Longitudinal Multiple-Group IRT Modeling: Covariance pattern selection using MCMC and RJMCMC

Caio L. N. Azevedo

Department of Statistics,
State University of Campinas,
Campinas, SP, Brazil
Fax: +55 19 3521-5921. E-mail: cnaber@ime.unicamp.br

Jean-Paul Fox

Department of Research Methodology, Measurement and Data Analysis,
University of Twente,
Enschede, Netherlands

Dalton F. Andrade

Department of Informatic and Statistics, Federal University of Santa Catarina, Florianópolis, Brazil

Abstract: Longitudinal studies in Psychometric assessment are often focused on latent traits of subjects, who are clustered in different groups (e.g., gender, grade, social level). The corresponding type of data can be characterized as longitudinal multiple-group item responses. For this type of data, a longitudinal multiple-group IRT (LMGIRT) model is proposed, where group-specific dependencies between latent traits can be modeled using the appropriate covariance structure. The multiple-group specification together with the developed MCMC-based algorithms make it possible to handle the scaling process simultaneously with the estimation of latent traits, item and population parameters. Also, a reversible-jump MCMC (RJMCMC) algorithm is proposed for joint parameter estimation and covariance matrix selection for a single-group longitudinal IRT model. The MCMC-based algorithms can handle identification rules, scaling issues and selection between restricted covariance structures. Simulation studies reveal that not only all parameters are accurately recovered, but also the correct underlying covariance pattern model is selected. Two real data sets are used to illustrate the longitudinal IRT models and the MCMC algorithms for estimation and model fit assessment. One study concerns the health condition of Dutch students from AGHLS (Amsterdam Growth and Health Longitudinal Study) and the other study a longitudinal research program of the Brazilian federal government.

Keywords: Longitudinal data, multiple group, item response theory, Bayesian inference, MCMC algorithm, RJMCMC algorithm

1 Introduction

Longitudinal studies in Psychometric assessment are often focused on measuring traits of subjects who are clustered in different groups (e.g., gender, grade, social level). The corresponding type of response data can be characterized as longitudinal multiple group item responses. For analyzing multiple group response data the multiple group model (MGM) proposed by Bock and Zimowski (1997) offers a unified approach to handle different aspects of interest. A fully Bayesian framework, including parameter estimation and model fit assessment of the MGM, was recently developed by Azevedo et al. (2012). Furthermore, Azevedo et al. (2014) developed a longitudinal (single-group) IRT (LIRT) model, which extends in several directions the longitudinal IRT models of Andrade and Tavares (2005) and Tavares and Andrade (2006), for a homogeneous group of respondents.

A longitudinal multiple-group IRT (LMGIRT) model is proposed, to model longitudinal item response data of respondents who are clustered in higher level units. The LMGIRT model and MCMC estimation method make it possible to handle the scaling process simultaneously with the estimation of latent traits, item and population parameters, allowing different dependency structures for the different groups of respondents. The modeling framework was motivated by combining the modeling approaches of Azevedo et al. (2014) and Azevedo et al. (2012). The LMGIRT model ensures that student performances across groups are measured on the same scale, while the dependencies in each group are modeled using the appropriate covariance structure. The entire estimation process can be carried out by an efficient MCMC algorithm.

For the LIRT model, the selection of the appropriate covariance structure would require fitting several competing models with different covariance structures. Subsequently, a model-comparison method is required to identify the optimal higher-level covariance structure among the competing models. This complex two-stage procedure can be optimized by including the selection of the optimal covariance pattern model in the MCMC algorithm. Therefore, a reversible-jump MCMC (RJMCMC) algorithm has been developed to estimate simultaneously all LIRT model parameters including the selection of the covariance matrix.

For both models, the LIRT and the LMGIRT models, simulation studies show that with the developed MCMC-based algorithms all parameters can be accurately recovered and that the correct underlying covariance model can be identified with the RJMCMC algorithm. In addition, the studies show that the LMGIRT modeling approach avoids the use of the LIRT model with a posterior equating approach to relate measurements of subjects from different groups to each other.

This paper is outlined as follows. After introducing the Bayesian LMGIRT model, the MCMC algorithms are presented, which can handle group-specific covariance structures for the latent variable distribution. Then, a reversible-jump MCMC algorithm is presented, which handles the selection of the appropriate covariance structure, together with the estimation of the model parameters. Then, three simulation studies are given to show the performance of the MCMC and RJMCMC algorithm for the LMGIRT and LIRT models, respectively, and to show the comparison of the LMGIRT model with the LIRT model combined with a posterior equating approach. Subsequently, data from the AGHLS (Amsterdam Growth and Health Longitudinal Study) and a Brazilian longitudinal research study are analyzed. In the last section, the results and some model extensions are discussed.

2 The Model

One or more tests are administered to subjects clustered into different groups, which are followed along several time-points. The subjects are randomly selected at the first time-point from their respective groups. Subjects are measured at different time-points, which can be related to different grades or different months in a grade year, for example. At each measurement occasion, the groups of subjects can, for example, represent different grades, different genders and different geographic regions. Dropouts or inclusion of subjects during the study are allowed (even though we will treat only the complete data case). At each time-point *t*, $t = 1, ..., T$, a test of $\mathcal{I}_{kt}$ items, from a total of $I \leq \sum_{k=1}^{K} \sum_{t=1}^{T} \mathcal{I}_{kt}$ items, is administered to each group k $(k = 1, ..., K)$ of n_{kt} subjects. Here, a complete data case is assumed, $n_{kt} = n_k, \forall k$. Across measurement occasions common items are used, which defines an incomplete test design, see Montgomery (2004). For example, for three tests, each of 20 items, a configuration of an incomplete test design would be where test one and two have five common items, and test two and three six common items.

Let θ_{jkt} denote the latent trait of subject j of group k at time-point t, by $\boldsymbol{\zeta}_i$ the vector of parameters of item *i*, by $\boldsymbol{\eta}_{\boldsymbol{\theta}_k}$ the vector with the population parameters of group k (i.e., the population means, variances and correlation parameters), and by Y_{ijkt} the response to item *i* of examinee *j* of group *k* at time-point *t*. Then, the general model for longitudinal multiple group item response data is given by:

$$
\begin{aligned}
Y_{ijkt} \mid (\theta_{jkt}, \boldsymbol{\zeta}_i) &\sim \text{Bernoulli}(P_{ijkt}) \\
P_{ijkt} = P(Y_{ijkt} = 1 \mid \theta_{jkt}, \boldsymbol{\zeta}_i) &= \Phi(a_i \theta_{jkt} - b_i) \qquad (1) \\
\boldsymbol{\theta}_{jk.} | \boldsymbol{\eta}_{\boldsymbol{\theta}_k} &\sim N_T(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}), \qquad (2)
\end{aligned}
$$

where $\boldsymbol{\theta}_{jk.} = (\theta_{jk1}, ..., \theta_{jkT})^t$, $\Phi(.)$ stands for a cdf of the standard normal distribution, $\boldsymbol{\zeta}_i = (a_i, b_i)^t$ and $N_T(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}), k = 1, 2, ..., K$, denotes a T-dimensional normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{\theta}_k}$ and covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}$, where

$$
\boldsymbol{\mu}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \mu_{\theta_{k1}} \\ \mu_{\theta_{k2}} \\ \vdots \\ \mu_{\theta_{kT}} \end{bmatrix} , \boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \psi_{\theta_{k12}} & \dots & \psi_{\theta_{k1T}} \\ \psi_{\theta_{k12}} & \psi_{\theta_{k2}} & \dots & \psi_{\theta_{k2T}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{k1T}} & \psi_{\theta_{k2T}} & \dots & \psi_{\theta_{kT}} \end{bmatrix} , \qquad (3)
$$

$\boldsymbol{\eta}_{\boldsymbol{\theta}_k} = (\boldsymbol{\mu}_{\boldsymbol{\theta}_k}^t, v(\boldsymbol{\Psi}_{\boldsymbol{\theta}_k})^t)^t$, $v(\Psi_{\boldsymbol{\theta}_k}) = \text{vecd}(\Psi_{\boldsymbol{\theta}_k})$ and *vecd(.)* stands for the different elements of $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}$. Note that a multivariate normal distribution with a flexible variance-covariance structure is considered to model the within-subject latent trait dependencies. A total of $K\frac{T(T+1)}{2}$ parameters need to be estimated for the unstructured covariance model in each group.

An important assumption is the measurement invariance of the items. Measurement invariant items will function in the same way over groups and occasions (time-points), see Millsap (2010). The LMGIRT model will assume that the common items are measurement invariant. This assumption, within the IRT framework, is related to differential item functioning (DIF) and item parameter drift, see Bock and Zimowski (1997). Here, it will be assumed that the assumption of measurement invariance holds for the common items. This

aspect certainly deserves more investigation, but is beyond the scope of the present paper. The LMGIRT model extends the multiple-group model of Bock and Zimowski (1997), and it offers, in the context of longitudinal response data, a unified approach to handle different aspects of interest (differential item functioning, group specific latent trait distribution, item parameter drift, nonequivalent groups equating among others).

2.1 Unstructured covariance matrix with identification restrictions

To identify the latent scale, a reference group is required. Therefore, the mean and variance of the latent variable distribution at the first time-point of the reference group (i.e., the first group measured at the first occasion) will be fixed to zero and one, respectively. Furthermore, an incomplete test design is used such that common items are administered to different groups at different measurement occasions. The incomplete test design defines a common latent scale across measurement occasions and groups.

The restrictions on the parameters of the latent trait distribution complicates the specification of priors. In the proposed latent variable framework, the prior modeling approach developed by Azevedo et al. (2014), based on the work of McCulloch et al. (2000), to account for a restricted covariance structure, will be extended in order to accommodate multiple groups. Besides the identifiability restrictions for the reference group, the reparameterization presented in the Appendix 7.2, which also includes the details of this approach, will be useful to handle restricted covariance matrices for all groups.

2.2 Restricted Covariance Pattern Structures

By correctly modeling the correlation among the latent traits across measurement occasions, more accurate statistical inferences can be made. A time-heteroscedastic covariance structure can be considered to describe more complex patterns over time, where population variances of measurements can differ over time-points. A more parsimonious modeling of the group-specific covariance structures will decrease the number of parameters to be estimated, which can be estimated more accurately, especially in the presence of the multiple groups. In the simulation and real data studies, four of the most important covariance structures used in longitudinal data modeling are considered: the heteroscedastic uniform model (HU), the heteroscedastic toeplitz model (HT), the first-order heteroscedastic autoregressive model (ARH), and the first-order heteroscedastic autoregressive moving-average model (ARMAH). The structures of the matrices ARH, ARMAH and UH are presented in Appendix 7.1. A brief overview of other covariance structures can be found in Azevedo et al. (2014) and Tavares and Andrade (2006).

3 Bayesian inference and MCMC estimation

First, an MCMC algorithm for the LMGIRT (longitudinal multiple group) model is presented, where the covariance matrix for each group is defined in advance. Second, an RJMCMC algorithm for selecting the covariance matrix of the LIRT model is presented. More technical details are given in the Appendix (subsections 7.2, 7.3 and 7.4).

3.1 LMGIRT model: MCMC Algorithm

A full Gibbs sampling (FGS) algorithm is developed to estimate simultaneously all parameters. The MCMC algorithm can be used to obtain values from the marginal posteriors of the model parameters. Subsequently, inferences can be made from the sampled parameter values, see Gamerman and Lopes (2006).

This approach is an extension of the algorithm discussed by Azevedo et al. (2014) who considered one group of respondents. A review on MCMC methods for longitudinal and multivariate probit models is given by Azevedo et al. (2014). The idea is to define sets of augmented data ($\boldsymbol{Z}$), following the data augmentation procedure of Albert (1992) and Azevedo et al. (2014), and to define indicator variables ($\boldsymbol{V}$) and ($\boldsymbol{I}$) which handle the not-selective missing responses due to uncontrolled events and the incomplete block design, respectively. For a more thorough discussion the reader is referred to Azevedo et al. (2014) and for technical details see subsections 7.2 and 7.3. Here, the sequence of steps related to the MCMC algorithm are given below. Let (.) denote the set of all necessary parameters. Then, the full Gibbs sampling algorithm is defined as follows:

1. Start the algorithm by choosing suitable initial values.

   Repeat steps 2–10.

2. Simulate Z_{ijkt} from $Z_{ijkt} \mid (.), t = 1, .., T, i = 1, ..., \mathcal{I}_{kt}$, $j = 1, ..., n_k$, where $\mathcal{I}_{kt}$ is the set of items that compose the test applied to group k at the time-point t.

3. Simulate $\boldsymbol{\theta}_{jk.}$ from $\boldsymbol{\theta}_{jk.} \mid (.), j = 1, ..., n_k, k = 1, ..., K$.

4. Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i \mid (.)$, i =1,...,I.

5. Simulate $\boldsymbol{\mu}_{\boldsymbol{\theta}_k}$ from $\boldsymbol{\mu}_{\boldsymbol{\theta}_k}|(.), k = 1, ..., K$.

6. Simulate $\psi_{\theta_{k1}}$ from $\psi_{\theta_{k1}} \mid (.), k = 1, ..., K$.

7. Simulate $\boldsymbol{\psi}_k^*$ from $\boldsymbol{\psi}_k^* \mid (.), k = 1, ..., K$.

8. Simulate $\boldsymbol{\Psi}_k^*$ from $\boldsymbol{\Psi}_k^* \mid (.), k = 1, ..., K$.

9. Compute, for each group, the unstructured covariance matrix using the sampled covariance components from Steps 6-8 and Equations (8), (9) and (13)

10. Through a parameter transformation method using sampled unstructured covariance parameters, compute restricted covariance components of interest, for each group. The sampled restricted covariance structures $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}, k = 1, ..., K$ are used when repeating steps 2–8.

Appendix 7.2 shows how to handle the restrictions $\mu_{\theta_{11}} = 0$ and $\psi_{\theta_{11}} = 1$. Specifically, the expression in Equation (10) is used to simulate $\boldsymbol{\mu}_{\boldsymbol{\theta}_{1(1)}}$. Therefore, to simulate $(\mu_{\theta_{11}}, \psi_{\theta_{11}})^t$, the following decomposition is used (In Appendix 7.3, Equation (25)),

$$p(\boldsymbol{\theta}_{j1.}|\boldsymbol{\eta}_{\boldsymbol{\theta}_1}) = p(\boldsymbol{\theta}_{j1(1)}|\boldsymbol{\eta}_{\boldsymbol{\theta}_1}, \theta_{j11})p(\theta_{j11}|\boldsymbol{\eta}_{\boldsymbol{\theta}_{11}}),$$

where $\boldsymbol{\eta}_{\boldsymbol{\theta}_{11}} = (\mu_{\theta_{11}}, \psi_{\theta_{11}})^t$. To identify the model, the scale of the latent variable for the reference group (in this case, the first time-point of the the first group) is transformed to

mean zero and variance one. It is also possible to restrict the parameters $(\mu_{\theta_{11}}, \psi_{\theta_{11}})^t$ to other values.

In Step 9, MCMC samples of $\mathbf{\Psi}_k^*$ are drawn from an inverse-Wishart distribution, and each sampled covariance matrix is restricted to be positive definite. When considering the following relationship,

$$det(\mathbf{\Psi}_{\boldsymbol{\theta}_k}) = det(\psi_{\theta_{k1}}) det\left(\mathbf{\Psi}_{\boldsymbol{\theta}_{k(1)}} - \psi_{\theta_{k1}}^{-1} \boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}} \boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}}^{t}\right) = \psi_{\theta_{k1}} det(\mathbf{\Psi}_k^*),$$

$k = 1, ..., K$, using Equations (8) and (11), it can be seen that positive definite samples of $\mathbf{\Psi}_{\boldsymbol{\theta}_k}$ are obtained. When using a property of the determinant of block matrices, it follows that the determinant of $\mathbf{\Psi}_{\boldsymbol{\theta}_k}$ is greater than zero, since both the determinant of $\mathbf{\Psi}_k^*$ and $\psi_{\theta_{k1}}$ are greater than zero.

In each MCMC iteration, parameters of a specific covariance pattern (for each group) are computed using sampled unstructured covariance parameters. Each covariance pattern is nested in the most general unstructured pattern. Therefore, in MCMC Step 10, parameters of a specific covariance structure are computed. Each simulated covariance matrix will be positive definite, since it is based on a positive definite unstructured covariance matrix. This whole process is carried out for each group.

3.2 RJMCMC for covariance matrix selection

One of the most important issues concerning longitudinal data analysis is the choice of the most appropriate covariance matrix. For IRT models, the MCMC algorithms require a large amount of computational time. Therefore, it is important to consider an efficient mechanism of covariance structure selection. It is computationally cumbersome to estimate different models, considering different types of covariance matrices for each group, and choose the best one by using information criteria such as the BIC and DIC, see Azevedo et al. (2014). Also, as pointed out by these authors, these statistics did not provide good results. Instead, it would be preferable to select the most appropriate covariance structures while estimating the model parameters. The RJMCMC algorithm allows selecting optimal covariance structures, while sampling the model parameters from the marginal posterior distributions, see Green (1995).

To ease the notation, we will consider two possible covariance matrices, the ARH and the ARMAH and one group of respondents. In fact, we have implemented only these two covariance matrices, for the RJMCMC algorithm. However, this procedure can be extended to more covariance pattern models. The idea is to include an additional step in the MCMC algorithm presented in subsection 3.1, which performs the covariance pattern selection. Let (.) denote the set of all necessary parameters. Then, the RJMCMC algorithm is defined as follows:

1. Start the algorithm by choosing suitable initial values.

   Repeat steps 2–11.

2. Simulate Z_{ijt} from $Z_{ijt} \mid (.), t = 1, .., T, i = 1, ..., \mathcal{I}_t, j = 1, ..., n$, where $\mathcal{I}_t$ is the set of items that compose the test applied at the time-point t.

3. Simulate $\boldsymbol{\theta}_{j.}$ from $\boldsymbol{\theta}_{j.} \mid (.), j = 1, ..., n$.

4. Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i \mid (.)$, i =1,...,I.
5. Simulate $\boldsymbol{\mu_\theta}$ from $\boldsymbol{\mu_\theta}|(.)$.
6. Simulate ψ_{θ_1} from $\psi_{\theta_1} \mid (.)$.
7. Simulate $\boldsymbol{\psi}^*$ from $\boldsymbol{\psi}^* \mid (.)$.
8. Simulate $\boldsymbol{\Psi}^*$ from $\boldsymbol{\Psi}^* \mid (.)$.
9. Compute the unstructured covariance matrix using the sampled covariance components from Steps 6-8 and Equations (8), (9) and (13) (considering only one group).
10. Through a parameter transformation method using sampled unstructured covariance parameters, compute the current restricted covariance matrix (selected in the previous MCMC iteration.)
11. Select the covariance matrix using the steps presented in subsection 7.4.

4 Simulation studies

The parameter recovery and model selection procedure were analyzed using simulated data. In the first study, we explore a longitudinal multiple group structure with predefined covariance matrices for each group. In the second study, we consider a single group longitudinal study where the covariance matrix will be selected using the RJMCMC algorithm. In the third study, we compare the LMGIRT model with the LIRT model combined with a posterior equating technique.

The usual tools (traceplots, Geweke and Gelman-Rubin statistics) for monitoring MCMC convergence are used. A burn-in of 16,000 iterations are considered, a thin of 30 iterations and a total of 46,000 iterations are simulated, which produce a valid sample of size 1,000. For RJMCMC, the total number of iterations was set at 47,000 in order to obtain a valid sample of approximately 1,000 iterations. Different statistics were used to compare the results: mean of the estimates (M. Est.), correlation (Corr), the absolute bias (ABias), variance (Var), and the root mean squared error (RMSE). Let ϑ_l be an element of $(\theta_{jkt}, a_i, b_i, \eta_{\theta_{kt}})^t$, where l is a convenient index or a combination of them (*i*, *t*, *jt*, *kt*, *jkt*) and $\widehat{\vartheta}_{lr}$ its respective estimate obtained in the replica *r*, $r = 1, ..., R$. Define also $\widehat{\overline{\vartheta}}_l = \frac{1}{R}\sum_{r=1}^{R}\widehat{\vartheta}_{lr}$. The aforementioned statistics are, Corr: correlation between $\widehat{\overline{\vartheta}}_l$ and ϑ_l, Bias: $\left(\widehat{\overline{\vartheta}}_l - \vartheta_l\right)$, Var: $\frac{1}{R-1}\sum_{r=1}^{R}\left(\widehat{\vartheta}_{lr} - \widehat{\overline{\vartheta}}_l\right)^2$, RMSE: $\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\widehat{\vartheta}_{lr} - \vartheta_l\right)^2}$. To evaluate the accuracy of the MCMC estimates, a total of ten replicated data sets were generated, based on Azevedo and Andrade (2010) and De Ayala and Bolesta (1999). For the item and latent trait parameters, average statistics were computed by averaging across data sets, items and persons.

4.1 Parameter recovery of the longitudinal multiple-group study

In this study, two groups were assessed at three occasions. For the first group (for which the first time-point corresponds to the reference group), the latent means were $\boldsymbol{\mu}_{\boldsymbol{\theta}_1} = (0.0, 1.0, 2.0)^t$, which imply growth in mean latent averages. Furthermore, a heteroscedastic

Toeplitz matrix with $\boldsymbol{\psi}_{\boldsymbol{\theta}_1} = (1.00, 0.90, 0.95)^t$ was assumed, which imply a decrease and then an increase in the variability, with $\rho_{\theta_1} = 0.6$, which implies a moderate magnitude for the between-time correlations. For the second group, the latent means were set to $\boldsymbol{\mu}_{\boldsymbol{\theta}_2} = (0.2, 1.3, 2.5)^t$ and an ARMAH matrix was assumed with $\boldsymbol{\psi}_{\boldsymbol{\theta}_2} = (0.90, 0.80, 0.85)^t$, with $\rho_\theta = 0.80$ and $\gamma_\theta = 0.88$. These values induce similar behavior for the second group compared with the first group. Furthermore, $n_k = 1,000; k = 1, 2$, latent traits were simulated according to appropriate three-variate normal distributions. Following DeMars (2003), the sampled latent traits were transformed to the scale of the simulated latent traits using,

$$\boldsymbol{\theta}^{**}_{jk.} = Chol\left(\boldsymbol{\Psi}_{\theta_k}\right) Chol\left(\boldsymbol{S}_{\boldsymbol{\theta}_k}\right)^{-1} \left(\boldsymbol{\theta}^{*}_{jk.} - \overline{\boldsymbol{\theta}}_k\right) + \boldsymbol{\mu}_{\boldsymbol{\theta}_k},$$

where $\boldsymbol{\theta}^{*}_{jk.}$ are the simulated latent traits, $\overline{\boldsymbol{\theta}}_k$ and $\boldsymbol{S}_{\boldsymbol{\theta}_k}$ the sampled mean vector and covariance matrix of group k, respectively, and *Chol* represents the Cholesky decomposition. Finally, Table 1 represents the values of the item parameters and the structure of the adopted incomplete test design. For example, item 1 appears in tests 1 and 4, and so forth. Tests 1, 2 and 3 were administered to the first group at time-points (measurement occasions) 1, 2 and 3, respectively. Similarly, tests 4, 5 and 6 were administered to the second group at measurement occasions 1, 2 and 3, respectively.

Table 1 Test structures (linked test design) and item parameter values for the multiple group longitudinal study

Item	Test(s)	a	b	Item	Tests	a	b	Item	Test(s)	a	b
1	1-4	0.8	-2.0	37	2-3	1.2	1.3	73	4-5	1.4	0.0
2	1	1.1	-1.8	38	2-3	1.3	1.4	74	4-5	0.8	0.2
3	1	1.4	-1.6	39	2-3	1.4	1.6	75	4-5	1.1	0.4
4	1	0.8	-1.4	40	2-3	0.8	1.8	76	4-5	1.4	0.6
5	1	1.1	-1.2	41	2-3	1.1	2.0	77	4-5	0.9	0.8
6	1-4	1.4	-1.0	42	2-3	1.4	2.1	78	4-5	1.2	1.0
7	1	0.8	-0.8	43	3-6	1.3	0.0	79	5	0.9	-0.6
8	1	1.1	-0.6	44	3	1.0	0.2	80	5	1.3	2.2
9	1	1.4	-0.4	45	3	1.1	0.4	81	5	1.0	2.4
10	1-4	0.8	-0.1	46	3	1.2	0.6	82	5	1.3	0.8
11	1-4	0.9	1.0	47	3	1.3	0.8	83	5	1.0	1.2
12	1	1.2	1.2	48	3-6	1.4	1.0	84	5	1.1	1.1
13	1	1.4	1.4	49	3-6	0.9	1.2	85	5-6	1.2	1.3
14	1	1.0	1.6	50	3	1.0	1.4	86	5-6	1.3	1.4
15	1	1.1	1.8	51	3-6	1.1	1.6	87	5-6	1.4	1.6
16	1-4	1.4	2.0	52	3	1.2	1.8	88	5-6	0.8	1.8
17	1	0.9	0.8	53	3	0.8	3.4	89	5-6	1.1	2.0
18	1-4	1.1	0.0	54	3	1.1	2.2	90	5-6	1.4	2.1
19	1-2	1.4	0.0	55	3-6	1.4	2.1	91	6	1.0	0.2
20	1-2	0.8	0.2	56	3	0.9	2.3	92	6	1.1	0.4
21	1-2	1.1	0.4	57	3	1.2	3.0	93	6	1.2	0.6
22	1-2	1.4	0.6	58	3	1.3	2.6	94	6	1.3	0.8
23	1-2	0.9	0.8	59	3	1.1	2.8	95	6	1.0	1.4
24	1-2	1.2	1.0	60	3-6	1.4	3.8	96	6	1.2	1.8
25	2-5	1.3	-0.8	61	4	1.1	-1.8	97	6	0.8	3.4
26	2	0.9	-0.6	62	4	1.4	-1.6	98	6	1.1	2.2
27	2-5	1.2	-0.4	63	4	0.8	-1.4	99	6	0.9	2.3
28	2-5	1.3	-0.2	64	4	1.1	-1.2	100	6	1.2	3.0
29	2-5	1.0	0.0	65	4	0.8	-0.8	101	6	1.3	2.6
30	2-5	1.2	2.0	66	4	1.1	-0.6	102	6	1.1	2.8
31	2	1.3	2.2	67	4	1.4	-0.4	-	-	-	-
32	2	1.0	2.4	68	4	1.2	1.2	-	-	-	-
33	2	1.3	0.8	69	4	1.4	1.4	-	-	-	-
34	2-5	1.2	1.0	70	4	1.0	1.6	-	-	-	-
35	2	1.0	1.2	71	4	1.1	1.8	-	-	-	-
36	2	1.1	1.1	72	4	0.9	0.8	-	-	-	-

The following hyperparameter settings were used in the simulation study (see equations (15)-(18)):

$$\boldsymbol{\Psi}_k = \tau_k \boldsymbol{I}_{T-1} \tag{4}$$

$$\boldsymbol{\Psi}_{\boldsymbol{\Psi}_k} = (\nu_{\boldsymbol{\Psi}_k} - T + 1)\,(\boldsymbol{I}_{T-1} - \boldsymbol{\Psi}_k)\,, \tag{5}$$

where $\boldsymbol{I}_{T-1}$ stands for an identity matrix of order $T-1$, $\nu_{\boldsymbol{\Psi}_k} = 5$, $\tau_k = 1/8, k = 1, 2$, and the hyperparameters for the item parameters were specified as: $\boldsymbol{\mu}_{\boldsymbol{\zeta}} = (1, 0)^\top$, $\boldsymbol{\Psi}_{\boldsymbol{\zeta}} = \text{diag}(0, 5, 3)$, $\mu_{0k} = (0, 0, 0)^t$, $\boldsymbol{\Psi}_{0k} = \text{diag}(2, 2, 2)$, and $p(\psi_{\theta_{k1}}) \propto \mathbb{1}_{(0,\infty)}(\psi_{\theta_{k1}})$, $p(\boldsymbol{\psi}_k^*) \propto \mathbb{1}_{I\!R^{T-1}}(\boldsymbol{\psi}_k^*)$, for $k = 1, 2$.

The results presented in Table 2 indicate that the item parameters and the latent traits, for all groups and time-points, were properly recovered. Similar conclusions can be drawn about the estimates of the latent trait population parameters, see Table 3.

Table 2 Longitudinal multiple-group study: results for the estimated latent traits and item parameters.

Parameter	**Statistic**			
	Corr	**Abias**	**Var**	**RMSE**
latent trait	.994	.121	.057	.274
discrimination parameter	.985	.036	.010	.105
difficulty parameter	$>$.999	.029	.016	.114

Table 3 Longitudinal multiple-group study: results for the estimated latent trait population parameters.

Group	**Param.**	**True value**	**M.est.**	**Abias**	**Var**	**RMSE**
1	$\mu_{\theta_{12}}$	1.00	.974	.026	.001	.043
	$\mu_{\theta_{13}}$	2.00	1.939	.061	.009	.112
	$\psi_{\theta_{12}}$	.90	.819	.081	.009	.123
	$\psi_{\theta_{13}}$	.95	.916	.034	.017	.136
	$\rho_{\theta_{11}}$	.60	.608	.008	$<$.001	.019
2	$\mu_{\theta_{21}}$	.20	.193	.007	.001	.026
	$\mu_{\theta_{22}}$	1.30	1.253	.047	.002	.066
	$\mu_{\theta_{23}}$	2.50	2.424	.076	.012	.134
	$\psi_{\theta_{21}}$	.90	.891	.009	.004	.065
	$\psi_{\theta_{22}}$	.80	.768	.032	.008	.097
	$\psi_{\theta_{23}}$	.85	.820	.030	.019	.142
	$\rho_{\theta_{21}}$	.80	.802	.002	$<$.001	.020
	$\gamma_{\theta_{21}}$	.88	.872	.008	$<$.001	.016

4.2 Parameter recovery and covariance selection for the longitudinal single group study

A second simulation study is presented concerning the performance of the RJMCMC algorithm in recovering the parameters and matrix covariance selection of a longitudinal

(single group) study. The set-up of this study corresponds to the one in subsection 4.1, but attention was restricted to the first group measured at three time-points. With respect to the underlying covariance structure, three different scenarios were studied;

1. ARH covariance matrix with $\rho_{\theta(ARH)} = 0.80$
2. ARMAH covariance matrix with $\rho_{\theta(ARMAH)} = 0.80$ and $\gamma_\theta = 0.80$.
3. ARMAH covariance matrix with $\rho_{\theta(ARMAH)} = 0.80$ and $\gamma_\theta = 0.88$.

This setting of the covariance structure ensure that all within-subject latent trait correlations are high. Note that the second scenario is equivalent to an ARH matrix with $\rho_{\theta(ARH)} = \rho_{\theta(ARMAH)}\gamma_\theta$, since the ARH matrix is a particular case of the ARMAH structure, when $\rho_{\theta(ARMAH)} = \gamma_\theta$. The structure of the incomplete test design and the values of the item parameters can be found in Table 1, considering the tests 1, 2 and 3. There were six common items between test 1 and 2, and six between test 2 and 3. The item parameter values varied in terms of discrimination and difficulty. Also, we have used the same specification of hyperparameters, as presented in subsection 4.1, besides $\mu_\rho = \mu_\gamma = 0$, $\psi_\rho = \psi_\gamma = 1$ (see equations (28)-(30)).

Table 4 presents the average proportion of times that the true model was selected over ten replications. It can be seen that the RJMCMC algorithm always chose the true model (i.e., the true model was visited at least 50% of the iterations), leading to high values for the averaged proportions (higher than 73,9%). It shows that the RJMCMC algorithm was able to select the true model.

Table 5 shows the results for the latent traits and item parameters estimates. Table 6 presents the results for the latent trait population parameter estimates. The item parameters, the latent traits and the population parameters, for all groups, time-points and scenarios, were properly recovered. The results were less accurate than those in the previous subsection. This was expected since, in this case, the true covariance matrix was unknown and selected concurrently with the estimation of the parameters. Notice that in scenario 3, where the within-subject latent trait correlations are higher compared to the others, the higher within-subject latent trait correlations led to more accurate results.

Table 4 Longitudinal single group study: averaged proportion of visits for each model

Scenario	**True model**	**Model**	
		ARH	**ARMAH**
1	ARH	.977	.023
2	ARH	.980	.020
3	ARMAH	.261	.739

4.3 Implicit scaling (LMGIRT model) compared to posterior equating (LIRT model)

In the multiple-group model (MGM) of Bock and Zimowski (1997), subjects of different groups are assessed using tests with a linked design, where common (anchor) items are used to link the scales of the different groups. In a joint estimation procedure, the estimates of the latent traits are measured directly on the same scale using the linked design property.

Table 5 Longitudinal single group study: results for the estimated latent traits and item parameters.

Parameter	Scenario	Statistic			
		Corr	**Abias**	**Var**	**RMSE**
Latent trait	1	.994	.133	.052	.274
	2	.994	.113	.058	.271
	3	.982	.028	.009	.094
Discrimination parameter	1	.993	.135	.055	.280
	2	.994	.115	.061	.278
	3	.999	.057	.019	.135
Difficulty parameter	1	.967	.051	.009	.105
	2	.979	.030	.010	.099
	3	.999	.044	.016	.120

Table 6 Longitudinal single group study: results for the estimated latent trait population parameters.

Scenario	Time-point	True value	M. Est.	Abias	Var	RMSE
			μ_θ			
1	2	1.00	.956	.044	.001	.053
	3	2.00	1.899	.101	.003	.114
2	2	1.00	.974	.026	.001	.035
	3	2.00	1.952	.048	.007	.095
3	2	1.00	.971	.029	.002	.048
	3	2.00	1.954	.046	.006	.089
			ψ_θ			
1	2	.90	.825	.075	.006	.105
	3	.95	.850	.100	.004	.119
2	2	.90	.881	.019	.003	.061
	3	.95	.973	.023	.013	.118
3	2	.90	.879	.021	.003	.061
	3	.95	.937	.013	.012	.109
			ρ_θ			
1		.80	.804	.004	$< .001$	.006
2		.80	.794	.006	$< .001$	.009
3		.80	.801	.001	$< .001$	.015
			γ_θ			
3		.88	.874	.006	$< .001$	.080

Furthermore, more accurate estimation results can be obtained using the pooled information from both groups compared to a single-group estimation and a posterior test equating approach to get all parameter estimates on the same scale, see Kolen and Brennan (2004).

Here, subjects from different groups were assessed at different occasions. Each group at each occasion was considered to be a single group. The latent trait estimates under the LMGIRT model, using the linked design property, were all estimated on the same scale. In the single-group approach, the LIRT model was used for each group, see Azevedo et al. (2014), and all parameters were estimated for each group at each occasion. Then, through

the common item structure, the latent trait and item parameter estimates were transformed to the same scale using the Mean-Sigma procedure as the equating method, see Kolen and Brennan (2004).

The results of both procedures are compared to each other. Let Method 1 refer to a joint estimation of all parameters on a common scale under the LMGIRT model, and Method 2 to the two-step approach, where first the traits and item parameters are estimated under the LIRT model, and second common scale estimates are obtained using the Mean-Sigma procedure. Data were generated according to the setup described in subsection 4.1. The sum of the absolute bias (Abias), and the relative absolute bias (Rabias) were computed. The results are given in Table 7 and Figure 1. It can be concluded that the gain in accuracy is substantial, mainly in the estimation of the traits, when using the pooled information over groups. The split-up pattern, observed in the upper-right figure (Figure 1) is, probably, due to items, with no high item difficulties (see Table 1), presented to the subjects with the highest latent trait values.

Table 7 Sum of the absolute bias and the absolute relative bias for the parameter estimates for the two methods of scaling process (equating): Method 1 - joint estimation of all parameters on a common scale under the LMGIRT model; Method 2 - first the parameters are estimated under the LIRT model, and second common scale estimates are obtained using the Mean-Sigma procedure.

Parameter	**Abias**		**Rabias**	
	Method 1	**Method 2**	**Method 1**	**Method 2**
discrimination parameter	6.62	12.10	5.97	10.85
difficulty parameter	5.50	13.38	42.14	75.66
latent trait	1289.87	3982.28	5605.28	10695.73

5 Real data analysis

5.1 Amsterdam Growth and Health Longitudinal Study

Data were analyzed from the AGHLS (Amsterdam Growth and Health Longitudinal Study), which is a multidisciplinary longitudinal cohort study, originally set up to examine growth and health among teenagers (Kemper et al., 1978). The AGHLS is focused on research questions related to relationships between anthropometry (Hoekstra et al., 2011), physical activity (Douw et al., 2014), cardiovascular disease risk (Wijnstok et al., 2012, 2013), lifestyle (Twisk et al., 1997, 1998), musculoskeletal health, psychological health (Hoekstra et al., 2013) and wellbeing. The presented sample consists of 443 participants who were followed over the period 1993-2006 with a maximum of three measurement points for each individual. A subscale of the STAI-DY questionnaire was used to measure the latent variable "state anxiety", using a total of thirteen items.

Data from three years were used in the analysis; 1993, 2000 and 2006, referred to, respectively, as years 1, 2 and 3. Two groups were considered, male students (group 1) and female students (group 2). Therefore, two groups were assessed at three occasions (similar to the design of the first simulation study). A total of 59 male students and 72 female students were assessed at each measurement occasion, providing 393 responses per item.

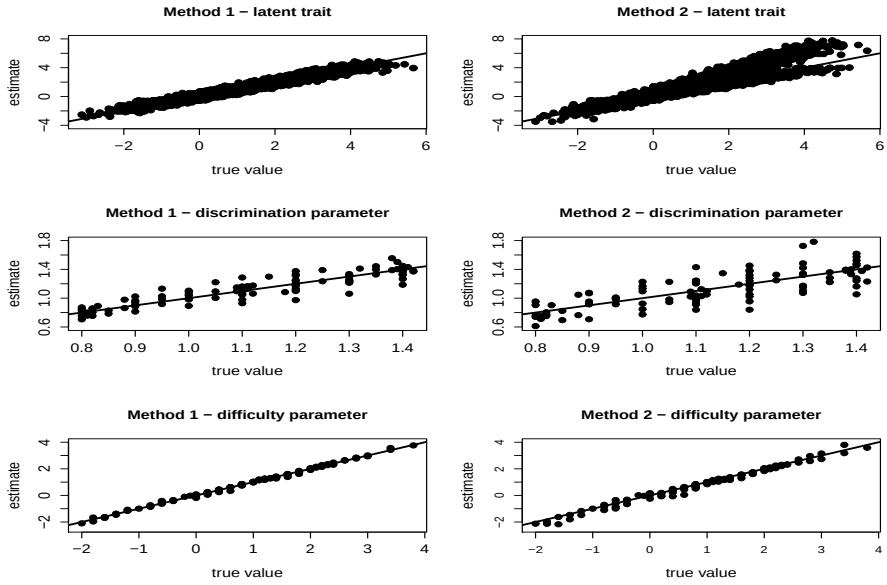

Figure 1 Posterior mean for the latent traits, discrimination and difficulty parameter using scaling/equating methods 1 and 2.

In an explanatory analysis, the multiple-group model (MGM) (Azevedo et al., 2012), was used to estimate the latent traits given the response data according to a cross-sectional design. A total of six groups were considered (the male and the female groups measured at each of the three occasions). The MGM for cross-sectional data assumed that students were nested in groups and latent traits were assumed to be independently distributed over occasions given the mean level of the group. Subsequently, Pearson's correlations were estimated for the pairs of estimated latent traits corresponding to years one to three. The results are presented in Table 8. It can be seen that the unstructured covariance pattern model can describe the relation between traits over occasions.

Therefore, a longitudinal multiple group model (LMGM) with an unstructured covariance matrix for males and females was considered. The 95% HPD intervals of the correlation parameters showed that they were not significantly different from each other over occasions. Hence, an LMGM with common correlations over occasions for the latent traits was assumed for the male and female group (i.e., an heteroscedastic uniform covariance structure for each group).

Furthermore, model-assessment tools developed by Azevedo et al. (2014) were considered to evaluate the fit of the model. The Pearson chi-squared discrepancy measure (CHDM) and the predicted distribution for the scores, were used to evaluate the fit of the model. The overall Bayesian p-value (related to the CHDM) of p = .3922 indicated that the model fitted well. Figure 2 presents the observed, predicted and the 95% credibility intervals related to the score distributions. It can be seen that almost all observed scores fall within the credibility intervals, for each gender and year, which confirmed that the model fitted well. The results of the item-fit analysis are shown in Figure 3. When looking at the chi-square statistics, it can be seen that all items were fitted well by the model.

Figure 4 represents the population parameter estimates and 95% HPD credible intervals of the population parameters for males and females. For both groups, a slight growth in the means from the first to the second year and a slight decrease from the second to the third year were detected. Most likely due to the small sample size, the estimated mean differences were not significant. The measurement-occasion specific variances appeared to be common over years for males and females. Furthermore, the correlations among the latent traits appear to be equal for males and females. In conclusion, the covariance matrix of the latent traits for the males and for the females has time-homogenous variances and correlations.

Finally, Figures 5 and 6 represent the posterior means and 95% credible intervals of the discrimination and difficulty parameter estimates, respectively. The discrimination parameter estimates indicate that, in general, the items were adequate since 75% of them had sufficient discriminating power ($\geq .60$). In addition, by comparing the difficulty parameter estimates with the population mean estimates, it follows that the symptoms measured by the items were likely to be observed for high latent trait values, since all difficulty values were significantly positive.

Table 8 Pearson's between-year correlations of the latent trait scores of males and females.

	Male		**Female**	
	Year 2	**Year 3**	**Year 2**	**Year 3**
Year 1	0.616	0.499	0.596	0.635
Year 2	-	0.709	-	0.748

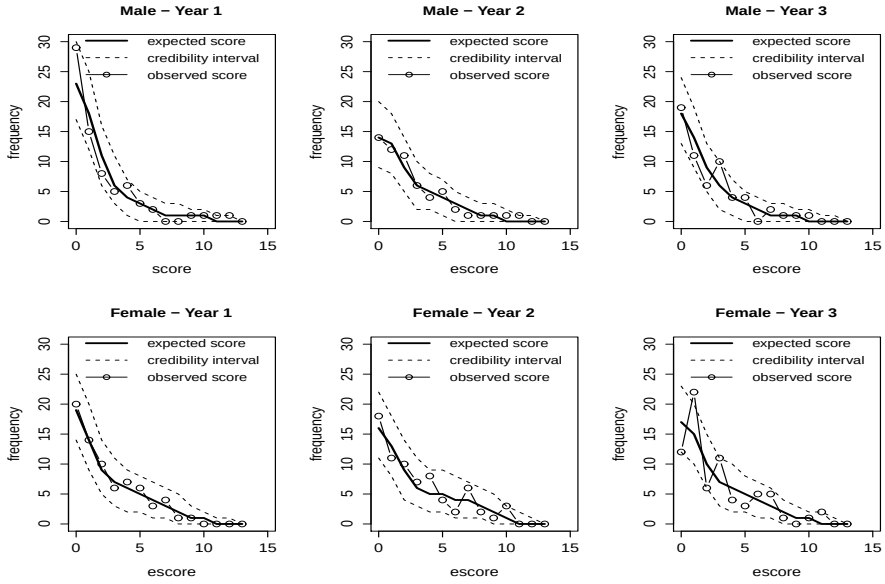

Figure 2 Observed and predicted score distribution and 95% central credible intervals.

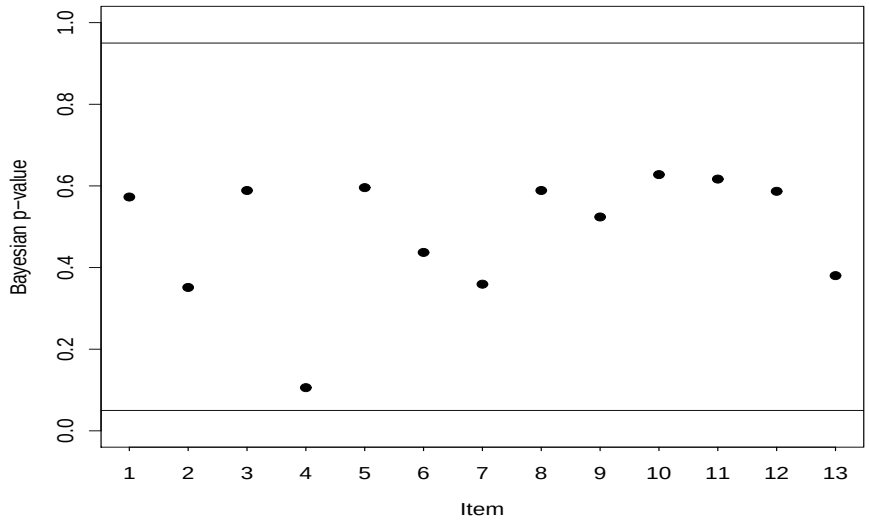

Figure 3 Posterior predictive p-values corresponding to the item-based discrepancy measure based on Pearson (chi-square statistic).

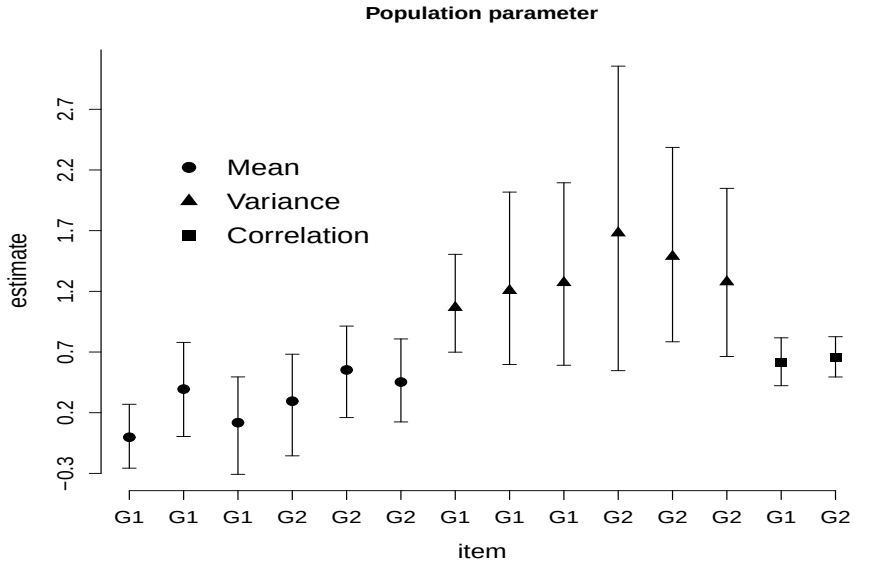

Figure 4 Posterior means and HPD intervals of the male and female population distribution (G1=male; G2=female).

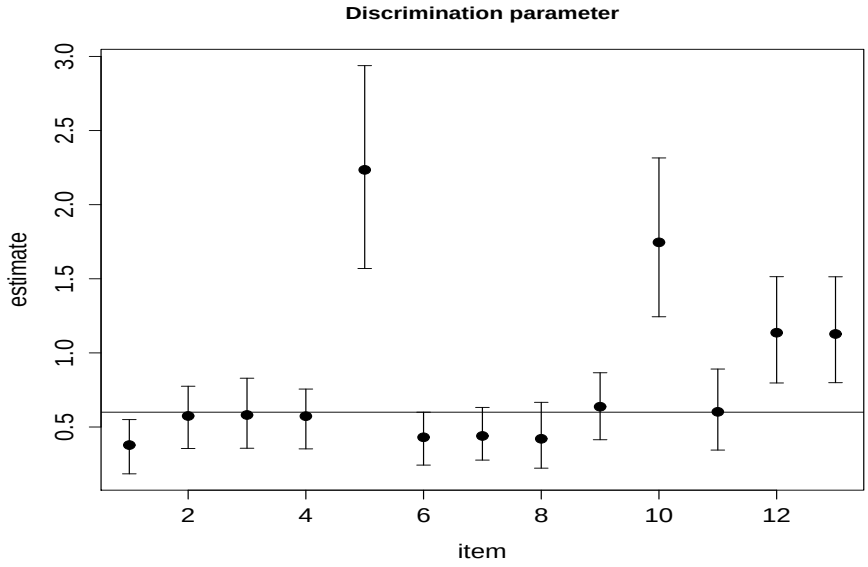

Figure 5 Posterior means and HPD intervals for the discrimination parameters.

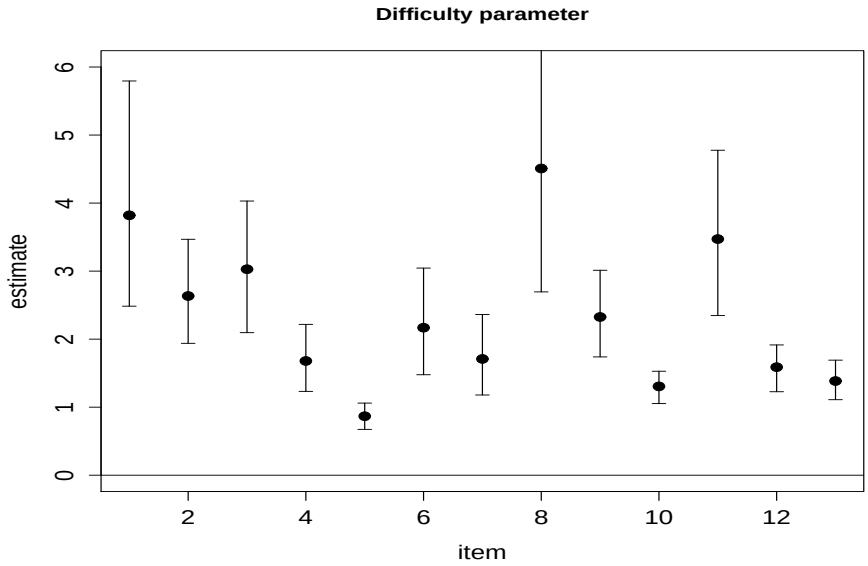

Figure 6 Posterior means and HPD intervals for the difficulty parameters.

5.2 The Brazilian longitudinal educational study

This study was conducted from 1999 to 2003. At the start, 158 public schools were monitored, where 55 schools were selected for the program. The sampled schools were

located over six Brazilian states with two states in each of three Brazilian regions (North, Northeast, and Center West). The schools had at least 200 students enrolled for the daytime educational programs, were located at urban zones and offered an educational program to the eighth grade. At baseline, a total of 12,580 students were sampled. From 2000 to 2003, the cohort consisted of students from the baseline sample who were approved to the fifth grade and did not switch schools. Students enrolled in the fifth grade but coming from another school, and students not assessed in former grades constituted a second cohort, which was followed the four subsequent years. Other cohorts were defined in the same way. The longitudinal test design allowed dropouts and inclusions along the time-points. Besides achievements, social-cultural information was collected. The selected students were tested each year. The data has been analyzed by Azevedo et al. (2014) using the longitudinal single-group IRT model, and they provide more details about this study.

Here, the mathematic performance of 1,500 randomly selected students, who were assessed in the fourth, fifth, and sixth grade, were considered. A total of 72 test items was used, where 23, 26, and 31 items were used in the tests in grade four, grade five, and grade six, respectively. Five anchor items were used in all three tests. Another common set of five items was used in the test in grade four and five. Furthermore, four common items were used in the tests in grade five and six.

To investigate the time-heterogeneous covariance structure, two competing covariance models were considered, the ARH and ARMAH, which have been implemented in the MCMC-based algorithms. The RJMCMC algorithm results showed that the ARMAH model was visited 80.1% of the iterations and, thus, this covariance pattern model was selected. To have a valid sample size for the selected model of approximately 1,000 iterations, a burn-in of 16,000 iterations was used, a thin of 30 iterations, and a total of 53,500 MCMC iterations was considered.

The model assessment tools developed by Azevedo et al. (2014) were used to evaluate the fit of the model. The overall Bayesian p-value (related to the CHDM) of $p = .36$, indicated a well-fitting model. Figure 7 represents the observed, predicted and the 95% credibility intervals, related to the score distributions. It can be seen that most of the observed scores fall within the credibility intervals, for each grade. The results of the item-fit analysis are shown in Figure 8. When looking at the chi-square statistics, it can be seen that almost all of the items were well fitted by the model (57 of 72). Further investigation could be considered to look for a more appropriate item response function for the other 15 items, for example, using an asymmetric function as presented in Bazan et al (2006).

Table 9 represents the population parameter estimates and 95% HPD credible intervals of the three grade levels while accounting for a time-heterogenous correlation structure among latent traits. A significant growth in latent trait means was detected given the non-overlapping credible intervals. As expected, the mean growth of math achievement over grade years was significant. The within-grade variability was relatively small, but the between-grade correlations were significant. Each within-examinee latent growth was computed, while accounting for the complex dependencies, which showed a comparable pattern compared to the mean latent growth over grade years.

Figures 9 and 10 represent the posterior means and posterior standard deviations of the item discrimination and difficulty parameters, respectively. The discrimination parameter estimates are relatively low, where approximately 50% of the items have sufficient discrimination power ($\geq .60$). In addition, by comparing the difficulty parameter estimates with the population mean estimates, it follows that the tests were relatively easy, since most

of the difficulty values are below zero, which corresponds to the lowest value among all latent trait population averages.

In Figure 11, the parameter estimates obtained using MCMC (and the ARMAH covariance matrix) were compared to the estimates obtained using the RJMCMC algorithm (which includes the selection of the covariance matrix of latent traits). It can be seen that both posterior means and posterior standard deviations were quite similar. Therefore, the model uncertainty related to the selection of the appropriate covariance structure did not seem to affect the accuracy of the estimates (provided that for both cases the valid MCMC samples had comparable sizes).

Table 9 Population parameter estimates and 95% HPD intervals.

		Mean	
Grade	**Mean**	**SD**	**HPD 95%**
four (reference)	-	-	-
five	0.212	0.036	[0.137 ; 0.291]
six	0.683	0.046	[0.594 ; 0.776]
		Variance	
Grade	**Mean**	**SD**	**HPD 95%**
Grade	-	-	-
four (reference)	0.856	0.064	[0.732;0.990]
five	0.739	0.067	[0.617 ; 873]
		Correlation (ρ_θ)	
Grades	**Mean**	**SD**	**HPD 95%**
-	0.932	0.016	[0.899 ; 0.963]
		Correlation (γ_θ)	
Grades	**Mean**	**SD**	**HPD 95%**
-	0.855	0.013	[0.834 ; 0.884]

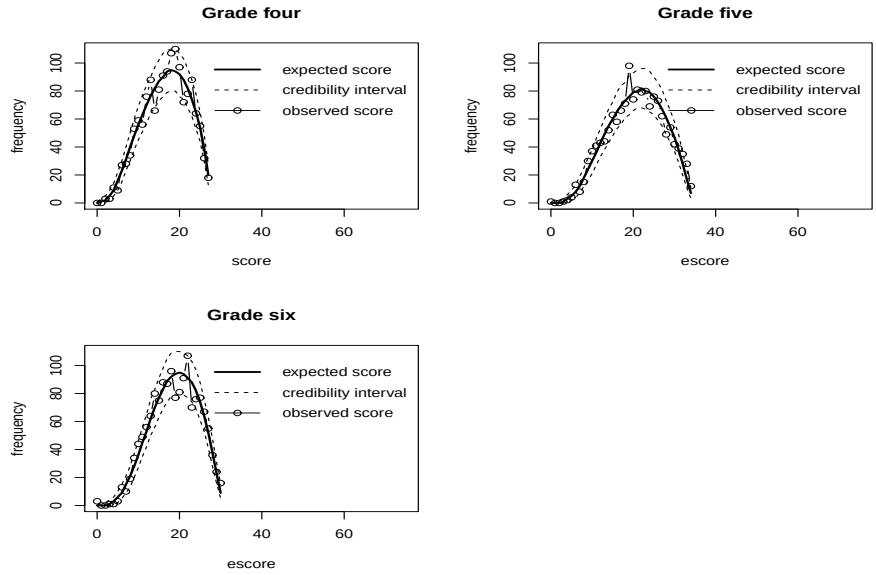

Figure 7 Observed score distribution, predicted score distribution, and 95% central credible intervals.

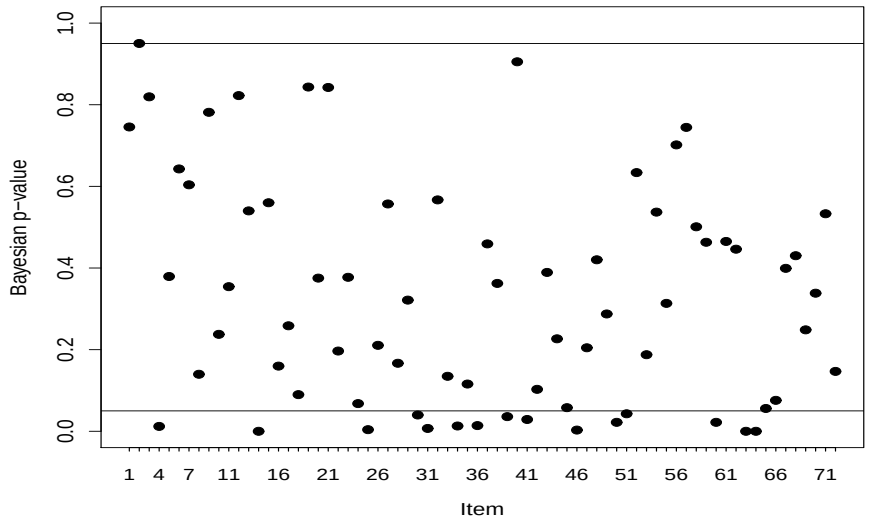

Figure 8 Posterior predictive p-values corresponding to the item-based discrepancy measure based on Pearson (chi-square statistic).

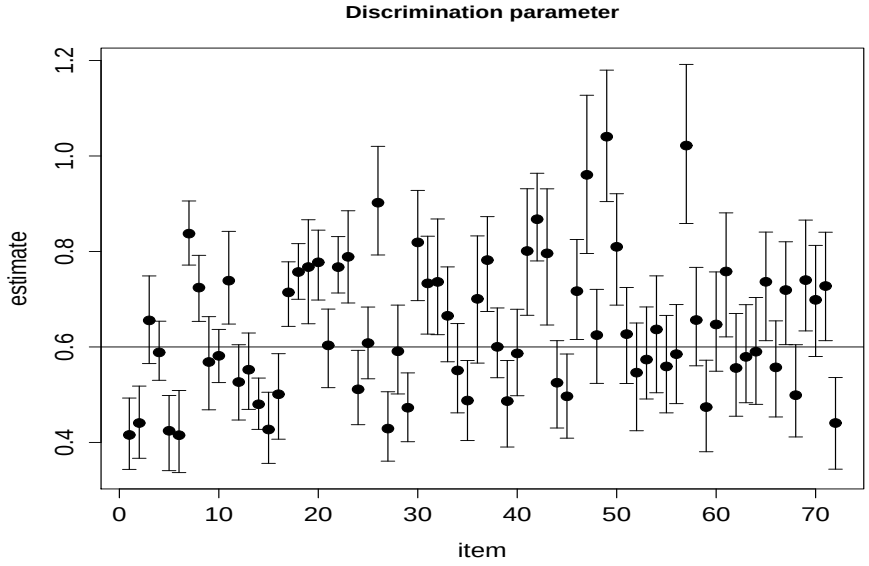

Figure 9 Posterior means and HPD intervals for the discrimination parameters.

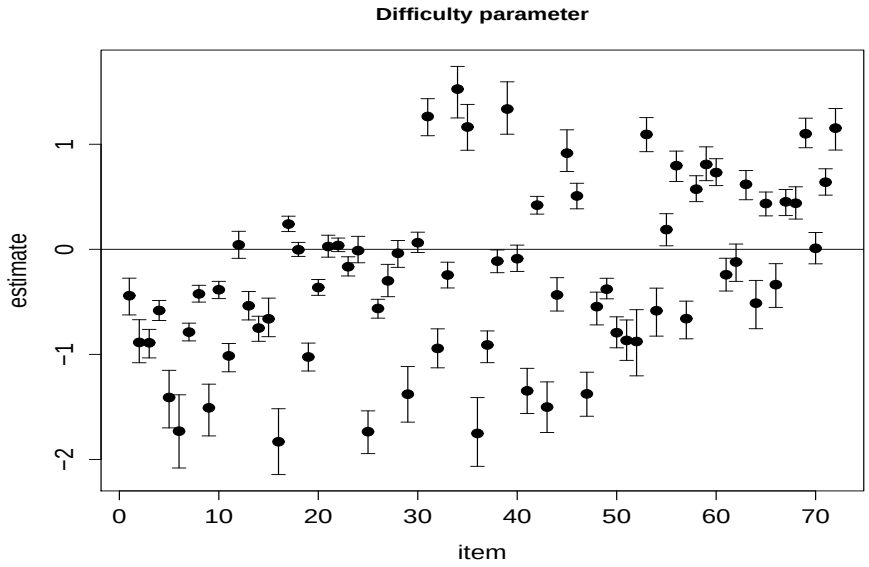

Figure 10 Posterior means and HPD intervals for the difficulty parameters.

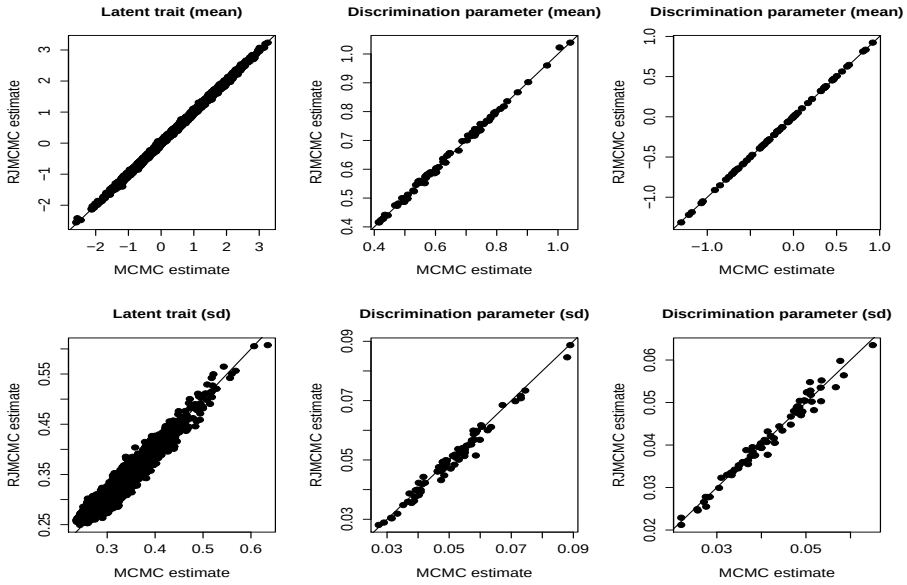

Figure 11 Posterior means (mean) and posterior standard deviations (sd) for the latent traits, discrimination and difficulty parameters using RJMCMC and MCMC algorithm.

6 Conclusions and Comments

A longitudinal multiple-group IRT model with group-specific time-heterogenous covariance structures for the latent trait distribution has been proposed. The developed MCMC algorithm can handle identification issues, the scaling process and the selection of restricted covariance patterns of latent traits across groups. Furthermore, an RJMCMC algorithm for the joint parameter estimation and covariance pattern selection have been proposed for the single-group longitudinal IRT model.

The simulation studies showed a good recovery of the model parameters using MCMC-based algorithms, and also showed that the correct underlying covariance pattern model can be selected using the RJMCMC method. In addition, it was shown that the use of a posterior equating method can be avoided, as expected, similar to the results obtained by Andrade (2001) (who compared the multiple group model in combination with posterior equating methods for cross-sectional data). Therefore, the proposed modeling approach will be very useful for analyzing longitudinal multiple-group IRT data. The proposed explicit modeling of covariance patterns of latent variables could also be useful in related works, for example, in the longitudinal IRT models of Entink al. (2011) and van den Hout et al. (2011), and in the longitudinal single-group IRT models of von Davier and Xu (2011), te Marveld et al. (2006) and Embretson (1991).

Other model extensions of the LMGIRT model can be considered to obtain a more realistic description of the longitudinal response data. The assumed multivariate normal latent variable distribution can be adjusted to account for skewness. For example, by

using a multivariate skewed latent variable distribution to model asymmetric latent trait distributions. The skewed latent variable approach of Azevedo et al. (2011) could be used. The extension to nominal and ordinal response data can be made by defining a more flexible item response model. Dropout and inclusion of examinees were not allowed in the current data analysis. A multiple imputation method could be developed to support this situation, see for example, Azevedo (2008). In general, the LIRT model can be adapted to accommodate incomplete designs, latent growth curves, collateral information for latent traits, informative mechanisms of non-response, mixture structures on latent traits and/or item and population parameters, and flexible latent trait distributions, among other things. This requires defining a general IRT model for the response data using flexible priors that can include the different extensions. The RJMCMC algorithm can be adapted to allow other covariance structures (as discussed in Azevedo et al. (2014)).

7 Appendix

7.1 ARH, ARMAH and UH matrices

The strutcture of the ARMAH is given by

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{2k}}}\gamma_{\theta_k} & \cdots & \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-2} \\ \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{k2}}}\gamma_{\theta_k} & \psi_{\theta_{k2}} & \cdots & \sqrt{\psi_{\theta_{k2}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta}^{T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-2} & \sqrt{\psi_{\theta_{k2}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-3} & \cdots & \psi_{\theta_{kT}} \end{bmatrix}. \quad (6)$$

The ARH matrix is obtained by doing $\rho_{\theta_k} = \gamma_{\theta_k}, \forall\, k$ in Equation (6). On the other hand, the structure of the UH matrix is given by:

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k2}}}\rho_{\theta_k} & \cdots & \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} \\ \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k2}}}\rho_{\theta_k} & \psi_{\theta_{k2}} & \cdots & \sqrt{\psi_{\theta_{k2}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} & \sqrt{\psi_{\theta_{k2}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} & \cdots & \psi_{\theta_{kT}} \end{bmatrix}. \quad (7)$$

7.2 Restricted Unstructured Covariance Structure

Following Azevedo et al. (2014) and McCulloch et al. (2000), a parametrization of the latent trait's covariance structure is considered. Therefore, the following partition of the latent traits structure is defined,

$$\begin{aligned} \boldsymbol{\theta}_{jk.} &= (\theta_{j11}, \theta_{jk2}, \ldots, \theta_{jkT})^t = (\theta_{jk1}, \boldsymbol{\theta}_{jk(1)})^t, \\ \boldsymbol{\mu}_{\boldsymbol{\theta}_k} &= (\mu_{\theta_{k1}}, \mu_{\theta_{k2}}, \ldots, \mu_{\theta_{kT}})^t = (\mu_{\theta_{k1}}, \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}})^t, \end{aligned}$$

where, $\boldsymbol{\theta}_{jk(1)} = (\theta_{jk2}, \ldots, \theta_{jkT})^t$, $\boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}} = (\mu_{\theta_{k2}}, \ldots, \mu_{\theta_{kT}})^t$. Furthermore, the covariance structure is partitioned as,

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \boldsymbol{\psi}^t_{\boldsymbol{\theta}_{k(1)}} \\ \boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}} & \boldsymbol{\Psi}_{\boldsymbol{\theta}_{k(1)}} \end{bmatrix}, \tag{8}$$

where $\boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}} = (\psi_{\theta_{k12}}, \ldots, \psi_{\theta_{k1T}})^t$ and

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_{k(1)}} = \begin{bmatrix} \psi_{\theta_{k2}} & \ldots \psi_{\theta_{k2T}} \\ \vdots & \ddots & \vdots \\ \psi_{\theta_{k2T}} & \ldots & \psi_{\theta_{kT}} \end{bmatrix}. \tag{9}$$

From properties of the multivariate normal distribution, see Rencher (2002), it follows that

$$\boldsymbol{\theta}_{jk(1)} | \theta_{jk1} \sim N_{(T-1)} \left(\boldsymbol{\mu}^*_k, \boldsymbol{\Psi}^*_k \right), \tag{10}$$

where, $\boldsymbol{\mu}^*_k = \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}} + \psi^{-1}_{\theta_{k1}} \boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}} (\theta_{jk1} - \mu_{\theta_{k1}})$,
and

$$\boldsymbol{\Psi}^*_k = \boldsymbol{\Psi}_{\boldsymbol{\theta}_{k(1)}} - \psi^{-1}_{\theta_{k1}} \boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}} \boldsymbol{\psi}^t_{\boldsymbol{\theta}_{k(1)}}. \tag{11}$$

This is equivalent to

$$\boldsymbol{\theta}_{jk(1)} \mid \theta_{jk1} = \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}} + \frac{\boldsymbol{\psi}^*_k}{\sqrt{\psi_{\theta_{k1}}}} (\theta_{jk1} - \mu_{\theta_{k1}}) + \boldsymbol{\xi}_{jk}, \tag{12}$$

where $\boldsymbol{\xi}_{jk} \sim N_{(T-1)} (\mathbf{0}, \boldsymbol{\Psi}^*_k)$ and $\boldsymbol{\psi}^*_k$ is given by (13). That is, the $\boldsymbol{\theta}_{jk(1)}$ are conditionally multivariate normally distributed given the first component θ_{jk1}, with an unrestricted covariance matrix. Equation (12) defines a linear multivariate regression model with independent variable $(\theta_{jk1} - \mu_{\theta_{k1}})$, intercept $\boldsymbol{\mu}_{\boldsymbol{\theta}_{k1}}$, and regression parameters

$$\boldsymbol{\psi}^*_k = \boldsymbol{\psi}_{\boldsymbol{\theta}_{k(1)}} / \sqrt{\psi_{\theta_{k1}}}. \tag{13}$$

The reparameterization in Equation (13) is considered to handle the restriction $\psi_{\theta_{11}} = 1$, in the MCMC and RJMCMC algorithms. In addition, the matrix $\boldsymbol{\Psi}^*_k$ is an unstructured covariance matrix without any identifiability restrictions, see Singer and Andrade (2000). As a result, the common modeling (e.g., using an Inverse-Wishart prior) and estimation approaches can be applied for Bayesian inference, see Gelman et al. (2004).

Therefore, the parameters,

$$(\boldsymbol{\psi}^{*^t}_k, \boldsymbol{\Psi}^{*^t}_{\boldsymbol{\theta}_k})^t, \tag{14}$$

define a one-to-one relation with the free parameters of the original covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}$. As a result, the estimates of the population parameters for each group (combination of the

original group with time-point) can be obtained from the estimates of the parameters in Equation (14). The latent variable distribution of the reference group (the combination of the original group with time-point) will be restricted to identify the model. That is, the mean and the variance of the latent trait distribution are fixed to arbitrary values. In our case, without loss of generality, the reference group is the first group measured at the first occasion. This is done by re-scaling the respective vector of latent variable values to a pre-specified scale in each MCMC (RJMCMC) iteration. The latent variable population distribution of subsequent measurement occasions for this group are conditionally specified according to Equation (12), given the restricted population distribution parameters of the first measurement occasion. Subsequently, the covariance parameters of the latent multivariate model are not restricted for identification purposes, which will facilitate a straightforward specification of the prior distributions.

7.3 MCMC estimation for the longitudinal MGM

Following Azevedo et al. (2014) and McCulloch et al. (2000), conditional conjugate prior distributions are considered, see Gelman et al. (2004) and Gelman (2006). Remembering that the parameters of interest are $(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}^t, \psi_{\theta_{k1}}, \boldsymbol{\psi}_k^{*^t})$ and $\boldsymbol{\Psi}_k, k = 1, ..., K$ and conditional conjugate priors are specified as,

$$\boldsymbol{\mu}_{\boldsymbol{\theta}_k} \sim N_T(\boldsymbol{\mu}_{0k}, \boldsymbol{\Psi}_{0k}), \tag{15}$$

$$\psi_{\theta_{k1}} \sim IG(\nu_{0k,}\kappa_{0k}), \tag{16}$$

$$\boldsymbol{\psi}_k^* \sim N_{T-1}(\boldsymbol{\mu}_{\boldsymbol{\psi}_k}, \boldsymbol{\Psi}_{\boldsymbol{\psi}_k}), \tag{17}$$

$$\boldsymbol{\Psi}_k^* \sim IW_{T-1}(\nu_{\boldsymbol{\Psi}_k}, \boldsymbol{\Psi}_{\boldsymbol{\Psi}_k}), \tag{18}$$

where $IG(\nu_{0k}, \kappa_{0k})$ stands for the inverse-gamma distribution with shape parameter ν_{0k} and scale parameter κ_{0k}, and $IW_{T-1}(\nu_{\boldsymbol{\Psi}_k}, \boldsymbol{\Psi}_{\boldsymbol{\Psi}_k})$ for the inverse-Wishart distribution with degrees of freedom $\nu_{\boldsymbol{\Psi}_k}$ and dispersion matrix $\boldsymbol{\Psi}_{\boldsymbol{\Psi}_k}$.

For the item parameters, the prior is specified as

$$p(\boldsymbol{\zeta}_i \mid \boldsymbol{\mu}_{\boldsymbol{\zeta}}, \boldsymbol{\Psi}_{\boldsymbol{\zeta}}) \propto \exp\Bigg(-0.5 (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{\boldsymbol{\zeta}})^t \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1} \times (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{\boldsymbol{\zeta}}) \Bigg) \mathbb{1}_{(a_i>0)} \mathbb{1}_{(b_i \in (-\infty,\infty))}, \tag{19}$$

where $\boldsymbol{\mu}_{\boldsymbol{\zeta}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\zeta}}$ are the hyperparameters, and $\mathbb{1}$ is the usual indicator function. The hyperparameters are fixed and often set in such a way that they represent reasonable values for the prior parameters.

As in Azevedo et al. (2014), to facilitate the FGS implementation, and to account for missing response data, an augmented data scheme will be introduced, see also Albert (1992). It corresponds to sample normally distributed latent response data $\boldsymbol{Z}_{....} = (Z_{1111}, ..., Z_{\mathcal{I}_{KT} n_K KT})^t$, given the discrete observed response data; that is,

$$Z_{ijkt} | (\theta_{jkt}, \boldsymbol{\zeta}_i, Y_{ijkt}) \sim N(a_i \theta_{jkt} - b_i, 1), \tag{20}$$

where Y_{ijkt} is the indicator of Z_{ijkt} being greater than zero.

To handle incomplete block designs, an indicator variable $\boldsymbol{I}_{....} = (I_{1111}, ..., I_{\mathcal{I}_{KT} n_K KT})^t$ is defined that defines the set of administered items for each occasion and subject. This indicator variable is defined as follows,

$$I_{ijkt} = \begin{cases} 1, \text{item } i \text{ administered for examinee } j, \text{belonging to the group } k \\ \quad \text{at time-point } t \\ 0, \text{missing by design.} \end{cases} \tag{21}$$

The not-selective missing responses due to uncontrolled events as dropouts, inclusion of examinees, non-response, or errors in recoding data are marked by another indicator, which is defined as,

$$V_{ijkt} = \begin{cases} 1, \text{observed response of examinee } j, \text{belonging to the group } k \\ \quad \text{at time-point } t \text{ on item } i \\ 0, \text{otherwise.} \end{cases} \tag{22}$$

It is assumed that the missing data are missing at random (MAR), such that the distribution of patterns of missing data does not depend on the unobserved data. When the MAR assumption does not hold and the missing data are non-ignorable, a missing data model can be defined to model explicitly the pattern of missingness. In case of MAR, the observed data can be used to make valid inferences about the model parameters.

For the sake of simplicity, let indicator matrix $\boldsymbol{I}_{....}$ represent both cases of missing data. Then, under the above assumptions, the distribution of augmented data $\boldsymbol{Z}_{....}$ (conditioned on all other quantities) is given by

$$p(\boldsymbol{z}_{....} \mid \boldsymbol{y}_{....}, \boldsymbol{I}_{....}, \boldsymbol{\theta}_{...}, \boldsymbol{\zeta}, \boldsymbol{\eta_\theta}) \propto \prod_{k=1}^{K} \prod_{t=1}^{T} \prod_{j=1}^{n_k} \prod_{i|I_{ijkt}=1} \left\{ \exp\left\{ -0.5 \left(z_{ijkt} - a_i \theta_{jkt} + b_i \right)^2 \right\} \times \mathbb{1}_{(z_{ijkt}, y_{ijkt})} \right\}, \tag{23}$$

where $\boldsymbol{z}_{....}$ and $\boldsymbol{y}_{....}$ are the vectors with all augmented data and observed responses availabe, respectively, and $\mathbb{1}_{(z_{ijkt}, y_{ijkt})}$ represents the restriction that z_{ijkt} is greater (lesser) than zero when y_{ijkt} equals one (zero), according to Equation (20).

Given the augmented data likelihood in Equation (23) and the prior distributions in Equations (19), (15), (16), (17), (18) and (2), the joint posterior distribution is given by:

$$\begin{aligned} p(\boldsymbol{\theta}_{...}, \boldsymbol{\zeta}, \boldsymbol{\mu_\theta}, \boldsymbol{\psi}^*, \boldsymbol{\Psi}^* | \boldsymbol{z}_{....}, \boldsymbol{y}_{....}) \propto\ & p(\boldsymbol{z}_{....} \mid \boldsymbol{y}_{....}, \boldsymbol{I}_{....}, \boldsymbol{\theta}_{...}, \boldsymbol{\zeta}, \boldsymbol{\eta_\theta}) p(\boldsymbol{\theta}_{...} | \boldsymbol{\eta_\theta}) \\ & \times p(\boldsymbol{\zeta} | \boldsymbol{\mu_\zeta}, \boldsymbol{\Psi_\zeta}) p(\boldsymbol{\eta_\theta}), \end{aligned} \tag{24}$$

where

$$p(\boldsymbol{\theta}_{...} | \boldsymbol{\eta_\theta}) = \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(\boldsymbol{\theta}_{jk.} | \boldsymbol{\eta}_{\boldsymbol{\theta}_k}), \tag{25}$$

$$p(\boldsymbol{\zeta}|\boldsymbol{\mu}_{\boldsymbol{\zeta}}, \boldsymbol{\Psi}_{\boldsymbol{\zeta}}) = \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i|\boldsymbol{\mu}_{\boldsymbol{\zeta}}, \boldsymbol{\Psi}_{\boldsymbol{\zeta}}),$$

$$p(\boldsymbol{\eta}_{\boldsymbol{\theta}}) = \prod_{k=1}^{k} p(\boldsymbol{\eta}_{\boldsymbol{\theta}_k}) = \prod_{k=1}^{K} p(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}) p(\psi_{\theta_{k1}}) p(\boldsymbol{\psi}_k^*) p(\boldsymbol{\Psi}_k^*),$$

and $\boldsymbol{\theta}_{...} = (\theta_{111}, ..., \theta_{n_K KT})^t$, $\boldsymbol{\eta}_{\boldsymbol{\theta}} = (\boldsymbol{\eta}_{\boldsymbol{\theta}_1}^t, ..., \boldsymbol{\eta}_{\boldsymbol{\theta}_K}^t)^t$, $\boldsymbol{\psi}^* = (\boldsymbol{\psi}_k^{*^t}, ..., \boldsymbol{\psi}_k^{*^t})^t$ and $\boldsymbol{\Psi}^* = (\boldsymbol{\Psi}_1, ..., \boldsymbol{\Psi}_k)$. The posterior distribution, given by equation (24), has an intractable form but, the full conditionals are known and easy to sample from. More specifically, we have:

- **Step 1 :** Simulate the augmented data using $Z_{ijkt}|(.)$, according to Equation (20).
- **Step 2 :** Simulate the latent traits using

$$\boldsymbol{\theta}_{jk.}|(.) \sim N_T(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}_{jk}} \widehat{\boldsymbol{\theta}}_{jk}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}_{jk}})$$

  where

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{jk} &= \sum_{i|I_{ijkt}=1} a_i b_i \mathbf{1}_T + \sum_{i|I_{ijkt}=1} a_i \boldsymbol{z}_{ijk.} + \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{-1} \boldsymbol{\mu}_{\boldsymbol{\theta}_k}\,, \\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}_{jk}} &= \left(\sum_{i|I_{ijkt}=1} a_i^2 \boldsymbol{I}_T + \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{-1} \right)^{-1}, \end{aligned}$$

  $\boldsymbol{z}_{ijk.} = (z_{ijk1}, \ldots, z_{ijkT})^t$, $\mathbf{1}$ is a unit vector of size T and $\boldsymbol{I}_T$ is a identity matrix of order T.

- **Step 3 :** Simulate the item parameters by using $\boldsymbol{\zeta}_i|(.) \sim N(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i} \widehat{\boldsymbol{\zeta}}_i, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i})$, mutually indepedently, where

$$\begin{aligned} \widehat{\boldsymbol{\zeta}}_i &= \boldsymbol{H}_{i..}^t \boldsymbol{z}_{i...} + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\zeta}}\,, \\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i} &= \left(\boldsymbol{H}_{i...}^t \boldsymbol{H}_{i...} + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1} \right)^{-1}, \\ \boldsymbol{H}_{i...} &= [\boldsymbol{\theta}_{...} \;\; -\mathbf{1}] \bullet \mathbf{I}_i\,, \end{aligned} \tag{26}$$

  where $\boldsymbol{z}_{i...} = (\boldsymbol{z}_{1...}, ..., \boldsymbol{z}_{I...})^t$, $\mathbf{I}_i$ is the indicator vector of item i, which indicates the subjects responding to item i and "$\bullet$" is the Hadamard product.

- **Step 4 :** Simulate the population mean vectors, for $k = 1, ..., K$, by using

$$\begin{aligned}\mu_{\theta_{k1}}|(.) &\sim N(\widetilde{\mu}_{\theta_{k1}}, \widehat{\psi}_{\mu_k})\,,\\ \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}}|(\mu_{\theta_{k1}},(.)) &\sim N_T(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_{k(T-1)}}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}_{k(T-1)}})\,,\end{aligned}$$

where

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}_k} &= \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{-1}\sum_{j=1}^{n_k}\boldsymbol{\theta}_{jk.} + \boldsymbol{\Psi}_{0k}^{-1}\boldsymbol{\mu}_{0k} = (\widehat{\mu}_{\theta_{k1}}, \widehat{\mu}_{\theta_{k2}}, \ldots, \widehat{\mu}_{\theta_{kT}})^t = (\widehat{\mu}_{\theta_{k1}}, \widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}_k}^{(T-1)})^t\,,\\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}_k} &= \left(n_k\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{-1} + \boldsymbol{\Psi}_{0k}^{-1}\right)^{-1} = \begin{bmatrix}\widehat{\psi}_{\mu_k} & \widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}_k}^{t\,(T-1)}\\ \widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}_k}^{(T-1)} & \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}_k}^{(T-1)}\end{bmatrix}\,,\\ \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_k} &= \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}_k}\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}_k} = (\widetilde{\mu}_{\theta_{k1}}, \widetilde{\mu}_{\theta_{k2}}, \ldots, \widetilde{\mu}_{\theta_{kT}})^t = (\widetilde{\mu}_{\theta_{k1}}, \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_k}^{(T-1)})^t\,,\\ \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_{k(T-1)}} &= \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_k}^{(T-1)} + \widehat{\psi}_{\mu_k}^{-1}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}_k}^{(T-1)}(\mu_{\theta_{k1}} - \widetilde{\mu}_{\theta_{k1}})\,,\\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}_{k(T-1)}} &= \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}_k}^{(T-1)} - \widehat{\psi}_{\mu_k}^{-1}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}_k}^{(T-1)}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}_k}^{t\,(T-1)}\,.\end{aligned}$$

- **Step 5 :** Simulate the first time-point variance, for $k = 1, ..., K$, using $\psi_{\theta_{k1}}|(.) \sim IG(\widehat{\upsilon}_{0k}, \widehat{\kappa}_{0k})$, where

$$\begin{aligned}\widehat{\upsilon}_{k1} &= \frac{n_k + \upsilon_{0k}}{2}\,,\\ \widehat{\kappa}_{k1} &= \frac{\sum_{j=1}^{n_k}(\theta_{jk1} - \mu_{\theta_{k1}})^2 + \kappa_{0k}}{2}\,.\end{aligned}$$

- **Step 6 :** Simulate the vector of covariances, for $k = 1, ..., K$, using $\boldsymbol{\psi}_k^* \sim N_{T-1}(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\psi}_k}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\psi}_k}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\psi}_k})$, where

$$\begin{aligned}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\psi}_k} &= \psi_{\theta_{k1}}^{-1/2}\left(\boldsymbol{\Psi}_k^*\right)^{-1}\sum_{j=1}^{n_k}\left(\boldsymbol{\theta}_{jk(1)} - \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}}\right)(\theta_{jk1} - \mu_{\theta_{k1}}) + \boldsymbol{\Psi}_{\boldsymbol{\psi}_k}^{-1}\boldsymbol{\mu}_{\boldsymbol{\psi}_k}\,,\\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\psi}_k} &= \left(\psi_{\theta_{k1}}^{-1}\left(\boldsymbol{\Psi}_k^*\right)^{-1}\sum_{j=1}^{n_k}(\theta_{jk1} - \mu_{\theta_{k1}})^2 + \boldsymbol{\Psi}_{\boldsymbol{\psi}_k}^{-1}\right)^{-1}\,.\end{aligned}$$

- **Step 7 :** Simulate the covariance matrix, for $k = 1, ..., K$, using $\boldsymbol{\Psi}_k^* \sim IW_{T-1}(\widehat{\nu}_{\boldsymbol{\Psi}_k}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\Psi}_k})$, where

$$\begin{aligned}\widehat{\nu}_{\boldsymbol{\Psi}_k} &= n_k + \nu_{\boldsymbol{\Psi}_k}\,,\\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\Psi}_k} &= \boldsymbol{\Psi}_{\boldsymbol{\Psi}_k} + \sum_{j=1}^{n_k}\left(\boldsymbol{\theta}_{jk(1)} - \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}}\right)\left(\boldsymbol{\theta}_{jk(1)} - \boldsymbol{\mu}_{\boldsymbol{\theta}_{k(1)}}\right)^t\,.\end{aligned}$$

- **Step 8 :** Calculate the original covariance matrix, for $k = 1, ..., K$, using (8) and $\boldsymbol{\Psi}_{\boldsymbol{\theta}_{k(1)}} = \boldsymbol{\Psi}_k^* + \boldsymbol{\psi}_k^* \boldsymbol{\psi}_k^{*^t}$.
- **Step 9 :** Calculate the population variances, for $(k = 1, ..., K)$, using

$$(\psi_{\theta_{k2}}, \ldots, \psi_{\theta_{kT}})^t = \boldsymbol{\psi}^*_{\boldsymbol{\theta}_{k(1)}} = Diag(\boldsymbol{\Psi}_k^* + \boldsymbol{\psi}_k^* \boldsymbol{\psi}_k^{*t}), \tag{27}$$

  where $Diag$ extracts the main diagonal of a square matrix.
- **Step 10 :** For $(k = 1, ..., K)$, depending on the restricted covariance structure of interest, transformations are defined for unrestricted parameters to facilitate draws of restricted model parameters. These details can be found in Azevedo et al. (2014).

7.4 RJMCMC algorithm for the longitudinal one-group model

The Steps 1 - 10 of the RJMCMC algorithm are easily obtained from the developments presented in subsection 7.3 considering K=1. To implement the Step 11 of the RJMCMC algorithm we need to define prior distributions for each parameter of the covariance structure corresponding to the competing models (i.e., the ARH and ARMAH models). We consider:

$$\rho_{\theta(ARH)} \sim N_{(-1,1)}(\mu_\rho, \psi_\rho) \tag{28}$$
$$\rho_{\theta(ARMAH)} \sim N_{(-1,1)}(\mu_\rho, \psi_\rho) \tag{29}$$
$$\gamma_\theta \sim N_{(-1,1)}(\mu_\gamma, \psi_\gamma), \tag{30}$$

where $N_{(a,b)}(\mu, \psi)$ stands for a normal distribution truncated to the interval (a, b), μ and ψ are, respectively, the mean and the variance of the original normal distribution, and $\rho_{\theta(.)}$ the correlation parameter of the covariance structure (.).

For the RJMCMC algorithm presented in subsection 3.2, when considering two covariance matrices as the possible states of the chain (in our case ARH and ARMAH matrices), and given the prior distributions (28), (29) and (30), the Step 11 of the RJMCMC algorithm is given by:

1. If the current model is ARH
   - **Step 1** Simulate $\gamma_\theta = u \sim U(0, 1)$ and calculate $\rho_{\theta(ARMAH)} = \frac{\rho_{\theta(ARH)} + \gamma_\theta}{2}$.
   - **Step 2** Generate the matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}(ARMAH)}$ by using $(\rho_{\theta(ARMAH)}, \gamma_\theta)^t$ and $\boldsymbol{\psi}_{\boldsymbol{\theta}}$ (the vector with current simulated values for the population variances).
   - **Step 3** Calculate the acceptance rate:

$$R_{(ARH,ARMAH)} = \frac{\exp\left\{-0.5 \sum_{j=1}^{n} (\boldsymbol{\theta}_{j.} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^t \boldsymbol{\Psi}^{-1}_{\boldsymbol{\theta}(ARMAH)} (\boldsymbol{\theta}_{j.} - \boldsymbol{\mu}_{\boldsymbol{\theta}})\right\}}{\exp\left\{-0.5 \sum_{j=1}^{n} (\boldsymbol{\theta}_{j.} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^t \boldsymbol{\Psi}^{-1}_{\boldsymbol{\theta}(ARH)} (\boldsymbol{\theta}_{j.} - \boldsymbol{\mu}_{\boldsymbol{\theta}})\right\}} \times$$
$$\frac{\exp\left\{-\frac{0.5}{\psi_\rho}\left(\rho_{\theta(ARMAH)} - \mu_\rho\right)^2\right\} \exp\left\{-\frac{0.5}{\psi_\gamma}\left(\gamma_\theta - \mu_\gamma\right)^2\right\}}{\exp\left\{-\frac{0.5}{\psi_\rho}\left(\rho_{\theta(ARH)} - \mu_\rho\right)^2\right\}}$$
$$\times \frac{1}{\sqrt{\pi \psi_\gamma}} \frac{|\boldsymbol{\Psi}_{\boldsymbol{\theta}(ARMAH)}|^{-n/2}}{|\boldsymbol{\Psi}_{\boldsymbol{\theta}(ARH)}|^{-n/2}}.$$

Accept the model ARMAH with probability $\pi_{(ARH,ARMAH)} = \min\left\{R_{(ARH,ARMAH)}, 1\right\}$, otherwise, continue with ARH.

2. If the current model is ARMAH

- **Step 1** Calculate $u = \gamma_\theta$ and $\rho_{\theta(ARH)} = 2\rho_{\theta(ARMAH)} - \gamma_\theta$.
- **Step 2** Generate the matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}(ARH)}$ by using $\rho_{\theta(ARH)}$ and $\boldsymbol{\psi_\theta}$ (the vector with current simulated values for the population variances).
- **Step 3** Calculate the acceptance rate :

$$R_{(ARMAH,ARH)} = \frac{\exp\left\{-0.5\sum_{j=1}^{n}\left(\boldsymbol{\theta}_{j.}-\boldsymbol{\mu_\theta}\right)^t\boldsymbol{\Psi}^{-1}_{\boldsymbol{\theta}(ARH)}\left(\boldsymbol{\theta}_{j.}-\boldsymbol{\mu_\theta}\right)\right\}}{\exp\left\{-0.5\sum_{j=1}^{n}\left(\boldsymbol{\theta}_{j.}-\boldsymbol{\mu_\theta}\right)^t\boldsymbol{\Psi}^{-1}_{\boldsymbol{\theta}(ARMAH)}\left(\boldsymbol{\theta}_{j.}-\boldsymbol{\mu_\theta}\right)\right\}} \times \frac{\exp\left\{-\frac{0.5}{\psi_\rho}\left(\rho_{\theta(ARH)}-\mu_\rho\right)^2\right\}}{\exp\left\{-\frac{0.5}{\psi_\rho}\left(\rho_{\theta(ARMAH)}-\mu_\rho\right)^2\right\}\exp\left\{-\frac{0.5}{\psi_\gamma}\left(\gamma_\theta-\mu_\gamma\right)^2\right\}} \times \sqrt{\pi\psi_\gamma}\frac{|\boldsymbol{\Psi}_{\boldsymbol{\theta}(ARH)}|^{-n/2}}{|\boldsymbol{\Psi}_{\boldsymbol{\theta}(ARMAH)}|^{-n/2}}$$

Accept the model ARH with probability $\pi_{(ARMAH,ARH)} = \min\left\{R_{(ARMAH,ARH)}, 1\right\}$, otherwise, continue with the ARMAH.

References

Albert, J. (1992), "Bayesian estimation of normal ogive item response curves using Gibbs sampling", *Journal of Educational and Behavioral Statistics*, 17, 3, 251-269.

Andrade, D. F. (2001) . "Comparing the performance of groups of students through item response theory", In Portuguese, *Estudos em Avaliação Educacional, São Paulo*, 23, 31-70.

Andrade, D. F., & Tavares, H. R. (2005), "Item Response Theory for Longitudinal Data: population parameter estimation", *Journal of Multivariate Analysis*, 95, 1-22.

Azevedo, C. L. N. (2008), "Multilevel multiple group longitudinal models in Item Response Theory: estimation methods and structural selection under a Bayesian perspective", *Unpublished PhD Thesis, In Portuguese*.

Azevedo, C.L.N., Andrade, D.F. (2010), "An estimation method for latent trait and population parameters in nominal response model", *Brazilian Journal of Probability and Statistics*, 24, 415–433.

Azevedo, C.L.N., Bolfarine H. and Andrade, D.F. (2011), "Bayesian inference for a skew-normal IRT model under the centred parameterization", *Computational Statistics and Data Analysis*, 55, 353–365.

Azevedo, C. L. N.,Andrade, D. F. and Fox,J.-P. (2012), "A Bayesian generalized multiple group IRT model with model-fit assessment tools", *Computational Statistics and Data Analysis*, 56, 4399-4412.

Azevedo, C. L. N., Fox,J.-P. & Andrade, D. F. (2014), "Bayesian longitudinal item response modeling with restricted covariance pattern structures", *Statistics & Computing, DOI: 10.1007/s11222-014-9518-5*.

Bazán, J. L., Branco, M. D. and Bolfarine, H. (2006), *A Skew Item Response Model*, Bayesian analysis, 1, 4, 861-892.

Bock, D. R. and Zimowski, M. F. (1997), "The Multiple Groups IRT", *in Handbook of modern Item Response Theory, Wim J. van der Linden and Ronald K. Hambleton eds, Springer-Verlag*

De Ayala, R., Sava-Bolesta, M. (1999), "Item parameter recovery for the nominal response model", *Applied Psychological Measurement*, 23, 3–19.

DeMars, C. E. (2003), "Sample size and the recovery of nominal response model item parameters", *Applied Psychological Measurement*, 27, 4, 275-288.

Douw L, Nieboer D, van Dijk B, Stam C and Twisk J. (2014), "A healthy brain in a healthy body: brain network correlates of physical and mental fitness". *PLoS One. 2014;9(2):e88202. doi:10.1371/journal.pone.0088202*.

Entink, R. H. K., Fox, J.-P., van den Hout, A. (2011), "A mixture model for the joint analysis of latent developmental trajectories and survival", *Statistics in Medicine*, 30, 18, 2310-2325.

Embretson, S. E. (1991), "A multidimensional latent trait model for measuring learning and change". *Psychometrika*, 56, 3, 495-515.

Gamerman, D., & Lopes, H. (2006), *Markov Chain Monte Carlo : Stochastic simulation for bayesian inference, second edition*, Chapman & Hall/CRC, London

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004), *Bayesian data analysis, second edition*, Chapman & Hall, London.

Gelman, A. (2006), "Prior distribution for variance parameters in hierarchical models", *Bayesian Analysis*, 1, 3, 515-533.

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, 99, 4, 5711-732.

Hoekstra T., Barbosa-leiker C., Koppes L. and Twisk J. (2011), "Developmental trajectories of body mass index throughout the life course: an application of Latent Class Growth (Mixture) Modelling", *Longitudinal Life Course Studies: International Journal*, 2, 3, 319-330.

Hoekstra T, Barbosa-Leiker C, Twisk J. (2013), "Vital exhaustion and markers of low-grade inflammation in healthy adults: the Amsterdam Growth and Health Longitudinal Study", *Stress Heal J Int Soc Investig Stress, 29, 5, 392–400*.

Kemper H., Hof M. van't. (1978), "Design of a multiple longitudinal study of growth and health in teenagers". *European Journal of pediatrics*, 155, 147–155.

Kolen M.J., Brennan, R. L. (2004), *Test Equating, Scaling, and Linking: Methods and Practices, second edition*. Springer-Verlag.

McCulloch, R., Polson, N. G. & Rossi, P. E. (2000), "A Bayesian analysis of the multinomial probit model with fully identified parameters", *Journal of Econometrics*, 99, 1, 173-193.

Millsap, R. E. (2000), "Testing Measurement Invariance Using Item Response Theory in Longitudinal Data: An Introduction", *Child Development Perspectives*, 4,1, 5-9.

Montgomery, D. C. (2004), *Design and Analysis of Experiments, sixth edition*, Chapman & Hall, London.

Rencher, R. C. (2002), *Methods of Multivariate Analysis, first edition*, Wiley Series, New York-NY.

Singer, J. M., & Andrade, D. F. (2000), "Analysis of longitudinal data", *in Handbook of Statistics, 18*, P. K. Sen and C. R. Rao eds, Elsevier, The Netherlands.

Tavares, H. R. & Andrade, D. F. (2006), "Item response theory for longitudinal data: Item and population ability parameters estimation", *Test*, 15, 1, 97-123.

te Marvelde, J.M., Glas, C. A. W., van Damme, J. (2006), "Application of Multidimensional Item Response Theory Models to Longitudinal Data", *Educational and Psychological Measurement*, 66, 1, 5-34.

Twisk J, Kemper H, van Mechelen W, Post G. (1997), "Tracking of Risk Factors for Coronary Heart Disease over a 14-Year Period: A Comparison between Lifestyle and Biologic Risk Factors with Data from the Amsterdam Growth and Health Study", *American Journal of Epidemiology. 145(10):888–898. doi:10.1093/oxfordjournals.aje.a009048*.

Twisk J, Staal B, Brinkman M, Kemper H, van Mechelen W. (1998), "Tracking of lung function parameters and the longitudinal relationship with lifestyle", *European Respiratory Journal, 12, 3, 627-624*.

van den Hout, A, Fox, J.-P., Entink, R. H. K. (2011), "Bayesian inference for an illnessdeath model for stroke with cognition as a latent time-dependent risk factor", *Statistical Methods in Medical Research*, 11, 1-19.

Von Davier, H., Xu, X. (2011), "Measuring growth in a longitudinal large-scale asessment with a general latent variable model", *Psychometrika*, 76, 2, 318-336.

Wijnstok N, Hoekstra T, Eringa E, Smulders Y, Twisk J, Serne E. (2012), "The relationship of body fatness and body fat distribution with microvascular recruitment: The Amsterdam Growth and Health Longitudinal Study", *Microcirculation, 19, 3, 273-279*.

Wijnstok N, Serné E, Hoekstra T, Schouten F, Smulders Y, Twisk J. (2013), "The relationship between 30-year developmental patterns of body fat and body fat distribution and its vascular properties: the Amsterdam Growth and Health Longitudinal Study", *Nutrition & Diabetes, 2013;3:e90. doi:10.1038/nutd.2013.31*.