

## Bayesian longitudinal item response modeling with multivariate asymmetric serial dependencies

José Roberto Silva dos Santos, Caio Lucidius Naberezny Azevedo & Jean-Paul Fox

To cite this article: José Roberto Silva dos Santos, Caio Lucidius Naberezny Azevedo & Jean-Paul Fox (2021): Bayesian longitudinal item response modeling with multivariate asymmetric serial dependencies, Journal of Statistical Computation and Simulation, DOI: [10.1080/00949655.2021.1965604](https://doi.org/10.1080/00949655.2021.1965604)

To link to this article: <https://doi.org/10.1080/00949655.2021.1965604>



Published online: 19 Aug 2021.



Submit your article to this journal [↗](#)



Article views: 47






View related articles [↗](#)



View Crossmark data [↗](#)



# Bayesian longitudinal item response modeling with multivariate asymmetric serial dependencies

José Roberto Silva dos Santos <sup>a</sup>, Caio Lucidius Naberezny Azevedo <sup>b</sup> and Jean-Paul Fox <sup>c</sup>

<sup>a</sup>Departament of Statistics and Applied Mathematics, Federal University of Ceara, Fortaleza, Brazil;

<sup>b</sup>Departament of Statistics, State University of Campinas, Campinas, Brazil; <sup>c</sup>Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, Netherlands

## ABSTRACT

It is usually impossible to impose experimental conditions in large-scale longitudinal (observational) studies in education. This increases the risk of bias due to for instance unobserved heterogeneity, missing background variables, and dropouts. A flexible statistical model is required for the nature of the observational assessment data and to account for the unexplained heterogeneity. A general class of longitudinal item response theory (IRT) models is proposed, where growth in performance can be monitored using a skewed multivariate normal distribution for the latent variables. Change in performance and unexplained heterogeneity is addressed through structured covariance patterns and skewed multivariate latent variable distributions. The Cholesky decomposition of the covariance matrix is considered to model the dependence structure. A novel multivariate skew-normal distribution is defined by the antedependence model with centered skew-normal distributed errors. A hybrid MCMC approach is developed for parameter estimation, model-fit assessment, and model comparison. Results of simulation studies show good parameter recovery. A longitudinal assessment study by the Brazilian federal government is considered to show the performance of the general LIRT model.

## ARTICLE HISTORY

Received 17 August 2020

Accepted 4 August 2021


## KEYWORDS

Longitudinal IRT; Bayesian inference; antedependence models; MCMC algorithms; Cholesky decomposition

## 1. Introduction

In psychometric longitudinal research, the data consists of responses of subjects to items, which belong to a measurement instrument (e.g. cognitive test, psychiatric questionnaire) along different measurement occasions (e.g. grades, assessment or multiple screening time points). Furthermore, the item responses are indicators of an underlying trait or latent variable (e.g. ability, health status). The tests are usually unidimensional, where the latent variable explains the dependence between the subject's time-specific item responses. Most often the change in latent variable measurements represents the relevant change over

**CONTACT** José Roberto Silva dos Santos  [silvadosjr@dema.ufc.br](mailto:silvadosjr@dema.ufc.br)

 Supplemental data for this article can be accessed here. <https://doi.org/10.1080/00949655.2021.1965604>

time. For instance, in latent growth curve modeling, there is an interest in the individual trajectories of the measured latent variables.

Repeated measurements of the same subject are correlated, since the same subject is measured multiple times on the same measurement scale. There is less information available in the correlated measurements, which makes it essential to model the serial correlation between subject's measurements. The authors in Refs [1,2] considered a structured covariance matrix for the serial dependence between the within-subject latent variables. They argue that a suitable specification of the correlation pattern is very important to explain the growth in latent variables [2].

Different models have been proposed to analyze longitudinal IRT data and change in the latent variables. Next to modelling occasion-specific latent variables [3,4], auto-regressive error distributions have been introduced [5–7]. The authors in Refs. [1,8] introduced a multivariate normal distribution for the latent variables and used an unrestricted covariance matrix for the dependences across time. Recently, Azevedo [2] proposed a more general Bayesian class of LIRT models with a multivariate normal distribution for the latent variables. This class takes into account important features of the longitudinal data, such as time-heterogeneous latent variable variances and serial correlation.

The multivariate modelling approach of the longitudinal latent variable can lead to a rapid increase of the number of model parameters, as the number of subjects and/or the number of measurement occasions increases. Although Azevedo et al. [2] explored several covariance pattern structures, in the estimation procedure an unstructured full covariance matrix is considered. Despite modelling a restricted correlation structure, the number of estimated parameters still increases, as the number of measurement occasions increases. Another limitation of their approach is the restriction to a balanced design, where all subjects are measured on all occasions.

Furthermore, the serial dependence is modeled conditionally on a symmetric normal population distribution for the latent variables. The pattern of serial dependence can interfere with the strict assumption of a symmetric latent variable distribution. [9,10] show that the statistical inferences can be misleading, when the latent variable distribution is incorrectly assumed to be symmetric. In longitudinal psychometric data studies, it is more common to have an asymmetric latent variable distribution across measurement occasions. For instance, negative or positive asymmetry of the latent variable (population) distribution can occur due to inclusion or exclusion of subjects during the study. Furthermore, unexplained (non)linear growth in the longitudinal latent variable can also lead to a skewed latent variable distribution. For instance, unexplained above-average performances of a subset of subjects can affect the symmetry of the latent variable distribution.

We propose a generalized LIRT modeling approach using flexible covariance pattern structures to model serial correlation of repeatedly measured latent variables. The LIRT model allows for unbalanced data, where the number of subjects and items can differ across measurement occasions. Extensions to more complex regression structures for the latent variables – latent growth curves – are accommodated. Furthermore, diagnostic tools were developed using posterior predictive techniques. With respect to the item response modelling component, the effect of random guessing is modelled through the three-parameter probit model [11] to define the relationship between the observed response data and the latent variables.

## Antedependence model

The serial correlation in latent variables is modelled using the antedependence model. This component of the LIRT model builds on the Cholesky decomposition of the structured covariance matrix, which represents the relationship between the latent variables across time-points [12,13]. For most covariance structures the analytical expression of the Cholesky decomposition is unknown. However, the one-to-one correspondence between the parameters of the Cholesky decomposition and the covariance matrix makes it possible to compute the Cholesky decomposition given sampled covariance parameters. Our antedependence modelling approach is very flexible, includes a wide range of structured covariance patterns, and allows for handling multivariate distributions through univariate conditional distributions. This can reduce the computational costs in running MCMC algorithms, compared to a multivariate modelling approach. The dependence structure can be modelled without requiring any random effects, which often leads to highly parameterized models (see for instance [5,6]).

Prior distributions can be defined in a straightforward way in the antedependence model. Furthermore, flexible priors can be specified for the correlation parameters which also ensures that the resulting covariance matrix is positive definite.

To model asymmetric latent variable distributions, a generalization of the antedependence model is proposed by defining a skew-normal distribution for the independently distributed errors. The centered skew-normal (CSN) distribution is considered [14]. Furthermore, Henze's stochastic representation is used [15], where the error term is constructed from a half-normal and a normal random variable. It is shown that this representation facilitates a novel MCMC simulation method. As pointed out in different previous works, such as, [16–18], the CSN distribution supports model identification in a straightforward way, in contrast to the usual skew normal, see [19]. The proposed MCMC method to estimate all model parameters includes different sampling techniques to construct an efficient algorithm. Furthermore, the forward-filtering backward-sampling technique, based on the work of [20,21], is extended to deal with CSN-distributed latent variables.

This paper is outlined as follows. In Section 2, the CSN distribution is presented together with the statistical properties. In Section 3, the LIRT antedependence model is presented and the modelling of the serial correlation between latent variables is explained. In Sections 4 and 5, the MCMC algorithm is described with several novel sampling steps, and results of several simulation studies are presented to show the numerical performance of the MCMC algorithm. In Section 7, model fit assessment tools are given, and results from our Brazilian school development program study are presented. Finally, in Section 8, some conclusions and remarks are given.

## 2. The skew-normal distribution

The skew-normal distribution is a subclass of the elliptical distributions [22]. It has been used for modelling asymmetric data in many research fields [9]. A random variable  $\theta$  follows the skew-normal (SN) distribution with location parameter  $\alpha \in \mathbb{R}$ , scale parameter  $\omega \in \mathbb{R}^+$  and shape parameter  $\lambda \in \mathbb{R}$  (notation:  $\theta \sim \text{SN}(\alpha, \omega, \lambda)$ ) if its density is

given by

$$p(\theta; \alpha, \omega, \lambda) = 2\omega^{-1}\phi_1\left(\frac{y - \alpha}{\omega}\right)\Phi_1\left(\lambda\frac{\theta - \alpha}{\omega}\right), \text{ for all } \theta \in \mathbb{R},$$

where  $\phi_1$  and  $\Phi_1$  denote the probability density function (pdf) and the cumulative density function (cdf) of the univariate standard normal distribution, respectively. The mean and variance of  $\theta$  are given by, respectively,  $\mathbb{E}(\theta) = \alpha + \omega\delta r$  and  $\text{Var}(\theta) = \omega^2(1 - r^2\delta^2)$ , where,  $r = \sqrt{\frac{2}{\pi}}$  and  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ . The parameter  $\delta$  lies in the interval  $(-1, 1)$  and can be used instead of  $\lambda$ .

Azzalani [14] introduced the so-called centered parametrization which is defined as  $\theta_c = \alpha + \omega\theta_z = \mu + \sigma\theta_0$ , where  $\theta_z \sim \text{SN}(0, 1, \lambda)$ ,  $\theta_0 = \sigma_z^{-1}(\theta_z - \mu_z)$  and  $\mu_z = \mathbb{E}(\theta_z) = r\delta$ ,  $\sigma_z^2 = \text{Var}(Z) = 1 - \mu_z^2$ . Then,  $\theta_c$  follows a CSN distribution,  $\theta_c \sim \text{CSN}(\mu, \sigma^2, \gamma)$ , with mean, standard deviation and skewness coefficient, given by, respectively:

$$\mu = \alpha + \omega\mu_z; \quad \sigma = \omega\sqrt{1 - \mu_z^2}; \quad \gamma = r\delta^3 \left[ \frac{4}{\pi} - 1 \right] [1 - \mu_z^2]^{-3/2}, \quad (1)$$

where  $\gamma \in (-.99527, .99527)$ . The closer  $\gamma$  is to -1 (1) the more negatively (positively) skewed is the CSN distribution. It is (approximately) symmetric when  $\gamma \in (-.13, .13)$ . Furthermore, the Fisher information matrix obtained through the CSN distribution is non singular for all  $\gamma$  and the likelihood is well behaved, unlike the usual skew-normal distribution. In addition, its density is given by

$$\begin{aligned} p(\theta_c) &= 2\frac{\sqrt{\sigma_z^2}}{\sigma}\phi_1\left[\frac{\sqrt{\sigma_z^2}}{\sigma}\left(\theta_c - \mu + \frac{\sigma}{\sqrt{\sigma_z^2}}\mu_z\right)\right]\Phi_1\left(\lambda\left[\frac{\sqrt{\sigma_z^2}}{\sigma}\left(\theta_c - \mu + \frac{\sigma}{\sqrt{\sigma_z^2}}\mu_z\right)\right]\right) \\ &= 2\omega^{-1}\phi_1(\omega^{-1}(\theta_c - \alpha))\Phi_1[\lambda(\omega^{-1}(\theta_c - \alpha))], \end{aligned} \quad (2)$$

which corresponds to the density of the usual skew-normal distribution with parameters defined as:  $\alpha = \mu - \sigma\gamma^{1/3}\kappa$ ;  $\omega = \sigma\sqrt{1 + \gamma^{2/3}\kappa^2}$ ;  $\lambda = \frac{\gamma^{1/3}\kappa}{\sqrt{r^2 + \kappa^2\gamma^{2/3}(r^2 - 1)}}$ , where  $\kappa = (\frac{2}{4-\pi})^{1/3}$ . For more details, see [9,19].

According to Ref. [23], the stochastic representation of Henze [15] can be used to enable a simple and efficient implementation of MCMC algorithms, while avoiding the use of the complex structure of the respective density, given in Equation (2). In fact, if  $\theta \sim \text{CSN}(\mu, \sigma^2, \gamma)$  then,  $\theta = \alpha + \omega(\delta X_1 + \sqrt{1 - \delta^2}X_2)$ , where  $X_1 \sim \text{HN}(0, 1)$  and  $X_2 \sim N(0, 1)$  are independent random variables with a half-normal and normal distribution, respectively. Therefore, we have that  $\theta|X_1 \sim N(\alpha + \tau X_1, \zeta^2)$ , where  $\tau = \omega\delta$  and  $\zeta = \omega\sqrt{1 - \delta^2}$ .

### 3. The longitudinal IRT model

A longitudinal test design with anchor items is considered. Subject's item responses are observed along the different time-points. Specifically, we consider  $T$  time-points, and on each occasion a test with  $I_t$  items ( $t = 1, 2, \dots, T$ ) is administered to  $n_t$  subjects. Common items (anchors) are defined across tests to define a linked design (i.e. incomplete block design). This ensures that the latent variables are defined on a common scale. The total

number of items equals  $I \leq \sum_{t=1}^T I_t$  and the total number of latent variables equals  $n = \sum_{t=1}^T n_t$ .

Dropouts and inclusions of subjects along the study are allowed. The test design includes the particular situation where subjects are administered identical tests across occasions. This is common in health studies, where the same measurement instrument (e.g. a psychiatric inventory) is used across assessments in order to monitor a clinical condition.

The following notation is used. Let  $\theta_{jt}$  denote the latent variable of subject  $j$  ( $j = 1, 2, \dots, n_t$ ) measured at time-point  $t$ ,  $\theta_j$  is the latent variables of subject  $j$ , and  $\theta_{..}$  represent the total set of latent variables. Let  $Y_{ijt}$  denote the response of subject  $j$  to item  $i$  ( $i = 1, 2, \dots, I$ ) at time-point  $t$ ,  $\mathbf{Y}_{jt} = (Y_{1jt}, \dots, Y_{Ijt})'$  is the response vector of subject  $j$  at time-point  $t$ ,  $\mathbf{Y}_{..} = (\mathbf{Y}'_{..1}, \dots, \mathbf{Y}'_{..T})'$  is the entire response data matrix, and  $(y_{ijt}, \mathbf{y}'_{jt}, \mathbf{y}'_{..})'$  represents the respective observed values. Let  $\xi_i$  represent the vector of item parameters of item  $i$ ,  $\xi$  the vector of all item parameters and  $\eta_\theta$  the vector of population parameters, related to the latent variable distribution.

For level 1 of the LIRT model, a Probit three-parameter IRT model is considered to model the dichotomous responses given item parameters and time-specific latent variables, which is given by

$$Y_{ijt} | \theta_{jt}, \xi_i \sim \text{Bernoulli}(P_{ijt}),$$

$$P_{ijt} = \mathbb{P}(Y_{ijt} | \theta_{jt}, \xi_i) = c_i + (1 - c_i) \Phi(a_i \theta_{jt} - b_i), \tag{3}$$

where,  $a_i$  denotes the discrimination parameter,  $b_i = a_i b_i^*$ , with  $b_i^*$  is the original difficulty parameter and  $c_i$  is the so-called guessing parameter [11].

### 3.1. Modelling the latent variables: structured covariance patterns

A multivariate normal distribution is assumed for the longitudinal latent variables  $\theta_{jt}$  with mean  $\mu_\theta$  and a covariance matrix that represents the dependence structure over time denoted by  $\Sigma_\theta$ . An unstructured covariance matrix is not necessarily the most appropriate choice, since it requires a high number of parameters to model the dependence structure and it cannot be used to identify a longitudinal pattern. The unstructured covariance matrix is often appropriate when the design is balanced, and the number of measurement occasions is relatively small. However, in other scenarios correlations often change over time. For instance, the correlation between latent variables tend to decrease to zero when the time between occasions increases.

For equally or near equally spaced time points, a common covariance between the latent variables could better describe the dependence structure, which is also more efficient than an unstructured approach. When a structured covariance pattern is supported by the data, the number of parameters is reduced and the model fit can be improved compared to the unstructured covariance model. Also, the unstructured pattern is often not appropriate in more complex situations as an unbalanced design, small sample sizes (with respect to the number of subjects and items) and many measurement occasions (or time-points) [2,24]. Furthermore, more accurate statistical inferences can be made when dependences between latent variables are correctly modelled.

In Table 1, different covariance patterns are presented which will be adopted in our LIRT modelling approach. The (heteroscedastic) Toeplitz structure is most appropriate, when

**Table 1.** Overview of structured covariance patterns;  $\sigma$ -parameters represent (co)variances and  $\rho$ -parameters represent correlations.

Structure	Matrix form
First-order Heteroscedastic Autoregressive: ARH(1)	$\begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta} & \dots & \sigma_{\theta_1}\sigma_{\theta_T}\rho_{\theta}^{T-1} \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta} & \sigma_{\theta_2}^2 & \dots & \sigma_{\theta_2}\sigma_{\theta_T}\rho_{\theta}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_1}\sigma_{\theta_T}\rho_{\theta}^{T-1} & \sigma_{\theta_2}\sigma_{\theta_T}\rho_{\theta}^{T-2} & \dots & \sigma_{\theta_T}^2 \end{pmatrix}$
First-order Heteroscedastic Autoregressive Moving-Average: ARMAH(1,1)	$\begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta_1} & \dots & \sigma_{\theta_1}\sigma_{\theta_T}\rho_{\theta_1}\rho_{\theta_2}^{T-2} \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta_1} & \sigma_{\theta_2}^2 & \dots & \sigma_{\theta_2}\sigma_{\theta_T}\rho_{\theta_1}\rho_{\theta_2}^{T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_1}\sigma_{\theta_T}\rho_{\theta_1}\rho_{\theta_2}^{T-2} & \sigma_{\theta_2}\sigma_{\theta_T}\rho_{\theta_1}\rho_{\theta_2}^{T-3} & \dots & \sigma_{\theta_T}^2 \end{pmatrix}$
Heteroscedastic Toeplitz: HT	$\begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta_1} & \dots & \sigma_{\theta_1}\sigma_{\theta_T}\rho_{\theta(T-1)} \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta_1} & \sigma_{\theta_2}^2 & \dots & \sigma_{\theta_2}\sigma_{\theta_T}\rho_{\theta(T-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_1}\sigma_{\theta_T}\rho_{\theta(T-1)} & \sigma_{\theta_2}\sigma_{\theta_T}\rho_{\theta(T-2)} & \dots & \sigma_{\theta_T}^2 \end{pmatrix}$
Antependence Matrix: AD	$\begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta_1} & \dots & \sigma_{\theta_1}\sigma_{\theta_T}\prod_{t=1}^{T-1}\rho_{\theta_t} \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta_1} & \sigma_{\theta_2}^2 & \dots & \sigma_{\theta_2}\sigma_{\theta_T}\prod_{t=2}^{T-1}\rho_{\theta_t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_1}\sigma_{\theta_T}\prod_{t=1}^{T-1}\rho_{\theta_t} & \sigma_{\theta_2}\sigma_{\theta_T}\prod_{t=2}^{T-1}\rho_{\theta_t} & \dots & \sigma_{\theta_T}^2 \end{pmatrix}$

measurements are made at equally spaced time intervals. The (heteroscedastic) first-order autoregressive structure is a special case of the (heteroscedastic) Toeplitz structure. The heteroscedasticity assumption allows variances to differ across measurement occasions. More details about the covariance patterns can be found in [2].

### 3.2. Antependence covariance pattern

Pourahmadi [12] showed that a covariance matrix can be decomposed to model explicitly the covariance structure. In this approach, the Cholesky decomposition of the inverse of the covariance matrix (precision matrix) is modelled through generalized autoregressive parameters and a set of variance parameters (innovation variances). A similar approach is followed to model a structured covariance pattern using the antependence model. It is shown that this offers a flexible way to model changes in mean and covariance structures under a multivariate normal distribution.

In the antependence model, the conditional distribution of the latent variable of subject  $j$  at time-point  $t$  is defined conditionally on the latent variables before this time-point,

$$\theta_{jt} = \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) + \varepsilon_{jt}, \quad (4)$$

where  $\phi_{tk}$  are unrestricted generalized autoregressive parameters and  $\sum_{k=1}^0 k = 0$ . Note that the parameters  $\phi_{tk}$  should not be confused with the standard normal density function

$\phi_1$ . Then, the antedependence model for the latent variables of subject  $j$  is represented by

$$\boldsymbol{\varepsilon}_j = \mathbf{L}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta), \tag{5}$$

where the generalized autoregressive parameters are stored in the lower-triangular matrix  $\mathbf{L}$  ( $T \times T$ ), represented by

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{21} & 1 & 0 & \cdots & 0 \\ -\phi_{31} & -\phi_{32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\phi_{T1} & -\phi_{T2} & \cdots & -\phi_{T(T-1)} & 1 \end{pmatrix}. \tag{6}$$

In the unstructured antedependence model in Equation (5) are the error terms  $\boldsymbol{\varepsilon}_j = (\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{jT})'$  uncorrelated with covariance matrix  $\mathbf{D}$ , where  $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_T)$  is a diagonal matrix with the innovation variances on the diagonal [25].

The Cholesky decomposition of the  $\boldsymbol{\Sigma}_\theta$  is defined as  $\mathbf{L}\boldsymbol{\Sigma}_\theta\mathbf{L}' = \mathbf{D} = \mathbf{L}\text{Cov}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta)\mathbf{L}'$ , where  $\mathbf{D}$  is a diagonal matrix with the innovation variances, and  $\mathbf{L}$  is a lower-triangular matrix with  $-\phi_{tk}$  as the  $(t, k)$ th entry. For the antedependence model (Equation (5)), it is often not possible to obtain closed-form expressions for  $\mathbf{L}$  and  $\mathbf{D}$ . However, the Cholesky decomposition can be numerically computed, given values for the covariance parameters, which can be used to compute both  $\mathbf{L}$  and  $\mathbf{D}$ . Furthermore, there is a one-to-one correspondence between the parameters  $(d_1, d_2, \dots, d_T, \phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$  and the parameters  $(\sigma_{\theta_1}^2, \sigma_{\theta_2}^2, \dots, \sigma_{\theta_T}^2, \rho_\theta)'$ . Prior distributions for the covariance parameters can be translated to the generalized autoregressive parameters and innovation variances of  $\mathbf{L}$  and  $\mathbf{D}$ , respectively.

**Antedependence with CSN-error distribution**

The LIRT model with the antedependence structure for the multivariate distributed latent variables is extended to allow for a skewed multivariate normal distribution. This leads to a novel general class of LIRT models. Changes in the latent variable can be modelled across time, in the mean and covariance structure, while accounting for time-specific possibly skewed-error distributions. Furthermore, it will be shown that the implied marginal distributions for the latent variable also defines a new class of distributions. Extensions to include different covariance pattern structures are also possible.

In order to model the multivariate skew-normal structure of the latent variables, the errors in Equation (4) are assumed to be CSN distributed,

$$\varepsilon_{jt} \stackrel{i.i.d}{\sim} \text{CSN}(0, d_t, \gamma_{\varepsilon_t}), \quad t = 1, 2, \dots, T \text{ and } j = 1, \dots, n_t. \tag{7}$$

When  $\gamma_{\varepsilon_t} = 0, \forall t$ , the (symmetric) normal distribution is assumed, where  $\boldsymbol{\theta}_j \sim N_T(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ , with marginal distribution  $\theta_t \sim N(\mu_{\theta_t}, \sigma_{\theta_t}^2)$ . The multivariate normal distribution of the latent variable  $\boldsymbol{\theta}_j$  of subject  $j$  according to the antedependence model with CSN distributed errors is given by,

$$p(\boldsymbol{\theta}_j) = p(\theta_{j1}) \prod_{t=2}^T p(\theta_{jt} | \theta_1, \dots, \theta_{t-1})$$



$$= 2^T \prod_{t=1}^T \omega_t^{-1} \phi_1(\omega^{-t}(\theta_{jt} - \beta_{jt})) \Phi_1[\lambda_t(\omega_t^{-1}(\theta_{jt} - \beta_{jt}))], \quad (8)$$

where  $\beta_{jt}$  is defined as

$$\beta_{jt} = \alpha_{jt} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}), \quad (9)$$

with  $\beta_{j1} = \alpha_{j1}$ .

A closed-form expression for the marginal variances of the latent traits can be derived, given the inverse of the lower-triangular matrix. From Equation (5), it follows that  $(\theta_j - \mu_{\theta}) = \mathbf{L}^{-1} \boldsymbol{\varepsilon}_j$ , and as a result,

$$\sigma_{\theta_1}^2 = d_1; \quad \sigma_{\theta_t}^2 = d_t + \sum_{k=1}^{t-1} l_{tk}^2 d_k, \quad t = 2, \dots, T. \quad (10)$$

where  $l_{tk}$  is the  $(t, k)$ th entry of  $\mathbf{L}^{-1}$ .

Pearson's skewness coefficient can also be used to characterize the asymmetry of the latent variables at time-point  $t$ . This coefficient is given in terms of the (standardized) third moment of the random variable  $\theta_{jt}$ . The following lemma provides a general expression for the third central moment of  $\theta_{jt}$ . It can be used to compute the marginal skewness coefficient of each latent variable distribution for each measurement occasion.

**Lemma 3.1:** *Consider the antedependence model with CSN-distributed errors for multivariate normally distributed latent variables  $\theta_{jt}$  with parameters  $d_t$  and  $\gamma_{\varepsilon_t}$ , and  $l_{tk}$  the entries of the lower triangular matrix  $\mathbf{L}^{-1}$  of the Cholesky decomposition of covariance matrix  $\text{Cov}(\boldsymbol{\theta}_j) = \boldsymbol{\Sigma}_{\theta}$ ; the third centered moment of  $\theta_{jt}$  is given by*

$$\mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] = d_t^{3/2} \gamma_{\varepsilon_t} + \sum_{k=1}^{t-1} l_{tk}^3 d_k^{3/2} \gamma_{\varepsilon_k}. \quad (11)$$

The proof is given in Appendix A. According to lemma 3.1, the marginal skewness coefficient, represented by  $\gamma_{\theta_t} = \mathbb{E}((\theta_{jt} - \mu_{\theta_t})/\sigma_{\theta_t})^3$ , is given by

$$\gamma_{\theta_1} = \gamma_{\varepsilon_1}$$

$$\gamma_{\theta_t} = \frac{1}{\sigma_{\theta_t}^3} \left[ d_t^{3/2} \gamma_{\varepsilon_t} + \sum_{k=1}^{t-1} l_{tk}^3 d_k^{3/2} \gamma_{\varepsilon_k} \right], \quad t = 2, \dots, T, \quad (12)$$

with marginal variances  $\sigma_{\theta_t}^2$  (Equation (10)).

### 3.3. Model identification

A linked design is considered, where sets of common items are used in subsequent test occasions. This ensures that a common metric for the latent variables is defined, and that all results related to the tests across time are measured on the same scale. The scale is identified

by restricting the mean and variance of the latent variable to zero and one, respectively, at a measurement occasion (e.g. the first occasion is often chosen as the reference point).

When the first occasion serves as the reference group, the latent variable of the first measurement occasion is (standard) skew-normal,  $\theta_{j1} \sim \text{CSN}(0, 1, \gamma_{\theta_1})$ , with a free skewness parameter. This defines the metric of the scale. In conclusion, with a linked test design and by restricting the mean and variance of the distribution of the latent variable on the first measurement occasion, the LIRT model, with a skewed multivariate normal latent variable distribution and a structured covariance pattern, is identified and the metric is defined.

#### 4. MCMC algorithm

An MCMC algorithm is developed to estimate all parameters, since the posterior distributions have intractable analytical forms. The MCMC samples are used to obtain empirical approximations of the marginal posterior distributions of interest and to estimate the characteristics of the marginal posteriors.

The MCMC algorithm is described for an LIRT model with multivariate normally distributed latent variables with a structured covariance matrix according to an antedependence model with CSN-distributed errors. Furthermore, a three-parameter Probit model is used for the dichotomous item response data. Extensions of the MCMC algorithm concerning the type of covariance structure for the latent variables are discussed.

First, data augmentation schemes are discussed to address the sampling of the guessing parameters and to deal with dichotomous response observations. Second, the sampling of the parameters (mean, variance and skewness) of the marginal skewed normal distribution of the latent variables is discussed. Third, the sampling of the latent variables is described, and in the fourth part, the sampling of the correlation parameters is presented.

#### Data augmentation scheme

In order to facilitate the implementation of the MCMC algorithms, a well-known augmented data approach is used. For the three-parameter Probit model, the augmented data scheme proposed by [26] is used. A vector of binary variables  $W_{ijt}$  is introduced:

$$W_{ijt} = \begin{cases} 1, & \text{subject } j, \text{ time-point } t \text{ knows response to item } i \\ 0, & \text{subject } j, \text{ time-point } t \text{ doesn't know response to item } i. \end{cases}$$

The conditional distribution of  $W_{ijt}$  given  $Y_{ijt} = y_{ijt}$  is given by,

$$\begin{aligned} \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto \Phi(a_i\theta_{jt} - b_i); \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto c_i(1 - \Phi(a_i\theta_{jt} - b_i)) \\ \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 0; \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 1. \end{aligned} \tag{13}$$

Another augmented continuous variable is defined, referred to as  $Z_{ijt}$ , which is defined conditionally on the  $W_{ijt}$ :

$$Z_{ijt} | (\theta_{jt}, \zeta_i, w_{ijt}) = \begin{cases} N(a_i\theta_{jt} - b_i, 1)\mathbb{I}_{(Z_{ijt} \geq 0)}, & \text{if } W_{ijt} = 1 \\ N(a_i\theta_{jt} - b_i, 1)\mathbb{I}_{(Z_{ijt} < 0)}, & \text{if } W_{ijt} = 0, \end{cases} \tag{14}$$

where  $\mathbb{I}$  denotes the usual indicator function. An indicator variable  $\mathbf{I}$  is defined to deal with the incomplete test design:

$$\mathbf{I}_{ijt} = \begin{cases} 1, & \text{item } i \text{ administrated to respondent } j \text{ at time-point } t, \\ 0, & \text{item } i \text{ not administrated to respondent } j \text{ at time-point } t. \end{cases}$$

Then, following the usual assumptions of conditional independence, the augmented likelihood is given by

$$L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}) \propto \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in I_{jt}} p(z_{ijl} | w_{ijl}, \theta_{jt}, \boldsymbol{\zeta}_i) p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i), \quad (15)$$

where  $I_{jt}$  is the set of items answered by subject  $j$  at time-point  $t$ .

### Prior distributions

The following priors are defined, where independence is assumed between latent variables and item parameters given hyperparameters  $\boldsymbol{\eta}$ :

$$\begin{aligned} p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \boldsymbol{\eta}_{\zeta}, \boldsymbol{\eta}_{\eta}) \\ = p(\theta_{j1} | \boldsymbol{\eta}_{\theta_1}) \prod_{t=2}^T \prod_{j=1}^{n_t} p(\theta_{jt} | \theta_{j[1:(t-1)]}, \boldsymbol{\eta}_{\theta_t}) \left( \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_{\zeta}) \right) \left( \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_{\eta}) \right), \end{aligned}$$

and  $\boldsymbol{\eta}_{\zeta}$  and  $\boldsymbol{\eta}_{\eta}$  are the hyperparameters associated with  $\boldsymbol{\zeta}$  and  $\boldsymbol{\eta}_{\theta}$ , respectively. The subscript  $[1 : (t-1)]$  indicates the range of latent variables and the respective prior distribution is defined in Equation (8). Let  $\boldsymbol{\zeta}_{i(-c_i)} = (a_i, b_i)$ , for the item parameters the priors are given by

$$\begin{aligned} p(\boldsymbol{\zeta}_{i(-c_i)}) &\sim N(\boldsymbol{\mu}_{\zeta}, \boldsymbol{\Psi}_{\zeta}) \mathbb{I}_{(a_i > 0)}, \\ c_i &\sim \text{Beta}(a_c, b_c). \end{aligned}$$

The covariance structure of the latent variables is specified by the antedependence model Equation (4) with CSN distributed errors according to Equation (7). Then, using Henze's stochastic representation, this skewed multivariate normal distribution for the latent variables is represented by

$$\theta_{j1} | h_{j1} \stackrel{i.i.d.}{\sim} N(\alpha_1 + \tau_1 h_{j1}, \varsigma_1^2), \quad \theta_{jt} | h_{jt} \stackrel{i.i.d.}{\sim} N\left(\alpha_t + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_k) + \tau_t h_{jt}, \varsigma_t^2\right), \quad (16)$$

where  $h_{jt} \sim \text{HN}(0, 1)$  with  $\text{HN}(\mu, \psi)$  the half-normal distribution with (non-truncated) mean  $\mu$  and variance  $\psi$ . Furthermore, the following conjugate priors are assumed,

$$\alpha_t \sim N(\mu_{\alpha}, \sigma_{\alpha}^2); \quad \tau_t \sim N(\mu_{\tau}, \sigma_{\tau}^2); \quad \varsigma_t^2 \sim \text{IG}(a_{\varsigma}, b_{\varsigma}). \quad (17)$$

These parameters are used only to facilitate the implementation of the MCMC algorithm. However, note that the original parameters can still be recovered. Let  $\omega_t = \sqrt{\tau_t^2 + \varsigma_t^2}$ ,  $\delta_t =$

$\frac{\lambda_t}{\sqrt{1+\lambda_t^2}}$ , and  $\lambda_t = \frac{\tau_t}{\zeta_t}$ , then the original parameters can be obtained from:

$$\mu_t = \alpha_t + r\delta_t\omega_t; \quad d_t = \omega_t^2(1 - r^2\delta_t^2)^2; \quad \gamma_t = r\delta_t^3 \left[ \frac{4}{\pi} - 1 \right] [1 - r^2\delta_t^2]^{-3/2}. \quad (18)$$

The original parameters are recovered in the final stage of the estimation process.

For the generalized autoregressive parameters, the following prior is assumed:

$$\phi_{tk} \sim N(\mu_\phi, \sigma_\phi^2) \quad t = 2, \dots, T \text{ and } k = 1, \dots, t - 1. \quad (19)$$

For a structured matrix, the prior distribution for correlation parameters is a truncated normal distribution on the interval  $[-1, 1]$ :

$$\rho_{\theta_t} \sim N(\mu_\rho, \sigma_\rho^2) \mathbb{I}_{[-1,1]}. \quad (20)$$

Depending on the application, the truncation can be adjusted to consider for instance only positive correlations. This is a more common situation in educational assessment, where a positive growth in abilities can be expected across measurement occasions.

The joint posterior distribution is given by:

$$p(\boldsymbol{\theta}_{\dots}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, | \mathbf{z}_{\dots}, \mathbf{w}_{\dots}, \mathbf{y}_{\dots}) \propto L(\boldsymbol{\theta}_{\dots}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \mathbf{z}_{\dots}, \mathbf{w}_{\dots}, \mathbf{y}_{\dots}) p(\boldsymbol{\theta}_{\dots}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta)$$

This posterior distribution has an intractable analytical form, and an MCMC algorithm is used to estimate the model parameters.

### Posterior distributions

For convenience in notation, let  $(.)$  denote the set of all others parameters. The full conditional posterior distribution of the item parameters is given by:

$$\boldsymbol{\zeta}_{i(-c_i)} | (.) \sim N(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i} \widehat{\boldsymbol{\zeta}}_i, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}) \quad (21)$$

with  $\widehat{\boldsymbol{\zeta}}_i = (\boldsymbol{\Theta}_i)' \mathbf{z}_i + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\zeta}}$ ,  $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i} = [(\boldsymbol{\Theta}_i)'(\boldsymbol{\Theta}_i) + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1}]^{-1}$ , and  $\boldsymbol{\Theta}_i = [\boldsymbol{\theta} - \mathbf{1}_n]$ , taking into account which responses were collected for item  $i$ .

The posterior distribution for the *guessing parameter* is given by,

$$c_i | (.) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1) \quad (22)$$

where  $s_i = \sum_{j=1}^n \mathbb{I}(w_{ijt} = 0) y_{ij}$  represents the number of correct responses through guessing and  $t_i = \sum_{j=1}^n \mathbb{I}(w_{ijt} = 0)$  the number of subjects who did not know the correct answer to item  $i$  and guessed the response.

The posterior distributions of the structured covariance parameters are discussed. First, consider  $\beta_{jt}$ , the variables  $h_{jt}$  can be sampled from

$$h_{jt} | (\cdot) \sim N \left[ \frac{\tau_t(\theta_{jt} - \beta_{jt})}{\tau_t^2 + \zeta_t^2}, \frac{\zeta_t^2}{\tau_t^2 + \zeta_t^2} \right] \mathbb{I}(h_{jt} > 0). \quad (23)$$

Second, the transformed population parameters are sampled as follows:

$$\alpha_t | (\cdot) \sim N(aA; A), \quad (24)$$

where

$$a = \frac{1}{\zeta_t^2} \sum_{j=1}^{n_t} \left[ \theta_{jt} - \sum_{k=1}^{t-1} \phi_{tk} (\theta_{jk} - \mu_{\theta_k}) - \tau_t h_{jt} \right] + \frac{\mu_\alpha}{\sigma_\alpha^2} \quad \text{and} \quad A = \left( \frac{n_t}{\zeta_t^2} + \frac{1}{\sigma_\alpha^2} \right)^{-1}$$

and

$$\tau_t | (\cdot) \sim N(bB; B), \quad (25)$$

where  $b = \frac{1}{\zeta_t^2} \sum_{j=1}^{n_t} (\theta_{jt} - \beta_{jt}) h_{jt} + \frac{\mu_\tau}{\sigma_\tau^2}$  and  $B = \left( \frac{\sum_{j=1}^{n_t} h_{jt}^2}{\zeta_t^2} + \frac{1}{\sigma_\tau^2} \right)^{-1}$ , and

$$\zeta_t^2 | (\cdot) \sim \text{IG} \left\{ \frac{n_t}{2} + a_\zeta; \frac{1}{2} \sum_{j=1}^{n_t} (\theta_{jt} - \beta_{jt} - \tau_t h_{jt})^2 + b_\zeta \right\}. \quad (26)$$

Third, the original population parameters can be easily recovered, since  $(\alpha_t, \tau_t, \zeta_t^2)$  are a one-to-one mapping of  $(\mu_{\theta_t}, d_t, \gamma_{\varepsilon_t})$ . Furthermore, the generalized autoregressive parameters are sampled from a normal distribution:

$$\begin{aligned} \phi_{tk} | (\cdot) &\sim N(Q_{tk} q_{tk}, Q_{tk}), \\ Q_{tk} &= \left( \sum_{j=1}^{n_t} (\theta_{j(t-1)} - \mu_{\theta_{t-1}})^2 / \zeta_t^2 + 1 / \sigma_\phi^2 \right)^{-1}, \\ q_{tk} &= \frac{1}{\zeta_t^2} \sum_{j=1}^{n_t} (\theta_{jk} - \mu_{\theta_k}) (\theta_{jt} - \alpha_t - \sum_{k \neq t} \phi_{tk} (\theta_{jk} - \mu_{\theta_k}) - \tau_t h_{jt}). \end{aligned} \quad (27)$$

### Forward Filtering Backward Sampling of Latent Variables

The straightforward sampling of time-specific latent variables can lead to highly autocorrelated chains, in specific, when there are many time points [27]. In that case, samples from the posterior distribution of  $\theta_j$  are obtained by sampling from the univariate full conditional distributions,  $\theta_{jt} | \theta_{j(-t)}$  with  $\theta_{j(-t)}$  the latent variables excluding the  $t$ -th component. Therefore, [20,21] proposed a sampling scheme for dynamic models where so-called state parameters are jointly sampled using the Kalman filter. The procedure is referred to as the Forward Filtering Backward Sampling (FFBS) method.

Following [27], for the augmented data the LIRT model can be presented as a dynamic linear model with an observation equation and a system equation. In this representation, the model is given by,

$$Z_{jit} = a_i \theta_{jt} - b_i + \xi_{jit}, \tag{28}$$

$$\theta_{jt} = \beta_{jt} + \tau_t h_{jt} + \varepsilon_{jt}, \tag{29}$$

where the error components are normally distributed,  $\xi_{jit} \sim N(0, 1)$ ,  $\varepsilon_{jt} | h_{jt} \sim N(0, \varsigma_t^2)$  and  $h_{jt} \sim HN(0, 1)$ . The  $\beta_{jt}$  are defined according to Equation (9). The model is completed with a prior  $\theta_{j1} \sim CSN(\mu_{\theta_1}, d_1, \gamma_{\varepsilon_1})$ .

One of the main aspects of a dynamic model is that sequential inference can be based on the update distribution of  $\theta_{jt}$  given  $\mathbf{z}_j^t$ , where  $\mathbf{z}_j^{t-1}$  represents the information until  $t$ . Consider that at time-point  $t-1$ , the update distribution is  $\theta_{j(t-1)} | \mathbf{z}_j^{t-1}, h_{jt} \sim N(m_{j(t-1)}, C_{j(t-1)})$ . The system equations (29) can be written as  $\theta_{jt} | \theta_{j(t-1)}, h_{jt} \sim N(\beta_{jt} + \tau_t h_{jt}, \varsigma_t^2)$ . By properties of the normal distribution, these specifications can be combined leading to the marginal distribution  $\theta_{jt} | \mathbf{z}_j^{t-1}, h_{jt} \sim N(a_{jt}, R_{jt})$ , where  $a_{jt} = \alpha_t + \sum_{k=1}^{t-1} \phi_{tk}(m_{jk} - \mu_k) + \tau_t h_{jt}$  and  $R_{jt} = \varsigma_t^2 + \sum_{k=1}^{t-1} \phi_{tk}^2 C_{jk}$ .

According to the Bayes' theorem, the updated posterior distribution is obtained by  $p(\theta_{jt} | \mathbf{Z}_j^t, h_{jt}) \propto p(\mathbf{Z}_j^t | \theta_{jt})p(\theta_{jt} | \mathbf{Z}_j^{t-1}, h_{jt})$ , resulting in

$$\theta_{jt} | \mathbf{z}_j^t, h_{jt} \sim N(m_{jt}, C_{jt}), \tag{30}$$

where

$$C_{jt} = \left( \sum_{i \in I_{jt}} a_i^2 + \frac{1}{R_{jt}} \right)^{-1} \quad \text{and} \quad m_{jt} = \left( \sum_{i \in I_{jt}} a_i(z_{ij t} + b_i) + \frac{a_{jt}}{R_{jt}} \right) C_{jt}.$$

Equation (30) is referred in the literature as the *Kalman Filter*. A sample from the posterior distribution of  $(\theta_j | \mathbf{z}_j^t, h_{jt})$  is obtained by noting that

$$p(\theta_j | \mathbf{z}_j^T, h_{jt}) = p(\theta_{jT} | \mathbf{z}_j^T, h_{jt}) \prod_{t=1}^{T-1} p(\theta_{jt} | \theta_{j(t+1)}, \mathbf{z}_j^t, h_{jt}).$$

We can show that

$$\theta_{jt} | \theta_{j(t+1)}, \mathbf{z}_j^t, h_{jt} \sim N(m_{\theta_t}, C_{\theta_t}), \tag{31}$$

where

$$C_{\theta_{jt}} = \left( \frac{\phi_{t+1,t}^2}{\varsigma_{t+1}^2} + \frac{1}{C_{jt}} \right)^{-1}, \quad m_{\theta_{jt}} = \left( \frac{\phi_{t+1,t} (\theta_{j(t+1)} - \alpha_{j(t+1)})}{\varsigma_{t+1}^2} + \frac{m_{jt}}{C_{jt}} \right) C_{\theta_{jt}}$$

and

$$\alpha_{j(t+1)} = \alpha_{t+1} - \phi_{t+1,t} \mu_t + \tau_{t+1} h_{j(t+1)} + \sum_{k=1}^{t-1} \phi_{t+1,k} (\theta_{jk} - \mu_k).$$

Therefore, a scheme to sample from the full conditional distribution of  $\theta_j$  is given in algorithm 4.1. Step 1 is obtained by running the Kalman filter from  $t = 1$  to  $t = T$ . When running the filter, the updated means  $m_{jt}$  and variance  $C_{jt}$ , are stored for use in step 2.

**Algorithm 4.1** FFBS algorithm

- 
- 1: Sample  $\theta_{jT}$  from its updated distribution given in (30) and set  $t = T - 1$ .
  - 2: Sample  $\theta_{jt}$  from the distribution (31).
  - 3: Decrease  $t$  to  $t - 1$  and return to step 2 until  $t = 1$ .
- 

**Sample parameters covariance pattern**

Beside the antedependence model, a prior for the correlation parameter of another structured covariance matrix is directly specified, see Equation (20). When a specific covariance pattern is imposed, it is not possible to sample directly from the full conditional distribution. Therefore, the Metropolis-Hastings (M-H) algorithm is used with a uniform proposal density:

$$p(\rho_{\theta_t}^{(m)} | \rho_{\theta_t}^{(m-1)}) = U\left(v_1(\rho_{\theta_t}^{(m-1)}), v_2(\rho_{\theta_t}^{(m-1)})\right), \quad (32)$$

where  $v_1 = \max\{-.99527, \rho_{\theta_t} - \Delta_\rho\}$  and  $v_2 = \max\{.99527, \rho_{\theta_t} + \Delta_\rho\}$ ,  $\Delta_\rho > 0$ . The superscript ( $m$ ) refers to the MCMC iteration number. The M-H step is illustrated for the ARH(1) covariance matrix (see Table 1). Let  $\theta_{..}$  represent the set of all latent variables, as defined in Section 3 and  $p(\theta_{..} | \mu_\theta, \phi, \mathbf{d}, \boldsymbol{\gamma}_\theta)$  represent the conditional posterior of the latent variables,

$$\begin{aligned} p(\theta_{..} | \mu_\theta, \phi, \mathbf{d}, \boldsymbol{\gamma}_\theta) &= \prod_{j=1}^{n_t} p(\theta_{j1} | \mu_{\theta_1}, d_1, \gamma_{\theta_1}) \prod_{t=2}^T p(\theta_{jt} | \mu_{\theta_t}, \phi_t, d_t, \gamma_{\theta_t}) \\ &\propto 2^{Tn_t} \prod_{j=1}^{n_t} \prod_{t=1}^T \omega_t^{-1} \phi_1(\omega_t^{-1}(\theta_{jt} - \beta_{jt})) \Phi_1[\lambda_t(\omega_t^{-1}(\theta_{jt} - \beta_{jt}))], \end{aligned} \quad (33)$$

where  $\phi = (\phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$ ,  $\mathbf{d} = (d_1, \dots, d_T)'$ ,  $\boldsymbol{\gamma}_\theta = (\gamma_{\theta_1}, \dots, \gamma_{\theta_T})'$ ,  $\mu_\theta = (\mu_{\theta_1}, \dots, \mu_{\theta_t})$  and  $\phi_t$  denotes the elements correspondent to the time-point  $t$ . Furthermore,  $\beta_{j1} = \alpha_1$ ,  $\beta_{jt} = \alpha_t + \sum_{k=1}^{t-1} \phi_{jk}(\theta_{jk} - \mu_{\theta_k})$ , and the original parameters are defined as  $\alpha_t = \mu_{\theta_t} - \sqrt{d_t} \gamma_{\theta_t}^{1/3} \kappa$ ,  $\omega_t = \sqrt{d_t(1 + \gamma_{\theta_t}^{2/3} \kappa^2)}$ , and  $\lambda_t = \frac{\gamma_{\theta_t}^{1/3} \kappa}{\sqrt{r^2 + \kappa^2 \gamma_{\theta_t}^{2/3} (r^2 - 1)}}$ , where  $r$  and  $\kappa$  are the constants defined in Section 2. Finally, algorithm 4.2 presents the procedure for sampling correlation parameters of the ARH(1) structure.

The MCMC algorithm described in algorithm 4.3 presents the sampling steps for the LIRT model with the (unstructured) antedependence model and skewed multivariate normally distributed latent variables. To consider another structured covariance matrix, step 11 is replaced by Metropolis-Hastings steps to sample the necessary correlation parameters. For example, when using ARH(1), step 11 is replaced by algorithm 4.2. Note that slight modifications of the algorithm 4.2 are required to handle more complex covariance matrices. All algorithms were implemented by the authors in R programming language [28], and computational routines are available upon request from the authors.

---

**Algorithm 4.2** Procedure for sampling the correlation parameter for ARH(1)

---

**Require:** A function chol ( ) to perform the Cholesky decomposition

**Require:** A function ARH1.matrix ( ) to build the ARH(1) matrix

- 1: Draw  $\rho_\theta^{(m)} \sim U(v_1(\rho_\theta^{(m-1)}), v_2(\rho_\theta^{(m-1)}))$
- 2: Build the ARH(1) proposed matrix  $\Sigma_{\rho_\theta}^{(m)}$  using ARH1.matrix ( )
- 3: Perform the Cholesky decomposition of  $\Sigma_{\rho_\theta}^{(m)}$  to obtain the matrices  $\phi^{(m)}$  and  $\mathbf{d}^{(m)}$
- 4: Draw  $u \sim U(0, 1)$
- 5: **if**  $u < \min\{1, Q\}$  where

$$Q = \frac{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{..}^{(m-1)} | \boldsymbol{\mu}_\theta^{(m-1)}, \boldsymbol{\phi}^{(m)}, \mathbf{d}^{(m)}, \boldsymbol{\gamma}_\theta^{(m-1)}) p(\rho_\theta^{(m)}) [v_1(\rho_\theta^{(m)}) - v_2(\rho_\theta^{(m)})]}{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{..}^{(m-1)} | \boldsymbol{\mu}_\theta^{(m-1)}, \boldsymbol{\phi}^{(m-1)}, \mathbf{d}^{(m-1)}, \boldsymbol{\gamma}_\theta^{(m-1)}) p(\rho_\theta^{(m-1)}) [v_1(\rho_\theta^{(m-1)}) - v_2(\rho_\theta^{(m-1)})]}$$

6: **then**  $\rho_\theta^{(m-1)} = \rho_\theta^{(m)}$

7: **end if**

---



---

**Algorithm 4.3** Gibbs sampling with FFBS for unstructured matrix

---

- 1: Start the algorithm by choosing suitable initial values. Repeat steps 2-11.
  - 2: Simulate  $W_{ijt}$  from  $W_{ijt} | (\cdot)$  using equation (13) for all  $i, j, t$ .
  - 3: Simulate  $Z_{ijt}$  from  $Z_{ijt} | (\cdot)$  using equation (14) for all  $i, j, t$ .
  - 4: Simulate  $h_{jt}$  from  $h_{jt} | (\cdot)$  using equation (23) for all  $j, t$ .
  - 5: Simulate  $\theta_{jt}$  using algorithm 4.1 for all  $j, t$ .
  - 6: Simulate  $\boldsymbol{\zeta}_{i(-c_i)}$  from  $\boldsymbol{\zeta}_{i(-c_i)} | (\cdot)$  using equation (21) for all  $i$ .
  - 7: Simulate  $c_i$  from  $c_i | (\cdot)$  using equation (22) for all  $i$ .
  - 8: Simulate  $\alpha_t$  from  $\alpha_t | (\cdot)$  using equation (24) for all  $t$ .
  - 9: Simulate  $\tau_t$  from  $\tau_t | (\cdot)$  using equation (25) for all  $t$ .
  - 10: Simulate  $\zeta_t^2$  from  $\zeta_t^2 | (\cdot)$  using equation (26) for all  $t$ .
  - 11: Simulate  $\phi_{tk}$  from  $\phi_{tk} | (\cdot)$  using equation (27) for all  $t \geq 2$  and  $k = 1, \dots, t - 1$ .
- 

## 5. Parameter recovery study

Two simulation studies were conducted to examine the performance of the MCMC algorithm. In this study, without lose of generality, the AD model was used for the parameter recovery study. Responses of 1500 subjects for six time points were simulated according to the LIRT model with skewed multivariate latent variables and the AD modelled covariance matrix (see Section 3). Latent variables were also simulated under the misspecification settings using multivariate normal and multivariate skew-t distributions, see [29]. Item parameters to simulate data were sampled using the following intervals:  $a_i \in [0.7, 2.62]$ ,  $b_i^* \in [-1.95, 4]$ , and  $c_i \in (0.20, \dots, 0.25)$ . A total of six test were constructed from total 120 items. The order of the difficulty parameters were fixed from low to high to construct an increase in test difficulties, which also determined the order of the tests. The first test consisted of 20 items, each subsequent test contained 20 unique items and the last 20 items of the former test.



**Table 2.** Hyperparameter specifications for the prior parameters.

	Hyperparameters						
	$\mu_{\zeta}$	$\Psi_{\zeta}$	$(a_c, b_c)$	$(\mu_{\alpha}, \sigma_{\alpha}^2)$	$(\mu_{\tau}, \sigma_{\tau}^2)$	$(a_{\zeta}, b_{\zeta})$	$(\mu_{\rho}, \sigma_{\rho}^2)$
Prior 1	(1, 0)	(0.5, 16)	(5.6, 16.8)	(0,100)	(0,100)	(0.01, 0.01)	(0, 10)
Prior 2	(1, 0)	(2, 16)	(5.6, 16.8)	(0,50)	(0,50)	(0.02, 0.02)	(0, 5)
Prior 3	(1, 0)	(0.5, 9)	(3.8, 15.2)	(0,10)	(0,10)	(0.1, 0.1)	(0, 1)
Prior 4	(1, 0)	(0.2, 3)	(5.6, 16.8)	(0,1)	(0,1)	(1, 1)	(0, 1)

The latent variables were simulated from the antedependence model (Equation (4)) with both normal and CSN-distributed errors (Equation (7)). We also simulate data from the multivariate skew-t distribution with asymmetry parameters equal to 10 and degrees of freedom equal to 2 in order to have heavy-tailed and strongly asymmetric latent distributions. All results were scaled to have the same mean and covariance. For the six measurement occasions, the true population parameters: mean, variance, skewness, and correlation are given in Tables 5–7 for the tree different latent distributions. The skewness parameters represent strong, moderate and weak asymmetry. The population mean increased over occasions to simulate improvement in average performance. The true correlations were high to enforce impact of serial correlation.

Table 2 presents the hyperparameter settings for the prior distributions. In order to verify sensibility of the proposed algorithm to the prior choice, four set of hyperparameters were defined. The prior for the correlation parameter corresponds to a truncated normal distribution, according to Equation (20). The population-mean item discrimination and difficulty was set to one and zero, respectively. For the skewness parameter, more weight was given to values near zero, while allowing reasonable weight for the other non-zero values. The discrimination parameters were assumed to vary reasonably around a satisfactory discrimination power. The guessing parameters were assumed to be centered around 0.25 or 0.20 with a small between-item variance. In general, variances were fixed in order to have noninformative, informative and strong informative priors.

### **MCMC convergence**

The usual tools were used to examine the MCMC convergence: trace plots, Gelman-Rubin's and Geweke's statistics. For latent variables we consider the multivariate PSRF (*potential scale reduction factor*), see [30]. We generated three chains based on three different sets of starting values. The univariate and multivariate PSRF was close to one for all parameters, which showed no indication that the chains did not converge. The trace plots and Geweke's statistic indicated that a burn-in of 10,000 iterations was sufficient to reach convergence. Further, the correlograms (not showed) indicated that MCMC samples composed by storing every 40th iteration had negligible autocorrelation. For each run, a total of 50,000 MCMC iterations was done. Trace plots are presented in Appendix A.

### **Performance criteria**

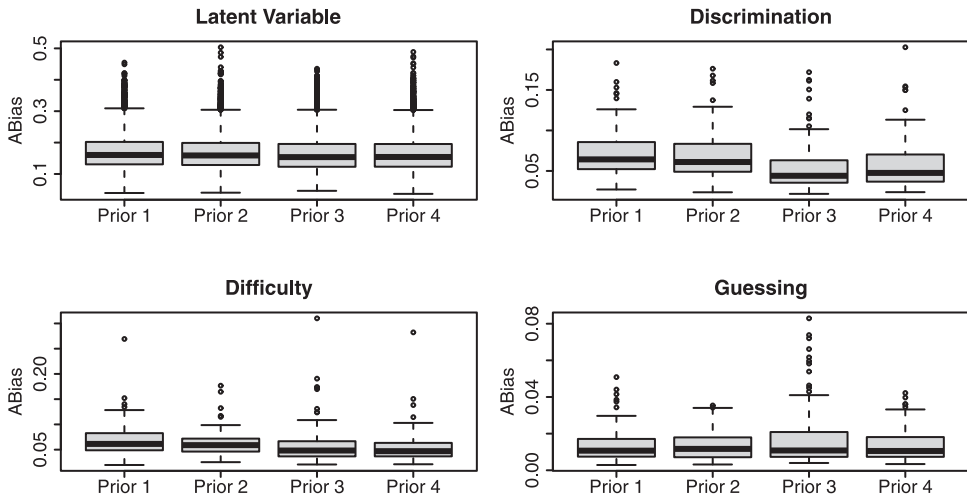
In order to assess the parameter recovery, we considered the following statistics: mean of the posterior mean estimates (M.Est.), correlation (Corr), absolute bias (ABias), variance

(VAR) and the root mean squared error (RMSE). A total of 10 data replications were simulated. For latent variable and item parameters, the statistics were computed by averaging across data sets, items and subjects. Let  $\vartheta$  be the parameter of interest, and let  $r$  refer to the data replication number. For one replication, the posterior mean estimate is represented by  $\hat{\vartheta}_r$ , and across  $R$  replications the final posterior estimate is represented by  $\hat{\vartheta} = \sum_{r=1}^R \hat{\vartheta}_r / R$ . The statistics can be represented as

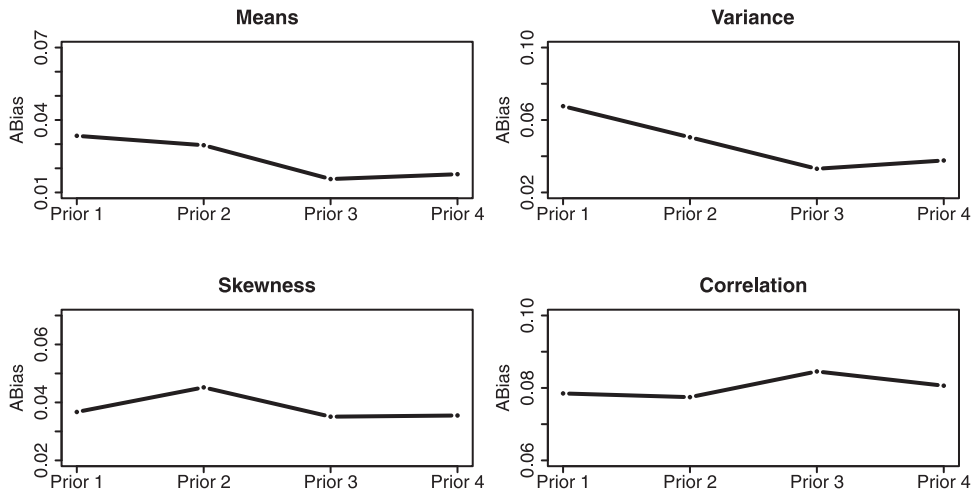
$$\begin{aligned} \text{Abias} &= \frac{1}{R} \sum_{r=1}^R \left| \hat{\vartheta}_r - \vartheta \right|, \\ \text{Corr} &= \frac{1}{R} \sum_{r=1}^R \text{Cor} \left( \hat{\vartheta}_r, \vartheta \right), \\ \text{VAR} &= \frac{1}{R-1} \sum_{r=1}^R \left( \hat{\vartheta}_r - \hat{\vartheta} \right)^2, \\ \text{RMSE} &= \sqrt{\frac{1}{R} \sum_{r=1}^R \left( \hat{\vartheta}_r - \vartheta \right)^2}. \end{aligned}$$

### Simulation results

Figures 1 and 2 present the results of the prior sensitivity study, considering the priors defined in Table 2. The amount of bias does not vary considerably over the different priors. One may notice a slight increase in the average bias of correlation parameters for more informative priors. Considering the mean and variance parameters, a slight reduction in average bias was observed for more informative priors. We can conclude that a misspecification of the hyperparameters will not lead to substantial bias in the posterior estimates. In Tables 3 and 4, the estimation results are presented for the latent variables and item parameters. For each scenario, both symmetric and skew models were fitted, in order to compare the results of the two adjustments. For the actual setting, i.e. skew-normal latent distributions, the statistic estimates show that those parameters were properly recovered for the skew model, since the variance and bias of the estimates are small. The correlation estimates for the latent variable, the discrimination and difficulty parameter estimates were all high. This is an usual pattern for IRT models (see for instance [2]). The correlation estimates of the guessing parameters estimates were omitted. The simulated data sets showed a large variation in guessing effects, this led to a low correlation between estimated and true guessing item parameters. However, the other statistics showed acceptable results. The results show an increasing in the average relative bias and RMSE when the symmetric model is considered. For the normal data, the results of the two adjustments are similar, being that, bias and RSME of the symmetric model are slightly smaller. We also see that, the estimates of the latent variables and item parameters under the skew model are more robust than those obtained fitting the symmetric model, for latent variables simulated from the multivariate skew-t distribution. In this scenario, bias and RMSE for the symmetric model are quite higher than those obtained using skew model. In Tables 5–7, the estimation results are shown for the latent variable population parameters. The population values for the mean



**Figure 1.** Prior sensitivity study: absolute bias of the latent variables and item parameters estimates.



**Figure 2.** Prior sensitivity study: averaged absolute bias of population parameters estimates.

and variance of the first occasion were set to zero and one, respectively, to identify the latent scale. Considering the fit of the skew model under the skew-normal assumption, it can be concluded that the estimates of the occasion-specific mean and variance of the latent variable were close to the true values. The estimates of skewness and correlation parameters were also close to the true values. Note that the parameters  $\rho_{\theta_1}$  until  $\rho_{\theta_5}$  refer to the correlation with the current and former latent variable estimates of occasions 1 to 6, as defined in Table 1 under the antedependence matrix. Estimates of the skew model are more accurate than those obtained considering the symmetric model for all considered scenarios, specially, estimates of means and variances.

**Table 3.** Simulation study: estimation results for the latent variable and item parameters across six time points.

		Normal simulated data				Skew-normal simulated data			
		Statistic				Statistic			
	Parameter	Corr	RMSE	VAR	ABias	Corr	RMSE	VAR	ABias
Skew	Latent variable	0.984	0.231	0.771	0.186	0.985	0.212	0.798	0.171
	Discrimination	0.983	0.101	0.005	0.086	0.983	0.078	0.005	0.064
	Difficulty	0.998	0.103	0.004	0.092	0.998	0.076	0.004	0.064
	Guessing	–	0.015	< 0.001	0.013	–	0.016	< 0.001	0.013
Symmetric	Latent variable	0.984	0.221	0.793	0.177	0.983	0.279	0.680	0.231
	Discrimination	0.982	0.085	0.005	0.070	0.982	0.187	0.005	0.176
	Difficulty	0.998	0.082	0.004	0.070	0.997	0.186	0.004	0.177
	Guessing	–	0.015	< 0.001	0.013	–	0.020	< 0.001	0.017

**Table 4.** Simulation study: estimation results for the latent variable and item parameters across six time points.

		Skew-t simulated data			
		Statistic			
	Parameter	Corr	RMSE	VAR	ABias
Skew	Latent variable	0.970	0.341	0.790	0.249
	Discrimination	0.974	0.113	0.011	0.093
	Difficulty	0.997	0.288	0.090	0.182
	Guessing	–	0.019	< 0.001	0.015
Symmetric	Latent variable	0.901	0.712	0.733	0.538
	Discrimination	0.933	0.198	0.032	0.158
	Difficulty	0.989	0.747	0.256	0.585
	Guessing	–	0.031	0.001	0.027

### Simulating dropouts

Dropouts were simulated by randomly removing subjects from the sample across occasions. The number of students was 1500 at the first occasion but dropped to 1470 in occasion 4, to 1410 in occasion 5 and 6. Except for including the effects of dropouts, the setup of this study was similar to the first simulation study. In Table 8, the estimation results are presented for the latent variable and item parameter estimates. It follows that these estimation results are again good and comparable to those of Table 3, where no dropouts were allowed. In Table 9, the results are presented of the population parameter estimates of the latent variable estimates. The parameters were recovered accurately, and the estimated value resembled the true value, where the estimated variance (VAR) was small. The RMSE of the skewness parameter estimates are comparable to those of the scenario of no dropouts, and only slightly higher for the last time points where there were more dropouts.

## 6. Model fit assessment tools

*Posterior Predictive Model Checking* is considered for model fit evaluation [31,32]. The main idea is to evaluate the extremeness of an observed discrepancy measure using simulated discrepancy values based on the posterior predictive distribution. More formally, let  $y^{obs}$  represent the response data, and  $y^{rep}$  the replicated response data. Then, the posterior

**Table 5.** Simulation study: estimation results for the population parameters of the latent variable for six occasions.

		Skew-Normal simulated data											
		True	M.est	RMSE	VAR	ABias	True	M.est	RMSE	VAR	ABias		
Skew	$\mu_{\theta_1}$	0.000	–	–	–	–	$\gamma_{\theta_1}$	0.800	0.706	0.122	0.007	0.104	
	$\mu_{\theta_2}$	1.000	0.993	0.019	< 0.001	0.015	$\gamma_{\theta_2}$	0.546	0.497	0.066	0.002	0.054	
	$\mu_{\theta_3}$	1.400	1.386	0.022	< 0.001	0.019	$\gamma_{\theta_3}$	0.118	0.087	0.040	0.001	0.033	
	$\mu_{\theta_4}$	2.000	1.965	0.043	0.001	0.039	$\gamma_{\theta_4}$	0.054	0.044	0.015	< 0.001	0.013	
	$\mu_{\theta_5}$	2.300	2.256	0.055	0.001	0.051	$\gamma_{\theta_5}$	0.116	0.106	0.017	< 0.001	0.014	
	$\mu_{\theta_6}$	2.500	2.445	0.066	0.002	0.058	$\gamma_{\theta_6}$	–0.016	–0.023	0.022	< 0.001	0.020	
	$\sigma_{\theta_1}^2$	1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.807	0.004	< 0.001	0.106	
	$\sigma_{\theta_2}^2$	1.270	1.183	0.107	0.004	0.089	$\rho_{\theta_2}$	0.750	0.758	0.009	< 0.001	0.077	
	$\sigma_{\theta_3}^2$	0.900	0.837	0.072	0.001	0.063	$\rho_{\theta_3}$	0.700	0.715	0.016	< 0.001	0.064	
	$\sigma_{\theta_4}^2$	0.880	0.812	0.077	0.001	0.068	$\rho_{\theta_4}$	0.650	0.672	0.024	< 0.001	0.066	
	$\sigma_{\theta_5}^2$	0.700	0.650	0.061	0.001	0.050	$\rho_{\theta_5}$	0.600	0.633	0.043	0.001	0.080	
	$\sigma_{\theta_6}^2$	0.650	0.593	0.065	0.001	0.058							
	Sym	$\mu_{\theta_1}$	< 0.001	–	–	–	–	$\gamma_{\theta_1}$	0.800	–	–	–	–
		$\mu_{\theta_2}$	1.000	0.947	0.056	< 0.001	0.053	$\gamma_{\theta_2}$	0.546	–	–	–	–
$\mu_{\theta_3}$		1.400	1.331	0.070	< 0.001	0.069	$\gamma_{\theta_3}$	0.118	–	–	–	–	
$\mu_{\theta_4}$		2.000	1.845	0.156	< 0.001	0.155	$\gamma_{\theta_4}$	0.054	–	–	–	–	
$\mu_{\theta_5}$		2.300	2.105	0.196	0.001	0.195	$\gamma_{\theta_5}$	0.116	–	–	–	–	
$\mu_{\theta_6}$		2.500	2.279	0.223	0.001	0.221	$\gamma_{\theta_6}$	–0.016	–	–	–	–	
$\sigma_{\theta_1}^2$		1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.806	0.004	< 0.001	0.106	
$\sigma_{\theta_2}^2$		1.270	1.034	0.239	0.002	0.236	$\rho_{\theta_2}$	0.750	0.756	0.007	< 0.001	0.076	
$\sigma_{\theta_3}^2$		0.900	0.628	0.274	0.001	0.272	$\rho_{\theta_3}$	0.700	0.712	0.013	< 0.001	0.065	
$\sigma_{\theta_4}^2$		0.880	0.661	0.220	< 0.001	0.219	$\rho_{\theta_4}$	0.650	0.667	0.021	< 0.001	0.072	
$\sigma_{\theta_5}^2$		0.700	0.525	0.178	0.001	0.175	$\rho_{\theta_5}$	0.600	0.631	0.040	0.001	0.080	
$\sigma_{\theta_6}^2$		0.650	0.474	0.177	< 0.001	0.176							

Note: Acronym *Sym* stands for symmetric.

predictive distribution of replicated data at the time-point  $t$  is given by

$$p(\mathbf{y}_t^{rep} | \mathbf{y}_t^{obs}) = \int p(\mathbf{y}_t^{rep} | \boldsymbol{\vartheta}_t) p(\boldsymbol{\vartheta}_t | \mathbf{y}_t^{obs}) d\boldsymbol{\vartheta}_t,$$

where  $\boldsymbol{\vartheta}_t$  denotes the parameters at the time-point  $t$ . Then, to examine the extremeness of an observed discrepancy measure, a *Bayesian p-value* can be calculated, which is defined as

$$\mathbb{P}(D(\mathbf{y}_t^{rep} | \boldsymbol{\vartheta}_t) \geq D(\mathbf{y}_t^{obs} | \boldsymbol{\vartheta}_t) | \mathbf{y}_t^{obs}) = \int_{D(\mathbf{y}_t^{rep}) \geq D(\mathbf{y}_t^{obs})} p(\mathbf{y}_t^{rep} | \mathbf{y}_t^{obs}) d\mathbf{y}_t^{rep}.$$

In practice,  $M$  MCMC draws from the posterior distribution  $p(\boldsymbol{\vartheta}_t | \mathbf{y}_t^{obs})$  of  $\boldsymbol{\vartheta}_t$  and  $M$  draws from the distribution  $p(\mathbf{y}_t^{rep} | \boldsymbol{\vartheta}_t)$  are used to compute the proportion of the  $M$  replications for which  $D(\mathbf{y}_t^{rep})$  exceeds  $D(\mathbf{y}_t^{obs})$  to estimate of the Bayesian  $p$ -value. Values close to zero or one indicate model misfit.

For the LIRT model, the fit of the observed score distribution can be evaluated per time-point using the posterior predictive score distribution. Furthermore, to evaluate item fit, the following discrepancy measure was used,

$$D_i = \sum_l \frac{|P_{li}^O - P_{li}^E|}{P_{li}^E}, \tag{34}$$

**Table 6.** Simulation study: estimation results for the population parameters of the latent variable for six occasions.

		Normal simulated data											
		True	M.est	RMSE	VAR	ABias		True	M.est	RMSE	VAR	ABias	
Skew	$\mu_{\theta_1}$	0.000	–	–	–	–	$\gamma_{\theta_1}$	0.000	–	–	–	–	
	$\mu_{\theta_2}$	1.000	0.971	0.030	< 0.001	0.029	$\gamma_{\theta_2}$	0.000	–	–	–	–	
	$\mu_{\theta_3}$	1.400	1.360	0.043	< 0.001	0.040	$\gamma_{\theta_3}$	0.000	–	–	–	–	
	$\mu_{\theta_4}$	2.000	1.922	0.083	0.001	0.078	$\gamma_{\theta_4}$	0.000	–	–	–	–	
	$\mu_{\theta_5}$	2.300	2.204	0.102	0.001	0.096	$\gamma_{\theta_5}$	0.000	–	–	–	–	
	$\mu_{\theta_6}$	2.500	2.394	0.113	0.002	0.106	$\gamma_{\theta_6}$	0.000	–	–	–	–	
	$\sigma_{\theta_1}^2$	1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.807	0.004	< 0.001	0.106	
	$\sigma_{\theta_2}^2$	1.270	1.144	0.148	0.007	0.128	$\rho_{\theta_2}$	0.750	0.758	0.009	< 0.001	0.077	
	$\sigma_{\theta_3}^2$	0.900	0.801	0.104	0.001	0.099	$\rho_{\theta_3}$	0.700	0.715	0.016	< 0.001	0.064	
	$\sigma_{\theta_4}^2$	0.880	0.780	0.119	0.005	0.102	$\rho_{\theta_4}$	0.650	0.672	0.024	< 0.001	0.066	
	$\sigma_{\theta_5}^2$	0.700	.608	0.098	0.001	0.092	$\rho_{\theta_5}$	0.600	0.633	0.043	0.001	0.080	
	$\sigma_{\theta_6}^2$	0.650	0.565	0.088	0.001	0.085							
	Sym	$\mu_{\theta_1}$	0.000	–	–	–	–	$\gamma_{\theta_1}$	0.000	–	–	–	–
		$\mu_{\theta_2}$	1.000	0.947	0.056	< 0.001	0.053	$\gamma_{\theta_2}$	0.000	–	–	–	–
$\mu_{\theta_3}$		1.400	1.331	0.070	< 0.001	0.069	$\gamma_{\theta_3}$	0.000	–	–	–	–	
$\mu_{\theta_4}$		2.000	1.845	0.156	< 0.001	0.155	$\gamma_{\theta_4}$	0.000	–	–	–	–	
$\mu_{\theta_5}$		2.300	2.105	0.196	0.001	0.195	$\gamma_{\theta_5}$	0.000	–	–	–	–	
$\mu_{\theta_6}$		2.500	2.279	0.223	0.001	0.221	$\gamma_{\theta_6}$	0.000	–	–	–	–	
$\sigma_{\theta_1}^2$		1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.807	0.004	< 0.001	0.107	
$\sigma_{\theta_2}^2$		1.270	1.034	0.239	0.002	0.236	$\rho_{\theta_2}$	0.750	0.758	0.008	< 0.001	0.075	
$\sigma_{\theta_3}^2$		0.900	0.628	0.274	0.001	0.272	$\rho_{\theta_3}$	0.700	0.712	0.012	< 0.001	0.064	
$\sigma_{\theta_4}^2$		0.880	0.661	0.220	< 0.001	0.219	$\rho_{\theta_4}$	0.650	0.670	0.022	< 0.001	0.065	
$\sigma_{\theta_5}^2$		0.700	0.525	0.178	0.001	0.175	$\rho_{\theta_5}$	0.600	0.636	0.039	< 0.001	0.078	
$\sigma_{\theta_6}^2$		0.650	0.474	0.177	< 0.001	0.176							

Note: Acronym *Sym* stands for symmetric.

where  $P_{li}^O$  and  $P_{li}^E$  denote, respectively, the observed and expected proportion of respondents with score  $l$ , who scored correctly to item  $i$ , where  $L$  denotes the maximum score.

### Model comparison

Model comparison statistic can be used to select the most appropriate covariance structure. The *Deviance information criteria* (DIC), the expected value of *Akaike’s information criteria* (EAIC) and the *Bayesian information criteria* (EBIC) can be used to compare the fit of different covariance patterns. The information criteria are based on a penalty term,  $\rho_D = \overline{D(\vartheta)} - D(\vartheta)$ . Let  $\vartheta$  denote a set of parameters of interest and  $D(\vartheta) = -2(\text{Log likelihood})$ . It follows that

$$D(\vartheta) = -2\text{Log}(L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta})P(\boldsymbol{\theta}_{..} | \boldsymbol{\eta}_{\theta})), \tag{35}$$

where  $L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta})$  represents the likelihood of the observed response data with  $P(\boldsymbol{\theta}_{..} | \boldsymbol{\eta}_{\theta})$  defined in Equation (33).

In practice, having  $M$  MCMC draws from the joint posterior distribution,  $\overline{D(\vartheta)}$  is evaluated using the posterior mean estimates, and the posterior mean deviance as  $\overline{D(\vartheta)} \approx \sum_m D(\vartheta^{(m)})/M$ . The statistics are computed according to:

$$\widehat{\text{DIC}} = D(\bar{\vartheta}) + 2\rho_D$$

**Table 7.** Simulation study: estimation results for the population parameters of the latent variable for six occasions.

		Skew-t simulated data										
		True	M.est	RMSE	VAR	ABias	True	M.est	RMSE	VAR	ABias	
Skew	$\mu_{\theta_1}$	0.000	–	–	–	–	$\gamma_{\theta_1}$	–	–	–	–	
	$\mu_{\theta_2}$	1.000	1.048	0.180	0.033	0.095	$\gamma_{\theta_2}$	–	–	–	–	
	$\mu_{\theta_3}$	1.400	1.456	0.223	0.052	0.120	$\gamma_{\theta_3}$	–	–	–	–	
	$\mu_{\theta_4}$	2.000	2.039	0.273	0.081	0.168	$\gamma_{\theta_4}$	–	–	–	–	
	$\mu_{\theta_5}$	2.300	2.334	0.295	0.095	0.188	$\gamma_{\theta_5}$	–	–	–	–	
	$\mu_{\theta_6}$	2.500	2.527	0.300	0.099	0.195	$\gamma_{\theta_6}$	–	–	–	–	
	$\sigma_{\theta_1}^2$	1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.793	0.035	0.001	0.092
	$\sigma_{\theta_2}^2$	1.270	1.096	0.209	0.015	0.174	$\rho_{\theta_2}$	0.750	0.744	0.028	0.001	0.062
	$\sigma_{\theta_3}^2$	0.900	0.736	0.210	0.019	0.164	$\rho_{\theta_3}$	0.700	0.698	0.030	0.001	0.057
	$\sigma_{\theta_4}^2$	0.880	0.730	0.203	0.021	0.159	$\rho_{\theta_4}$	0.650	0.658	0.039	0.002	0.079
	$\sigma_{\theta_5}^2$	0.700	0.567	0.163	0.010	0.133	$\rho_{\theta_5}$	0.600	0.620	0.045	0.002	0.094
	$\sigma_{\theta_6}^2$	0.65	.516	0.171	0.013	0.134						
Sym	$\mu_{\theta_1}$	0.000	–	–	–	–	$\gamma_{\theta_1}$	–	–	–	–	
	$\mu_{\theta_2}$	1.000	1.364	0.432	0.060	0.364	$\gamma_{\theta_2}$	–	–	–	–	
	$\mu_{\theta_3}$	1.400	1.873	0.569	0.112	0.473	$\gamma_{\theta_3}$	–	–	–	–	
	$\mu_{\theta_4}$	2.000	2.545	0.702	0.217	0.545	$\gamma_{\theta_4}$	–	–	–	–	
	$\mu_{\theta_5}$	2.300	2.875	0.758	0.271	0.575	$\gamma_{\theta_5}$	–	–	–	–	
	$\mu_{\theta_6}$	2.500	3.078	0.784	0.312	0.578	$\gamma_{\theta_6}$	–	–	–	–	
	$\sigma_{\theta_1}^2$	1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.760	0.063	0.002	0.074
	$\sigma_{\theta_2}^2$	1.270	0.996	0.489	0.182	0.444	$\rho_{\theta_2}$	0.750	0.714	0.069	0.004	0.056
	$\sigma_{\theta_3}^2$	0.900	0.596	0.503	0.178	0.470	$\rho_{\theta_3}$	0.700	0.660	0.114	0.013	0.089
	$\sigma_{\theta_4}^2$	0.880	0.505	0.437	0.055	0.426	$\rho_{\theta_4}$	0.650	0.610	0.184	0.036	0.140
	$\sigma_{\theta_5}^2$	0.700	0.393	0.369	0.047	0.360	$\rho_{\theta_5}$	0.600	0.475	0.491	0.251	0.266
	$\sigma_{\theta_6}^2$	0.650	0.335	0.361	0.035	0.352						

Note: Acronym *Sym* stands for symmetric

**Table 8.** Simulation study: estimation results for the latent variable and item parameters across six time points with dropouts.

		Statistic			
Parameter		Corr	RMSE	VAR	ABias
AD	Latent variable	0.970	0.300	00.760	0.250
	Discrimination	0.934	0.111	0.011	0.093
	Difficulty	0.997	0.128	0.017	0.101
	Guessing	–	0.027	0.001	0.023

$$\widehat{EAIC} = \overline{D(\vartheta)} + 2\rho_D$$

$$\widehat{EBIC} = \overline{D(\vartheta)} + 2\log(N_I),$$

where  $N_I$  represents the number of subjects times the number of items.

### 7. The Brazilian school development study

The real data study concerns a large-scale assessment program funded by the Brazilian Federal Government know as the School Development Program. It is aimed at monitoring teaching quality in Brazilian public schools. A more detailed description of the data can

**Table 9.** Simulation study: estimation results for the population parameters of the latent variable for six occasions with dropouts

	True	M.est	RMSE	VAR	ABias		True	M.est	RMSE	VAR	ABias
$\mu_{\theta_1}$	0.000	–	–	–	–	$\gamma_{\theta_1}$	0.800	0.794	0.042	0.002	0.032
$\mu_{\theta_2}$	1.000	0.995	0.042	0.002	0.034	$\gamma_{\theta_2}$	0.546	0.521	0.056	0.003	0.039
$\mu_{\theta_3}$	1.400	1.400	0.055	0.003	0.042	$\gamma_{\theta_3}$	0.181	0.217	0.042	0.001	0.036
$\mu_{\theta_4}$	2.000	2.019	0.091	0.009	0.061	$\gamma_{\theta_4}$	0.275	0.222	0.058	0.001	0.052
$\mu_{\theta_5}$	2.300	2.327	0.106	0.012	0.073	$\gamma_{\theta_5}$	0.221	0.147	0.079	0.001	0.074
$\mu_{\theta_6}$	2.500	2.552	0.107	0.010	0.075	$\gamma_{\theta_6}$	–0.039	0.034	0.077	0.001	0.073
$\sigma_{\theta_1}^2$	1.000	–	–	–	–	$\rho_{\theta_1}$	0.810	0.811	0.004	< 0.001	0.075
$\sigma_{\theta_2}^2$	1.270	1.305	0.115	0.013	0.091	$\rho_{\theta_2}$	0.890	0.866	0.024	< 0.001	0.063
$\sigma_{\theta_3}^2$	0.900	0.963	0.118	0.011	0.097	$\rho_{\theta_3}$	0.930	0.888	0.042	< 0.001	0.058
$\sigma_{\theta_4}^2$	0.880	0.942	0.122	0.012	0.101	$\rho_{\theta_4}$	0.730	0.793	0.063	< 0.001	0.086
$\sigma_{\theta_5}^2$	0.700	0.781	0.087	0.001	0.081	$\rho_{\theta_5}$	0.890	0.837	0.054	< 0.001	0.060
$\sigma_{\theta_6}^2$	0.650	0.738	0.097	0.002	0.088						

be found in [2]. The longitudinal study was focused on assessing children’s ability in math and Portuguese. The data concerning the math part was used in this analysis.

A total of 1987 students from public schools from different regions of the country were followed from fourth to eighth grade of primary school. The students made a math test at each of six occasions (1999/April, 1999/November, 2000/November, 2001/November, 2002/November and 2003/November). A preliminary analysis revealed that most of the items administered at the sixth occasion (i.e. 2003/November) did not fit properly. The fitted items showed high difficulty estimates and discrimination estimates were around zero. This results indicated inconsistencies in the test design. Therefore, data of the last occasion was not used in the current longitudinal analysis.

The set of 142 items administered at the first five occasions were used for the analysis. Table 10 presents the design of the tests, and shows the number of items per test and the number of common items. For example, test 1 had 34 items, and test 2 has 38 items, with 10 items from test 1 and so forth.

In the first analysis, there were no dropouts considered, and the response data of the 1,987 children across the five test occasions was used. The LIRT model with an unstructured covariance matrix was fitted to the data to explore the serial dependence between latent variables over time. The estimates of the unstructured covariance matrix are given in Equation (36). The latent variable variances are given on the main diagonal, correlations on the upper triangular, and covariances on the lower triangular. The estimates show a time-heteroscedastic pattern for the variances and a serial correlation structure. Notice

**Table 10.** The test design of the Brazilian School Development Program

	Test 1	Test 2	Test 3	Test 4	Test 5
Test 1	34	10	5	1	0
Test 2	10	38	5	3	0
Test 3	5	5	36	5	3
Test 4	1	3	5	34	7
Test 5	0	0	3	7	39



**Table 11.** Estimated information criteria to select the optimal latent variable dependence structure.

	$\rho_D$	DIC	EAIC	EBIC
ARH(1)	10,361.20	419,303.6	429,664.8	555,653.3
ARMAH(1, 1)	10,516.62	419,911.8	430,428.4	558,306.7
AD	10,474.12	419,307.9	429,782.0	557,143.6
HT	10489.82	419,703.6	430,193.4	557746.0

**Table 12.** Estimated information criteria to compare the skewed against the symmetric multivariate latent variable distribution.

	$\rho_D$	DIC	EAIC	EBIC
Skew ARH(1)	103,61.20	419,303.6	429,664.8	555,653.3
Symmetric ARH(1)	11,732.36	421,270.1	433,002.5	575,663.8

that correlations are high and decay slowly.

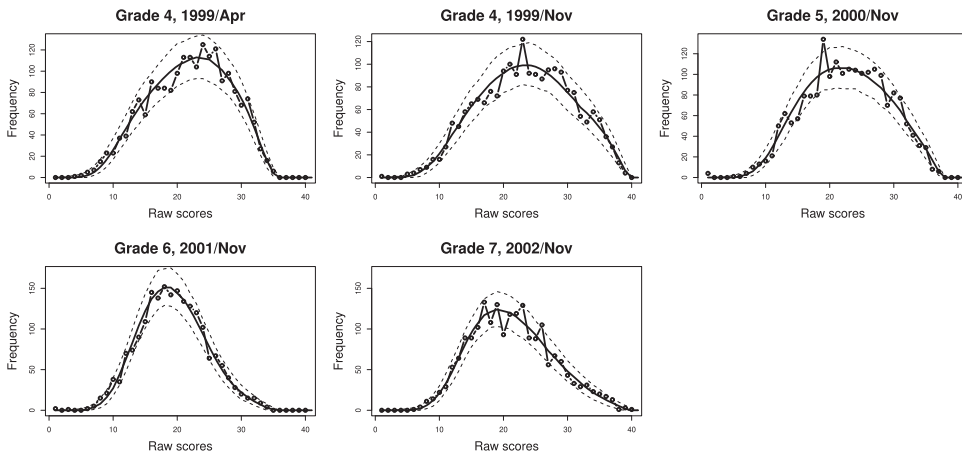
$$\begin{pmatrix} 1.000 & 0.868 & 0.772 & 0.711 & 0.660 \\ 0.744 & 0.733 & 0.885 & 0.814 & 0.755 \\ 0.571 & 0.560 & 0.547 & 0.915 & 0.846 \\ 0.710 & 0.695 & 0.675 & 0.994 & 0.922 \\ 0.806 & 0.788 & 0.764 & 1.122 & 1.488 \end{pmatrix} \quad (36)$$

In order to select the optimal covariance structure for the latent variables, the information criteria were computed for each covariance pattern. The estimated statistics indicated that the better fit was obtained with the ARH(1) covariance pattern, followed by the AD structure, see Table 11. Considering the LIRT with dependence structure ARH(1), the skewness assumption of the multivariate normal latent variable distribution was also evaluated. The results in Table 12 shows that the skewed error distribution improved the fit of the model, when comparing it to a symmetric error distribution.

In Figure 3, the observed and predicted scores for each time point are plotted with the respective 95% equi-tailed credibility intervals. It can be seen that the majority of the observed scores are well within the corresponding credible intervals, indicating a good model fit. Table 13 presents the population estimates of a skewed and symmetric latent variable distribution with the ARH(1) covariance structure. It is apparent that the population variances are smaller under a skewed latent variable distribution, which shows that the symmetric assumption leads to an overestimation of the latent variable variance. There is less variance between the latent variable estimates when assuming asymmetric latent variable distributions.

The estimates of the population means show an increase of the average ability across grades. The estimated correlation is around 0.89, indicating a high correlation between any pair of consecutive grades. Furthermore, the latent variable distributions at the first three time-points present moderate asymmetry. The skewed LIRT model describes properly the change in performance across students and occasions.

Figure 4 presents the estimates of discrimination parameters, under both skewed and symmetric latent variable distributions with the ARH(1) covariance matrix. The vertical bars stands for the 95% credibility intervals. In general, it can be concluded that all test items present good discriminating power (estimates higher than 0.6). Furthermore, the



**Figure 3.** Observed and predicted score distributions with 95% credibility intervals.

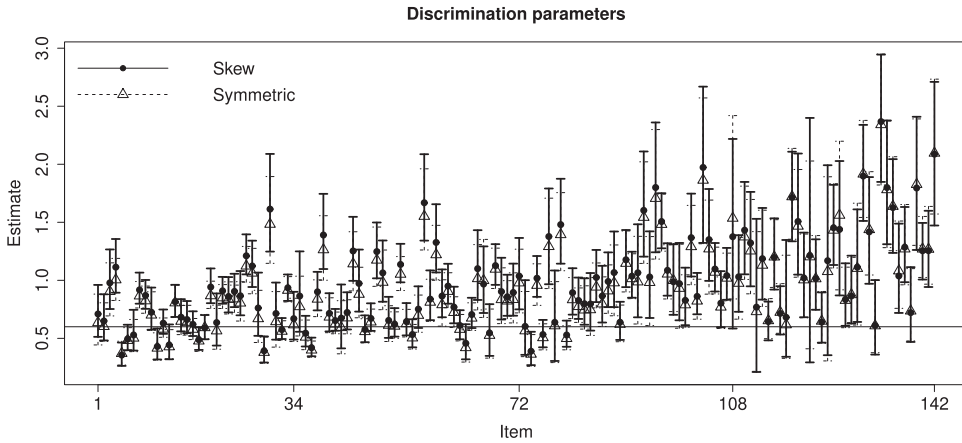
**Table 13.** Estimates of the population parameters under ARH(1) with both symmetric and skewed latent variable distributions.

	Symmetric			Skew		
	Mean	SD	CI (95%)	Mean	SD	CI (95%)
$\mu_{\theta_1}$	0.000	–	–	0.000	–	–
$\mu_{\theta_2}$	0.232	0.036	[0.160, 0.303]	0.230	0.037	[0.155, 0.300]
$\mu_{\theta_3}$	0.766	0.038	[0.689, 0.840]	0.745	0.036	[0.677, 0.814]
$\mu_{\theta_4}$	1.435	0.051	[1.342, 1.539]	1.390	0.050	[1.300, 1.489]
$\mu_{\theta_5}$	1.691	0.059	[1.582, 1.812]	1.632	0.057	[1.528, 1.737]
$\sigma_{\theta_1}^2$	1.000	–	–	1.000	–	–
$\sigma_{\theta_2}^2$	1.426	0.173	[1.152, 1.807]	1.306	0.162	[1.025, 1.667]
$\sigma_{\theta_3}^2$	1.243	0.151	[0.980, 1.554]	1.125	0.161	[0.831, 1.451]
$\sigma_{\theta_4}^2$	0.622	0.079	[0.488, 0.798]	0.574	0.074	[0.448, 0.744]
$\sigma_{\theta_5}^2$	0.407	0.053	[0.316, 0.522]	0.423	0.065	[0.323, 0.569]
$\gamma_{\theta_1}$	–	–	–	–0.426	0.076	[–0.562, –0.271]
$\gamma_{\theta_2}$	–	–	–	–0.319	0.056	[–0.418, –0.199]
$\gamma_{\theta_3}$	–	–	–	–	0.039	[–0.304, –0.152]
$\gamma_{\theta_4}$	–	–	–	–0.168	0.032	[–0.231, –0.104]
$\gamma_{\theta_5}$	–	–	–	–0.146	0.030	[–0.205, –0.092]
$\rho_{\theta_1}$	0.890	0.006	[0.878, 0.901]	0.891	0.006	[0.879, 0.902]

discrimination parameter estimates tend to be slightly higher under the skewed distribution in comparison to a symmetric distribution. Similar graphs, concerning to difficulty and guessing parameters estimates, are shown in the supplementary material.

### Artificial dropouts

The real data study did not had dropouts, therefore artificial dropouts were created by randomly excluding response data of some children. This was done in a similar way as the simulation study. The following number of children per time-point were assessed,  $n_1 = 1987$ ,  $n_2 = 1937$ ,  $n_3 = 1937$ ,  $n_4 = 1917$  and  $n_5 = 1917$ , where along the way some



**Figure 4.** Posterior means and 95% central credibility intervals for discrimination parameters.

children were no longer tested. The objective was to compare the estimates of the parameters under both scenarios and to examine the influence of the dropouts.

The skewed LIRT model with an ARH(1) covariance matrix was fitted. Table 14 shows the estimates of the population parameters for a skewed and symmetric multivariate latent variable distribution. The estimates of the population variances are slightly higher than those of Table 13, as well as the respective posterior standard deviations. The estimates are less accurate when children dropout, since data information is lost while estimating the same number of parameters. The estimated correlation and Pearson’s skewness coefficients are very close to the ones without dropouts. The dropouts hardly influenced the population

**Table 14.** Estimates of the population parameters under ARH(1) with symmetric and skewed multivariate latent variable distributions and dropouts.

	Symmetric			Skew		
	Mean	SD	CI (95%)	Mean	SD	CI (95%)
$\mu_{\theta_1}$	0.000	–	–	0.000	–	–
$\mu_{\theta_2}$	0.226	0.041	[0.140, 0.302]	0.224	0.037	[0.149, 0.292]
$\mu_{\theta_3}$	0.765	0.041	[0.684, 0.844]	0.750	0.037	[0.678, 0.819]
$\mu_{\theta_4}$	1.435	0.056	[1.328, 1.552]	1.406	0.049	[1.309, 1.502]
$\mu_{\theta_5}$	1.681	0.063	[1.563, 1.811]	1.657	0.054	[1.552, 1.762]
$\sigma_{\theta_1}^2$	1.000	–	–	1.000	–	–
$\sigma_{\theta_2}^2$	1.459	0.185	[1.151, 1.872]	1.346	0.169	[1.042, 1.702]
$\sigma_{\theta_3}^2$	1.328	0.177	[1.035, 1.720]	1.179	0.163	[0.898, 1.546]
$\sigma_{\theta_4}^2$	0.572	0.070	[0.445, 0.722]	0.604	0.085	[0.443, 0.780]
$\sigma_{\theta_5}^2$	0.392	0.053	[0.300, 0.509]	0.445	0.066	[0.332, 0.590]
$\gamma_{\theta_1}$	–	–	–	–0.424	0.076	[–0.567, –0.265]
$\gamma_{\theta_2}$	–	–	–	–0.312	0.055	[–0.415, –0.196]
$\gamma_{\theta_3}$	–	–	–	–0.220	0.041	[–0.297, –0.135]
$\gamma_{\theta_4}$	–	–	–	–0.179	0.037	[–0.247, –0.106]
$\gamma_{\theta_5}$	–	–	–	–0.151	0.032	[–0.211, –0.084]
$\rho_{\theta}$	0.892	0.006	[0.880, 0.904]	0.893	0.006	[0.881, 0.904]

estimates, which was to be expected, since the dropouts were randomly selected. Furthermore, results of the predicted scores and item parameter estimates are also similar to those without any dropouts, and they are shown in the supplementary material.

### 8. Concluding remarks

A general class of longitudinal three-parameter IRT models with latent variable covariance patterns was proposed. In our approach, a general Cholesky decomposition was used to model multivariate serial dependence between the latent variables across occasions. A wide range of (serial) dependence structures has been described, while also allowing for asymmetric multivariate distributions.

An MCMC algorithm based on the FFBS and Metropolis-Hasting steps was developed for parameter estimation. It showed to be efficient in terms of parameter recovery, which was investigated through a simulation study. The results also indicate that the skew model proposed outperforms the symmetric model when the latent distributions are asymmetric and asymmetric/heavy-tailed.

The MCMC algorithm also reduced the number of parameters to be estimated independently of the number of time-points, when comparing to the multivariate approach of [2,33]. Computational time was evaluated for 10,000 iterations of the MCMC algorithm for both skew and symmetric models and different covariance matrices, considering the setting described in Section 5. The computer used has the following settings: i3-4160 CPU 3.60 GHz, RAM 8.00 GB, OS 64bits. Computational times are presented in Table 15. The computation time under skew model was around 3 h and a little bit less when the symmetric model is fitted. It suggest that the estimation of the skewness coefficients have not increase, considerably, the computational coast.

Furthermore, the real data example concerning the Brazilian school development study, showed a good performance of the LIRT to examine longitudinal growth. Model fit assessment tools were considered to investigate the fit of each model. Estimation results showed that the latent variables highly correlated across time-points. Also, the majority of the marginal latent variable distributions showed asymmetric behaviour.

In addition, the LIRT model with an asymmetric latent variable distribution showed a better fit to the data than the (restricted) symmetric distribution. Our simulation study showed that the accuracy of the estimates of the discrimination parameters, the population parameter variances and the skewness parameters were affected by dropouts. In conclusion, our approach showed promising results to analyze longitudinal IRT data with complex serial dependence between the latent variables. In future research, the object is to explore model extensions focused on asymmetric growth curves, multiple group structures, and

**Table 15.** Computational time (in hours) for skew and symmetric models under different covariance matrices.

	ARH(1)	ARMAH(1, 1)	AD	HT
Skew	3 h04 m	3 h01 m	3 h00 m	2 h53 m
Symmetric	2 h55 m	2h56	2 h56 m	2 h53 m

regression patterns for the multivariate latent variable distribution. Furthermore, the modelling framework needs to be extended to handle mixed response types, thereby including a wider variety of IRT models.

## Acknowledgements

Authors are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from Brazil, for the financial support (Grants: 140415/2014-9 and 142486/2015-9) of the doctorate of the first author under guidance of the second author, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support (PDSE-99999.003757/2014-05) of the internship abroad of the first author under supervision of the third author.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The financial support of this study was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from Brazil [grant numbers 140415/2014-9 and 142486/2015], and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [grant number PDSE-99999.003757/2014-05].

## ORCID

José Roberto Silva dos Santos  <http://orcid.org/0000-0003-4769-9149>

Caio Lucidius Naberezny Azevedo  <http://orcid.org/0000-0001-9535-292X>

Jean-Paul Fox  <http://orcid.org/0000-0002-0058-1496>

## References

- [1] Andrade D, Tavares H. Item response theory for longitudinal data: population parameter estimation. *J Multivar Anal.* 2005;95(1):1–22. doi:10.1016/j.jmva.2004.07.005
- [2] Azevedo CLN, Fox JP, Andrade DF. Bayesian longitudinal item response modeling with restricted covariance pattern structures. *Stat Comput.* 2016;26(1):443–460. doi:10.1007/s11222-014-9518-5
- [3] Andersen EB. Estimating latent correlations between repeated testings. *Psychometrika.* 1985;50(1):3–16. doi:10.1007/BF02294143
- [4] Embretson SE. A multidimensional latent trait model for measuring learning and change. *Psychometrika.* 1991;56(3):495–515. doi:10.1007/BF02294487
- [5] Dunson DB. Dynamic latent trait models for multidimensional longitudinal data. *J Am Stat Assoc.* 2003;98(463):555–563. doi:10.1198/016214503000000387
- [6] Vasdekis VGS, Cagnone S, Moustaki I. A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika.* 2012;77(3):425–441. doi:10.1007/s11336-012-9264-6
- [7] Jeon M, Rabe-Hesketh S. An autoregressive growth model for longitudinal item analysis. *Psychometrika.* 2016;81(3):830–850. doi:10.1007/s11336-015-9489-2
- [8] Tavares HR, Andrade DF. Item response theory for longitudinal data: item and population ability parameters estimation. *TEST.* 2006;15(1):97–123. doi:10.1007/BF02595420
- [9] Azevedo CL, Bolfarine H, Andrade DF. Bayesian inference for a skew-normal IRT model under the centred parameterization. *Comput Stat Data Anal.* 2011;55(1):353–365. doi:10.1016/j.csda.2010.05.003

- [10] Azevedo CL, Bolfarine H, Andrade DF. 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.* **2012**;82(11):1679–1699. doi:10.1080/00949655.2011.591798
- [11] Baker FB, Kim SH. *Item response theory : parameter estimation techniques*, 2nd ed., New York, NY: Marcel Dekker Inc; **2004**.
- [12] Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika.* **1999**;86(3):677–690. doi:10.1093/biomet/86.3.677
- [13] Núñez Antón V, Zimmerman DL. Modeling nonstationary longitudinal data. *Biometrics.* **2000**;56(3):699–705. doi:10.1111/j.0006-341X.2000.00699.x
- [14] Azzalini A. A class of distributions which includes the normal ones. *Scand J Stat.* **1985**; 12:171–178.
- [15] Henze N. A probalistic representation of the skew-normal distribution. *Scand J Stat.* **1986**; 13:271–275.
- [16] Azevedo CL, Andrade DF, Fox JP. A Bayesian generalized multiple group IRT model with model-fit assessment tools. *Comput Stat Data Anal.* **2012**;56(12):4399–4412. doi:10.1016/j.csda.2012.03.017
- [17] Santos JRS, Azevedo CLN, Bolfarine H. A multiple group item response theory model with centred skew normal latent trait distributions under a Bayesian framework. *J Appl Stat.* **2013**;40(10):2129–2149. doi:10.1080/02664763.2013.807331
- [18] Padilla JL, Azevedo CL, Lachos VH. Multidimensional multiple group IRT models with skew normal latent trait distributions. *J Multivar Anal.* **2018**;167:250–268. doi:10.1016/j.jmva.2018.05.003
- [19] Pewsey A. Problems of inference for Azzalini’s skewnormal distribution. *J Appl Stat.* **2000**;27(7):859–870. doi:10.1080/02664760050120542
- [20] Carter CK, Kohn R. On ibbs sampling for state space models. *Biometrika.* **1994**;81:541–553. doi:10.1093/biomet/81.3.541
- [21] Frühwirth-Schnatter S. Data augmentation and dynamic linear models. *J Time Seri Anal.* **1994**;15:183–202. doi:10.1111/j.1467-9892.1994.tb00184.x
- [22] Genton M. *Skew-elliptical distributions and their applications: A journey beyond normality*. 1st ed., New York: Chapman & Hall; **2004**. doi:10.1201/9780203492000
- [23] Bayes CL, Branco MD. Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Braz J Prob Stat.* **2007**;21(2):141–163.
- [24] Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics.* **1986**;42(4):805–820. doi:10.2307/2530695
- [25] Zimmerman D, Núñez Antón V. *Antependence models for longitudinal data*. 1 ed.. New York: Chapman and Hall/CRC; **2009**.
- [26] Béguin AA, Glas CAW. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika.* **2001**;66(4):541–561. doi:10.1007/BF02296195
- [27] Gamerman D, Lopes HF. *Markov chain Monte Carlo : stochastic simulation for Bayesian inference*. 2nd ed., London: Chapman & Hall/CRC; **2006**.
- [28] R Core Team, *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (2020). <https://www.R-project.org/>.
- [29] Azzalini A, Capitanio A. *The skew-normal and related families*. New York: Cambridge University Press; **2014**. (IMS Monographs series), doi:10.1017/CBO9781139248891.
- [30] Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat.* **1998**;7(4):434–455. doi:10.1080/10618600.1998.10474787
- [31] Sinharay S. A Bayesian item fit analysis for unidimensional item response theory models. *Br J Math Stat Psychol.* **2006**;59:429–449. doi:10.1348/000711005X66888
- [32] Sinharay S, Johnson MS, Stern H. Posterior predictive assessment of item response theory models. *Appl Psychol Meas.* **2006**;30(4):298–321. doi:10.1177/0146621605285517
- [33] Azevedo CL, Fox JP, Andrade DF. Longitudinal multiple-group IRT modelling: covariance pattern selection using MCMC and RJMCMC. *Int J Quant Res Educ.* **2015**;2(3-4):213–243. doi:10.1504/IJQRE.2015.071737

## Appendices

### Appendix 1. Proof of Lemma 3.1

**Proof:** According to Equation (5), we can write,  $(\theta_j - \mu_\theta) = \mathbf{L}^{-1} \boldsymbol{\varepsilon}_j$ .

By definition, matrix  $\mathbf{L}$  is a lower triangular matrix with ones on the main diagonal (see matrix (6)). Therefore,  $\mathbf{L}^{-1}$  has the same form. Then, we can write, for  $t = 1, \dots, T$ ,

$$\theta_{jt} - \mu_{\theta_t} = (\varepsilon_{jt} + \sum_{k=1}^{t-1} l_{tk} \varepsilon_{jk}) \Leftrightarrow (\theta_{jt} - \mu_{\theta_t})^3 = (\varepsilon_{jt} + \sum_{k=1}^{t-1} l_{tk} \varepsilon_{jk})^3.$$

Taking expectations on both sides of the expression. It follows that,

$$\mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] = \mathbb{E}[(\varepsilon_{jt} + \sum_{k=1}^{t-1} l_{tk} \varepsilon_{jk})^3].$$

The multinomial theorem is used to rewrite the term on the right-hand side:

$$\mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] = \mathbb{E} \left[ \sum_{k_1+k_2+\dots+k_t=3} \binom{3}{k_1, k_2, \dots, k_t} \varepsilon_{jt}^{k_t} \prod_{1 \leq m \leq t-1} l_{tm}^{k_m} \varepsilon_{jm}^{k_m} \right],$$

Indices  $k_1$  through  $k_t$  are non-negative integers, such that the sum of all  $k_m$  equals three. For each term in the expansion, the exponents of  $\varepsilon_{jm}$  adds up to three. When expanding again the term on the right-hand side,

$$\begin{aligned} & \mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] \\ &= \mathbb{E} \left[ \varepsilon_{jt}^3 + \sum_{k=1}^{t-1} l_{tk}^3 \varepsilon_{jk}^3 + \sum_{\substack{k_1+k_2+\dots+k_t=3, \\ k_m \in \{0,1,2\}}} \binom{3}{k_1, k_2, \dots, k_t} \varepsilon_{jt}^{k_t} \prod_{1 \leq m \leq t-1} l_{tm}^{k_m} \varepsilon_{jm}^{k_m} \right] \\ &= \mathbb{E}(\varepsilon_{jt}^3) + \sum_{k=1}^{t-1} l_{tk}^3 \mathbb{E}(\varepsilon_{jk}^3) + \sum_{\substack{k_1+k_2+\dots+k_t=3, \\ k_m \in \{0,1,2\}}} \binom{3}{k_1, k_2, \dots, k_t} \mathbb{E}(\varepsilon_{jt}^{k_t}) \prod_{1 \leq m \leq t-1} l_{tm}^{k_m} \mathbb{E}(\varepsilon_{jm}^{k_m}). \end{aligned}$$

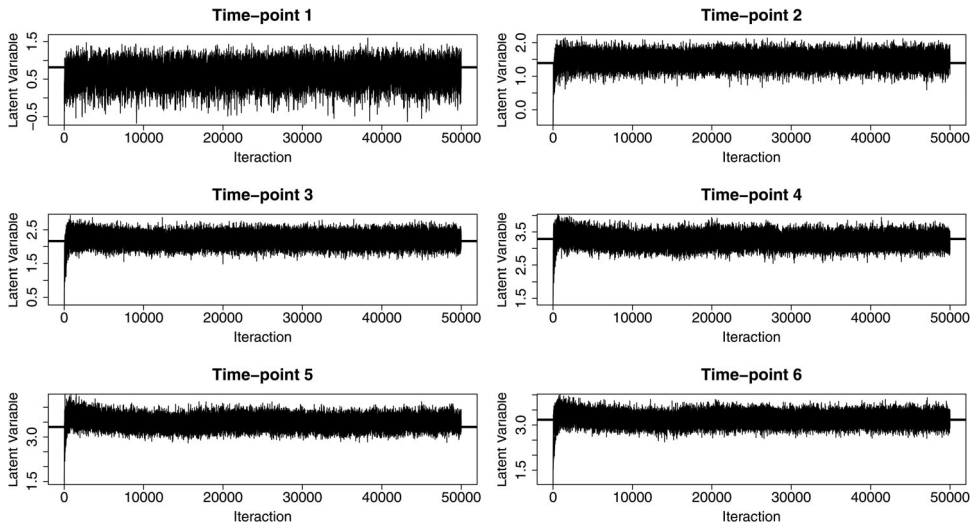
The  $\varepsilon_{jt}$  are independently distributed random variables with mean zero. Thus, the last term on the right-hand side is equal to zero. It follows that,

$$\begin{aligned} \mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] &= \mathbb{E}(\varepsilon_{jt}^3) + \sum_{k=1}^{t-1} l_{tk}^3 \mathbb{E}(\varepsilon_{jk}^3) \\ &= d_t^{3/2} \gamma_{\varepsilon_t} + \sum_{k=1}^{t-1} l_{tk}^3 d_k^{3/2} \gamma_{\varepsilon_k}, \end{aligned}$$

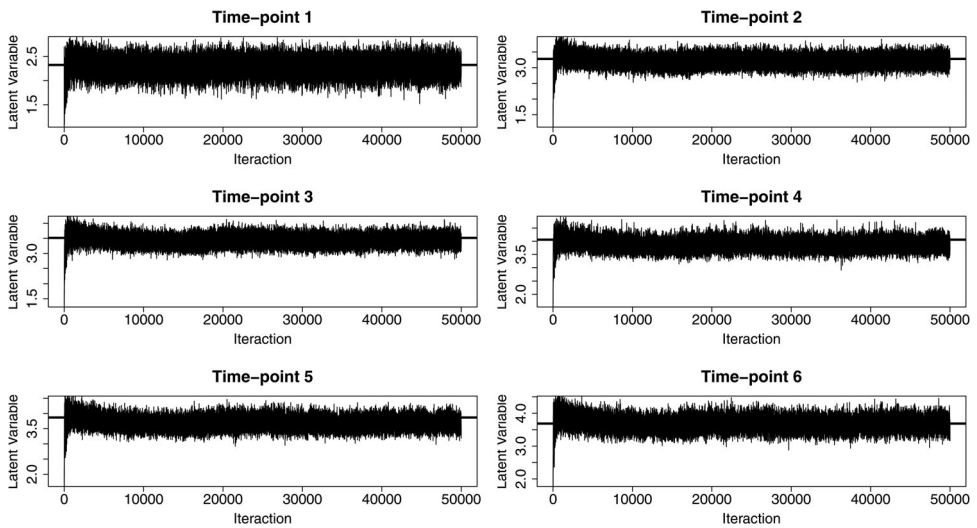
where  $\mathbb{E}(\varepsilon_{jt}^3) = d_t^{3/2} \gamma_{\varepsilon_t}$ , which follows directly from the definition of the Pearson's skewness coefficient. ■

### Appendix 2. Convergence monitoring

Below we present some trace plots related to the MCMC runs, considering the setup described in Section 5. The results for latent variables and item parameters were randomly selected from all possibilities (a total of 1500 subjects in 6 time-points and 120 items with 3 parameters). The horizontal lines in the plots represents the true value of the parameters. In each trace-plot three chains based on three different sets of initial values are presented. The results on each set of initial values are distinguished by different aspects of the lines (solid line, dashed line and dotted-dashed line).

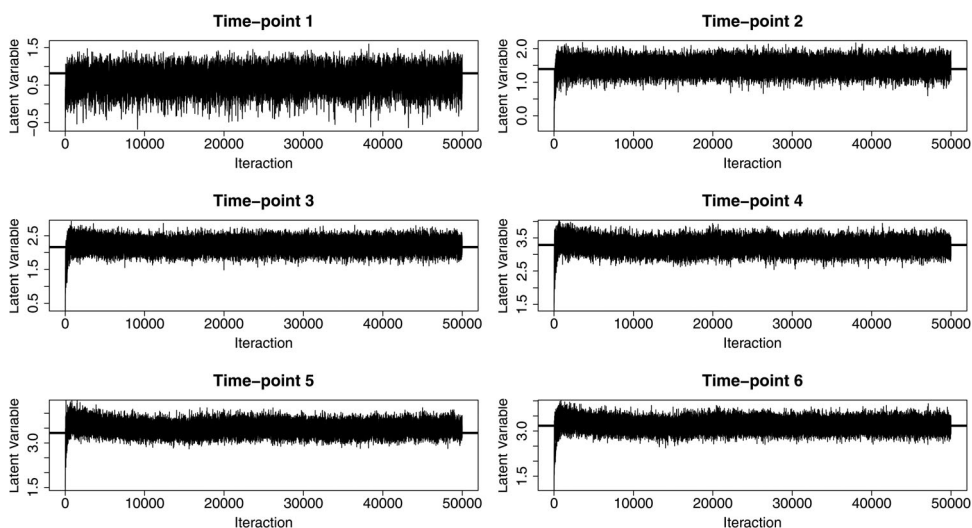


**Figure A1.** Traces-plots of the latent variables of the subject 420. Multivariate PSRF resulted 1.01.

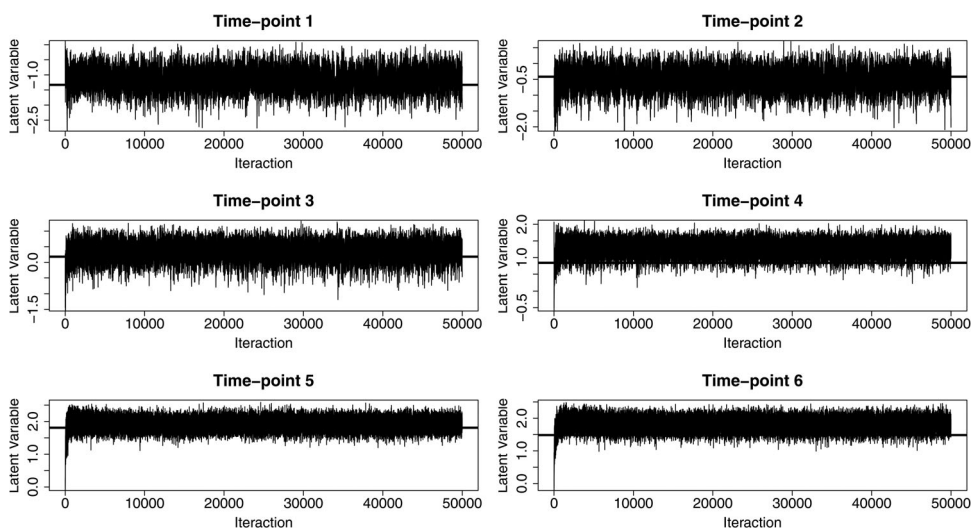


**Figure A2.** Traces-plots of the latent variables of the subject 453. Multivariate PSRF resulted 1.01.

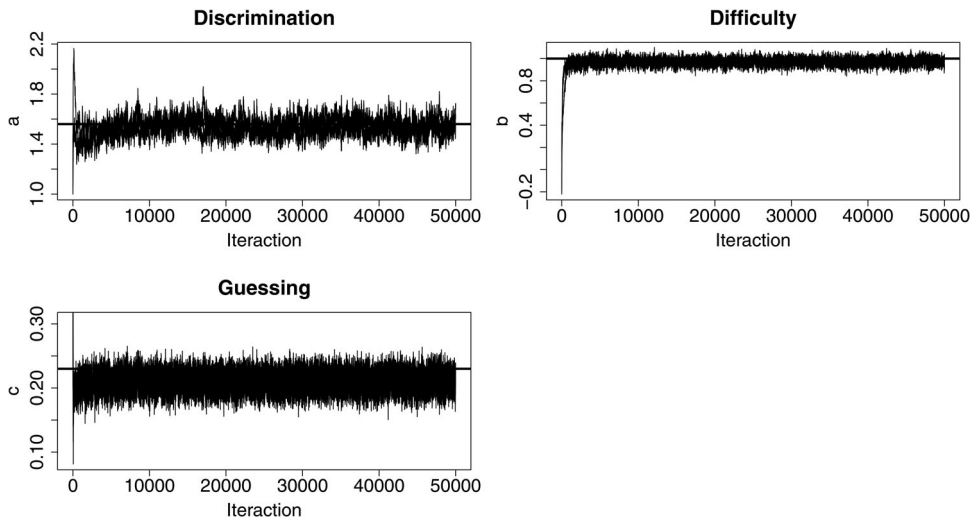




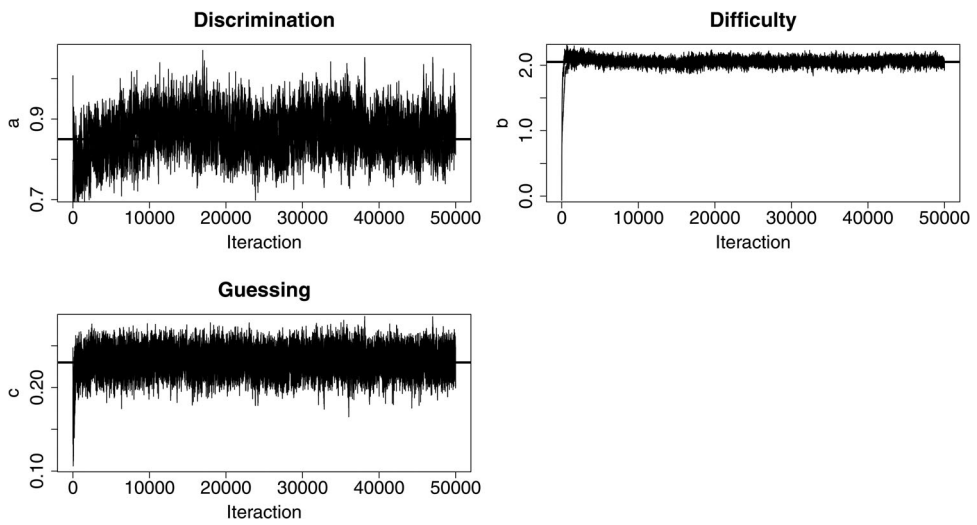
**Figure A3.** Traces-plots of the latent variables of the subject 1123. Multivariate PSRF resulted 1.01.



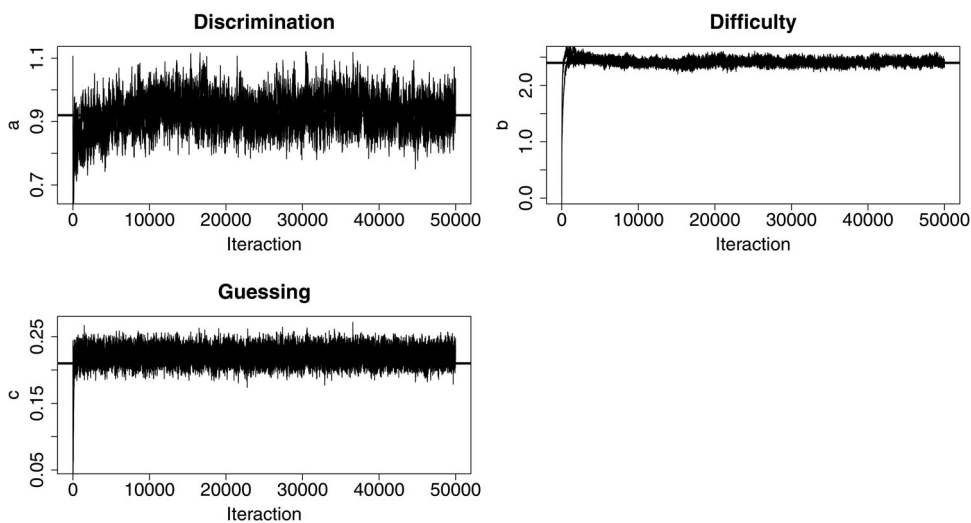
**Figure A4.** Traces-plots of the latent variables of the subject 963. Multivariate PSRF resulted 1.00.



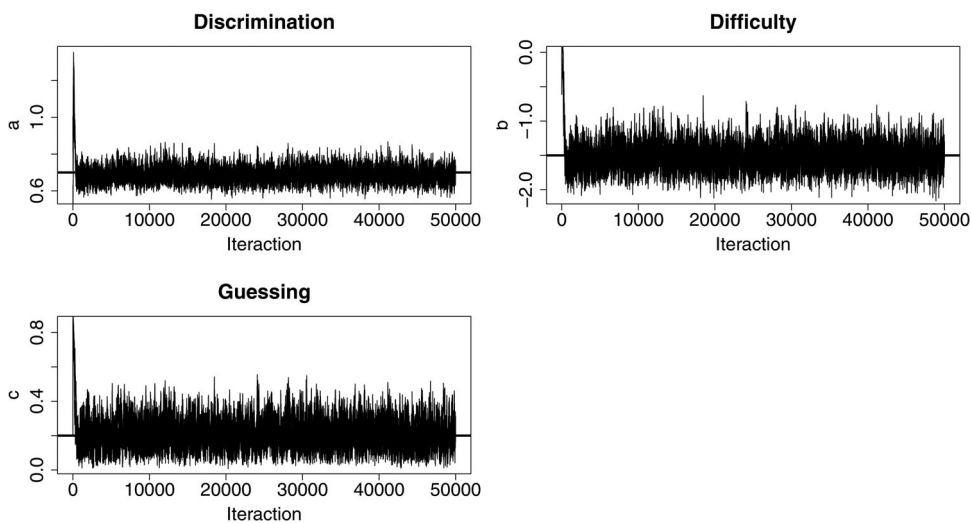
**Figure A5.** Traces-plots of item 40. Univariate PSRF resulted 1.00, 1.01 and 1.00 for the discrimination, difficulty and guessing parameter, respectively.



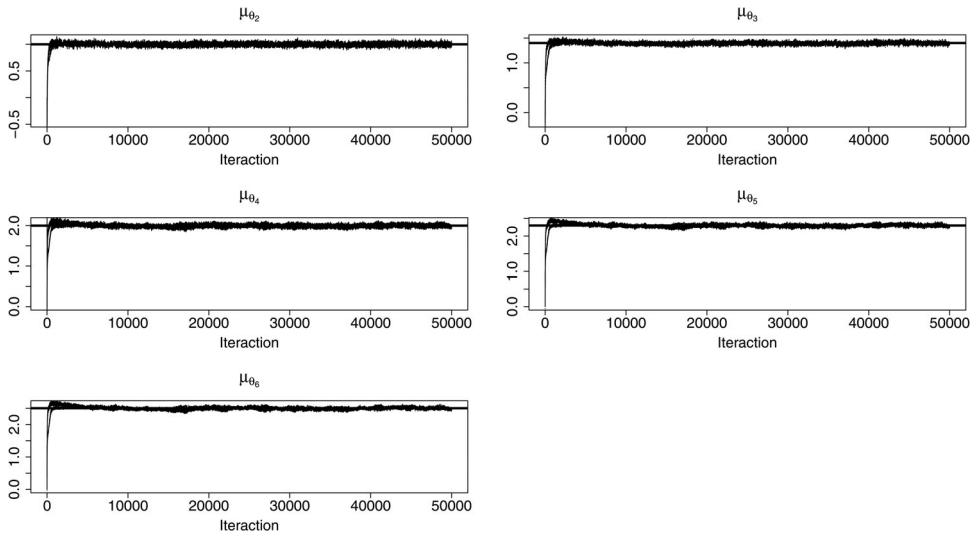
**Figure A6.** Traces-plots of item 94. Univariate PSRF resulted 1.01, 1.01 and 1.00 for the discrimination, difficulty and guessing parameter, respectively.



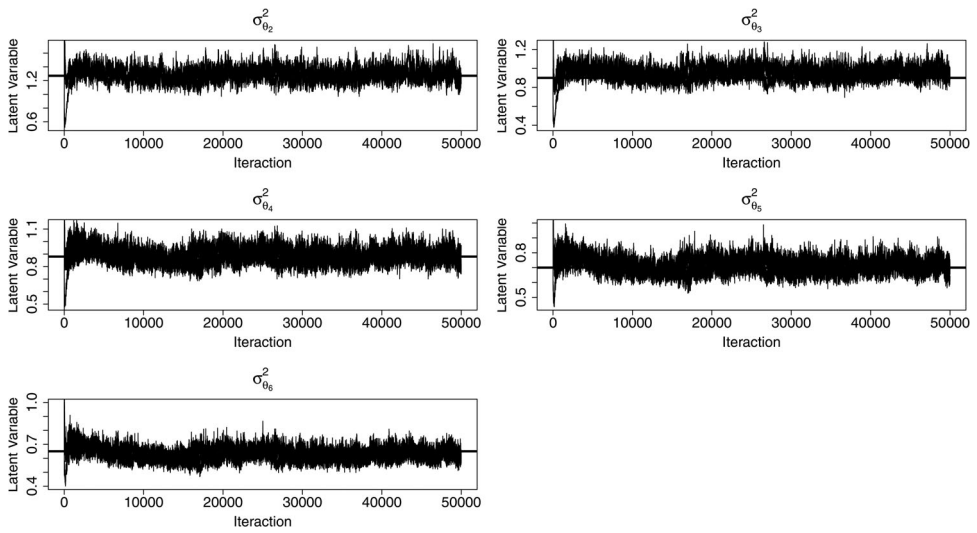
**Figure A7.** Traces-plots of item 56. Univariate PSRF resulted 1.01, 1.04 and 1.00 for the discrimination, difficulty and guessing parameter, respectively.



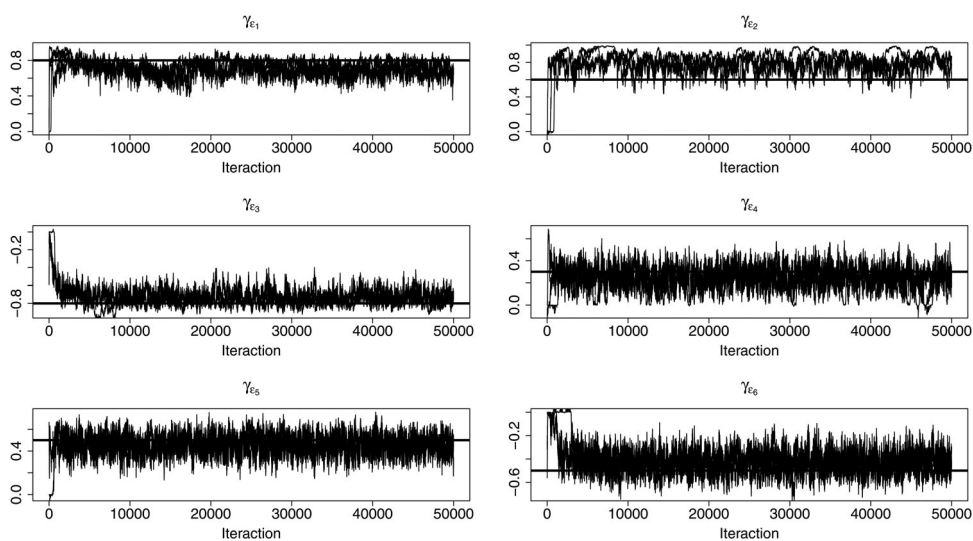
**Figure A8.** Traces-plots of item 1. Univariate PSRF resulted 1.00, 1.00 and 1.00 for the discrimination, difficulty and guessing parameter, respectively.



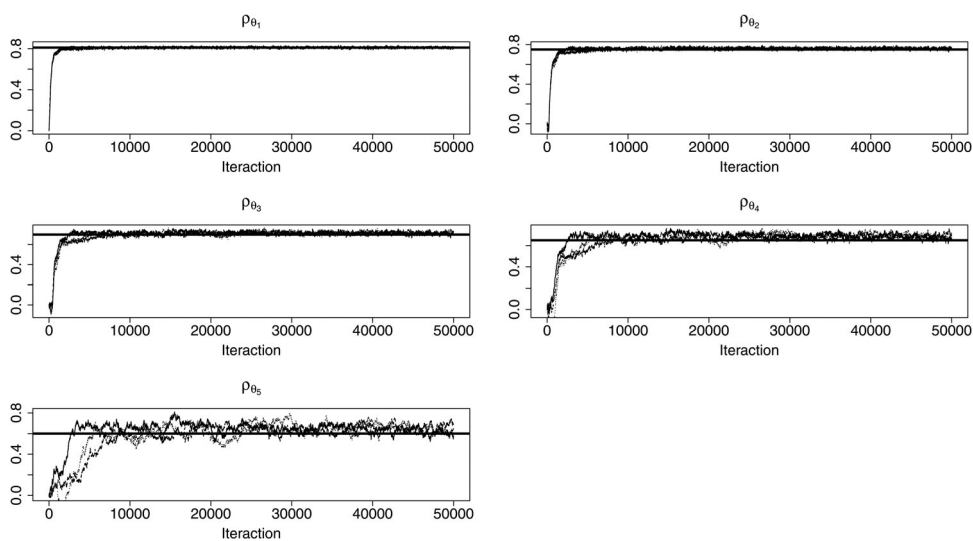
**Figure A9.** Traces-plots of the marginal means. Univariate PSRF ranged between 1.00 and 1.04.



**Figure A10.** Traces-plots of the marginal variances. Univariate PSRF ranged between 1.00 and 1.04.



**Figure A11.** Traces-plots of the conditional asymmetry coefficients. Univariate PSRF ranged between 1.00 and 1.07.



**Figure A12.** Traces-plots of the correlation parameters. Univariate PSRF ranged between 1.00 and 1.04.