.star[j, 1:(K[j]-1)],k)	# obtain k from k.star
		}
	}

	for (i in 1:n) {

		theta[i, 1:m] ~ dmnorm(m.theta[],Pr.theta[,])	# theta-prior: multivariate normal

			for (j in 1:p){			# cumulative probabilities computation
				for (k in 1:(K[j]-1)){	# probit model according to
							# additive structure

				P[i,j,k] <- phi(kappa[j,k] - alpha[j] * theta[i, vv[j]+1] - alpha0[j]) *  theta[i,1])
				}
				P[i,j,K[j]] <- 1
			}

			for(j in 1:p){			# probability computation from
							# cumulative probabilities
				prob[i,j,1] <- P[i,j,1]
				for (k in 2:K[j]){
					prob[i,j,k] <- P[i,j,k] - P[i,j,k-1]
				}
				Y[i,j] ~ dcat(prob[i,j, 1:K[j]])	# likelihood: categorical (generalized Bernoulli)
							# distribution
			}
	}
}
```

Appendix B

R procedures for the simulation study

The following sections report the codes used to perform the simulation study for both the multiunidimensional and the additive GRMs. For each model, the procedure about a single scenario (i.e. with particular simulation conditions that can be set at the beginning of the procedure) is described.

The simulation study has been conducted by using an R procedure to generate the objects of interest, and by recalling OpenBUGS trough the R package BRugs. The main advantage of the combined use of R and OpenBUGS consists in the possibility to create an automatic routine to complete all replications within a distinct scenario.

For further details about all the available functions and features of the package BRugs, see Thomas et al. (2006).

B.1 Multiunidimensional GRM: R code

```
# Multiunidimensional Graded Response Model – Simulation procedure (R code)
library(msm)
library(BRugs)
library(MASS)

# Simulation conditions:
n = 1000                                  # number of examinees – Indexed by i
m = 2                                     # number of latent traits (and subtests) – Indexed by v
p1 = 5                                    # first subtest length
p2 = 10                                   # second subtest length
p <- p1 + p2; p                           # test length – Indexed by j
K = c(3,3,3,3,3,3,3,3,3,3,3,3,3,3,3)      # Number of categories for each item
vv = c(1,1,1,1,1,2,2,2,2,2,2,2,2,2,2)     # Mapping vector, used to associate each item
                                          # to its corresponding subtest

# Input values for hyperparameters in prior distributions (according to the identification constraints)
m.alpha <- 0                              # alpha mean
s.alpha <- 1                              # alpha variance
sd.alpha <- sqrt(s.alpha)                 # alpha standard deviation

m.kappa <- 0                              # kappa mean
s.kappa <- 1                              # kappa variance
sd.kappa <- sqrt(s.kappa)                 # kappa standard deviation

m.theta <- rep(0,m)                       # theta mean
one <- rep(1,m)                           # variance/covariance matrix for latent traits (for uncorrelated traits)
Sigma.theta <- diag(one)
# Sigma.theta <- matrix(c(1, 0.4, 0.4, 1),2)        # (use this line for correlated traits)

# Initialization of auxiliary objects
kappa.star <- array(rep(0), c(p,(max(K)-1)))   # Initialization matrices of threshold values (unconstrained)
kappa <- array(rep(0), c(p,(max(K)-1)))        # Initialization matrices of threshold values (constrained)

eta <- array(rep(0), c(n,p,max(K)))            # Initialization array with predictors, eta
P <- array(rep(0), c(n,p,max(K)))              # Initialization array with cumulative probabilities, P
prob <- array(rep(0), c(n,p,max(K)))           # Initialization array with probabilities, prob

Y_q <- array(rep(0), c(n,p))              # Initialization (n x p) data matrix with generated responses

# Generation of:
# 1) theta, (n x m)-matrix , from its prior distribution (multivariate normal distribution)
set.seed(1)
theta <- mvrnorm(n, m.theta, Sigma.theta)
```

```
# 2) alpha, p-vector, from its prior distribution (truncated normal distribution)
set.seed(2)
alpha <- rtnorm(p, mean=m.alpha, sd=sd.alpha, lower=0); alpha

# 3) kappa.star, (p, max(K)-1)-matrix, from its prior distribution (normal distribution)
# The matrix kappa of constrained values is then obtained by sorting kappa.star
set.seed(5)
for(j in 1:p){
  kappa.star[j,1:(K[j]-1)] <- rnorm((K[j]-1),mean=m.kappa, sd=sd.kappa)
  kappa[j,1:(K[j]-1)] <- sort(kappa.star[j,1:(K[j]-1)])
}

# Storing of the generated (real) values
write.table(theta,'theta.txt')
write.table(alpha,'alpha.txt')
write.table(kappa,'kappa.txt')

# Computation of P and prob, according to the multiunidimensional GRM
set.seed(54)
for(i in 1:n){
	for(j in 1:p){
		for(k in 1:(K[j]-1)){
			eta[i,j,k] <- kappa[j,k] - alpha[j]*theta[i,vv[j]]	# predictor
			P[i,j,k] <- pnorm(eta[i,j,k])			# P = Phi(predictor)
		}
		P[i,j,K[j]] <- 1

		prob[i,j,1] <- P[i,j,1]
		for(k in 2:K[j]){
			prob[i,j,k] <- P[i,j,k] - P[i,j,k-1]
		}
	}
}

# Data containing two different set of initial values
theta01 <- matrix(0, n, m)
alpha01 <- rep(0, p)
kappa.star01 <- array(0, c(p,(max(K)-1)))
bugsData(list(alpha=alpha01, kappa.star=kappa.star01, theta=theta01),"Inits1.txt")

theta02 <- matrix(1, n, m)
alpha02 <- rep(1, p)
kappa.star02 <- array(1, c(p,(max(K)-1)))
bugsData(list(alpha=alpha02, kappa.star=kappa.star02, theta=theta02),"Inits2.txt")
```

 # Multiunidimensional GRM model estimation by using OpenBUGS (for Q replications)

```
Q <- 10          # Number of replication for each scenario (distinct simulation conditions)
for(q in 1:Q){
	dataY_q <- paste(q,"-Y.txt",sep="")		# To add the replication number in files with results
	set.seed(q)					# Seed must changes at each iteration
							# to obtain different samples
	for(i in 1:n){
		for(j in 1:p){
			Y_q[i,j] <- sample(1:K[j], size = 1, prob = prob[i,j,])   # Generate a data matrix
		}							     # by using the stored prob
	}
	write.table(Y_q,quote=FALSE,sep=" ",dataY_q)    # Storing of matrix Y_q

	bugsData(list(Y=Y_q, n=n, p=p, K=K, m=m, vv=vv, m.theta=m.theta, Sigma.theta=Sigma.theta,
	m.alpha=m.alpha, s.alpha=s.alpha, m.kappa=m.kappa, s.kappa=s.kappa),"Data.txt")

	stats_q <- paste(q,"-stats.txt",sep="")		# To add the replication number in files with results
	dic_q <- paste(q,"-dic.txt",sep="")

	sim_q <- BRugsFit(modelFile = "ModelloGRMmultiuni.txt", data = "Data.txt", inits = c("Inits1.txt",
	"Inits2.txt"), numChains = 2, parametersToSave = c("theta", "alpha", "kappa"), nBurnin = 15000,
	nIter = 30000, nThin = 1, DIC = TRUE, working.directory = NULL, digits = 5)

	str(sim_q)
	write.table(sim_q$Stats,quote=FALSE,sep=", ",stats_q)		# Save results
	write.table(sim_q$DIC,quote=FALSE,sep=", ",dic_q)		# Save results
}
```

 # Summary matrices with estimated parameters (posterior mean and sd).
 # Each column is referred to a replication.

```
stat_1 <- read.table("1-stats.txt",header=TRUE)
mean <- matrix(stat_1[,1])
sd <- matrix(stat_1[,2])
for(q in 2:Q) {
	infile <- paste(q,"-stats.txt",sep="")
	stat_q <- read.table(infile,header=TRUE)
	mean <- cbind(mean,matrix(stat_q[,1]))
	sd <- cbind(sd,matrix(stat_q[,2]))
}
write.table(mean, 'SummaryMean.txt')
write.table(sd, 'SummarySd.txt')
```

B.2 Additive GRM: R code

```
# Additive Graded Response Model – Simulation procedure (R code)
library(msm)
library(BRugs)
library(MASS)

# Simulation conditions:
n = 1000                        # number of examinees – Indexed by i
m = 2                           # number of specific latent traits (and subtests) – Indexed by v
p1 = 5                          # first subtest length
p2 = 10                         # second subtest length
p <- p1 + p2; p                 # test length – Indexed by j
K = c(3,3,3,3,3,3,3,3,3,3,3,3,3,3,3)  # Number of categories for each item
vv = c(1,1,1,1,1,2,2,2,2,2,2,2,2,2,2) # Mapping vector, used to associate each item
                                # to its corresponding subtest

# Input values for hyperparameters in prior distributions (according to the identification constraints)
m.alpha <- 0                    # alpha mean
s.alpha <- 1                    # alpha variance
sd.alpha <- sqrt(s.alpha)       # alpha standard deviation

m.alpha0 <- 0                   # alpha0 mean
s.alpha0 <- 1                   # alpha0 variance
sd.alpha0 <- sqrt(s.alpha0)     # alpha0 standard deviation

m.kappa <- 0                    # kappa mean
s.kappa <- 1                    # kappa variance
sd.kappa <- sqrt(s.kappa)       # kappa standard deviation

m.theta <- rep(0,m+1)           # theta mean (m latent abilities + 1 general ability)
one <- rep(1,m+1)               # variance/covariance matrix for latent traits (for uncorrelated traits)
Sigma.theta <- diag(one)
# Sigma.theta <- matrix(c(1, 0.4, 0.3, 0.4, 1, 0.2, 0.3, 0.2, 1),3)     # (use this line for correlated traits)

# Initialization of auxiliary objects
kappa.star <- array(rep(0), c(p,(max(K)-1)))   # Initialization matrices of threshold values (unconstrained)
kappa <- array(rep(0), c(p,(max(K)-1)))        # Initialization matrices of threshold values (constrained)

eta <- array(rep(0), c(n,p,max(K)))            # Initialization array with predictors, eta
P <- array(rep(0), c(n,p,max(K)))              # Initialization array with cumulative probabilities, P
prob <- array(rep(0), c(n,p,max(K)))           # Initialization array with probabilities, prob

Y_q <- array(rep(0), c(n,p))    # Initialization (n x p) data matrix with generated responses
```

```
# Generation of:
# 1) theta, (n x m)-matrix , from its prior distribution (multivariate normal distribution)
set.seed(1)
theta <- mvrnorm(n, m.theta, Sigma.theta)

# 2) alpha, p-vector, from its prior distribution (truncated normal distribution)
set.seed(2)
alpha <- rtnorm(p, mean=m.alpha, sd=sd.alpha, lower=0)

# 3) alpha0, p-vector, from its prior distribution (truncated normal distribution)
set.seed(4)
alpha0 <- rtnorm(p, mean=m.alpha0, sd=sd.alpha0, lower=0)

# 4) kappa.star, (p, max(K)-1)-matrix, from its prior distribution (normal distribution)
# The matrix kappa of constrained values is then obtained by sorting kappa.star
set.seed(5)
for(j in 1:p){
  kappa.star[j,1:(K[j]-1)] <- rnorm((K[j]-1),mean=m.kappa, sd=sd.kappa)
  kappa[j,1:(K[j]-1)] <- sort(kappa.star[j,1:(K[j]-1)])
}

# Storing of the generated (real) values
write.table(theta,'theta.txt')
write.table(alpha,'alpha.txt')
write.table(alpha0,'alpha0.txt')
write.table(kappa,'kappa.txt')

# Computation of P and prob, according to the multiunidimensional GRM
set.seed(54)
for(i in 1:n){
        for(j in 1:p){
                for(k in 1:(K[j]-1)){
                        eta[i,j,k] <- kappa[j,k] - alpha[j]*theta[i,vv[j]+1] - alpha0[j]*theta[i, 1]  # predictor
                        P[i,j,k] <- pnorm(eta[i,j,k])                                                 # P = Phi(predictor)
                }
                P[i,j,K[j]] <- 1

                prob[i,j,1] <- P[i,j,1]
                for(k in 2:K[j]){
                        prob[i,j,k] <- P[i,j,k] - P[i,j,k-1]
                }
        }
}
```

```
# Data containing two different set of initial values
theta01 <- matrix(0, n, m)
alpha01 <- rep(0, p)
alpha001 <- rep(0, p)
kappa.star01 <- array(0, c(p,(max(K)-1)))
bugsData(list(alpha=alpha01, alpha0=alpha001, kappa.star=kappa.star01, theta=theta01),"Inits1.txt")

theta02 <- matrix(1, n, m)
alpha02 <- rep(1, p)
alpha002 <- rep(0, p)
kappa.star02 <- array(1, c(p,(max(K)-1)))
bugsData(list(alpha=alpha02, alpha0=alpha002, kappa.star=kappa.star02, theta=theta02),"Inits2.txt")

# Additive GRM model estimation by using OpenBUGS (for Q replications)

Q <- 10        # Number of replication for each scenario (distinct simulation conditions)
for(q in 1:Q){
    dataY_q <- paste(q,"-Y.txt",sep="")        # To add the replication number in files with results
    set.seed(q)                                 # Seed must changes at each iteration
                                                # to obtain different samples
    for(i in 1:n){
        for(j in 1:p){
            Y_q[i,j] <- sample(1:K[j], size = 1, prob = prob[i,j,])   # Generate a data matrix
        }                                                              # by using the stored prob
    }
    write.table(Y_q,quote=FALSE,sep=" ",dataY_q)    # Storing of matrix Y_q

    bugsData(list(Y=Y_q, n=n, p=p, K=K, m=m, vv=vv, m.theta=m.theta, Sigma.theta=Sigma.theta,
    m.alpha=m.alpha, s.alpha=s.alpha, m.alpha0=m.alpha0, s.alpha0=s.alpha0, m.kappa=m.kappa,
    s.kappa=s.kappa),"Data.txt")

    stats_q <- paste(q,"-stats.txt",sep="")     # To add the replication number in files with results
    dic_q <- paste(q,"-dic.txt",sep="")

    sim_q <- BRugsFit(modelFile = "ModelloGRMmultiuni.txt", data = "Data.txt", inits = c("Inits1.txt",
    "Inits2.txt"), numChains = 2, parametersToSave = c("theta", "alpha", "alpha0", "kappa"),
    nBurnin = 15000, nIter = 30000, nThin = 1, DIC = TRUE, working.directory = NULL, digits = 5)

    str(sim_q)
    write.table(sim_q$Stats,quote=FALSE,sep=", ",stats_q)       # Save results
    write.table(sim_q$DIC,quote=FALSE,sep=", ",dic_q)           # Save results
}
```

```
# Summary matrices with estimated parameters (posterior mean and sd).
# Each column is referred to a replication.

stat_1 <- read.table("1-stats.txt",header=TRUE)
mean <- matrix(stat_1[,1])
sd <- matrix(stat_1[,2])
for(q in 2:Q) {
	infile <- paste(q,"-stats.txt",sep="")
	stat_q <- read.table(infile,header=TRUE)
	mean <- cbind(mean,matrix(stat_q[,1]))
	sd <- cbind(sd,matrix(stat_q[,2]))
}

write.table(mean, 'SummaryMean.txt')
write.table(sd, 'SummarySd.txt')
```

Appendix C

Survey questionnaire

In this section we report the questionnaire submitted to residents in the Romagna area and in the State of San Marino (Italy). The questionnaire has been created to investigate residents' evaluations about costs and benefits of the tourism industry, a personal benefit from tourism, the quality of life in the area, the degree of involvement in the tourism industry, the residents' satisfaction with either their leisure or non-leisure domains, their quality of life, and the degree of support for future development of the tourism industry.

QUESTIONARIO RIMINESI

Intervistatrice num.___ Intervistato num.____

Buongiorno, questa è un'indagine coordinata da docenti dell'Università di Rimini per conoscere le opinioni dei cittadini sulla qualità della vita. Possiamo avere anche il suo parere? La disturberemo solo pochi minuti e le sue risposte rimarranno completamente anonime ... (*passare subito alla domanda successiva*).

1. Giudichi la Romagna rispetto a: / Con un voto da 1 a 7, come giudica questi aspetti della Romagna (1 min sodd, 7 max sodd):

1. tenore di vita ____
2. dotazione di servizi pubblici ____
2. traffico ____
3. pulizia della città e verde ____
4. ospitalità e accoglienza ____
5. possibilità di lavoro/carriera ____
6. attività ricreative e culturali ____
7. sicurezza ____

2. Dia un voto da 1 a 7 ai seguenti vantaggi che il turismo porta nella Romagna (1 min vantaggio, 7 max vantaggio) :

1. sviluppa l'economia della città ____
2. migliora lo standard/qualità di vita ____
3. sviluppa i servizi pubblici ____
4. aumenta le opportunità lavorative ____
5. migliora le attività culturali ____

3. Dia un voto da 1 a 7 ai seguenti problemi che il turismo porta nella Romagna (1 min problema, 7 max problema):

1. Aumenta il costo della vita e delle case ____
2. Aumenta il disordine e la criminalità ____
3. Danneggia l'ambiente e il paesaggio ____
4. Aumenta il traffico ____
5. Aumenta l'inquinamento ____

4. Con un voto da 1 a 7, quanto si ritiene soddisfatto dei seguenti aspetti della sua vita (ultimo anno):

1. Situazione economica ____
2. Salute ____
3. Relazioni famigliari ____
4. Relazioni con amici ____
5. Lavoro ____
6. spiritualità/religione ____

5. Con un voto da 1 a 7, quanto si ritiene soddisfatto dalle attività che svolge nel tempo libero (ultimo anno):

1. relazioni sociali ____
2. attività sportive/fitness ____
3. hobby personali ____
4. attività culturali (cinema, teatro, ecc.) ____
5. attività ricreative (ristoranti, discoteche, ecc.) ____
6. fare shopping ____
7. andare in spiaggia/mare ____

6. Con un voto da 1 a 7 (1 min sodd, 7 max sodd), quanto (ultimo anno):

1. è soddisfatto di come sono andate le cose nella sua vita ____
2. è soddisfatto della maggior parte degli aspetti della sua vita ____
3. trova soddisfazione nel pensare a quello che è riuscito a fare nella vita ____
4. è soddisfatto per quello che è quando si confronta con amici e familiari ____

7. Con un voto da 1 a 7 (1 min accordo, 7 max accordo), quanto è d'accordo con le seguenti affermazioni: l'industria turistica:

1. ha migliorato la qualità della mia vita ____
2. ha reso Rimini il posto migliore dove trascorrere il mio tempo libero ____
3. ha reso Rimini una città che mi consente di realizzarmi ____

8. Con un voto da 1 a 7 (1 min accordo, 7 max accordo), quanto è in accordo con le seguenti affermazioni:

1. sono a favore dello sviluppo del turismo balneare ____
2. sono a favore dello sviluppo delle attività culturali e ricreative della mia città ____
3. sono a favore dello sviluppo delle manifestazioni fieristiche e sportive ____

9. Complessivamente è a favore dello sviluppo dell'industria turistica nella Romagna: ☐si ☐no

10. Per migliorare la qualità della vita in Romagna cosa suggerisce di fare (1 sola proposta) ……………………………………………………………………………

11. La sua professione è in qualche modo legata al mercato turistico:

Sì, svolgo un'attività legata al settore turistico .... ☐si☐no
Saltuariamente o in passato ho svolto attività legate al settore turistico .... ☐si ☐no
L'attività dei miei familiari è legata al settore turistico .... ☐si ☐no

12. Lei...:

1. legge abitualmente quotidiani ....☐si ☐no
3. fa sport regolarmente ....☐si ☐no
4. va spesso a mostre d'arte, musei o teatro ....☐si ☐no
5. viaggia spesso per vacanza ....☐si ☐no
6. naviga spesso su Internet da casa ....☐si ☐no
7. acquista su internet ....☐si ☐no
8. fa volontariato e/o politica..............................☐si ☐no
9. va in Chiesa o altro luogo di culto religioso.......☐si ☐no

La ringrazio per la sua cortese collaborazione, per concludere posso ancora chiederle:

13. La sua età: ________

14. Genere

Maschio ....☐1
Femmina ....☐2

15. In quale città risiede? _______________________

16. Da quanti anni vive in Romagna? ________

---

17. Il suo stato civile

nubile/ celibe... ☐1
coniugato/a .... ☐2
separato/divorziato .... ☐3
vedovo/.. ☐4

18. Qual è il suo titolo di Studio:

Licenza Elementare.. ☐1
Licenza Media.. ☐2
Diploma.. ☐3
Laurea.. ☐4

19. E la sua professione? (1 sola risposta):

Dirigente / Funzionario / Professionista d'albo................................... ☐1
Imprenditore/ Lavoratore in proprio/Artigiano.................................... ☐2
Impiegato/a o quadro... ☐3
Insegnante (professore, maestro, ecc.) .... ☐4
Operaio/a... ☐5
Casalinga... ☐6
Studente/ssa.. ☐7
Pensionato/a.. ☐8
In cerca di lavoro... ☐9
Altro.. ☐10
Specificare_______________________________________