Bayesian general multivariate latent variable modeling of longitudinal item response data

Caio L. N. Azevedo[1], Jean-Paul Fox[2], Dalton F. Andrade [3]

1 Department of Statistics, University of Campinas,2 Department of Research Methodology, University of Twente, The Netherlands,3 Department of Informatics and Statistics, Federal University of Santa Catarina, Brazil

Abstract

Longitudinal item response data are characterized by examinees that are assessed at different time points or measurement occasions such that time-specific measurements are nested within examinees (Singer & Andrade (2000)). Besides the usual nesting of response observations within examinees, the time-specific latent traits are also nested within examinees, see Andrade & Tavares (2006) and Andrade & Tavares (2005). In the well-known hierarchical modeling approach, the complex dependencies due to the nested structure of the data are commonly modeled by introducing random effects such that observations and latent traits are conditionally independently distributed. However, the implied compound symmetry structure is often not sufficient to model the complex time-heterogenous dependencies.

Therefore, a Bayesian general multivariate item response modeling framework is proposed that accounts for the complex within-examinee latent trait dependencies. Flexible parametric covariance structures are considered to model specific within-examinee dependencies. Furthermore, it can handle many measurement occasions, different item response functions, and different latent trait population distributions, and generalizes the modeling framework of Andrade & Tavares (2005).

Due to identification rules and restricted parametric covariance structures, a conditional modeling approach is pursued to specify proper priors for the unrestricted parameters and to implement an efficient MCMC algorithm, by conditioning on baseline population parameters. 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.

Key words: `elsart`, longitudinal item response theory, covariance patterns, Bayesian inference, MCMC

1 Introduction

Longitudinal data are characterized by sample units that are followed along different time points such that time-specific observations are nested within the sample units (Singer & Andrade , 2000). This principle also holds for longitudinal item response theory (LIRT), where the data consist of response patterns of different examinees responding to different tests at different measurement occasions (e.g. grades). Typically, in LIRT, besides the usual nesting of response observations within subjects, within-examinee dependency is to be expected since time-specific latent traits are nested within the examinee (Andrade & Tavares , 2006).

Various longitudinal item response models have been proposed to handle the nesting of time-variant measurements in subjects. The popular mixed-effects regression modeling approach is often considered, where the subjects are treated as random effects to model the between-subject and within-subject variances. Conoway (1990) proposed a Rasch LIRT model to analyze panel data and proposed a marginal maximum likelihood method (Bock & Aitkin , 1981) for parameter estimation. Liu & 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 (1990) analyzed longitudinal response data from a quality of life instrument using a joint model, which consisted of Cox and the graded item response response model.

In the mentioned work, the common assumption is made that the time-variant measurements are conditionally independent given the subject's time-invariant mean latent trait level. The time-specific latent traits are assumed to be independently normally distributed around the subject's mean value. However, in practice the within-subject latent trait dependencies are often not completely modeled and the model errors are still correlated over time. Furthermore, the basic assumption of conditional independence given the random subject effects leads to a simple compound symmetry covariance structure for the errors. This covariance structure, consisting of just two variance-covariance parameters, assumes equal variances and covariances over time, which is most often not very realistic.

Andrade & Tavares (2006) proposed a logistic three-parameter model for longitudinal dichotomous responses with a multivariate normal population distribution for the latent traits and different covariance matrices to model the within-examinee dependency. They fitted the model using marginal maximum likelihood estimation. In their modeling framework, the latent traits covariance

Email address: Caio L. N. Azevedo: cnaber@ime.unicamp.br.
URL: http://www.ime.unicamp.br/∼cnaber.

structure models the between-subject and within-subject variances, where the variances are allowed to vary over time. This model has two serious restrictions; the covariance structures are assumed to be time-homoscedastic and the number of time-points is restricted to three. The limitation in the number of time-points is due to the fact that multivariate integrals are evaluated using Gauss-Hermite quadrature techniques. Schilling & Bock (2005) suggested adaptive quadrature or Monte carlo integration methods, depending on the number of time-points, but this will still limit the model to Gaussian latent trait distributions.

Motivated by the complex structure of a large-scale longitudinal assessment study conducted by the Brazilian Federal government, a more general LIRT model is proposed for multivariate dichotomous response data, which can handle time-heterogenous latent trait correlations. At each time-point or measurement occasion, the multivariate response observations are assumed to measure a time-specific latent trait. At the level of observations, the time-specific latent trait is considered to be a random effect. The dichotomous response observations are assumed to be conditionally independently distributed given the random effect. The time-specific latent traits are assumed to be nested within the subject. However, the subjects are not treated as random effects. Instead, the time-specific latent traits are assumed to be multivariate normally distributed, where different covariance structures can be used to capture time-invariant and time-variant dependencies. In this general flexible framework, the errors at the time-specific latent trait level are allowed to correlate and different variance-covariance structures are proposed to capture the time-specific between-subject variability and time-heterogenous longitudinal dependencies between latent traits. An important aspect is that the covariance matrices considered allow for time-heterogenous variances and covariances.

A joint estimation procedure is proposed, where an MCMC algorithm is developed for estimating simultaneously all model parameters. 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 (2011) have shown that a 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. The modeling framework is extended with various Bayesian model-fit assessment tools to enable a full Bayesian analysis of longitudinal item response data. 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 residuals can be used to evaluate the normality assumptions.

This paper is outlined as follows. After introducing the Bayesian LIRT model, the full Gibbs sampling method is given, which can handle different variance-covariance matrices. 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 eight 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, which includes the implementation of a guessing parameter.

2 The Model

One or more tests are administered to different examinees at different points in time. At each time point t, $t = 1, ..., T$, there are n_t examinees and each test has I_t items. The tests have common items and the test design can be typed as an incomplete block design. For a complete design, $n_t = n, \forall t$ and the total number of items equals $I = \sum_{t=1}^{T}$. 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 $(j = 1, \ldots, n,)$ at time-point or measurement occasion t $(t = 1, \ldots, T)$, $\boldsymbol{\theta}_{j.} = (\theta_{j1}, \ldots, \theta_{jT})^t$ the vector of the latent traits of the examinee j, and $\boldsymbol{\theta}_{..} = (\boldsymbol{\theta}_{.1}, \ldots, \boldsymbol{\theta}_{.T})^t$ the vector of all latent traits. Let Y_{ijt} represent the response of examinee j to item i $(i = 1, \ldots, I)$, $\boldsymbol{Y}_{.jt} = (Y_{1jt}, \ldots, Y_{Ijt})^t$ the response vector of examinee j, $\boldsymbol{Y}_{..t} = (\boldsymbol{Y}^t_{.1t}, \ldots, \boldsymbol{Y}^t_{.nt})^t$ the response vector of all examinees, $\boldsymbol{Y}_{...} = (\boldsymbol{Y}^t_{.1.}, \ldots, \boldsymbol{Y}^t_{.n.})^t$ the entire response set and $(y_{ijt}, \boldsymbol{y}^t_{.jt}, \boldsymbol{y}^t_{..t}, \boldsymbol{y}^t_{...})^t$ the corresponding observed values, at time-point t. Let $\boldsymbol{\zeta}_i$ denote the vector of parameters of item i, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^t_1, \ldots, \boldsymbol{\zeta}^t_I)^t$ the whole set of item parameters, and $\boldsymbol{\eta_\theta}$ the vector with population parameters.

A general LIRT model will be 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. The general LIRT model is stated by:

$$
\begin{aligned}
Y_{ijt} \mid (\theta_{jt}, \boldsymbol{\zeta}_i) &\sim \text{Bernoulli}(P_{ijt}) \\
P_{ijt} = P(Y_{ijt} = 1 \mid \theta_{jk}, \boldsymbol{\zeta}_i, \boldsymbol{\phi}) &= \sum_{l=1}^{L} \prod_{h=1}^{H} F_{lh}\left(\theta_{jt}, \boldsymbol{\zeta}_i, \boldsymbol{\phi}\right) \qquad (1) \\
\boldsymbol{\theta}_{j.} | \boldsymbol{\eta}_{\boldsymbol{\theta}} &\sim N_T(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Psi}_{\boldsymbol{\theta}}), \qquad (2)
\end{aligned}
$$

where $F(.)$ stands for a convenient cumulative distribution function with parameters $\boldsymbol{\phi}$. The above framework includes the one-, two- and three- parameter model using a probit, logit, log-log, Student t, skew probit, or skew probit link function. In addition, at stage two, other latent traits distributions can be considered such as the multivariate Student-t, multivariate skew normal, multivariate skew Studen-t and finite mixture of multivariate normals.

Furthermore, $N_T(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ denotes a T-dimensional normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$. The within-subject dependencies among the latent traits are modeled through the covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$. Let

$$
\boldsymbol{\mu}_{\boldsymbol{\theta}} = \begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \vdots \\ \mu_{\boldsymbol{\theta}_T} \end{bmatrix}, \boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{11(\boldsymbol{\theta})} & \psi_{12(\boldsymbol{\theta})} & \cdots & \psi_{1T(\boldsymbol{\theta})} \\ \psi_{21(\boldsymbol{\theta})} & \psi_{22(\boldsymbol{\theta})} & \cdots & \psi_{2T(\boldsymbol{\theta})} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1T(\boldsymbol{\theta})} & \psi_{2T(\boldsymbol{\theta})} & \cdots & \psi_{TT(\boldsymbol{\theta})} \end{bmatrix}, \qquad (3)
$$

where $\boldsymbol{\eta}_{\boldsymbol{\theta}} = (\boldsymbol{\mu}_{\boldsymbol{\theta}}^t, v(\Psi_{\boldsymbol{\theta}})^t)^t$, $v(\Psi_{\boldsymbol{\theta}}) = \text{vecd}(\Psi_{\boldsymbol{\theta}})$, *vecd(.)* stands for the different elements of $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$. Note that a multivariate normal distribution with a flexible variance-covariance structure is considered to model the within-subject latent dependencies.

This flexible LIRT model, combined with MCMC estimation methods, can 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. In addition, prediction of latent traits between observed time-points and of future time-points are easily obtained due to the conditional structure of the multivariate normal distribution. Also, model fit assessment tools such as p-values based on posterior predictive distributions and discrepancy measures can be easily implemented.

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. Secondly, an incomplete test design is required such that common items are administered at different measurement occasions. In this way, a common latent scale is defined across measurement occasions.

The restrictions on the parameters of the latent trait distribution of the first time-point complicate the specification of priors. In this latent variable framework, a novel prior modeling approach will be followed to account for the restricted covariance structure. Following McCulloh, Polson & Rossi (2000), a parametrization of the latent trait's covariance structure is considered. Therefore, the following partition of the latent traits structure is defined,

$$\begin{aligned}\boldsymbol{\theta}_{j.} &= (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jT})^t = (\theta_{j1}, \boldsymbol{\theta}_{j.}^{(T-1)})^t,\\ \boldsymbol{\mu}_{\boldsymbol{\theta}} &= (\mu_{\theta_1}, \mu_{\theta_2}, \ldots, \mu_{\theta_T})^t = (\mu_{\theta_1}, \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)})^t,\end{aligned}$$

where, $\boldsymbol{\theta}_{j.}^{(T-1)} = (\theta_{j2}, \ldots, \theta_{jT})^t$, $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)} = (\mu_{\theta_2}, \ldots, \mu_{\theta_T})^t$. Furthermore, the covariance structure is partitioned as,

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \boldsymbol{\psi}_{\boldsymbol{\theta}}^{t\,(T-1)} \\ \boldsymbol{\psi}_{\boldsymbol{\theta}}^{(T-1)} & \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{(T-1)} \end{bmatrix}, \tag{4}$$

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

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}}^{(T-1)} = \begin{bmatrix} \psi_{22(\boldsymbol{\theta})} & \cdots & \psi_{2T(\boldsymbol{\theta})} \\ \vdots & \ddots & \vdots \\ \psi_{2T(\boldsymbol{\theta})} & \cdots & \psi_{TT(\boldsymbol{\theta})} \end{bmatrix}. \tag{5}$$

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

$$\boldsymbol{\theta}_{j.}^{(T-1)} | \theta_{j1} \sim N_{(T-1)}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}(T-1)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}\right), \tag{6}$$

where, $\boldsymbol{\mu}_{\boldsymbol{\theta}(T-1)} = \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)} + \psi_{\theta_1}^{-1} \boldsymbol{\psi}_{\boldsymbol{\theta}}^{(T-1)} (\theta_{j1} - \mu_{\theta_1})$ and $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} = \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{(T-1)} - \psi_{\theta_1}^{-1} \boldsymbol{\psi}_{\boldsymbol{\theta}}^{(T-1)} \boldsymbol{\psi}_{\boldsymbol{\theta}}^{t\,(T-1)}$. This is equivalent to

$$\boldsymbol{\theta}_{j.}^{(T-1)} \mid \theta_{j1} = \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)} + \psi_{\theta_1}^{-1} \boldsymbol{\psi}_{\boldsymbol{\theta}}^{(T-1)} \left(\theta_{j1} - \mu_{\theta_1}\right) + \boldsymbol{\xi}_{j.}^{(\boldsymbol{\theta})(T-1)}, \tag{7}$$

where $\boldsymbol{\xi}_{j.}^{(\boldsymbol{\theta})(T-1)} \sim N_{(T-1)}\left(\mathbf{0}_{(T-1)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}\right)$.

As a result, the $\boldsymbol{\theta}_{j.}^{(T-1)}$ are conditionally multivariate normally distributed given the first component θ_{j1}, with an unrestricted covariance matrix. Equation (7) defines a linear multivariate regression model with independent variable $(\theta_{j1} - \mu_{\theta_1})$, intercept $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)}$, and regression parameters $\boldsymbol{\psi}_{\boldsymbol{\theta}}^{(T-1)}$. The matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}$ is an unstructured covariance matrix without any identifiability restrictions, see Singer & Andrade (2000). As a result, the common modeling (e.g., using an Inverse-Wishart prior) and estimation approaches can be applied for Bayesian inference, veja Gelman et al (2004).

For estimation purposes, the conditional distribution of the latent variables is expressed as

$$\boldsymbol{\theta}_{j.}^{(T-1)} | \theta_{j1} = \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)} + \psi_{\theta_1}^{-1/2} \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)} \left(\theta_{j1} - \mu_{\theta_1}\right) + \boldsymbol{\xi}_{j.}^{(\boldsymbol{\theta})(T-1)} \ , \tag{8}$$

where

$$\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)} = (\sqrt{\psi}_{\theta_2} \rho_\theta, \ldots, \sqrt{\psi}_{\theta_T} \rho_\theta^{(T-1)})^t \,. \tag{9}$$

In this way, the covariance and variance parameters,

$$(\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^t, \boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}^t)^t \,, \tag{10}$$

define an one-to-one relation with the free parameters of the original covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$. As a result, the estimates of the population parameters $(\boldsymbol{\psi}_{\boldsymbol{\theta}}, \rho_\theta)$ can be obtained from the estimates of Equation (10). 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 Equation (8), 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.

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 within-subject latent measurements over time. 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. Furthermore, specific covariance pattern models can provide insight in the residual correlation between latent measurements over time and account for unexplained latent dependencies to improve the fit of the model.

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 more complex patterns of latent residuals over time., where population variances of latent measurements can be differ over time-points.

Among others, Hedeker & Gibbons (2006) and Fitzmaurice et al (2008) have shown that the analysis of longitudinal multivariate response data require 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.

2.2.1 First-order heteroscedastic autoregressive model

The correlations between subject's latent traits may decrease, when distances between instants of evaluation increase for example. One of the most popular covariance structures to capture such a decrease in correlation is the first-order autoregressive, AR(1), which is given by,

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_2}}\rho_\theta & \cdots & \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_T}}\rho_\theta^{T-1} \\ \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_2}}\rho_\theta & \psi_{\theta_2} & \cdots & \sqrt{\psi_{\theta_2}}\sqrt{\psi_{\theta_T}}\rho_\theta^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_T}}\rho_\theta^{T-1} & \sqrt{\psi_{\theta_2}}\sqrt{\psi_{\theta_T}}\rho_\theta^{T-2} & \cdots & \psi_{\theta_t} \end{bmatrix}, \tag{11}$$

where $\boldsymbol{\psi_\theta} = (\psi_{\theta_1}, \ldots, \psi_{\theta_T})^t$, $\psi_{\theta_t} \in (0, \infty)$ and $\rho_\theta \in (-1, 1)$.

2.2.2 Heteroscedastic uniform model

This covariance structure allows for time-heteroscedastic variances over occasions, but time-homogenous correlations overtime. The occasion-specific variance terms are used to construct the covariance terms but cancel out time-specific correlation effects. The heteroscedastic uniform covariance matrix is given by,

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k2}}}\rho_{\theta_k} & \cdots & \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} \\ \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k2}}}\rho_{\theta} & \psi_{\theta_{k2}} & \cdots & \sqrt{\psi_{\theta_{k2}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} \\ \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k3}}}\rho_{\theta} & \sqrt{\psi_{\theta_{k2}}}\sqrt{\psi_{\theta_{k3}}}\rho_{\theta_k} & \cdots & \sqrt{\psi_{\theta_{k3}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} & \sqrt{\psi_{\theta_{k2}}}\sqrt{\psi_{\theta_{kT}}}\rho_{\theta_k} & \cdots & \psi_{\theta_{kT}} \end{bmatrix}. \tag{12}$$

2.2.3 Heteroscedastic Toeplitz model

In some settings, the covariance between subject's latent traits of two nonconsecutive instants will be equal to zero. This might be suitable when correlations decay quickly due to relatively large time-spaces between non-consecutive measurement occasions. This covariance pattern relates to the first-order moving-average process of latent traits, and is represented by,

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k2}}}\rho_{\theta_k} & 0 & \cdots & 0 \\ \sqrt{\psi_{\theta_{k1}}}\sqrt{\psi_{\theta_{k2}}}\rho_{\theta_k} & \psi_{\theta_{2k}} & \sqrt{\psi_{\theta_{2k}}}\sqrt{\psi_{\theta_{3k}}}\rho_{\theta_k} & \cdots & 0 \\ 0 & \sqrt{\psi_{\theta_{2k}}}\sqrt{\psi_{\theta_{3k}}}\rho_{\theta_k} & \psi_{\theta_{3k}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \psi_{\theta_{kT}} \end{bmatrix}. \tag{13}$$

2.2.4 Hankel model

Opposite to a heterogenous toeplitz model, time-heterogenous variances and correlations are specified while assuming a common covariance term. The variance terms specify the time-heterogenous correlations using a common covariance term across occasions and measurements. Subsequently, relatively high time-specific latent trait variances will specify a low correlation between them. The Hankel covariance structure is represented by,

$$\mathbf{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \psi_{\theta_k} & \psi_{\theta_k} & \cdots & \psi_{\theta_k} \\ \psi_{\theta_k} & \psi_{\theta_{k2}} & \psi_{\theta_k} & \cdots & \psi_{\theta_k} \\ \psi_{\theta_k} & \psi_{\theta_k} & \psi_{\theta_{k3}} & \cdots & \psi_{\theta_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_k} & \psi_{\theta_k} & \psi_{\theta_k} & \ddots & \psi_{\theta_{kT}} \end{bmatrix} . \tag{14}$$

2.2.5 First-order autoregressive moving-average model

As the first-order autoregressive AR(1) 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 ARMA(1,1), represents an extension of an AR(1) structure since allows a more flexible modeling of the time-specific correlations. The ARMA(1,1) covariance matrix is represented by,

$$\mathbf{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{k2}}}\gamma_{\theta_k} & \cdots & \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-2} \\ \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{k2}}}\gamma_{\theta_k} & \psi_{\theta_{k2}} & \cdots & \sqrt{\psi_{\theta_{k2}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-2} & \sqrt{\psi_{\theta_{k2}}\psi_{\theta_{kT}}}\gamma_{\theta_k}\rho_{\theta_k}^{T-3} & \cdots & \psi_{\theta_{kT}} \end{bmatrix} . \tag{15}$$

2.2.6 Ante-dependence model

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 variability present. The ante-dependence model implies that the correlations decrease as long as the distance between time-measurements increase. This structure has more parameters than the ARMA(1) and supports a more dynamic modeling of covariance patterns. It is represented by,

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{k2}}}\rho_{\theta_{k1}} & \cdots & \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{kT}}}\prod_{t=1}^{T-1}\rho_{\theta_{kt}} \\ \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{k2}}}\rho_{\theta_{k1}} & \psi_{\theta_{k2}} & \cdots & \sqrt{\psi_{\theta_{k2}}\psi_{\theta_{kT}}}\prod_{t=2}^{T-1}\rho_{\theta_{kt}} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_{k1}}\psi_{\theta_{kT}}}\prod_{t=1}^{T-1}\rho_{\theta_{kt}} & \sqrt{\psi_{\theta_{k2}}\psi_{\theta_{kT}}}\prod_{t=2}^{T-1}\rho_{\theta_{kt}} & \cdots & \psi_{\theta_{kT}} \end{bmatrix} \quad (16)$$

3 Bayesian inference and Gibbs sampling methods

The marginal posterior distributions comprise the main tool to perform Bayesian inference. Unfortunately, it is impossible 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 & Lopes (2006). More specifically, we will develop a full Gibbs sampling (FGS) algorithm to estimate simultaneously all parameters.

MCMC methods for longitudinal and multivariate probit models have been developed by, among others, Chib & Green (1998), Chib & Carlin (1999), Imai & van Dyk (2005), and McCulloh, Polson & Rossi (2000). A particular problem in Bayesian modeling of longitudinal multivariate response data is the prior specification of covariance matrixes. When covariance parameters are not functionally dependent, an Inverse-Wishart distribution is the most suitable option, see Tiao & Zellner (1964). When this is not the case, the specification of an appropriate prior becomes complicated. In the present case, identification rules impose restrictions on the latent trait distribution. Therefore, the prior distribution for $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$ should take this restriction and other functional dependencies into account. To accomplish that, the covariance structure is modeled conditionally given the population parameters of the first measurement occasion. Following McCulloh, Polson & Rossi (2000), this approach supports a proper implementation of the identifying restrictions and a full Gibbs sampling implementation.

We will consider prior distributions which lead to conditional conjugate families, see Gelman et al (2004) and Gelman (2006). According to the approach presented in Section 2, the parameters of interest are $(\boldsymbol{\mu}_{\boldsymbol{\theta}}^t, \psi_{\theta_1}, \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^t, \boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}^t)^t$. Conjugate priors are specified as,

$$\boldsymbol{\mu_\theta} \sim N_T(\boldsymbol{\mu}_0, \boldsymbol{\Psi}_0)\,, \tag{17}$$
$$\psi_{\theta_1} \sim IG(\nu_0, \kappa_0)\,, \tag{18}$$
$$\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)} \sim N_{T-1}(\boldsymbol{\mu}_{\boldsymbol{\psi}}, \boldsymbol{\Psi}_{\boldsymbol{\psi}})\,, \tag{19}$$
$$\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} \sim IW_{T-1}(\nu_{\boldsymbol{\Psi}}, \boldsymbol{\Psi}_{\boldsymbol{\Psi}})\,, \tag{20}$$

where $IG(\nu_0, \kappa_0)$ stands for the inverse-gamma distribution with shape parameters ν_0 and scale parameter κ_0, $IW_{T-1}(\nu_{\boldsymbol{\Psi}}, \boldsymbol{\Psi}_{\boldsymbol{\Psi}})$ stands 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

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

where $\boldsymbol{\mu}_{\boldsymbol{\zeta}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\zeta}}$ are the hyperparameters, and $\mathbb{1}$ the 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). An augmented scheme is introduced to sample normally distributed latent response data $\boldsymbol{Z}$, given the discrete observed response data; that is,

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

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

To handle incomplete block designs, and indicator variable $\boldsymbol{I}$ 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, nonresponse, or errors in recoding data are marked by another indicator, which is defined as,

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

It is assumed that the random 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.

Under the above assumptions, the joint distribution of $(\boldsymbol{Z}'_{..}, \boldsymbol{V}'_{..})'$ (conditioned on all other quantities) is given by

$$\begin{aligned} p(\boldsymbol{z}_{...}, \boldsymbol{v}_{...} | \boldsymbol{y}_{...}, \boldsymbol{\zeta}, \boldsymbol{\theta}_{..}, \boldsymbol{\eta}_{\boldsymbol{\theta}}) \propto \prod_{t=1}^{T} \prod_{j=1}^{n} \prod_{i \in I_{jt}} \Big\{ & \exp\left\{-0.5\left(z_{ijt} - a_i\theta_{jt} + b_i\right)^2\right\} \\ & \times \mathbb{1}_{(z_{ijt}, y_{ijt})} \Big\}. \end{aligned} \tag{25}$$

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

$$\begin{aligned} p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} | \boldsymbol{z}_{...}, \boldsymbol{y}_{...}) \propto\, & p(\boldsymbol{z}_{...} | \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{y}_{...}) p(\boldsymbol{\theta}_{..} | \boldsymbol{\eta}_{\boldsymbol{\theta}}) \\ & \times p(\boldsymbol{\zeta} | \boldsymbol{\mu}_{\boldsymbol{\zeta}}, \boldsymbol{\Psi}_{\boldsymbol{\zeta}}) p(\boldsymbol{\eta}_{\boldsymbol{\theta}}). \end{aligned} \tag{26}$$

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 (.) denote the set of all necessary parameters. 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_{ijt} from $Z_{ijt} \mid (.), t = 1, .., T, i = 1, ..., I, j = 1, ..., n$.
(3) Simulate $\boldsymbol{\theta}_{j.}$ from $\boldsymbol{\theta}_{j.} \mid (.), j = 1, ..., n_k$.
(4) Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i \mid (.)$, i =1,...,I.
(5) Simulate $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ from $\boldsymbol{\mu}_{\boldsymbol{\theta}}$.
(6) Simulate ψ_{θ_1} from $\psi_{\theta_1} \mid (.)$.
(7) Simulate $\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}$ from $\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)} \mid (.)$.
(8) Simulate $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}$ from $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} \mid (.)$.
(9) Compute the unstructured covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} + \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)} \boldsymbol{\psi}^t_{\boldsymbol{\theta}(T-1)}$.

(10) Compute specific covariance pattern model parameters given sampled unstructured covariance parameters.

To handle the restriction $\mu_{\theta_1} = 0$, the expression in Equation (6) is used to simulate $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(T-1)}$. To simulate $(\mu_{\theta_1}, \psi_{\theta_1})$, the following decomposition is used,

$$p(\boldsymbol{\theta}|\boldsymbol{\eta}) = p(\boldsymbol{\theta}_{j.}^{(T-1)}|\boldsymbol{\eta}_{\boldsymbol{\theta}})p(\theta_{j1}|\boldsymbol{\eta}_{\theta_1}).$$

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 $(\mu_{\theta_1}, \psi_{\theta_1})$ to specific values.

In Step 9, MCMC samples of $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}$ 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,

$$det(\boldsymbol{\Psi}_{\boldsymbol{\theta}}) = det(\psi_{\theta_1})det\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}}^{(T-1)} - \psi_{\theta_1}^{-1}\boldsymbol{\psi}_{\boldsymbol{\theta}}^{(T-1)}\boldsymbol{\psi}_{\boldsymbol{\theta}}^{t(T-1)}\right) = \psi_{\theta_1}det(\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}),$$

using Equation (4) and a property of the determinant of block matrices. As a result, the $det(\boldsymbol{\Psi}_{\boldsymbol{\theta}})$ is larger than zero since the determinant of $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)}$ is larger than zero and ψ_{θ_1} is not equal to zero. This implies positive definite samples of $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$.

In this MCMC implementation, different covariance patterns can be considered, which are treated as restricted versions of the most general restricted unstructured covariance matrix. That is, in each MCMC iteration, a specific covariance pattern model parameters is computed using the sampled unstructured covariance parameters. This is based on the notion that the different covariance patterns are nested in the most general unstructured pattern. Therefore, in MCMC Step 10, parameters of a specific covariance structure are computed. The simulated covariance matrices will be positive definite, since they are based on a positive definite unstructured covariance matrix. The selection of the most optimal covariance structure will be accomplished using Bayesian measures of model complexity as in Spiegelhalter et al (2002), pseudo-Bayes factors as in Kass & Raftery (1995), or using reversible Jump MCMC algorithms, see Azevedo (2008).

4 Simulation study

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

$$\boldsymbol{\mu}_{\boldsymbol{\psi}} = \mathbf{0}_{T-1} \, , \boldsymbol{\Psi}_{\boldsymbol{\psi}} = \tau \boldsymbol{I}_{T-1} \tag{27}$$

$$\boldsymbol{\Psi}_{\boldsymbol{\Psi}} = (\nu_{\boldsymbol{\Psi}} - T + 1) \, (\boldsymbol{I} - \boldsymbol{\Psi}_{\boldsymbol{\psi}}) \, , \tag{28}$$

and the hyperparameters for the item parameters were specified as: $\mu_{\boldsymbol{\zeta}} = (1, 0)$ and $\boldsymbol{\Psi}_{\boldsymbol{\zeta}} = \text{diag}(0, 5; 3.0)$.

4.1 Convergence and autocorrelation assessment

Following Gamerman & 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.

Responses of $n = 1000$ examinees were simulated for three measurement occasions. At each occasion or time point, 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. For each examinee, in total 60 items were administered.

The latent traits were generated from a three-variate normal distribution with $\mu_\theta = (0, 1, 2)$. The within-subject latent traits were assumed to be correlated according to an AR(1) covariance structure, where $\psi_\theta = (1, 0.9, 0.95)$ and $\rho_\theta = 0.8$. This implies a latent growth in the mean structure, variable latent variance across time, and a strong within-subject correlation over time.

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

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

where $\boldsymbol{\theta}^{*}_{jk.}$ are the transformed latent traits, $\overline{\boldsymbol{\theta}}_k$ and $\boldsymbol{S}_{\boldsymbol{\theta}_k}$ 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 1.00, for all parameters. In general it can be concluded that convergence can be established easily without requiring informative initial parameter values or long burn-in periods.

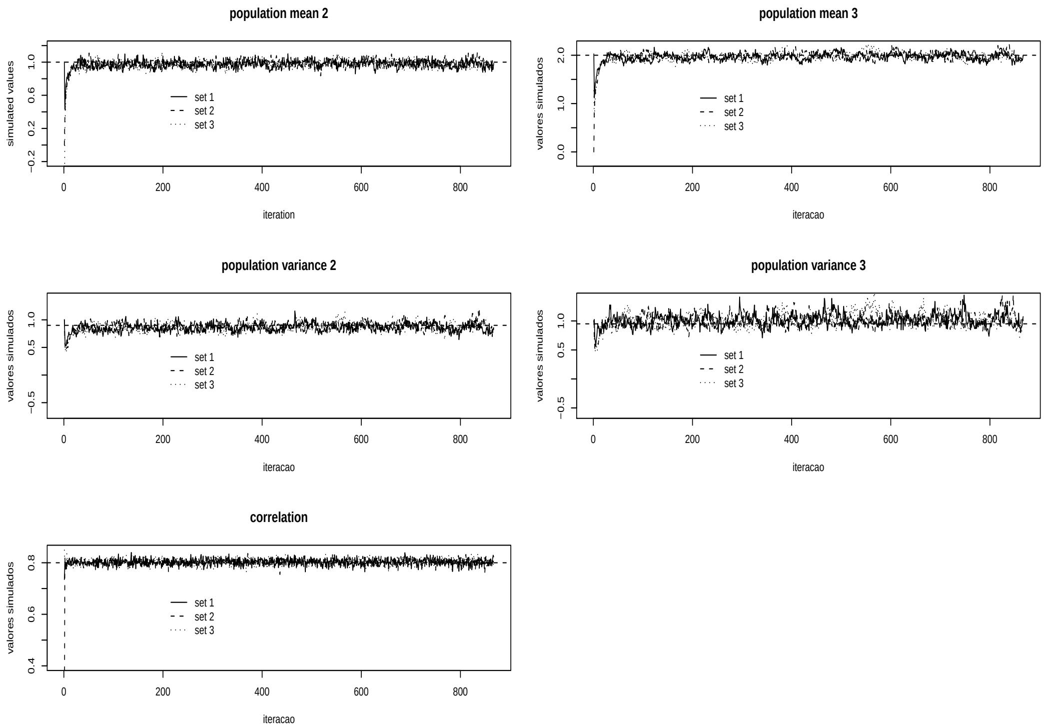

Fig. 1. For different starting values, trace plots of the simulated values of the population parameters.

4.2 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 lead to an additional set of 5 parameters. De Ayala & 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

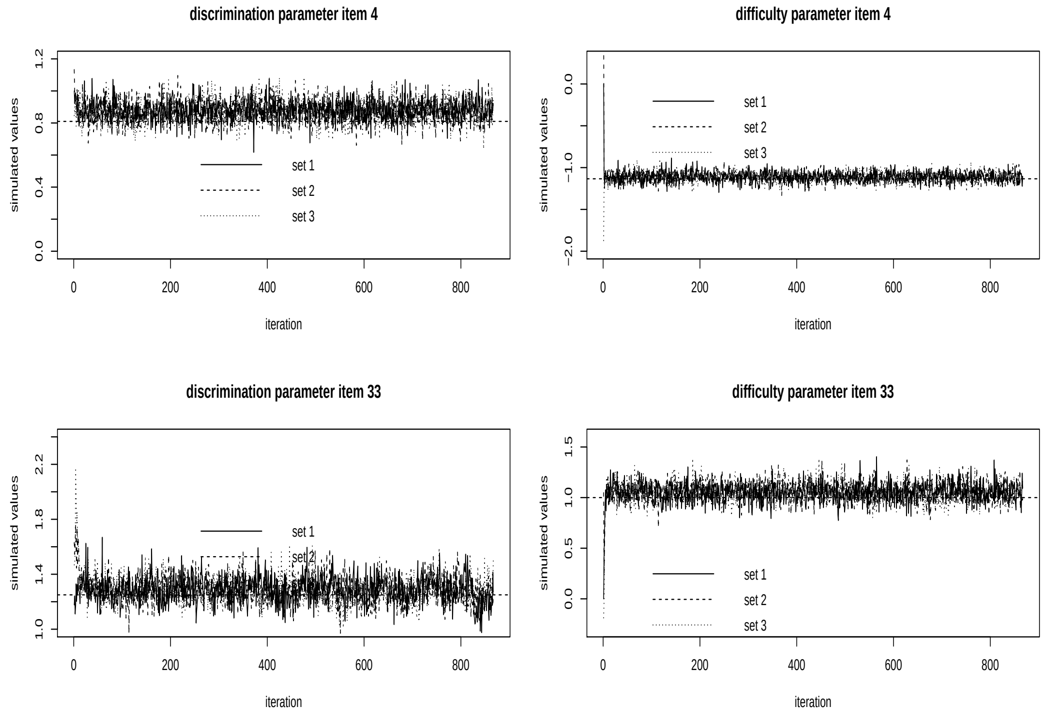

Fig. 2. For different starting values, trace plots of the simulated values for parameters of item 4 and 33.

real data study described further on, will resemble those of the simulated data study.

Different statistics were used to compare the results: correlation (Corr), mean of the standard error (MSE), the squared bias (SBias), the variance (Var) 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 & Andrade (2010) and De Ayala & Bolesta (1999). For the item and latent trait parameters, average statistics were computed by averaging across data sets, and items and persons, respectively. Accurate results are related to high values of the correlation and small values for the other statistics.

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. Therefore, we can conclude that the MCMC algorithm recovered all parameters properly. In addition, the results obtained by using other values for the hyperparameters for the prior distributions of the population parameters were very close to that ones presented. This indicate that the model is not sensitive to the choice of the prior distributions.

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

Parameter	**Statistic**				
	Corr	**SE**	**SBias**	**Var**	**RMSE**
Latent trait	.994	.280	.022	.080	.319
Discrimination	.983	.094	.001	.010	.105
Difficulty	.999	.121	.004	.020	.154

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

Parameter	**Statistics**				
	M. Est.	**SE**	**SBias**	**Var**	**RMSE**
μ_{θ_2}	.994	.046	$< .001$	$< .001$	.026
μ_{θ_3}	1.998	.079	.002	.002	.063
ψ_{θ_2}	.896	.079	.003	.003	.076
ψ_{θ_3}	1.001	.110	.007	.005	.113
ρ_{θ}	.797	.012	$< .001$	$< .001$	.008

5 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 the 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 80 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. The specific within-subject dependencies due to the longitudinal nature of the study is ignored. 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 3.

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 the AR(1) covariance structure, among others, to account for the specific dependencies.

The general LIRT model with different covariance structures was used to model the response data. 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. Different model selection criteria were used to identify the most suitable covariance structure. Table 4 represents the estimated number of parameters, AIC, BIC, and DIC, for

Table 3
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

each covariance structure. The information criteria are represented such that a smaller value corresponds to a better model fit. Although the differences are relatively small, the heterogenous uniform covariance structure describes best the latent dependency structure.

Table 4
Selecting the optimal covariance structure: Estimated Bayesian information criteria

Model	ρ_D	$\mathbb{E}(AIC)$	$\mathbb{E}(BIC)$	$\mathbb{E}(DIC)$
Uniform	**3484.07**	**147603.06**	**148331.89**	**150935.13**
Toeplitz	3486.19	147607.58	148336.41	150941.77
AR(1)	5350.19	156834.32	157563.15	162032.51
ARMA(1.1)	3488.93	147618.15	148346.98	150955.08
Hankel	3489.30	147612.24	148341.07	150949.54
Ante-dependence	3485.94	147609.86	148338.70	150943.80

Different model fit assessment tools, based on posterior predictive densities of different quantities were used to evaluate the LIRT model with a heterogenous 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 Azevedo (2008) and Azevedo et al (2011).

The Bayesian p-value was p = .482, 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 four, see Figure 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 5 represents the population parameter estimates and 95% credible intervals of the three grade levels while accounting for a uniform heterogenous correlation structure among latent traits. A significant growth in latent trait means was detected given the non-overlapping credible intervals. As expected,

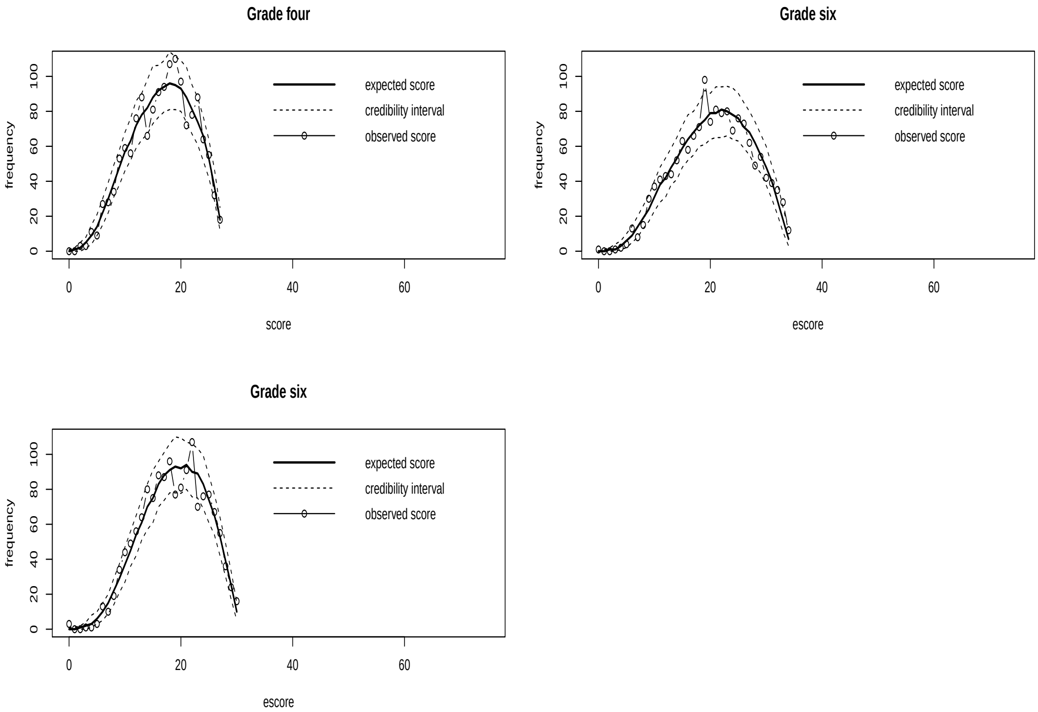

Fig. 3. Observed score distribution, predicted score distribution, and 95% central credible intervals

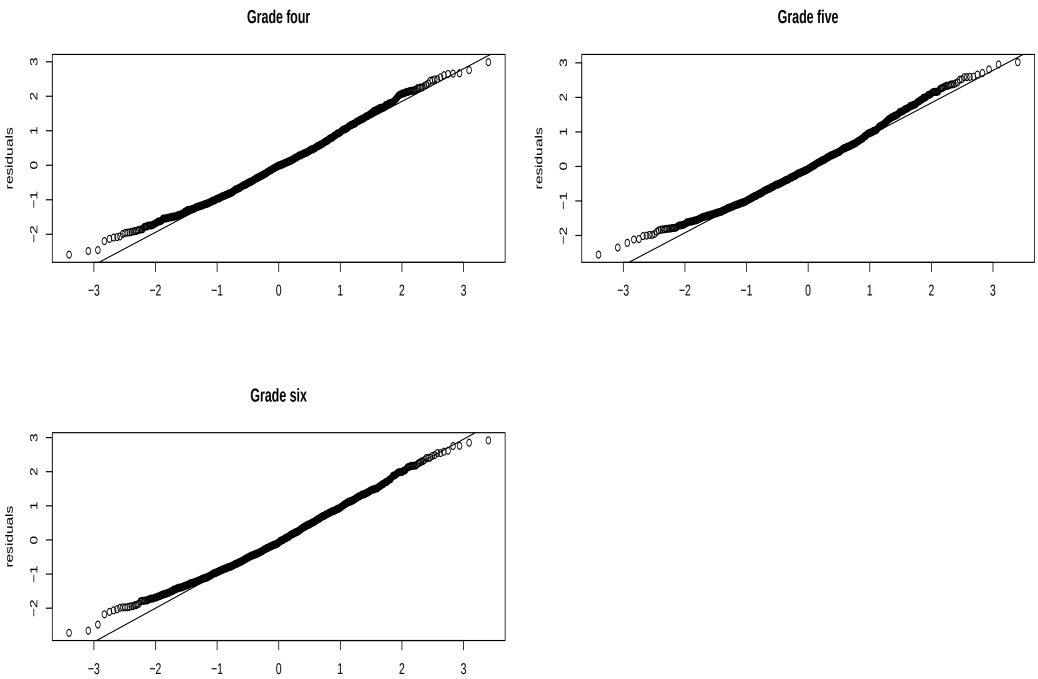

Fig. 4. For each grade, Quantile-Quantile plot of estimated latent trait residuals

the mean growth of math achievement over grade years is significant. The within-grade variability is relatively small, but the between-grade correlation is 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.

Table 5
Population parameter estimates and 95% credible intervals

	Mean			Variance		
Grade	**Mean**	**SD**	**CI 95%**	**Mean**	**SD**	**CI 95%**
four (Reference)	0	-	-	1	-	-
five	.242	.042	[.159 , .327]	1.038	.079	[.890 , 1.195]
six	.769	.050	[.672 , .866]	.972	.082	[.821 , 1.144]
Correlation						
-	0.864	.009	[.847 , .881]			

Finally, Figures 5 and 6 represent the posterior means and 95% credible intervals of the item discrimination and difficulty estimates, respectively. The discrimination parameter estimates are relatively low, where approximately 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 Figure 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.

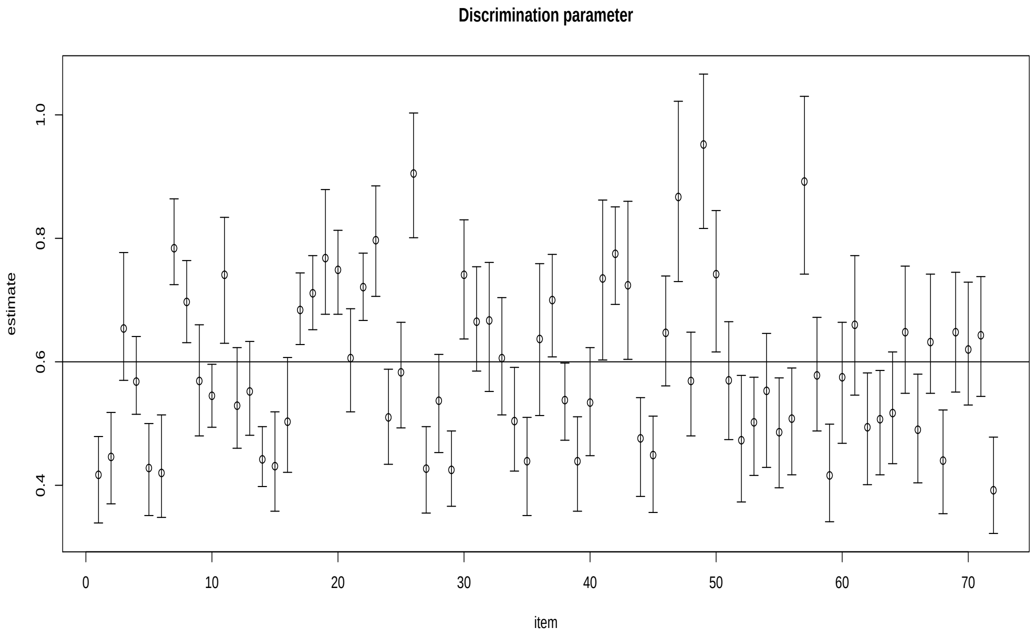

Fig. 5. Posterior means and HPD intervals for the discrimination parameters

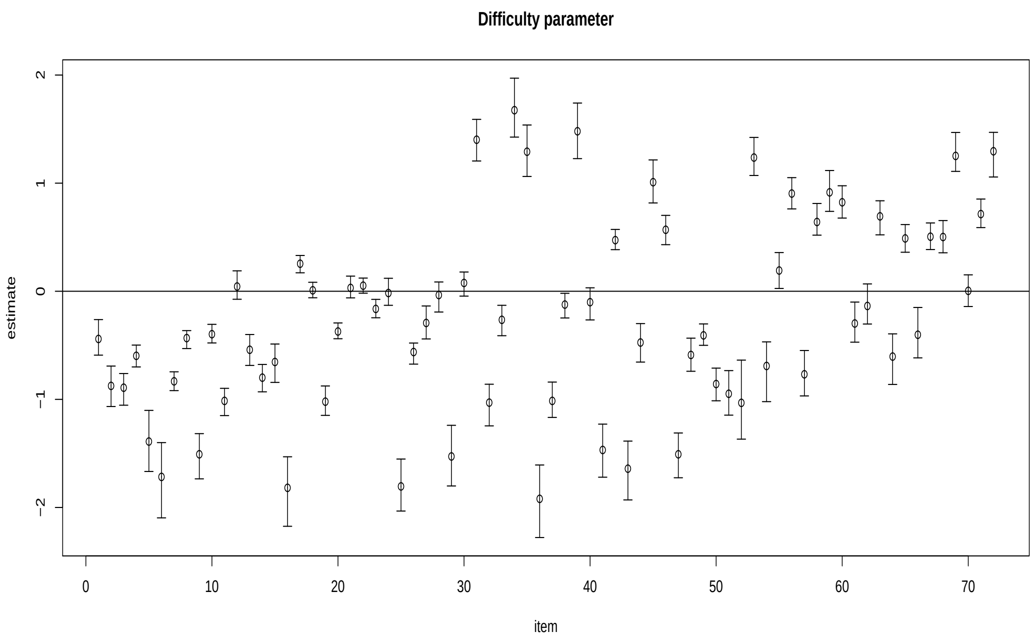

Fig. 6. Posterior means and HPD intervals for the difficulty parameters

6 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-heterogenous 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 group, 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 explicitly used in 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).

7 Acknowledgments

The authors are thankfull 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.

8 Appendix

- **Step 1 :** Simulate the augmented data by using $Z_{ijt}|(.)$, according to equation (22).
- **Step 2 :** Simulate the latent trait, independently, by using,

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

where

$$\widehat{\boldsymbol{\theta}}_j = \sum_{i \in I_{jt}} a_i b_i \mathbf{1}_T + \sum_{i \in I_{jt}} a_i \boldsymbol{z}_{ij.} + \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}} ,$$

$$\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}_j} = \left(\sum_{i \in I_{jt}} a_i^2 \boldsymbol{I}_T + \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{-1} \right)^{-1} ,$$

e $\boldsymbol{z}_{ij.} = (z_{ij1}, \ldots, z_{ijT})^t$.

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

$$\widehat{\boldsymbol{\zeta}}_i = \boldsymbol{H}_{i..}^t \boldsymbol{z}_{i..} + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\zeta}} ,$$
$$\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i} = \left(\boldsymbol{H}_{i..}^t \boldsymbol{H}_{i..} + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1} \right)^{-1} ,$$
$$\boldsymbol{H}_{i..} = [\boldsymbol{\theta} \quad -\mathbf{1}] \bullet \mathbb{1}_i , \tag{29}$$

where $\mathbb{1}_i$ is an $n \times 2$ indicator matrix with rows equal to the $\mathbf{0}$ or $\mathbf{1}$ vectors according to the subject j in the time-point t answers or not the item i and "$\bullet$" denotes the Hadamard's product, which pointwisly multiples the elements of matrices of same dimension, see Horn & Johnson (1991).

- **Step 4 :** Simulate the population mean vector by using

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

where

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{-1} \sum_{j=1}^{n} \boldsymbol{\theta}_{j.} + \boldsymbol{\Psi}_0^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}} = (\widehat{\mu}_{\theta_1}, \widehat{\mu}_{\theta_2}, \ldots, \widehat{\mu}_{\theta_T})^t = (\widehat{\mu}_{\theta_1}, \widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^{(T-1)})^t ,$$

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

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = (\widetilde{\mu}_{\theta_1}, \widetilde{\mu}_{\theta_2}, \ldots, \widetilde{\mu}_{\theta_T})^t = (\widetilde{\mu}_{\theta_1}, \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^{(T-1)})^t ,$$
$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}(T-1)} = \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^{(T-1)} + \widehat{\psi}_{\mu}^{-1}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}}^{(T-1)}(\mu_{\theta_1} - \widetilde{\mu}_{\theta_1}) ,$$
$$\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}(T-1)} = \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\mu}}^{(T-1)} - \widehat{\psi}_{\mu}^{-1}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}}^{(T-1)}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\mu}}^{t\,(T-1)} .$$

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

$$\widehat{\upsilon}_1 = \frac{n + \upsilon_0}{2} ,$$
$$\widehat{\kappa}_1 = \frac{\sum_{j=1}^{n}(\theta_{j1} - \mu_{\theta_1})^2 + \kappa_0}{2} .$$

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

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

- **Step 7 :** Simulate the covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} \sim IW_{T-1}(\widehat{\nu}_{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\Psi}})$, where

$$\widehat{\nu}_{\boldsymbol{\Psi}} = n + \nu_{\boldsymbol{\Psi}},$$
$$\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}_{\boldsymbol{\Psi}} + \sum_{j=1}^{n}\left(\boldsymbol{\theta}_{j.}^{(T-1)} - \boldsymbol{\mu}_{\boldsymbol{\theta}(T-1)}\right)\left(\boldsymbol{\theta}_{j.}^{(T-1)} - \boldsymbol{\mu}_{\boldsymbol{\theta}(T-1)}\right)^t.$$

- **Step 8 :** Calculate the original covariance matrix by using (4) and $\boldsymbol{\Psi}_{\boldsymbol{\theta}}^{(T-1)} = \boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} + \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^t$.
- **Step 9 :** Calculate the population variances by using

$$(\psi_{\theta_2}, \ldots, \psi_{\theta_T})^t = \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^* = Diag(\boldsymbol{\Psi}_{\boldsymbol{\theta}(T-1)} + \boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^t), \tag{30}$$

where $Diag$ extracts the main diagonal of a square matrix.
- **Step 10 :** Calculate the correlation coefficiente by using

$$\rho_\theta = \frac{1}{T-1}\mathbf{1}_{T-1}^t\left(\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)} \bullet (\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^*)^{-1/2}\right)^{1/t}, \tag{31}$$

where $\boldsymbol{\psi}_{\boldsymbol{\theta}(T-1)}^*$ is given by (30).

The Steps 9 and 10 above, concern the AR(1) matrix. The respective calculations for the other matrices are straightforward, see Azevedo (2008).

References

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.

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

Fitzmaurice, G., Davidian, M., Verbeke, D. & Molenberghs, G. (2008) *Longitudinal Data Analysis, first edition*, Chapman & Hall/CRC, London.

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

Andrade, D. F., & Tavares, H. R., (2006) Item Response Theory for Longitudinal Data: Item and Population Ability Parameters Estimation, *Test, 15, 1, 97-123.*

Baker, F. B. & Seock-Ho, Kim. (2004) *Item response theory : Parameter Estimation Techniques, second edition*, Marcel Dekker, Inc, New York, NY.

Bock, R. D., & Aitkin, M., (1981) Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm, *Psychometrika, 46, 2, 317-328.*

Conoway, M. R., A (1990) Random Effects Model for Binary Data, *Biometrics, 46, 2, 1990, 317-328.*

Douglas, J. A., (1990) Item response models for longitudinal quality of life data in clinical trials, *Statistics in Medicine, 18, 21, 1999, 2917-2931.*

Sahu, S. K., (2002) Bayesian Estimation and Model Choice in Item Response Models, *Journal of Statistical Computation and Simulation, 72, 3, 2002, 217-232.*

Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F., (2011), *Journal of Statistical Computation and Simulation, DOI:10.1080/00949655.2011.591798.*

Eid, M., (1996) Longitudinal Confirmatory Factor Analysis for Polytomous Item Responses : Model Definition and Model Selection on the Basis of Stochastic Measurement Theory, *Methods of Psychological Research Online, 1, 4, 65-85.*

Liu, L. C., & Hedeker, D., A, (2006) fixed-effects regression model for longitudinal multivariate ordinal data, *Biometrics, 62, 1, 261-268.*

Cochran, W. G., (2001) *Sampling Techniques, third edition*, Wiley Series, New York, NY.

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

Hedeker, D. and Gibbons, R. D., (2006) *Longitudinal data analysis, first edition*, Wiley Series, New York, NY.

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

Albert, J., (1992) Bayesian estimation of normal ogive item response curves using Gibbs sampling, *Journal of educational statistics, 17, 3, 251-269.*

Chib, S. & Greenberg, E., (1998) Analysis of multivariate probit models, *Biometrics, 85, 2, 347-361.*

Chib, S. & Greenberg, E., (1999) On MCMC sampling in hierarchical longitudinal models, *Statistics and Computing, 9, 1, 17-26.*

McCulloh, 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.*

Imai, K. & van Dyk, D. A., (2005) A Bayesian analysis of the multinomial probit model using marginal data augmentation, Journal of *Econometrics, 124, 2, 311-334.*

Tiao, G. C. & Zellner, A., (1964) On the bayesian estimation of multivariate regression, *Journal of the Royal Statistical Society, B, 26, 2, 277-285.*

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.*

De Ayala, R. & Bolesta, M. Sava, (1999) Item Parameter Recovery for the Nominal Response Model, *Applied psychological measuremente, 23, 3-19.*

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

Schafer, J. L., (1997) *Analysis of Incomplete Multivariate Data, Monographs on Statistics and Applied Probablity 72, first edition*, Chapman & Hall, London.

Little, R. J. & Rubin, D., (2002) *Statistical analysis with missing data, second edition*, John Wiley & Sons, New Jersey.

Gelman, A., Carlin, J. B., Stern, H. S. & 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.*

Nunez-Anton, V. & Zimmerman, D. L., (2000) Modellinng nonstationary longitudinal data, *Biometrics, 56, 3, 699-705.*

Horn, R. A. and Johnson, C. R., (1991) *Topics in Matrix Analysis*, Cambridge university press, London.

DeMars, C. E., (2003) Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement, 27, 4, 275288.*

Schilling, S. and Bock. R. D. (2005) High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70, 3, 533555.*

Azevedo, C. L. N., Andrade, D. F. and Fox, J.-P. (2011), Gibbs Sampling Based Estimation Procedure and Model-Fit Assessment for the Multiple-Group IRT Model, *submitted to Computational Statistics & Data Analysis.*

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 & Data Analysis, 55, 353-365.*

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit, *Journal of Royal Statistical Society, 64, 4, 583-639.*

Kass, R. E. and Raftery, A. E. (1995) Bayes Factors, *Journal of the American Statistical Association, 90, 430, 773-795.*