

# Bayesian general Cholesky decomposition based modeling of longitudinal multiple-group IRT data with skewed latent distributions and growth curves.

José Roberto Silva dos Santos;\*Caio Lucidius Naberezny Azevedo †

## Abstract

In this work we introduce a multiple-group longitudinal IRT model considering skewed latent traits distribution, based on the work of [Pourahmadi \(1999\)](#), which uses the Cholesky decomposition of the matrix of variance and covariance (dependence) of interest related to the latent traits. A kind of multivariate skew-normal distribution for the latent traits is induced by an antedependence model with centered skew-normal errors. In addition, we consider growth curve models for the mean of the latent traits. A three parameters probit model for dichotomous items is considered. We assume tests administered to subjects clustered into independent groups, which are followed along several time-points (not necessarily equally spaced). Test have common items and may differ among groups and or time-points. Using an appropriate augmented data structure, a longitudinal IRT model is developed through the Pourahmadi's approach. The parameter estimation, model fit assessment and model comparison were implemented through a hybrid MCMC algorithm, such that when the full conditionals are not known, the SVE (Single Variable Exchange) and Metropolis-Hastings algorithms are used. Simulation studies indicate that the parameters are well recovered. Furthermore, a longitudinal study extracted from the Amsterdam Growth and Health Longitudinal Study (AGHLS), that monitor health and life-style of Dutch teenagers, was analyzed to illustrate our model.

**keywords:** longitudinal IRT data, Bayesian inference, antedependence models, SVE algorithm, MCMC algorithms, Cholesky decomposition.

## 1 Introduction

Very often in scientific studies involving latent variables there is an interest in studying subjects belonging different groups, along different assessment occasions. For example: the assessment of students from public and private schools followed along scholar grades, the quality of life of men and women measured along weeks, the psychiatric condition of male and female patients along years, among others. This kind of data was named by [Azevedo et al. \(2015\)](#) as Longitudinal Multiple Group Item Response Data. In this kind of IRT data the group heterogeneity can reflect different behaviors, as well as, the longitudinal structure can induce a correlation pattern between the measures of the same

---

\*Department of Statistics, University of Campinas, Brazil

†Department of Statistics, University of Campinas, Brazil

subject, which can differ among groups. The multiple group longitudinal IRT model proposed by [Azevedo et al. \(2015\)](#) unifies two methodologies. The multiple group IRT model ([Bock and Zimowski, 1997](#)), that allows a simultaneously equating estimation process which leads to more accurate results than a posterior equating, see [Kolen and Brennan \(2004\)](#) and the longitudinal IRT model ([Azevedo et al., 2016](#)), that takes into account the within-subject latent trait dependencies. However, this methodology do not consider unbalanced data design caused by dropouts and/or inclusion of subjects, which is very common in longitudinal studies. Another important feature of longitudinal IRT data is the asymmetry of the latent trait distributions induced by inclusion/exclusion of the subjects along the study, growing and/or decreasing in the latent traits, resulting in negative/positive asymmetry, see [Azevedo et al. \(2011\)](#) and [Santos et al. \(2013\)](#) and references therein. The goal of this paper is to extend the model proposed by [Azevedo et al. \(2015\)](#) in order to handle unbalanced data and asymmetric latent trait distributions. Our model will be build considering a general Cholesky decomposition based modeling of longitudinal IRT data, which are related to the Antedependence Models, see [Pourahmadi \(1999\)](#) and [Nunez-Anton and Zimmerman \(2000\)](#). This approach is very flexible and allows to handle multivariate distributions through the univariate conditional distributions. It allows to represent properly a wide range of specific correlation patterns and different latent traits distributions. The centered skew-normal distribution ([Azzalani, 1985](#)) is considered for the error term, in order to characterize asymmetric behaviors of the latent trait distributions. In addition, this kind of modeling is quite useful for developing diagnostic tools. We will also consider the modeling of the mean of the latent traits through specific parametric growth curves, which can reduce the number of parameters to be estimated.

This paper is outlined as follows. In section 2 we introduce the model. In section 3 we describe the MCMC algorithm developed for parameter estimation. In section 4 a parameter recovery study is presented. In section 5 some model fit assessment tools are introduced and a real data is analyzed. Finally, in section 6 we presented some comments and conclusions.

## 2 Modeling

We will consider one or more tests administered to subjects clustered into  $K$  groups ( $k = 1, \dots, K$ ), which are followed along  $T_k$  time-points, not necessarily equally spaced, say,  $t_1, t_2, \dots, t_{T_k}$ . For brevity we will denote  $t = 1, 2, \dots, T_k$ . For each time-point  $t$  of  $n_{kt}$  subjects, a test of  $I_{kt}$ , from a total of  $I \leq \sum_{k=1}^K \sum_{t=1}^{T_k} I_{kt}$  items, is administered. The tests have common items and the structure can be recognizable as an incomplete block design (the tests may differ among groups and/or time-points). Dropouts and inclusions of the respondents during the study are allowed. Let us define the following notation:  $\theta_{jkt}$  is the latent trait of the subject  $j$  ( $j = 1, 2, \dots, n_{kt}$ ) from group  $k$  at the time-point  $t$ ,  $\boldsymbol{\theta}_{jk.} = (\theta_{jk1}, \dots, \theta_{jkT})'$  is the latent traits vector of the respondent  $j$  from group  $k$ ,  $\boldsymbol{\theta}_{.k.} = (\theta_{.k1}, \dots, \theta_{.kT})'$  is the vector of all latent traits of the group  $k$  and  $\boldsymbol{\theta}_{...} = (\theta_{.1.}, \dots, \theta_{.K.})$  is the vector of all latent traits. The total of latent traits is  $n = \sum_{k=1}^K \sum_{t=1}^T n_{kt}$ . Let  $Y_{ijkt}$  denoting the response of the subject  $j$  from group  $k$  to the item  $i$  ( $i = 1, 2, \dots, I$ ) at the time-point  $t$ ,  $\mathbf{Y}_{.jkt} = (Y_{1jkt}, \dots, Y_{I_{kt}jkt})'$  is the response vector of subject  $j$  from group  $k$  at the time-point  $t$ ,  $\mathbf{Y}_{..kt} = (\mathbf{Y}'_{.1t}, \dots, \mathbf{Y}'_{.n_{kt}})'$  is the response vector of all respondents from group  $k$  at time-point  $t$  and  $\mathbf{Y}_{...} = (\mathbf{Y}'_{.1.}, \dots, \mathbf{Y}'_{.n.})'$  is the entire response matrix. The vector  $(y_{ijkt}, \mathbf{y}'_{.jkt}, \mathbf{y}'_{..kt}, \mathbf{y}'_{...})'$  are the respective observed values, respectively. Let  $\boldsymbol{\zeta}_i$  be the vector of item parameters of item  $i$ ,  $\boldsymbol{\zeta}$  the vector of all item parameters. The  $\boldsymbol{\eta}_{\theta_k}$  is the vector of population parameters of group  $k$  and  $\boldsymbol{\eta}_{\theta}$  the

vector of all population parameters.

The longitudinal multiple group IRT model is defined in two levels: the level of responses and the level of latent traits. At the first level is considered a probit three-parameter IRT model. This model is properly for dichotomous or dichotomized responses. The second level consists in some kind of multivariate skew-normal distribution induced by the antedependence structure, assuming centered skew-normal errors. That is,

$$\begin{aligned} Y_{ijkt} | \theta_{jkt}, \boldsymbol{\zeta}_i &\sim \text{Bernoulli}(P_{ijkt}), \\ P_{ijkt} = \mathbb{P}(Y_{ijkt} | \theta_{jkt}, \boldsymbol{\zeta}_i) &= c_i + (1 - c_i) \Phi(a_i \theta_{jkt} - b_i), \\ \boldsymbol{\theta}_{jk.} | \boldsymbol{\eta}_\theta &\sim D_{T_k}(\boldsymbol{\eta}_{\theta_k}), \end{aligned} \quad (1)$$

where  $D_{T_k}(\cdot)$  stands for some  $T_k$ -dimensional skew-normal distribution indexed by the parameters  $\boldsymbol{\eta}_{\theta_k}$ . In equation (1),  $a_i$  denote the discrimination parameter,  $b_i = a_i b_i^*$ , where  $b_i^*$  is the original difficulty parameter and  $c_i$  is the so called guessing parameter, see [Baker and Kim \(2004\)](#).

An important issue in longitudinal data analysis, concerns to the appropriate modeling of the covariance structure. A suitable specification of the correlation pattern is very important to explain the growth in latent traits, as pointed out by [Azevedo et al. \(2016\)](#). In this work, we will adapt for IRT context, the general procedure of covariance matrix estimation proposed by [Pourahmadi \(1999\)](#). Such approach is based on the Cholesky decomposition of the inverse of the covariance matrix (precision matrix) and allows to represent a wide range of the variance-covariance structures.

## 2.1 Latent traits modeling

To handling the multivariate structure of latent traits, we consider the so-called antedependence models, see [Zimmerman and Núñez-Antón \(2009\)](#). This approach offers a flexible way to deal with multivariate distribution and to represent covariance structures. We assume that  $\mathbb{E}(\boldsymbol{\theta}_{jk.}) = \boldsymbol{\mu}_{\theta_k}$  and  $\text{Cov}(\boldsymbol{\theta}_{jk.}) = \boldsymbol{\Sigma}_{\theta_k}$ .

Then, we can write the latent trait of the subject  $j$  ( $j = 1, \dots, n_{kt}$ ) from group  $k$  at the time-point  $t$  as:

$$\theta_{jkt} = \mu_{\theta_{kt}} + \sum_{r=1}^{t-1} \phi_{ktr} (\theta_{jkr} - \mu_{\theta_{kr}}) + \varepsilon_{jkt}, \quad t = 1, 2, \dots, T_k \text{ and } k = 1, \dots, K, \quad (3)$$

where  $\phi_{ktr}$  are the so-called generalized autoregressive parameters, see [Pourahmadi \(1999\)](#). In addition, consider  $\sum_{r=1}^0 r = 0$ .

In matrix form, we have:

$$\boldsymbol{\varepsilon}_{jk.} = \mathbf{L}_k(\boldsymbol{\theta}_{jk.} - \boldsymbol{\mu}_{\theta_k}), \quad (4)$$

where  $\boldsymbol{\mu}_{\theta_k} = (\mu_{\theta_{k1}}, \dots, \mu_{\theta_{kT_k}})'$  is the mean vector of the latent traits of the group  $k$ . This model was named by [Zimmerman and Núñez-Antón \(2009\)](#) unstructured antedependence model. The random variables  $\boldsymbol{\varepsilon}_{jk.} = (\varepsilon_{jk1}, \varepsilon_{jk2}, \dots, \varepsilon_{jkT_k})'$  are uncorrelated with  $\text{Cov}(\boldsymbol{\varepsilon}_{jk.}) = \mathbf{D}_k$ , where  $\mathbf{D}_k$  is a diagonal matrix  $\text{diag}(d_{k1}, d_{k2}, \dots, d_{kT_k})$  and  $\mathbf{L}_k$  is a  $(T_k \times T_k)$  lower-triangular matrix having the following form,

$$\mathbf{L}_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{k21} & 1 & 0 & \cdots & 0 \\ -\phi_{k31} & -\phi_{k32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\phi_{kT_k1} & -\phi_{kT_k2} & \cdots & -\phi_{T_k(T_k-1)} & 1 \end{pmatrix}. \quad (5)$$

Then, from (4) and using the definition of  $\mathbf{D}_k$  we have that,

$$\text{Cov}(\boldsymbol{\varepsilon}_{jk.}) = \mathbf{L}_k \text{Cov}(\boldsymbol{\theta}_{jk.} - \boldsymbol{\mu}_{\theta_k}) \mathbf{L}'_k = \mathbf{L}_k \boldsymbol{\Sigma}_{\theta_k} \mathbf{L}'_k = \mathbf{D}_k. \quad (6)$$

Therefore the matrix  $\mathbf{L}_k$  diagonalize the covariance matrix  $\boldsymbol{\Sigma}_{\theta_k}$ . This result is related with a variant of the classical Cholesky decomposition (Newton, 1988) of the  $\boldsymbol{\Sigma}_{\theta_k}$  and  $\boldsymbol{\Sigma}_{\theta_k}^{-1}$ .

More parsimonious models, can be obtained by considering some specific correlation patterns. It can reduce considerably the number of parameters when the restricted correlation pattern is supported by the data. The structured matrices considered in this work can be seen in Table 1.

Table 1: Structured covariance matrices used in this work. The  $\sigma$ -parameters are related to variances, while  $\rho$ -parameters are used for correlations.

Structure	Matrix form
First-order Heteroscedastic Autoregressive: ARH(1)	$\begin{pmatrix} \sigma_{\theta_{k1}}^2 & \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_k} & \cdots & \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \rho_{\theta_k}^{T-1} \\ \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_k} & \sigma_{\theta_{k2}}^2 & \cdots & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \rho_{\theta_k}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \rho_{\theta_k}^{T-1} & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \rho_{\theta_k}^{T-2} & \cdots & \sigma_{\theta_{kT}}^2 \end{pmatrix}$
First-order Heteroscedastic Autoregressive Moving-Average: ARMAH(1,1)	$\begin{pmatrix} \sigma_{\theta_{k1}}^2 & \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_{k1}} & \cdots & \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \rho_{\theta_{k1}} \rho_{\theta_{k2}}^{T-2} \\ \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_{k1}} & \sigma_{\theta_{k2}}^2 & \cdots & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \rho_{\theta_{k1}} \rho_{\theta_{k2}}^{T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \rho_{\theta_{k1}} \rho_{\theta_{k2}}^{T-2} & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \rho_{\theta_{k1}} \rho_{\theta_{k2}}^{T-3} & \cdots & \sigma_{\theta_{kT}}^2 \end{pmatrix}$
Heteroscedastic Toeplitz: HT	$\begin{pmatrix} \sigma_{\theta_{k1}}^2 & \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_{k1}} & \cdots & \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \rho_{\theta_{k(T-1)}} \\ \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_{k1}} & \sigma_{\theta_{k2}}^2 & \cdots & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \rho_{\theta_{k(T-2)}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \rho_{\theta_{k(T-1)}} & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \rho_{\theta_{k(T-2)}} & \cdots & \sigma_{\theta_{kT}}^2 \end{pmatrix}$
Antedependence AD	$\begin{pmatrix} \sigma_{\theta_{k1}}^2 & \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_{k1}} & \cdots & \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \prod_{t=1}^{T-1} \rho_{\theta_{kt}} \\ \sigma_{\theta_{k1}} \sigma_{\theta_{k2}} \rho_{\theta_{k1}} & \sigma_{\theta_{k2}}^2 & \cdots & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \prod_{t=2}^{T-1} \rho_{\theta_{kt}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_{k1}} \sigma_{\theta_{kT}} \prod_{t=1}^{T-1} \rho_{\theta_{kt}} & \sigma_{\theta_{k2}} \sigma_{\theta_{kT}} \prod_{t=2}^{T-1} \rho_{\theta_{kt}} & \cdots & \sigma_{\theta_{kT}}^2 \end{pmatrix}$

Considering centered skew-normal distributions, see [Azzalani \(1985\)](#), for the errors of the model (3) we have,

$$\theta_{jkt} = \mu_{\theta_{kt}} + \sum_{r=1}^{t-1} \phi_{ktr}(\theta_{jkr} - \mu_{\theta_{kr}}) + \varepsilon_{jkt}, \quad \varepsilon_{jkt} \stackrel{i.i.d.}{\sim} CSN(0, d_{kt}, \gamma_{\varepsilon_{kt}}), \quad t = 1, 2, \dots, T_k \text{ and } k = 1, \dots, K, \quad (7)$$

where CSN stands for the centered skew-normal distribution. The multivariate distribution of the latent traits  $\boldsymbol{\theta}_{jk}$  of the subject  $j$  can be characterized by the density bellow. According to the density of the CSN distribution and (7) we have:

$$\begin{aligned} p(\boldsymbol{\theta}_{jk} | \boldsymbol{\eta}_{\theta_k}) &= p(\theta_{jk1}) \prod_{t=2}^{T_k} p(\theta_{jkt} | \theta_{jk(1:t-1)}) \\ &= 2^{T_k} \prod_{t=1}^{T_k} \omega_{kt}^{-1} \phi(\omega_{kt}^{-1}(\theta_{jkt} - \beta_{jkt})) \Phi[\lambda_t \omega_{kt}^{-1}(\theta_{jkt} - \beta_{jkt})], \quad j = 1, \dots, n_k \end{aligned} \quad (8)$$

where subscript  $(1 : t - 1)$  stands for the preceding latent traits and  $\beta_{jkt}$  is defined as:

$$\beta_{jk1} = \xi_{k1} \text{ and } \beta_{jkt} = \xi_{kt} + \sum_{r=1}^{t-1} \phi_{ktr}(\theta_{jkr} - \mu_{\theta_{kr}}), \quad \text{for all } t = 2, \dots, T_k. \quad (9)$$

To characterize marginal asymmetry, that is, the asymmetry of latent traits at the time-point  $t$ , we will also use the Pearson's skewness coefficient. This coefficient is given in terms of the centered moment of the random variable  $\theta_{jt}$ . It is possible to show that the marginal skewness coefficients are given by:

$$\begin{aligned} \gamma_{\theta_{kt}} &= \gamma_{\varepsilon_{k1}} \\ \gamma_{\theta_{kt}} &= \frac{1}{\sigma_{\theta_{kt}}^3} \left[ d_{kt}^{3/2} \gamma_{\varepsilon_{kt}} + \sum_{r=1}^{t-1} l_{ktr}^3 d_{kr}^{3/2} \gamma_{\varepsilon_{kr}} \right], \quad t = 2, \dots, T, \end{aligned} \quad (10)$$

where  $\sigma_{\theta_{kt}}^2$  are the marginal variances. That is,

$$\begin{aligned} \sigma_{\theta_{k1}}^2 &= d_{k1} \\ \sigma_{\theta_{kt}}^2 &= d_{kt} + \sum_{r=1}^{t-1} l_{ktr}^2 d_{kr}, \quad t = 2, \dots, T_k. \end{aligned} \quad (11)$$

## 2.2 Growth curves for population means

Here the population mean of each group and along the time-points is modeled by a growth curve. In many IRT applications we can identify some patterns, of the population means, that can be describe by some growth curve models. For example, in psychiatric studies a decreasing mean curve can indicate an inefficiency of the treatment to reverse some pathology. Generally, we consider

$$\boldsymbol{\mu}_{\theta_{jk.}} = (\mu_{\theta_k}(1), \dots, \mu_{\theta_k}(T_k))'; \quad k = 1, \dots, K, \quad (12)$$

where  $\mu_{\theta_k}(\cdot)$  is a parametric function of the time-points. Some curves have been proposed in longitudinal data analysis, see [Lindsey \(1993\)](#) for example. [Table 2](#) presents three important curves that can be useful in many situations. We must to consider some restrictions of the population parameters in order to ensure the identifiability of the model. For the multiple group model the identifiability can be obtained by considering some group as the reference one and then fix the mean and variance of some time-point (reference time-point) of the reference group. In this work we are using a  $(0, 1)$  scale, it means that, the mean and variance of the reference group, in the reference time-point is fixed in 0 and 1, respectively. [Table 3](#) presents the same curves from [Table 2](#) under the restriction  $\mu_{\theta_k}(1) = 0$ .

Table 2: Growth curves for population means

Name	Curve
Jenns	$\mu_{\theta_k}(t) = \alpha_{k1} + \alpha_{k2}t - \exp(\alpha_{k3} + \alpha_{k4}t)$
Count	$\mu_{\theta_k}(t) = \alpha_{k1} + \alpha_{k2} \log(t) + \alpha_{k3}t$
Mitscherlish	$\mu_{\theta_k}(t) = \alpha_{k1} - \alpha_{k3} \exp(-\alpha_{k2}t)$

Table 3: Restricted growth curves for population means

Name	Curve
Jenns	$\mu_{\theta_k}(t) = \alpha_{k1}[1 - \exp(\alpha_{k3}(t - 1))] + \alpha_{k2}(t - 1)$
Count	$\mu_{\theta_k}(t) = \alpha_{k1}(t - 1) + \alpha_{k2} \log(t)$
Mitscherlish	$\mu_{\theta_k}(t) = \alpha_{k1}[1 - \exp(-\alpha_{k2}(t - 1))]$

We can notice that the number of mean parameters to estimated, can be considerably reduced. For example, considering the count growth curve and  $T_k > 2$ , the number of parameters is reduced from  $K \times T_k - 1$  to  $K \times 2 - 1$  considering the identifiability restriction.

### 3 Bayesian Estimation and MCMC Algorithms

In order to facilitates the implementation of the MCMC algorithms, particularly, aiming to obtain full conditional distribution with know form and to develop properly model-fit assessment tools; we will use the augmented data approach to represent our IRT model, see [Tanner and Wong \(1987\)](#). For the three-parameter probit model we can use the augmented data scheme proposed by [Béguin and Glas \(2001\)](#). This methodology consist on define a vector of binary variables  $W_{ijkt}$  such that

$$W_{ijkt} = \begin{cases} 1, & \text{if the subject } j \text{ from group } k, \text{ at time-point } t \text{ knows the right response to the item } i \\ 0, & \text{if the subject } j \text{ from group } k, \text{ at time-point } t \text{ does not know the right response to the item } i. \end{cases}$$

Consequently, the conditional distribution of  $W_{ijkt}$  given  $Y_{ijkt} = y_{ijkt}$  is given by

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

Therefore the augmented variables  $\mathbf{Z} = (Z_{1111}, \dots, Z_{1n_{11}11}, \dots, Z_{In_k T K T_k})'$ , are given by

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

The original data can be represented by

$$Y_{ijkt} = \mathbb{I}(Z_{ijkt} > 0) \mathbb{I}(W_{ijkt} = 1) + \mathbb{I}(Z_{ijkt} \leq 0) \mathbb{I}(W_{ijkt} = 0), \tag{15}$$

where  $\mathbb{I}$  denotes the indicator function. To handle incomplete block design an indicator variable  $\mathbf{I}$  is defined as:

$$\mathbf{I}_{ijkt} = \begin{cases} 1, & \text{if item } i, \text{ was administered to the respondent } j \text{ from group } k \text{ at time-point } t, \\ 0, & \text{if item } i, \text{ was not administered to the respondent } j \text{ from group } k \text{ at time-point } t. \end{cases}$$

To describe possible omissions on the data, caused by uncontrolled events, such that, non-response or errors in recoding data, we defined another variable as follows,

$$V_{ijkt} = \begin{cases} 1, & \text{if observed response of respondent } j \text{ from group } k \text{ at time-point } t \text{ on item } i, \\ 0, & \text{otherwise.} \end{cases}$$

We assumed that the missing data are missing at random (MAR), such that the missing data patterns distribution does not depend on the unobserved data. Therefore, the augmented likelihood is given by

$$\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, |\mathbf{z}, \dots, \mathbf{w}, \dots, \mathbf{y}, \dots) &\propto \prod_{k=1}^K \prod_{t=1}^{T_k} \prod_{j=1}^{n_{tk}} \prod_{i \in I_{jkt}} \exp \{ -0.5 (z_{ijkt} - a_i \theta_{jkt} + b_i)^2 \} \mathbb{I}_{(z_{ijkt}, w_{ijkt})} \\
&\times p(w_{ijkt} | y_{ijkt}, \theta_{jkt}, \zeta_i),
\end{aligned} \tag{16}$$

where  $\mathbb{I}_{(z_{ijkt}, w_{ijkt})}$  stands for the indicator function  $\mathbb{I}_{(z_{ijkt} < 0, w_{ijkt} = 0)} + \mathbb{I}_{(z_{ijkt} \geq 0, w_{ijkt} = 1)}$  and  $I_{jkt}$  is the set of items answered by the subject  $j$  at time  $t$  and  $p(w_{ijkt} | y_{ijkt}, \theta_{jkt}, \zeta_i)$  follows from (13).

### 3.1 Prior and posterior distributions

The joint prior distribution of the unknown parameters is assumed to be

$$p(\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) = \left\{ \prod_{k=1}^K (\theta_{jk1} | \boldsymbol{\eta}_{\theta_{k1}}) \prod_{t=2}^{T_k} \prod_{j=1}^{n_{tk}} p(\theta_{jkt} | \theta_{jk(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_\zeta) \right\} \left\{ \prod_{k=1}^K \prod_{t=1}^{T_k} p(\boldsymbol{\eta}_{\theta_{kt}} | \boldsymbol{\eta}_\eta) \right\}, \quad (17)$$

where  $\boldsymbol{\eta}_\zeta$  and  $\boldsymbol{\eta}_\eta$  are hyperparameters associated with  $\boldsymbol{\zeta}$  and  $\boldsymbol{\eta}_\theta$ , respectively, and the subscript  $(1 : t - 1)$  denotes the preceding latent traits. In addition, we are assuming independence between items and population parameters. The prior distributions of the latent traits are defined in equation (7). For the item parameters we have:

$$p(\zeta_{i(-c_i)}) \propto \exp[-.5(\zeta_{i(-c_i)} - \boldsymbol{\mu}_\zeta)' \boldsymbol{\Psi}_\zeta^{-1} (\zeta_{i(-c_i)} - \boldsymbol{\mu}_\zeta)] \mathbb{I}_{(a_i > 0)} \quad (18)$$

and

$$c_i \sim \text{Beta}(a_c, b_c), \quad (19)$$

where  $\zeta_{i(-c_i)} = (a_i, b_i)$ . For population parameters (including the growth curve parameters) we are considering the following priors:

$$\begin{aligned} \alpha_{kl} &\sim N(\mu_\alpha, \sigma_\alpha^2); \quad l = 1, 2, 3, \\ \sigma_{\theta_{kt}}^2 &\sim IG(a_\sigma, b_\sigma), \\ \gamma_{\varepsilon_{kt}} &\sim N(\mu_\gamma, \sigma_\gamma^2) \mathbb{I}_{[-.99527, .99527]}, \quad k = 1, \dots, K \text{ and } t = 1, \dots, T_k. \end{aligned} \quad (20)$$

For the generalized autoregressive parameters we define:

$$\phi_{ktr} \sim N(\mu_\phi, \sigma_\phi^2) \quad t = 2, \dots, T_k \text{ and } r = 1, \dots, t - 1. \quad (21)$$

In the case of structured matrix, the prior distributions for correlation parameters are specified as:

$$\rho_{\theta_{kt}} \sim N(\mu_\rho, \sigma_\rho^2) \mathbb{I}_{[0,1]}, \quad t = 1, 2, \dots, T_k - 1. \quad (22)$$

That is, a truncated normal distribution on the interval  $[0, 1]$ . This interval was considered, since negative correlations are rarely observed in longitudinal studies.

Given the augmented likelihood in equation (16) and the prior distribution in equations (7), (18), (19), (20) and (21), the joint posterior distribution is given by:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, \mathbf{z}, \mathbf{w}, \mathbf{y}) &\propto \left\{ \prod_{k=1}^K \prod_{t=1}^{T_k} \prod_{j=1}^{n_{kt}} \prod_{i \in I_{jkt}} \exp\{-.5(z_{ijkt} - a_i \theta_{jkt} + b_i)^2\} p(w_{ijkt} | y_{ijkt}, \theta_{jkt}, \zeta_i) \mathbb{I}_{(z_{ijkt}, w_{ijkt})} \right\} \\ &\times \left\{ \prod_{k=1}^K \prod_{j=1}^{n_{kt}} p(\theta_{jk1} | \boldsymbol{\eta}_{\theta_{kt}}) \prod_{t=2}^{T_k} p(\theta_{jkt} | \theta_{jk(1:t-1)}, \boldsymbol{\eta}_{\theta_{kt}}) \right\} \\ &\times \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_\zeta) \right\} \left\{ \prod_{k=1}^K \prod_{t=1}^{T_k} p(\boldsymbol{\eta}_{\theta_{kt}} | \boldsymbol{\eta}_\eta) \right\}. \end{aligned} \quad (23)$$



Since the posterior distribution has an intractable analytical form, we will use MCMC algorithms in order to obtain empirical approximation. In the case of items and generalized autoregressive parameters we can find full conditional distributions easy to sample from, that is,

$$\zeta_{i(-c_i)} | (\cdot) \sim N(\widehat{\Psi}_{\zeta_i} \widehat{\zeta}_i, \widehat{\Psi}_{\zeta_i}), \quad (24)$$

where

$$\begin{aligned} \widehat{\zeta}_i &= (\Theta_{i..})' \mathbf{z}_i + \Psi_{\zeta_i}^{-1} \boldsymbol{\mu}_{\zeta_i}, \\ \widehat{\Psi}_{\zeta_i} &= \left[ (\Theta_{i..})' (\Theta_{i..}) + \Psi_{\zeta_i}^{-1} \right]^{-1}, \\ \Theta_{i..} &= [\boldsymbol{\theta} - \mathbf{1}_n] \bullet \mathbb{1}_i, \end{aligned}$$

$(\cdot)$  denotes the set of all others parameters,  $\mathbb{1}_i$  is a  $(n \times 2)$  matrix with lines, equals to 1 or 0, according to the response/missing response of the subject  $j$  to the item  $i$  at time-point  $t$  and  $\bullet$  denotes the *Hadamard* product and for the guessing parameters,

$$c_i | (\cdot) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1), \quad (25)$$

where

$$s_i = \sum_{j|w_{ijkt}=0}^n \mathbf{y}_{ijk}; \quad \sum_{j=1}^{n_{kt}} \mathbb{I}(w_{ijkt} = 0).$$

Considering unstructured covariance matrices we can sample generalized autoregressive parameters directly from:

$$\phi_{tr} | (\cdot) \sim N(Q_{tr} q_{tr}, Q_{tr}),$$

where

$$\begin{aligned} Q_{tr} &= \left( \frac{\sum_{j=1}^{n_{kt}} (\theta_{jk(t-1)} - \mu_{\theta_{k(t-1)}})^2}{\zeta_{kt}^2} + \frac{1}{\sigma_{\phi}^2} \right)^{-1} \\ q_{tr} &= \frac{1}{\zeta_{kt}^2} \sum_{j=1}^{n_{kt}} (\theta_{jkr} - \mu_{\theta_{kr}}) (\theta_{jkt} - \xi_{kt} - \sum_{r \neq t} \phi_{tr} (\theta_{jkr} - \mu_{\theta_{kr}}) - \tau_{kt} h_{jkt}), \end{aligned} \quad (26)$$

for all  $k = 1, \dots, K$ ,  $t = 2, \dots, T_k$  and  $r = 1, \dots, (t-1)$ .

For population parameters is not possible to obtain known full conditional, therefore some auxiliary algorithm is necessary. The Metropolis-Hastings algorithm is considered using the original likelihood of the skew-normal distribution. Let us denote the conditional distribution of  $\theta_{jt} | \theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_{kt}}$  as

$$L(\theta_{jkt} | \theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_{kt}}) \equiv L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}) = 2^{T_k} \omega_{kt}^{-1} \phi(\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt})) \Phi[\lambda_{kt} (\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt}))]. \quad (27)$$

The function  $\beta_{jk}$  is defined in equations (9). In order to perform the Metropolis-Hastings steps we need to defined the following proposal densities:

$$\begin{aligned} p(\alpha_{kl}^{(m)}, \alpha_{kl}^{(m-1)}) &= N(\alpha_{kl}^{(m-1)}, \sigma_0^2); \quad l = 1, 2, 3, \\ p((\sigma_{\theta_k}^2)^{(m)}, (\sigma_{\theta_k}^2)^{(m-1)}) &= \text{Lognormal}((\sigma_{\theta_k}^2)^{(m-1)}, \sigma_0^2) \\ p(\gamma_{\varepsilon_{kt}}^{(m)}, \gamma_{\varepsilon_{kt}}^{(m-1)}) &= U(\nu_1(\gamma_{\varepsilon_{kt}}^{(m-1)}), \nu_2(\gamma_{\varepsilon_{kt}}^{(m-1)})); \quad k = 1, \dots, K, \end{aligned} \quad (28)$$

where

$$\nu_1 = \max\{-.99527, \gamma_{\varepsilon_{kt}}^{(m-1)} - \Delta_\gamma\} \quad \text{and} \quad \nu_2 = \max\{.99527, \gamma_{\varepsilon_{kt}}^{(m-1)} + \Delta_\gamma\}, \quad \Delta_\gamma > 0.$$

The constant  $\Delta_\gamma$  is defined in a suitable value, see [Azevedo et al. \(2011\)](#). The superscript  $(m)$  indicate the estimate at the iteration  $m$  of the MCMC algorithm. The algorithms 3.1 to 3.3 describe the metropolis-hastings steps for each growth curve, variance and conditional skewness parameters.

---

**Algorithm 3.1** Growth curve parameters sampler

---

- 1: Specify the function  $\mu_{\theta_k}(t)$
- 2: Draw  $\alpha_{kl}^{(m)} \sim N(\alpha_{kl}^{(m-1)}, \sigma_0^2)$
- 3: Set  $\boldsymbol{\mu}_{\theta_k}^{(m)} = (\mu_{\theta_k}^{(m)}(1), \dots, \mu_{\theta_k}^{(m)}(T_k))'$
- 4: Draw  $u \sim U(0, 1)$
- 5: **if**  $u < \min\{1, R_\alpha\}$  where

$$R_\alpha = \frac{\prod_{t=1}^{T_k} \prod_{j=1}^{n_{kt}} L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}^{(m)}) N(\alpha_{kl}^{(m)}, \mu_\alpha, \sigma_\alpha^2)}{\prod_{t=1}^{T_k} \prod_{j=1}^{n_{kt}} L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}^{(m-1)}) N(\alpha_{kl}^{(m-1)}, \mu_\alpha, \sigma_\alpha^2)}$$

- 6: **then**  $\alpha_{kl}^{(m-1)} = \alpha_{kl}^{(m)}$
  - 7: **end if**
- 

---

**Algorithm 3.2** Variance parameters sampler

---

- 1: Draw  $(\sigma_{\theta_{kt}}^2)^{(m)} \sim \text{Lognormal}((\sigma_{\theta_{kt}}^2)^{(m-1)}, \sigma_0^2)$
- 2: Draw  $u \sim U(0, 1)$
- 3: **for**  $t=1$  to  $T_k$  **do**
- 4: **if**  $u < \min\{1, R_\sigma\}$  where

$$R_\sigma = \frac{\prod_{j=1}^{n_{kt}} L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}^{(m)}) IG((\sigma_{\theta_{kt}}^2)^{(m)}, a_\sigma, b_\sigma) LN((\sigma_{\theta_{kt}}^2)^{(m-1)}, (\sigma_{\theta_{kt}}^2)^{(m-1)}, \sigma_0^2)}{\prod_{j=1}^{n_{kt}} L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}^{(m-1)}) IG((\sigma_{\theta_{kt}}^2)^{(m-1)}, a_\sigma, b_\sigma) LN((\sigma_{\theta_{kt}}^2)^{(m)}, (\sigma_{\theta_{kt}}^2)^{(m-1)}, \sigma_0^2)}$$

- 5: **then**  $(\sigma_{\theta_{kt}}^2)^{(m-1)} = (\sigma_{\theta_{kt}}^2)^{(m)}$
  - 6: **end if**
  - 7: **end for**
- 

The notation  $N(\cdot, \mu_\alpha, \sigma_\alpha^2)$ ,  $IG(\cdot, a_\sigma, b_\sigma)$  and  $LN(\cdot, \sigma_{\theta_{kt}}^2, \sigma_0^2)$  stand for a density of a normal, inverse-gamma and log-normal distributions, respectively.

---

**Algorithm 3.3** Conditional skewness parameters sampler
 

---

- 1: Draw  $\gamma_{\varepsilon_{kt}}^{(m)} \sim U\left(\nu_1(\gamma_{\varepsilon_{kt}}^{(m-1)}), \nu_2(\gamma_{\varepsilon_{kt}}^{(m-1)})\right)$
- 2: Draw  $u \sim U(0, 1)$
- 3: for  $t=1$  to  $T_k$  **do**
- 4: **if**  $u < \min\{1, R_\gamma\}$  where

$$R_\gamma = \frac{\prod_{j=1}^{n_{kt}} L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}^{(m)}) N(\gamma_{\varepsilon_{kt}}^{(m)}, \mu_\gamma, \sigma_\gamma^2) [\nu_1(\gamma_{\varepsilon_{kt}}^{(m)}) - \nu_2(\gamma_{\varepsilon_{kt}}^{(m)})]}{\prod_{j=1}^{n_{kt}} L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}^{(m-1)}) N(\gamma_{\varepsilon_{kt}}^{(m-1)}, \mu_\gamma, \sigma_\gamma^2) [\nu_1(\gamma_{\varepsilon_{kt}}^{(m-1)}) - \nu_2(\gamma_{\varepsilon_{kt}}^{(m-1)})]}$$

- 5: **then**  $\gamma_{\varepsilon_{kt}}^{(m-1)} = \gamma_{\varepsilon_{kt}}^{(m)}$
  - 6: **end if**
  - 7: **end for**
- 

### 3.1.1 The latent trait sampler

To sample latent traits we consider the Henze's stochastic representation of the skew-normal distribution, see [Rodriguez \(2005\)](#) and [Santos et al. \(2017\)](#), for more details. Therefore, considering the augmented data structure, we have the following representation in terms of dynamic model:

$$Z_{jikt} = a_i \theta_{jkt} - b_i + \xi_{jikt}, \quad \xi_{jikt} \sim N(0, 1), \quad (29)$$

$$\theta_{jkt} = \beta_{jkt} + \tau_{kt} h_{jkt} + \varepsilon_{jkt}, \quad \varepsilon_{jkt} \sim N(0, \varsigma_{kt}^2), \quad (30)$$

where  $h_{jkt}$  and  $\varepsilon_{jkt}$  are independent and  $\beta_{jt}$  is define as before with  $H_{jkt} \sim HN(0, 1)$ . Then the forward step of the FFBS is described below. Following [Gamerman and Lopes \(2006\)](#) consider the conditional distribution  $\theta_{jk(t-1)} | \mathbf{z}_{jk}^{t-1} \sim N(m_{jk(t-1)}, C_{jk(t-1)})$ , where  $\mathbf{z}_{jk}^{t-1}$  refer to the information until  $t-1$ . The system equation (30) can be written as  $\theta_{jkt} | \theta_{jk(t-1)}, h_{jkt} \sim N(\beta_{jkt} + \tau_{kt} h_{jkt}, \varsigma_{kt}^2)$ . By properties of the normal distribution, these specifications can be combined leading to the marginal distribution:

$$\theta_{jkt} | \mathbf{z}_{jk}^{t-1}, h_{jkt} \sim N(a_{jkt}, R_{jkt}), \quad (31)$$

where

$$a_{jkt} = \xi_{kt} + \sum_{r=1}^{t-1} \phi_{tkr} (m_{jkr} - \mu_{\theta_{kr}}) + \tau_{kt} h_{jkt} \quad \text{and} \quad R_{jkt} = \varsigma_{kt}^2 + \sum_{r=1}^{t-1} \phi_{tr}^2 C_{jkr}.$$

Thus,

$$\theta_{jkt} | \mathbf{z}_{jk}^t, h_{jkt} \sim N(m_{jkt}, C_{jkt}), \quad (32)$$

where

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

Equation 32 is referred in the literature as *Kalman Filter*. Therefore, the backward distributions are given by:

$$\theta_{jkt} | \theta_{jk(t+1)}, \mathbf{z}_{jk.}^t, h_{jkt} \sim N(m_{\theta_{jkt}}, C_{\theta_{jkt}}), \quad (33)$$

where

$$C_{\theta_{jkt}} = \left( \frac{\phi_{t+1,t}^2}{\zeta_{t+1}^2} + \frac{1}{C_{jkt}} \right)^{-1} \quad \text{and} \quad m_{\theta_{jkt}} = \left( \frac{\phi_{t+1,t} (\theta_{jk(t+1)} - \alpha_{jk(t+1)})}{\zeta_{k(t+1)}^2} + \frac{m_{jkt}}{C_{jkt}} \right) C_{\theta_{jkt}},$$

where  $\alpha_{jk(t+1)} = \xi_{k(t+1)} - \phi_{t+1,t} \mu_{\theta_{kt}} + \tau_{kt} h_{jk(t+1)} + \sum_{r=1}^{t-1} \phi_{t+1,t} (\theta_{jkr} - \mu_{\theta_{kr}})$ , for all  $k = 1, \dots, K$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ . Then, a scheme to sample from the full conditional distribution of  $\theta_{jk.}$  is given by algorithm 3.4.

---

**Algorithm 3.4** FFBS algorithm

---

- 1: Sample  $\theta_{jkT}$  from  $\theta_{jkT} | \mathbf{z}_{jk.}^{T_k}, h_{jkt}$  and set  $t = T_k - 1$ .
  - 2: Sample  $\theta_{jkt}$  from  $\theta_{jkt} | \theta_{jk(t+1)}, \mathbf{z}_{jk.}^t, h_{jkt}$ .
  - 3: Decrease  $t$  to  $t - 1$  and return to step 2 until  $t = 1$ .
- 

Step 1 is obtained by running the Kalman filter from  $t = 1$  to  $t = T_k$ . The augmented data  $H_{jkt}$  is simulated from its full conditional distribution given by

$$H_{jkt} | (\cdot) \sim N \left[ \frac{\tau_{kt} (\theta_{jkt} - \beta_{jkt})}{\tau_{kt}^2 + \zeta_{kt}^2}, \frac{\zeta_{kt}^2}{\tau_{kt}^2 + \zeta_{kt}^2} \right] \mathbb{I}(h_{jkt} > 0). \quad (34)$$

The so-called Henze's parameters  $(\xi_{kt}, \tau_{kt}, \zeta_{kt}^2)$  are a one to one mapping of the population parameters  $(\mu_{\theta_{kt}}, \sigma_{\theta_{kt}}^2, \gamma_{\theta_{kt}})$ .

To sample the correlation parameters in the case of structured correlation matrices we are considering the SVE (Single Variable exchange) algorithm, see Marsman (2014). To illustrate this algorithm consider, for example, an ARH(1) matrix, see Table 1. Let  $\theta_{.k.}$  being the set of all latent traits of the group  $k$ , as defined in Section 2 and  $p(\theta_{.k.} | \boldsymbol{\mu}_{\theta_k}, \boldsymbol{\phi}_k, \mathbf{d}_k, \boldsymbol{\gamma}_{\theta_k})$  denoting the likelihood generated by the antedependence model defined in equation (7), that is

$$\begin{aligned} p(\theta_{.k.} | \boldsymbol{\mu}_{\theta_k}, \boldsymbol{\phi}_k, \mathbf{d}_k, \boldsymbol{\gamma}_{\theta_k}) &= \prod_{j=1}^{n_{kt}} p(\theta_{jk1} | \mu_{\theta_{k1}}, d_{k1}, \gamma_{\theta_{k1}}) \prod_{t=2}^{T_k} p(\theta_{jkt} | \mu_{\theta_{k1}}, \boldsymbol{\phi}_{kt}, d_{kt}, \gamma_{\theta_{kt}}) \\ &\propto 2^{Tn_t} \prod_{j=1}^{n_{kt}} \prod_{t=1}^{T_k} \omega_{kt}^{-1} \phi_d (\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt})) \\ &\times \Phi_d [\lambda_{kt} (\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt}))], \end{aligned} \quad (35)$$

where  $\boldsymbol{\phi}_k = (\phi_{k21}, \phi_{k31}, \phi_{k32}, \dots, \phi_{kT_k(T_k-1)})'$ ,  $\mathbf{d}_k = (d_{k1}, \dots, d_{kT_k})'$ ,  $\boldsymbol{\gamma}_{\theta_k} = (\gamma_{\theta_{k1}}, \dots, \gamma_{\theta_{kT_k}})'$ ,  $\boldsymbol{\mu}_{\theta_k} = (\mu_{\theta_{k1}}, \dots, \mu_{\theta_{kT_k}})$  and  $\boldsymbol{\phi}_{kt}$  denotes the elements correspondents to the time-point  $t$  and group  $k$ .

Algorithm 3.5 present the correlation parameters sampler for the ARH(1) model for all simulation  $m = 1, \dots, M$ .

---

**Algorithm 3.5** The SVE algorithm with oversampling to sample a correlation parameter considering ARH(1) matrix

---

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

**Require:** A function `AR1.matrix()` to build the ARH(1) matrix

- 1: **for**  $g = 1$  to  $G$  **do**
- 2: Draw  $\rho_{\theta_{kg}}^{(m)} \sim p(\rho_{\theta_k})$
- 3: Draw  $\boldsymbol{\theta}_{.k.g}^{(m)}$  from the model (7)
- 4: Compute  $r_1(\boldsymbol{\theta}_{.k.g}^{(m)})$  the first-order sample correlation
- 5: **end for**
- 6: Choose the  $\rho_{\theta_{kg}}^{(m)}$  and  $\boldsymbol{\theta}_{.k.g}^{(m)}$  whose  $r_1(\boldsymbol{\theta}_{.k.g}^{(m)})$  is closest to  $r_1(\boldsymbol{\theta}_{.k}^{(m-1)})$
- 7: Set  $\rho_{\theta_k}^{(m)} = \rho_{\theta_{kg}}^{(m)}$  and  $\boldsymbol{\theta}_{.k}^{(m)} = \boldsymbol{\theta}_{.k.g}^{(m)}$  the candidate values
- 8: Build the ARH(1) proposed matrix  $\Sigma_{\rho_{\theta_k}}^{(m)}$  using `ARH1.matrix()`
- 9: Perform the Cholesky decomposition of  $\Sigma_{\rho_{\theta_k}}^{(m)}$  to obtain the matrices  $\mathbf{L}_k^{(m)}$  and  $\mathbf{D}_k^{(m)}$
- 10: Draw  $u \sim U(0, 1)$
- 11: **if**

$$\left( u < \min \left\{ 1, \frac{p(\boldsymbol{\theta}_{.k}^{(m-1)} | \boldsymbol{\mu}_{\theta_k}^{(m-1)}, \boldsymbol{\phi}_k^{(m)}, \mathbf{d}_k^{(m)}, \boldsymbol{\gamma}_{\theta_k}^{(m-1)}) p(\boldsymbol{\theta}_{.k}^{(m)} | \boldsymbol{\mu}_{\theta_k}^{(m-1)}, \boldsymbol{\phi}_k^{(m-1)}, \mathbf{d}_k^{(m-1)}, \boldsymbol{\gamma}_{\theta_k}^{(m-1)})}{p(\boldsymbol{\theta}_{.k}^{(m-1)} | \boldsymbol{\mu}_{\theta_k}^{(m-1)}, \boldsymbol{\phi}_k^{(m-1)}, \mathbf{d}_k^{(m-1)}, \boldsymbol{\gamma}_{\theta_k}^{(m-1)}) p(\boldsymbol{\theta}_{.k}^{(m)} | \boldsymbol{\mu}_{\theta_k}^{(m-1)}, \boldsymbol{\phi}_k^{(m)}, \mathbf{d}_k^{(m)}, \boldsymbol{\gamma}_{\theta_k}^{(m-1)})} \right\} \right)$$

**then**

- 12:  $\rho_{\theta_k}^{(m-1)} = \rho_{\theta_k}^{(m)}$
  - 13: **end if**
-

For more than one correlation parameter, the algorithm 3.5 can be applied independently to each one by choosing suitable sufficient statistics. It is also possible to sample blocks of correlation parameters. This can be done by modifying line 11 of the algorithm 3.5 to allow accept/reject proposed values jointly. Also, different covariance matrices can be considered for different groups, as we will see in next section.

In summary, a general algorithm to estimate the parameter's model is a combination of Gibbs sampling, FFBS, SVE and Metropolis-Hastings algorithms as we can see in algorithms 3.6 and 3.7.

---

**Algorithm 3.6** General algorithm for unstructured matrix

---

- 1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
  - 2: Simulate  $W_{ijkt}$  from  $W_{ijkt}|\cdot$  for all  $K = 1, \dots, K$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 3: Simulate  $Z_{ijkt}$  from  $Z_{ijkt}|\cdot$  for all  $K = 1, \dots, K$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 4: Simulate  $H_{jkt}$  from  $H_{jkt}|\cdot$  for all  $K = 1, \dots, K$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 5: Simulate  $\theta_{jkt}$  using the algorithm 3.4 for all  $K = 1, \dots, K$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 6: Simulate  $\zeta_i$  from  $\zeta_i|\cdot$  for all  $i = 1, \dots, I$ .
  - 7: Simulate  $c_i$  from  $c_i|\cdot$  for all  $i = 1, \dots, I$ .
  - 8: Simulate  $\alpha_{kl}$  from  $\alpha_{kl}|\cdot$  using algorithm 3.1 for all  $k = 1, \dots, K$  and  $l = 1, \dots, 3$ .
  - 9: Simulate  $\sigma_{\theta_{kt}}^2$  from  $\sigma_{\theta_{kt}}^2|\cdot$  using algorithm 3.2 for all  $k = 1, \dots, K$  and  $t = 1, \dots, T_k$ .
  - 10: Simulate  $\gamma_{\varepsilon_{kt}}$  from  $\gamma_{\varepsilon_{kt}}|\cdot$  using algorithm 3.3 for all  $k = 1, \dots, K$  and  $t = 1, \dots, T_k$ .
  - 11: Simulate  $\phi_{tk}$  from  $\phi_{tk}|\cdot$  for all  $t = 2, \dots, T$  and  $k = 1, \dots, t - 1$ .
- 

---

**Algorithm 3.7** General algorithm for structured matrices

---

- 1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
  - 2: Simulate  $W_{ijkt}$  from  $W_{ijkt}|\cdot$  for all  $K = 1, \dots, K$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 3: Simulate  $Z_{ijkt}$  from  $Z_{ijkt}|\cdot$  for all  $K = 1, \dots, K$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 4: Simulate  $H_{jkt}$  from  $H_{jkt}|\cdot$  for all  $K = 1, \dots, K$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 5: Simulate  $\theta_{jkt}$  using the algorithm 3.4 for all  $K = 1, \dots, K$ ,  $j = 1, \dots, n_{kt}$  and  $t = 1, \dots, T_k$ .
  - 6: Simulate  $\zeta_i$  from  $\zeta_i|\cdot$  for all  $i = 1, \dots, I$ .
  - 7: Simulate  $c_i$  from  $c_i|\cdot$  for all  $i = 1, \dots, I$ .
  - 8: Simulate  $\alpha_{kl}$  from  $\alpha_{kl}|\cdot$  using algorithm 3.1 for all  $k = 1, \dots, K$  and  $l = 1, \dots, 3$ .
  - 9: Simulate  $\sigma_{\theta_{kt}}^2$  from  $\sigma_{\theta_{kt}}^2|\cdot$  using algorithm 3.2 for all  $k = 1, \dots, K$  and  $t = 1, \dots, T_k$ .
  - 10: Simulate  $\gamma_{\varepsilon_{kt}}$  from  $\gamma_{\varepsilon_{kt}}|\cdot$  using algorithm 3.3 for all  $k = 1, \dots, K$  and  $t = 1, \dots, T_k$ .
  - 11: Simulate  $\rho_{\theta_t}$  using a SVE procedure for all necessary correlation parameters.
-

## 4 Parameter recovery study

In this section we study the efficiency of our model and the proposed estimation algorithm concerning to parameter recovery. Our algorithm allows to consider different structured covariance matrices. Some examples can be seen in Table 1.

We considered a scenario with  $K = 2$  groups with  $T_1 = T_2 = 6$ , that is, the same number of time-points. Responses of  $n_{kt} = 500$  were simulated for  $k = 1, 2$  and  $t = 1, \dots, 6$  according to model (7) considering the ARH(1) matrix for the group 1 and the AD matrix for the group 2. We considered different growth curves for each group. For group 1 we considered a count curve with restriction  $\mu_{\theta_1}(1) = 0$  and parameters  $\alpha_{11} = \alpha_{12} = .1$ , see Table 3. Therefore the first time-point of group 1 is considered as the reference. For group 2, we choose a mitscherlich curve with parameters  $\alpha_{21} = 1$ ,  $\alpha_{22} = 2$  and  $\alpha_3 = .3$ , see Table 2. Figure 1 presents the simulated mean curve for each group. Population variances were fixed as  $\sigma_{\theta_1}^2 = (1.00, .77, .60, .57, .34, .40)'$  for group 1 and  $\sigma_{\theta_2}^2 = (1.10, 1.15, .80, .73, .70, .65)'$ . For marginal skewness coefficient we considered  $\gamma_{\theta_1} = (.80, .54, .10, .12, .17, -.02)'$  for group 1 and  $\gamma_{\theta_2} = (.85, .58, .19, .30, .24, -.08)'$  for group 2. For each group, the population means increase over time whereas the variances increase and then decrease. Also, we can see from Figure 1 that the population means of the groups are different in the first time-points and become equal along the study. For correlation parameters we have:  $\rho_{\theta_1} = .8$  for the group 1 and  $\rho_{\theta_2} = (.81, .89, .93, .73, .69)'$  for group 2.

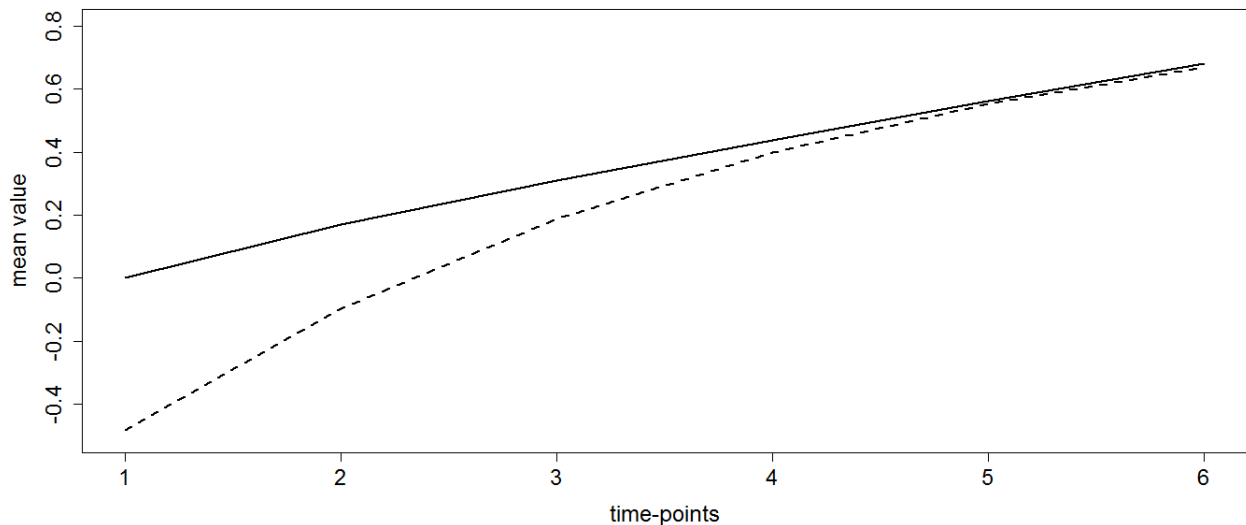


Figure 1: Underlying growth curves for the simulation study. Legend: Group 1 (solid line), Group 2 (dashed line)

The tests structure consists on a unique test with  $I = 25$  items applied for each group along all time-points. It means that, each subject is submitted to the same items in every occasions. This design is very common in psychiatric studies and resembles the real data analyzed in next section. Table 4 the values for the item parameters.

Table 5 presents the hyperparameters for the prior distributions. The prior distribution for the population mean and variance are concentrated around zero and one, respectively. For the skewness parameter, we are assuming more probability for values near zero but

Table 4: Item parameters for the recovery parameters study

Item	$a_i$	$b_i$	$c_i$	Item	$a_i$	$b_i$	$c_i$
1	1.410	-2.100	.200	13	.970	.360	.200
2	1.870	-1.900	.210	14	.810	.560	.210
3	1.470	-1.690	.220	15	1.580	.770	.220
4	1.380	-1.490	.230	16	1.190	.970	.230
5	.980	-1.280	.240	17	1.600	1.180	.240
6	1.160	-1.080	.250	18	1.860	1.380	.250
7	1.970	-.870	.200	19	1.690	1.590	.200
8	1.740	-.670	.210	20	1.700	1.790	.210
9	1.010	-.460	.220	21	.850	2.000	.220
10	1.120	-.260	.230	22	.810	2.200	.230
11	1.310	-.050	.240	23	1.280	2.410	.240
12	1.980	.150	.250	24	.910	2.610	.250
				25	1.140	2.820	.200

allowing reasonable probabilities for the others. The discrimination parameters are assumed to vary reasonably around a satisfactory discrimination power and for the difficulty parameter we assume a value above the mean of the reference time-point.

Table 5: The priors distributions hyperparameters

Hyperparameters					
$\mu_\zeta$	$\Psi_\zeta$	$(\mu_\alpha; \sigma_\alpha^2)$	$(a_\sigma, b_\sigma)$	$(\mu_\phi, \sigma_\phi^2)$	$(\mu_\rho, \sigma_\rho^2)$
(1, 0)	(.5, 16)	(0, 10)	(2.1, 1.1)	(0, 10)	(0, 10)

The usual tools to investigate the MCMC algorithms convergence, that is, trace plots, Gelman-Rubin's and Geweke's statistics were monitored. We generate three chains based on three different sets of starting values. The Gelman-Rubin's statistic were close to one for all parameters, indicating convergence. The trace plots and Geweke's monitoring indicate that a Burn-in of 10000 iterations was enough to reach the convergence. Further, the correlograms indicate that the samples composed by storing every 40th iteration have negligible autocorrelation. Therefore, we will work with valid samples with size 1000.

In order to assess the parameter recovery we consider the following statistics: correlations (Corr), mean of the bias (MBias), mean of the absolute bias (MABias) and mean of the absolute relative bias (MAVRB). Also, mean of the variances (MVAR) and mean of the root squared mean error (MRMSE), whose definitions can be seen bellow. Let  $\vartheta$  and  $\hat{\vartheta}$  a parameter and its estimate (posterior mean), respectively. The comparison statistics are defined as: Mcorr:  $\text{cor}(\vartheta_l, \hat{\vartheta}_l)$ , MBias:  $\frac{1}{n_p} \sum_{l=1}^{n_p} (\vartheta_l - \hat{\vartheta}_l)$ , MABias:  $\frac{1}{n_p} \sum_{l=1}^{n_p} |\vartheta_l - \hat{\vartheta}_l|$ , MAVRB:  $\frac{1}{n_p} \sum_{l=1}^{n_p} \frac{|\vartheta_l - \hat{\vartheta}_l|}{|\vartheta_l|}$ , MVAR:  $\frac{1}{n_p} \sum_{l=1}^{n_p} (\hat{\vartheta}_l - \vartheta_l)^2$  and MRMSE:  $\sqrt{\frac{1}{n_p} \sum_{l=1}^{n_p} (\text{MVAR} + (\vartheta_l - \vartheta)^2)}$ , with  $l = 1, \dots, n_p$  where  $n_p$  denotes the number of parameters.

Tables 6 and 7 present the population parameters estimates of the group 1 and 2, respectively. We can see that estimate were very close to the true values, and most of the 95% credibility intervals include the true parameters values. The estimates of some skewness and correlation parameters of the group 2 present a slight deviation of the true



values. It is certainly due a random fluctuation of the generated sample. Further, Table 8 and Figure 2 present the results for the latent traits and item parameters. In general the estimates presents small bias and small standard error, indicating that our algorithm recovered the parameters properly.

Table 6: Population parameters estimates of the Group 1

	True value	Mean	SD	CI(95%)
$\alpha_{11}$	.100	.105	.025	[.048, .147]
$\alpha_{12}$	.100	.078	.065	[-.041, .227]
$\sigma_{\theta_{11}}^2$	1.000	—	—	—
$\sigma_{\theta_{12}}^2$	.770	.708	.097	[.531, .909]
$\sigma_{\theta_{13}}^2$	.600	.601	.083	[.459, .787]
$\sigma_{\theta_{14}}^2$	.570	.572	.074	[.438, .728]
$\sigma_{\theta_{15}}^2$	.340	.354	.055	[.268, .483]
$\sigma_{\theta_{16}}^2$	.400	.383	.056	[.273, .496]
$\gamma_{\theta_{11}}$	.800	.738	.094	[.525, .895]
$\gamma_{\theta_{12}}$	.539	.483	.217	[.046, .864]
$\gamma_{\theta_{13}}$	.103	-.154	.202	[-.601, .175]
$\gamma_{\theta_{14}}$	.118	.008	.188	[-.405, .373]
$\gamma_{\theta_{15}}$	.168	.032	.196	[-.359, .438]
$\gamma_{\theta_{16}}$	-.022	-.043	.237	[-.594, .426]
$\rho_{\theta_1}$	.800	.791	.014	[.761, .815]

Table 7: Population parameters estimates of the Group 2

	True value	Mean	SD	CI(95%)		True value	Mean	SD	CI(95%)
$\alpha_{21}$	1.000	.908	.117	[.731, 1.212]	$\gamma_{\theta_{22}}$	.583	.033	.185	[-.318, .438]
$\alpha_{22}$	2.000	2.095	.122	[1.880, 2.373]	$\gamma_{\theta_{23}}$	.193	-.682	.148	[-.914, -.318]
$\alpha_3$	.300	.350	.047	[.247, .438]	$\gamma_{\theta_{24}}$	.298	.067	.187	[-.298, .483]
$\sigma_{\theta_{21}}^2$	1.100	1.296	.162	[1.022, 1.632]	$\gamma_{\theta_{25}}$	.241	.131	.200	[-.220, .590]
$\sigma_{\theta_{22}}^2$	1.150	1.318	.151	[1.059, 1.627]	$\gamma_{\theta_{26}}$	-.082	-.077	.194	[-.500, .306]
$\sigma_{\theta_{23}}^2$	.800	.814	.099	[.618, 1.019]	$\rho_{\theta_{21}}$	.810	.804	.011	[.779, .828]
$\sigma_{\theta_{24}}^2$	.730	.775	.093	[.629, .972]	$\rho_{\theta_{22}}$	.890	.857	.014	[.831, .881]
$\sigma_{\theta_{25}}^2$	.700	.716	.090	[.531, .864]	$\rho_{\theta_{23}}$	.930	.891	.022	[.845, .927]
$\sigma_{\theta_{26}}^2$	.650	.675	.062	[.555, .796]	$\rho_{\theta_{24}}$	.730	.782	.030	[.723, .838]
$\gamma_{\theta_{21}}$	.850	.435	.171	[.074, .744]	$\rho_{\theta_{25}}$	.690	.724	.052	[.622, .819]

Table 8: Results for the estimated latent traits and item parameters.

	Statistic				
	Corr	MBias	MABias	MVAR	MRMSE
Latent trait	.939	.024	.250	.107	.675
Discrimination	.785	.033	.193	.020	.551
Difficulty	.995	.046	.121	.014	.472
Guessing	.359	-.017	.027	.002	.241

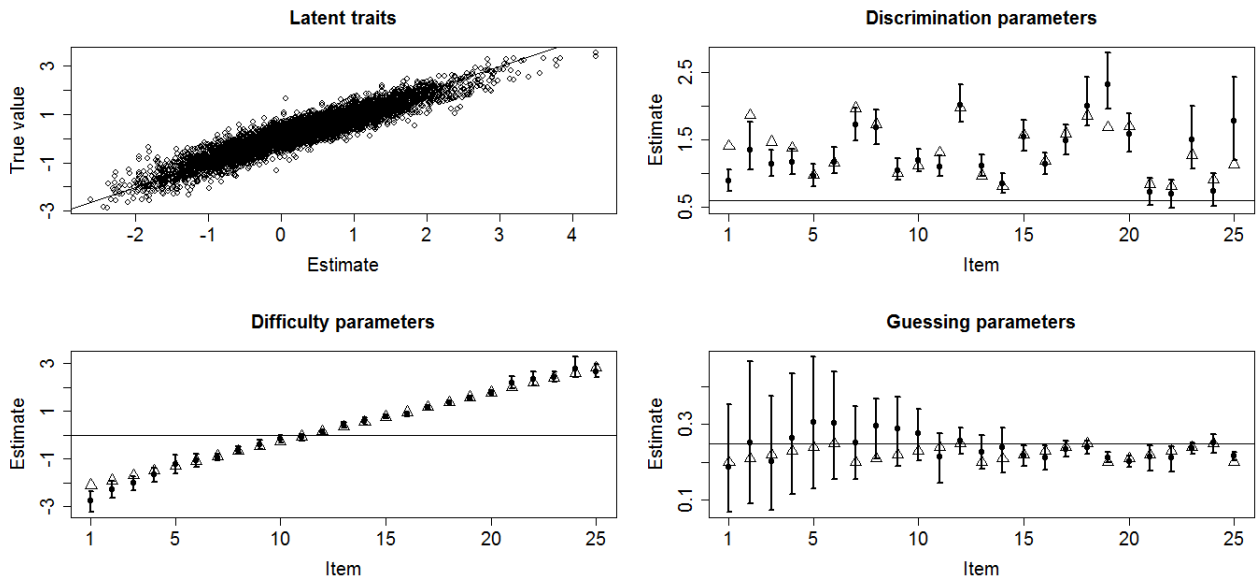


Figure 2: Estimates of latent traits and item parameters. Legend: circles denotes estimates, triangles denotes true values and vertical bars denote 95% credibility intervals

## 5 Real Data Analysis and Model fit Assessment

### 5.1 Model fit assessment tools

For model fit assessment we consider the so-called *Posterior Predictive Model Checking*, see [Sinharay \(2006\)](#) and [Sinharay et al. \(2006\)](#) for more details. The main idea is to compare the observed and simulated data, where the former is generated by using the posterior predictive distribution. Let  $\mathbf{y}^{obs}$  be the response matrix, and  $\mathbf{y}^{rep}$  be the replicated response matrix. Then, the *posterior predictive distribution* of replicated data at the time-point  $t$  is given by

$$p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs}) = \int p(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t)p(\boldsymbol{\vartheta}_t|\mathbf{y}_t^{obs})d\boldsymbol{\vartheta}_t, \quad (36)$$

where  $\boldsymbol{\vartheta}_t$  denotes the parameters at the time-point  $t$ . An usual method to compare the replicated and observed data, is to calculate the *Bayesian p-value* defined as

$$\mathbb{P}(D(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t) \geq D(\mathbf{y}_t^{obs}|\boldsymbol{\vartheta}_t)|\mathbf{y}_t^{obs}) = \int_{D(\mathbf{y}_t^{rep}) \geq D(\mathbf{y}_t^{obs})} p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs})d\mathbf{y}_t^{rep}, \quad (37)$$

where  $D$  denotes a suitable statistic defined to address some aspect of interest. In practice, if we have  $M$  draws from the posterior distribution  $p(\boldsymbol{\vartheta}_t|\mathbf{y}_t^{obs})$  of  $\boldsymbol{\vartheta}_t$  and  $M$  draws from the likelihood distribution  $p(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t)$ , the proportion of the  $M$  replications for which  $D(\mathbf{y}_t^{rep})$  exceeds  $D(\mathbf{y}_t^{obs})$  provides an estimate of the Bayesian  $p$ -value. Values close to 1, or 0, indicate model misfit.

For IRT models, [Béguin and Glas \(2001\)](#) have proposed a posterior predictive check to compare the observed score distribution with the posterior predictive score distribution. For the longitudinal IRT model, the observed score distribution can be evaluated per time-point. Specifically, to evaluate items fit we defined the following statistic:

$$D_i = \sum_l \frac{|P_{li}^O - P_{li}^E|}{P_{li}^E}, \quad (38)$$

where  $P_{li}^O$  and  $P_{li}^E$  denote, respectively, the observed and expected proportion of respondents with scores  $l$ , that scored correctly the item  $i$ , for all  $l = 1, 2, \dots, L$  and  $i = 1, 2, \dots, I$ , where  $L$  denotes the maximum score.

### 5.2 Model Comparison

For model comparison, where the main interest lies on the choice of the most appropriated covariance matrix, we used the approach of [Spiegelhalter et al. \(2002\)](#). The related statistics are *Deviance information criteria* (DIC), and the expected values of the *Akaike's information criteria* (EAIC) and *Bayesian information criteria* (EBIC). These statistics are based on the  $\rho_D$  statistics defined as  $\overline{D(\vartheta)} - D(\hat{\vartheta})$ . In our case we have,

$$D(\vartheta) = -2\text{Log}(L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta)P(\theta|\boldsymbol{\eta}_\theta)). \quad (39)$$

In practice, having  $M$  MCMC draws from the posterior distributions, the quantity  $\overline{D(\vartheta)}$  can be estimated as:

$$\overline{D(\vartheta)} = \frac{1}{M} \sum_{m=1}^M D(\vartheta^{(m)}), \quad (40)$$

and  $D(\bar{\vartheta})$  is evaluated on the estimates. Then, the estimates of the comparison statistics are give by

$$\begin{aligned}\widehat{\text{DIC}} &= D(\bar{\vartheta}) + 2\rho_D, \\ \widehat{\text{EAIC}} &= \overline{D(\vartheta)} + 2\rho_D, \\ \widehat{\text{EBIC}} &= \overline{D(\vartheta)} + 2\log(n \times I),\end{aligned}\tag{41}$$

where  $n$  and  $I$  are, respectively, the number of latent traits and the number of items.

### 5.3 Amsterdam Growth and Health Longitudinal Study

Data were analyzed from the AGHLS (Amsterdam Growth and Health Longitudinal Study), a longitudinal cohort study planned to monitor the growth, health and life-style of teenagers from secondary school (Kemper and van 't Hof, 1978). The AGHLS is focused on items concerning to the relationships between anthropometry, physical activity, cardiovascular disease risk, life-style, musculoskeletal health, psychological health and wellbeing. The sample considered corresponds to 452 participants who were followed over the period of 1990-2006 with a maximum of four assessment conditions (1991, 1993, 2000 and 2006) for each subject. The measurement years will be referred as 1, 2, 3 and 4, respectively. Two groups ( $k = 2$ ) were considered, male students and female students. The frequency of students in each time-point according to gender is presented in Table 9. A Dutch version of the Spielberger state trait anxiety inventory (STAI-DY) (Spielberger, 1989) was used to measure the trait "state anxiety", using a total of thirteen items with four response categories. High scores on this test are related to high anxiety.

Table 9: Number of subjects according to gender

Year	Female	Male
1991	112	70
1993	99	68
2000	222	185
2006	193	148

A explanatory analysis was conducted through the symmetric multiple group IRT model (Azevedo et al., 2016). This model assumes that subjects were nested in groups and latent traits are independently distributed over groups. Then we considered the measurement years as four groups. Table 10 presents the Pearson's correlations estimated for the complete pairs of estimated latent traits corresponding to years 1 to 4 according by gender. The results suggests that the latent traits are correlated over time-points. Figure 3 presents the estimated population means for each gender. The results indicate a difference among population means over the years and between groups, being the female's means higher than the male's ones.

Some multiple group longitudinal skew IRT models are compared by using the statistics of model comparison discussed before. Firstly, we considered the count growth curve for both groups and compare two correlation structures: AD and HT. These matrices seem to be the most appropriated ones, since the time-points are not equally spaced. We used the same correlation matrix for both groups, for simplicity. The results are presented in Table 11. The AD model outperforms the HT according to all statistics. Figures 4 and 5

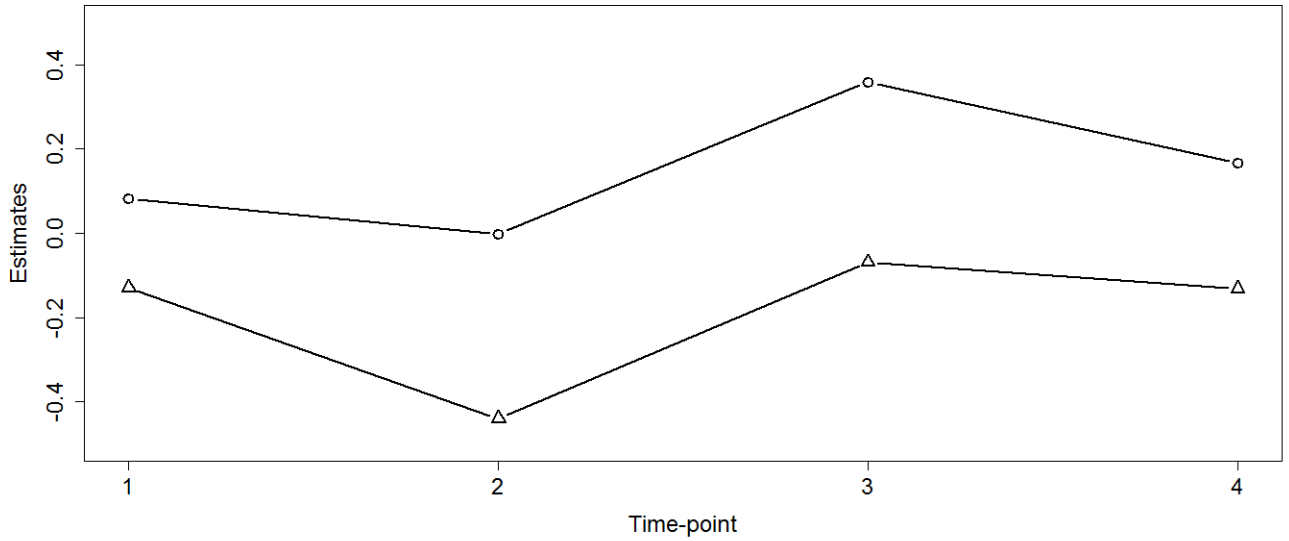


Figure 3: Estimates of the population means considering the Multiple Group Model.  
 Legend: Female (— o —), Male (— Δ —)

Table 10: Estimated correlation structure for female and male group

		1	2	3	4
Female	1	1.000	.565	.394	.546
	2	.565	1.000	.563	.619
	3	.394	.563	1.000	.743
	4	.546	.619	.743	1.000
Male	1	1.000	.594	.631	.436
	2	.594	1.000	.567	.415
	3	.631	.567	1.000	.714
	4	.436	.415	.714	1.000

present the estimated growth curves considering the AD matrix and the estimated means with 95% central credibility intervals. The results are in agreement with the explanatory analysis. Table 12 presents the results of the comparison of the curves. The count was selected according to all comparison statistics. Therefore, the selected model considers a count growth curve and an AD covariance matrix for the two groups.

Table 11: Statistics for comparison of the correlation structures

	DIC	EAIC	EBIC
AD	<b>12960.41</b>	<b>13795.53</b>	<b>20113.33</b>
HT	14423.47	15588.61	24403.07

Table 12: Statistics for comparison of the growth curves

	DIC	EAIC	EBIC
Count	<b>12960.41</b>	<b>13795.53</b>	<b>20113.33</b>
Mit.	13162.11	14293.39	22851.69
Jenns	13230.42	14334.07	22683.37

Figures 6 and 7 present the observed and predicted scores with 95% credibility intervals. We can see that most of the observed scores distribution are well within the intervals. Some discrepancy is observed considering the Year 2000 of the male group. However, in general the results are indicating that the model is well fitted. Figures 8 and 9 present smoothed histograms of the latent trait estimates according to the skew AD model, with theoretical curves. The plot of the theoretical curves of the skew AD model was based on latent traits simulated via model 7. The estimated latent trait distributions are heavy-tailed and multimodal in some cases. This behavior is probably due to the few number of observations available. However, our model provided a good description of the latent trait distributions behavior. The population parameters estimates can be seen in Tables 13 and 14. These Tables also present the estimates of  $\mathbb{P}(\gamma_{\theta_{kt}} \notin [-.13, .13])$ , denoted by  $\hat{p}_\gamma$ . Values of the skewness coefficient in the interval  $[-.13, .13]$ , indicate symmetric latent trait distribution. We considered that the asymmetry coefficient of the latent traits distribution is not null when we had more than 60% of the sampled asymmetry coefficient values outside the interval  $[-.13, .13]$ . Therefore, only time-points 1 and 3 of the female group can be consider asymmetric.

The correlation parameters estimates were high and statistically significant. Figure 10 presents the differences between the group means over the measurement years. We can see that mean of the groups are statistically different, except in time-point 1. Figure 11 presents the Bayesian  $p$ -values for the item parameters. The results indicate that all items are well fitted. The estimates of the item parameters are presented in Figure 12. In general items presented a good discrimination (estimates greater than .6) and high difficulty. The guessing parameter estimates were low and concentrated around .05, indicating that the two parameters probit model could be preferable.

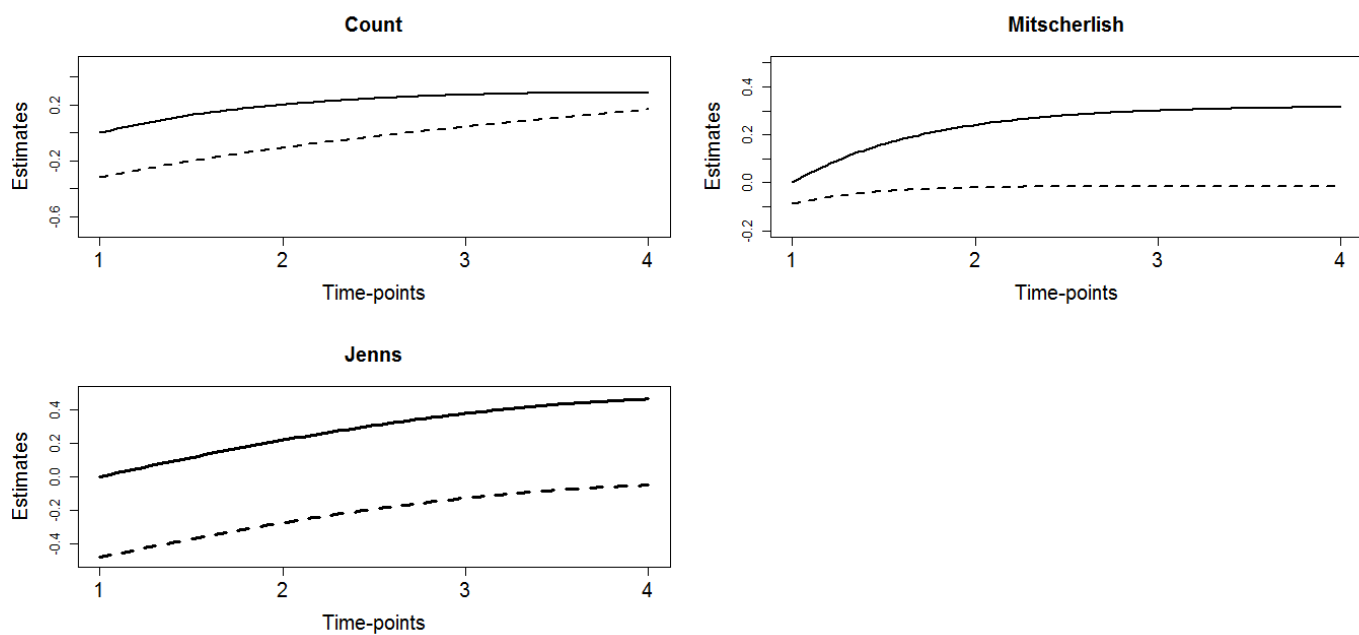


Figure 4: Estimated growth curves. Legend: Female (solid line), Male (dashed line)

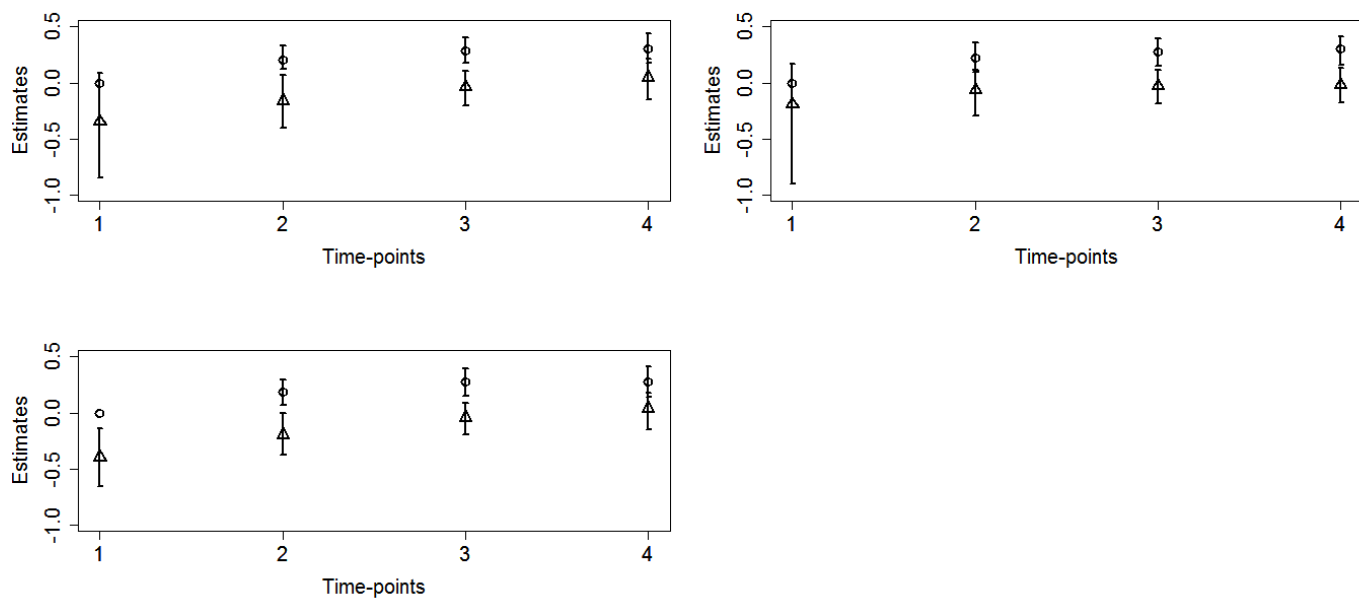


Figure 5: Estimated means with 95% central credibility intervals. Legend: Female (—○—), Male (—△—).

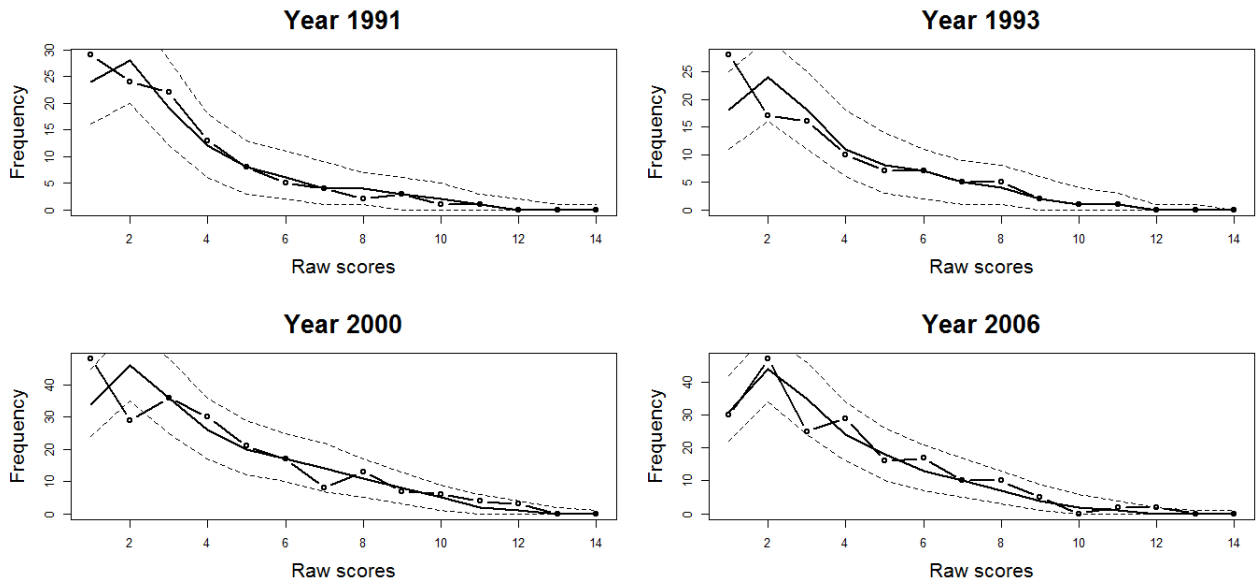


Figure 6: Observed and predicted scores distributions with 95% credibility intervals for female students

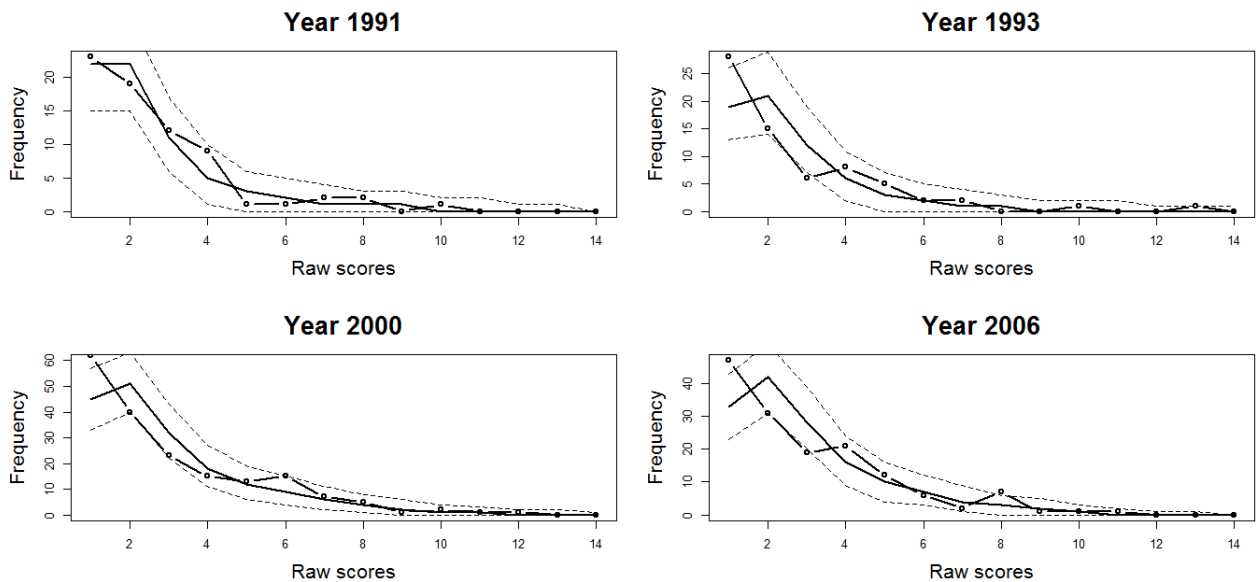


Figure 7: Observed and predicted scores distributions with 95% credibility intervals for male students



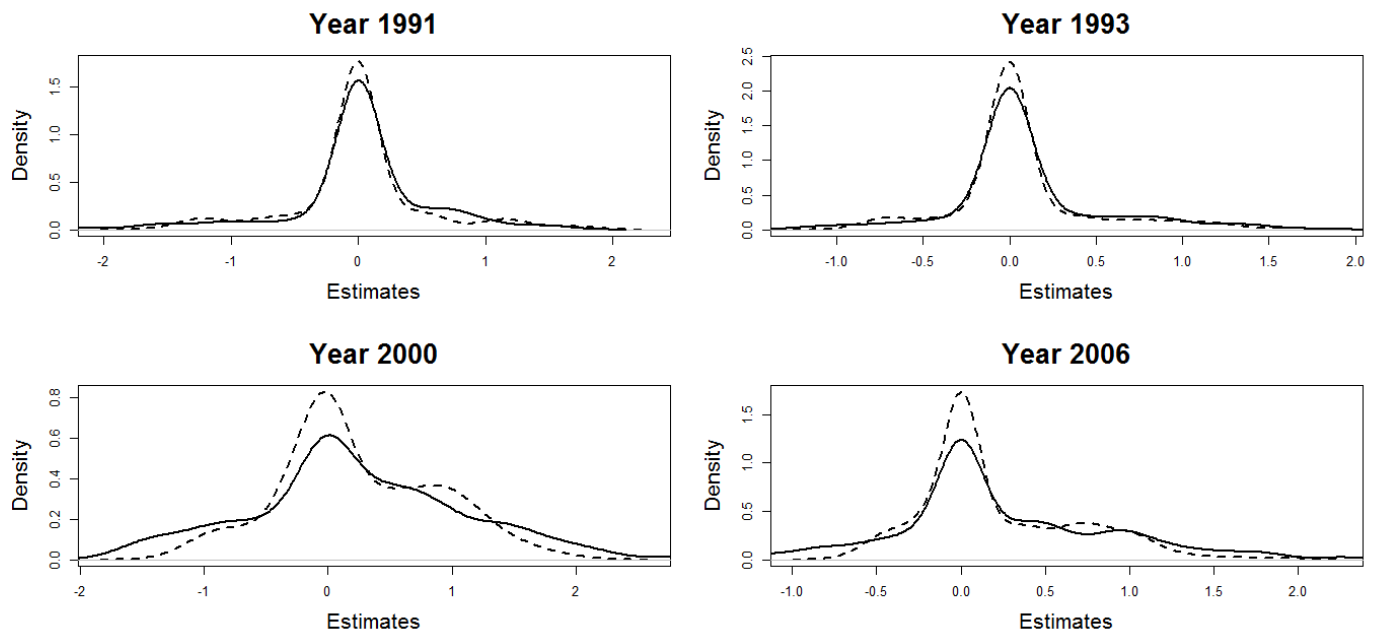


Figure 8: Estimated latent trait distributions for female students. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

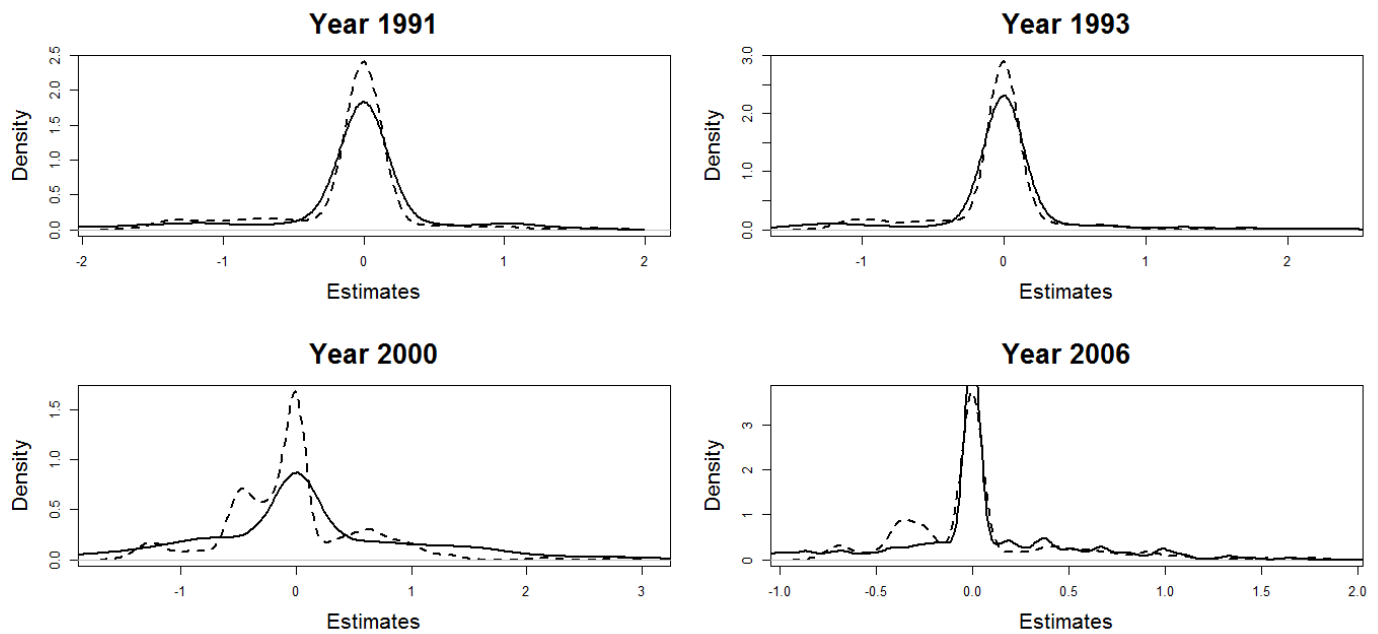


Figure 9: Estimated latent trait distributions for male students. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

Table 13: Estimates of the population parameters for the female students

	Mean	SD	CI(95%)	$\hat{p}_\gamma$
$\alpha_{\theta_{11}}$	-.126	.096	[-.352, .014]	–
$\alpha_{\theta_{21}}$	.490	.206	[.201, .990]	–
$\sigma_{\theta_{11}}^2$	1.000	–	–	–
$\sigma_{\theta_{21}}^2$	.518	.189	[.267, .990]	–
$\sigma_{\theta_{31}}^2$	1.016	.300	[.563, 1.704]	–
$\sigma_{\theta_{41}}^2$	.480	.136	[.266, .768]	–
$\gamma_{\theta_{11}}$	-.433	.205	[-.771, -.007]	.911
$\gamma_{\theta_{21}}$	.062	.267	[-.455, .645]	.554
$\gamma_{\theta_{31}}$	-.007	.215	[-.496, .438]	.446
$\gamma_{\theta_{41}}$	.168	.253	[-.297, .701]	.619
$\rho_{\theta_{11}}$	.776	.037	[.694, .839]	–
$\rho_{\theta_{21}}$	.834	.052	[.714, .924]	–
$\rho_{\theta_{31}}$	.853	.086	[.659, .983]	–

Table 14: Estimates of the population parameters for the male students

	Mean	SD	CI(95%)	$\hat{p}_\gamma$
$\alpha_{\theta_{12}}$	-.370	.231	[-.815, .135]	–
$\alpha_{\theta_{22}}$	.297	.386	[-.224, 1.296]	–
$\alpha_{\theta_3}$	.001	.144	[-.307, .174]	–
$\sigma_{\theta_{12}}^2$	.997	.534	[.278, 2.354]	–
$\sigma_{\theta_{22}}^2$	.730	.425	[.229, 1.760]	–
$\sigma_{\theta_{32}}^2$	1.327	.471	[.713, 2.476]	–
$\sigma_{\theta_{42}}^2$	.540	.274	[.236, 1.181]	–
$\gamma_{\theta_{12}}$	-.057	.272	[-.626, .508]	.576
$\gamma_{\theta_{22}}$	-.001	.265	[-.572, .558]	.531
$\gamma_{\theta_{32}}$	.019	.237	[-.446, .524]	.504
$\gamma_{\theta_{42}}$	-.108	.264	[-.660, .412]	.557
$\rho_{\theta_{12}}$	.842	.041	[.731, .890]	–
$\rho_{\theta_{22}}$	.794	.062	[.655, .877]	–
$\rho_{\theta_{32}}$	.791	.112	[.594, .971]	–

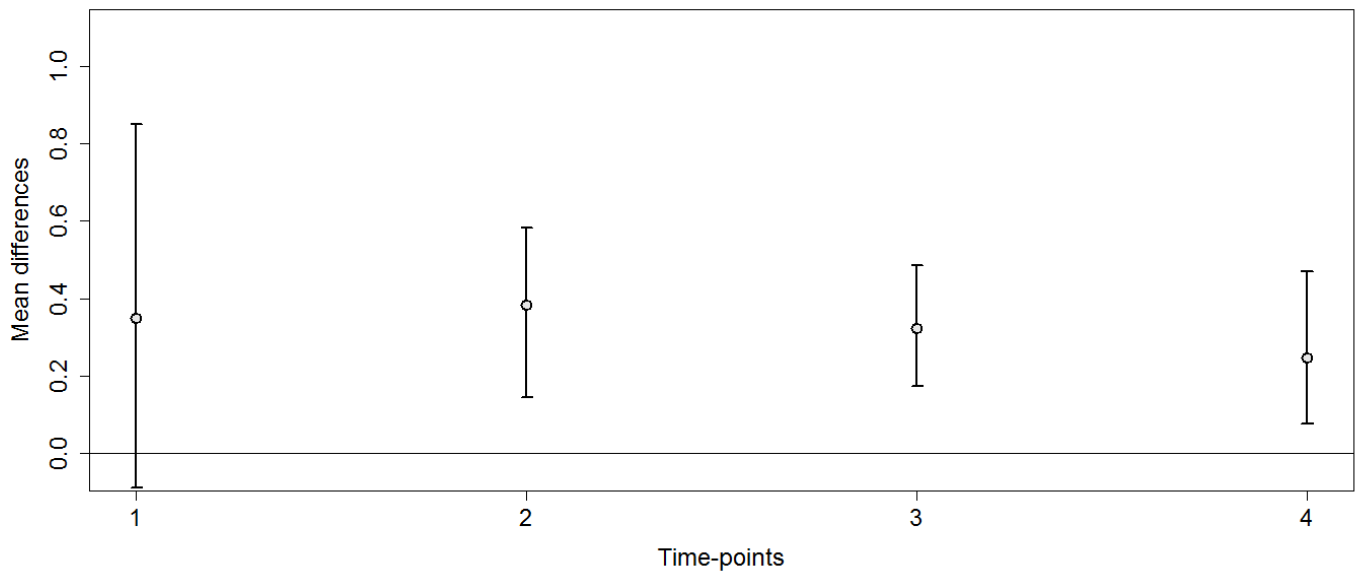


Figure 10: Estimated difference between the group means along the years. Legend: Estimate (circles), 95% credibility intervals (vertical bars)

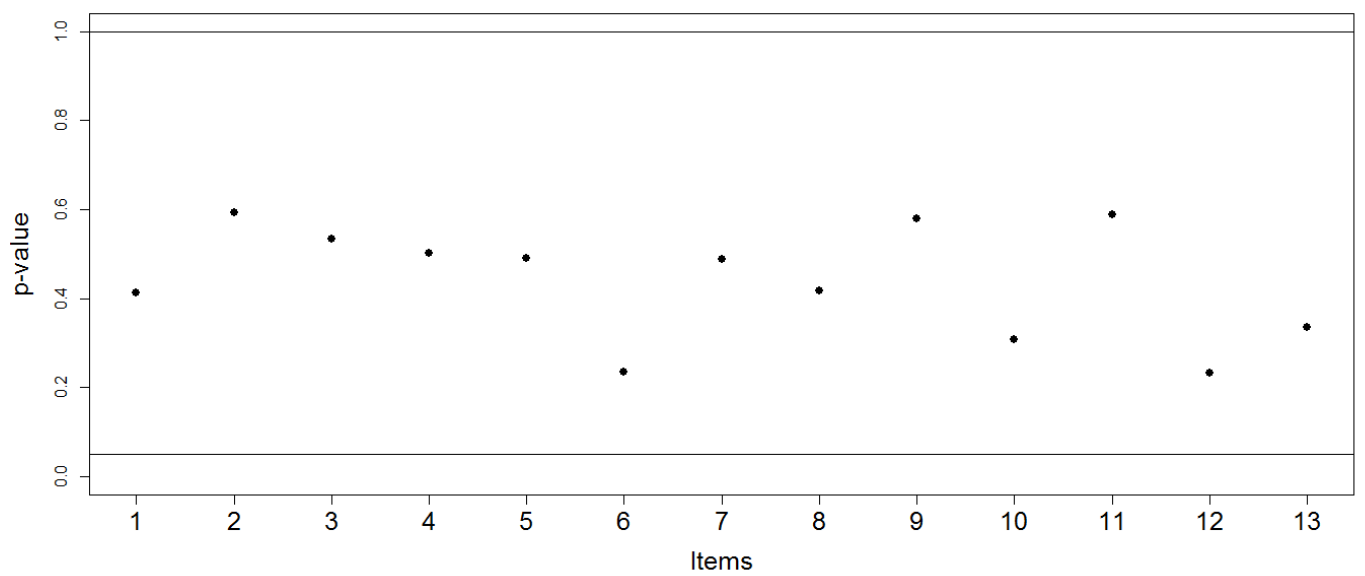


Figure 11: Bayesian  $p$ -values for item parameters

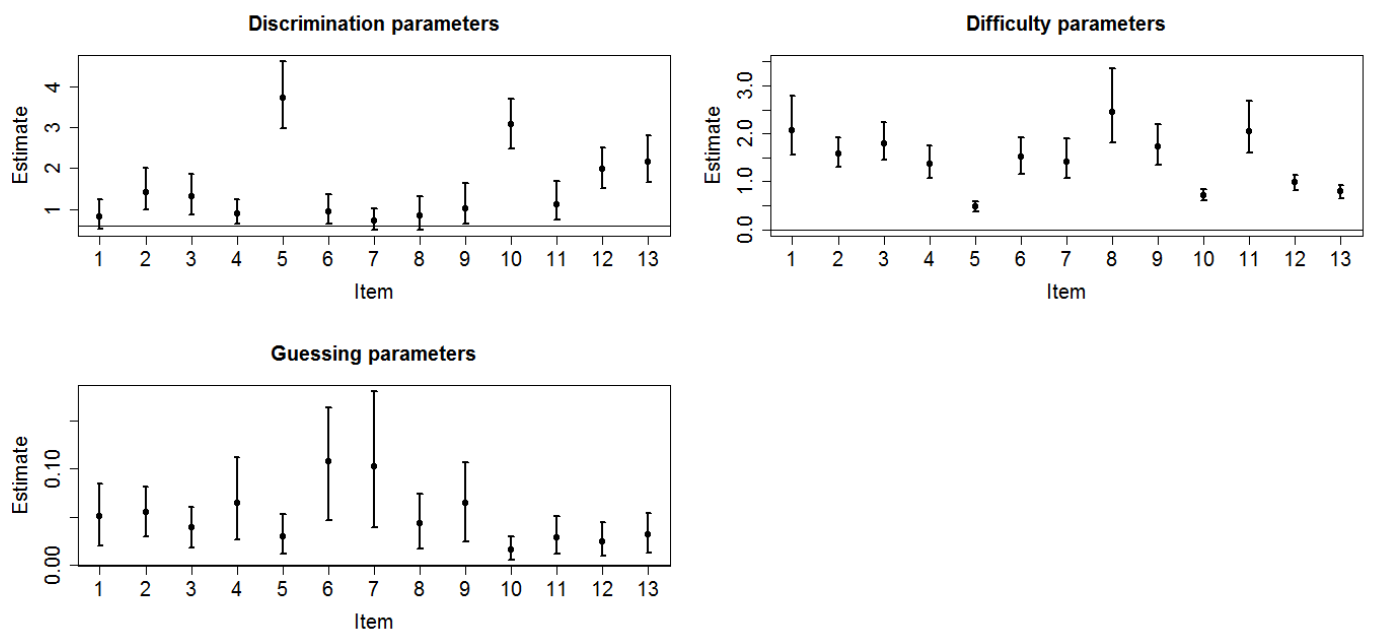


Figure 12: Item parameters estimates. Legend: estimates (circles), 95% credibility intervals (vertical bars)

## 6 Concluding Remarks

We introduced a skew multiple group longitudinal IRT model based on a general Cholesky decomposition procedure, with growth curve modeling for population means. The developed MCMC algorithm can handle model identification issues, scaling process, nonresponses and unbalanced data. Our modeling also allows different covariance latent trait patterns across groups and different skew-normal latent distributions over the time-points. Furthermore, our MCMC algorithm presented a good parameter recovery according to the simulated study. In addition, a real data concerning to a growth and health study with teenagers participants was analyzed. This data were also analyzed by [Azevedo et al. \(2015\)](#) with a symmetric multiple group longitudinal IRT model. Differently from these author, we considered the unbalanced design and identified significant difference between groups in terms of the population means. We also found a significant asymmetry of the marginal latent trait distribution in most of cases. In fact, the small number of observations in each cell (combination of group and time-point) become difficult the study of the latent trait distributions. In general, our model fitted the data properly according to some specific model fit assessment tools. In conclusion, our approach revealed to be a promising alternative to the usual ones in analyzing multiple group longitudinal IRT data. In future research we intend to explore more growth curves and others regression structures for the mean of the latent trait distributions, including covariates. Other response models to allow ordinal, nominal or gradual responses could also be considered.

## References

- Azevedo, C., Bolfarine, H., and Andrade, D. (2011). Bayesian inference for a skew-normal irt model under the centred parameterization. *Computational Statistics & Data Analysis*, 55:353–365. [2](#), [10](#)
- Azevedo, C. L., Fox, J.-P., and Andrade, D. F. (2015). Longitudinal multiple-group irt modelling: covariance pattern selection using mcmc and rjmc. *International Journal of Quantitative Research in Education*, 2(3-4):213–243. [1](#), [2](#), [29](#)
- Azevedo, C. L. N., Fox, J.-P., and Andrade, D. F. (2016). Bayesian longitudinal item response modeling with restricted covariance pattern structures. *Statistics and Computing*, 26(1):443–460. [2](#), [3](#), [20](#)
- Azzalani, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 12:171–178. [2](#), [5](#)
- Baker, F. B. and Kim, S.-H. (2004). *Item response theory : Parameter Estimation Techniques*. Marcel Dekker, Inc, New York, NY, second edition. [3](#)
- Béguin, A. A. and Glas, C. A. W. (2001). Mcmc estimation and some model-fit analysis of multidimensional irt models. *Psychometrika*, 66(4):541–561. [6](#), [19](#)
- Bock, D. R. and Zimowski, M. F. (1997). *The Multiple Groups IRT, in Handbook of modern Item Response Theory*. Wim J. van der Linden and Ronald K. Hambleton eds, Springer-Verlag, New York, NY, first edition. [2](#)
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo : Stochastic simulation for bayesian inference*. Chapman & Hall/CRC, London, second edition. [11](#)
- Kemper, H. C. G. and van 't Hof, M. A. (1978). Design of a multiple longitudinal study of growth and health in teenagers. *European Journal of Pediatrics*, 129(3):147–155. [20](#)
- Kolen, M. J. and Brennan, R. L. (2004). *Teste Equating - Methods and Practices*. Springer-Verlag, New York, second edition. [2](#)
- Lindsey, J. K. (1993). *Models for Repeated Measurements*. Oxford University Press, 2nd edition. [6](#)
- Marsman, M. (2014). *Plausible Values in Statistical Inference*. PhD Thesis University of Twente, Enschede., 1 edition. [12](#)
- Newton, H. (1988). *TIMESLAB: A time series analysis laboratory*. Pacific Grove, CA: Wadsworth & Brooks/Cole., 1 edition. [4](#)
- Nunez-Anton, V. and Zimmerman, D. L. (2000). Modelinng nonstationary longitudinal data. *Biometrics*, 56(3):699–705. [2](#)
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690. [1](#), [2](#), [3](#)
- Rodriguez, C. L. B. (2005). *Bayesian Inference on the Skew-normal Model*. Master dissertation (In portugues), São Paulo, SP, ime-usp edition. [11](#)

- Santos, J., Azevedo, C., and Fox, J. (2017). A general cholesky decomposition based modeling of longitudinal irt data: Handling skewed latent traits distributions. Thecnical report. [11](#)
- Santos, J. R. S., Azevedo, C. L. N., and Bolfarine, H. (2013). A multiple group item response theory model with centred skew normal latent trait distributions under a bayesian framework. *Journal of Applied Statistics*, 40(10):2129–2149. [2](#)
- Sinharay, S. (2006). A bayesian item fit analysis for unidimensional item response theory models. *British journal of mathematical and statistical psychology*, 59:429–449. [19](#)
- Sinharay, S., Johnson, M. S., and Stern, H. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, 30(4):298–321. [19](#)
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linden, A. (2002). Bayesian measures of model complexity and fit. *Journal Royal Statistical Society*, 64(3):583–639. [19](#)
- Spielberger, C. (1989). *State-Trait Anxiety Inventory: Bibliography*. Palo Alto, CA: Consulting Psychologists Press, 2 edition. [20](#)
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of American Statistical Association*, 82(398):5258–550. [6](#)
- Zimmerman, D. and Núñez-Antón, V. (2009). *Antedependence Models for Longitudinal Data*. Chapman and Hall/CRC, 1 edition. [3](#)