

Bayesian longitudinal item response modeling with restricted covariance pattern structures

Caio L. N. Azevedo · Jean-Paul Fox · Dalton F. Andrade

Received: 6 August 2012 / Accepted: 25 September 2014 / Published online: 25 October 2014
© Springer Science+Business Media New York 2014

Abstract Educational studies are often focused on growth in student performance and background variables that can explain developmental differences across examinees. To study educational progress, a flexible latent variable model is required to model individual differences in growth given longitudinal item response data, while accounting for time-heterogenous dependencies between measurements of student performance. Therefore, an item response theory model, to measure time-specific latent traits, is extended to model growth using the latent variable technology. Following Muthén (Learn Individ Differ 10:73–101, 1998) and Azevedo et al. (Comput Stat Data Anal 56:4399–4412, 2012b), among others, the mean structure of the model represents developmental change in student achievement. Restricted covariance pattern models are proposed to model the variance–covariance structure of the student achievements. The main advantage of the extension is its ability to describe and explain the type of time-heterogenous dependency between student achievements. An efficient MCMC algorithm is given that can handle identification rules and restricted parametric covariance structures. A reparameterization technique is used, where unrestricted model parameters are sampled and transformed to obtain MCMC samples under the implied

restrictions. The study is motivated by a large-scale longitudinal research program of the Brazilian Federal government to improve the teaching quality and general structure of schools for primary education. It is shown that the growth in math achievement can be accurately measured when accounting for complex dependencies over grades using time-heterogenous covariances structures.

Keywords Longitudinal item response theory · Covariance patterns · Bayesian inference · MCMC

1 Introduction

Longitudinal item response data occur when students are assessed at several time points (Singer and Andrade 2000). This kind of data consist of response patterns of different examinees responding to different tests at different measurement occasions (e.g. grades). This leads to a complex dependence structure that arises from the fact that measurements from the same student are typically correlated (Tavares and Andrade 2006).

Various longitudinal item response theory (LIRT) models have been proposed to handle the correlation between measurements made over time. The popular mixed-effects regression modeling approach is often considered, where random effects are used to model the between-student and within-student dependencies. Conoway (1990) proposed a Rasch LIRT model to analyze panel data and proposed a marginal maximum likelihood method (Bock and Aitkin 1981) for parameter estimation. Liu and Hedeker (2006) developed a comparable three-level model to analyze LIRT data for ordinal response data. Eid (1996) defined a LIRT model for polytomous response data. Douglas (1999) analyzed longitudinal response data from a quality of life instrument using a joint

C. L. N. Azevedo (✉)
Department of Statistics, University of Campinas,
Campinas, São Paulo, Brazil
e-mail: cnaber@ime.unicamp.br

J.-P. Fox
Department of Research Methodology, University of Twente,
Enschede, The Netherlands

D. F. Andrade
Department of Informatics and Statistics, Federal University of Santa
Catarina, Florianópolis, Brazil

model, which consisted of proportional odds model and the graded item response model.

In these mixed effects models, the assumption of conditional independence is achieved by the random effects. That is, the assumption is made that the time-variant measurements are conditionally independent given the student's latent trait. The random effects imply a compound symmetry covariance structure, which assumes equal variances and covariances over time. In practice, the within-student latent trait dependencies are often not completely modeled and the errors are still correlated over time. Furthermore, in regression using repeated measurements it is common for the errors to show a time series structure (e.g., an autoregressive dependence) (Fitzmaurice et al. 2008; Hedeker and Gibbons 2006). If the dependence structure of the errors is not correctly specified, the parameter estimates and their standard errors will be biased.

Therefore, following the work of Jennrich and Schluchter (1986), Muthén (1998) and Tavares and Andrade (2006), restricted covariance pattern models are considered to model the time series structure of the errors. That is, the errors are allowed to correlate over time, and different variance-covariance structures are proposed to capture time-specific between-student variability and time-heterogeneous longitudinal dependencies between latent traits. An important aspect is that the covariance matrices considered allow for time-heterogeneous variances and covariances. The covariance pattern modeling framework is integrated in the LIRT modeling approach. At the student level, the time-specific latent traits are assumed to be multivariate normally distributed, and the within-student correlation structure is modeled using a covariance pattern model. This makes it possible to model the specific type of time-invariant and time-variant dependencies.

This modeling framework builds on the work of Tavares and Andrade (2006), who proposed a logistic three-parameter IRT model with a multivariate normal population distribution for the latent traits. They used a covariance matrix to model the within-examinee dependency, where the variances are allowed to vary over time but the covariance structure is assumed to be time-homoscedastic. The modeling framework also relates to the generalized linear latent trait model for longitudinal data of Dunson (2003), where the latent variable covariance structure is modeled via a linear transition model using observed predictors and an autoregressive component.

A full Gibbs sampling (FGS) algorithm is developed, which avoids the use of MCMC methods that require adaptive implementations, like the Metropolis–Hastings algorithm, to regulate the convergence of the algorithm. Furthermore, Sahu (2002) and Azevedo et al. (2012a) have shown that an FGS algorithm tends to perform better, in terms of parameter recovery, than a Metropolis–Hastings-

within-Gibbs sampling algorithm when dealing with IRT models. The proposed MCMC algorithm recovers all parameters properly and accommodates a wide range of variance-covariance structures. Using a parameter transformation method, MCMC samples are obtained of restricted parameters by transforming MCMC samples of unrestricted parameters. The proposed modeling framework is extended with various Bayesian model-fit assessment tools. Among other things, a Bayesian p-value is defined based on a suitable discrepancy measure for a global model-fit assessment and it is shown how Bayesian latent residuals can be used to evaluate the normality assumptions (Albert and Chib 1995).

This paper is outlined as follows. After introducing the Bayesian LIRT model, the FGS method is given, which can handle different variance-covariance structures. Then, the accuracy of the MCMC estimation method as well as the prior sensitivity are assessed. Subsequently, a real data study is presented, where the data set comes from a large-scale longitudinal study of children from the fourth to the eighth grade of different Brazilian public schools. One of the objects of the study is to analyze the student achievements across different grade levels. The model assessment tools are used to evaluate the fit of the model. In the last section, the results and some model extensions are discussed.

2 The model

A longitudinal test design is considered, where tests are administered to different examinees at different points in time. For each measurement occasion at time point t , $t = 1, \dots, T$, n_t examinees complete a test consisting of I_t items. The design can be typed as an incomplete block design such that common items are defined across tests and the total number of items equals $I \leq \sum_{t=1}^T I_t$. For a complete design, $n_t = n$, for all t . Dropouts and inclusion of students during the study are allowed.

The following notation will be introduced. Let θ_{jt} represent the latent trait of examinee j ($j = 1, \dots, n$) at time-point or measurement occasion t ($t = 1, \dots, T$), $\theta_j = (\theta_{j1}, \dots, \theta_{jT})^t$ the vector of the latent traits of the examinee j , and $\theta_{..} = (\theta_{.1}, \dots, \theta_{.n})^t$ the vector of all latent traits. Let Y_{ijt} represent the response of examinee j to item i ($i = 1, \dots, I$) in time-point t , $Y_{.jt} = (Y_{1jt}, \dots, Y_{I_tjt})^t$ the response vector of examinee j in time-point t , $Y_{..t} = (Y_{.1t}^t, \dots, Y_{.n_t t}^t)^t$ the response vector of all examinees in time-point t , $Y_{...} = (Y_{..1}^t, \dots, Y_{..T}^t)^t$ the entire response set. Let ζ_i denote the vector of parameters of item i , $\zeta = (\zeta_1^t, \dots, \zeta_I^t)^t$ the whole set of item parameters, and η_θ the vector with population parameters (related to the latent traits distribution).

A LIRT model is proposed that consists of two stages. At the first stage, a time-specific two-parameter IRT model is

considered for the measurement of the time-specific latent traits given observed dichotomous response data. The item-specific response probabilities are assumed to be independently distributed given the item and time-specific latent trait parameters. At the second stage, the subject-specific latent traits are assumed to be multivariate normally distributed with a time-heterogenous covariance structure, that is:

$$Y_{ijt} \mid (\theta_{jt}, \zeta_i) \sim \text{Bernoulli}(P_{ijt})$$

$$P_{ijt} = P(Y_{ijt} = 1 \mid \theta_{jt}, \zeta_i) = \Phi(a_i\theta_{jt} - b_i) \tag{1}$$

$$\theta_{j \cdot} \mid \eta_\theta \sim N_T(\mu_\theta, \Psi_\theta), \tag{2}$$

where η_θ consists on μ_θ and Ψ_θ and $\Phi(\cdot)$ stands for the cumulative normal distribution function. In this parametrization, the difficulty parameter $b_i = a_i b_i^*$ is a transformation of the original difficulty parameter denoted by b_i^* .

The within-subject dependencies among the time-specific latent traits are modeled using a T -dimensional normal distribution, denoted as $N_T(\mu_\theta, \Psi_\theta)$, with mean vector μ_θ and unstructured covariance matrix Ψ_θ , where

$$\mu_\theta = \begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \vdots \\ \mu_{\theta_T} \end{bmatrix} \text{ and } \Psi_\theta = \begin{bmatrix} \psi_{\theta_1} & \psi_{\theta_{12}} & \dots & \psi_{\theta_{1T}} \\ \psi_{\theta_{12}} & \psi_{\theta_2} & \dots & \psi_{\theta_{2T}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{1T}} & \psi_{\theta_{2T}} & \dots & \psi_{\theta_T} \end{bmatrix}, \tag{3}$$

respectively. A total of $\frac{T(T+1)}{2}$ parameters need to be estimated for the unstructured covariance model.

2.1 Restricted covariance pattern structures

In the LIRT model, the mean component of the multivariate population distribution for the latent traits can be extended to allow latent growth curves and include explanatory information. Besides modeling the mean structure, it is also possible to model the correlation structure between latent traits. Therefore, different covariance patterns are considered, which are restricted versions of the unrestricted covariance matrix (Eq. 3). Each restricted covariance pattern can address specific dependencies between the latent traits.

Several arguments can be given to explicitly model the covariance structure of the errors. First, the unstructured covariance model for the latent variables measured at different occasions will allow one parameter for every unique covariance term. There are no assumptions made about the nature of the residual correlation between the latent traits over time. However, for unbalanced data designs, small sample sizes with respect to the number of subjects and items, and many measurement occasions, the unstructured covariance model may lead to unstable covariance parameter estimates with large posterior variances (Hedeker and Gibbons 2006; Jennrich and Schluchter 1986). In practical test situations, the test length differs over occasions and students,

and the measurement error associated with the traits differ over students and measurement occasions (e.g., NELS, the national education longitudinal study; student monitoring systems for pupils in the Netherlands). As noted by Muthén (1998), the available longitudinal test data are of complex multivariate form, often involving different test forms, attrition, and students sampled hierarchically within schools.

Second, a fitted covariance pattern can provide insight in the residual correlation between latent measurements over time. When a covariance pattern is identified, information about the underlying growth process will be revealed. That is, on top of the change in latent traits modeled in the mean structure, the fitted covariance pattern can further explain the growth in latent traits. For example, when covariances of latent traits change over time, a fitted covariance pattern can be used to identify and describe the type of change.

Third, by correctly modeling the subject-specific correlated residuals across measurement occasions, more accurate statistical inferences can be made from the mean structure. Here, time-heteroscedastic covariance structures are considered to model complex patterns of residuals over time, where population variances of latent measurements can differ over time-points. A restricted covariance pattern model can lead to a decrease in model fit compared to the unstructured covariance pattern model, when the data do not fully support the restriction. When the restricted covariance model is supported by the data, it can lead to an improved fit compared to the unstructured model. Furthermore, due to the decrease in the number of model parameters, model selection criteria might prefer the restricted covariance model over the unstructured covariance model. Muthén (1998) and Jennrich and Schluchter (1986), among others, already noted that the efficiency of the mean structure parameters can be improved by modeling the covariance structure parsimoniously. For sample sizes and unbalanced data, the parameter estimates are most likely to be improved.

Muthén (1998) explained in more detail that the growth model consists of two components, a mean and a covariance structure component. Both components need to be modeled properly to describe the growth, and data interpretations depend on the specification of each component. Hedeker and Gibbons (2006) and Fitzmaurice et al. (2008) have shown that the analysis of longitudinal multivariate response data requires more complex covariance structures to capture the often complex dependency structures. Here, different covariance pattern models will be considered. For all cases, the sampling design is allowed to be unbalanced, where subjects can vary in the number of measurement occasions and response observations per measurement. The measurement times can vary over subjects and are not restricted to be equally spaced over subjects.

2.1.1 First-order heteroscedastic autoregressive model: ARH

This structure assumes that the correlations between subject’s latent traits decrease, when distances between instants of evaluation increase. However the magnitude of the correlations depend only on the distance between the time-points, not on their values. In addition, the variances are not assumed to follow any specific pattern. The form of the covariance matrix is given by,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta} & \dots & \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\rho_{\theta}^{T-1} \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta} & \psi_{\theta_2} & \dots & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\rho_{\theta}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\rho_{\theta}^{T-1} & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\rho_{\theta}^{T-2} & \dots & \psi_{\theta_T} \end{bmatrix}, \tag{4}$$

where $\psi_{\theta_t} \in (0, \infty)$, for $t = 1, \dots, T$, and $\rho_{\theta} \in (-1, 1)$. Note that the number of parameters for the ARH(1) is $T + 1$, which is much lower than the number of parameters for the unstructured covariance matrix. For more details, see Singer and Andrade (2000), Tavares and Andrade (2006), Andrade and Tavares (2005) and Fitzmaurice et al. (2008).

2.1.2 Heteroscedastic uniform model: HU

This is a special case of the ARH, which also assumes time-heterogenous variances over measurement occasions but time-homogenous correlations over time. So, the HU model assumes equal correlations between all pairs of time-specific latent trait measurements, independently of the distance between them. The heteroscedastic uniform covariance matrix is given by,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta} & \dots & \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\rho_{\theta} \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta} & \psi_{\theta_2} & \dots & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\rho_{\theta} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\rho_{\theta} & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\rho_{\theta} & \dots & \psi_{\theta_T} \end{bmatrix}, \tag{5}$$

again, $\psi_{\theta_t} \in (0, \infty)$, for $t = 1, \dots, T$, and $\rho_{\theta} \in (-1, 1)$. This covariance structure reduces the number of covariance parameters to one, while allowing T variance parameters, which resembles the total number of parameters used for the ARH. See Singer and Andrade (2000), Tavares and Andrade (2006), Andrade and Tavares (2005) and Fitzmaurice et al. (2008) for more details.

2.1.3 Heteroscedastic Toeplitz model: HT

As a special case of covariance pattern HU, the heteroscedastic Toeplitz model assumes a zero covariance between subject’s latent traits of two nonconsecutive instants. This might

be suitable when correlations decay quickly due to relatively large time-spaces between non-consecutive measurement occasions. This covariance pattern is represented by,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta} & 0 & \dots & 0 \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta} & \psi_{\theta_2} & \sqrt{\psi_{\theta_2}\psi_{\theta_3}}\rho_{\theta} & \dots & 0 \\ 0 & \sqrt{\psi_{\theta_2}\psi_{\theta_3}}\rho_{\theta} & \psi_{\theta_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \psi_{\theta_T} \end{bmatrix}. \tag{6}$$

In this case, $\psi_{\theta_t} \in (0, \infty)$ for $t = 1, \dots, T$, but $\rho_{\theta} \in (-k, k)$, where k depends on the value of T . For instance, $k \approx 1/\sqrt{2}$, for $T = 3$ and $k = 1/2$ for large T . For more details see Singer and Andrade (2000), Andrade and Tavares (2005), and Tavares and Andrade (2006).

2.1.4 Heteroscedastic covariance model: HC

The heteroscedastic covariance (HC) model is a restricted version of the heteroscedastic uniform model, where in the HC model, a common covariance is assumed across time points. As in the other covariance models, time-heterogenous variances are assumed. The HC model also corresponds to an unstructured covariance matrix with equal covariances. Note that the time-heterogeneous variances define time-heterogenous correlations, while assuming a common covariance term across time. Subsequently, relatively high time-specific latent trait variances will specify a low correlation between them. The HC covariance structure is represented by,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \rho_{\theta} & \dots & \rho_{\theta} \\ \rho_{\theta} & \psi_{\theta_2} & \dots & \rho_{\theta} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\theta} & \rho_{\theta} & \dots & \psi_{\theta_T} \end{bmatrix}, \tag{7}$$

where, $\psi_{\theta_t} \in (0, \infty)$ for $t = 1, \dots, T$, and $|\rho_{\theta}| < \min_{i,j} \sqrt{\psi_{\theta_i}\psi_{\theta_j}}$. Fore more details, see Andrade and Tavares (2005) and Tavares and Andrade (2006).

2.1.5 First-order autoregressive moving-average model: ARMAH

As the first-order autoregressive ARH structure, correlations between subject’s latent traits decrease as long as the distances between the instants of evaluation increase. However, the decrease is further parameterized due to the additional covariance parameter γ_{θ} . This covariance matrix, denoted as ARMAH, generalizes the ARH model, since it supports a

more flexible modeling of the time-specific correlations. The ARMAH covariance matrix is represented by,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\gamma_{\theta} & \dots & \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\gamma_{\theta}\rho_{\theta}^{T-2} \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\gamma_{\theta} & \psi_{\theta_2} & \dots & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\gamma_{\theta}\rho_{\theta}^{T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\gamma_{\theta}\rho_{\theta}^{T-2} & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\gamma_{\theta}\rho_{\theta}^{T-3} & \dots & \psi_{\theta_T} \end{bmatrix}. \tag{8}$$

In this case, $\psi_{\theta_t} \in (0, \infty)$ for $t = 1, \dots, T$, and $(\gamma_{\theta}, \rho_{\theta}) \in (-1, 1)^2$. For more details, see Singer and Andrade (2000) and Rochon (1992).

2.1.6 Ante-dependence model: AD

The last covariance structure model that will be considered is specifically useful when time points are not equally spaced and/or there is an additional source of variability present. This (first-order) AD model is a general covariance model that allows for changes in the correlation structure over time and unequally spaced measurement occasions.

The AD model generalizes the ARH using time-specific covariance parameters. It also generalizes the ARMAH model, since the covariance structure of the ARMAH is defined with $\rho_{\theta_1} = \gamma_{\theta}$ and $\rho_{\theta_2}, \dots, \rho_{\theta_T} = \rho_{\theta}$. The AD model supports a more dynamic modeling of the covariance pattern compared to ARH and ARMAH. The AD covariance model is represented by,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta_1} & \dots & \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\prod_{i=1}^{T-1}\rho_{\theta_i} \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta_1} & \psi_{\theta_2} & \dots & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\prod_{i=2}^{T-1}\rho_{\theta_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\prod_{i=1}^{T-1}\rho_{\theta_i} & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\prod_{i=2}^{T-1}\rho_{\theta_i} & \dots & \psi_{\theta_T} \end{bmatrix}, \tag{9}$$

where, $\psi_{\theta_t} \in (0, \infty)$ and $\rho_{\theta_t} \in (-1, 1)$, for $t = 1, 2, \dots, T - 1$. In conclusion, the AD model permits variances and correlations to change over time, and uses $2T - 1$ parameters. For more details, see Singer and Andrade (2000) and Nunez-Anton and Zimmerman (2000).

2.2 A restricted unstructured covariance structure

The latent variable framework will require a reference time-point to identify the latent scale. To accomplish that, the latent mean and variance of the first time-point will be fixed to zero and one, respectively. This way, the latent trait estimates across time will be estimated on one common scale, since an incomplete test design is used such that common items are administered at different measurement occasions (time-

points). The common items, also known as anchors, make it possible to measure the latent traits on one common scale.

However, the estimation of the covariance parameters is complicated due to the identifying constraints. Note that even for the unstructured covariance matrix, a restriction is implied on a variance parameter, which leads to a restricted unstructured covariance matrix. Furthermore, the restrictions on the parameters of the latent trait distribution also complicate the specification of priors. In this case, assuming an inverse-Wishart distribution for the unstructured covariance matrix is not possible, when the variance parameter of the first time-point is restricted to one.

In the present latent variable framework, a novel prior modeling approach will be followed to account for the restricted covariance structure. Following McCulloh 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,

$$\theta_j = (\theta_{j1}, \theta_{j2}, \dots, \theta_{jT})^t = (\theta_{j1}, \boldsymbol{\theta}_{j(1)})^t, \\ \boldsymbol{\mu}_{\theta} = (\mu_{\theta_1}, \mu_{\theta_2}, \dots, \mu_{\theta_T})^t = (\mu_{\theta_1}, \boldsymbol{\mu}_{\theta(1)})^t,$$

where, $\boldsymbol{\theta}_{j(1)} = (\theta_{j2}, \dots, \theta_{jT})^t$ and $\boldsymbol{\mu}_{\theta(1)} = (\mu_{\theta_2}, \dots, \mu_{\theta_T})^t$. In this notation, the index (1) indicates that the first component is excluded. It follows that the covariance structure, see definition in Eq. (3), is partitioned as,

$$\Psi_{\theta} = \begin{bmatrix} \psi_{\theta_1} & \boldsymbol{\psi}_{\theta(1)}^t \\ \boldsymbol{\psi}_{\theta(1)} & \boldsymbol{\Psi}_{\theta(1)} \end{bmatrix}, \tag{10}$$

where $\boldsymbol{\psi}_{\theta(1)} = (\psi_{\theta_{12}}, \dots, \psi_{\theta_{1T}})^t$ and

$$\boldsymbol{\Psi}_{\theta(1)} = \begin{bmatrix} \psi_{\theta_2} & \dots & \psi_{\theta_{2T}} \\ \vdots & \ddots & \vdots \\ \psi_{\theta_{2T}} & \dots & \psi_{\theta_T} \end{bmatrix}. \tag{11}$$

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

$$\theta_{j(1)}|\theta_{j1} \sim N_{(T-1)}(\boldsymbol{\mu}^*, \boldsymbol{\Psi}^*), \tag{12}$$

where

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_{\theta(1)} + \psi_{\theta_1}^{-1}\boldsymbol{\psi}_{\theta(1)}(\theta_{j1} - \mu_{\theta_1}),$$

and

$$\boldsymbol{\Psi}^* = \boldsymbol{\Psi}_{\theta(1)} - \psi_{\theta_1}^{-1}\boldsymbol{\psi}_{\theta(1)}\boldsymbol{\psi}_{\theta(1)}^t. \tag{13}$$

As a result, when conditioning on the restricted first-time point parameter, θ_{j1} , the remaining $\boldsymbol{\theta}_{j(1)}$ are conditionally multivariate normally distributed given θ_{j1} , with an unrestricted covariance matrix. The matrix $\boldsymbol{\Psi}^*$ 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).

For estimation purposes (using the restriction $\psi_{\theta_1} = 1$), it is convenient to eliminate the restricted parameter ψ_{θ_1} from the vector of covariances, $\boldsymbol{\psi}_{\theta(1)}$ (see the covariance structures in Equations from (4) to (9)) between the first component, θ_{j1} , and the remaining components $\theta_{j(1)}$. This new vector is denoted as $\boldsymbol{\psi}^*$ and is equal to

$$\boldsymbol{\psi}^* = \boldsymbol{\psi}_{\theta(1)} / \sqrt{\psi_{\theta_1}}. \tag{14}$$

Subsequently, the conditional distribution of the unrestricted latent variables is expressed as

$$\theta_{j(1)} | \theta_{j1} = \boldsymbol{\mu}_{\theta(1)} + \frac{\boldsymbol{\psi}^*}{\sqrt{\psi_{\theta_1}}} (\theta_{j1} - \mu_{\theta_1}) + \boldsymbol{\xi}_j. \tag{15}$$

where $\boldsymbol{\xi}_j \sim N(\mathbf{0}, \boldsymbol{\Psi}^*)$. The variance and correlation parameters,

$$\boldsymbol{\psi}^* \text{ and } \boldsymbol{\Psi}^*, \tag{16}$$

define an one-to-one relation with the free parameters of the original covariance matrix $\boldsymbol{\Psi}_{\theta}$, since the parameter ψ_{θ_1} is restricted to 1. As a result, the estimates of the population variances and covariances can be obtained from the estimates of Eq. (16). The latent variable distribution of the first measurement occasion will be restricted to identify the model. This is done by re-scaling the vector of latent variable values of the first measurement occasion to a pre-specified scale in each MCMC iteration. The latent variable population distribution of subsequent measurement occasions are conditionally specified according to Eq. (15), 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.

3 Bayesian inference and Gibbs sampling methods

The marginal posterior distributions comprise the main tool to perform Bayesian inference. Unfortunately, it is not possible to obtain closed-form expressions of the marginal posterior distributions. An MCMC algorithm will be used to obtain samples from the marginal posteriors, see [Gamerman and Lopes \(2006\)](#). More specifically, we will develop a FGS algorithm to estimate all parameters simultaneously.

MCMC methods for longitudinal and multivariate probit models have been developed by, among others, [Albert and Chib \(1993\)](#), [Chib and Greenberg \(1998\)](#), [Chib and Carlin \(1999\)](#), [Imai and Dyk \(2005\)](#), and [McCulloh et al. \(2000\)](#). A particular problem in Bayesian modeling of longitudinal multivariate response data is the prior specification for covariance matrices. An Inverse-Wishart prior distribution is plausible when covariance parameters are not functionally dependent, see [Tiao and Zellner \(1964\)](#). When this is

not the case, the prior specification of covariance parameters becomes much more complicated. Here, identification rules impose restrictions on the covariance parameters of the latent trait distribution. Therefore, the covariance structure is modeled by conditioning on the restricted parameters, which are related to the first measurement occasion. Following [McCulloh et al. \(2000\)](#), this approach supports a proper implementation of the identifying restrictions and a FGS implementation.

Conjugate prior distributions are considered, see [Gelman et al. \(2004\)](#) and [Gelman \(2006\)](#). According to the approach presented in Sect. 2, the parameters of interest are $(\boldsymbol{\mu}_{\theta}^t, \psi_{\theta_1}, \boldsymbol{\psi}^{*t})^t$ and $\boldsymbol{\Psi}^*$. Conjugate priors are specified as,

$$\boldsymbol{\mu}_{\theta} \sim N_T(\boldsymbol{\mu}_0, \boldsymbol{\Psi}_0), \tag{17}$$

$$\psi_{\theta_1} \sim IG(\nu_0, \kappa_0), \tag{18}$$

$$\boldsymbol{\psi}^* \sim N_{T-1}(\boldsymbol{\mu}_{\boldsymbol{\psi}}, \boldsymbol{\Psi}_{\boldsymbol{\psi}}), \tag{19}$$

$$\boldsymbol{\Psi}^* \sim IW_{T-1}(\nu_{\boldsymbol{\psi}}, \boldsymbol{\Psi}_{\boldsymbol{\psi}}), \tag{20}$$

where $IG(\nu_0, \kappa_0)$ stands for the inverse-gamma distribution with shape parameter ν_0 and scale parameter κ_0 , and $IW_{T-1}(\nu_{\boldsymbol{\psi}}, \boldsymbol{\Psi}_{\boldsymbol{\psi}})$ for the inverse-Wishart distribution with degrees of freedom $\nu_{\boldsymbol{\psi}}$ and dispersion matrix $\boldsymbol{\Psi}_{\boldsymbol{\psi}}$.

The prior for the item parameters is specified as

$$p(\boldsymbol{\xi}_i = (a_i, b_i) | \boldsymbol{\mu}_{\boldsymbol{\xi}}, \boldsymbol{\Psi}_{\boldsymbol{\xi}}) \propto \exp\left(-0.5 (\boldsymbol{\xi}_i - \boldsymbol{\mu}_{\boldsymbol{\xi}})^t \boldsymbol{\Psi}_{\boldsymbol{\xi}}^{-1} \times (\boldsymbol{\xi}_i - \boldsymbol{\mu}_{\boldsymbol{\xi}})\right) \mathbb{I}_{(a_i > 0)}, \tag{21}$$

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

In order to facilitate an FGS approach, and to account for missing response data, an augmented data scheme will be introduced, see [Albert \(1992\)](#) and [Albert and Chib \(1993\)](#). An augmented scheme is introduced to sample normally distributed latent response data $Z_{...} = (Z_{111}, \dots, Z_{I_T n T})^t$, given the discrete observed response data; that is,

$$Z_{ijt} | (\theta_{jt}, \boldsymbol{\xi}_i, Y_{ijt}) \sim N(a_i \theta_{jt} - b_i, 1), \tag{22}$$

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

To handle incomplete block designs, and indicator variable \mathbf{I} is defined that defines the set of administered items for each occasion and subject. This indicator variable is defined as follows,

$$I_{ijt} = \begin{cases} 1, & \text{item } i \text{ administered for examinee } j \text{ at time point } t \\ 0, & \text{missing by design.} \end{cases} \tag{23}$$

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,

$$V_{ijt} = \begin{cases} 1, & \text{observed response of examinee } j \text{ at time point } t \text{ on item } i \\ 0, & \text{otherwise,} \end{cases} \quad (24)$$

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.

To ease the notation, let indicator matrix \mathbf{I} represent both cases of missing data. Then, under the above assumptions, the distribution of augmented data \mathbf{Z}_{\dots} (conditioned on all other quantities) is given by

$$p(\mathbf{z}_{\dots} \mid \mathbf{y}_{\dots}, \boldsymbol{\zeta}, \boldsymbol{\theta}_{\dots}, \boldsymbol{\eta}_{\theta}, \mathbf{I}) \propto \prod_{t=1}^T \prod_{j=1}^n \prod_{i|I_{ijt}=1} \left\{ \exp \left\{ -0.5 (z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} \mathbb{I}_{(z_{ijt}, y_{ijt})} \right\}, \quad (25)$$

where $\mathbb{I}_{(z_{ijt}, y_{ijt})}$ represents the restriction that z_{ijt} is greater (lesser) than zero when y_{ijt} equals one (zero), according to Eq. (22).

Given the augmented data likelihood in Eq. (25) and the prior distributions in Eqs. (2), (17), (18), (19), (20) and (21), the joint posterior distribution is given by:

$$p(\boldsymbol{\theta}_{\dots}, \boldsymbol{\zeta}, \boldsymbol{\mu}_{\theta}, \psi_{\theta_1}, \boldsymbol{\Psi}^*, \boldsymbol{\Psi}_{\theta}^* \mid \mathbf{z}_{\dots}, \mathbf{y}_{\dots}) \propto p(\mathbf{z}_{\dots} \mid \boldsymbol{\theta}_{\dots}, \boldsymbol{\zeta}, \mathbf{y}_{\dots}) p(\boldsymbol{\theta}_{\dots} \mid \boldsymbol{\eta}_{\theta}) \times p(\boldsymbol{\zeta} \mid \boldsymbol{\mu}_{\zeta}, \boldsymbol{\Psi}_{\zeta}) p(\boldsymbol{\eta}_{\theta}). \quad (26)$$

where

$$p(\boldsymbol{\theta}_{\dots} \mid \boldsymbol{\eta}_{\theta}) = \prod_{j=1}^n p(\boldsymbol{\theta}_j \mid \boldsymbol{\eta}_{\theta}), \quad (27)$$

and

$$p(\boldsymbol{\eta}_{\theta}) = p(\boldsymbol{\mu}_{\theta}) p(\psi_{\theta_1}) p(\boldsymbol{\Psi}^*) p(\boldsymbol{\Psi}_{\theta}^*).$$

This posterior distribution (26) has an intractable form but, as shown in the Appendix, the full conditionals are known and easy to sample from. Let (\cdot) denote the set of all necessary parameters. The FGS algorithm is defined as follows:

1. Start the algorithm by choosing suitable initial values. Repeat steps 2–10.
2. Simulate Z_{ijt} from $Z_{ijt} \mid (\cdot), i = 1, \dots, I_t, j = 1, \dots, n, t = 1, \dots, T$.
3. Simulate θ_j from $\theta_j \mid (\cdot), j = 1, \dots, n$.
4. Simulate ζ_i from $\zeta_i \mid (\cdot), i = 1, \dots, I$.
5. Simulate $\boldsymbol{\mu}_{\theta}$ from $\boldsymbol{\mu}_{\theta} \mid (\cdot)$.
6. Simulate ψ_{θ_1} from $\psi_{\theta_1} \mid (\cdot)$.

7. Simulate $\boldsymbol{\Psi}^*$ from $\boldsymbol{\Psi}^* \mid (\cdot)$.
8. Simulate $\boldsymbol{\Psi}_{\theta}^*$ from $\boldsymbol{\Psi}_{\theta}^* \mid (\cdot)$.
9. Compute the unstructured covariance matrix using the sampled covariance components from Steps 6–8 and Eqs. (10), (13) and (14).
10. Through a parameter transformation method using sampled unstructured covariance parameters, compute restricted covariance components of interest. The sampled restricted covariance structure $\boldsymbol{\Psi}_{\theta}$ is used when repeating steps 2–8.

To handle the restriction $\mu_{\theta_1} = 0$, the expression in Eq. (12) is used to simulate $\boldsymbol{\mu}_{\theta(1)}$. To simulate $(\mu_{\theta_1}, \psi_{\theta_1})^t$, the following decomposition is used in (27),

$$p(\boldsymbol{\theta}_j \mid \boldsymbol{\eta}_{\theta}) = p(\boldsymbol{\theta}_{j(1)} \mid \boldsymbol{\eta}_{\theta}, \theta_{j1}) p(\theta_{j1} \mid \boldsymbol{\eta}_{\theta_1}),$$

where $\boldsymbol{\eta}_{\theta_1} = (\mu_{\theta_1}, \psi_{\theta_1})^t$. To identify the model, the scale of the latent variable of measurement occasion one is transformed to mean zero and variance one. It is also possible to restrict the parameters $(\mu_{\theta_1}, \psi_{\theta_1})^t$ to specific values.

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

$$\begin{aligned} \det(\boldsymbol{\Psi}_{\theta}) &= \det(\psi_{\theta_1}) \det \left(\boldsymbol{\Psi}_{\theta(1)} - \psi_{\theta_1}^{-1} \boldsymbol{\Psi}_{\theta(1)} \boldsymbol{\Psi}_{\theta(1)}^t \right) \\ &= \psi_{\theta_1} \det(\boldsymbol{\Psi}^*), \end{aligned}$$

using Eqs. (10) and (13) and a property of the determinant of block matrices. As a result, the $\det(\boldsymbol{\Psi}_{\theta})$ is greater than zero since both the determinant of $\boldsymbol{\Psi}^*$ and ψ_{θ_1} are greater than zero. This implies positive definite samples of $\boldsymbol{\Psi}_{\theta}$.

In MCMC Step 10, parameters of a posited covariance pattern structure are computed given an MCMC sample of the unrestricted unstructured covariance parameters. Each simulated covariance matrix will be positive definite, since it is based on a positive definite unstructured covariance matrix. In the Appendix, the reparameterization for each covariance structure is specified, which facilitates the sampling of the parameters of the restricted covariance matrices. That is, in each MCMC iteration, parameters of a specific covariance pattern are computed using sampled unstructured covariance parameters. This procedure is based on the notion that each restricted covariance pattern is nested in the most general unstructured pattern, and that in the MCMC procedure parameter transformations can be used to achieve draws from the transformed parameter distribution.

4 Selection of covariance structure

Accurate inferences are obtained when selecting the most appropriate covariance pattern. Selecting a too simple covari-

ance pattern can lead to underestimated standard errors and biased parameter estimates, see [Singer and Andrade \(2000\)](#) and [Singer and Andrade \(1994\)](#). Selecting a too complex covariance pattern can lead to a decrease in power and efficiency. The general method to select an appropriate covariance structure is based on some Bayesian optimality criterion. The different covariance structures are viewed as competing covariance models and the one that optimizes the Bayesian model criterion is selected. Attention is focused on three criteria for model selection, which are widely used in the literature; the deviance information criterion (DIC), posterior expectation of the Aikake’s information criterion (AIC) and the posterior expectation of the Bayesian information criterion (BIC), see [Spiegelhalter et al. \(2002\)](#). For each criterion, the covariance pattern is selected with the smallest criterion value. All competing covariance structures are time heterogenous, which generalizes the work of [Andrade and Tavares \(2005\)](#) and [Tavares and Andrade \(2006\)](#).

The general form of the different information criteria is the deviance (i.e. minus two times the log-likelihood) plus a penalty term for model complexity, which includes the number of model parameters. Let ϑ denote the set of relevant parameters, that is, the latent traits, the item and the population parameters, then the following deviance is considered to define the model selection criteria,

$$\begin{aligned}
 D_1(\vartheta) &= -2 \left[\log p(y \mid \theta, \zeta) + \log p(\theta \mid \mu_\theta, \Psi_\theta) \right] \\
 &= -2 \left(\sum_{t=1}^T \sum_{j=1}^n \sum_{i|I_{ijt}=1} \log P(Y_{ijt} = y_{ijt} \mid \theta_{jt}, \zeta_i) \right. \\
 &\quad \left. + \sum_{j=1}^n \log p(\theta_j \mid \mu_\theta, \Psi_\theta) \right) \\
 &= -2(LL + LLLT),
 \end{aligned}$$

where $p(\theta_j \mid \mu_\theta, \Psi_\theta)$ represents the density of the multivariate normal distribution, $LL = \sum_{t=1}^T \sum_{j=1}^n \sum_{i|I_{ijt}=1} \log P(Y_{ijt} = y_{ijt} \mid \theta_{jt}, \zeta_i)$ and $LLL T = \sum_{j=1}^n \log p(\theta_j \mid \mu_\theta, \Psi_\theta)$.

The deviance depends highly on the estimated latent traits. The covariance structure will influence the latent trait estimates, although they are mostly influenced by the data. The terms LL and $LLL T$ both emphasize the fit of the latent traits, and will diminish the importance of the fit of the covariance structure. Therefore, the deviance term $D_2(\vartheta) = LL$ is also considered to evaluate the fit of the covariance structure by evaluating the fit of the latent traits in the likelihood term.

Let $\overline{D_i(\vartheta)}$ denote the posterior mean deviance and $D_i(\hat{\vartheta})$ the deviance at the posterior mean. Then, the DIC is defined as,

$$DIC_i = 2\overline{D_i(\vartheta)} - D_i(\hat{\vartheta}), i = 1, 2,$$

where the penalty function for model complexity is determined by an estimate of the effective number of model parameters, which allows nonzero covariance among model parameters.

For the AIC and the BIC, the penalty function for model complexity is determined by the effective number of parameters in the model, which is difficult or impossible to ascertain when random effects are involved. Following [Spiegelhalter et al. \(2002\)](#) and [Congdon \(2003\)](#), the posterior mean deviance is used as a penalized fit measure, which includes a measure of complexity. Then, the following specification is made for the AIC and the BIC,

$$AIC_i = \overline{D_i(\vartheta)} + 2(2(I - 1) + (T + n_{\Psi_\theta})),$$

and

$$BIC_i = \overline{D_i(\vartheta)} + (2(I - 1) + (T + n_{\Psi_\theta})) \ln(n^*),$$

$i = 1, 2$, respectively, where n_{Ψ_θ} is the total number of covariance parameters, T is the number of time points and $n^* = \sum_{j=1}^n \sum_{t=1}^T \sum_{i=1}^{I_{ijt}} V_{ijt}$.

The AIC and BIC results are not guaranteed to lead to the same model, see [Spiegelhalter et al. \(2002\)](#) and [Ando \(2007\)](#). The BIC has a much higher penalty term for model complexity than the AIC. Therefore, a relatively more concise description of the covariance structure can be expected from the BIC. When different results of the two criteria are obtained, the model selected by the BIC is preferred over the one selected by the AIC, see [Spiegelhalter et al. \(2002\)](#) and [Ando \(2007\)](#).

The deviance can be approximated using the MCMC output, and using G MCMC iterations the posterior mean of the deviance is estimated by

$$\overline{D_i(\vartheta)} = \frac{1}{G} \sum_{g=1}^G D_i(\vartheta^g),$$

and the deviance at the posterior mean by,

$$D_i(\hat{\vartheta}) = D_i \left(\frac{1}{G} \sum_{g=1}^G \vartheta^g \right),$$

with index g representing the g -th value of the valid MCMC sample (considering the burn-in and the thin value).

Here, the selection of the most optimal covariance structure is carried out using Bayesian measures of model complexity as in [Spiegelhalter et al. \(2002\)](#). It is also possible to use pseudo-Bayes factors as in [Kass and Raftery \(1995\)](#), or reversible Jump MCMC algorithms, see [Green \(1995\)](#) and [Azevedo \(2008\)](#), which would require a different computational implementation.

4.1 Model assessment: posterior predictive checks

Besides using model selection criteria for selecting the covariance structure, the fit of the general LIRT model can be evaluated using Bayesian posterior predictive tests and Bayesian residual analysis techniques (Albert and Chib 1995). The literature about posterior predictive checks for Bayesian item response models shows several diagnostics for evaluating the model fit. A general discussion can be found in, among others, Stern and Sinharay (2005), Sinharay (2006), Sinharay et al. (2006), and Fox (2004, 2005, 2010).

The common posterior predictive tests can be generalized to make them applicable for the LIRT model. Each posterior predictive test is based on a discrepancy measure, where this discrepancy measure is defined in such a way that a specific assumption or general fit of the model can be evaluated. The main idea is to generalize the well known discrepancy measures to a longitudinal structure.

In general, let y^{obs} be the matrix of observed responses, and y^{rep} the matrix of replicated responses generated from its posterior predictive distribution. The posterior predictive distribution of the response data of time-point t is represented by

$$p(y_t^{rep} | y_t^{obs}) = \int p(y_t^{rep} | \theta_t) p(\theta_t | y_t^{obs}) d\theta_t,$$

where θ_t denotes the set of model parameters corresponding time-point t . Generally, given a discrepancy measure $D(y_t, \theta_t)$, the replicated data are used to evaluate whether the discrepancy value given the observed data is typical under the model. A p-value can be defined that quantifies the extremeness of the observed discrepancy value in time-point t ,

$$p_0(y_t^{(obs)}) = P(D(y_t^{(rep)}, \theta_t) \geq D(y_t^{(obs)}, \theta_t) | y_t^{(obs)}),$$

where the probability is taken over the joint posterior of $(y_t^{(rep)}, \theta_t)$. In some cases, the discrepancy measure can be generalized from the time-point level to the population level. In that case, the discrepancy measure can be used to evaluate model fit at the time-point and population level.

Here, p-values based on a chi-square distance, predictive distributions of latent scores, and Bayesian latent residuals are considered (Fox 2004, 2010; Azevedo et al. 2011). The chi-square posterior predictive check is defined to evaluate the predictive score distribution with the observed score distribution. The discrepancy measure for evaluating the score distribution is defined as,

$$D(y_t) = \sum_l \frac{(n_{l,t} - E(n_{l,t}))^2}{V(n_{l,t})},$$

where $n_{l,t}$ is the number of subjects with a score l at measurement occasion t , and $E(\cdot)$ and $V(\cdot)$ stand for the expectation

and the variance, respectively. The posterior predictive check based on the score distribution is evaluated using MCMC output.

The predictive score distribution is easily calculated using the MCMC output. In each iteration, a sample of the score distribution is obtained. This is accomplished by generating response data for the sampled parameters. Subsequently, the number of subjects can be calculated for each possible score at each time-point. For each possible score, the median and 95 % credible interval is calculated to evaluate the score distribution.

A general approach for model adequacy assessment using Bayesian (latent) residuals is described by Albert and Chib (1995). Here, Bayesian residuals are analyzed for the latent traits at each time-point. The following quantity is considered,

$$\frac{\hat{\theta}_j - \hat{\mu}_{\theta_t}}{\sqrt{\hat{\psi}_{\theta_t}}},$$

for $t = 1, 2, \dots, T$, using posterior mean estimates. Subsequently, the normality assumption is evaluated using box and/or Q-Q plots.

5 Simulation study

Convergence properties and parameter recovery were analyzed using simulated data. The following hyperparameter settings were used in the simulation study:

$$\mu_\psi = \mathbf{0}_{T-1}, \Psi_\psi = \tau \mathbf{I}_{T-1} \tag{28}$$

$$\Psi_\psi = (\nu_\psi - T + 1) (\mathbf{I} - \Psi_\psi), \tag{29}$$

where $\nu_\psi = 5$, $\tau = 1/8$ and the hyperparameters for the item parameters were specified as: $\mu_\zeta = (1, 0)^\top$ and $\Psi_\zeta = \text{diag}(0, 5, 3)$.

Responses of $n = 1,000$ examinees were simulated for three measurement occasions. At each occasion, data were simulated according to a test of 24 items. There were six common items between test one and two, and six between test two and three. The item parameter values vary in terms of discrimination power and difficulty, properly. For each examinee, a total of 60 items were administered.

Examinees' latent traits were generated from a three-variate normal distribution with $\mu_\theta = (0, 1, 2)^t$. The within-subject latent traits were correlated according to an ARH covariance structure, where $\psi_\theta = (1, 0.9, 0.95)^t$ and $\rho_\theta = 0.75$. This implies latent growth in the mean structure, weak heterogeneous latent trait variance across time, and a strong within-subject correlation over time.

5.1 Convergence and autocorrelation assessment

Following [Gamerman and Lopes \(2006\)](#), the convergence of the MCMC algorithm was investigated by monitoring trace plots generated by three different sets of starting values, and by evaluating Geweke's and Gelman and Rubin's convergence diagnostics.

Following [DeMars \(2003\)](#), the sampled latent traits were transformed to the scale of the simulated latent traits according to

$$\theta_{j.}^{**} = \text{Chol}(\Psi_{\theta}) \text{Chol}(S_{\theta})^{-1} (\theta_{j.}^* - \bar{\theta}) + \mu_{\theta},$$

where $\theta_{j.}^*$ are the simulated latent traits, $\bar{\theta}$ and S_{θ} are the sample mean vector and covariance matrix, respectively, and *Chol* stands for the Cholesky decomposition.

Figure 1 represents trace plots of latent trait population parameters for occasions two and three. The population parameters of time point one were fixed for identification. Figure 2 represents trace plots of parameters of two randomly selected items. Sampled values were

stored every 30th iteration. The MCMC sample composed by storing every 30th value showed negligible autocorrelation. Posterior density plots (not shown) using the sampled values showed that symmetric behavior of the posteriors, which support the posterior mean as a Bayesian point estimate.

In each plot, three different chains are plotted, which correspond to three different initial values. From a visual inspection it can be concluded that within 100 (thinned) iterations each chain of simulated values reached the same area of plausible parameter values. Each MCMC chain mixed very well, which indicates that the entire area of the parameter space was easily reached. The Geweke diagnostic, based on a burn-in period of 16,000 iterations, indicated convergence of the chains of all model parameters. Furthermore, the Gelman–Rubin diagnostic were close to one, for all parameters. Convergence was established easily without requiring informative initial parameter values or long burn-in periods. Therefore, the burn-in was set to be 16,000, and a total of 46,000 values were simulated, and samples were collected at a spacing of 30 iterations.

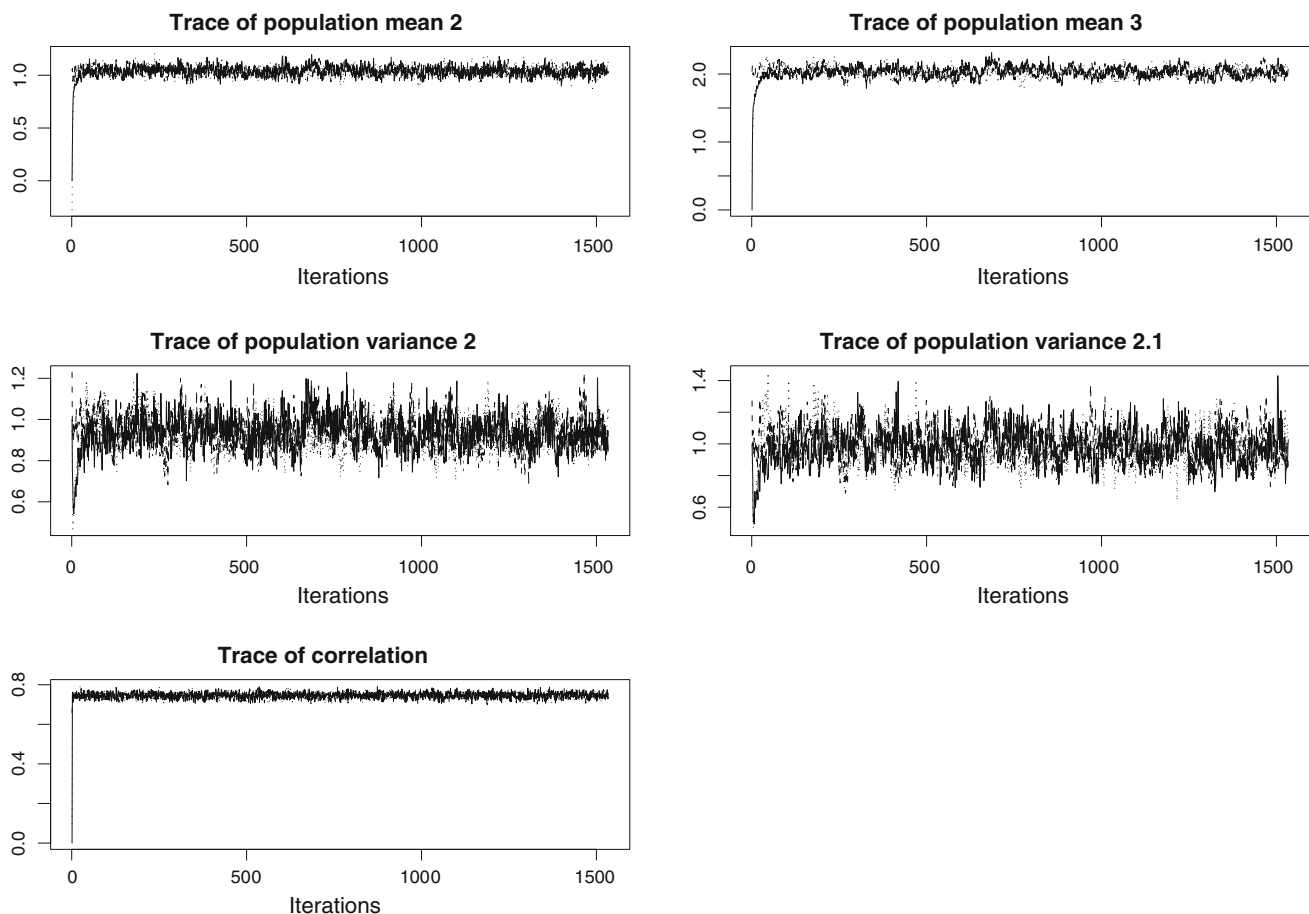


Fig. 1 For different starting values, trace plots of the simulated values of the population parameters

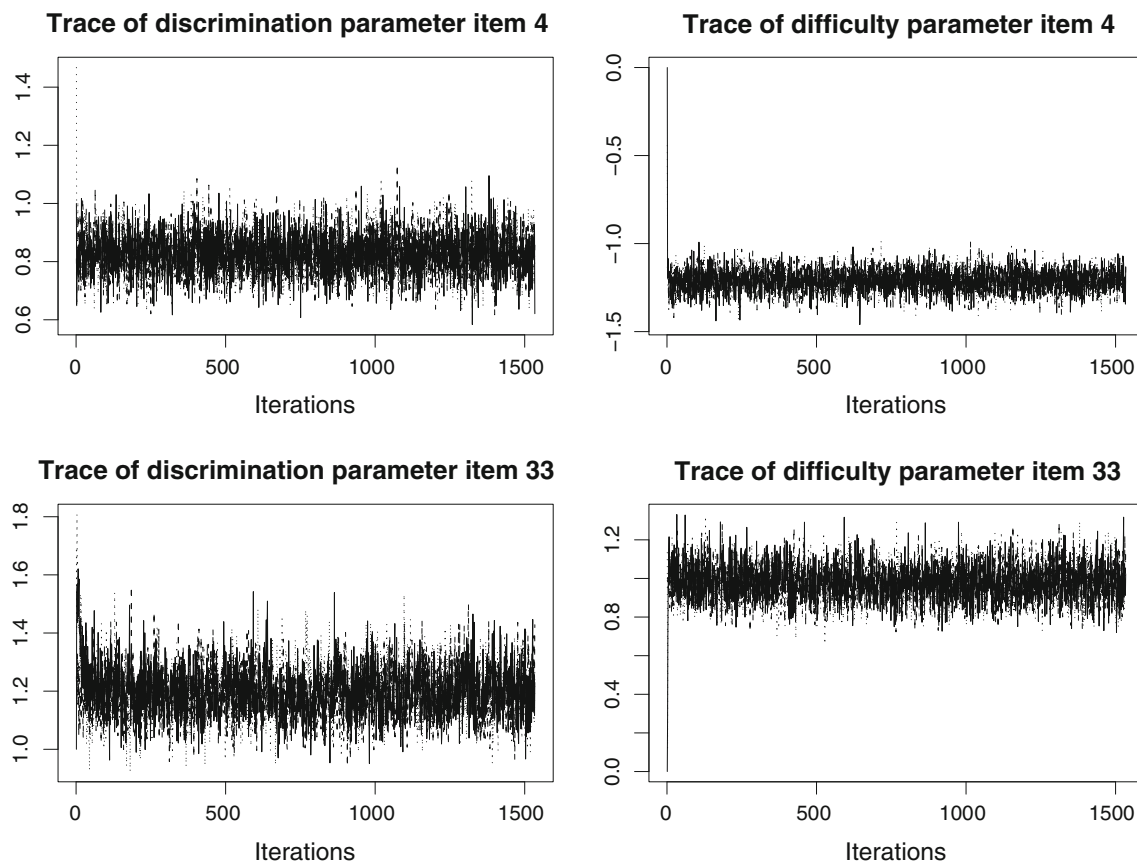


Fig. 2 For different starting values, trace plots of the simulated values for parameters of item 4 and 33

5.2 Parameter recovery

The linked test design contains 60 items such that 120 item parameters need to be estimated and 3,000 person parameters. The general population model for the person parameters leads to an additional set of five parameters, since two population parameters were restricted. [Ayala and Sava-Bolesta \(1999\)](#) suggest to consider around 1,200 subjects per item to obtain accurate parameter estimates. Here, 1,000 responses per item were simulated since the specification of a correct prior structure of the LIRT becomes more important when less data are available. Furthermore, the characteristics of the real data study described further on, will resemble those of the simulated data study.

Different statistics were used to compare the results: mean of the estimates (M. Est.), correlation (Corr), mean of the standard error (MSE), variance (VAR), the absolute bias (ABias) and the root mean squared error (RMSE). To evaluate the accuracy of the MCMC estimates, a total of ten replicated data sets were generated, which was based on [Azevedo and Andrade \(2010\)](#) and [Ayala and Sava-Bolesta \(1999\)](#). For the item and latent trait parameters, average statistics were computed by averaging across data sets, and items and persons, respectively.

Table 1 Replication study: results for the estimated latent trait and item parameters

Parameter	Statistic				
	Corr	MSE	ABias	Var	RMSE
Latent trait	.993	.086	.114	.061	.278
Discrimination	.982	.011	.028	.010	.097
Difficulty	.999	.023	.042	.016	.122

Table 1 represents the results for the latent traits and item parameters. The estimated values of the statistics indicate that the MCMC algorithm recovered all parameters properly. Furthermore, the estimated posterior means of the discrimination and difficulty parameters were also close to the true values. Similar conclusions can be drawn about the estimates of the latent trait population parameters, see Table 2. The estimated posterior means are close to the true values, and the biases are relatively small.

5.3 Covariance structure selection

The information criteria were used to compare the fit of the different covariance models. The results are given in Table 3, which includes the information criteria for model

Table 2 Replication study: results for the estimated latent trait population parameters

Parameter	Statistics				
	M. est.	MSE	ABias	Var	RMSE
μ_{θ_2}	.968	.003	.032	.002	.053
μ_{θ_3}	1.968	.009	.032	.008	.094
ψ_{θ_2}	.864	.006	.036	.005	.078
ψ_{θ_3}	.948	.009	.002	.009	.092
ρ_{θ}	.749	<.001	.001	<.001	.007

comparison, as presented in Sect. 4. The information criteria results for the heteroscedastic toeplitz (HT) model were much higher in comparison to the other covariance models, since the dependency structure was restricted to correlations between two adjacent time measurements. Therefore, to avoid distraction these results were not included in Table 3.

From Table 3 it follows that the DIC_2 selects the true covariance model (ARH), where the AIC_2 and BIC_2 select the UH structure. However, note that the ARH model was ranked second by these two criteria. It is not surprising that the BIC_2 selects the UH above the ARH, since the DIC tends to prefer simpler models. However, the AIC_2 , which tends to select more complex models, also selected the UH model, even though the difference from the related statistic for the ARH model is quite small. This behavior could be caused by sampling fluctuation.

As expected, the results of the AIC_1 , BIC_1 and DIC_1 were inflated by the values of the LLLT, by emphasizing the fit of the latent trait estimates. The quantification of the fit of each particular covariance structure is not well represented by these criteria, since the latent trait estimates dominate the deviance term. Although the results show some consistency when considering the DIC_2 , a more thorough study is necessary, which is beyond the scope of the present study.

6 The Brazilian school development study

The data set analyzed stems from a major study initiated by the Brazilian Federal Government known as the School

Development Program. The aim of the program is to improve the teaching quality and the general structure (classrooms, libraries, laboratory informatics etc) in Brazilian public schools. A total of 400 schools in different Brazilian states joined the program. Achievements in mathematics and Portuguese language were measured over five years (from fourth to eight grade of primary school) from students of schools selected and not selected for the program.

The 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.

In the present study, mathematic performances 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 test 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.

In an exploratory analysis, the multiple group model (MGM), described in Azevedo et al. (2011), was used to estimate the latent student achievements given the response data. The MGM for cross-sectional data assumes that students are nested in groups and latent traits are assumed to be independent given the mean level of the group. Typical

Table 3 Selecting the optimal covariance structure for the real data set: estimated Bayesian information criteria. Bold values indicates models chosen by the statistics

Model	AIC_1	BIC_1	DIC_1	LL	LLLT	AIC_2	BIC_2	DIC_2
HU	61,696	62,844	64,610	-25,593	-3,547	54,102	55,250	56,518
ARH	61,000	62,148	63,920	-25,636	-3,154	54,113	55,261	56,453
ARMAH	61,076	62,234	63,996	-25,631	-3,195	54,114	55,271	56,461
HC	59,275	60,423	62,326	-26,442	-1,420	55,359	56,507	57,334
AD	60,834	61,991	63,759	-25,641	-3,061	54,122	55,279	56,457
Unst.	60,719	61,885	63,637	-25,652	-2,995	54,135	55,302	56,459

Table 4 Estimated posterior variances, covariances, and correlations among estimated latent traits are given in the diagonal, lower and upper triangle, respectively

	Grade four	Grade five	Grade six
Grade four	1.000	.723	.629
Grade five	.659	1.152	.681
Grade six	.540	.641	1.071

Table 5 Selecting the optimal covariance structure for the real data set: estimated Bayesian information criteria. Bold values indicates models chosen by the statistics

Model	LL	AIC_2	BIC_2	DIC_2
HU	-71,980	147,477	148,941	150,398
ARH	-72,164	147,693	149,157	150,462
ARMAH	-72,179	147,727	149,201	150,496
HC	-72,840	148,707	150,171	151,139
AD	-72,184	147,723	149,196	150,477
Unst.	-71,984	147,470	148,954	150,368

for the longitudinal nature of the study, a positive correlation among latent traits from the same examinee is to be expected, but this aspect was ignored in this explanatory analysis. Pearson’s correlations, variances, and covariances were estimated among the vectors of estimated latent traits corresponding to grade four to six. The estimates are represented in Table 4.

The results show significant between-grade dependencies. That is, the latent traits are not conditionally independently distributed over grades given the grade-specific means. The estimated variances increased after grade four, which indicates the presence of time-heterogenous variances. Furthermore, given the estimates of covariances, time-heteroscedastic covariances and time-decreasing correlations are to be considered to account for within-subject (between-grade) dependencies among latent traits. Therefore, the LIRT model was estimated using each one of the covariance structures to account for the specific dependencies.

The response data were modeled according to the LIRT model using different covariance structures. First, attention was focused on selecting the optimal covariance structure. Second, a more detailed model fit assessment was carried out using the selected covariance structure. The three model selection criteria were used to identify the most suitable covariance structure. As in the simulation study, the heteroscedastic toeplitz model did not fit the data and produced much higher information criteria estimates. For each other covariance structure, Table 5 represents the estimated values for the AIC_i , BIC_i , and DIC_i , $i = 1, 2$. The information criteria are represented such that a smaller value corresponds to a better model fit.

The AIC_2 and DIC_2 preferred the unstructured covariance model, where the BIC_2 preferred the more parsimonious HU. However, the unstructured model was ranked sec-

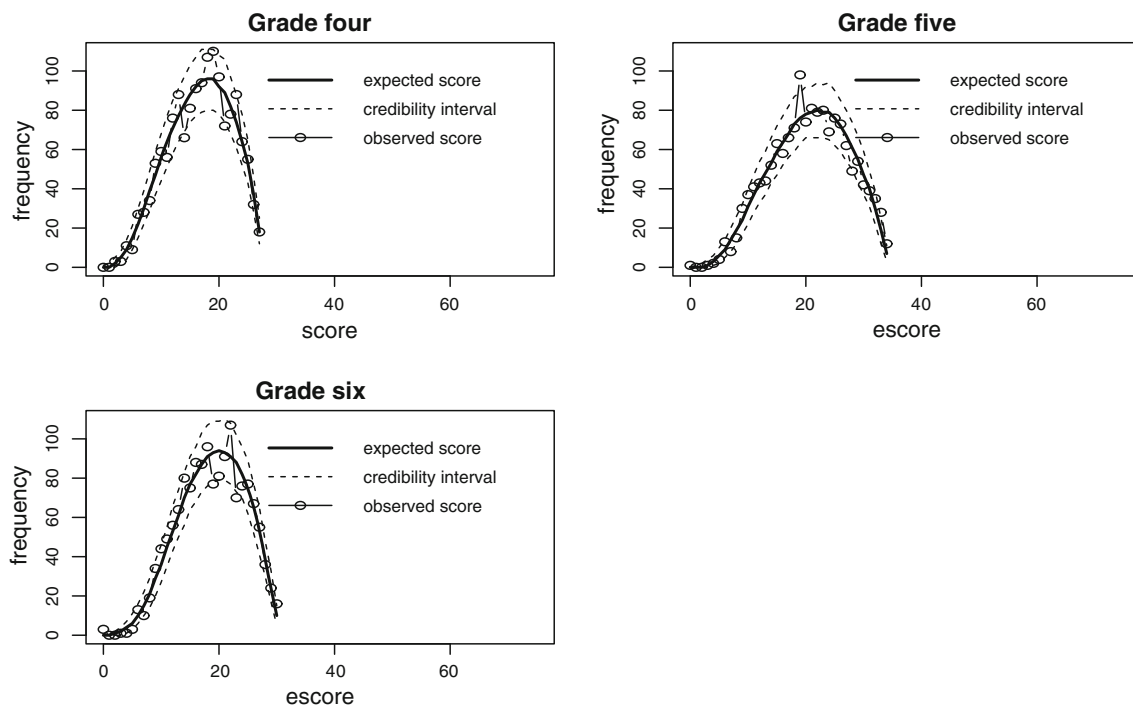


Fig. 3 Observed score distribution, predicted score distribution, and 95 % central credible intervals

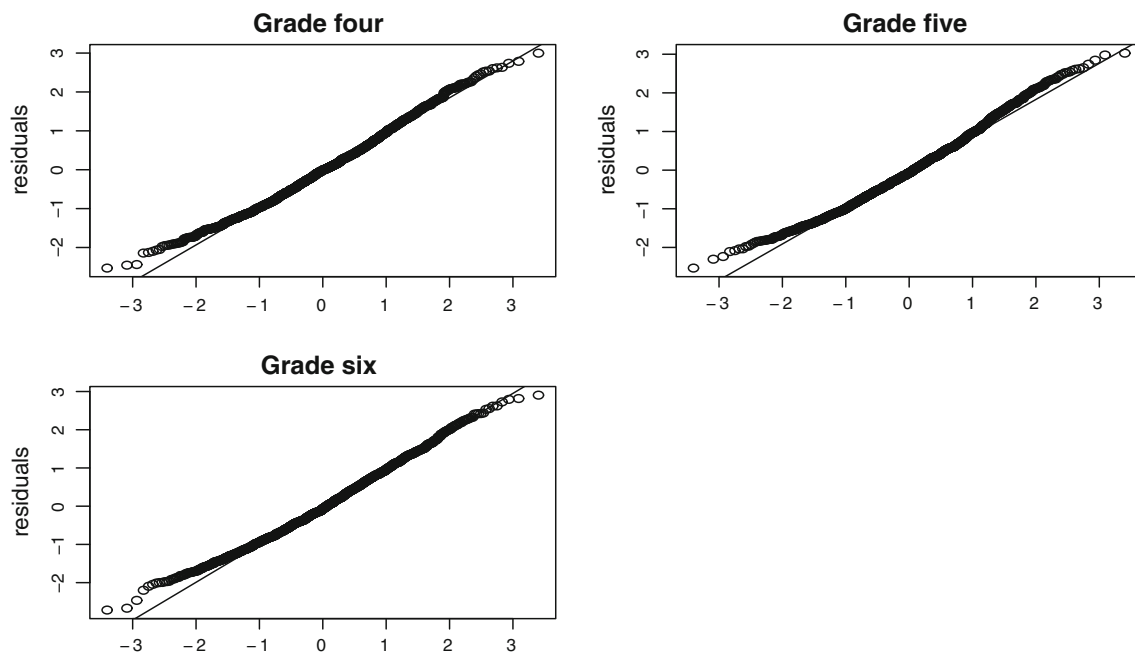


Fig. 4 For each grade, Quantile-Quantile plot of estimated latent trait residuals

ond by BIC_2 , whereas the UH was second ranked by AIC_2 and DIC_2 . There were only three measurement occasions, and the correlations between grade years were high. This made the comparison between the unstructured and UH difficult. In the presence of high between grade correlations and a few time points, the information criteria results preferred the UH (the most parsimonious model) and the unstructured covariance matrix. From the various competing covariance structures, the unstructured covariance matrix was used for further analysis.

Different model fit assessment tools, based on posterior predictive densities of different quantities were used to evaluate the LIRT model with the ARMAH covariance structure. The p-value based on a chi-squared distance, and predictive distributions of latent scores and Bayesian latent residuals were considered, see for more details about the posterior checks Albert and Chib (1995), Azevedo (2008), and Azevedo et al. (2011).

The Bayesian p-value was $p = .398$, which indicates that the model fitted well. In addition, the observed scores fall almost all within the credible intervals for each grade, except for observed scores equal to 20 in grade five, see Fig. 3. Figure 4 represents an estimated quantile-quantile plot of the latent trait residuals of each grade. In general, from visual inspection follows that the assumed normal probability distribution in each grade seems to be appropriate.

Table 6 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

Table 6 Population parameter estimates and 95 % credible intervals

Grade	Mean	SD	HPD 95%
<i>Mean</i>			
Four (reference)	0	–	–
Five	.240	.040	[.170, .319]
Six	.763	.048	[.680, .862]
<i>Variance</i>			
Four (reference)	1	–	–
Five	1.032	.081	[.876, 1.183]
Six	.969	.087	[.794, 1.131]
<i>Correlations</i>			
Four and five	.857	.012	[.832, .879]
Four and six	.759	.017	[.724, .790]
Five and six	.810	.015	[.784, .840]

trait means was detected given the non-overlapping credible intervals. As expected, the mean growth of math achievement over grade years is significant. The within-grade variability is relatively small, but the between-grade correlations are 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.

Finally, Figs. 5 and 6 represent the posterior means and 95 % credible intervals of the item discrimination and original difficulty estimates (b_i^*), respectively. The discrimination parameter estimates are relatively low, where approx-

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

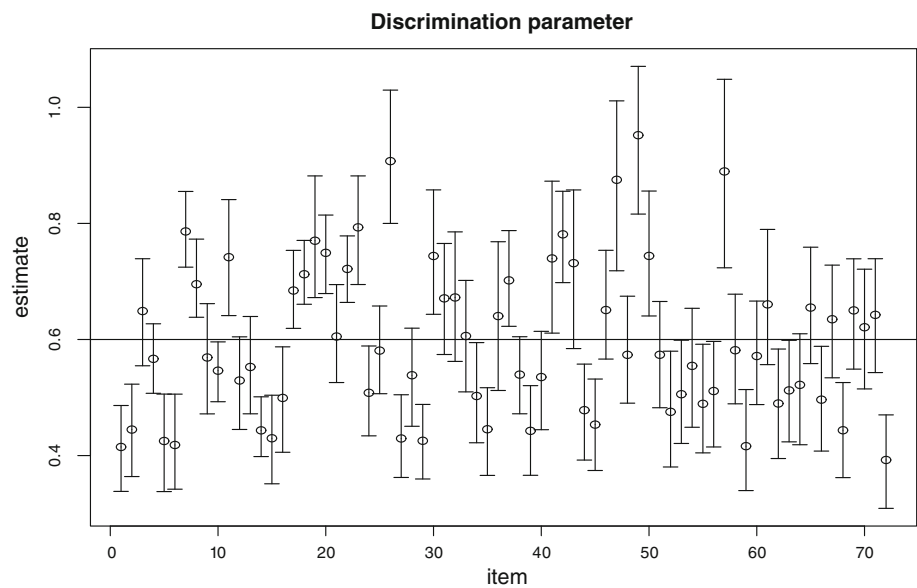
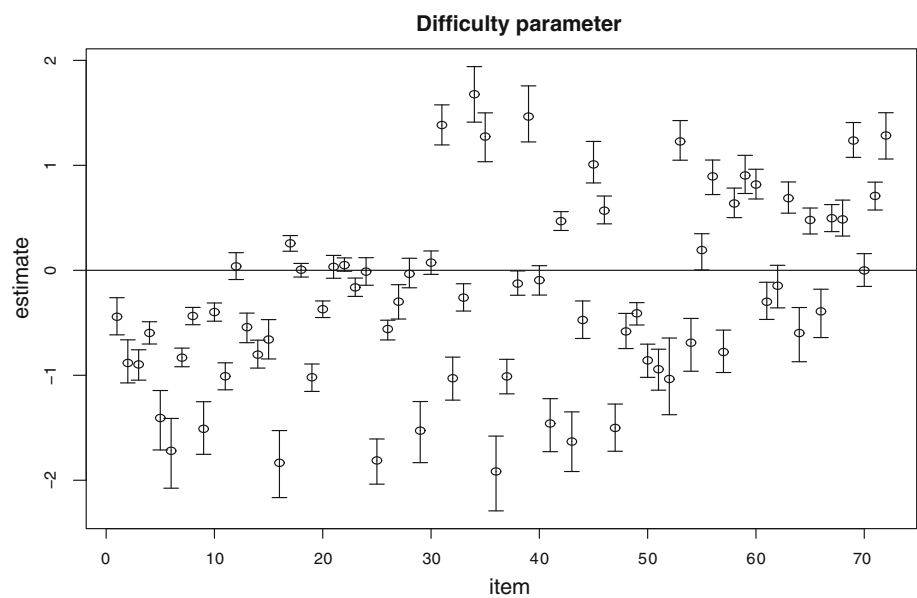


Fig. 6 Posterior means and HPD intervals for the difficulty parameters



imately 50 % of the items have sufficient discriminating power. 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. To obtain more accurate estimates of latent growth of well-performing and excellent examinees, more difficult test items are needed. The relatively easy items led to skewed population distributions (see Fig. 3), where a lot of students performed very well, which makes it difficult to accurately measure the math performances of these students. However, note that the within-examinee dependency structure over time contributes to an improved estimate of subject-specific latent trait, since it supports the use of information from other grade years to estimate the achievement level.

7 Conclusions and comments

A longitudinal item response model is proposed, where the within-examinee latent trait dependencies are explicitly modeled using different covariance structures. The time-heterogeneous covariance structures allow for time-varying latent trait variances, covariances, and correlations. The complex dependency structure across time and identification issues lead to restrictions on the covariance matrix, which complicates the specification of priors and implementation of an MCMC algorithm. By conditioning on a reference or baseline time-point, an unrestricted unstructured covariance matrix was specified given the baseline population parameters. Furthermore, the restricted structured covariance models were handled as restricted versions of the unstructured

restricted covariance model, which was estimated through the developed MCMC method.

The developed Bayesian methods include an MCMC estimation method, and different posterior predictive assessment tools. In a simulation study, the MCMC algorithm showed a good recovery of the model parameters. The assessment tools were shown to be useful in evaluating the fit of the model.

Various model extensions of the LIRT model can be considered. The latent variable distribution is assumed to be multivariate normal. This can be adjusted for example by using a multivariate skewed latent variable distribution to model asymmetric latent trait distributions. Furthermore, 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 response model at level 1 of the longitudinal model. Dropouts and inclusions of examinees were not allowed in the present data study. A multiple imputation method could be developed to support this situation, see Azevedo (2008). More 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 more general IRT model for the response data using flexible priors that can include the different extensions.

Acknowledgments The authors are thankful to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) from Brazil, for the financial support through a Doctoral Sandwich Scholarship granted to the first author under the guidance of the two others

Appendix

- **Step 1:** Simulate the augmented data using $Z_{ijt}|(\cdot)$, according to Eq. (22).
- **Step 2:** Simulate the latent traits using

$$\theta_{j\cdot}|(\cdot) \sim N_T(\widehat{\Psi}_{\theta_j} \widehat{\theta}_j, \widehat{\Psi}_{\theta_j})$$

where

$$\widehat{\theta}_j = \sum_{i|I_{ijt}=1} a_i b_i \mathbf{1}_T + \sum_{i|I_{ijt}=1} a_i z_{ij} + \Psi_{\theta}^{-1} \mu_{\theta},$$

$$\widehat{\Psi}_{\theta_j} = \left(\sum_{i|I_{ijt}=1} a_i^2 \mathbf{I}_T + \Psi_{\theta}^{-1} \right)^{-1},$$

where $\mathbf{z}_{ij\cdot} = (z_{ij1}, \dots, z_{ijT})^t$.

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

$$\widehat{\zeta}_i = \mathbf{H}_{i\cdot}^t \mathbf{z}_{i\cdot} + \Psi_{\zeta}^{-1} \mu_{\zeta},$$

$$\widehat{\Psi}_{\zeta_i} = \left(\mathbf{H}_{i\cdot}^t \mathbf{H}_{i\cdot} + \Psi_{\zeta}^{-1} \right)^{-1},$$

$$\mathbf{H}_{i\cdot} = [\boldsymbol{\theta} \quad -\mathbf{1}] \bullet \mathbf{I}_i, \tag{30}$$

where \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 vector by using

$$\mu_{\theta_1}|(\cdot) \sim N(\widetilde{\mu}_{\theta_1}, \widehat{\Psi}_{\mu}),$$

$$\mu_{\theta(1)}|(\mu_{\theta_1}, (\cdot)) \sim N_T(\widetilde{\boldsymbol{\mu}}_{\theta(T-1)}, \widehat{\Psi}_{\mu(T-1)}),$$

where

$$\widehat{\boldsymbol{\mu}}_{\theta} = \Psi_{\theta}^{-1} \sum_{j=1}^n \theta_j + \Psi_0^{-1} \mu_{\theta}$$

$$= (\widehat{\mu}_{\theta_1}, \widehat{\mu}_{\theta_2}, \dots, \widehat{\mu}_{\theta_T})^t (\widehat{\mu}_{\theta_1}, \widehat{\boldsymbol{\mu}}_{\theta}^{(T-1)})^t,$$

$$\widehat{\Psi}_{\mu} = \left(n \Psi_{\theta}^{-1} + \Psi_{\mu}^{-1} \right)^{-1} = \begin{bmatrix} \widehat{\Psi}_{\mu} & \widehat{\Psi}_{\mu}^{(T-1)} \\ \widehat{\Psi}_{\mu}^{(T-1)} & \widehat{\Psi}_{\mu}^{(T-1)} \end{bmatrix},$$

$$\widetilde{\boldsymbol{\mu}}_{\theta} = \widehat{\Psi}_{\mu} \widehat{\boldsymbol{\mu}}_{\theta} = (\widetilde{\mu}_{\theta_1}, \widetilde{\boldsymbol{\mu}}_{\theta}^{(T-1)})^t$$

$$= (\widetilde{\mu}_{\theta_1}, \widetilde{\boldsymbol{\mu}}_{\theta}^{(T-1)})^t,$$

$$\widetilde{\boldsymbol{\mu}}_{\theta(T-1)} = \widetilde{\boldsymbol{\mu}}_{\theta}^{(T-1)} + \widehat{\Psi}_{\mu}^{-1} \widehat{\Psi}_{\mu}^{(T-1)} (\mu_{\theta_1} - \widetilde{\mu}_{\theta_1}),$$

$$\widehat{\Psi}_{\mu(T-1)} = \widehat{\Psi}_{\mu}^{(T-1)} - \widehat{\Psi}_{\mu}^{-1} \widehat{\Psi}_{\mu}^{(T-1)} \widehat{\Psi}_{\mu}^{(T-1)}.$$

- **Step 5:** Simulate the first time point variance using $\psi_{\theta_1}|(\cdot) \sim IG(\widehat{\nu}_0, \widehat{\kappa}_0)$, where

$$\widehat{\nu}_1 = \frac{n + \nu_0}{2},$$

$$\widehat{\kappa}_1 = \frac{\sum_{j=1}^n (\theta_{j1} - \mu_{\theta_1})^2 + \kappa_0}{2}.$$

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

$$\widehat{\boldsymbol{\psi}} = \psi_{\theta_1}^{-1/2} \Psi_{\theta}^{*-1} \sum_{j=1}^n (\theta_{j(1)} - \mu_{\theta(1)}) (\theta_{j1} - \mu_{\theta_1})$$

$$+ \Psi_{\boldsymbol{\psi}}^{-1} \mu_{\boldsymbol{\psi}},$$

$$\widehat{\Psi}_{\boldsymbol{\psi}} = \left(\psi_{\theta_1}^{-1} \Psi_{\theta}^{*-1} \sum_{j=1}^n (\theta_{j1} - \mu_{\theta_1})^2 + \Psi_{\boldsymbol{\psi}}^{-1} \right)^{-1}.$$

- **Step 7:** Simulate the covariance matrix $\Psi^* \sim IW_{T-1}(\hat{v}_\Psi, \hat{\Psi}_\Psi)$, where

$$\begin{aligned} \hat{v}_\Psi &= n + v_\Psi, \\ \hat{\Psi}_\Psi &= \Psi_\Psi + \sum_{j=1}^n (\theta_{j(1)} - \mu_\theta^*) (\theta_{j(1)} - \mu_\theta^*)^t. \end{aligned}$$

- **Step 8:** Calculate the original covariance matrix using (10) and $\Psi_{\theta(1)} = \Psi^* + \psi^* \psi^{*t}$.
- **Step 9:** Calculate the population variances using

$$(\psi_{\theta_2}, \dots, \psi_{\theta_T})^t = \psi_{\theta(1)}^* = \text{Diag}(\Psi^* + \psi^* \psi^{*t}), \quad (31)$$

where *Diag* extracts the main diagonal of a square matrix.

- **Step 10:** Depending on the restricted covariance structure of interest, transformations are defined for unrestricted parameters to facilitate draws of restricted model parameters. Below, in each subitem, the following notation is used: $\psi_{\theta(1)}^*$ is given by (31), “ \bullet ” denotes the Hadamard product, $(\cdot)^{-1/2}$ is an inverse-square-root pointwise operator, and $A[t]$ and $A[t :]$ denotes the t -th component and the remaining values of the vector A , starting at t , respectively.
- ARH and UH: Calculate the correlation coefficient using

$$\rho_\theta = \frac{1}{T-1} \mathbf{1}_{T-1}^t \left(\psi^* \bullet (\psi_{\theta(1)}^*)^{-1/2} \right). \quad (32)$$

- HT: Calculate the correlation coefficient using

$$\rho_\theta = \psi^*[1] \times (\psi_{\theta(1)}^*[1])^{-1/2}. \quad (33)$$

- HC: Calculate the covariance parameter using

$$\rho_\theta = \frac{1}{T-1} \mathbf{1}_{T-1}^t \left(\sqrt{\psi_{\theta_1}} \psi^* \right). \quad (34)$$

- ARMAH: Calculate the moving average parameter (γ_θ) using

$$\gamma_\theta = \psi^*[1] \times (\psi_{\theta(1)}^*[1])^{-1/2} \quad (35)$$

and the correlation parameter (ρ_θ) using

$$\rho_\theta = \frac{1}{T-2} \mathbf{1}_{T-1}^t \left(\psi^*[T-2:] \bullet (\psi_{\theta(1)}^*[T-2:])^{-1/2} \right). \quad (36)$$

- AD: Calculate the correlation parameter using

$$\rho_{\theta_1} = \psi^*[1] \times (\psi_{\theta(1)}^*[1])^{-1/2} \quad (37)$$

and, for $t = 2, \dots, T - 1$, using

$$\rho_{\theta_t} = \frac{\psi^*[t:] \times (\psi_{\theta(1)}^*[t:])^{-1/2}}{\prod_{t'=1}^{t-1} \rho_{\theta_{t'}}}. \quad (38)$$

- **Step 11:** A specific covariance pattern model is computed using the appropriate restriction on the free parameters sampled from their joint distribution. The computed restricted covariance matrix is used in the repeating MCMC Steps.

The unstructured covariance matrix is the least restrictive version, and assumes unique variance and covariance parameters for the measurements of theta over time. Each structured covariance pattern is a restricted version of the unrestricted covariance pattern. The parameter space defined by the unstructured covariance pattern model represents all possible combinations of the different parameters. Therefore, this parameter space will contain all possible combinations of parameters of each restricted covariance pattern model. This property is explicitly used in the present sampling procedure. That is; each restriction will be used to imply a relationship between the parameters sampled from their joint distribution. Each relationship is implied to restrict the free parameters, which are sampled from their joint distribution, where the restriction implies a common covariance or a function of the common covariance parameter, which is defined by the set of free covariance parameters.

By sampling parameters of the unrestricted covariance pattern, potentially all possible restricted versions can be drawn. In the procedure, a restricted version is computed from the unstructured sampled covariance parameters and the restricted set of parameters are considered to be the implied restricted sample from the unrestricted sample. Since all possible restricted samples are generated from all free possible combinations of parameters, the restricted sample is obtained from the parameter space of all possible combinations of the different parameters of the restricted covariance pattern model.

References

Albert, J.: Bayesian estimation of normal ogive item response curves using Gibbs sampling. *J. Educ. Behav. Stat.* **17**, 251–269 (1992)

Albert, J.A., Chib, S.: Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* **88**, 669–679 (1993)

Albert, J.A., Chib, S.: Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747–769 (1995)

Ando, T.: Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical Bayes models. *Biometrika* **94**, 443–458 (2007)

- Andrade, D.F., Tavares, H.R.: Item response theory for longitudinal data: population parameter estimation. *J. Multivar. Anal.* **95**, 1–22 (2005)
- Azevedo, C.L.N.: Multilevel multiple group longitudinal models in item response theory: estimation methods and structural selection under a Bayesian perspective. unpublished PhD thesis, in Portuguese (2008)
- Azevedo, C.L.N., Andrade, D.F.: An estimation method for latent trait and population parameters in nominal response model. *Braz. J. Probab. Stat.* **24**, 415–433 (2010)
- Azevedo, C.L.N., Bolfarine, H., Andrade, D.F.: Bayesian inference for a skew-normal IRT model under the centred parameterization. *Comput. Stat. Data Anal.* **55**, 353–365 (2011)
- Azevedo, C.L.N., Andrade, D.F., Fox, J.-P.: A Bayesian generalized multiple group IRT model with model-fit assessment tools. *Comput. Stat. Data Anal.* **56**, 4399–4412 (2012b)
- Azevedo, C.L.N., Bolfarine, H., Andrade, D.F.: Parameter recovery for a skew-normal IRT model under a bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size. *J. Stat. Comput. Simul.* **82**, 1679–1699 (2012a)
- Bock, R.D., Aitkin, M.: Marginal maximum likelihood estimation of item parameters: an application of an EM algorithm. *Psychometrika* **46**, 317–328 (1981)
- Chib, S., Greenberg, E.: Analysis of multivariate probit models. *Biometrics* **85**, 347–361 (1998)
- Chib, S., Carlin, B.P.: On MCMC sampling in hierarchical longitudinal models. *Stat. Comput.* **9**, 17–26 (1999)
- Congdon, P.: *Applied Bayesian Modelling*. Wiley, Chichester (2003)
- Conoway, M.R.: A: random effects model for binary data. *Biometrics* **46**(1990), 317–328 (1990)
- De Ayala, R., Sava-Bolesta, M.: Item parameter recovery for the nominal response model. *Appl. Psychol. Meas.* **23**, 3–19 (1999)
- DeMars, C.E.: Sample size and the recovery of nominal response model item parameters. *Appl. Psychol. Meas.* **27**, 275288 (2003)
- Douglas, J.A.: Item response models for longitudinal quality of life data in clinical trials. *Stat. Med.* **18**, 2917–2931 (1999)
- Dunson, D.B.: Dynamic latent trait models for multidimensional longitudinal data. *J. Am. Stat. Assoc.* **98**, 555–563 (2003)
- Eid, M.: Longitudinal confirmatory factor analysis for polytomous item responses : model definition and model selection on the basis of stochastic measurement theory. *Methods Psychol. Res. Online* **1**, 65–85 (1996)
- Fitzmaurice, G., Davidian, M., Verbeke, D., Molenberghs, G.: *Longitudinal Data Analysis*, 1st edn. Chapman & Hall/CRC, London (2008)
- Fox, J.-P.: Multilevel IRT assessment. In: van der Ark, M., Sijtsma, K. (eds.) *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*, pp. 227–252. Lawrence Erlbaum Associates, Inc., London (2004)
- Fox, J.-P., Glas, C.A.W.: Bayesian modification indices for IRT models. *Stat. Neerl.* **59**, 95–106 (2005)
- Fox, J.-P.: *Bayesian Item Response Modeling: Theory and Applications*, 1st edn. Springer, New York (2010)
- Gamerman, D., Lopes, H.: *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference*, 2nd edn. Chapman & Hall/CRC, London (2006)
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: *Bayesian Data Analysis*, 2nd edn. Chapman & Hall/CRC, London (2004)
- Gelman, A.: Prior distribution for variance parameters in hierarchical models. *Bayesian Anal.* **1**, 515–533 (2006)
- Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732 (1995)
- Hedeker, D., Gibbons, R.D.: *Longitudinal Data Analysis*, 1st edn. Wiley Series, New York (2006)
- Imai, K., van Dyk, D.A.: A Bayesian analysis of the multinomial probit model using marginal data augmentation. *J. Econom.* **124**, 311–334 (2005)
- Jennrich, R.I., Schluchter, M.D.: Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805–820 (1986)
- Kass, R.E., Raftery, A.E.: Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995)
- Liu, L.C., Hedeker, D.: A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics* **62**, 261–268 (2006)
- McCulloch, R., Polson, N.G., Rossi, P.E.: A Bayesian analysis of the multinomial probit model with fully identified parameters. *J. Econom.* **99**, 173–193 (2000)
- Muthén, B.O.: Longitudinal studies of achievement growth using latent variable modeling. *Learn. Individ. Differ.* **10**, 73–101 (1998)
- Nunez-Anton, V., Zimmerman, D.L.: Modeling nonstationary longitudinal data. *Biometrics* **56**, 699–705 (2000)
- Rencher, R.C.: *Methods of Multivariate Analysis*, 1st edn. Wiley Series, New York (2002)
- Rochon, J.: Arma covariance structures with time heterocedasticity for repeated measures experiments. *J. Am. Stat. Assoc.* **87**, 777–784 (1992)
- Sahu, S.K.: Bayesian estimation and model choice in item response models. *J. Stat. Comput. Simul.* **72**, 217–232 (2002)
- Singer, J.M., Andrade, D.F.: On the choice of appropriate error terms in profile analysis. *Statistician* **43**, 259–266 (1994)
- Singer, J.M., Andrade, D.F.: Analysis of longitudinal data. In: Sen, P.K., Rao, C.R. (eds.) *Handbook of Statistics 18, Bioenvironmental and Public Health Statistics*, pp. 115–160. Elsevier, Amsterdam (2000)
- Sinharay, S.: A Bayesian item fit analysis for unidimensional item response theory models. *Br. J. Math. Stat. Psychol.* **59**, 429–449 (2006)
- Sinharay, S., Johnson, M.S., Stern, H.: Posterior predictive assessment of item response theory models. *Appl. Psychol. Meas.* **30**, 298–321 (2006)
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A.: Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **64**, 583–639 (2002)
- Stern, H.S., Sinharay, S.: Bayesian model checking and model diagnostics. In: Dey, D.D., Rao, C.R. (eds.) *Handbook of Statistics 25, Bayesian Modelling, Thinking and Computation*, pp. 171–192. Elsevier, Amsterdam (2005)
- Tavares, H.R., Andrade, D.F.: Item response theory for longitudinal data: item and population ability parameters estimation. *Test* **15**, 97–123 (2006)
- Tiao, G.C., Zellner, A.: On the Bayesian estimation of multivariate regression. *J. R. Stat. Soc. B* **26**, 277–285 (1964)